Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

TECHNISCHE UNIVERSITÄT WIEN

MAX-PLANCK-GESELLSCHAFT

DIPLOMARBEIT

Simulation and analysis of emerging structures in feasting *Escherichia coli* colonies

zur Erlangung des akademischen Grades

Diplom-Ingenieur/in

im Rahmen des Studiums

Technische Physik

eingereicht von

Sebastian Mair, B.Sc.
Matrikelnummer 1025731

ausgeführt am Institut für Theoretische Physik
der Fakultät für Physik der Technischen Universität Wien
(in Zusammenarbeit mit dem Max-Planck-Institut für Dynamik und Selbstorganisation, Göttingen)

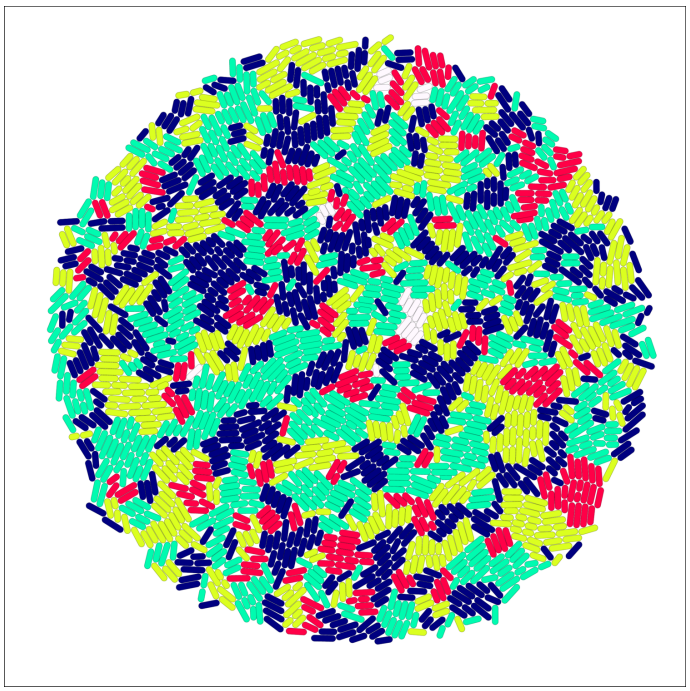
Betreuung
Betreuer: Marco G. Mazza, PhD
Betreuer: Ao. Univ .Prof. DI Dr. Gerhard Kahl

Wien, 10.09.2018

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

Simulation and analysis of emerging structures in feasting *Escherichia coli* colonies

Sebastian Mair, BSc.



Supervisors:
Marco G. Mazza
Gerhard Kahl

A thesis presented for the degree of
Master of Science

MAX-PLANCK-GESELLSCHAFT
Max-Planck-Institut für Dynamik
und Selbstorganisation
Nonequilibrium Soft Matter

Technische Universität Wien

Soft Matter and Complex Systems

Abstract

We study a theoretical model the growth of sessile *Escherichia coli* colonies in the exponential growth phase (*feasting*), by performing Discrete Element Simulations in two dimensions. We find that mechanical interactions are sufficient for the formation of highly ordered mesoscopic structures, *vulgo* microdomains. Basic tools of analysis, such as the contact angle distribution or the radial distribution function do not indicate the formation of these microdomains. For this purpose, we employ a community detection algorithm on contact networks representing the colonies. We compare three different variants with a range of threshold angles, and evaluate their overall performance and correlation with more conventional measures. We find that a good threshold angle to discriminate "cohesive" from "repulsive" contacts between particles i and j, to be $c_t = |\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j| \approx 0.96$. For high threshold angles, all three variants performed comparably well, whereas the naive dichotomization variant was clearly outperformed for lower threshold angles. We believe that this method opens new avenues to study morphogenesis and can be equally beneficially applied in related fields such as systems of anisometric particles.

Declaration

I hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award. Where other sources of information have been used, they have been acknowledged.

Dedication

Für meine Eltern Gernot und Rosalinde die mich immer aus vollem Herzen und bedingungslos unterstützt haben.

Acknowledgements

I want to thank my advisor Marco for his valuable inputs, his trust and patience. I want to thank my dear colleagues at the MPI Arne, Chamkor, Fabian, Hannes, Jan, Jana, Jeremy, Johannes, Rebekka and Tine, for their helpful discussions and especially for the joyful times we spent. Also, I want to thank my flatmates and friends Laura, Nina and Till, who were the most important people in Göttingen to me, my Vienna flatmates and friends Lisanne, Magdalena and Nora who I was missing while being away and Martina who I happily met after coming back.

Special thanks are given to my Vienna study group and friends Marko, Martin, Stefan and Vimal, without whom I would have never gotten so far.

Finally, I want to thank my dear friends Angelika, Silke, Sylwia who will always be important to me, no matter how far we are apart, my brothers Johannes, Andreas and Wolfgang, and of course Sanjay and Suchart who are like brothers to me.

Contents

List of Symbols **xii**

1 Introduction **1**
1.1 Basic physiology of *E. coli* . . . 2
1.1.1 Planctonic state . . . 4
1.1.2 Surface colonialization and growth . . . 5
1.1.3 Biofilm formation . . . 6
1.2 Morphogenesis . . . 8
1.2.1 Significance of mechanical interactions . . . 9

2 Model **10**
2.1 Particle model . . . 10
2.2 Pair-wise interaction . . . 11
2.3 Reduced units and parametrization . . . 12

3 Methods **14**
3.1 Discrete Element Simulations . . . 14
3.1.1 Force calculation . . . 15
3.2 Simulation setup . . . 17

4 Microdomains **18**
4.1 Community detection . . . 19
4.2 Graph representation . . . 19
4.3 Community similarity . . . 21
4.3.1 Normalized mutual information . . . 22
4.3.2 Consensus matrix . . . 22
4.4 Validation of communities . . . 23
4.4.1 Particle neighborhood . . . 24
4.4.2 Elbow method . . . 24
4.4.3 Clustering index . . . 25
4.4.4 Feature size matching . . . 26
4.5 Community detection . . . 28

4.5.1 Visualization 28

5 Results **29**
5.1 Comparison with Previous results 29
5.1.1 Particle growth 29
5.1.2 Local packing fraction 29
5.1.3 Radial nematic order structure 32
5.2 Local structure 34
5.2.1 Particle contacts 34
5.2.2 Pair-correlation function 34
5.3 Microdomains 37
5.3.1 Similarity of partitions 37
5.4 Domain quality 41
5.5 Clustering index 41
5.6 Domain sizes 42
5.7 Experimental data 46

6 Conclusion **48**

A Appendix **50**
A.1 Modularity for mutli-layer networks 50
A.2 Logistic sigmoidal function 50
A.3 Time evolution of the modularity Q 51
A.4 Correlation of equal-time polar-pair correlation C_2 and the average domain number-size N_σ 51

List of Figures

1.1 (a) Light-microscopy picture of an *E. coli* biofilm, taken from Ref. [1]. (b) Sketch of the architecture of a single *E. coli* bacterium, from Microbiology Notes [2]. (c) Sketch of the physiological life-cycle of *E. coli*, inspired by Ref. [3] 3

1.2 Architecture of a small *E. coli* colony. Light-microscopy image taken from Ref. [4]. 6

1.3 Structure of (a,c) wild type and (b,d) EPS-defective *E. coli* biofilms, grown for 72h under similar conditions. In contrast to the EPS-defective strain, the wild type *E. coli* form a three dimensional biofilm. Taken from [5]. 7

1.4 Emerging nematic order in a geometrically confined, growing *E. coli* colony. Subfigures (A-C) show the same colony at subsequent points in time. Taken from [6]. 9

2.1 Sketch of two interacting particles: illustration of forces, torques, particle properties and geometry. 11

3.1 At regular intervals, particles are assigned to bins on a fixed grid. The pair-wise interactions are only computed for particles that reside in bins that are in proximity to one another. The red dot in the sketch indicates the center of mass of a particle i, the dashed lines indicate the bin-boundaries, and the red solid line marks the outline of the considered bins. The interaction potential between the central particle (with the red dot) and the particles whose center of mass lies outside the region, is not calculated. 16

4.1 Illustration of the proposed contact network mappings. The red and yellow dots indicate a particle's node representation and the blue lines the edges that two particles at contact share, where the saturation indicates the contact angle depended edge-weight. The deficits of dichotomization (Heaviside-mapping (1)) compared to the other mappings (2&3) become evident, as the resulting network topology sensitively depends on the orientation of the particle marked with the red dot. In contrast, the mappings (2&3) preserve the topology independently of the orientation of any individual particle. This yields a more robust and intuitive notion of microdomains. Here, possible partitionings into microdomains are indicated by the blue and red coloring of the particles. 21

4.2 The NMI for two at-first-identical partitions is calculated, where one partition is continuously modified from the original form. In panel (a) $\mathcal{S} = \mathcal{P} = (0, 1, \cdots, n)$ is calculated, where $n = 100$ is the total number of particles and the first l particles in partition $\mathcal{S}$ are reassigned to a new community σ_i. In (b) $\mathcal{S} = \mathcal{P} = (0, 0, \cdots 0, 1, 1, \cdots, 1)$ with the first half of particles assigned to community $\sigma_i = 0$ and the second half to $\sigma_i = 1$, these two assignment vectors are shifted with respect to each other. 23

4.3 Sketch to illustrate the smallest enclosing circle, with radius R_σ around a group of particles. 27

5.1 Growth of an unconstrained colony of particles with an average splitting length of $\langle l_{split} \rangle = 4$. Coloring indicates the orientation of the particles. 30

5.2 The relative difference between the observed number of particles in the colony and a continuous exponential count $((N - N_{continuous})/N_{continuous})$ oscillates due to synchronized growth, and it decreases in magnitude over time. 31

5.3 Local packing-fraction of colonies of particles with an average splitting length of $\langle l_{split} \rangle = 4$, at different times. 32

5.4 Radial nematic order structure of colonies of particles with an average splitting length of $\langle l_{split} \rangle = 4$. Particles in the bulk orient randomly $S_r^{random} = \frac{2}{\pi} \approx 0.637$, whereas particles at the colony edge preferentially orient perpendicular to the vector pointing to the colony center $S_r^{edge} < S_r^{random}$. 33

5.5 Time resolved distribution of (a) the contact number N_c, and (b) the contact angle $|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_i|$ in the bulk of colonies of particles with an average splitting length of $\langle l_{split} \rangle = 4$. 35

5.6 Time resolved radial distribution function $g(r)$ (a) for an entire time-range and (b) at a single point in time. 36
5.7 Partitioning of a colony of particles with $\langle l_{split} \rangle = 4$, in (a,c,e) generation 9 and (b,d,f) generation 11, with different threshold-values . . 38
5.8 Boxplots showing the normalized mutual information of partitions of colonies in generation 9, obtained using the network mapping in the column-header, and (a) the Heaviside-, (b) the signum-, (c) the linear-network mapping. These are all the possible combinations, as the NMI is symmetric and the Heaviside domain detection is deterministic and will always lead to the same partitioning, i.e. $NMI = 1$, i.e. there are a total of 5 distinct combinations convey information. . 39
5.9 Consensus matrix obtained from (a) Heaviside-, (b) signum- and (c) linear network-mappings of colony snapshots at generation 7, generation 9, generation 11, with a threshold angle of $c_t = 0.81$. 40
5.10 The averaged scalar nematic order for partitions obtained from linear network mappings of colonies in generation 9. The error-bars indicate the standard deviation. 41
5.11 The silhouette index for partitions of colonies in generation 9 shows a maximum for the linear- and the signum-mapping at $c_t = 0.96$. . . . 42
5.12 Equal-time polar-pair function at sequential times. The red dots indicate the first zero-crossing of the function that serves as an estimate for the size of the nematic features in the colony. The red line is the best linear function fit through the zero-crossing points. 43
5.13 Average domain number-size and radius of smallest domain-enclosing circle for domains of obtained from (a) Heaviside-, (b) signum- and (c) linear-network mappings. 44
5.14 Correlation matrix, showing the Pearson correlation coefficient r of ensemble-averaged time series of the equal-time polar-pair correlation C_2 and the radius of the smallest domain-enclosing circle r_σ. 45
5.15 (a-c) Extraction of *E. coli* coordinates from light microscopy image [4], and (d) the probability density functions for community sizes in experimentally observed and simulated colonies, at a threshold angle of $c_t = 0.96$ and linear network mapping. 46
5.16 Comparison of microdomain size distributions in simulation and experiment for a range of threshold angles c_t and linear network mapping, at two different colony generations, (a) generation 7-8 and (b) generation 8-9. 47

A.1 The modularity Q of partitions obtained from a linear network mapping (a) for a single time-stamp (b) for a sequence of time-stamps. . . 52

A.2 Correlation matrix, showing the Pearson correlation coefficient r of ensemble-averaged time series of the equal-time polar-pair correlation C_2 and the domain number-size N_σ. 53

List of Tables

1.1 Composition ranges of biofilm matrices, from [7]. 7

List of Symbols

$\mathbf{x}$	particle position
$\mathbf{u}$	particle orientation
l	particle length
τ	time increment
α	growth rate
δ_{ij}	distance at closest approach of particles i and j
α_{ij}	enclosing angle of particles i and j
δ_t	threshold distance of closest approach, for particles at contact
c_t	threshold contact angle
$\mathcal{H}$	Heaviside function
δ	Kronecker-Delta function
Q	quality function, modularity function
A_{ij}	adjacency matrix
p_{ij}	null-model
J_{ij}	adjacency matrix minus null-model
$\mathcal{S}$, $\mathcal{P}$	graph partitions
Σ, Π	sets of community assignments
σ_k	community assignment for particle k
N_c	contact number
η	local packing-fraction
S_r	radial nematic order structure
$g(r)$	radial distribution function
S	scalar nematic order parameter
$\mathcal{C}$	silhouette index
C_2	equal-time polar-pair correlation function
r_σ	radius of smallest enclosing circle
r	pearson correlation

Chapter 1

Introduction

Microbes were first observed by Antonie van Leeuwenhoek, who found these "animalcules" in the plaque of his own teeth in 1676. Numerous species have been observed displaying a vast range of interesting properties, still only a tiny fraction of all microorganism-species have been classified. Many questions of basic and applied sciences are immediately related to understanding the individual and collective properties of microbial forms of life. They are the simplest and the oldest living organisms on our planet. Also, due to their simplicity, they are perfectly suited to study evolutionary processes and to use them as model for active matter systems.

In the last few years, a multitude of phenomena of the motile microorganisms have attracted great attention in the physics community, and a lot of work has been done to understand this form of active matter. In contrast, much less work has been devoted to understanding the fundamental physics governing growing sessile colonies [8].

Sessile microbial colonies might provide insights into the mechanisms to create complex, highly specialized structures, such as biofilms or human tissues. Understanding microbial colony growth could help to address the standing questions of emerging anisotropy in embryonic development [9] and give way to novel applications.

We want to contribute with this work by performing discrete element simulations of *Escherichia coli* colonies in the exponential growth phase (*feasting*). We choose *E. coli*, because they take a special position in microbial research. They have been studied for over 150 years; their genome was fully sequenced and they serve as a 'workhorse' for numerous laboratories around the world [10].

Our simulation model only incorporates mechanical interactions, which contrasts traditional microbial research, which focusses primarily on chemical processes. In fact, only recently the significance of mechanical interactions has been recognized [6, 9, 11].

In accordance with recent results published by You *et al.* [12], we find that mechanical interactions suffice to form mesoscale structures, *vulgo* microdomains. These mesoscale structures might *in vivo* prepattern tissues and as such significantly effect their development and final structures. We have devised a novel method to identify microdomains, that is based on community detection, and studied their growth behavior.
After completing this thesis, we found that the problem formulation and objective is closely related to image segmentation tasks [13], while this late insight does not effect the outcomes of this thesis, it might well inform the choice of methods in future studies.

In the following section, we want to recap some results of *E. coli* research, putting the present work into perspective of the current state of knowledge and providing the reader new to the field with an overview to the basic processes and phenomena. In Chapter 2 we will introduce the particle-based model for our numerical simulations. Chapter 3 is devoted to the simulation methods employed. Chapter 4 introduces the notion of microdomains and motivates a networked approach to analyze these. Moreover, we will outline pathways to evaluate identified micro-domains. In Chapter 5 we discuss the results, of our simulations and their significance in relation to prior investigations. In Chapter 6 we conclude the work with a summary and indicate possible questions for further research.

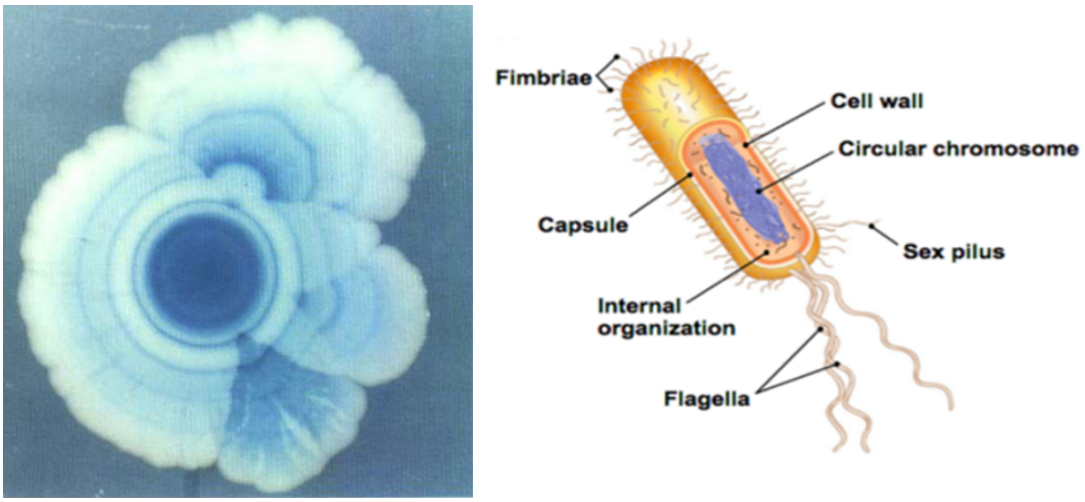
In the following we will describe basic properties of *E. coli*; this will aid the layout of our model in Chapter 2, as well as indicate possible directions for future studies.

1.1 Basic physiology of *E. coli*

Bacterial cells commonly exhibit one of the three basic shapes: *bacilli*, *cocci* and *spirilla*, that refer to rod-like, spherical and curved shapes, respectively. Some species aggregate these basic cell architectures to form more complex multicellular consortia, see Figure 1.1a. *E. coli* is a *bacillus*, i.e. single cell rod-shaped bacteria. The sizes and aspect ratios depend on the specific strain, the physiological state and the experienced environmental conditions. Governed by the same influences, *E. coli* grow a range of functional appendages, that serve to perform most different tasks, see Figure 1.1b for illustration.

Now we want to briefly introduce the reader to the physiological states of *E. coli*, following the order given by the sketch in Figure 1.1c.

(a) (b)



(c)

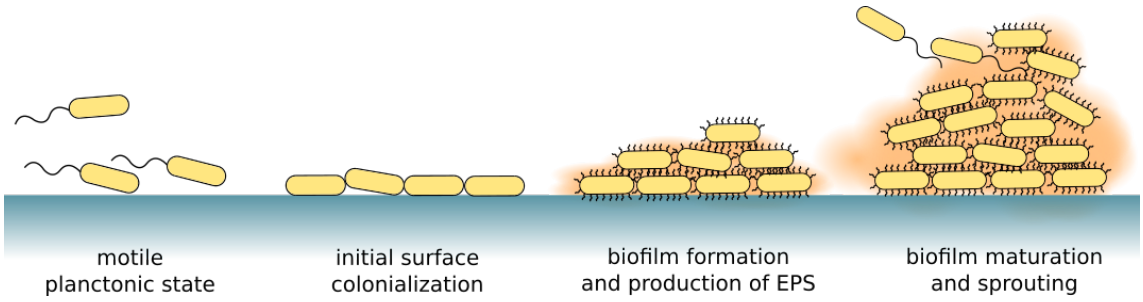


Figure 1.1: (a) Light-microscopy picture of an *E. coli* biofilm, taken from Ref. [1]. (b) Sketch of the architecture of a single *E. coli* bacterium, from Microbiology Notes [2]. (c) Sketch of the physiological life-cycle of *E. coli*, inspired by Ref. [3]

1.1.1 Planctonic state

Today the planctonic or motile state arguably attracts the highest attention from the physicists community. Swimming bacteria are along with flocks of birds and robot swarms [14, 15] a prototypic example of collective motion and more generally active matter [8, 16]. This state is closely related to the earliest systems that were studied in the field of active matter (*e.g.* by Viscek *et al.* [17]).

E. coli cells reach a length of up to roughly 8μm, a width of 1.5μm, and they develop multiple flagella. A flagellum comprises a rotary motor, that drives a long helical filament that extends several cell body lengths out into the external medium and is an interesting piece of nanomachinery of its own [18].

Bacteria use these flagella to propell themselves performing a *run-and-tumble* motion. Intervals of concerted flagellar movements, that result in straight paths (*runs*), are interrupted by non-synchronized flagellar movements, that lead to random reorientations of the bacterium (*tumbles*). The frequency of these *tumble* events is tuned to the environmental conditions, and result in an effective attraction towards favorable regions in terms of nutrient supply (*chemotaxis*) [19], temperature (*thermotaxis*) [20], and other stimuli.

Bacteria live in a low-Reynolds number environment, i.e. the viscous forces dominate over the inertial forces [21]. The effect of explicit hydrodynamic interactions (and consequently momentum conservation) is increasingly studied within the notion of *wet* active matter [22], while *dry* active matter research follows a coarse grained approach, to study larger systems [23,24]. While it is possible to study some phenomena by approximating the hydrodynamic fields with simple force dipoles, on the microscopic level some results sensitively depend on the explicit geometry and interactions [25–28]. The swimming mechanics of individual cells have been recently studied in great detail [29], and the impact of external fields attracts increasing attention [30]. A range of fluid-dynamic solvers can be used to study these systems [31], where *Multiparticle Collision Dynamics* [32–34] and its variants proove to be one of the most popular choice today.

Active motion systems do not obey detailed balance [35] and yield a range of interesting phenomena and potential applications. Among others, trapping [36] and cargo delivery [37], rectification [38], driving of micromachines [39] or the physiologically important *wall-hugging* have been studied. *Wall-hugging* denotes the effect that *E. coli* bacteria, like many other motile particles, experience an effective attraction towards surfaces. This attraction can be caused by hydrodynamic interactions and/or geometric effects. It is believed that this effective attraction to surfaces facilitates the invasion of surfaces and thus promotes the formation of sessile colonies.

1.1.2 Surface colonialization and growth

Sessile *E. coli* consume nutrients and grow in length, while maintaining their width. During this process new cell wall is inserted along the cylindrical midcell and insertion is decreased at the cell poles [40]. Once the cell reaches a certain length, a septum forms in the middle of the cell that closes and divides the bacterium into two basically equal daughter cells with complete nucleoids and at least one copy of the genome each [41]. This process is referred to as *binary fission*, and the details of the underpinning mechanisms are actively researched upon. Generally, the process comprises the coordinated operation of two mechanisms in the cytoskeleton. The growth and maintainance of the cell shape is attributed to the actin homologue MreB, that are filamentous proteins inside the cell envelope. The MreB cytoskeleton in *Escherichia coli* preferentially localizes at regions of negative curvature, directing growth away from the poles and actively straightening locally curved regions of the cell [42]. The second mechanism is related to the FtsZ protein. This protein is connected with the formation of the Z-ring, that spans the cell circumference and contributes to the cell-division [43].

When strictly constraint, *E. coli* can occupy all sorts of bent shapes, squeeze through pores as narrow as half their diameter. Within a certain range of distortions, *E. coli* bacteria can recover their original shape after the constrictions are released, beyond this regime a wide range of morphologies can persist [44].

At optimum conditions, i.e. sufficient supply of nutrients, oxygen and the optimum temperature *E. coli* colonys grow exponentially (*exponential-* or *log phase*) at a doubling time of approximately 20 minutes [45]. After a short time the nutrient sources are usually fully consumed or poisonous waste products hinder further bacterial growth. In the superseding *stationary phase*, bacteria grow smaller, with more spherical shapes, and become more resistent to various environmental challenges and starvation. Bacterial growth is characterized by long periods of nutritional deprivation punctuated by short periods that allow fast growth, a feature that is commonly referred to as the *feast-or-famine* lifestyle. See Ref. [46] for a more detailed overview.

Unconstrained *E. coli* colonies grow initially as a circular monolayer of particles with a complex internal organization [4, 12]. Particles at the colony perimeter predominantely orient perpendicular to the vector pointing from its position to the colony center [4]. Experiments with two differently dyed sub-populations showed that regions inhabited by distinct sub-populations shared fractal boundaries [47,48]. At a more detailed level, the formation of highly ordered micro-domains have been studied [12], which will also be the subject matter of our study. As a colony grows on, bacteria in the colony center perceive growth constraining pressure, as they have to push the neighboring bacteria to the side, and eventually buckle to form a second, and subsequently third layer [49,50], cf. Figure 1.2. A sufficiently large colony depletes the nutrients at the colony edge, bulk bacteria enter the *stationary phase*

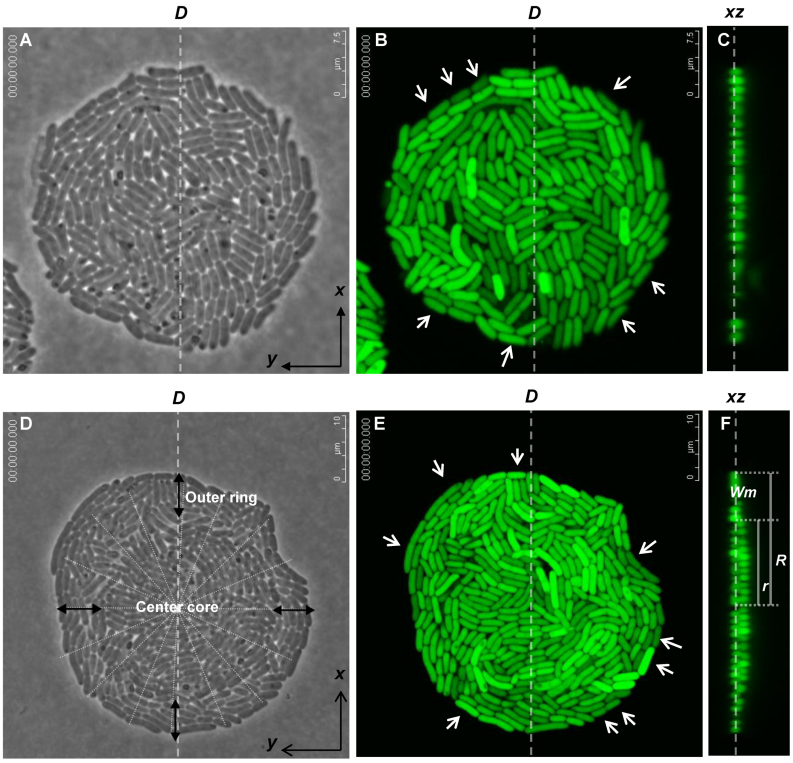


Figure 1.2: Architecture of a small *E. coli* colony. Light-microscopy image taken from Ref. [4].

while a leading edge keeps expanding [51]. The formation of rough colony edges (i.e. branching instability) were previously attributed to chemical and mechanical interactions [52, 53]. If multiple different species or mutants are present distinct sectors form. Also gene-surfing was observed [54, 55].

E. coli produce chemicals that diffuse and possibly trigger colony-level synchronized processes [56]. In response to environmental conditions, bacteria can develop a range of adhesive appendages (*fimbriae*, *pili*) to adhere to invaded surfaces and to engage in more complex interactions.

1.1.3 Biofilm formation

The transition from a surface-bound microcolony to a full-fledged biofilm is gradual. A biofilm is an equally complex and fascinating multi-cellular consortium, where the structural and functional properties drastically differ from the single-cell and colony properties. The initial structure of the microcolony prepatterns the biofilm architecture. At the onset of biofilm formation, bacteria cells produce large amounts of biopolymers, collectively termed extracellular polymeric substance (EPS). Despite the fact that EPS is not necessarily important for the initial colonialization [57], it proves pivotal for the development of a three-dimensional framework (biofilm matrix, see Figure 1.3) [5]. This matrix builds a highly heterogeneous structure of aggregates

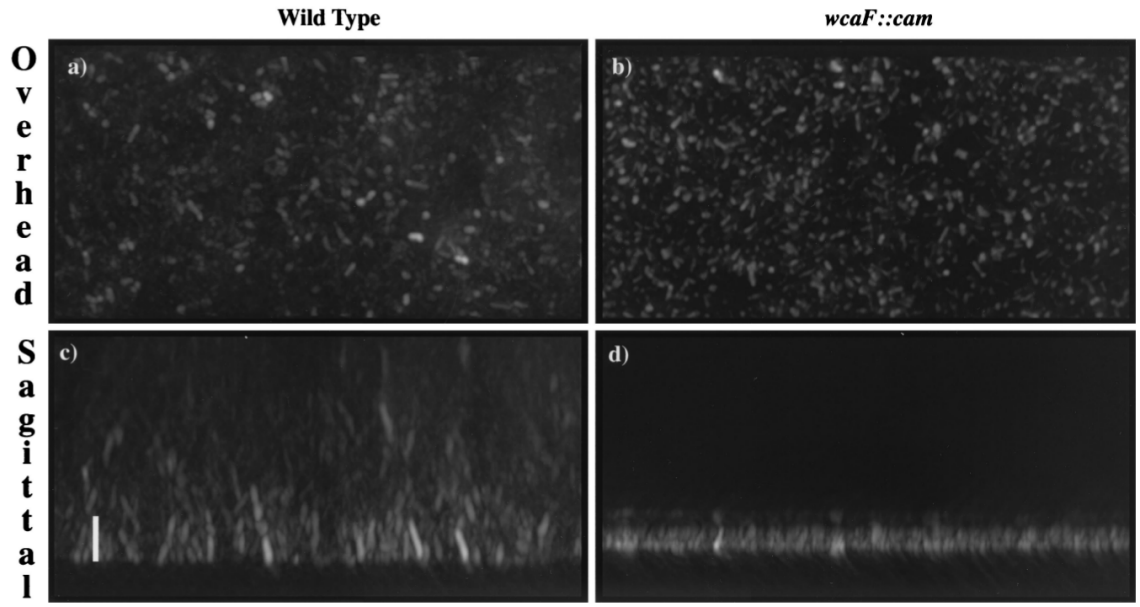


Figure 1.3: Structure of (a,c) wild type and (b,d) EPS-defective *E. coli* biofilms, grown for 72h under similar conditions. In contrast to the EPS-defective strain, the wild type *E. coli* form a three dimensional biofilm. Taken from [5].

Table 1.1: Composition ranges of biofilm matrices, from [7].

Component	**% of matrix**
Water	up to 97%
Microbial cells	2-5% (Many species)
Polysaccharides	1-2%
Proteins	<1-2%
DNA and RNA	<1-2%
Ions	unkown

of microbial cells, interstitial voids and channels, and provides protection against physical, chemical and biological agents, facilitates cell-cell signaling and horizontal gene transfer [58]. EPS might even provide a nutrient source for some of the cells [59]. Microbes account only for a small fraction of a biofilms mass [60] that is mostly made up by water [61], cf. Table 1.1 for a detailed listing.

Biofilms are commonly referred to as microbial communities, which reflects on the fact, that biofilms often comprise microorganisms of many different species [7,62]. Analyzing the composition of a microbial community is the common starting point for a detailed ecological analysis [7]. Each species can perform important tasks within the community and engage in a range of inter-species interactions. Symbiotic interactions include *parasitism* (i.e. one organism benefits at the expense of

the other), *commensalism* (i.e. one organism benefits without effecting the other) or *mutualism* (i.e. each organism benefits). Different species can also interact antagonistically, for instance, if they compete for the same nutrients or if they have a predator-prey relationship [63]. Evolutionary processes like point mutations and gene rearrangements can introduce additional heterogeneities [64].

An important feature of microbial communities is their robustness, which has structural and functional components. Biofilms continuously adapt to changing environmental conditions and can in turn modify their habitat by increasing or decreasing porosity of geologic media, altering the electrochemical properties, surface roughness, elastic moduli, stiffness, as well as seismic and magnetic properties of minerals through the precipitation of bacterial magnetosomes [65].

The earliest biofilm simulations employed Cellular Automata (i.e. local update rules on a discrete grid) [66] or Diffusion Limited Aggregation (i.e. random accumulation of biomass on the biofilm surface) [67] and seeked to study a biofilm on the scale of small volume elements. These models mostly neglected explicit growth and interaction mechanisms between individual cells, or volume elements. In contrast, the popular agent- and individual based models [68–70] incorporate particle-particle interactions, where the focus typically lies upon the chemical interactions. From a physicists view-point, biofilms resemble complex fluids: the rigid bacteria are analogous to colloids, and the EPS is a cross-linked polymer gel [65, 71]. This analogy has inspired polymer network-based models [72, 73].

The inherently complex nature of biofilms have caused some authors to raise doubts whether it is possible, after all, to gain insights into biofilms via purely reductionistic methods, but advocated a systems biology approach instead [63]. However, the large range of experimental and theoretical tools to study microbial communities are continuously advancing [62, 65, 67]. It will be necessary to focus the efforts on yet to be established model systems, to meet the many challenges that are connected with understanding the mechanisms that govern the complex structure and dynamics of microbial communities [62].

1.2 Morphogenesis

Morphogenesis is the biological process that causes an organism to develop its shape. The process controls the organized spatial distribution of cells during the embryonic development of an organism. Advances in data recording and data processing allow more detailed studies, coined “morphological profiling” [74]. This includes subpopulation identification and aggregation. Little is known about the influence of mechanical interactions on morphogenesis.

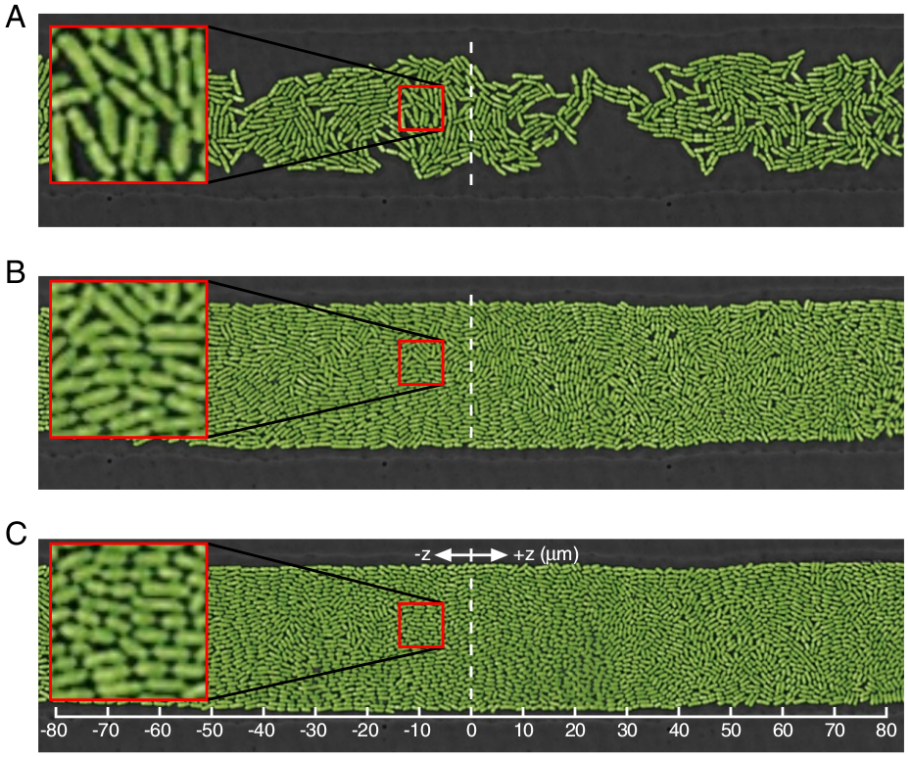


Figure 1.4: Emerging nematic order in a geometrically confined, growing *E. coli* colony. Subfigures (A-C) show the same colony at subsequent points in time. Taken from [6].

1.2.1 Significance of mechanical interactions

Traditionally, microbiological research focused on chemical processes in microbial colonies. Only recently the significance of mechanical interactions has been recognized. Today readily available microfluidic devices, make it possible to more closely study the mechanical interactions together with other effects [75]. About a decade ago, Volfson *et al.* [6] and Cho *et al.* [11] published pioneering works in this field. Both contributions showed that the emergence of order, in geometrically confined *E. coli* colonies, can be directly attributed to pairwise mechanical interactions, cf. Figure 1.4. Analogies to granular physics have been drawn. These initial results inspired further research, focussing on different confining geometries [76–78], sub-population boundaries [47], cell-morphologies [79], social interactions [80], nutrient distribution [52], buckling into a second layer [50], the structural influence of EPS [81], and the mesoscale structure [12].

Simulations are usually performed using Discrete Element methods. A comprehensive review on the mechanistic world of bacteria was recently given by Persat *et al.*, cf. Ref. [82].

Chapter 2

Model

Following an increasingly popular granular material approach, we study growing *E. coli* bacteria at the individual particle level, taking only mechanical interactions into account. Our model is closely related to the pioneering paper by Volfson *et al.* [6]. Our model can be used to perform detailed simulations in 3 dimensions, however, in this thesis we limit our studies to 2 dimensional systems. We will use the terms bacteria, cell and particle synonymously.

2.1 Particle model

A colony is a system of neutrally bouyant [83], rigid, granular particles, with an elongated spherocylindrical shape. The mass of a particle i computes to

$$m_i = \rho \left[\frac{(l_i - d)d^2\pi}{4} + \frac{d^3\pi}{6} \right], \tag{2.1}$$

and the diagonal elements of the mass moment of inertia tensor in the body frame approximate to (cf. Ref. [84])

$$I_{i,z} = \pi\rho d^5 \left[(\gamma_i - 1) + \frac{8}{15} \right] \quad \text{along the symmetry axis, and} \tag{2.2}$$

$$I_{i,x} = \pi\rho d^5 \left\{ \frac{\gamma_i - 1}{6} \left[3 + 4(\gamma_i - 1)^2 \right] + \frac{4}{3} \left[\frac{83}{320} + \left((\gamma_i - 1) + \frac{3}{8} \right)^2 \right] \right\} \tag{2.3}$$

along the off-symmetry axes. Here, ρ denotes the particles density and d the length of the particles off-symmetry axes. The end-to-end length of the spherocylinder i is denoted with l_i and the aspect ratio by $\gamma_i = \frac{l_i}{d}$.

Particles grow along their symmetry axis at an exponential rate, while keeping the dimensions of their off-symmetry axes constant [45].

$$\frac{dl_i}{dt} = \alpha l_i \tag{2.4}$$

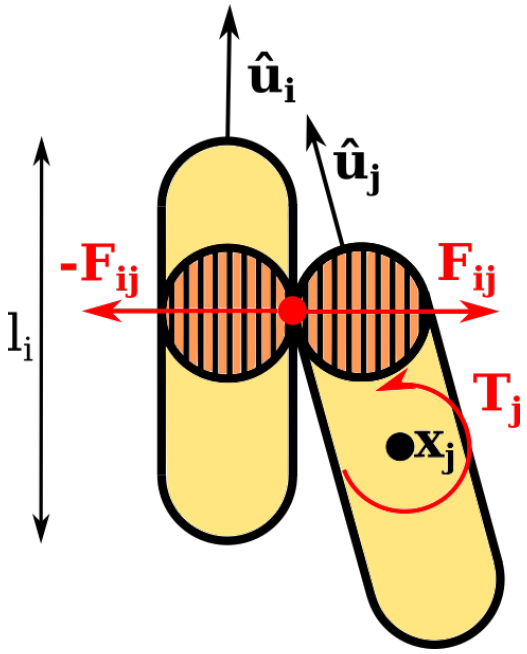


Figure 2.1: Sketch of two interacting particles: illustration of forces, torques, particle properties and geometry.

A particle grows up to a certain length l_i^{split} and splits into two particles of close to similar length. The mean splitting length $\langle l^{split} \rangle$ is chosen as a Gaussian-distributed random variable with an average of $l_i^{split} = 4d$, and with standard deviation $\sigma_l = 0.3d$. The two resulting colinear daughter cells have lengths of $l_i^{(1)} = pl_i$ and $l_i^{(2)} = (1-p)l_i$, where $p \in [0.45, 0.55]$ is a uniformly distributed random value. In the studied *exponential* growth regime, the proliferation behavior is robust and the effect of apoptosis can be readily neglected [45].

A particle's position is exhaustively described by the vector pointing from the coordinate origin to the center of mass $\mathbf{x}_i$, its length l_i and the orientational vector $\hat{\mathbf{u}}_i$. Particles interact via contact-forces (see Figure 2.1 for illustration).

2.2 Pair-wise interaction and equations of motion

The pair-wise particle-interaction was modelled as the interaction of two virtual spheres located at the points of closest approach along the particles' main axes. The two virtual spheres of diameter d, with centers at $\mathbf{r}_i$ and $\mathbf{r}_j$, and with velocities $\mathbf{v}_i$ and $\mathbf{v}_j$, interact via a Hertzian force of elastic contact and Coloumb friction

$$\mathbf{F}_{ij} = F_n \mathbf{n}_{ij} + \mathbf{F}_t, \tag{2.5}$$

$$F_n = k_n \delta^{3/2} - \gamma_n M_e \delta v_n, \tag{2.6}$$

$$\mathbf{F}_t = -\gamma_t M_e \delta^{1/2} \mathbf{v}_t, \tag{2.7}$$

where F_n is the normal force and F_t is the tangential force. The magnitude of the tangential force is bounded from above by the sliding friction force: $\mu_{cc}F_n$ for cell-cell contacts and $\mu_{cw}F_n$ for cell-wall contacts. M_e is the reduced mass for cell-cell interaction. $\delta = d - r_{ij}$ and $v_n = \mathbf{v}_{ij} \cdot \mathbf{n}_{ij}$ are the overlap and the relative velocity in the direction of the normal $\mathbf{n}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/r_{ij}$, respectively. The tangential velocity direction $\mathbf{t}_{ij} = \mathbf{v}_t/v_t$ is specified by the relative tangential velocity $\mathbf{v}_t = \mathbf{v}_{ij} - v_n\mathbf{n}_{ij}$.

Furthermore, particles experience a Stokesian drag force through the surrounding fluid

$$\mathbf{F}_{Stokes} = -\beta m\mathbf{v} \tag{2.8}$$

These forces enter Newton's equations of motion[1]

$$m\ddot{\mathbf{x}} = \mathbf{F}_e + \textstyle\sum_c \mathbf{F}_c, \tag{2.9}$$

$$\mathbf{I} \cdot \dot{\omega} = \textstyle\sum_c (\mathbf{x}_c - \mathbf{x}) \times \mathbf{F}_c, \tag{2.10}$$

where m and $\mathbf{I}$ are the mass and the tensor of intertia, respectively. $\mathbf{x}$ denotes the center of mass of the particle, ω is the angular velocity, $\mathbf{F}_e$ represents external body forces, the sums run over the contact forces applied at every contact a given particle has, and $\mathbf{x}_c$ defines the vector pointing from the particle's center of mass to the contact point.

2.3 Reduced units and parametrization

There is no unique choice of reduced units. In principle, the goal is to simplify the equations of motion and to generalize the outlined problem to describe a class of related problems. This procedure allows to easily identify qualitatively equal configurations and helps to prevent running redundant simulations.

In accordance with the original paper by Volfson *et al.* [6], we normalized all quantities by an appropriate combination of the diameter, d, mass of a virtual cube, $M = \rho d^3$, and the gravitational acceleration, g.

$$\text{distance} \quad r^* = r/d \tag{2.11}$$

$$\text{mass} \quad m^* = m/M \tag{2.12}$$

$$\text{time} \quad t^* = t\sqrt{g/d} \tag{2.13}$$

$$\text{force} \quad F^* = F/(Mg) \tag{2.14}$$

The characteristic mass computes to $M = 10^{-18}$kg, the characteristic length is 1μm and the characteristic timescale computes to roughly $\sqrt{g/d} \approx 50$ minutes, i.e. twice the experimentally observed doubling time of *E. coli*. We have chosen the

[1] In literature, it is also common to integrate a simplified set of equations in the over-damped limit.

growth rate accordingly to $\alpha = 2 \ln 2 \approx 1.4$. The force parameters were chosen to be the same values as published by Volfson: $k_n = 2 \times 10^6 (Mg/d)$ and $\gamma_n = \gamma_t = 2.2 \times 10^2 (g/d)^{1/2}$. The coefficients of friction for cell-cell and cell-wall interactions are $\mu_{cc} = 0.1$ and $\mu_{cw} = 0.8$, respectively. The Stokes drag factor was chosen to be smaller than in the original paper to be $\beta = 0.04$.

Chapter 3

Methods

Based on the forces described in Chapter 2, we performed Discrete Element Simulations to simulate freely growing *E. coli* bacteria on a two dimensional plane. The simulations were seeded with two particles of equal length, and random relative orientations and separation. The time integration was implemented in parallel code, running on GPUs using the python `pycuda` package.

3.1 Discrete Element Simulations

The chosen simulation approach is based on the well known method of molecular dynamics (MD) simulations [6]. The main idea of the method is to follow the dynamics of the individual particles that constitute the system under study. In the case of a gas, these particles are molecules, and in the case of a granular material, each particle represents a single grain.

In the general case, for the integration of the equations of motion we need to calculate the particle's moment of inertia around any axis. Following the integration scheme outlined in Ref. [85], [1] we calculate the moment of inertia in the particle's body frame and transform it into the laboratory frame, Therefore, the orientation of a particle is described with a rotation quaternion $\mathbf{q} = (q_0, q_1, q_2, q_3)$ [85, 86], that is related to a rotation matrix

$$\mathbf{D}(\mathbf{q}) = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix}. \tag{3.1}$$

We distinguish vectors in the laboratory frame and body-fixed frame by a superscript, i.e. $\mathbf{v}^s$ is a vector in the laboratory (or space-fixed) frame, while

$$\mathbf{u}^b = \mathbf{D}\mathbf{u}^s \tag{3.2}$$

[1] The integration scheme was originally chosen, to study microbial growth in an explicit fluid. For the final simulation setup and the questions we eventually addressed, a simpler integration scheme would be sufficient.

is the corresponding vector in the body-fixed frame. For vectors in the laboratory frame, we will frequently omit the superscript. The orientation 3-vector of a particle is $\mathbf{u} = \mathbf{D}^\intercal \mathbf{u}^b = \mathbf{D}^\intercal (0,0,1)^\intercal$. The moment of inertia tensor in the body-fixed frame $\mathbf{I}^0$ is a constant diagonal matrix. The angular velocity is calculated as $\omega^s = \mathbf{D}^\intercal (\mathbf{I}^0)^{-1} \mathbf{D} \mathbf{l}^s$, were $\mathbf{l}$ is the angular momentum.

This leads us to the following definition of the equations of motion

$$
\begin{aligned}
M\ddot{\mathbf{x}} &= \mathbf{F}, && (3.3)\\
\ddot{\mathbf{q}} &= \frac{1}{2}\left[\mathbf{Q}(\dot{\mathbf{q}}) \begin{pmatrix} 0 \\ \boldsymbol{\omega}^b \end{pmatrix} + \mathbf{Q}(\mathbf{q}) \begin{pmatrix} 0 \\ \dot{\boldsymbol{\omega}}^b \end{pmatrix}\right], && (3.4)\\
\dot{\mathbf{q}} &= \frac{1}{2}\mathbf{Q}(\mathbf{q}) \begin{pmatrix} 0 \\ \boldsymbol{\omega}^b \end{pmatrix}, && (3.5)\\
\frac{d\omega_\alpha^b}{dt} &= I_\alpha^{-1}\left[T_\alpha^b + (I_\beta - I_\gamma)\omega_\beta^b \omega_\gamma^b\right]. && (3.6)
\end{aligned}
$$

Here, $\mathbf{Q}(\mathbf{q})$ is defined as

$$
\mathbf{Q}(\mathbf{q}) = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \qquad (3.7)
$$

To numerically integrate the equations of motion, we employ a Verlet algorithm. For each time-step $\tau = 0.001\sqrt{d/g}$ we update $\mathbf{x}$, $\mathbf{q}$ and l according to

$$
\begin{aligned}
\mathbf{x}(t+\tau) &= \mathbf{x}(t) + \dot{\mathbf{x}}(t)\tau + \frac{\tau^2}{2M}\mathbf{F}^s(t), && (3.8)\\
\mathbf{q}(t+\tau) &= (1-\tilde{\lambda})\mathbf{q}(t) + \dot{\mathbf{q}}\tau + \frac{\tau^2}{2}\ddot{\mathbf{q}}, && (3.9)\\
\tilde{\lambda} &= 1 - \frac{\dot{\mathbf{q}}^2\tau^2}{2} - \sqrt{1 - \dot{\mathbf{q}}^2\tau^2 - \dot{\mathbf{q}}\ddot{\mathbf{q}}\tau^3 - (\ddot{\mathbf{q}}^2 - \dot{\mathbf{q}}^4\tau^4/4)}, \text{ and} && (3.10)\\
l &= l\,(1+\alpha\tau)\,. && (3.11)
\end{aligned}
$$

3.1.1 Force calculation

Force calculations are typically the computationally most expensive tasks, when performing Molecular Dynamics or Discrete Elements simulations. If one calculates the particle-particle interactions in a naive brute-force manner, the required computational effort scales with the particle number N like $\mathcal{O}(N^2)$. A common procedure to improve the performance of the force calculation requires to create a neighbor list to keep track of the position of each particle in the simulation box and calculate the pair-wise forces only for particles within the interaction range, cf. Figure 3.1. We implemented a grid of cubical bins with side-length $L = (\langle l^{split} \rangle/2 + \sigma_l)$, where we

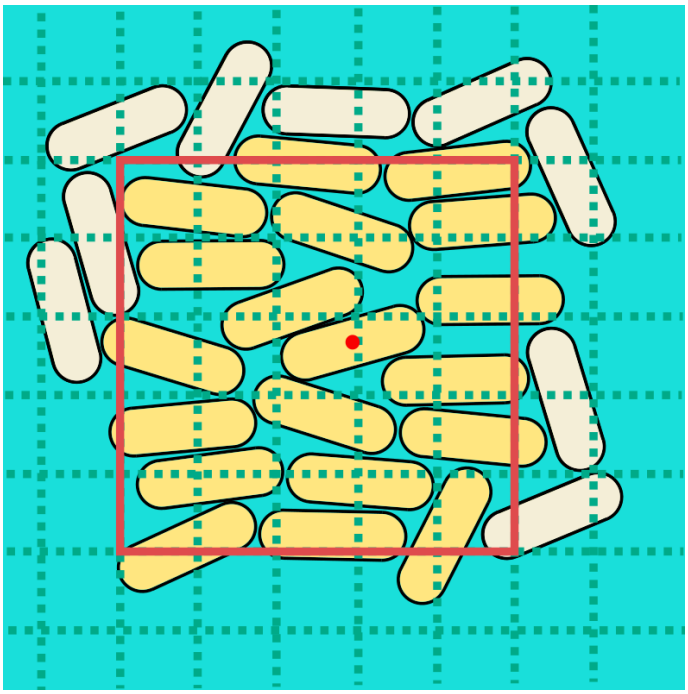


Figure 3.1: At regular intervals, particles are assigned to bins on a fixed grid. The pair-wise interactions are only computed for particles that reside in bins that are in proximity to one another. The red dot in the sketch indicates the center of mass of a particle i, the dashed lines indicate the bin-boundaries, and the red solid line marks the outline of the considered bins. The interaction potential between the central particle (with the red dot) and the particles whose center of mass lies outside the region, is not calculated.

calculated the interactions between particles that resided in the next 2 bins to each side. We assigned particles to bins at an interval of $\Delta t_{bin} = 0.01\sqrt{d/g}$.

We stored bins in a linked list data-structure for which the computational effort for calculating forces between binned particles scales like $\mathcal{O}(N \log(N))$. For the precise distance calculation between two rods at the points of closest approach, we implemented the algorithm proposed by Vega and Lago [87].

3.2 Simulation setup

Simulations were seeded with two bacteria in close proximity, of equal length $l^{initial} = 3d$ and random relative orientations. We performed simulations in two dimensions only, with freely growing colonies, i.e. bacteria do not experience boundaries, and do not buckle into next layers.

Chapter 4

Microdomains

The basic objective of this thesis is to gain insights into the local organization of micro-colonies and to study regions that have some degree of autonomy within a colony. An intuitive approach is to group together objects that are similar and separate dissimilar objects. You *et al.* [12] proposed to study the mesoscopic structure of growing *E. coli* colonies, by grouping adjacent particles to *microdomains* if they are in contact (*i.e.* the distance between the points of closest approach along their main axes is below a threshold, $\delta_{ij} \leq \delta_t$) and the enclosing angle α_{ij} between their main axes is below a certain threshold: $\cos\alpha_{ij} = |\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j| \geq c_t$ [12]. This partition can be easily obtained by interpreting the colony as an undirected, spatially embedded network [88], where each particle corresponds to a node and the network can be described with an adjacency matrix. The adjacency matrix J_{ij} describes the topology of a network, where each node i is represented by both the ith-row and the ith-column of the matrix, and the matrix values describe the link "strengths", where 0 means no link. The adjacency matrix of an undirected network is symmetric, and for the outlined case it is constructed as

$$J_{ij} = f(\delta_{ij}, \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = \mathcal{H}(\delta_{ij} - \delta_t)\mathcal{H}(|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j| - c_t), \tag{4.1}$$

with the Heaviside-function $\mathcal{H}$. The proposed hard thresholding criterion results in a graph of several disconnected components $\Sigma = \{\sigma_k\}$, where each component corresponds to a microdomain.

This simplistic approach is a special case of *modularity maximization*, that is part of the more general framework of *community detection* methods. In the following sections, we want to discuss the creation of a graph-representation and elaborate the general ideas behind community detection. Furthermore, we critically discuss the specifics and weaknesses of the techniques we used, and introduce means to validate the identified microdomains. We use standard network nomenclature; for a comprehensive review please refer to Ref. [89].

4.1 Community detection

Given a system's graph representation, an expanding body of scholarly literature has addressed the partitioning problem. This is commonly refered to as *community detection* and the identified groups are consequently refered to as *communities* [90–93]. Community detection is frequently used to study social or biological networks, and only recently community detection was used to study force-chain networks in granular media [94–96].

Modern interpretations of communities focus on the *probability* that vertices share edges with a subgraph. Fortunato [90] stated: "The existence of communities implies that vertices interact more strongly with the other members of their community than they do with vertices of the other communities. Consequently, there is a preferential linking pattern between vertices of the same group."

This definition is the foundation of the popular *modularity maximization* approach [97], that is based on maximizing a so called modularity or quality function

$$Q(\{\sigma_k\}) = \frac{1}{2m}\sum_{i\neq j} J_{ij}\delta(\sigma_i,\sigma_j) = \frac{1}{2m}\sum_{i\neq j}(A_{ij}-p_{ij})\delta(\sigma_i,\sigma_j), \tag{4.2}$$

where $\delta(\sigma_i,\sigma_j)$ is the Kronecker delta, σ_i, σ_j denote the communities of the particles i and j, A_{ij} the adjacency matrix, m the sum over all edge-weights and p_{ij} the null-model (see next Section). This maximization problem is analogous to finding the ground state of an infinite range Potts spin glass [98] and an interaction is called *ferromagnetic* when $J_{ij} > 0$ and *antiferromagnetic* when $J_{ij} < 0$ [99].

The modularity function can be easily extended to multi-layer networks (time sequence of network-snapshots), cf. Appendix A.1.

4.2 Graph representation and null model

A system's graph representation and the corresponding null model are closely related. The null model usually contains important information about the way a graph is constructed. It is possible to include any form of prior knowledge about the network to construct p_{ij} [93]. Choosing a suitable null-model is a crucial step in modularity maximization to find good partitions.

As we want to group similar particles, it is important to note that a range of similarity measures may appear plausible. There is no commonly agreed-upon procedure to construct a systems graph representation. In accordance with You *et al.* [12] and Bassett *et al.* [95], we construct our graph representation such that two particles are only connected via an edge, if they are touching. Due to their spatial extent, particles can only be in contact with their direct neighbors [95]. Furthermore, we calculate the edge weight between two particles as a function of

the relative orientation. We will describe three different simple monotonic functions as possible choices to define the J_{ij} matrix, that incorporates the edge weights ω_{ij} and the corresponding null models p_{ij}[1], for illustration cf. Figure 4.1:

- Heaviside [12]
  A hard threshold angle $c_t > |\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j|$ is defined. All pairs of adjacent particles, that have a smaller enclosing angle form ferromagnetic links, the others do not form links; the adjacency matrix reads

$$J_{ij} = \mathcal{H}(\delta_{ij} - \delta_t)\mathcal{H}(|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j| - c_t) \tag{4.3}$$

  This is the simplest way to identify microdomains, but the dichotomization profoundly alters the network structure, that was reported to potentially yield a range of problems [100]. Specifically, this definition is very sensitive to small perturbations and denies to study the time evolution of microdomains[2]. Furthermore it is important to note that in this procedure only ferromagnetic links are taken into account, while antiferromagnetic links are neglected.

- Signum
  A hard threshold angle $c_t > |\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j|$ is defined. All pairs of adjacent particles, that have a smaller enclosing angle form *ferromagnetic* links, the others form *antiferromagnetic* links; the adjacency matrix reads

$$J_{ij} = \mathcal{H}(\delta_{ij} - \delta_t)\left[2\mathcal{H}(|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j| - c_t) - 1\right] \tag{4.4}$$

  Besides taking antiferromagnetic links into account, this approach shares many properties with the Heaviside mapping.

- Polynomial
  A simple choice that retains link weights is to use a polynomial function [93]

$$J_{ij} = \mathcal{H}(\delta_{ij} - \delta_t)(|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j|^n - c_t^n). \tag{4.5}$$

  This yields a continuous range of edge-weights. *Ferromagnetic* and *antiferromagnetic* links are taken into account, and the identified domains are more

[1] Besides the described choices, the sigmoidal function features favorable properties. This approach was not studied in detail, but the formulation and the rationale are provided in the Appendix, cf. Section A.2.

[2] We found it not feasible to study the microdomain dynamics using the Heaviside network mapping and performing an *ad hoc* 2-step approach [91]. In the first step, we identified the colonies' microdomains at two subsequent time-steps t and $t+\Delta t$. In the second step, microdomains σ_i were associated between time-steps by using a genetic algorithm to maximize the mutual information between the two partitions. cf. Figure 4.1.

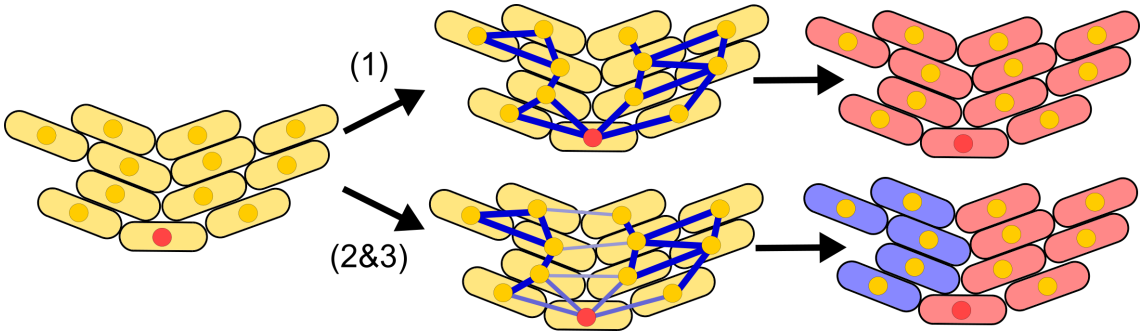


Figure 4.1: Illustration of the proposed contact network mappings. The red and yellow dots indicate a particle's node representation and the blue lines the edges that two particles at contact share, where the saturation indicates the contact angle depended edge-weight. The deficits of dichotomization (Heaviside-mapping (1)) compared to the other mappings (2&3) become evident, as the resulting network topology sensitively depends on the orientation of the particle marked with the red dot. In contrast, the mappings (2&3) preserve the topology independently of the orientation of any individual particle. This yields a more robust and intuitive notion of microdomains. Here, possible partitionings into microdomains are indicated by the blue and red coloring of the particles.

robust against small perturbations. However, finding a community structure is computationally more expensive. For our further studies we have chosen a linear ansatz, $n = 1$.

Modularity maximization is known to have deficits in certain contexts, like a resolution limit [101] and a high computational complexity (NP-hard) [102]. Most severely, the quality function might have many near-optimum degeneracies [103], that means one cannot generally assume a community structure of a network is uniquely defined. There may exist several but very different partitions that all have a comparably high value of modularity. Quality optimizing algorithms can emphasize existing difficulties, or even introduce new ones.

4.3 Similarity of identified communities

The aforementioned subtleties of community detection ask for means to evaluate its effectiveness. Applying the same algorithm several times to a single network potentially yields structurally very different partitions at comparable modularity values. To address this issue, we generally repeat the community detection algorithm 10 times for a single network and threshold-angle c_t, and aggregate the results. Furthermore, we will introduce two methods to evaluate the self-similarity and the similarity of communities detected across different network mappings.

4.3.1 Normalized mutual information

Similarity of two partitions can be estimated by the *normalized mutual information*, that is the "amount of information" that can be obtained about one partition through the other partition [90]. The *normalized mutual information* is not a metric.

This quantity is derived from information theory, where the entropy is the fundamental measure for information, or surprise that is carried by an event. For instance, the flip of a fair coin with equal probability of heads and tails, carries one bit of information and the entropy of m tosses is m bits. Less likely events provide more information, conversely, more likely events provide less information. Events that are certain to occur do not provide any information.

Let $\Sigma = \{\sigma_i\}$ and $\Pi = \{\pi_i\}$ be community assignments, where given two alternative partitions $\mathcal{S}$ and $\mathcal{P}$ of the same colony, σ_i and π_i indicate the community labels for each particle i within partition $\mathcal{S}$ and $\mathcal{P}$, respectively. The entropy of a community assigment Σ is defined as

$$H(\Sigma) = -\sum_{\sigma} P(\sigma) \log P(\sigma), \tag{4.6}$$

where $P(\sigma) = P(\Sigma = \sigma) = n_\sigma^\Sigma / n$ is the probability of particle i being assigned to community σ, and for $P(\pi) = P(\Pi = \pi) = n_\pi^\Pi / n$ analogously.

The mutual information of two random variables Σ and Π is a measure of the mutual dependence between the two variables. More specifically, it quantifies the "amount of information" obtained about one random variable, through the other random variable. It is defined as

$$MI(\Sigma, \Pi) = H(\Sigma) - H(\Sigma|\Pi), \tag{4.7}$$

where $H(\Sigma|\Pi) = -\sum_{\sigma,\pi} P(\sigma, \pi) \log P(\sigma|\pi)$ is the conditional entropy of Σ given Π.

Dividing the mutual information by the arithmetic average of the entropies of $\mathcal{S}$ and $\mathcal{P}$, yields the *normalized mutual information* (NMI)

$$NMI(\mathcal{S}, \mathcal{P}) = \frac{2MI(\Sigma, \Pi)}{H(\Sigma) + H(\Pi)}. \tag{4.8}$$

The NMI equals 1 if and only if the partitions are identical, whereas it has an expected value of 0 if they are independent. Figure 4.2 gives an intuition of this measure. In the following, we used the implementation that was published in the python `igraph` package [104].

4.3.2 Consensus matrix

Another method to visualize the robustness of a partitioning is by plotting the consensus matrix. For this purpose the community detection algorithm is performed

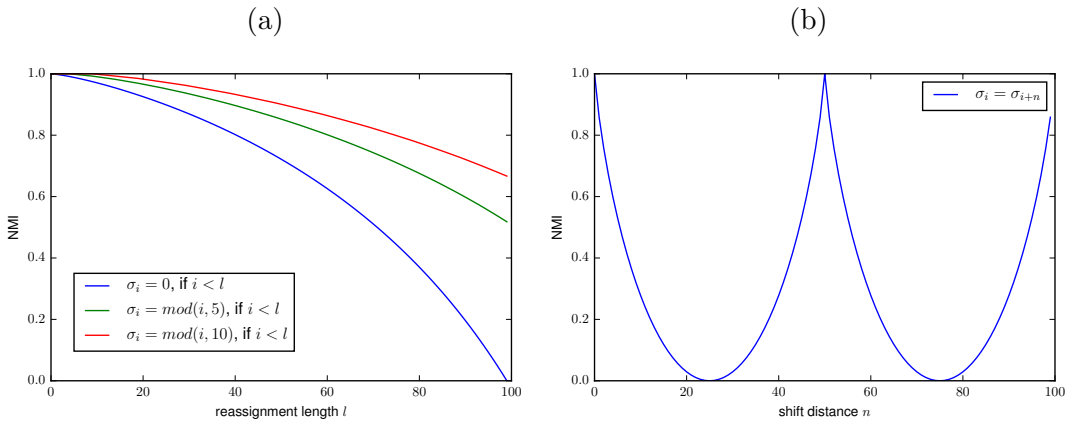


Figure 4.2: The NMI for two at-first-identical partitions is calculated, where one partition is continuously modified from the original form. In panel (a) $\mathcal{S} = \mathcal{P} = (0, 1, \cdots, n)$ is calculated, where $n = 100$ is the total number of particles and the first l particles in partition $\mathcal{S}$ are reassigned to a new community σ_i. In (b) $\mathcal{S} = \mathcal{P} = (0, 0, \cdots 0, 1, 1, \cdots, 1)$ with the first half of particles assigned to community $\sigma_i = 0$ and the second half to $\sigma_i = 1$, these two assignment vectors are shifted with respect to each other.

multiple times on a single network. Then a matrix is created, similarly to an adjacency matrix, where particles share a edge if they have been assigned to the same community, the edge-weight being the fraction of times of this co-assignment. In the next step, the columns and rows have to be rearranged, such that cohesive blocks are grouped. This is achieved by applying the Reverse-Cuthill McKee ordering [105]. The resulting - in many cases blockdiagonal - consensus matrix provides a good overview of the domain-size distribution and the detection robustness. We used an implementation of the Reverse-Cuthill McKee ordering that is provided in the python `scipy` package.

4.4 Validation of communities

This leaves us with the fundamental challenge of evaluating the results, without auxiliary information, or differently put:

How shall we choose the threshold angle c_t to find good partitions? How does a good partition look like?

In analogy to the data-science task of cluster validation, we suggest a number of approaches to resolve this problem.

4.4.1 Particle neighborhood

A common way to parameterize the modularity function (Eq. 4.2) is such, that a fixed fraction (e.g. 50 percent) of the edges are rendered ferromagnetic, the remaining edges are antiferromagnetic. Information about a particles neighborhood such as the contact angle distribution $|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j|$ may inform a similar choice.

Contact number

We define the contact number N_c of a particle i, as the number of adjacent particles that are in contact with particle i. As the particles interact purely via contact-forces, the system offers an intuitive notion of particles in contact. For the data analysis, we considered particles to be in contact at a distance of closest approach δ, slightly larger than a particle-width ($\delta^2 \leq 1.2d$). This procedure allows to observe fully relaxed contacts.

Radial-Distribution function

We calculate the *radial-distribution function* $g(r)$, which provides a measure of the local spatial ordering in the bulk of a fluid. It quantifies the probability of finding a particle within a shell of radius r, around a central particle at the coordinate origin ($r_i = 0$), normalized by the mean particle density ρ_0. In 2D it can be calculated from

$$g(r) = \frac{1}{2\pi r(N-1)\phi_0} \langle \sum_j \sum_{i \neq j} \delta(|\mathbf{x}_i - \mathbf{x}_j| - r) \rangle \tag{4.9}$$

where ϕ_0 and N are the total area fraction and number of particles, respectively. Structured systems exhibit peaks in $g(r)$, these indicate favored interparticle distances. In contrast, uncorrelated particle systems do not have favored distances and the radial distribution function is a constant, $g(r) \approx 1$. More specifically, a perfect crystal structure would yield a sequence of delta-peaks, marking the lattice points, whereas fluids typically exhibit locally ordered structures, that result in a sequence of peaks with decreasing magnitude and no long-range correlation.

To calculate the radial distribution function, only central particles i in the bulk of your colony are taken into account.

4.4.2 Elbow method

The elbow method is a visually aided technique to choose a reasonable parameter for a clustering task. It is commonly used to choose the correct cluster count k, for a k-means algorithm. Clustering is performed using different values of k, and the variance is calculated. As more clusters are introduced, the variance decreases,

here the variance is zero for a single individual in each cluster, which is obviously overfitting the data. However, the variance over k sometimes shows a point of rapidly changing slope, from this point on the variance decreases slower with increasing cluster-number k than previously. The location of the elbow is accepted as good choice for k.

Translated to our problem, we vary the threshold value c_t and evaluate: (i) the modularity function as defined above, and (ii) the domain averaged nematic order parameter and search for an elbow in the plot.

Nematic order parameter

The scalar nematic order parameter S is a measure for the parallel alignment of particles in a system. $0 \leq S \leq 1$, where larger values indicate higher order and smaller values lower order. The tensor order parameter $Q_{\alpha\beta}$ for m nematogens in a d-dimensional system reads

$$Q_{\alpha\beta} = \frac{d\langle \hat{u}_{m\alpha}\hat{u}_{m\beta}\rangle - \delta_{\alpha\beta}}{d-1}. \tag{4.10}$$

The scalar nematic order parameter is defined as the largest eigenvalue of $Q_{\alpha\beta}$.

We calculate particle number-weighted average of the scalar nematic order parameter of individual domains as

$$S_\sigma = \frac{\sum_\sigma N_\sigma S}{\sum_\sigma N_\sigma}, \tag{4.11}$$

where S denotes the scalar nematic order parameter evaluated for the given domain σ, and N_σ the domain number-size.

4.4.3 Clustering index

A large number of clustering indices have been proposed to evaluate the quality of a clustering [106, 107]. We adapted the following definition of the silhouette index from Ref. [106].

Silhouette index

The silhouette index $\mathcal{C}$ compares the distance of a particle $i \in \sigma_k$ to the other particles j in the same domain, with the distance to the other adjacent domains. At the optimum resolution, it is maximal, where $\mathcal{C} = +1$ indicates highest proximity within the domains, at maximal distance to the adjacent domains, and $\mathcal{C} = -1$ the contrary.

First, we define the distance between two particles i and j as

$$d(i,j) = 1 - |\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j|. \tag{4.12}$$

The within-domain mean distance $a(i)$ reads

$$a(i) = \frac{1}{N_{\sigma_k} - 1} \sum_{\substack{i,j \in \sigma_k \\ j \neq i}} d(i,j), \tag{4.13}$$

The mean distance $\mathfrak{d}(i, \sigma_{k'})$ of i to the particles of each of the adjacent domains $\sigma_{k'} \in \partial\sigma_k$

$$\mathfrak{d}(i, \sigma_{k'}) = \frac{1}{N_{\sigma_{k'}}} \sum_{i \in \sigma_k j \in \sigma_{k'}} d(i,j). \tag{4.14}$$

Let us also denote by $b(i)$ the smallest of these mean distances

$$b(i) = \min_{k' \neq k} \mathfrak{d}(i, \sigma_{k'}) \tag{4.15}$$

For each particle i, one then forms the quotient

$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)}. \tag{4.16}$$

That is called the "silhouette width" of the particle. Extending this standard procedure, we assigned "orphaned" particles that constitute their own domain a silhouette width of $s_{orphaned} = -1$.

The mean of the silhouette widths for a given micro-domain σ_k is called the cluster mean silhouette and is denoted as $\mathfrak{s}_k$

$$\mathfrak{s}_k = \frac{1}{n_k} \sum_{i \in \sigma_k} s(i) \tag{4.17}$$

Finally, the global silhouette index is the mean of the cluster mean silhouettes

$$\mathcal{C} = \frac{1}{K} \sum_{k=1}^{K} \mathfrak{s}_k \tag{4.18}$$

4.4.4 Feature size matching

Another option to find a useful parameter value range is to match domain sizes with feature sizes derived from a different method. In the results section we will compare the feature sizes found with the *equal-time polar-pair correlation function*, with the average domain particle count N_σ and the average radius of the smallest domain enclosing circle ${r_\sigma}$[3].

[3] A number of other geometrical measures are possible [108, 109].

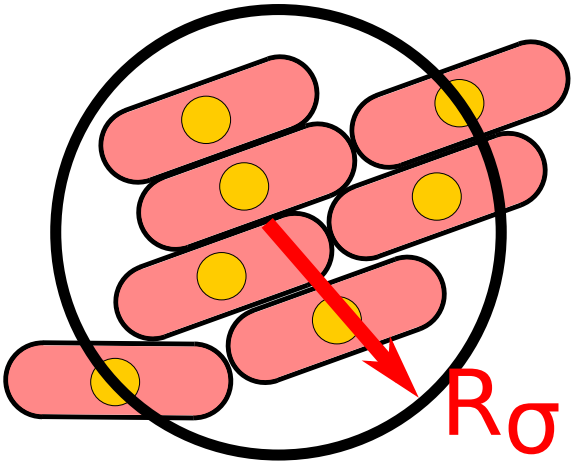


Figure 4.3: Sketch to illustrate the smallest enclosing circle, with radius R_σ around a group of particles.

Equal-time polar pair-correlation function

The spatial extent of a nematic microdomain can be approximated from the spatial correlation of the orientations of a center particle with particles at distance r, at time t, and can be described with the *equal-time polar pair-correlation function* [110]. That is defined as

$$C_2(r,t) = \frac{\langle \sum_{i\neq j} \left[d(\hat{\mathbf{u}}_i(t)\cdot\hat{\mathbf{u}}_j(t))^2 - 1 \right] \delta(|\mathbf{r}_i - \mathbf{r}_j| - r) \rangle}{(d-1)\langle \sum_{i\neq j} \delta(|\mathbf{r}_i - \mathbf{r}_j| - r)\rangle}. \tag{4.19}$$

The average is taken over all pairs of particles (i,j) in the bulk of the studied system, and $d = 2$ denotes the dimensionality of the studied system. $C_2 = 1$ indicates a parallel alignment, while $C_2 = 0$ for random relative alignment and $C_2 < 0$ for more perpendicular relative alignments.

Smallest enclosing circle

Analogously, the spatial extent of a microdomain can be estimated by the smallest circle that contains all center-points of the particles that contitute the microdomain, cf. Figure 4.3.[4] We calculate the average of the radii R_σ of the smallest enclosing circles, for all domains weighted by the number of particles

$$r_\sigma = \frac{\sum_\sigma N_\sigma R_\sigma}{\sum_\sigma N_\sigma}. \tag{4.20}$$

[4]We used a python implementation available at https://www.nayuki.io/page/smallest-enclosing-circle.

Pearson correlation

We use the Pearson correlation coefficient to measure the linear correlation between feature sizes x and y found using different methods. Strictly speaking, Pearson's correlation requires that each dataset be normally distributed, and not necessarily zero-mean. A value of $r = 1$ implies an exact linear positive correlation, $r = -1$ an exact linear negative correlation and $r = 0$ no correlation.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{4.21}$$

where $\bar{x}$ and $\bar{y}$ denote the respective mean values. We have used the implementation provided in the python `scipy` package.

4.5 Community detection algorithm

To optimize the quality function, we use an implementation of the popular Louvain method that was published by Mucha *et al.* [111]. The Louvain method is known to be computationally very efficient and to perform well in practical contexts [112]. The algorithm optimizes the quality function in a two step procedure. In the first step, the method locally optimizes the quality function by creating a partition with "small" communities. In the second step, the algorithm aggregates nodes that belong to the same community and builds a new network, where each community is a node. These steps are repeated iteratively until a maximum of modularity is attained. Eventually, the algorithm will partition the graph into non-overlapping communities [90]. The Louvain method performs best if the analyzed network comprises communities of approximately equal size.

4.5.1 Visualization

We found a simple way to visually distinguish adjacent microdomains. Similarly to the Louvain algorithm, we first aggreated nodes that belong to the same community and build a new network where each community is a node. Then we made use of a greedy color search algorithm, that is based on the four-color theorem, to make sure adjacent communities are colored in distinct colors. We used an implementation that was published in the python `networkx` package [113, 114].

Chapter 5

Results

5.1 Comparison with previously published results for growing *E. coli* colonies

To test our implementation of the discrete element simulations, we compare results obatained from our simulation to results that were reported in related experimental and simulation studies, and find good agreement.

5.1.1 Particle growth

In agreement with reported results, particles quickly form circular colonies [12, 52], and orientationally ordered patches evolve [12], cf. Figure 5.1. Visual inspection of the configurations shown in Fig. 5.1 reveals a strong degree of nematic alignment that is, however limited to short length scales. The local environment around any particle within the bulk of the system exhibits a microdomain of nearly parallel particles. As we move away from the reference particle discontinuous changes in orientation are often visible. These changes are the physical basis for our definition of microdomains.

The number of particles in the system grows at an exponential rate. All particles descend from the two seed particles, and their growth behavior is uniform except for a small variations in splitting ratios and splitting lengths. This yields synchronized particle growth within the system, however, relative growth synchrony is continuously decreasing over time, cf. Figure 5.2.

5.1.2 Local packing fraction

Colonies at unconstrained exponential growth behave like a compressible fluid, where the density in the center of the colony increases continuously [12, 52]. In the chosen simulation layout, as the colony size increases, particle overlap becomes inevitable.

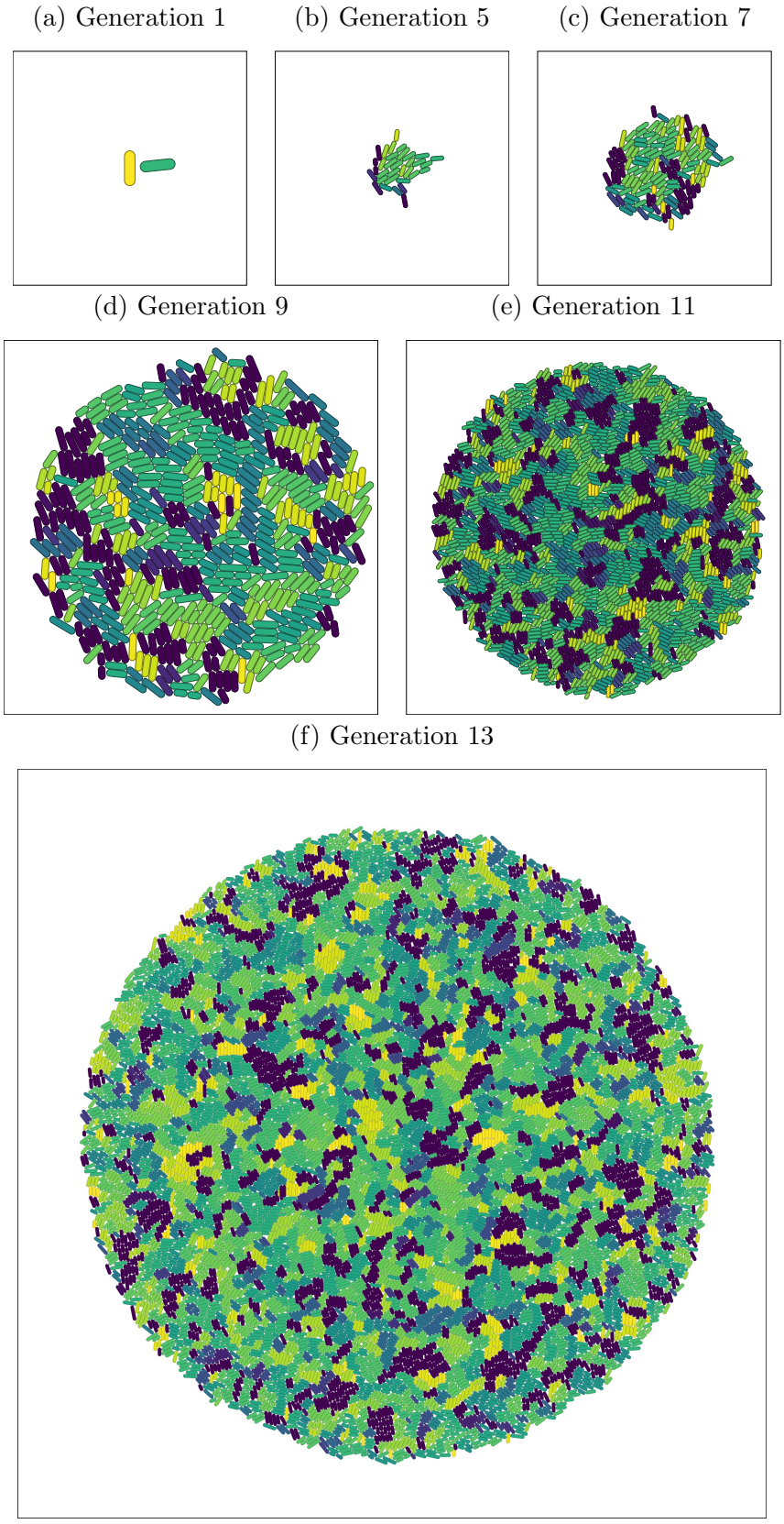


Figure 5.1: Growth of an unconstrained colony of particles with an average splitting length of $\langle l_{split} \rangle = 4$. Coloring indicates the orientation of the particles.

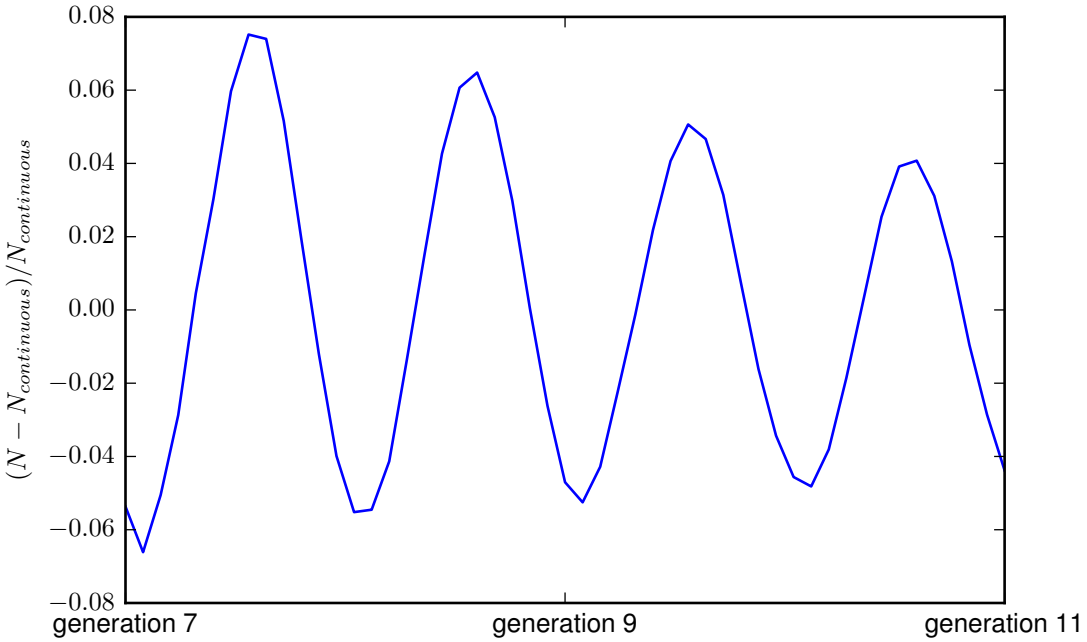


Figure 5.2: The relative difference between the observed number of particles in the colony and a continuous exponential count ($(N - N_{continuous})/N_{continuous}$) oscillates due to synchronized growth, and it decreases in magnitude over time.

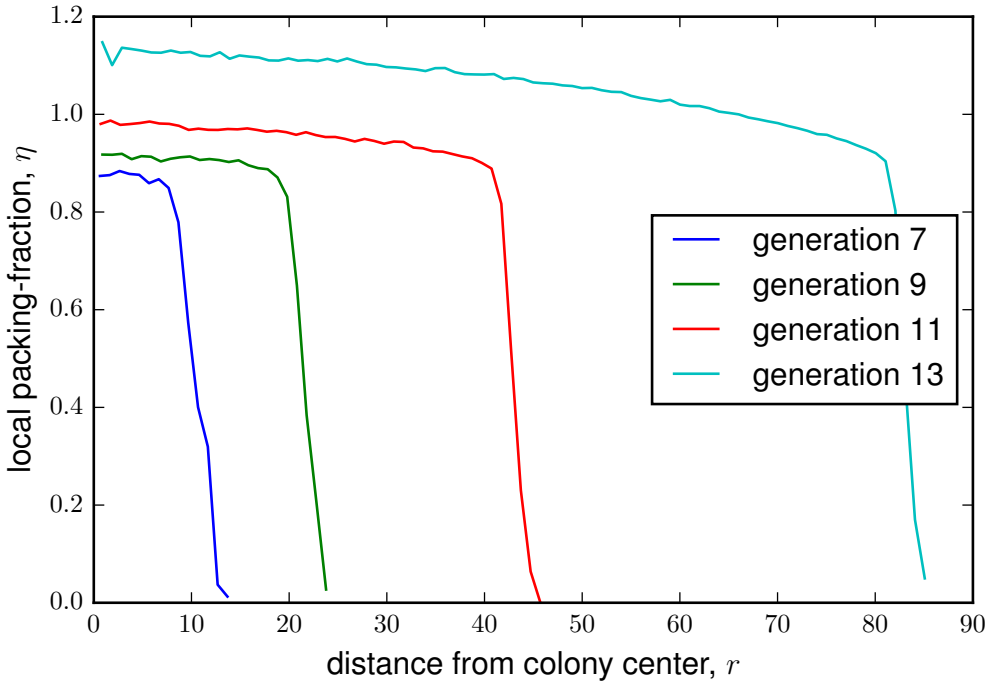


Figure 5.3: Local packing-fraction of colonies of particles with an average splitting length of $\langle l_{split} \rangle = 4$, at different times.

In vivo, bacteria will both buckle and change their physiological state in response to the growth pressure, therefore, we limit our observations to smaller colonies, before the onset of strong overlaps.

To calculate the local packing fraction, we approximated each particle by a set of $n_p = 10$ evenly distributed points along its main axis and performed a Voronoi tessellation. The local packing fraction can then be obtained from

$$\eta(r_i) = \frac{V_i}{\sum_{j=1}^{n_p} v_j^{(i)}} \tag{5.1}$$

with particle i's volume V_i, and the volume of a Voronoi cell $v_j^{(i)}$ of the point j along the main axis of particle i. Our observations are again in agreement with the previously published results. Figure 5.3 shows the dependence of the local packing-fraction η on distance from the colony center at successive generations. As the time (generations) increases, η increases monotonically within the bulk, but always shows a sharp drop in proximity of the colony boundary.

5.1.3 Radial nematic order structure

In agreement with experimental observations, we observe in our simulations that bacteria at the colony boundaries preferentially orient perpendicularly with respect

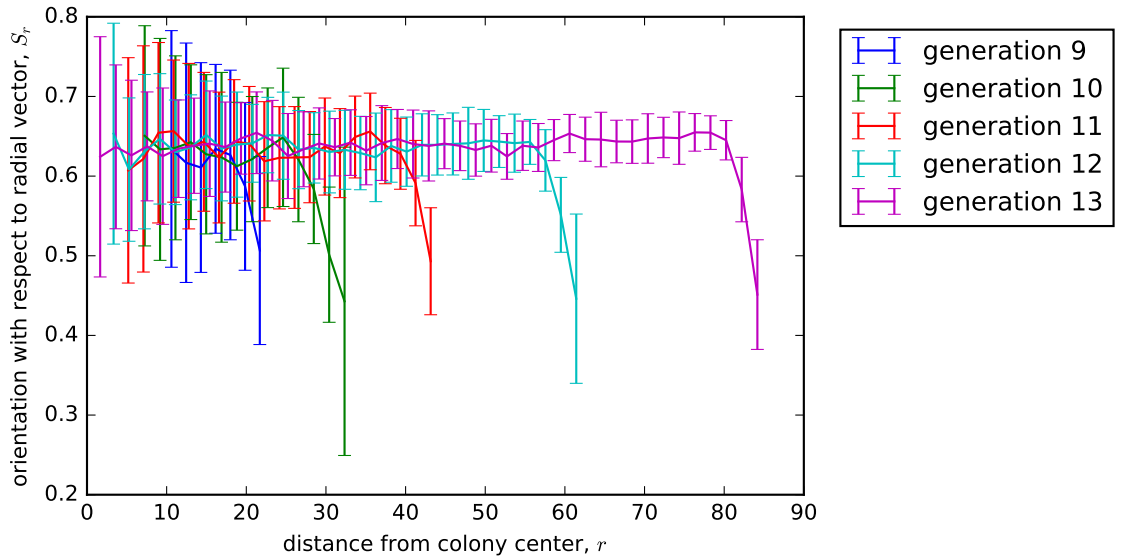


Figure 5.4: Radial nematic order structure of colonies of particles with an average splitting length of $\langle l_{split} \rangle = 4$. Particles in the bulk orient randomly $S_r^{random} = \frac{2}{\pi} \approx 0.637$, whereas particles at the colony edge preferentially orient perpendicular to the vector pointing to the colony center $S_r^{edge} < S_r^{random}$.

to the vector pointing from the colonies center of mass to the bacteria $\hat{\mathbf{r}}_{i,c} = \mathbf{x}_i - \mathbf{x}_c$ [4]. This can be quantified using the radial nematic order structure

$$S_r(r) = \frac{\langle \sum_{i \neq j} |\hat{\mathbf{u}}_i(t) \cdot \hat{\mathbf{r}}_{i,c}(t)| \delta(r_{i,c} - r) \rangle}{\langle \sum_{i \neq j} \delta(r_{i,c} - r) \rangle}. \tag{5.2}$$

We observe that, while in the bulk there is no global orientation and microdomains are equally likely to exhibit any alignment, at the boundary is always normal to the radial vector $\hat{r}_{i,c}$. Furthermore, the transition to this orientation occurs at a location that strongly correlates with the location of the sudden drop of η in Fig. 5.4.

5.2 Local structure

The local structure in complex systems is frequently studied, as it allows comparison across different domains. It aids further analysis and will serve as a basis for the subsequent study of micro-domains.

5.2.1 Particle contacts

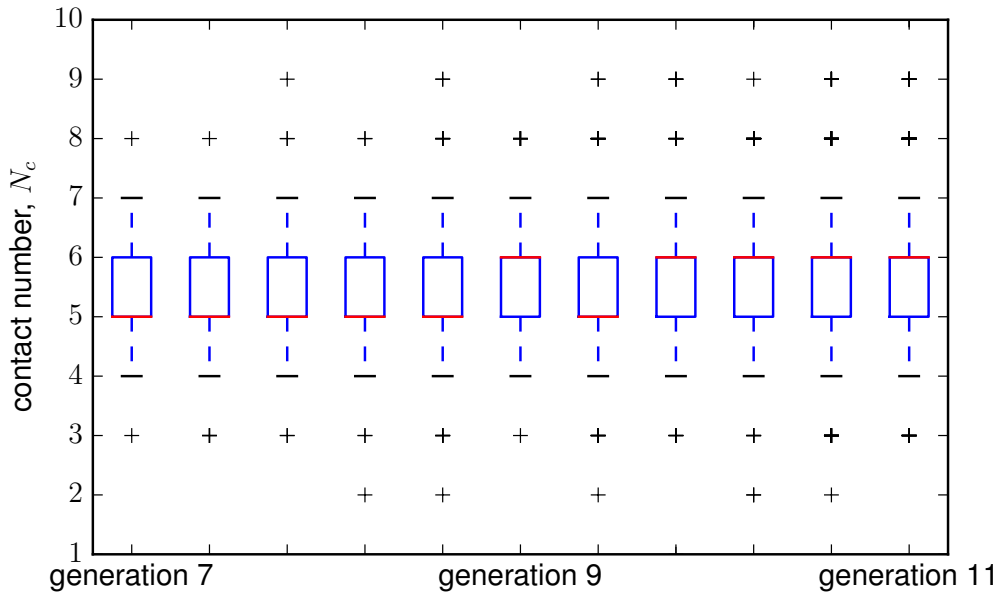
The contact angle distribution is skewed towards more parallel contact angles, as hinted by the inspection of Fig. 5.1. The contact number, as well as the contact angle distribution remain roughly constant over time, cf. Figure 5.5, where the data are displayed using box-plots. [1] This suggests that a chosen threshold angle c_t for community detection, is equally valid for the entire time span studied.

5.2.2 Pair-correlation function

To our knowledge, we are the first to publish the radial-distribution function of an *E. coli* colony, cf. Figure 5.6. The radial distribution function exhibits the typical behavior as found in fluids. It shows a pronounced first peak, corresponding to the closest neighbors in the direction of the particle's short axis. This first peak is followed by smaller peaks at the characteristic distances of the successive shells of neighbors. At large distances, $g(r)$ approaches unity, which indicates that the system has a homogeneous density. It is interesting that the position of the first peak slowly shifts towards smaller values of r as time increases, and that is in accordance with the previously described increase in density over time. However, we were not able to extract additional information, concerning microdomain properties from the radial distribution function.

[1] A box-plot provides information about the distribution of data. The center line indicates the median, the box indicates the lower and upper quartile, the whiskers describe the values outside the box, where the maximal length is limited to 1.5 times the inter-quartile range. Data-points further off are represented as outlier points.

(a)



(b)

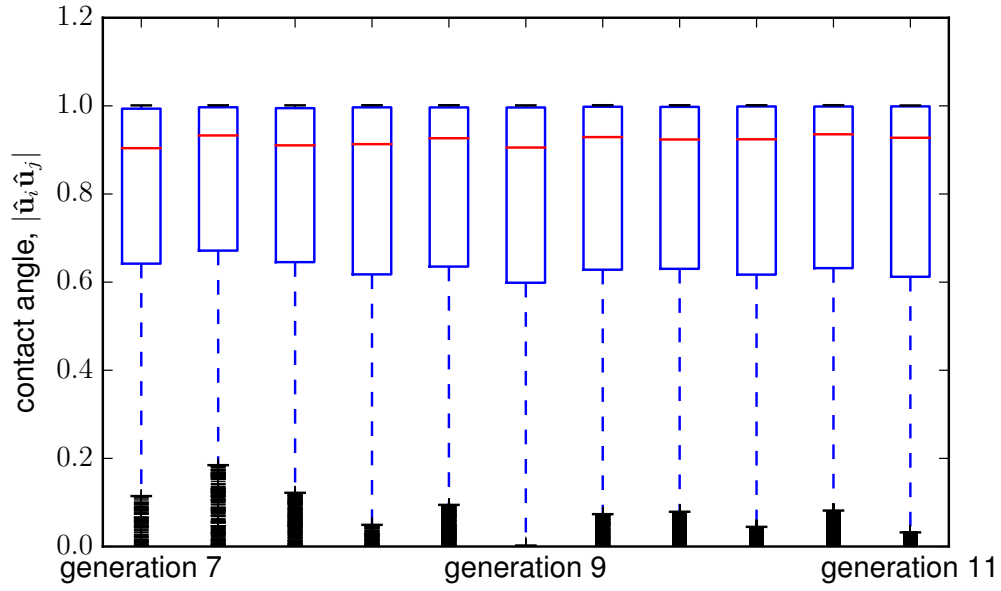
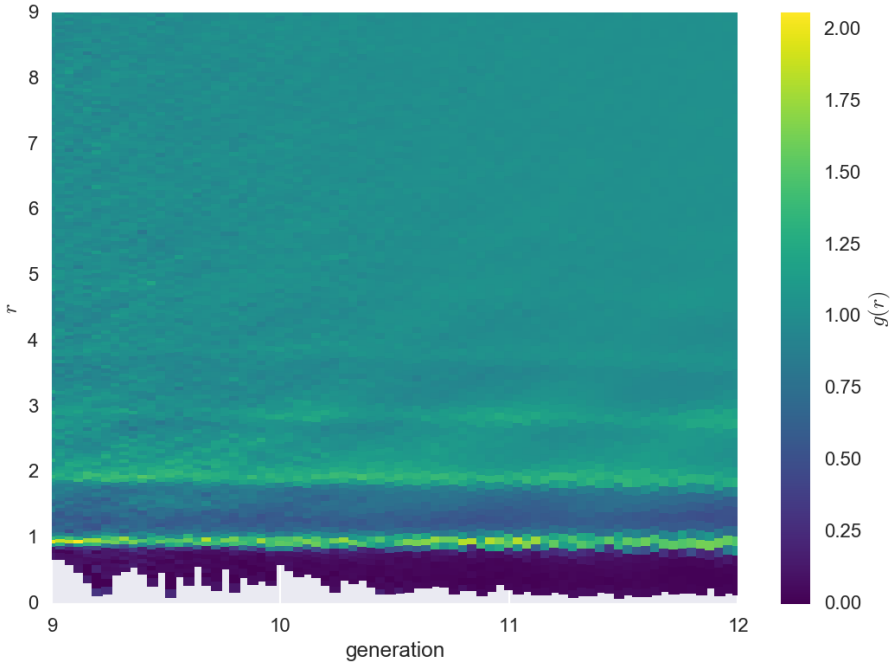


Figure 5.5: Time resolved distribution of (a) the contact number N_c, and (b) the contact angle $|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_i|$ in the bulk of colonies of particles with an average splitting length of $\langle l_{split} \rangle = 4$.

(a)



(b)

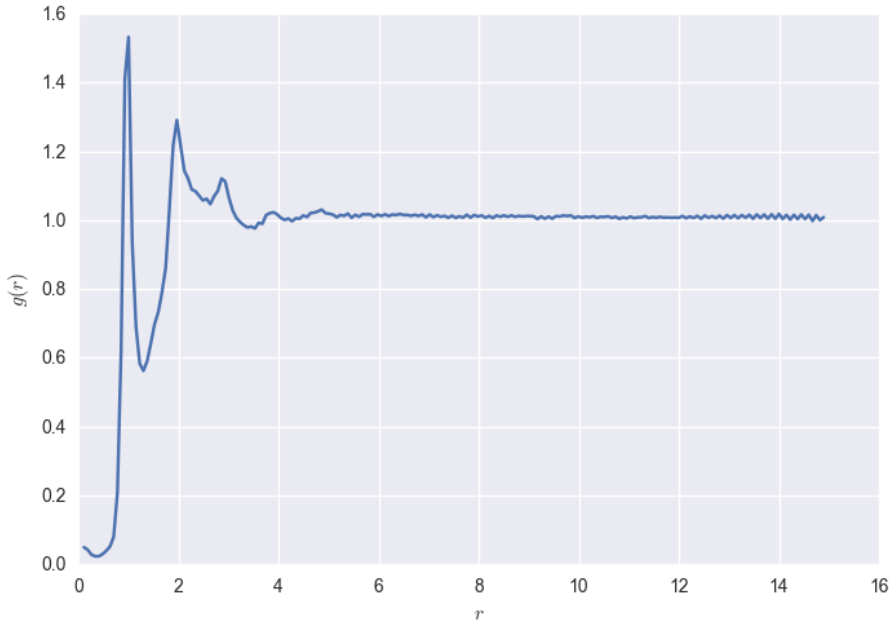


Figure 5.6: Time resolved radial distribution function $g(r)$ (a) for an entire time-range and (b) at a single point in time.

5.3 Microdomains

We identified microdomains using the methods introduced in Chapter 4 and compare the three different variants of network mappings: Heaviside, signum and linear. An impression of the isolated domains is provided in Figure 5.7. We use configurations from two simulation snapshots, that were colored according to partitionings using a set of threshold angles $c_t \in \{0.56, 0.86, 0.96\}$ and the linear network mapping (see Eq. 4.5, where we set $n = 1$). As can be seen in Fig. 5.7, a low threshold value ($c_t = 0.56$) yields large structures of low order, whereas the domain-sizes decrease and the order increases for higher threshold angles $c_t \in \{0.86, 0.96\}$.

From visual inspection only, we cannot decide for the optimal choice of threshold angle, neither can we provide a rationale for the most beneficial network mapping. In the following sections we want to present a method to infer possible answers to these open questions.

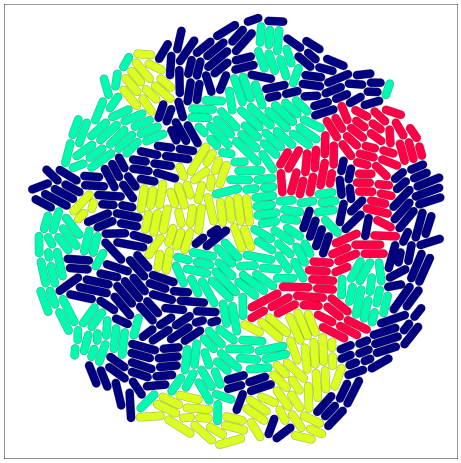
5.3.1 Similarity of partitions

The community detection algorithm is not deterministic and applying the algorithm on a single network will potentially yield different results (this is not the case for the Heaviside-mapping). To quantify the similarity of partitions obtained from different network-mappings of colonies in generation 9, we calculated the normalized mutual information, cf. Figure 5.8. For low threshold angles $c_t \lesssim 0.2$ the linear-mapping ($c_t \lesssim 0.1$ the signum-mapping) the colony is consistently partitioned into a single large community comprising all particles, therefore $NMI = 1$. For slightly higher threshold angles $c_t \in [0.3, 0.8]$ the linear-mapping ($c_t \in [0.1, 0.9]$) the partitions comprise multiple communities and become increasingly similar, for even higher threshold angles, the the signum and the linear mapping yield close to deterministic results.

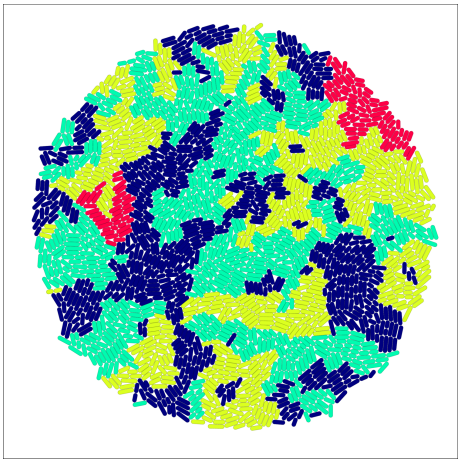
Inspecting the similarity of partitions obtained from different mappings, shows that the Heaviside-mapping yields drastically different partitions from the other two, and for very high threshold angles the partitions across mappings are almost identical, $NMI \approx 1$.

The important conclusion of this analysis is that, for low threshold angles c_t, the Heaviside-mapping yields a very unbalanced distribution of domain sizes, characterized by a single huge domain, and a number of much smaller domains. This can be easily seen from the consensus matrix, cf. Figure 5.9.

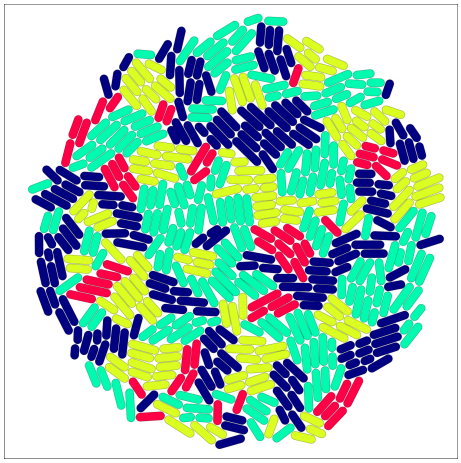
(a) Generation 9, $c_t = 0.56$



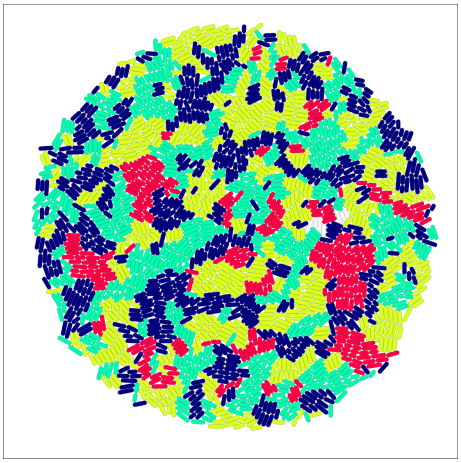
(b) Generation 11, $c_t = 0.56$



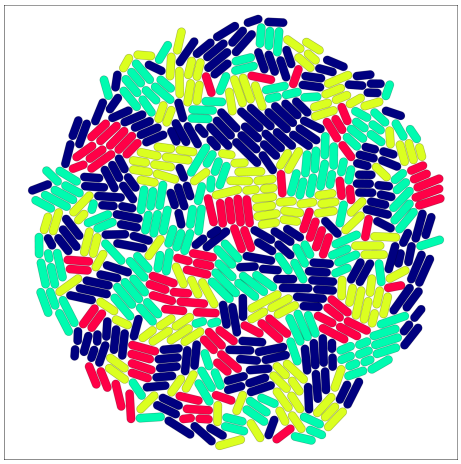
(c) $c_t = 0.86$



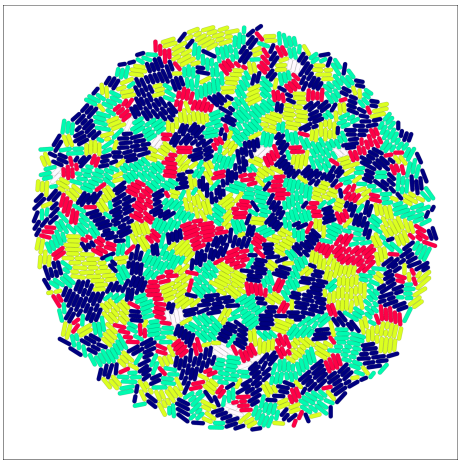
(d) $c_t = 0.86$



(e) $c_t = 0.96$



(f) $c_t = 0.96$



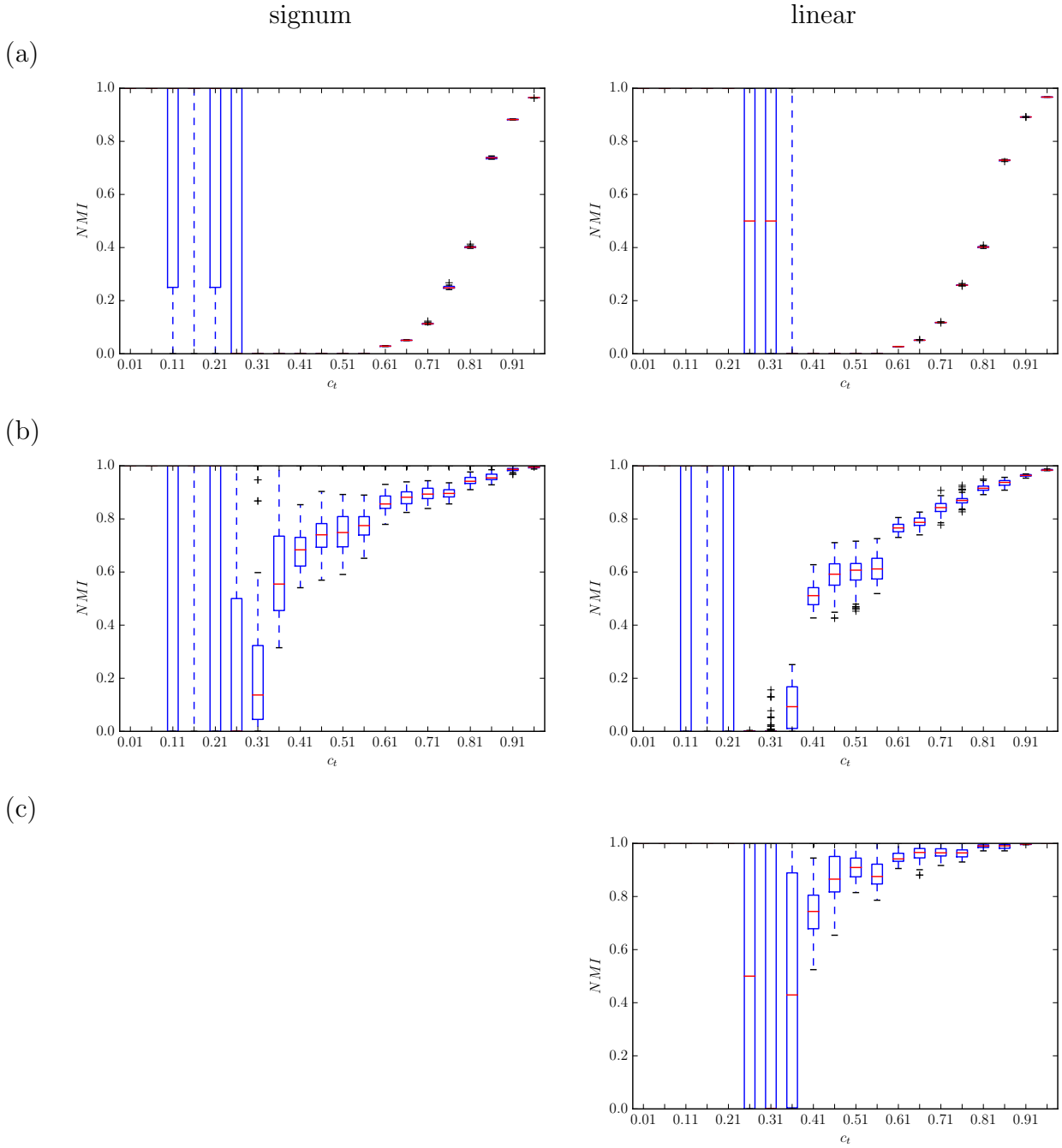


Figure 5.8: Boxplots showing the normalized mutual information of partitions of colonies in generation 9, obtained using the network mapping in the column-header, and (a) the Heaviside-, (b) the signum-, (c) the linear-network mapping. These are all the possible combinations, as the NMI is symmetric and the Heaviside domain detection is deterministic and will always lead to the same partitioning, i.e. $NMI = 1$, i.e. there are a total of 5 distinct combinations convey information.

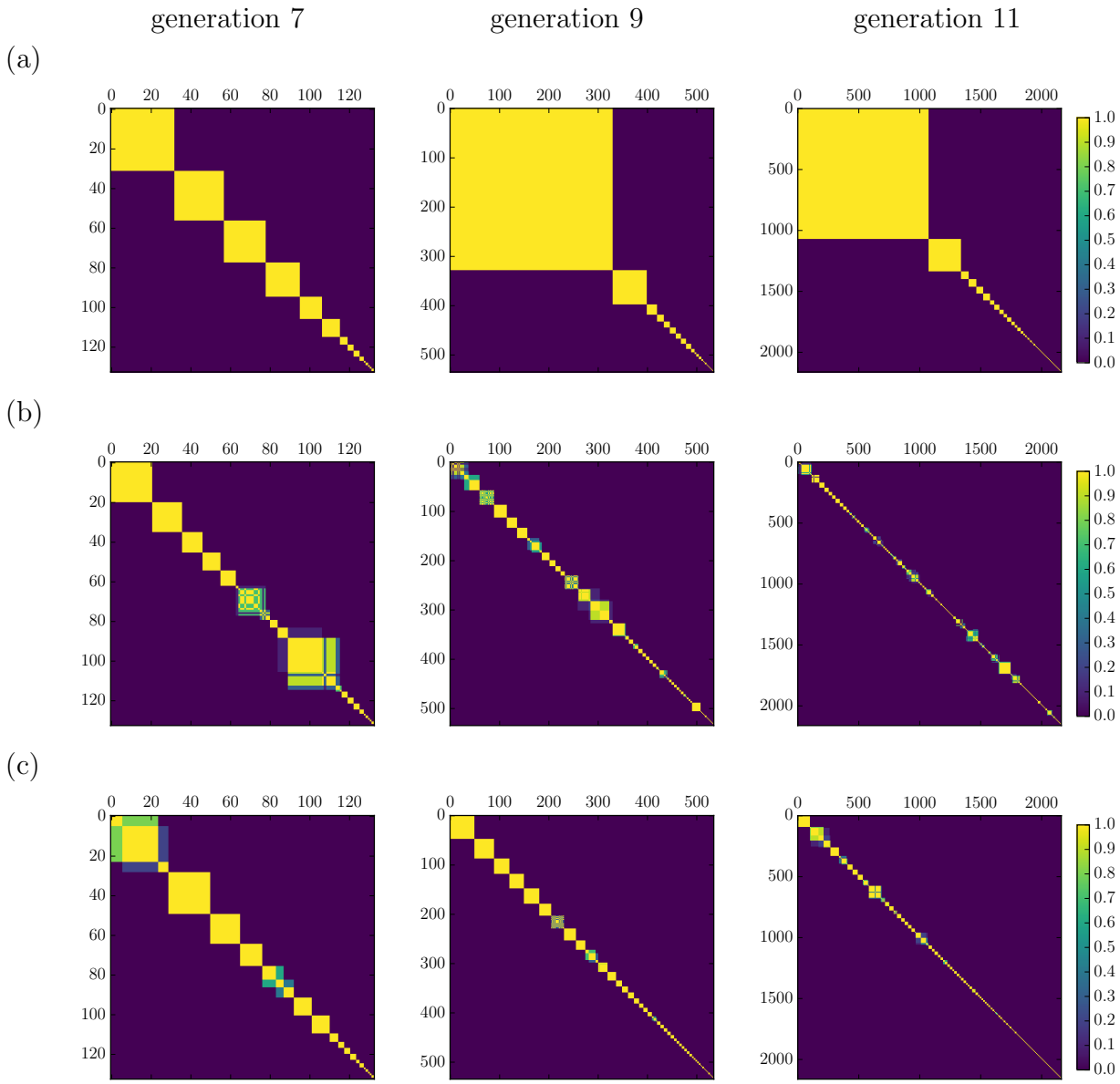


Figure 5.9: Consensus matrix obtained from (a) Heaviside-, (b) signum- and (c) linear network-mappings of colony snapshots at generation 7, generation 9, generation 11, with a threshold angle of $c_t = 0.81$.

5.4 Domain quality

The modularity function is only interpretable for the linear mapping, and does not show a clear elbow. The detailed time-evolution of the modularity function is provided in the appendix A.3. It is interesting to consider the average domain nematic order, weighted by the domain number-sizes. As can be seen from Figure 5.10, both the linear and the signum network mapping outperform the heaviside-mapping in identifying nematic domains. This is especially the case for lower threshold angles, e.g., when demanding a weaker local similarity criterium. Also for high threshold angles, the linear and the signum network mapping yield a higher average nematic order. Again, we found no clear cut elbow point, where the "linear" mapping at $c_t = 0.66$ most closely resembles one.

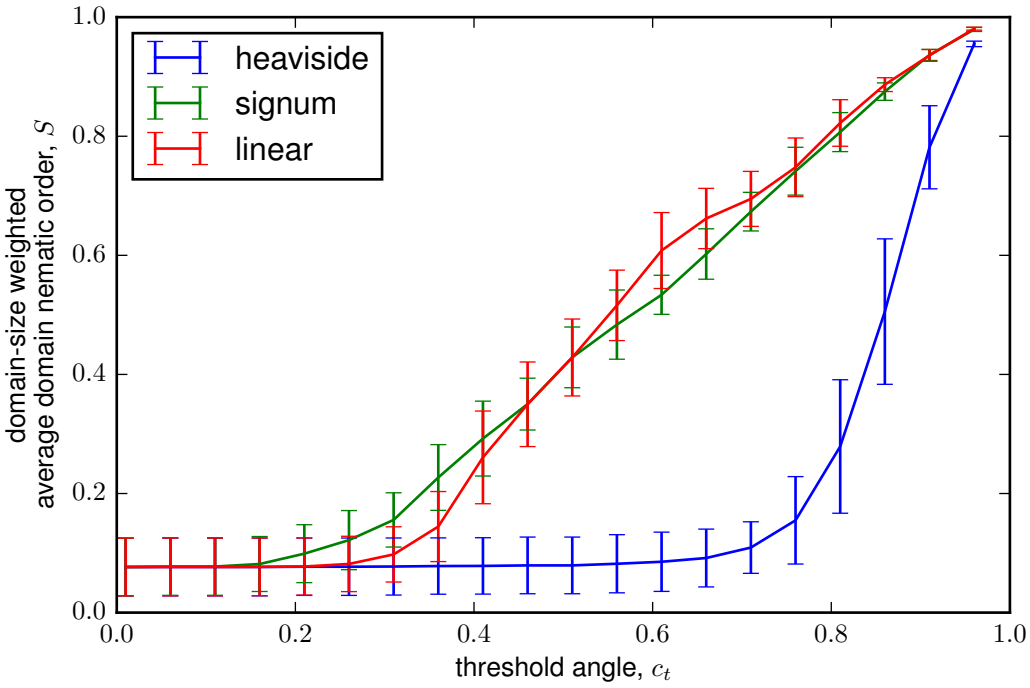


Figure 5.10: The averaged scalar nematic order for partitions obtained from linear network mappings of colonies in generation 9. The error-bars indicate the standard deviation.

5.5 Clustering index

Following a popular procedure to evaluate clusterings in machine learning applications, we calculated the silhouette index $\mathcal{C}$, and is shown in Fig. 5.11. The maximum indicates the best ratio of intra-domain-distance over adjacent-domain-distance and

occurs for both the linear- and signum- mappings at a threshold-angle of $c_t = 0.96$. For the Heaviside-mapping no global maximum was observed for $c_t \leq 0.98$, but only a local maximum at $c_t = 0.81$, that coincides with a local minimum for the linear mapping, cf. Figure 5.11.

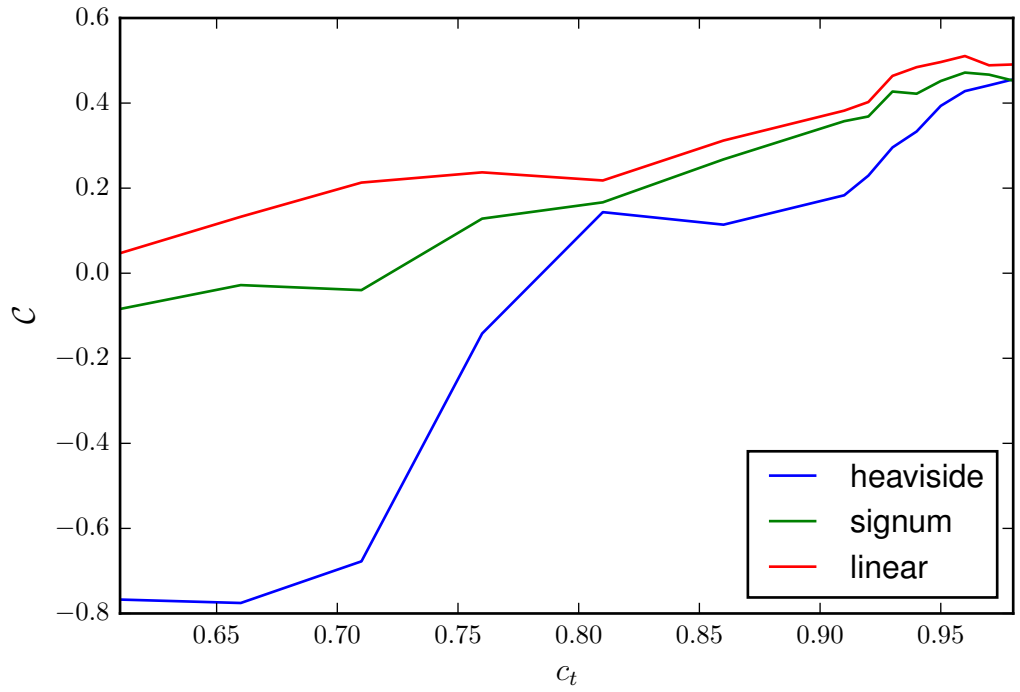


Figure 5.11: The silhouette index for partitions of colonies in generation 9 shows a maximum for the linear- and the signum-mapping at $c_t = 0.96$.

5.6 Domain sizes

We estimate the size of the nematic features in a colony by the first zero-crossing of the the equal-time polar-pair correlation function C_2. These zero-crossings show a growing trend with time, cf. Figure 5.12. This growing trend can also be found for extensive properties of the domains that were isolated with the Heaviside-, signum- and linear-network mapping. We calculated the average particle count per domain and the average radius of the smallest domain-enclosing circle, cf. Figure 5.13. We calculated the Pearson's correlation coefficient r between the different size-measures at a range of threshold angles c_t. We found moderate correlation ($r \approx 0.5$) between the mean domain number-size N_σ and C_2, cf. Figure A.2 in the appendix A.4, strong correlation ($r \approx 0.7$) between the radius of smallest domain-enclosig circle r_σ and C_2, cf. Figure 5.14, and very strong correlation ($r \approx 1.0$) across the network mapping variants measuring N_σ and r_σ respectively (see Fig. 5.14). Furthermore,

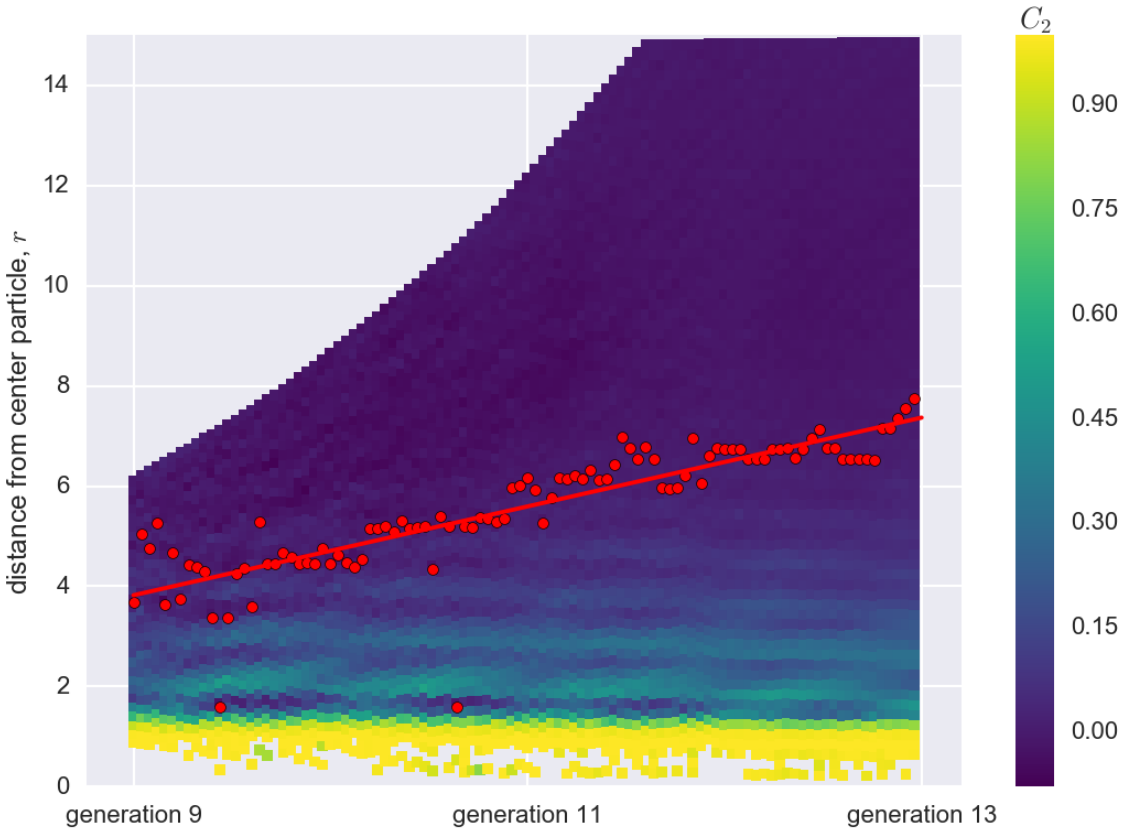


Figure 5.12: Equal-time polar-pair function at sequential times. The red dots indicate the first zero-crossing of the function that serves as an estimate for the size of the nematic features in the colony. The red line is the best linear function fit through the zero-crossing points.

we performed a one-way analysis of variance comparing the domain-sizes found in the colonies with resampled networks that have the same topology but randomly redistributed edge-weights. The domains in the colonies are significantly larger ($p \ll 1$, for all times, methods and contact-angle thresholds) than domains isolated in the resampled networks. This indicates a non-trivial relationship between the contact-network topology and the contact-angles. Thus, it will be necessary to quantify this relationship, in order to design a generative model for colony networks.

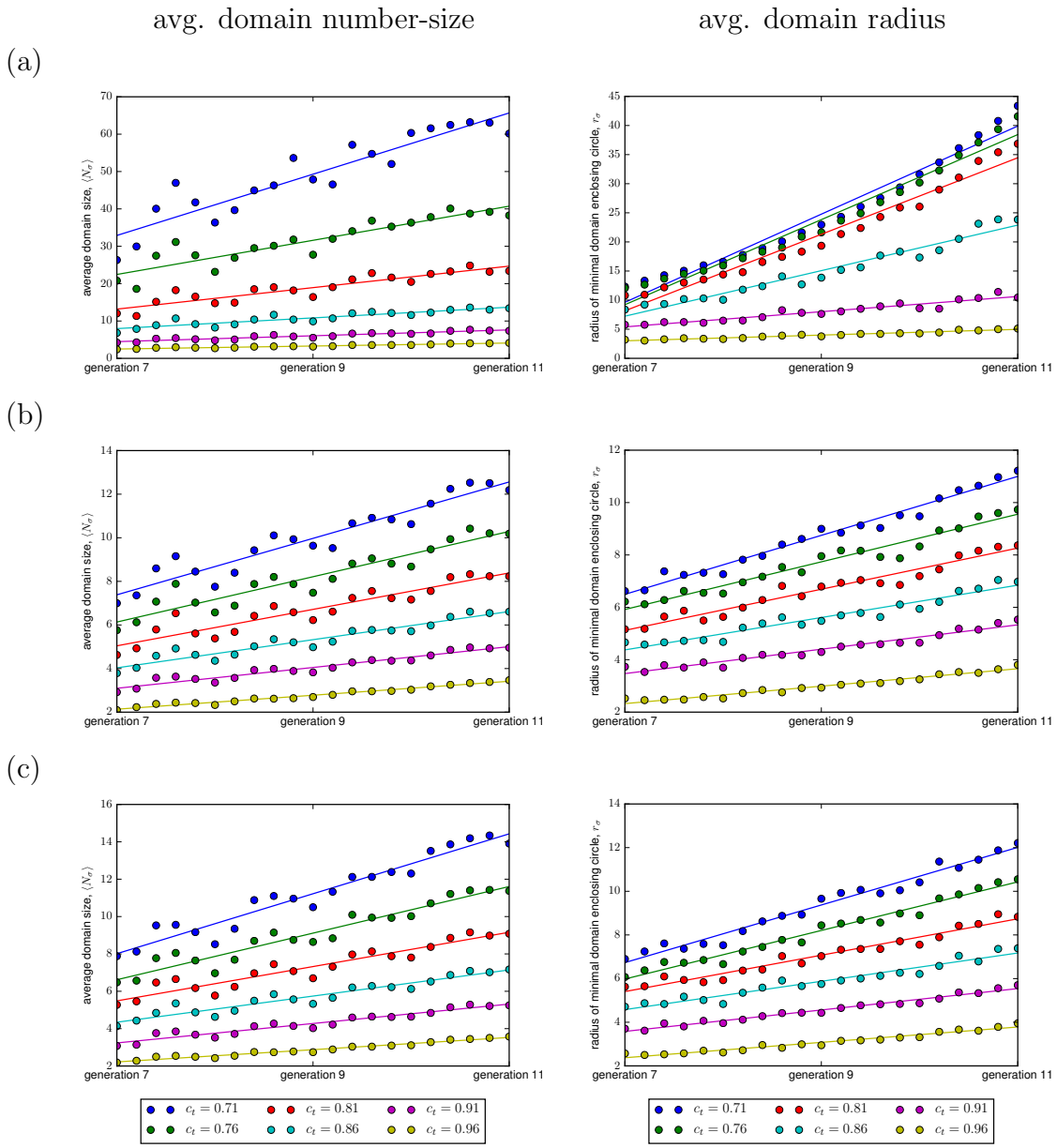


Figure 5.13: Average domain number-size and radius of smallest domain-enclosing circle for domains of obtained from (a) Heaviside-, (b) signum- and (c) linear-network mappings.

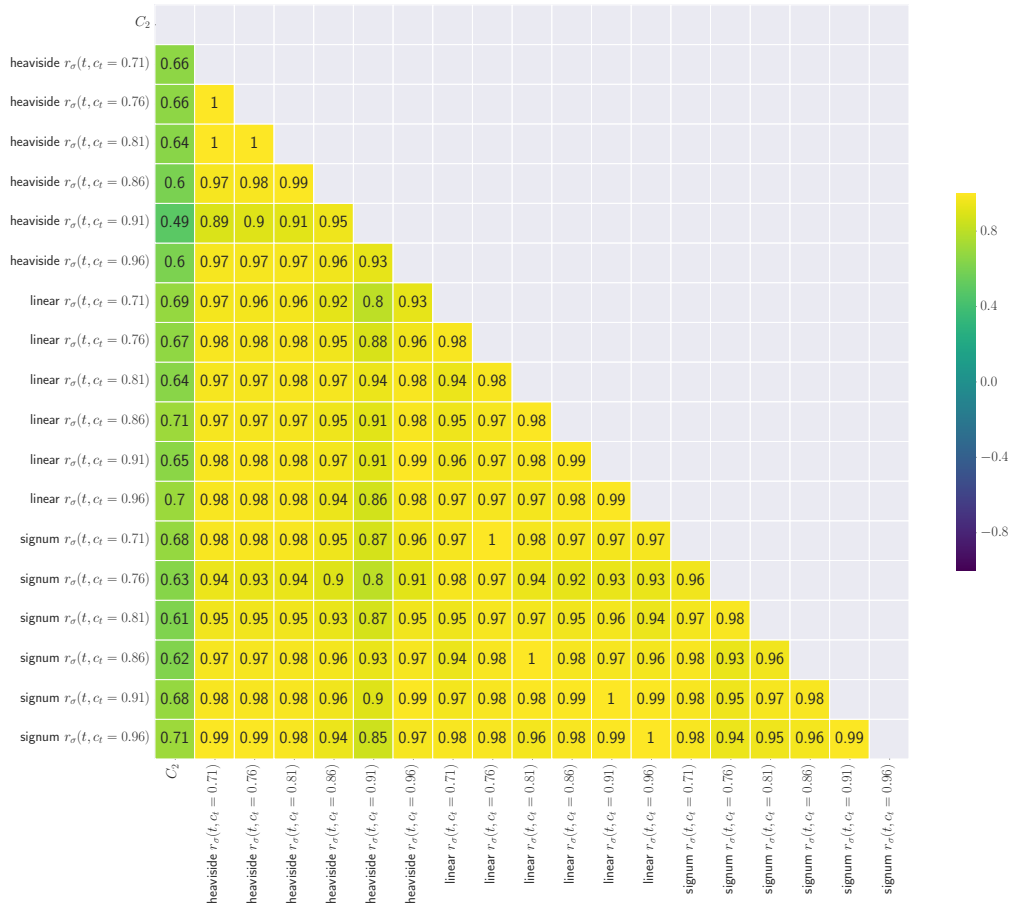


Figure 5.14: Correlation matrix, showing the Pearson correlation coefficient r of ensemble-averaged time series of the equal-time polar-pair correlation C_2 and the radius of the smallest domain-enclosing circle r_σ.

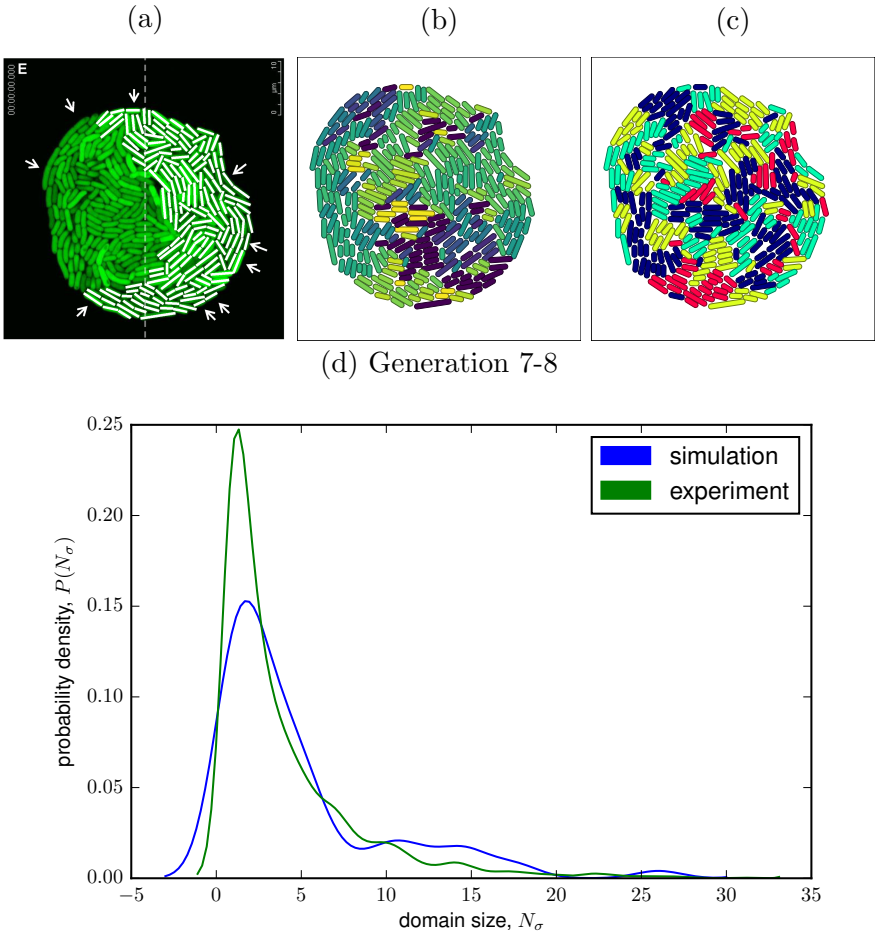
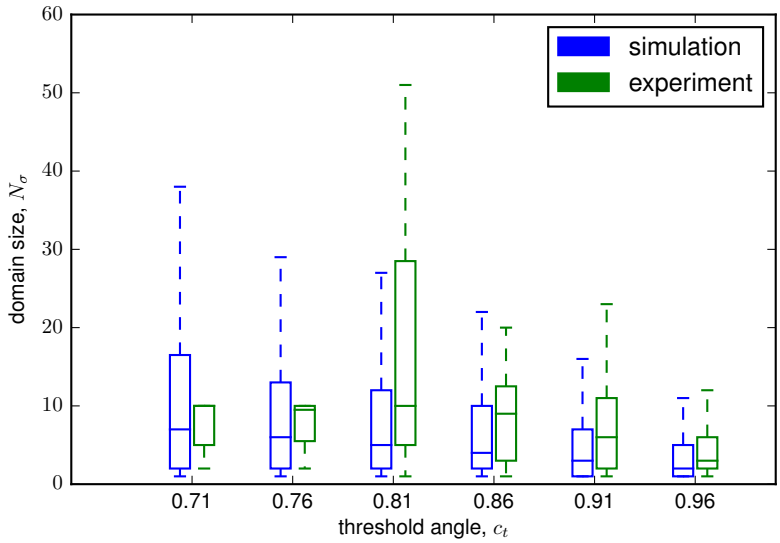


Figure 5.15: (a-c) Extraction of *E. coli* coordinates from light microscopy image [4], and (d) the probability density functions for community sizes in experimentally observed and simulated colonies, at a threshold angle of $c_t = 0.96$ and linear network mapping.

5.7 Comparison with experimental data

Exemplarily, we compare the microdomain structure obtained from the numerical model with data extracted from previously published light microscopy images of *E. coli* colonies [4] (cf. Figure 1.2). We manually tagged two light microscopy images with the `ImageJ` software [115, 116] and find that the domain size distributions in experimental data for different threshold angles c_t and linear network mapping are in reasonable agreement with the domain size distributions found in simulated colonies of similar size, cf. Figure 5.15 and Figure 5.16.

(a) Generation 7-8



(b) Generation 8-9

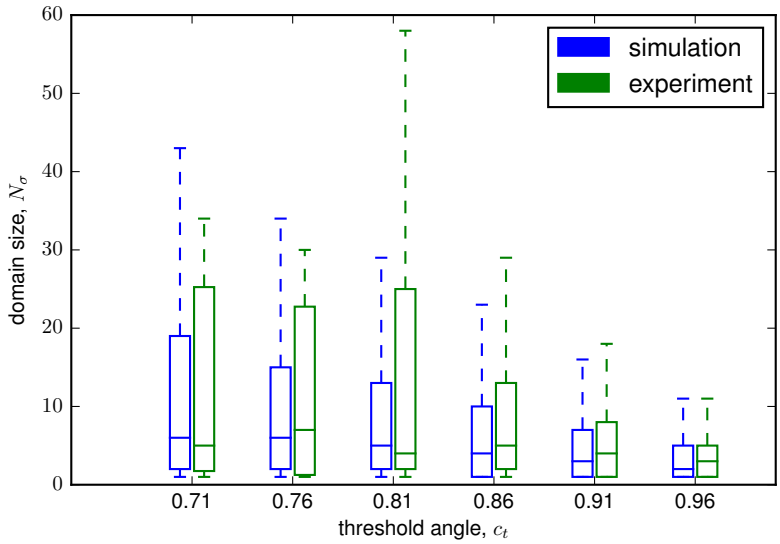


Figure 5.16: Comparison of microdomain size distributions in simulation and experiment for a range of threshold angles c_t and linear network mapping, at two different colony generations, (a) generation 7-8 and (b) generation 8-9.

Chapter 6

Conclusion

In this thesis, we studied the growth of sessile *Escherichia coli* colonies in the exponential growth phase (*feasting*), by performing Discrete Element Simulations in two dimensions. We have found that mechanical interactions are sufficient for the formation of highly ordered mesoscopic structures, *vulgo* microdomains. The basic tools of analysis, such as the contact angle distribution or the radial distribution function do not indicate the formation of these structures. For this purpose we employed a modularity maximizing community detection algorithm on contact network representations of the colonies. Hereby, the particles are represented by nodes, forming edges with particles at contact. The edge-weight is determined from the contact-angle and the community detection algorithm partitions the network into cohesive subgraphs for microdomains. We compared three different variants (defining the modularity based on Heaviside-, Signum- and Linear-mappings) with a range of threshold angles, and evaluated their overall performance and correlation with more conventional measures. We found that a good threshold angle to discriminate "cohesive" from "repulsive" contacts between particles i and j, to be $c_t = |\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j| \approx 0.96$. For high threshold angles, all three variants performed comparably well, whereas the simple dichotomization variant (*i.e.* Heaviside based modularity) was clearly outperformed for lower threshold angles. Eventually, we compared the microdomain structure of previously reported experimental data with our simulations and found decent agreement. This opens new avenues to understanding morphogenesis and tissue growth. Beyond the field of morphogenesis, this methodology can be applied in related fields, for instance to study systems of anisometric particles.

Only for the last decade bacterial colonies have been studied using Discrete Element Simulations and, as indicated in the introduction, a myriad of different research pathways appear fruitful. Immediately linked to our present work, we can evaluate more elaborate domain-separation mechanisms, or in detail trace the evolution of individual domains; both are outlined in the Appendix. We have a variety of additional measures at hand to quantify the properties of microdomains

(e.g. Minkowski metrics, network diameter, ...). We can study the mesoscopic features of three dimensional colonies, by allowing particle to buckle into subsequent layers. Also the effects of resource limitations, chemical interactions, growth-limiting pressure, apoptosis, external confinement or explicit hydrodynamic interactions can be incorporated into the existing model.

Appendix A

Appendix

A.1 Modularity for mutli-layer networks

Over the course of time communities may grow or shrink, they may merge with each other or split into smaller clusters, or do all of the above. To study the dynamics of communities, Mucha *et al.* extended the modularity quality function, to study the temporal evolution of communities [111].

$$Q_{temporal} = \sum_{ijsr} \{ (A_{ijs} - p_{ijs}) \, \delta_{sr} + \delta_{ij} C_{jsr} \} \delta \left(\sigma_{is}, \sigma_{jr} \right), \tag{A.1}$$

where the matrix C_{jsr}, couples nodes with themselves across network snapshots at times s and r.

Choosing the most truthful value for C_{jsr} is similarly to choosing c_t a non-trival task, that is left for future studies. However, interesting phenomena can occur: sufficiently high time cross-layer correlations C_{jsr} can yield spatially seperated subdomains (i.e. domains with enclaves).

A.2 Logistic sigmoidal function

Next to the in detail-studied network representations with heaviside-, signum- and linear-edge weights, several other functions are plausible. One such possible option is a logistic sigmoidal function, that is frequently used for decision-problems in machine learning applications [117]. This approach combines features of the linear and the signum edge weights: it is continuous and allows to finely tune a threshold,

$$J_{ij} = \mathcal{H}(\delta_{ij} - \delta_t) \left(\frac{1}{1 + \exp\left(-a(|\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j| - c_t)\right)} - \frac{1}{2} \right). \tag{A.2}$$

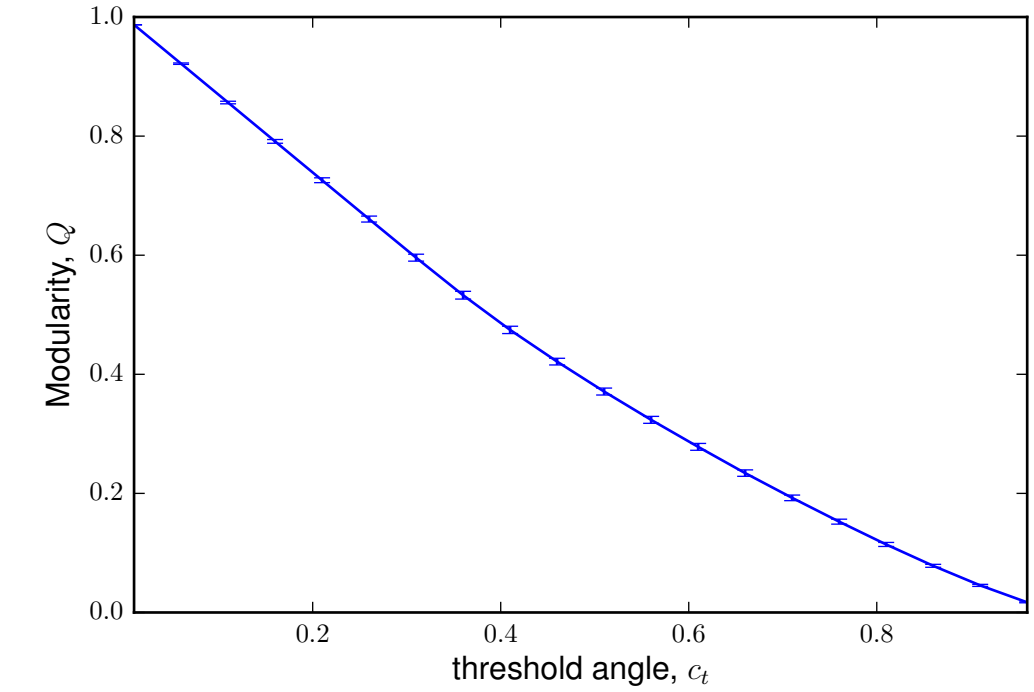
The additional parameter a tunes the steepness of the curve, and has to be chosen appropriately.

A.3 Time evolution of the modularity Q

The modularity Q does not show an elbow, and therefore no indication for an optimal threshold angle c_t, c.f. figure A.1a. The modularity slightly increases with colony age, c.f. figure A.1.

A.4 Correlation of equal-time polar-pair correlation C_2 and the average domain number-size N_σ

(a)



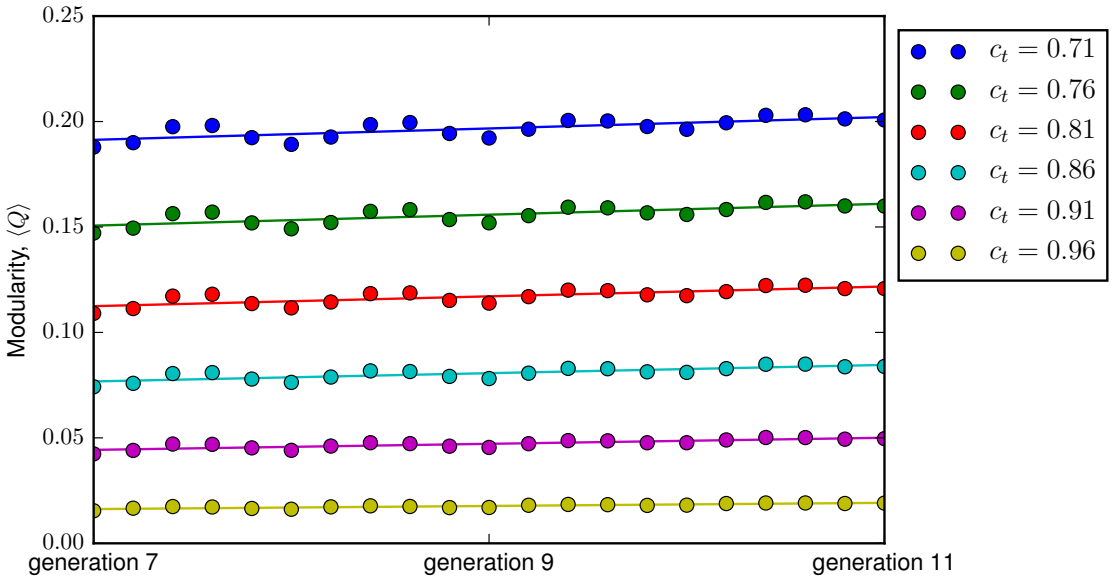


Figure A.1: The modularity Q of partitions obtained from a linear network mapping (a) for a single time-stamp (b) for a sequence of time-stamps.

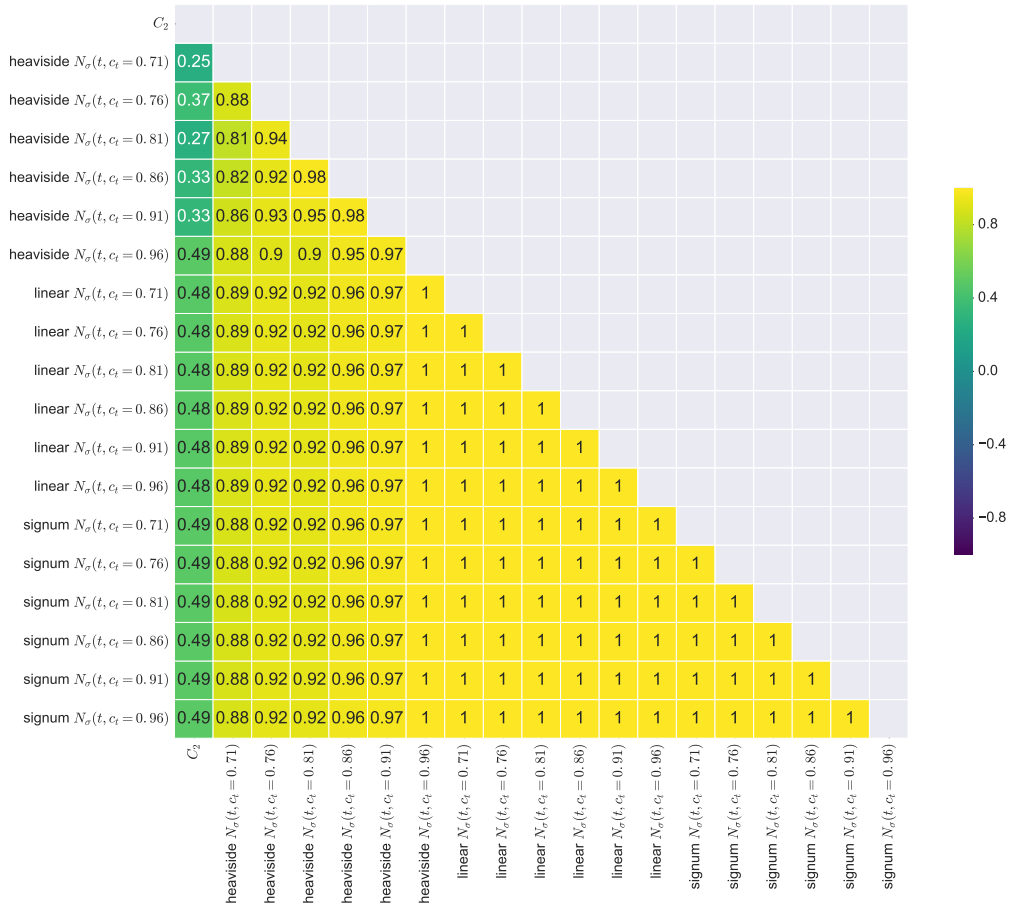


Figure A.2: Correlation matrix, showing the Pearson correlation coefficient r of ensemble-averaged time series of the equal-time polar-pair correlation C_2 and the domain number-size N_σ.

Bibliography

[1] J. A. Shapiro, *The significances of bacterial colony patterns*, Bioessays **17**, 597 (1995).

[2] Pratiksha Pokhrel, *Differences between fimbriae and pili*, [Online; accessed June 9, 2017] (2016).

[3] H. Vlamakis, Y. Chai, P. Beauregard, R. Losick, and R. Kolter, *Sticking together: building a biofilm the bacillus subtilis way*, Nature Reviews Microbiology **11**, 157 (2013).

[4] P.-T. Su, C.-T. Liao, J.-R. Roan, S.-H. Wang, A. Chiou, and W.-J. Syu, *Bacterial colony from two-dimensional division to three-dimensional development*, PLoS One **7**, e48098 (2012).

[5] J. Wingender, T. R. Neu, and H.-C. Flemming, *What are bacterial extracellular polymeric substances?*, in *Microbial extracellular polymeric substances*, 1–19, (Springer 1999).

[6] D. Volfson, S. Cookson, J. Hasty, and L. S. Tsimring, *Biomechanical ordering of dense cell populations*, Proceedings of the National Academy of Sciences USA **105**, 15346 (2008).

[7] I. W. Sutherland, *The biofilm matrix–an immobilized but dynamic microbial environment*, Trends in Microbiology **9**, 222 (2001).

[8] T. Vicsek and A. Zafeiris, *Collective motion*, Physics Reports **517**, 71 (2012).

[9] J. D. Murray, *Mathematical Biology. II Spatial Models and Biomedical Applications*, (Springer 2001).

[10] J. Hacker and G. Blum-Oehler, *In appreciation of theodor escherich* (2007).

[11] H. Cho, H. Jönsson, K. Campbell, P. Melke, J. W. Williams, B. Jedynak, A. M. Stevens, A. Groisman, and A. Levchenko, *Self-organization in high-density bacterial colonies: efficient crowd control*, PLoS Biology **5**, e302 (2007).

[12] Z. You, D. J. G. Pearce, A. Sengupta, and L. Giomi, *Geometry and mechanics of micro-domains in growing bacterial colonies*, arXiv:1703.04504 (2017).

[13] A. Browet, P.-A. Absil, and P. Van Dooren, *Community detection for hierarchical image segmentation*, in *International Workshop on Combinatorial Image Analysis*, 358–371, (Springer 2011).

[14] N. Tarcai, C. Virágh, D. Ábel, M. Nagy, P. L. Várkonyi, G. Vásárhelyi, and T. Vicsek, *Patterns, transitions and the role of leaders in the collective dynamics of a simple robotic flock*, Journal of Statistical Mechanics: Theory and Experiment **2011**, P04010 (2011).

[15] N. J. Suematsu, S. Nakata, A. Awazu, and H. Nishimori, *Collective behavior of inanimate boats*, Physical Review E **81**, 056210 (2010).

[16] G. Popkin, *The physics of life*, Nature **529**, 16 (2016).

[17] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, *Novel type of phase transition in a system of self-driven particles*, Physical Review Letters **75**, 1226 (1995).

[18] H. C. Berg, *Motile behavior of bacteria*, Physics Today **53**, 24 (2000).

[19] H. C. Berg, D. A. Brown, et al., *Chemotaxis in escherichia coli analysed by three-dimensional tracking*, Nature **239**, 500 (1972).

[20] K. Maeda, Y. Imae, J.-I. Shioi, and F. Oosawa, *Effect of temperature on motility and chemotaxis of escherichia coli.*, Journal of Bacteriology **127**, 1039 (1976).

[21] E. M. Purcell, *Life at low reynolds number*, American Journal of Physics **45**, 3 (1977).

[22] M. C. Marchetti, J. Joanny, S. Ramaswamy, T. Liverpool, J. Prost, M. Rao, and R. A. Simha, *Hydrodynamics of soft active matter*, Reviews of Modern Physics **85**, 1143 (2013).

[23] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, and C. Bechinger, *Active brownian motion tunable by light*, Journal of Physics: Condensed Matter **24**, 284129 (2012).

[24] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, *Active particles in complex and crowded environments*, Reviews of Modern Physics **88**, 045006 (2016).

[25] D. Smith, E. Gaffney, J. Blake, and J. Kirkman-Brown, *Human sperm accumulation near surfaces: a simulation study*, Journal of Fluid Mechanics **621**, 289 (2009).

[26] J. Elgeti, U. B. Kaupp, and G. Gompper, *Hydrodynamics of sperm cells near surfaces*, Biophysical Journal **99**, 1018 (2010).

[27] D. J. Smith, E. Gaffney, H. Shum, H. Gadêlha, and J. Kirkman-Brown, *Comment on the article by j. elgeti, ub kaupp, and g. gompper: hydrodynamics of sperm cells near surfaces* (2011).

[28] J. Elgeti, U. B. Kaupp, and G. Gompper, *Response to comment on article: Hydrodynamics of sperm cells near surfaces* (2011).

[29] J. Hu, M. Yang, G. Gompper, and R. G. Winkler, *Modelling the mechanics and hydrodynamics of swimming e. coli*, Soft Matter **11**, 7867 (2015).

[30] H. Stark, *Swimming in external fields*, The European Physical Journal Special Topics **11**, 2369 (2016).

[31] R. Van der Sman, *Simulations of confined suspension flow at multiple length scales*, Soft Matter **5**, 4376 (2009).

[32] R. Kapral, *Multiparticle collision dynamics: simulation of complex systems on mesoscales*, Advances in Chemical Physics **140**, 89 (2008).

[33] G. Gompper, T. Ihle, D. Kroll, and R. Winkler, *Multi-particle collision dynamics: a particle-based mesoscale simulation approach to the hydrodynamics of complex fluids*, Advanced computer simulation approaches for soft matter sciences III 1–87 (2009).

[34] J. Padding and A. Louis, *Hydrodynamic interactions and brownian forces in colloidal suspensions: Coarse-graining over time and length scales*, Physical Review E **74**, 031402 (2006).

[35] M. Cates, *Diffusive transport without detailed balance in motile bacteria: does microbiology need statistical physics?*, Reports on Progress in Physics **75**, 042601 (2012).

[36] A. Kaiser, H. Wensink, and H. Löwen, *How to capture active particles*, Physical Review Letters **108**, 268307 (2012).

[37] N. Koumakis, A. Lepore, C. Maggi, and R. Di Leonardo, *Targeted delivery of colloids by swimming bacteria*, Nature Communications **4** (2013).

[38] S. E. Hulme, W. R. DiLuzio, S. S. Shevkoplyas, L. Turner, M. Mayer, H. C. Berg, and G. M. Whitesides, *Using ratchets and sorters to fractionate motile cells of escherichia coli by length*, Lab on a Chip **8**, 1888 (2008).

[39] R. Di Leonardo, L. Angelani, D. Dell'Arciprete, G. Ruocco, V. Iebba, S. Schippa, M. Conte, F. Mecarini, F. De Angelis, and E. Di Fabrizio, *Bacterial ratchet motors*, Proceedings of the National Academy of Sciences **107**, 9541 (2010).

[40] F. Chang and K. C. Huang, *How and why cells grow as rods*, BMC Biology **12**, 54 (2014).

[41] K. D. Young, *Bacterial shape: two-dimensional questions and possibilities*, Annual Review of Microbiology **64**, 223 (2010).

[42] T. S. Ursell, J. Nguyen, R. D. Monds, A. Colavin, G. Billings, N. Ouzounov, Z. Gitai, J. W. Shaevitz, and K. C. Huang, *Rod-like bacterial shape is maintained by feedback between cell curvature and cytoskeletal localization*, Proceedings of the National Academy of Sciences **111**, E1025 (2014).

[43] S. X. Sun and H. Jiang, *Physics of bacterial morphogenesis*, Microbiology and Molecular Biology Reviews **75**, 543 (2011).

[44] J. Männik, R. Driessen, P. Galajda, J. E. Keymer, and C. Dekker, *Bacterial growth and motility in sub-micron constrictions*, Proceedings of the National Academy of Sciences **106**, 14861 (2009).

[45] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, *Robust growth of escherichia coli*, Current Biology **20**, 1099 (2010).

[46] J. D. Wang and P. A. Levin, *Metabolism, cell growth and the bacterial cell cycle*, Nature Reviews. Microbiology **7**, 822 (2009).

[47] T. J. Rudge, F. Federici, P. J. Steiner, A. Kan, and J. Haseloff, *Cell polarity-driven instability generates self-organized, fractal patterning of cell layers*, ACS Synthetic Biology **2**, 705 (2013).

[48] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson, *Genetic drift at expanding frontiers promotes gene segregation*, Proceedings of the National Academy of Sciences **104**, 19926 (2007).

[49] M. A. Grant, B. Wacław, R. J. Allen, and P. Cicuta, *The role of mechanical forces in the planar-to-bulk transition in growing escherichia coli microcolonies*, Journal of The Royal Society Interface **11**, 20140400 (2014).

[50] D. Boyer, W. Mather, O. Mondragón-Palomino, S. Orozco-Fuentes, T. Danino, J. Hasty, and L. S. Tsimring, *Buckling instability in ordered bacterial colonies*, Physical Biology **8**, 026008 (2011).

[51] O. Hallatschek and D. R. Nelson, *Life at the front of an expanding population*, Evolution **64**, 193 (2010).

[52] F. D. C. Farrell, O. Hallatschek, D. Marenduzzo, and B. Waclaw, *Mechanically driven growth of quasi-two-dimensional microbial colonies*, Physical Review Letters **111**, 168101 (2013).

[53] C. Giverso, M. Verani, and P. Ciarletta, *Branching instability in expanding bacterial colonies*, Journal of The Royal Society Interface **12**, 20141290 (2015).

[54] D. Fusco, M. Gralka, J. Kayser, A. Anderson, and O. Hallatschek, *Excess of mutational jackpot events in expanding populations revealed by spatial luria-delbruck experiments*, Nature Communications **7** (2016).

[55] M. Gralka, F. Stiewe, F. Farrell, W. Moebius, B. Waclaw, and O. Hallatschek, *Allele surfing promotes microbial adaptation from standing variation*, Ecology Letters **19**, 889 (2016).

[56] T. Danino, O. Mondragon-Palomino, L. Tsimring, and J. Hasty, *A synchronized quorum of genetic clocks*, Nature **463**, 326 (2010).

[57] D. G. Allison and I. W. Sutherland, *The role of exopolysaccharides in adhesion of freshwater bacteria*, Microbiology **133**, 1319 (1987).

[58] J. S. Madsen, M. Burmølle, L. H. Hansen, and S. J. Sørensen, *The interconnection between biofilm formation and horizontal gene transfer*, FEMS Immunology & Medical Microbiology **65**, 183 (2012).

[59] P. N. Danese, L. A. Pratt, and R. Kolter, *Exopolysaccharide production is required for development of escherichia coli k-12 biofilm architecture*, Journal of Bacteriology **182**, 3593 (2000).

[60] B. Frølund, R. Palmgren, K. Keiding, and P. H. Nielsen, *Extraction of extracellular polymers from activated sludge using a cation exchange resin*, Water Research **30**, 1749 (1996).

[61] X. Zhang, P. L. Bishop, and M. J. Kupferle, *Measurement of polysaccharides and proteins in biofilm extracellular polymers*, Water Science and Technology **37**, 345 (1998).

[62] S. Widder, R. J. Allen, T. Pfeiffer, T. P. Curtis, C. Wiuf, W. T. Sloan, O. X. Cordero, S. P. Brown, B. Momeni, W. Shou, et al., *Challenges in microbial ecology: building predictive understanding of community function and dynamics*, The ISME Journal **10**, 2557 (2016).

[63] A. E. Little, C. J. Robinson, S. B. Peterson, K. F. Raffa, and J. Handelsman, *Rules of engagement: interspecies interactions that regulate microbial communities*, Annual Review Microbiology **62**, 375 (2008).

[64] D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot, *Genomic evolution during a 10,000-generation experiment with bacteria*, Proceedings of the National Academy of Sciences **96**, 3807 (1999).

[65] M. G. Mazza, *The physics of biofilms—an introduction*, Journal of Physics D: Applied Physics **49**, 203001 (2016).

[66] J. W. Wimpenny and R. Colasanti, *A unifying hypothesis for the structure of microbial biofilms based on cellular automaton models*, FEMS Microbiology Ecology **22**, 1 (1997).

[67] Q. Wang and T. Zhang, *Review of mathematical models for biofilms*, Solid State Communications **150**, 1009 (2010).

[68] T. E. Gorochowski, A. Matyjaszkiewicz, T. Todd, N. Oak, K. Kowalska, S. Reid, K. T. Tsaneva-Atanasova, N. J. Savery, C. S. Grierson, and M. Di Bernardo, *Bsim: an agent-based tool for modeling bacterial populations in systems and synthetic biology*, PloS One **7**, e42790 (2012).

[69] C. Picioreanu, M. C. Van Loosdrecht, J. J. Heijnen, et al., *Effect of diffusive and convective substrate transport on biofilm structure formation: a two-dimensional modeling study*, Biotechnology and Bioengineering **69**, 504 (2000).

[70] L. A. Lardon, B. V. Merkey, S. Martins, A. Dötsch, C. Picioreanu, J.-U. Kreft, and B. F. Smets, *idynomics: next-generation individual-based modelling of biofilms*, Environmental Microbiology **13**, 2416 (2011).

[71] J. N. Wilking, T. E. Angelini, A. Seminara, M. P. Brenner, and D. A. Weitz, *Biofilms as complex fluids*, MRS bulletin **36**, 385 (2011).

[72] E. Alpkvist and I. Klapper, *Description of mechanical response including detachment using a novel particle model of biofilm/flow interaction*, Water Science and Technology **55**, 265 (2007).

[73] A. E. Ehret and M. Böl, *Modelling mechanical characteristics of microbial biofilms by network theory*, Journal of The Royal Society Interface **10**, 20120676 (2013).

[74] J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, et al., *Data-analysis strategies for image-based cell profiling*, Nature Methods **14**, 849 (2017).

[75] R. Rusconi, M. Garren, and R. Stocker, *Microfluidics expanding the frontiers of microbial ecology*, Annual Review of Biophysics **43**, 65 (2014).

[76] W. Mather, O. Mondragón-Palomino, T. Danino, J. Hasty, and L. S. Tsimring, *Streaming instability in growing cell populations*, Physical Review Letters **104**, 208101 (2010).

[77] T. J. Rudge, P. J. Steiner, A. Phillips, and J. Haseloff, *Computational modeling of synthetic microbial biofilms*, ACS Synthetic Biology **1**, 345 (2012).

[78] J. Sheats, B. Sclavi, M. C. Lagomarsino, P. Cicuta, and K. D. Dorfman, *Role of growth rate on the orientational alignment of escherichia coli in a slit*, Royal Society Open Science **4**, 170463 (2017).

[79] W. P. Smith, Y. Davit, J. M. Osborne, W. Kim, K. R. Foster, and J. M. Pitt-Francis, *Cell morphology drives spatial patterning in microbial communities*, Proceedings of the National Academy of Sciences 201613007 (2016).

[80] A. E. Blanchard and T. Lu, *Bacterial social interactions drive the emergence of differential spatial colony structures*, BMC Systems Biology **9**, 59 (2015).

[81] P. Ghosh, J. Mondal, E. Ben-Jacob, and H. Levine, *Mechanically-driven phase separation in a growing bacterial colony*, Proceedings of the National Academy of Sciences **112**, E2166 (2015).

[82] A. Persat, C. D. Nadell, M. K. Kim, F. Ingremeau, A. Siryaporn, K. Drescher, N. S. Wingreen, B. L. Bassler, Z. Gitai, and H. A. Stone, *The mechanical world of bacteria*, Cell **161**, 988 (2015).

[83] W. W. Baldwin, R. Myer, T. Kung, E. Anderson, and A. L. Koch, *Growth and buoyant density of escherichia coli at very low osmolarities.*, Journal of Bacteriology **177**, 235 (1995).

[84] Doru Constantin, *Mass moment of inertia of a hemisphere*, [Online; accessed July 21, 2017] (2014).

[85] M. Theers, E. Westphal, G. Gompper, and R. G. Winkler, *Modeling a spheroidal microswimmer and cooperative swimming in a narrow slit*, Soft Matter **12**, 7372 (2016).

[86] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, (Oxford university press 1989).

[87] C. Vega and S. Lago, *A fast algorithm to evaluate the shortest distance between rods*, Computers & Chemistry **18**, 55 (1994).

[88] M. Barthélemy, *Spatial networks*, Physics Reports **499**, 1 (2011).

[89] M. Newman, *Networks: an introduction*, (Oxford university press 2010).

[90] S. Fortunato and D. Hric, *Community detection in networks: A user guide*, Physics Reports **659**, 1 (2016).

[91] T. Aynaud, E. Fleury, J.-L. Guillaume, and Q. Wang, *Communities in evolving networks: definitions, detection, and analysis techniques*, in *Dynamics On and Of Complex Networks, Volume 2*, 159–200, (Springer 2013).

[92] D. S. Bassett, M. A. Porter, N. F. Wymbs, S. T. Grafton, J. M. Carlson, and P. J. Mucha, *Robust detection of dynamic community structure in networks*, Chaos: An Interdisciplinary Journal of Nonlinear Science **23**, 013142 (2013).

[93] S. Fortunato, *Community detection in graphs*, Physics Reports **486**, 75 (2010).

[94] D. S. Bassett, E. T. Owens, K. E. Daniels, and M. A. Porter, *Influence of network topology on sound propagation in granular materials*, Physical Review E **86**, 041306 (2012).

[95] D. S. Bassett, E. T. Owens, M. A. Porter, M. L. Manning, and K. E. Daniels, *Extraction of force-chain network architecture in granular materials using community detection*, Soft Matter **11**, 2731 (2015).

[96] Y. Huang and K. E. Daniels, *Friction and pressure-dependence of force chain communities in granular materials*, Granular Matter **18**, 85 (2016).

[97] M. E. Newman and M. Girvan, *Finding and evaluating community structure in networks*, Physical Review E **69**, 026113 (2004).

[98] J. Reichardt and S. Bornholdt, *Statistical mechanics of community detection*, Physical Review E **74**, 016110 (2006).

[99] J.-P. Onnela, D. J. Fenn, S. Reid, M. A. Porter, P. J. Mucha, M. D. Fricker, and N. S. Jones, *Taxonomies of networks from community structure*, Physical Review E **86**, 036104 (2012).

[100] A. C. Thomas and J. K. Blitzstein, *Valued ties tell fewer lies: Why not to dichotomize network edges with thresholds*, arXiv preprint arXiv:1101.0788 (2011).

[101] S. Fortunato and M. Barthélemy, *Resolution limit in community detection*, Proceedings of the National Academy of Sciences **104**, 36 (2007).

[102] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, *On modularity clustering*, IEEE transactions on knowledge and data engineering **20**, 172 (2008).

[103] B. H. Good, Y.-A. de Montjoye, and A. Clauset, *Performance of modularity maximization in practical contexts*, Physical Review E **81**, 046106 (2010).

[104] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, *Comparing community structure identification*, Journal of Statistical Mechanics: Theory and Experiment **2005**, P09008 (2005).

[105] E. Cuthill and J. McKee, *Reducing the bandwidth of sparse symmetric matrices*, in *Proceedings of the 1969 24th national conference*, 157–172, (ACM 1969).

[106] B. Desgraupes, *Clustering indices*, University of Paris Ouest-Lab Modal'X **1**, 34 (2013).

[107] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, *Internal versus external cluster validation indexes*, International Journal of Computers and Communications **5**, 27 (2011).

[108] D. Legland, K. Kiêu, and M.-F. Devaux, *Computation of minkowski measures on 2d and 3d binary images*, Image Analysis & Stereology **26**, 83 (2007).

[109] D. Legland and J. Beaugrand, *Automated clustering of lignocellulosic fibres based on morphometric features and using clustering of variables*, Industrial Crops and Products **45**, 253 (2013).

[110] D. Saintillan and M. J. Shelley, *Orientational order and instabilities in suspensions of self-locomoting rods*, Physical Review Letters **99**, 058102 (2007).

[111] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community structure in time-dependent, multiscale, and multiplex networks*, Science **328**, 876 (2010).

[112] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *Fast unfolding of communities in large networks*, Journal of Statistical Mechanics: Theory and Experiment **2008**, P10008 (2008).

[113] A. Kosowski and K. Manuszewski, *Classical coloring of graphs*, in *Graph Colorings*, 2–19, (Springer 2004).

[114] M. M. Syslo, M. Deo, and J. S. Kowalik, *Discrete Optimization Algorithms with Pascal Programs*, (Dover Books on Mathematics 1983).

[115] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, *Nih image to imagej: 25 years of image analysis*, Nature methods **9**, 671 (2012).

[116] T. J. Collins et al., *Imagej for microscopy*, Biotechniques **43**, 25 (2007).

[117] C. M. Bishop, *Pattern recognition and machine learning*, (Springer 2006).