Topological dynamics of small degree networks

Dissertation submitted towards the degree
Doctor of Natural Sciences
– Dr. rer. nat. –

Adrian Fessel

Bremen, December 17, 2018

Referees: Prof. Dr. Hans-Günther Döbereiner
Prof. Dr. Stefan Bornholdt

Universität Bremen

"Wisdom comes from experience. Experience is often a result of lack of wisdom."

— Sir Terry Pratchett

Abstract

In self-organizing networks, structure and dynamics interact in a unique way. Local activity sustained on the network organizes in global dynamic patterns to which the network topology adapts. It is believed that simple organisms lacking neural structures may derive their competence for complex behavior from this type of mechanistic interaction. One such organism is the slime mold *Physarum polycephalum*, which has emerged as an experimentally accessible model system for self-organizing transport networks over the recent years. Its remarkable features include the formation of a network structure optimized for transport, and various higher-level behaviors such as efficient foraging decisions and learning. Although significant progress has been made in uncovering these behaviors and the origins of dynamic processes taking place on the network, their relation via structural remodeling still lacks detailed understanding.

Combining experimental approaches with modeling and simulation, this work focuses on reorganization of the *P. polycephalum* network after fragmentation, a process occurring in a percolation transition in an expanding system. Data analysis and modeling are based on a novel method for graph extraction from spatial networks, which regards nodes as extended objects and fixes the number of degree two nodes in a graph.

Results suggest that the network formation process is separated into four functionally distinctive phases that are closely related to the topological state. Prominently, topological development is concluded once *P. polycephalum* reaches a steady state in which it continues to expand while maintaining a constant degree distribution, characterized by nodes of degrees one to four, with a negligible number of large degree nodes. Using this small degree property, analytical solutions to the random graph and configuration models of graph theory are derived. Comparison to experimental data reveals a shift of the percolation transition, which through modeling and simulation is attributed to active growth processes in the slime mold.

A model consisting of a deterministic rate equation and a stochastic master equation is devised based on the concept that the topological evolution can be decomposed into a sequence of elementary processes of four distinct classes, each representing one possible type of interaction between nodes. The model, characterized by a set of rate constants obtained from experimental data, describes the topological dynamics with excellent accuracy. In a simplified setting, the influence of interaction types is analyzed, and via simulations it is found that system growth shifts the percolation transition according to a power law.

Furthermore, the percolation critical exponents are determined, concluding that the simulated master equation model shares a universality class with mean-field percolation, whereas preliminary two-dimensional simulations, and by extension, network formation in *P. polycephalum*, are characterized by the exponents for two-dimensional percolation.

Contents

Abstract **iv**

Contents **vi**

List of figures **viii**

List of tables **xi**

List of algorithms **xi**

1 Introduction **1**
1.1 Past and present progress on *P. polycephalum* 3
1.2 Motivation, objectives and outline 9
1.3 Structure of this work 14

I Background **17**

2 Percolation, universality & complex networks **18**
2.1 Introduction to percolation theory 18
2.1.1 Site and bond percolation 19
2.1.2 Universal behavior: critical exponents 21
2.1.3 Finite size scaling 23
2.2 Introduction to complex networks 26
2.2.1 Fundamental concepts 27
2.2.2 Models of network structure 30
2.3 Percolation on graphs 35
2.3.1 Percolation in the random graph model 35
2.3.2 Percolation in the configuration model 36
2.3.3 The generating function formalism 38

3 Stochastic dynamics **42**
3.1 Markov processes and the Chapman-Kolmogorow equation 43
3.1.1 The Markov property 43
3.1.2 The Chapman-Kolmogorov equation 44
3.1.3 Derivation of the master equation 44

3.2 The master equation . . . 46
3.3 The chemical master equation . . . 46
3.3.1 Elementary reactions . . . 47
3.3.2 The propensity function . . . 48
3.3.3 Deterministic simplification: the rate equation . . . 49
3.4 The stochastic simulation algorithm . . . 50

II Results **54**

4 Analytical models for percolation in small-degree networks **55**
4.1 Small degree solutions in the random graph model . . . 56
4.1.1 Condition for the phase transition & geometric significance of the driving parameter . . . 57
4.1.2 Closed form solutions . . . 58
4.2 Small degree solutions in the configuration model . . . 64
4.2.1 Closed form solutions . . . 65
4.3 Comparison of closed form solutions . . . 67
4.3.1 Position of the phase transition . . . 69

5 A graph model for planar *P. polycephalum* networks **72**
5.1 Requirements for a suitable graph model . . . 73
5.2 The Apollonian graph approximation . . . 76
5.2.1 Obtaining the set of nodes . . . 79
5.2.2 Obtaining the set of edges . . . 82
5.2.3 Criticism, improvements and future directions . . . 83
5.3 Tracking topological modifications . . . 84

6 Dynamics of *P. polycephalum* network formation **87**
6.1 Experimental procedure & available data . . . 88
6.1.1 Experimental procedure . . . 88
6.1.2 Available data . . . 89
6.1.3 Influence of glucose . . . 97
6.2 Topological evolution of *P. polycephalum* . . . 99
6.2.1 *P. polycephalum* evolves towards a steady state . . . 100
6.2.2 Graph-based percolation in *P. polycephalum* . . . 103
6.3 Identifying phases of growth . . . 113
6.4 Tracking of topological changes . . . 121

7 A probabilistic model for topological dynamics **129**
7.1 Topological modifications & the relevance system growth . . . 130
7.2 Elementary topological modifications . . . 132
7.3 Modeling topological dynamics . . . 135
7.3.1 Rate equation . . . 135
7.3.2 Master equation . . . 139
7.3.3 The stochastic network simulation algorithm . . . 141
7.3.4 Towards a planar, stochastic network growth model . . . 145

7.4 Analysis of the model . . . 147
7.4.1 The GRFS model . . . 147
7.4.2 Analysis of the deterministic solution . . . 147
7.5 Analysis of stochastic simulations . . . 156
7.5.1 Simulating the GRFS model in two dimensions . . . 158
7.5.2 Percolation in the GRFS model . . . 160
7.6 Finite size scaling & critical exponents . . . 167
7.6.1 Influence of a bounded degree . . . 170
7.6.2 Influence of growth . . . 173
7.6.3 Finite size scaling with *P. polycephalum* rate constants and in two dimensions . . . 176

8 Summary **179**

Bibliography **183**

List of Figures

1.1.1 Plasmodium of *P. polycephalum* . . . 3
1.2.1 Information processing in *P. polycephalum* . . . 10
1.2.2 Graphical representation of relations between parts of this work . . . 12

2.1.1 Site and bond percolation . . . 19
2.1.2 Renormalization group coarse-graining . . . 22
2.2.1 A simple graph . . . 27
2.3.1 Self-consistency relation in the random graph model . . . 36
2.3.2 Self-consistency relation in the configuration model . . . 37

4.1.1 Giant component size in the random graph model with $n_k = 4$ and $p_2 = 0$ 60
4.1.2 Giant component size in the random graph model with $n_k = 5$ and $p_2 = 0$ 61
4.1.3 Phase diagram random graph model $n_k = 4$; illustration of the driving parameter . . . 63
4.3.1 Overview of closed form solutions . . . 69
4.3.2 Critical point in the configuration model . . . 71

5.1.1 Significance of p_2 nodes . . . 76
5.1.2 Color code for node degrees . . . 76
5.2.1 Apollonian Gasket in a triangular shape . . . 79
5.2.2 Apollonian graph approximation . . . 81

6.1.1 Initial configuration of experimental data . . . 90
6.1.2 Typical formation of a *P. polycephalum network* . . . 91
6.1.3 Anatomy of a *P. polycephalum* network; cluster analysis of nodes identifies structurally distinct regions . . . 94
6.1.4 Distribution of node size . . . 96
6.1.5 Adult networks at different nutrient concentrations . . . 97
6.1.6 Influence of the nutrient concentration . . . 98
6.2.1 Selected steps of a typical *P. polycephalum* graph evolution . . . 100
6.2.2 Topological evolution of a typical *P. polycephalum* network . . . 101
6.2.3 Topological steady state of *P. polycephalum* . . . 102
6.2.4 Evolution of component sizes in *P. polycephalum* . . . 106
6.2.5 Percolation of *P. polycephalum* and human endothelial cells . . . 110
6.2.6 Eliminating second degree nodes significantly improves the agreement between *P. polycephalum* percolation data and the configuration model . . . 111

6.2.7 Determination of the critical point by fitting with a heuristic model 112
6.3.1 Mean square displacement is a robust measure for different types of foraging behavior 114
6.3.2 Mean square displacement indicates phases of growth 116
6.3.3 Temporal correlation of *P. polycephalum* growth 118
6.3.4 *P. polycephalum* network at notable time points 120
6.4.1 Rate constants of characterizing the topological evolution 123
6.4.2 Clustering of rate constants 124
6.4.3 Rate equation model applied to experimental data 126

7.4.1 Numerical solutions of the rate equation 148
7.4.2 Degree distribution of trees produced by the rate equation model 149
7.4.3 Steady state of the rate equation: nodes 152
7.4.4 Steady state of the rate equation: degree distribution 153
7.4.5 Influence of rate constant magnitude on the dynamics of the degree distribution 155
7.5.1 Typical trajectories obtained by stochastic simulation 156
7.5.2 Comparison of rate equation, SSA and modified SSA for graph generation 157
7.5.3 Time series of network growth in a two-dimensional simulation 159
7.5.4 Topological evolution of a two-dimensional simulation 160
7.5.5 Component sizes in the GRFS model 161
7.5.6 Giant component size & percolation time; influence of separation 163
7.5.7 Influence of growth on the size of the giant component; influence of system size when growth is present 164
7.6.1 Data collapse of S, $< s >$ in the GRFS model with $n_k = 5$; no growth . . 169
7.6.2 Critical exponents in the GRFS model; influence of the degree bound . . . 172
7.6.3 Critical exponents in the GRFS model; influence of growth 174
7.6.4 Finite size scaling with rate constants from experimental data 176
7.6.5 Finite size scaling with of the spatial GRFS model 177

List of Tables

6.3.1 Topological characteristics of notable time points 117

List of Algorithms

3.4.1 Stochastic simulation algorithm 53

5.2.1 Apollonian graph approximation 78

7.3.1 Stochastic network simulation algorithm 144

Chapter 1

Introduction

Transport networks and systems are ubiquitous, affecting almost any aspect of our lives. We exchange information without delay via the internet, and logistics enable commuting or the distribution of goods on global and local scales. Similarly, our daily requirement for energy and water is met by local but interconnected networks. The most direct impact possibly comes from biological networks like the vascular system responsible for keeping our bodies supplied with all necessary compounds carried in the blood, or the neuronal network that constitutes the basis for information processing in higher organisms.

Although the examples given in this non-comprehensive list share important similarities, they also differ in many aspects: connections may be physically existent or abstract, depending on the quantity being transported on the network, and the networks may have been constructed by top-down planning or via self-organization. The latter difference initially seems to correspond to the distinction between artificial and natural systems, but a closer look reveals that many artificial systems also follow self-organizing principles. For example, road networks that were established in the past centuries have rarely been planned at the drawing board but have rather evolved by local changes deemed necessary due to fluctuations in demand, and were only later on modified by punctual, top-down planning (Barthelemy et al., 2013). Self-organization in this case may be attributed to the inability of implementing such large-scale projects in the past or to the lack of a central organizing authority.

However, present day technological advancement has rediscovered self-organization as an efficient way of implementing networking systems (e.g. Prehofer and Bettstetter (2005)), quite often even by direct bio-inspiration. Networks that are formed by the local interaction of autonomous agents are resilient against the failure of a central unit, and may adapt to changes in the network or changes in demand without the need for central reevaluation of network structure.

Nevertheless, networks constructed through local events need to be optimal with regard to some global task or set of properties. Lacking a central unit, the aforementioned principle encompassing the construction has to be implemented within a set of rules that govern, based on limited information, how the network structure should evolve locally.

Identifying a suitable set of rules remains a difficult task that may greatly benefit by learning from self-organizing biological systems that have been optimized through evolution. In many cases, obtaining the necessary information is challenging due to the vast complexity of the system at hand (e.g., neuronal networks), or due to its spatial embedding in an organism (e.g., the vascular network). Similarly, the time scales at which functional networks develop may be a hindrance (e.g., neuronal networks / road networks). This prompts study of model organisms that are more easily observed in a lab environment, such as the unicellular slime mold *Physarum polycephalum* (Oettmeier et al., 2017).

Physarum polycephalum forms a planar transportation network on a typical length scale of several centimeters that can alter its structure on a time-scale of hours. The network stage of *P. polycephalum*, termed macroplasmodium, moves by a treadmilling process in which distant parts of the network are deconstructed to provide the cellular material required for growth front regions to explore the slime mold's surroundings. The flow required for this process is driven by oscillations that give rise to a peristaltic wave pattern (Alim et al., 2013) coordinating the contraction of the slime mold's tubular veins. *P. polycephalum* historically emerged as a model organism for cell motility, but with the development of early genetic tools was abandoned in favor of organisms that, at the time, could be more easily modified.

Work on *P. polycephalum* was revitalized in the early 2000's when exciting properties of the network caught the attention of researchers: *P. polycephalum* optimizes its network structure for transport efficiency, as demonstrated via a maze solving experiment (Nakagaki, Yamada and Tóth, 2000). This property has since been exploited for various studies mimicking the structure of real-world transportation networks (e.g. Tero et al. (2010), Ntinas et al. (2017)), and is likely key to understanding other behavioral characteristics observed in the slime mold, such as learning and habituation (e.g. Saigusa et al. (2008), Beekman and Latty (2015), Boisseau et al. (2016)), that are generally associated with higher-level organisms.

Broadly speaking, current research projects in *P. polycephalum* pursue one or the other of two main objectives, either attempting to demonstrate and unravel the macroscopic behavior addressed in the previous paragraph, or to understand microscopic activity from a cell motility point of view. While the latter focuses on force generation and mechanochemical oscillation patterns in small, secluded units of the slime mold, the former neglects details of these mechanisms in favor of behavioral patterns shared between adult networks.

This dual focus of the *P. polycephalum* community leaves ample room in the middle ground, which can be occupied, e.g., by studying the formation of adult networks from small fragments and the remodeling and expansion of networks based on their dynamics. The best understood aspects in this regime describe the degradation or promotion of existing connections (Tero et al., 2007), and the organization of cytoplasmic flow in a developed network. However, both mechanisms do not explain the processes at the core of this work. First, the movement and expansion of network fragments based on

cytoplasmic flows is not well understood, and second the formation of new connections, especially during formation of adult networks from disconnected fragments, lacks a comprehensive description.

The latter process immediately suggests a fruitful analogy to the extremely well-developed field of graph theory and topology, which has proven to be tremendously helpful in the understanding of real-world networking systems (Newman, 2003). Therein, network formation is generally treated as a sequence of local events (Albert and Barabási, 2000), e.g. the placement of connections in a system of disconnected sites. Although these processes are generally stochastic in nature, they provide deep insights into the details of network formation, one of the most prominent example being a percolation transition at which the largest fraction of sites in a network forms a giant, inter-connected component.

1.1 Past and present progress on *P. polycephalum*

The present section is intended to provide the reader with an overview of phenomena observed in *P. polycephalum*, with past and current research efforts and future directions. Although the scope covered here is broader than what is required for following the display of this work, it is not intended as a comprehensive review. Several highly interesting topics are left out entirely, and those that are addressed are incomplete. For a more complete picture, the inclined reader is referred to the literature referenced below and in particular to the current review by Oettmeier et al. (2017).

Physarum polycephalum is a giant unicellular multinucleated slime mold. Belonging to the eukaryotic supergroup of amoebozoa, it is distinguished from the kingdoms of plants, animals or funghi. Slime molds are categorized into two subcategories, the acellular or 'true' slime molds such as *P. polycephalum*, which are single-celled but multinucleate, and the cellular slime molds such as the social amoeba *Dictyostelium discoideum* (Romeralo et al., 2012) that live as solitary organisms but may form multicellular structures as part of their reproductive cycle.

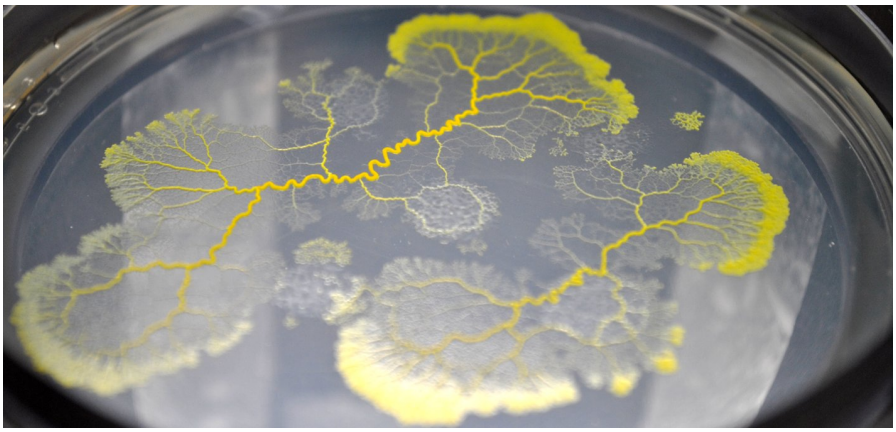


Figure 1.1.1: Plasmodium of the slime mold *P. polycephalum*, grown in a 9 cm petri dish containing agar. Image courtesy of C. Oettmeier.

Physarum polycephalum inhabits the moderate and tropical zones and is typically encountered in dark and moist environments, growing, for instance, on decaying wood, where it feasts preferably on bacteria and other microbes, and on a wide range of organic materials. *P. polycephalum*'s vegetative state is the plasmodium, in which it spans a tubular vein network with a typical length scale of several centimeters, although it is possible to grow plasmodia of larger size in a controlled environment. An example of a plasmodial vein network grown on an agar layer is shown in figure 1.1.1.

Although it is the plasmodial phase in which *P. polycephalum* spans networks and which is thus of primary interest to this work, it is worth noting that *P. polycephalum* has a diverse life cycle. This is detailed, for instance, in Sauer (1986). Very briefly, in response to selected conditions or stimuli, *P. polycephalum* can transition from the plasmodial phase into dormant stages intended for enduring adverse circumstances, or initiate the reproductive cycle. The dormant stage, which is entered upon desiccation of the plasmodium is referred to as the sclerotium. In this stage, all cellular function is interrupted and the plasmodium hardens, until the ambient conditions are returned to a more favorable state. In this dormant stage, *P. polycephalum* can endure for several years before being revived[1].

Although entering the sclerotium stage is a mechanism of differentiation in *P. polycephalum*, it is not a necessary stage in its life cycle. Similar to the sclerotium stage, the reproductive cycle of *P. polycephalum* is initiated in response to adverse conditions. Typically, it is initiated if the plasmodium is starved, but it is also possible to induce sporulation, the first stage of the reproductive cycle, with pulses of blue light (Werthmann and Marwan, 2017). Upon entering the sporulation stage, *P. polycephalum* forms stalks that raise above the plasmodium and carry sporangia at their tips. Sporangia are, unlike the brightly yellow colored plasmodium dark and mostly spherical. It is this peculiar stage which has provided *Physarum polycephalum* with its name meaning 'many-headed' (Oettmeier et al., 2017).

Sporangia hold spores, which are distributed by the wind. In favorable conditions, spores can germinate into uninucleate myxoamoebae, which, in contrast to the plasmodium, are haploid. Although *P. polycephalum* myxoamoebae show numerous fascinating behaviors (Sauer, 1986) of their own, they are not further considered here in order to focus on features of the plasmodium. Myxoamoebae can form a diploid zygote in a sexual fusion process. In the zygote, a synchronized nuclear division process commences at a rate of about one division per $8 - 10\,\mathrm{h}$. This nuclear division process, however, occurs without cell division, thus allowing the zygote to regrow into a multinucleate, diploid plasmodium.

The plasmodium forms an extended network that adapts to the topography of the surface it is grown on. Grown on a flat agar surface as shown in figure 1.1.1, the network is almost planar. Typically, two distinct regions exist in *P. polycephalum* networks. At

[1] Although this is useful for maintaining a *P. polycephalum* supply in the lab, the time span after which the slime mold can be successfully revived from sclerotia is short in comparison to the time after which *P. polycephalum* can be germinated from spores (Jump, 1954).

the rim of the structure, spacious growth fronts are visible that are extended by the slime mold in order to explore its surroundings. Growth fronts advance via amoeboid locomotion (Charras and Paluch, 2008; Lämmermann and Sixt, 2009), with velocities scaling with plasmodial height (Kuroda et al., 2015). Trailing the growth fronts is the actual vein network, which assumes a characteristic, hierarchical pattern: there exist few veins that stand out in terms of an above-average cross section and a length comparable to the size of the network. Branching from these central veins is a larger number of local veins with reduced cross section. Indeed, it has been demonstrated recently that the diameters of these branches are related to the diameter of the central vein via Murray's law (Akita et al., 2016).

Between two junctions, veins are approximately cylindrical in shape. Veins are separated into the solid, gel-like ectoplasm that forms the cortical layer of the vessel and the fluid endoplasm, which is enclosed by this wall. Endo- and ectoplasm do not differ in terms of composition and can be converted into one another (Isenberg and Wohlfarth-Bottermann, 1976) by a mechanism that is not known in detail. Furthermore, the endoplasm penetrates through the ectoplasm through a number of channels and internal veins that appear and disappear based on local pressure and shear (Guy et al., 2011). Unless the endoplasm is moving, it is typically not possible to clearly distinguish between the regions.

Whereas the endoplasm flows passively within the vein, the ectoplasm undergoes a cycle of active contraction and relaxation (Wohlfahrth-Bottermann, 1979). This active process is concerted by the interplay of force generation due to acto-myosin activity (Prost et al., 2015) in the cortical layer, and chemical regulation via calcium (Teplov, 2017). Local contractions along a vein lead to pressure differences, which in turn leads to a flow of the endoplasm. This process, which is characteristic for the *P. polycephalum* network is termed shuttle streaming. The contraction pattern giving rise to shuttle streaming are organized as a peristaltic wave with a wave length scaling with the size of the network (Alim et al., 2013). Local contractions repeat with a characteristic period of approximately 90 s (Bernitt et al., 2010), and flow within in the veins reverses on a similar time scale.

Shuttle streaming and oscillations are an essential ingredient for structure formation in *P. polycephalum* (Alim, 2018), its most basic function being the distribution of nutrients to remote regions of the network. Further, it has been observed, that upon encountering attractants or repellents, the slime mold would locally increase or decrease oscillation frequency and amplitude (Durham, 1976) prior to a reaction of the entire organism. It has since been shown, that the global wave pattern constituted by local oscillations in conjunction with the network structure are optimized for spreading signaling molecules that indicate the presence of a food source (Marbach et al., 2016; Alim et al., 2017).

The rich dynamics and characteristics of the slime mold in conjunction with the relative ease with which it can be cultivated and observed have naturally attracted researchers and made *P. polycephalum* a highly esteemed model system until the mid 1980s.

Following the historical account given by Oettmeier et al. (2017), topics of interest included, but were not limited to the study of differentiation between stages of the life cycle as outlined above, regulation of the cell cycle and naturally, cell motility, mechanics and locomotion. Furthermore, during this period the foundation has been laid for the study of macroscopic behavior in the slime mold, in particular in response to stimuli such as attractants and repellents or light exposure (Ueda et al., 1975; Häder and Schreckenbach, 1984).

However, with the development of improved techniques for cell culture and genetic manipulation for mammalian cells and bacteria, interest shifted away from *P. polycephalum* in favor of other systems, and the slime mold became dormant for almost two decades.[2] Following experiments performed by Nakagaki, Yamada and Tóth (2000), research in the field has been revitalized, but with a different focus. By demonstrating that *P. polycephalum* is capable of identifying the shortest path connecting two food sources in a maze, Nakagaki, Yamada and Tóth (2000) provided evidence that the network spanned by *P. polycephalum* is optimized to a degree largely unexpected in an unicellular organism. These findings have given rise to an active field focused on clarifying the mechanisms leading to the highly organized structure formation in the slime mold.

Numerous studies have since replicated and expanded knowledge on the behavioral intelligence in *P. polycephalum*. Following up on the maze-solving experiment by Nakagaki, Yamada and Tóth (2000), it has been demonstrated that *P. polycephalum* optimizes its network by favoring optimal motifs (Nakagaki et al., 2004; Shirakawa and Gunji, 2007), but can also compete with artificial transportation networks in terms of efficiency and redundancy (Tero et al., 2010). Even complex problems such as the traveling salesman problem can be solved by the slime mold (Zhu et al., 2013).

Closely related to the optimization of network structure is the decision-making behavior of *P. polycephalum* during foraging, as reviewed in Beekman and Latty (2015). For instance, it has been demonstrated that *P. polycephalum* is capable of selecting the food source with optimal protein to carbohydrate ratio in a complex nutritional challenge (Dussutour et al., 2010), and is capable of taking into account environmental risks during the decision-making process (Latty and Beekman, 2010*a*). Under stress, present through starvation or light exposure, foraging decisions made by the slime mold are subject to speed accuracy trade-offs comparable to those observed in higher organisms (Latty and Beekman, 2010*b*). The decision-making behavior of *P. polycephalum* has also been studied in the two-armed bandit problem, indicating that a non-neural organism is capable of efficiently solving decision-making challenges previously thought to be exclusive to higher organisms (Reid et al., 2016).

P. polycephalum shows different forms of learning. It has been shown that the slime secreted and deposited by the mold serves as an external, spatial memory, helping the slime mold to navigate its surroundings (Reid et al., 2012). The external memory serves

[2] Although *P. polycephalum*'s decline is linked to the advancement of methods in other systems, advancement of the same methods in the slime mold may return it to the focus of genetics due to a unique feature of the organism. All nuclei are identical, synchronized clones, which makes it possible to perform genetic readouts without disrupting or disturbing the system.

not only as a repellent marking regions that have been previously explored by the same slime mold or by others, but serves as an attractant if deposited in the vicinity of a food source (Reid et al., 2013), thus influencing the decision-making process. Furthermore, habituation, a learning behavior associated with higher organisms, has been reported in the behavior of *P. polycephalum* (Boisseau et al., 2016): upon frequent encounters with non-toxic, repellent substances, the slime mold becomes increasingly tolerant. Quite interestingly, a parallel search on the traveling salesman problem is performed significantly more efficient by amoeba that are fragmented from one specimen, when compared to the same task utilizing fragments of different specimen, indicating the existence of long term memory in the slime mold (Zhu et al., 2013). Coordinated behavior between non-fused fragments is also observed in mesoplasmodia, an emergency pattern in response to starvation in which fragments of the slime mold evacuate an area in a temporally correlated fashion (Lee et al., 2018).

Also emerged on the spectrum has the idea of unconventional computing, which comprises several concepts. For instance, *P. polycephalum* based algorithms have been developed that make use of the network optimization principle in order to complete computational tasks (Sun, 2017). In a more direct approach, it is tried to implement computing tasks on the slime mold (Adamatzky, 2016). First steps towards a biological computer have been taken by Tsuda et al. (2004), and the idea has been advanced since, for instance through the construction of logical gates via the slime mold (Adamatzky and Schubert, 2014). Slime molds have been employed for robot control in a number of scenarios, utilizing the concept that stimuli applied to the plasmodium lead to outputs in a controlled fashion (Tsuda et al., 2007; Jones et al., 2011). Although one may be rightfully skeptical about the usefulness of these implementations, they certainly highlight that much can be learned from the diverse behaviors of *P. polycephalum*.

All behaviors described above are, in one way or another, related to the dynamics of the *P. polycephalum* network, to the shuttle streaming within its veins and to the oscillations causing this streaming. Indeed, a close relation has been established between shuttle streaming and the adaptation and optimization of the network in the form of the current reinforcement model (Tero et al., 2007), which states that veins in the network will be degraded or promoted, depending on the amount of fluid transported through them. The model is used successfully in *P. polycephalum* algorithms and has successfully described the adaptation of the network in various scenarios (see, e.g., (Tero et al., 2010; Akita et al., 2017)). It is noteworthy that there exists a strong similarity between the current reinforcement model and the plastic adaptation of neural weights during reinforcement learning.

It has been outlined above that the oscillation pattern of the slime mold changes if new stimuli are encountered. Oscillations lead to flow, and changes thereof therefore are bound to modify details of the flow pattern on the network, thus directly affecting the structure of the network through current reinforcement. Furthermore, flows transport signaling molecules, and thus cause local stimuli to be presented to the entire organism. This indicates, that the oscillation dynamics of the network possibly lie at the very basis of decision making behavior. It thus appears worthwhile to study not only

the behaviors of the slime mold, but also the mechanistic processes that give rise to them.

Promising approaches relate macroscopic behavior observed in the slime mold to its structure without detailed knowledge of the underlying oscillator. One example is the emergence of the peristaltic wave pattern as detailed above, another is given by the analogy between hydrodynamic flows and elements of electrical circuits, termed hydraulic analogy (see, e.g., Oh et al. (2012)). Using a network of memristors, resistors that change state based on the history of currents flowing through them, Pershin et al. (2009) have been able to reproduce experimental results published by Saigusa et al. (2008). Therein, *P. polycephalum* is repeatedly subjected to unfavorable conditions that led to a decrease in oscillation frequency. After several repetitions, the slime mold would adapt its frequency even if the stimulus was not given. Furthermore, after an extended period of time and after the periodic adaptation had fully declined, a single stimulus would be sufficient to incite another set of adaptation periods. Subsequently, the same model has been successfully employed to solve a shortest path problem (Pershin and Ventra, 2013).

More traditionally oriented than research on the macroscopic behavior of the slime mold, research on oscillations, streaming and mechanics in the slime mold has become active again after its dormancy. As described above, contraction-relaxation oscillations in *P. polycephalum* networks are organized as a peristaltic wave (Alim et al., 2013). However, although this realization explains the emergence of shuttle streaming, it does not immediately provide further insights into the nature of local oscillations or into the mechanisms that control their coupling behavior. Although there exist promising candidates, both questions are, at large, still open. Numerous models have been proposed that recreate many wave phenomena observed in *P. polycephalum*. It is not attempted here to give a comprehensive review, and instead the reader is referred to current reviews such as Teplov (2017); Oettmeier et al. (2017). A biochemical model regarded as a promising candidate for the oscillator has been devised by Smith and Saldana (1992), in which contractions are due to myosin motor activity driven by the concentration of free calcium ions. The model has been extended to a spatial form by taking into account deformations of, and flow through the plasmodium. This is facilitated by modeling the plasmodium either as a poroelastic two-phase material (Radszuweit et al., 2014; Alonso et al., 2016), or as a one-phase active fluid material (Alonso et al., 2017). The model explains numerous wave phenomena observed in microdroplets of *P. polycephalum* (Takagi and Ueda, 2008, 2010). Other models are of reaction-diffusion-advection type (Nakagaki et al., 1999) or make use of the mechanical coupling rather than relying on the exchange of chemicals.

In any case, the dynamics of *P. polycephalum* are the outcome of a carefully orchestrated interplay of various processes and the unique physical properties of the slime mold. Investigating either are active fields, that have sparked numerous studies as seen above, and due to their complexity have raised the need for simpler model systems. Preparing these is possible due to *P. polycephalum*'s resilience to physical manipulation. For instance, microdroplets as mentioned in the previous paragraph are small volumes of cytosol extracted from a plasmodium using a syringe. After plating, these droplets form a new membrane and remain vital, showing a wide range of oscillation and wave

phenomena. Similarly, it is possible to create so-called microplasmodia by cutting, or by subjecting a plasmodium to shear forces in a liquid culture. These fragments range from $100-500\,\mu m$, are typically of an approximately globular shape unless migrating, and form vein networks only after growth or fusion to other microplasmodia. Microplasmodia have been employed to study the viscoelasticity and nonlinear elasticity of *P. polycephalum* (Fessel et al., 2017; Fessel and Döbereiner, 2017), but serve also as a model system for oscillations (Bernitt et al., 2010) and have been subject to ultrastructural investigation (Oettmeier et al., 2018). Further, the onset of peristaltic pumping and forces exerted onto the substrate have been studied and modeled in migrating microplasmodia, see, e.g., Lewis et al. (2015); Rodiek and Hauser (2015); Lewis and Guy (2017); Zhang et al. (2017).

1.2 Motivation, objectives and outline

Albeit its simple structure as a single-celled organism, *Physarum polycephalum* displays behavioral patterns typically associated with more complex organisms that are in possession of a refined central nervous system or at least some neuronal structures. Lacking such a system for information processing, *P. polycephalum* has to resort to a different method of conveying complex behavior. Although the exact mechanisms have been unraveled only in part, it is generally believed that the self-organized coordination of flows across the slime molds network, in conjunction with adaptation of the network structure to the flow pattern, drives information processing in *P. polycephalum* (see, e.g., Alim (2018)).

One possible, entirely mechanistic interpretation of information processing in *P. polycephalum* is illustrated in figure 1.2.1. Very briefly, *P. polycephalums* typical behavioral mode is foraging, which brings it in contact with diverse external stimuli such as chemotactic signals (Ueda et al., 1976), light exposure (Häder and Schreckenbach, 1984) or variation of humidity (Saigusa et al., 2008). To those stimuli, reactions are in principle possible in two ways. First, selected stimuli or prolonged exposure are bound to gradually influence transcription levels in the slime mold, or effectively switch between behavioral patterns. For instance, pulses of blue light induce sporulation (Werthmann and Marwan, 2017). However, these processes are slow in comparison to the reactions of the slime mold observed in response to some stimuli, such as the reaction to encountering a food source (Alim et al., 2017).

Typically, such an encounter directly leads to changes observed in the oscillation pattern (Alim et al., 2017). This suggests that dynamic processes on the network are regulated immediately in response to some stimuli, without the prior activation of a different transcriptional program. In any case, it has been observed in different experiments (e.g., (Tero et al., 2010; Akita et al., 2016)) that the structure of *P. polycephalum* is modified by flows resulting from the global wave pattern of the slime mold. Similarly, the domain on which flows and waves may propagate is constrained by the network topology. Thus, dynamic processes and structure formation are interacting, and it is hypothesized that within this mutual relationship lies the key to understanding how information is processed in *P. polycephalum*.

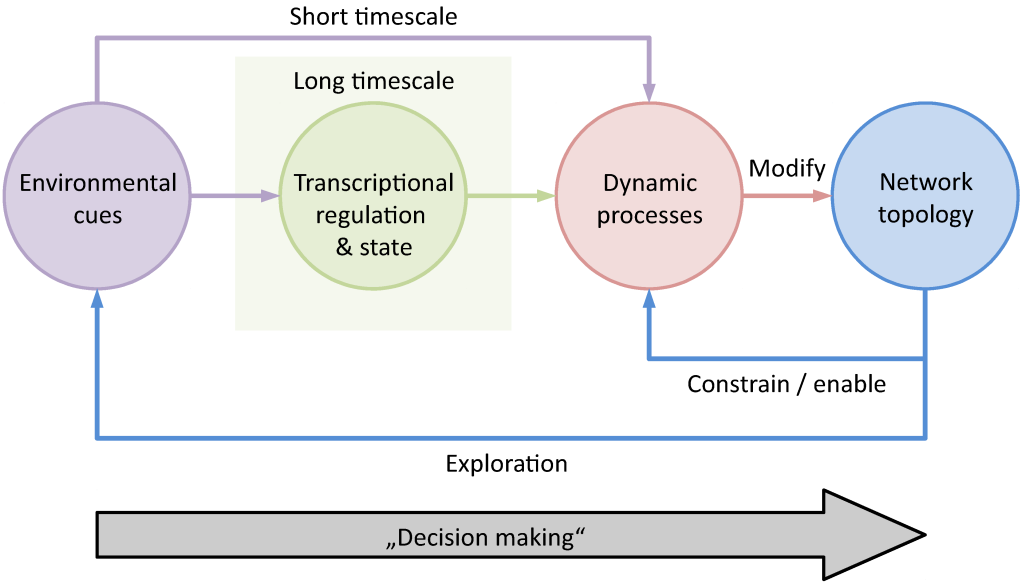


Figure 1.2.1: Simplified schematic for information processing in *P. polycephalum*. The slime mold reacts to external stimuli either by transcriptional regulation on a longer time scale, or immediately through adaptation of dynamic processes, such as flow or oscillation, taking place on the network topology. The network structure is modified by the dynamic processes and in turn influences these, as it functions as the domain on which the processes take place. Modifications in network structure can correspond to exploration and therefore bring the slime mold into contact with novel environmental cues.

However, verifying this hypothesis is not possible within in the scope of this work. Further understanding of the process described above requires detailed study of a wide range of aspects, including behavioral response to diverse stimuli from a decision-making point of view, coordination of network-wide oscillation patterns, and the remodeling and formation of network structure. This work aims to improve understanding of the latter aspect, remodeling and, in particular, formation of the slime mold network. Although no direct connection is made in the scope of this work, it is intended that an extended version of the framework developed here will pave the way for further systematic study of the interaction between dynamic processes occurring on, and modification of the network topology.

The process studied here in detail is the formation of a *P. polycephalum* network after fragmentation. Briefly, *P. polycephalum* can be severed into small fragments through a variety of methods, including the application of shear forces in a liquid shaking culture. As *P. polycephalum* is multinucleate, fragments of the slime mold are vital if sufficiently large, and if left by themselves proceed to regrow into a network through extension and fusion of fragments. This process is of special interest, as by the hypothesis stated above, a functional network structure is required by the slime mold in order to be able to exhibit

behavioral patterns such as foraging.

In order to transition from the fragmented state into the connected state required for foraging, *P. polycephalum* has to undergo numerous modifications of the structure or topology of the network. Although this process is in principle continuous, it can be regarded as a sequence of discrete events that can be classified as growth, fusion, or as a process reversing the aforementioned.

Interpreting structural modifications as discrete events appears most sensible if the modified structure is discrete as well. Therefore, states during the time evolution of a *P. polycephalum* network are, in the scope of this work, understood as graphs, i.e., as a collection of nodes, connected by a set of edges. In this picture, a growth process is represented by the addition of a novel, connected node to the graph, and fusion is the addition of a novel edge. However, in order to correctly represent these processes, the underlying graph model is required to fulfill a number of prerequisites, the most prominent being an unambiguous definition of the number of second degree nodes between junctions of the network. Therefore, the first objective of this work is the introduction of a graph model suitable for representing and analyzing *P. polycephalum* networks and the structural changes thereof.

It is the underlying hypothesis of this work, that the restructuring process described above can be understood along the lines of a percolation transition. In statistical physics, percolation processes are stochastic models that describe the emergence of global connectivity in an initially disconnected system. In particular, graph theory has adopted percolation as a means of describing the transition in which a connected graph forms if a set of unconnected nodes is gradually becoming connected through the stochastic introduction of novel edges.

The previous paragraphs highlight the close relation between the suggested models, i.e., percolation and graph formation via discrete events. However, there is one important difference. The models named above typically possess an inherent randomness, whereas the slime mold forms its structure as part of an active process. Nevertheless, as will be seen throughout this work, much can be learned through the application of stochastic models. On one hand, although active processes lead to structural changes in the slime mold network, there is also intrinsic randomness, which is why all *P. polycephalum* networks are organized similarly, but are never exactly equal. On the other hand, describing the active process with a stochastic model allows the discovery of properties that do not depend on the detailed construction history of a network, but are shared between very different realizations of the process.

The analysis of the network formation process in *P. polycephalum* in the light of the physical models stated above can be broken down as graphically organized in figure 1.2.2. The work is based on experimental data representing the network formation in *P. polycephalum*, and on the concept that construction of a network requires only a small number of elementary processes. Based on this, the first objective has already been stated above as the search for a suitable graph representation that approximates the *P.*

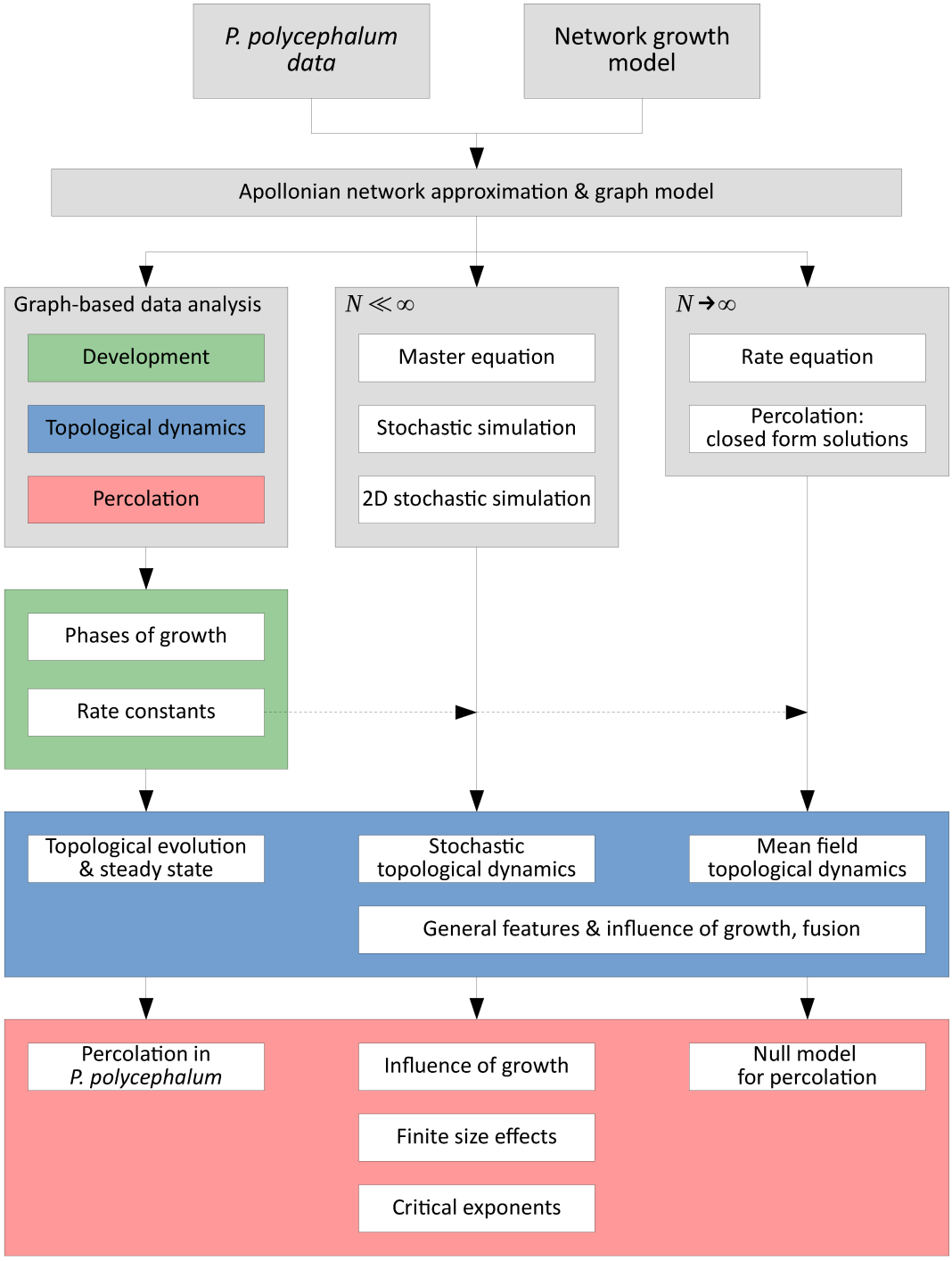


Figure 1.2.2: Graphical representation of relations between parts of this work. The leftmost pillar summarizes efforts and results related to data analysis, the center and right pillars indicate theoretical results, separated according to whether results are valid in the thermodynamic limit or in finite systems. Grey boxes collect methods and concepts, colored boxes categorize results.

polycephalum network and changes thereof well.

Building on this, the present work in essence consists of three pillars, the first being data analysis. Analyses are carried out with three central objectives, including, first, the identification and quantification of developmental phases during the network formation process. Second, the quantification of how network structure as characterized by the degree distribution and the distribution of component sizes evolves during the formation process. Finally, topological modifications such as growth or fusion are bound to occur at rates that can be measured experimentally.

The second and third pillar are based on theoretical modeling and simulation. Devising and testing the models and simulations as detailed below presents further objectives of this work. One possible way of structuring the results is given by system size. In the thermodynamic limit ($N \to \infty$), there are typically no fluctuations and it is possible to find deterministic solutions to physical problems. On the other hand, if $N \ll \infty$, systems become stochastic, and due to fluctuations the corresponding deterministic solution can only be recovered by averaging a significant number of stochastic trajectories.

Focusing first on the deterministic pillar, two different models are considered. Based on models stemming from graph theory, analytical closed form solutions to the percolation problem are derived. Generally, such solutions can only be found by numerical methods. However, for graphs possessing only a negligible number of large degree nodes, it is possible to solve the configuration and random graph models analytically. Comparing closed form solutions to experimental data highlights that in *P. polycephalum*, the critical point marking the phase transition is shifted. In comparison to the analytical solution, slime mold networks require a larger degree of structure in order to percolate.

In order to investigate possible causes for this discrepancy, a model is constructed that describes the formation of a network based on a small number of elementary processes resembling chemical reactions between different nodes. The model consists of two parts. The first is a deterministic rate equation describing the average evolution of the degree distribution and system size. As this average description holds either in the thermodynamic limit or for a large number of realizations, it is part of the second pillar. Belonging to the third pillar, a stochastic master equation describing the same quantities and their fluctuations is devised. The master equation is closely related to the stochastic simulation algorithm, which provides the basis for generating exact trajectories of stochastic processes. This algorithm is modified in order to simulate stochastic network growth, and stochastic network growth embedded in two dimensions. The model for network formation, unlike the model for percolation is not limited to small degree networks and can implement, for example, the formation of scale-free networks.

In summary, each pillar in figure 1.2.2 represents one way of modeling network formation as characterized by degree distribution and component sizes. Results are either obtained from experimental data, via stochastic simulation, or through combination of the two deterministic models for network formation and percolation. The models can incorporate either growth rates as measured in *P. polycephalum*, or arbitrary rates. This

makes it possible to compare results between the three pillars, or to investigate the influence of elementary processes. In particular, growth is investigated as a possible source for the observed shift. Furthermore, simulations enable variation of system size in a much broader range compared to the experiment, making it possible to investigate finite size effects, and to determine critical exponents of the simulated model.

The efforts outlined in this section formalize *P. polycephalum* network formation into a stochastic, graph-oriented framework, characterize the percolation phase transition and elucidate the influence features of the dynamical network growth have on the transition. In the previous section, dynamics and behavioral characteristics of the slime mold have been described in greater detail. It is hoped, that the framework developed here may serve as a discretization on which effective models, such as the current reinforcement model (Tero et al., 2007), or models for synchronization on networks (Arenas et al., 2008), can be brought together to simulate characteristics of the slime mold, or even serve as a model for decision making if interacting, for instance, with external gradients. Instead of initiating events that modify network structure through purely stochastic selection, their stochasticity can be influenced locally through the dynamic processes outlined above.

1.3 Structure of this work

The remainder of this work is structured as follows. Following this introduction, the main text is split into two parts. First, the **I Background** provides essential concepts adopted from the literature that are either required to follow the development of ideas presented in this work, or provide the means to critically assess the results detailed in the subsequent chapters. As this work is only in part based on the analysis of experimental data or simulations, and introduces theoretical ideas as well, it has been found useful to compile all chapters presenting own work into the part **II Results**.

Part **I Background** consists of two chapters, aimed toward the central theoretical concepts employed in this work. Chapter 2 introduces two related concepts. As outlined above, percolation theory encompasses the formation of connected structures from unconnected fragments. These fragments can be, for instance, fragments of an evolving graph structure. In chapter 2, the basic ideas underlying percolation theory and the concept of universality are outlined, followed by a brief introduction into the fundamentals and recent advances in complex networks. Finally, the two concepts are usefully combined, and the mathematical construct required to study percolation on graphs is outlined. Although networks and percolation are of significance in all parts of this work, chapter 2 is most closely related to chapter 4, where analytical solutions are devised for percolation in small degree networks.

Chapter 3 extends the theoretical background of this work to the theory of stochastic processes. With the intention of laying the basis for stochastic modeling and subsequent simulation of network formation, the master equation is derived building on the concept of Markov processes. The master equation describes the evolution of a stochastic system based on the gains and losses in occupation probability of the systems microstates. One of its many applications lies in the description of chemical reaction networks. With the

intent of later on formulating changes in graph topology as chemical reactions, the chemical master equation is introduced alongside the framework required for the stochastic description of a reaction network. As the master equation in many cases resists treatment by analytical methods, the stochastic simulation algorithm and the rationale behind it are present as a method of simulating exact trajectories of the master equation. In chapter 7, the stochastic simulation algorithm serves as the basis for the network simulation algorithm devised in this work. Chapter 3 concludes part **I Background**.

Part **II Results** presents own work and has been further subdivided into four chapters. The first chapter, chapter 4 continues the treatment of percolation on graphs. It is demonstrated that closed form solutions to the random graph and configuration models of graph theory are possible if the number of large degree nodes in a system is negligible. Solutions are derived and compared for a number of cases and constraints to the degree distribution near the transition are given.

Chapters 5 and 6 focus on the treatment of network data obtained as image series documenting the formation of *P. polycephalum* networks. First, in chapter 5 a graph model suitable for extracting graphs from *P. polycephalum* images is presented. Numerous requirements, including a sensible definition of second degree nodes for a transportation network are outlined and compared to the topological skeleton, which presents a more traditional approach to graph extraction from images. It is described how the outlined extraction process can be implemented. Briefly, the implementation makes use of a circle packing approach in which each network node is associated with an area on the network structure. Connectivity between nodes is then derived from the topological skeleton. The chapter is concluded with an account of a collaborative project that has led to the development of a method for tracking topological changes between graphs of the image series.

Chapter 6 reflects the data analysis performed as part of this work. Data analyzed are time series documenting the formation of *P. polycephalum* networks from scattered microplasmodia, represented as a series of graphs employing the graph model developed in chapter 5. Of pivotal interest to this work is the topological evolution of the slime mold, characterized by its degree distribution and the distribution of component sizes. For a clear display, the chapter has been structured as follows. First, the experimental procedures is outlined and typical results are described in detail in order to familiarize the reader with network formation in *P. polycephalum*, and with the final state of the network and the influence of initial conditions. Building on this, a systematic analysis of the topological evolution is performed, showing that there exist common features shared between evolving *P. polycephalum* networks, including the final state, which is characterized by a single degree distribution. Analyzing component sizes, it is observed that there are strong deviations between experimental data and analytical solutions as developed in chapter 4. Supported by simulations performed in chapter 7, it is concluded that growth and inaccuracy of the the configuration model for degree distributions with a large number of second degree nodes are the central reasons. Furthermore, a method is presented for distinguishing *p. polycephalum* growth into four functionally and structurally different phases. Finally, tracking of topological changes leads to the experimental

determination of a set of rate constants, which are parameters of the model devised in chapter 7 and drive the growth of *P. polycephalum*. It is shown that these constants correctly reflect the topological evolution and differ between the phases of growth.

The final chapter of part **II Results** focuses on modeling and simulations of topological network evolution. In chapter 7, concepts outlined in chapter 3 are taken up and two models are devised based on the concept that modifications of network topology can be expressed as chemical reactions. Four basic interaction types are defined, and based on them a deterministic rate equation for the degree distribution, and a master equation extending the model to a stochastic framework are devised. The model is characterized by a set of rate constants and can reproduce degree distributions of various network types, including random graphs or scale-free networks. In order to generate network topologies consistent with the models, the stochastic simulation algorithm is extended. The model is analyzed in two steps, focusing first on the rate equation, and in the second step on stochastic simulations. The analysis is performed in a simplified framework where only four rate constants characterizing growth, fusion and their reversals need to be defined. Working with the rate equation, typical behaviors of the model and its steady state are analyzed. Next, the analysis is extended to stochastic simulations, and excellent agreement is found between simulations and the rate equation. The analysis of stochastic simulations focuses on percolation. In the simulated framework it is possible to study the influence of the four basic interaction types on the percolation transition. In particular, the influences of growth and of the small degree property are studied, and it is observed that growth shifts the critical point according to a power law, leading to deviations from the configuration model. Systematic analysis makes use of finite size scaling as introduced in chapter 2. Percolation critical exponents are found to be in agreement with mean-field percolation and no evidence is found for an influence of growth or the degree bound. Finally, a preliminary two-dimensional implementation of the simulation is tested, leading to critical exponents in agreement with two-dimensional percolation and good agreement with the rate equation model.

Part I

Background

Chapter 2

Percolation, universality & complex networks

Concepts from percolation theory and graph theory are widely employed throughout this work. In the present chapter, the following steps are taken. Following this introduction, the basics of percolation theory are outlined in 2.1. Therein, following a brief historical account, the related concepts of bond and side percolation are introduced in section 2.1.1. Subsequently, the concept of universality is introduced in section 2.1.2 stating that near the percolation transition, very different systems scale according to the same critical exponents if they are of the same dimensionality. Dealing with data from finite systems in chapter 6 and 7, finite size effects make systematic corrections necessary in order to correctly determine critical exponents. The ideas underlining finite size scaling, which is a method to exploit finite size effects for the determination of critical exponents, are outlined in section 2.1.3.

Section 2.2 introduces the reader to basic ideas (section 2.2.1) underlying graph theory, and to a number of prominent graph models (section: 2.2.2). These include the random graph and configuration models employed in this work, which are further addressed in terms of graph percolation in section 2.3. The random graph model is addressed in section 2.3.1, and percolation in the configuration model is treated in section 2.3.2. In order to treat this problem properly, the generating function formalism is introduced in section 2.3.3.

2.1 Introduction to percolation theory

The term percolation, which has its etymological roots in the Latin word *percolatio*, meaning filtration, has been coined for a class of stochastic models in 1957 by Broadbent and Hammersley (1957), although there have been earlier works on comparable problems, for instance, gelation (Flory, 1941; Stockmayer, 1944). Broadbent and Hammersley (1957) proposed the percolation process as a model for the way in which a fluid or gas penetrates through a porous medium. In such a medium, microscopic pores may be either open and allow fluid through, or they are blocked with some probability and cannot admit any flow. This leads to the central question motivating the study of percolating systems: at what percentage of open passages will there be a way for the fluid to move freely from

one end of the medium to the other? Since then, numerous other physical problems have been mapped to the percolation process, such as the spread of epidemics (Grassberger, 1983) or forest fires (Henley, 1989), electrical conductance (Kirkpatrick, 1973), for the coverage of wireless sensor networks (Dousse et al., 2004), or even neural activity during decision-making (Kozma et al., 2009). Generally, most problems that are in one way or another related to connectivity in some underlying domain share similarities with percolation, highlighting its close relation to graph theory, which is of special interest to this work.

Percolation is one of the most simply stated models in the field of statistical physics that exhibits a critical phenomenon, or phase transition. A phase transition marks the point at which, upon variation of a driving parameter characterizing the microscopic structure of a system, macroscopic properties undergo spontaneous, drastic changes. With regard to the percolation problem introduced by Broadbent and Hammersley (1957) the critical point is marked by the fraction of open passages required on average to allow flow through the porous medium. However, typically this is stated differently but equivalently, as the search for a critical point at which a giant, connected cluster emerges, with physical dimensions on the order of the size of the system.

Percolation has first been considered on regular lattices, such as the square lattice in two or three dimensions, honeycomb or hexagonal lattices, or spatially non-homogeneous lattices such as the Bethe lattice (Fisher and Essam, 1961), which is a tree graph. This underlines the close relation between percolation and network theory. Therein, percolation is typically viewed as the existence of a globally connected configuration, based on how many of the edges connecting nodes are present in the graph.

2.1.1 Site and bond percolation

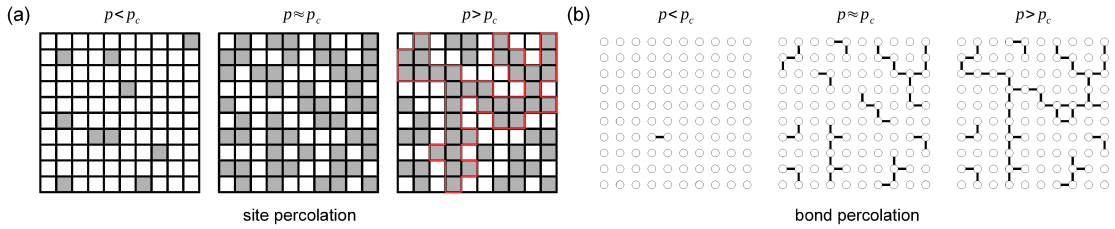


Figure 2.1.1: Site and bond percolation. (a) Site percolation. Sites can be either occupied or unoccupied. The driving parameter p measures the fraction of occupied sites. Continuously adding to the fraction of occupied sites drives the system first into the critical region ($p \approx p_c$), where the addition of a small number of occupied sites (gray) leads to the formation of a giant cluster (labeled in red). (b) Bond percolation. Instead of sites, bonds between sites are added, again leading to the formation of a spanning cluster. As each bond has six neighbors compared to four in the site model, the percolation threshold p_c is smaller in the bond percolation model.

Two underlying models for percolation have emerged in the literature that are most suitably compared on the two-dimensional square lattice, as shown in figure 2.1.1. The examples shown there are for a very small system size. Typically, analytical considerations of percolation problems are performed in systems of infinite size, i.e., lattices extending to infinity in both spatial dimensions in the present example. Finite sized systems require corrections, which are addressed further below.

In the site percolation model shown in panel (a), the driving parameter $p \in [0, 1]$ represents the fraction of occupied sites. Therefore, every site is occupied with probability p, and unoccupied with probability $1-p$. For small p, the lattice is only sparsely populated and there are few clusters in which more than one site is connected. Near $p = p_c$, the situation changes. Shortly before the critical point p_c, there are numerous average-sized clusters, and it is difficult to occupy further sites without merging these average-sized clusters. Doing so leads to the situation shown in the third panel ($p > p_c$), where there exists one giant cluster spanning almost the entire system. The threshold value p_c so far has not been determined analytically for site percolation on the square lattice, but has been determined numerically with a high degree of accuracy: $p_c \approx 0.59275$ (see, e.g., Newman and Ziff (2000)). The difficulty to obtain the threshold analytically illustrates, that percolation, although formulated as an extremely simple problem, requires in many cases complicated treatment.

In bond percolation as seen in figure 2.1.1 panel (b), the situation is quite similar. Instead of occupying sites, bonds (or edges) of the lattice are occupied. Notably, the percolation threshold p_c is different from the site percolation model. This arises from the fact, that each bond has six neighboring bonds, whereas each site on the square lattice has only four neighboring sites. The percolation threshold in this case is known to be $p_c = 1/2$, as has been proved by Kesten (1982). Although site percolation is typically considered to be a more general way of stating percolation problems, bond percolation is more closely related to graph theory.

Numerous details of percolation models have an influence on the exact value of the critical point. These include not only the dimension of the problem, but also microscopic details of the model constraining the local structure of the percolating system. As seen above, the percolation type, i.e., whether site or bond percolation is considered, has an influence that is directly tied to local structure via the connectivity of the unit cells. Similarly, the lattice or graph type has an influence, and so do details of the process in continuum percolation theory.

Percolation theory generally considers static systems and attempts to make statements on the global connectivity based on local structure as characterized by a driving parameter. However, in some cases it is instructive to take the dynamics of a process leading to percolation into consideration. For instance, throughout this work, the role of system growth during percolation is considered in a number of places, and it is found that the percolation threshold in a growing system is shifted in comparison to a non-growing null model. Explosive percolation on graphs (Achlioptas et al., 2009; D'Souza and Nagler, 2015) is another excellent example. Therein, development of a giant cluster

is suppressed by a simple rule: instead of placing a bond at random, two bonds are placed randomly but only the one which leads to the smaller resulting cluster size is kept. This dynamic construction process leads to a significant shift of the critical point.

2.1.2 Universal behavior: critical exponents

Unlike the location of the critical point, the behavior of critical quantities characterizing the percolation transition is independent of microscopic details of the system. In a percolating system, these quantities include, among others, the size of the giant cluster $S(p)$, which is the order parameter of the system. In an infinite system, S is equal to zero in the non-percolated regime and non-zero when $p > p_c$. At the transition, the giant component size is continuous but not continuously differentiable, indicating a second order phase transition.

The role of the fluctuations of the order parameter is taken by the average non-giant cluster size

$$< s(p) >= \frac{\sum_s' s^2 n_s(p)}{\sum_s' s n_s(p)} \ , \tag{2.1.1}$$

where $n_s(p)$ is the distribution of cluster sizes s for a given p, excluding the size of the giant cluster S. At the transition, $< s >$ diverges if the system is of infinite extension. Further, the correlation length ξ of the system, defined by

$$\xi(p)^2 = \frac{\sum_s' r_s^2 s^2 n_s(p)}{\sum_s' s^2 n_s(p)} \ , \tag{2.1.2}$$

where r_s is the average radius of gyration of non-giant clusters of size s, also diverges at the transition. The intuitive interpretation of the correlation length is the distance from the cluster center at which a site is still expected to belong to the same cluster.

The three quantities given above are said to exhibit critical scaling when approaching the transition, i.e., if $|p - p_c| \to 0$, as characterized by the power law behavior

$$S(p) \propto \begin{cases} 0 & p < p_c \\ (p - p_c)^\beta & p > p_c \end{cases} \ . \tag{2.1.3}$$

The average cluster size scales as

$$< s(p) > \propto |p - p_c|^{-\gamma} \ , \tag{2.1.4}$$

and the correlation length is characterized by

$$\xi(p) \propto |p - p_c|^{-\nu} \ . \tag{2.1.5}$$

The exponents β, γ, ν are referred to as critical exponents, and are expected depend only on the dimension of the percolating system. This principle is known as universality. The idea of universality in statistical physics is that systems with very different microscopic details, that exhibit a phase transition of the same type, share the same critical exponents if they their dimensionality is identical. An explanation for universality follows

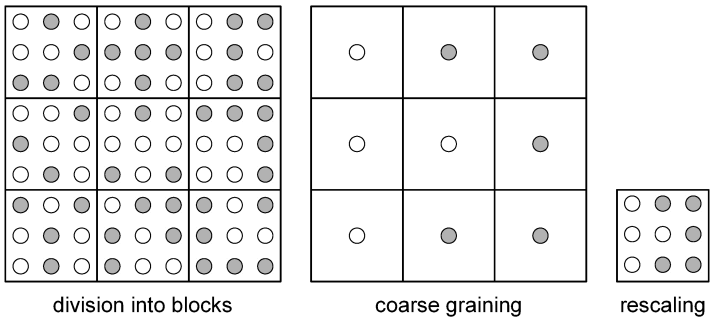


Figure 2.1.2: Coarse-graining in the renormalization group. The lattice is divided into blocks, in which the lattice values are summarized by some procedure. Finally, the result is rescaled.

from the renormalization group transform, which is outlined below upon omission of all technical details.

The renormalization group transform is a coarse-graining procedure, which attempts to reduce the number of degrees of freedom of a stochastic system. For instance, in site percolation there are as many degrees of freedom as there are lattice sites. As each lattice site can take one of two possible values, there is strong similarity to the Ising model, for which Kadanoff (1966) introduced the so called block spin model, a procedure by which the number of degrees of freedom is reduced, and which is shown in figure 2.1.2. First, the lattice is divided into equally-sized blocks. Next, within each block the lattice values are summarized by some method. In figure 2.1.2, the resulting value is selected according to the majority of lattice values within the block. Finally, the system is rescaled by a factor b, such that the dimensions of the new block agree with the dimensions of the old blocks.

This technique can be understood as a transformation in which a lattice configuration, characterized by the driving parameter p, is transformed into a novel configuration p'. This is denoted as

$$p' = R_b(p) \quad \text{and} \quad \xi' = \frac{\xi}{b} \tag{2.1.6}$$

where $R_c(\cdot)$ denotes the coarse-graining transformation which is characterized by the rescaling parameter b, or equivalently, by the selected block size. The correlation length ξ is the relevant length scale of the system, and it is rescaled as given above.

Now, at the critical point, clusters are fractals, and thus they are self-similar, implying that their properties are identical regardless of the scale at which they are considered. Viewing the coarse-graining procedure as a dynamical system, which has a flow in the phase space, implies that the critical point is a fixed point of the transformation. Due to the self similarity of the clusters, the critical point is characterized by $p_c = R_b(p_c)$ and $\xi' = \xi$. From the fixed point of the detailed transformation, the scaling relations and critical exponents for a system can be deduced.

The notion of universality arises from the consideration that the coarse-graining procedure removes effects that are only important on the micro-scale and maintains only those that influence the macroscopic behavior, which is characterized by the critical exponents, and which is shared between different systems with different microscopic behaviors. The formalization of the described process is due to Wilson (1975).

■ 2.1.3 Finite size scaling

The results above have been given for systems of infinite extension. In simulations or in experiments where the system size is small, these are only an idealization. As will be seen below, finite system size leads to systematic deviations of the macroscopic critical quantities that need to be correctly accounted for if simulations and experiments are to be interpreted correctly, and in particular if critical exponents are to be measured accurately. The finite size scaling method by Fisher and Barber (1972) provides a solution to this problem and will be outlined below. For a more rigorous derivation, it is recommended to consult Binder (1997).

Depending on the system under consideration, system size can either be given as the number of interacting units, which is the case on graphs that are the primary interest of this work, or as physical dimension. For instance, in the case of the two-dimensional square lattice, system size is typically given as the edge length L of the lattice. In an infinite system, at the phase transition the correlation length ξ diverges. In a finite system, this is impossible, as the correlation length needs to be cut off when approaching L. As long as $\xi < L$, it is not expected that there is any deviation from the typical behavior, which is the case before and after the transition. The most significant deviations are therefore expected when approaching $p = p_c$.

As a first result, it is noticed that there is a scaling contribution from the system size that shifts the location of the critical point. To distinguish between the critical point in infinite and finite systems, it appears sensible to refer to the latter as a pseudo-critical point. From the scaling of the correlation length $\xi \propto |p - p_c|^{-\nu}$ it follows that near the critical point

$$\begin{aligned} |p - p_c| &\propto \xi^{-1/\nu} \\ &\propto L^{-1/\nu} \quad \text{if} \quad \xi \to L \ , \end{aligned} \tag{2.1.7}$$

indicating that there is indeed a shift of the transition.

Using the first relation, the scaling of the quantities S and $< s >$ near the critical point can be expressed in terms of the correlation length ξ:

$$S \propto (p - p_c)^{\beta} \propto \xi^{-\beta/\nu} \quad \text{for} \quad p > p_c \tag{2.1.8}$$

$$< s > \propto |p - p_c|^{-\gamma} \propto \xi^{\gamma/\nu} \tag{2.1.9}$$

These results are no different from the results for infinite systems, as the cutoff is not yet in place. Nevertheless, the result indicates, that the correlation length directly influences the scaling of the of the other quantities. Typically, this is seen in the average non-giant

component size as a rounding of the divergence, such that the function has a maximum instead of a singularity. The giant component size, on the other hand, is smeared out at the transition.

The exact way, in which the rounding or smearing out affects the critical quantities can be implicitly added to the scaling relations by introducing an unknown, homogeneous function $F(\cdot)$, whose purpose is to correctly reflect the described behavior. $F(\cdot)$ is referred to as a scaling function, and the scaling of $< s >$ is written as

$$< s >= \xi^{\gamma/\nu} F(L/\xi) \ . \tag{2.1.10}$$

with

$$F(L/\xi) \begin{cases} = \text{const.} & L/\xi \to \infty \\ \propto (L/\xi)^{\gamma/\nu} & L/\xi \to 0 \end{cases} \tag{2.1.11}$$

At this point, a second unknown scaling function $\bar{F}(\cdot)$ is defined such that

$$\bar{F}(L/\xi) = (L/\xi)^{-\gamma} F((L/\xi)^{\nu}) \ . \tag{2.1.12}$$

Now, by using the scaling relation for the correlation length and by making use of the fact that $F(\cdot)$ has been defined as a homogeneous function, the correlation length can be eliminated in order to obtain a more useful form of the equation:

$$< s >= L^{-\gamma/\nu} \bar{F}(L^{1/\nu}|p - pc|) \tag{2.1.13}$$

A similar result can be derived for the the finite size behavior of the giant component size

$$S = L^{\beta/\nu} \bar{G}(L^{1/\nu}(p - pc)) \ , \tag{2.1.14}$$

where $\bar{G}$ is another unknown scaling function. Equations 2.1.13 and 2.1.14 are commonly referred to as the finite size scaling hypothesis. Due to the similar structure of the results, they can be stated as

$$X = N^{-\chi/\nu} F\left((p - p_c)N^{1/\nu}\right) \ , \tag{2.1.15}$$

where $F[\cdot]$ is an unknown scaling function and where χ is a critical exponent characterizing the scaling of the critical quantity X.

As can be gleaned from the equations, the finite size behavior of the critical quantities is controlled by the critical exponents. Therefore, working with the finite size scaling method presents a way of obtaining the critical exponents by studying the way in which experimental or computational data deviate due to finite size effects. However, for accurate results there remain certain requirements. A typical data set consists of simulations or experimental realizations of a system governed by the same physics, but for a number of system sizes. For each system size, there needs to be an appropriate number of realizations in order to minimize the effect of stochastic noise, and system sizes that are studied need to be sufficiently large to avoid further systematic errors (Binder, 1997).

Depending on the selected method and the quality of the data, it may be beneficial to have a-priori knowledge of the location of the critical point.

In order to determine critical exponents of a system, it is necessary to select an appropriate method that employs the scaling hypothesis. There exist numerous methods that typically have characteristic disadvantages or may not be well suited for all data.

For instance, it is possible to exploit that the power-law behavior of a critical quantity X on the system size, i.e., $X(p_c) \propto N^{\chi}$ where χ is a critical exponent associated with X, is only accurate at the transition. Therefore, power-law fitting $X(p)$ as a function of the system size N should result in the best fit when $p = p_c$ (see, e.g., Yi et al. (2013)). Fitting the data for numerous values of p and selecting the best fit thus gives an estimate of the critical point and the associated exponent. However, this method requires a large number of realizations, as noisy data is bound to have an influence on the quality of fit. Furthermore, it is helpful to provide a good estimate of the critical point.

In order to circumvent this limitation, Bastas et al. (2014) have proposed a method that is especially well-suited if the number of realizations is small. The requirement that data for numerous system sizes are available persists. Very briefly, Bastas et al. (2014) define the function

$$\Lambda(p, \chi) = \sum_{i \neq j} \left(H_{L_i} - H_{L_j} \right)^2 \tag{2.1.16}$$

where $H_{L_i} = XL_i^{\chi} + 1/(XL_i^{\chi})$. This function in theory equals zero if evaluated at the critical point with the correct critical exponent χ, as the rescaled critical quantity XL_i^{χ} shares either a point of inflection or a maximum at the critical point. In noisy data, it is assumed to be minimal, and the parameters $p \approx pc, \chi$ can be found by minimization of $\Lambda(p, \chi)$. The authors report accurate findings upon application of the method to graph percolation.

Finally, the most well-known method is the data collapse method. Although the scaling function $F(\cdot)$ in equation 2.1.15 is unknown, displaying $XN^{\chi/\nu}$ as a function of the argument $(p - p_c)N^{1/\nu}$ leads to a data collapse, where data for different system sizes collapse onto one unknown master curve if the critical point and critical exponents have the correct values. Thus, determining the parameters that lead to the best data collapse is a method for estimating the critical exponents and the critical point. However, as it will turn out in section 7.5.2, small variations of the critical point can lead to substantial changes of the critical exponents, and it is thus unclear whether it is more accurate to determine first the critical point, for instance by the method of Bastas et al. (2014) and determine only the critical exponents from the data collapse, or whether all parameters should be determined by the data collapse as suggested, for instance, by Melchert (2013).

The data collapse method as implemented here relies on optimization of a quality function proposed by Houdayer and Hartmann (2004), which reads

$$Q = \frac{1}{N} \sum_{i,j} \frac{(y_{ij} - Y_{ij})^2}{\mathrm{d}y_{ij}^2 - \mathrm{d}Y_{ij}^2} \ , \tag{2.1.17}$$

where y, $\mathrm{d}y$ are value and estimated error of the scaled data, and Y, $\mathrm{d}Y$ are value and estimated error of the unknown master curve, and where N represents the number of data points. The quality function is expected to be minimal for the optimal data collapse. The procedure of minimizing Q, however, requires good initial guesses for the parameters. The data collapse procedure used in this work in part relies on work by Melchert (2009); Sorge (2015) that implements the method by Houdayer and Hartmann (2004).

■ 2.2 Introduction to complex networks

Networks pose a ubiquitous aspect of our daily lifes, although in some cases they remain unnoticed. The reason behind this is that the term network describes an abstract concept that can be applied to a variety of systems of very different nature, and on very different length scales. The type of network that is maybe the least abstract are transportation networks. Selected examples that have been studied are the global airplane network (Verma et al., 2014), railway connections (Sen et al., 2003), street networks (Barthelemy et al., 2013) or the power grid (Watts and Strogatz, 1998). These large-scale examples all share that they are man-made. However, at smaller scales, biological transport networks can be discovered. One example is the slime mold network considered in this work, another is the human vascular network (Gamba et al., 2003). Although all examples have a similar purpose, their structures are very different, which is most easily seen when comparing the airplane network, where connections are not physically existent, to the others where rails, power lines or veins are permanently installed.

The abstract nature of networks becomes more obvious when turning to those found in information technology. Whereas the internet (Faloutsos et al., 1999) is a permanently installed network of servers, its various applications rather highlight interactions between services or users. For instance, the world wide web (Barabási and Albert, 1999) represents hyperlinks between web pages, and although each web page is stored on a server that is part of the internet, the hyperlinks do not reflect connections between servers. The same holds for the email network (Ebel et al., 2002), where emails exchanged between addresses can be interpreted as links. Email addresses are typically associated with a person and thus the email network can be viewed as a social network. Social networks are likely the most visible abstract network type, represented either as a network of real acquaintances (Davidsen et al., 2002) or as one of the many platforms enabled by information technology. However, there are less visible social networks that have huge impact on society and wealth, success in certain fields, or on health. Prominent examples are the network of company ownership (Newman et al., 2001), networks of citations between scientific publications (Redner, 1998), or even the network of human sexual contacts (Liljeros et al., 2001).

Returning to the biological sciences, various examples such as the metabolic network or the network of protein interactions (Jeong et al., 2000, 2001), or even ecological food webs (Sole and Montoya, 2001) highlight that abstract interactions in complex systems can be mapped to a network description.

The examples given above indicate that network approaches can be applied to a much

broader range of subjects than what is covered within this work, and it is therefore not intended to give a comprehensive review. Instead, it is recommended to consult the number of excellent reviews and books written in the period after network science boomed around 2000, for instance (Albert and Barabasi, 2002; Dorogovtsev and Mendes, 2002; Newman, 2003; Barabási and Oltvai, 2004; Boccaletti et al., 2006; Dorogovtsev et al., 2008; Barthelemy, 2018). Although research on networks and graph has been ongoing for centuries with the first modern models dating back to the 1950s, recent excitement is mainly due to a number of discoveries made around 2000, which are in part due to the increased availability of large data sets and the methods required to treat those.

Having addressed a number of applications for network science, before proceeding it is necessary to outline how networks are understood in a mathematical sense. This is done below in an abbreviated fashion. The picture is expanded throughout this work whenever novel concepts are required.

■ 2.2.1 Fundamental concepts

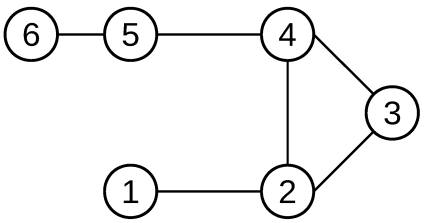


Figure 2.2.1: Example of a simple graph with one single connected component.

Mathematics views networks as graphs, i.e., the ordered set

$$G = (N, E) \ , \tag{2.2.1}$$

where N is the set of nodes of the graph, and E is the set of edges connecting these. A basic example is shown in figure 2.2.1. In the present graph, nodes have been labeled by numbers, and edges are the lines connecting these nodes. The sets of nodes and edges of this example are

$$N = \{1, 2, 3, 4, 5, 6\}$$
$$E = \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{4, 5\}, \{5, 6\}\} \ .$$

The abstract nature of this description lies in the fact that nodes and edges can represent very different entities, or even concepts and actions. For instance, with regard to the example given above, nodes could be individual persons, and edges might symbolize acquaintance. In a different picture nodes might still be persons, however represented by email addresses. In this case, edges could be emails exchanged between those addresses. This notion highlights that in some scenarios, it is not sufficient to denote only the sets of nodes and edges. For instance, if one node addressed another node in an email, but not vice versa, it might for some applications be sensible to denote the edge as a directed

edge, pointing only from the first node to the second. In this case, the set of edges cannot consist of unordered sets, but has to contain a number ordered sets. In the present work, only undirected graphs are considered, which simplifies a number of concepts.

There exist numerous ways of representing the structure of an undirected graph in terms of mathematical objects. One has been given above, another worth mentioning is the adjacency matrix (a_{ij}), whose matrix elements are defined as

$$a_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected} \\ 0 & \text{else} \end{cases} . \tag{2.2.2}$$

In most applications, the diagonal elements of this matrix are zero, as edges leading from one node to the same node are in many cases not desirable. Another strength of the adjacency matrix notation points to a property typically observed in spatial networks, which has not been considered so far. Edges may have a length, denoting, for instance, the euclidean distance between two junctions of a road network. This is easy to represent in the adjacency matrix notation, as it is possible to consider instead of the binary a_{ij} indicating connectivity, the pairwise distance matrix $d_{ij} \cdot a_{ij}$. Similarly, it would be possible to denote other edge properties in a matrix notation, for instance, resistance to flow along an edge.

Interpreting the graph in figure 2.2.1 as a spatial road network and nodes as junctions of these roads leads naturally to the question why the nodes labeled as three and five should be part of the graph, as they are not actually junctions but merely another point between nodes two and four, or four and six, respectively. The same road topology could be represented by removing nodes five and three and joining the adjacent nodes by longer edges. This, however, raises a number of problems in practical applications. One such problem arises if node three is removed: nodes two and four would then be connected by not one, but by two direct edges, which cannot be correctly represented by the adjacency matrix unless it is redefined accordingly. Joining the double edge into one single edge appears reasonable in terms of connectivity, but if the graph is meant to represent a road network, this solution would incorrectly suggest that only one route can be taken from two to four, although two roads exist. Further problems will be discussed in greater detail in chapter 5. For now, it appears more advisable to remain with the notion that nodes may take a different role, depending on how they are embedded in the graph.

The arguably most important property of a node with regard to how it is embedded in the graph is its degree k, defined as the number of direct neighbors to which the node is connected. In the given example, nodes one and six are of degree one, nodes three and five as considered in the previous paragraph have degree two and nodes two and four are actual junctions of degree three. Not shown are nodes of higher degree, or those that have no neighbors, although the latter will be of relevance to this work. Throughout this work, a quantity termed the degree distribution p_k will be considered, which is an extremely useful tool for characterizing the structure of a graph. It is defined as the probability to obtain a node of degree k, if randomly selecting one node from the graph. As the selected node is bound to have some degree (even if it is zero), there exists a

normalization condition for the degree distribution:

$$1 = \sum_k p_k \tag{2.2.3}$$

Numerous graph measures may be defined based on the degree distribution. For instance, it is possible to define the average degree $< k >$ of a network as

$$< k >= \sum_k k p_k \tag{2.2.4}$$

The average degree can be viewed as a scalar measure of how structured a graph is, although it hold less information than the degree distribution.

Including those based on the degree distribution, there exists a variety of measures intended for characterizing the structure of a graph. Again, it is not intended here to proved a complete review, as only few measures are made use of in this work. It is advised to consult the reviews referenced above, or the review by da F. Costa et al. (2007), which focuses on the diverse array of network measures. Notable measures include, for instance, the clustering coefficient

$$C = \frac{1}{|N|} \sum_i \frac{n_{i,\,\text{triangles}}}{n_{i,\,\text{triplets}}} \ , \tag{2.2.5}$$

which compares the number of triangles of which the i'th node is part (e.g., nodes $2, 3, 4$ in figure 2.2.1) to the number of triplets centered on that node. Selecting node $i = 2$ in figure 2.2.1 leads to three connected triplets (nodes $(1, 2, 3)$, $(1, 2, 4)$, $(3, 2, 4)$) centered on that node. Evaluating the clustering coefficient for an entire graph gives an indication of whether the graph has a tree-like structure or whether it is best viewed as a dense mesh.

A class of measures that is of obvious interest to spatial networks but is also highly relevant to others, is related to shortest paths on the graph topology. For instance, it is possible to compute the shortest paths l_{ij} between all pairs of nodes in a graph, and to calculate from those either the average shortest path length

$$< l >= \frac{2}{|N|(|N|+1)} \sum_{i \geq j} l_{ij} \ , \tag{2.2.6}$$

or a measure termed betweenness centrality which is defined as follows. For each individual node k in a graph, one can evaluate the quantity σ_{kij}, which is equal to one only if the shortest path from node i to j passes through node k. Computing

$$B(k) = \frac{2}{(|N|-1)(|N|-2)} \sum_{i \neq j} \sigma_{kij} \tag{2.2.7}$$

gives a single value for each node that indicates whether the node is central in a graph. This is based on the assumption that efficient function of a network is related to routing transport or information along shortest paths.

2.2.2 Models of network structure

Section 2.2 has introduced a number of very different systems sharing at first glance little more than the fascinating property that they can be viewed as networks. Historically, first models predicting the properties of a network emerged around 1950 in the form of the Erdös-Rényi random graph model (Erdös and Rényi, 1959), which remained the predominant model for studying network structure for about fifty years. Although the Erdös-Rényi model enabled numerous novel insights into the structure of complex networks, it was recognized with the increased availability of large data sets featuring networks, and with the emergence of methods for evaluating or simulating those, that in a variety of cases networks would possess structure beyond the inherent randomness of the random graph.

The field of complex networks research experienced a paradigm shift around 2000 due to developments driven in no small part by two novel concepts. The first is the extension of the small world phenomenon (Milgram, 1967) to complex networks by Watts and Strogatz (1998), indicating that the crossover from regular, lattice-like networks to random graphs is accompanied by a drastic drop in average in average path length.

Second, it was noticed that the degree distributions of real networks would significantly deviate from the Poisson-type distribution seen in random networks. In particular, it has been observed that a wide variety of systems share a degree distribution characterized by a power law tail. A mechanism thought responsible for the emergence of these scale-free networks has been proposed by (Barabási and Albert, 1999).

In this section the three models addressed above are briefly introduced. Furthermore, the configuration model Molloy and Reed (1995) is outlined. This model presents an algorithmic method of generating network topologies that obey a desired degree distribution. It is, for instance, possible, to generate a scale-free network by invoking the configuration model. The display is kept brief on all accounts, as the experimentally obtained network topologies analyzed in this work are neither expected to be scale-free nor possess the small-world property[1]. However, both concepts are central to current complex networks research and they are used for comparison in this work. In contrast, the random graph and configuration models are used extensively here. However, their relevant properties are addressed separately, in particular in section 2.3 and in chapter 4.

The random graph model The principle underlying the construction of random graphs in the Erdös-Rényi model (Erdös and Rényi, 1959) is straightforward. A system is initialized by defining a number N of initial nodes. In such a system, it would be possible to place $N(N-1)/2$ edges, and one defines a parameter $p \in [0,1]$ indicating the fraction of edges that is present in the graph. The parameter p thus characterizes an ensemble of graphs random graphs with N nodes and $E = pN(N-1)/2$ edges, and where each possible realization is equiprobable. Single realizations can be constructed by

[1]The latter is most likely not entirely true in the functional *P. polycephalum* network, if the hierarchical organization is taken into account. Backbone veins of the network have a long persistence length and can swiftly transport contents over large distances.

starting with N unconnected nodes and sequentially adding edges between random pairs of nodes until p has the desired value. This process is realized via simulations in chapter 7. Based on the definition given in the previous section, the average degree $< k >$ is related to the parameter p as

$$< k >= 2\frac{E}{N} = p(N-1) \ . \tag{2.2.8}$$

From the property that for one p all edges are present with equal probability, the degree distribution, which is of binomial type, can be derived. It is found that it takes the form (Newman, 2003)

$$p_k(p) = \binom{N}{k} p^k (1-p)^{N-k} \ . \tag{2.2.9}$$

Taking the limit $N \to \infty$ at constant average degree leads to the Poisson-type degree distribution in infinite random graphs:

$$p_k = < k >^k \frac{e^{-<k>}}{k!} \tag{2.2.10}$$

Depending on the value of p, the random graph will either be fragmented into disconnected components or have a single giant component. This highlights the similarity of graph construction to percolation processes as outlined in section 2.1. Percolation in the random graph model is further discussed in section 2.3.1. Analytical solutions to the percolation problem in random graphs with small degree are given in section 4.1.

The configuration model The random graph model is limited to the construction of random graphs with a degree distribution as given above. The configuration model presents a way of extending the construction of random graphs to arbitrary degree distributions.

Briefly, graph construction in the random graph model functions as follows. The degree distribution p_k needs to be specified, and a number of nodes N is defined. Then, N integer random numbers are drawn according to the probability distribution p_k, and each node is assigned one of these numbers. For each node, this number is equal to the number of connections it will have in the final configuration, and thus represents its degree k. Each node is said to have k 'stubs', sites to which veins can attach. For the construction procedure, random pairs of stubs are selected with equal probability and connected by an edges. This is repeated until no free stubs are left in the graph.

It is important to emphasize the difference to the random graph model, where random pairs of nodes are selected with equal probability. In contrast, in the configuration model, selecting random pairs of stubs instead of nodes means that a node with a larger degree will have a larger probability to obtain an edge. This will be of importance in section 2.3.2, where percolation in the configuration model is discussed. Analytical solutions to percolation in the configuration model with exclusively small node degrees are derived in section 4.2.

The small world property The term small world was first coined by Milgram (1967) for describing an experimentally observed property of social networks. The idea it conveys is that in many real networks, the average shortest path length is much shorter than the number of nodes in the network would suggest. Although this phenomenon has been studied empirically, its causes could not be traced back to graph theory until Watts and Strogatz (1998) presented a model that provided an explanation for the observed behavior. At the same time, this model was one of the first theoretical notions that statistical network models could lead to structures beyond the random graph model.

The following example illustrates the small world phenomenon. Consider a transportation network in two dimensions, for instance, a road network. On that network, distance when traveling between two points can be measured either in terms of nodes encountered on the way, or in terms of euclidean distance in the plane. If traveling between two points that are separated by a large euclidean distance, the number of nodes encountered in the shortest path will scale with the euclidean distance. Now, if the road network is augmented with a network of flight connections, the situation changes. Instead of searching the shortest path in terms of encountered nodes in the plane, it is much more efficient to use an airplane connection instead. These function as global edges, effectively decreasing the shortest path distance in the network significantly. Although the present example is instructive, it is clear that the shortest path distance is only reduced in terms of nodes encountered, and not in terms of euclidean distance. However, the example translates to networks that are not spatially embedded.

The model by Watts and Strogatz (1998) reflects the example given above. The steps required for the construction of a small world network can be stated as follows. As initial condition, a regular graph is generated with all node degrees being equal to the average degree. A possible example would be the triangular planar lattice. Following the initialization, depending on a parameter p a number of the edges present in the graph is rewired, meaning that for each chosen edge, one end is selected and connected to a random node in the graph. For small p, the resulting graph is unchanged from the original, regular graph, and for large p, the graph assumes a configuration comparable to the those created by the random graph model.

Watts and Strogatz (1998) characterize the behavior of the system upon crossing over from the regular to the random regimes by the average shortest path length and by the clustering coefficient. In regular graphs, as seen in the example given above, the shortest path length is typically very large, whereas it is small in random graphs. Similarly, the clustering coefficient is large in regular graphs, and small in random graphs. However, it takes a large number of rewiring steps to decrease the clustering coefficient, whereas it is observed that the shortest path length drops rapidly when only a small number of edges has been re-targeted.

The small world effect is observed in a large number of real world networks and has important implications, for instance, for the spread of infectious diseases, computer viruses, or even information. All of these 'travel' on a network and can reach distant regions significantly faster if there are global edges that can be used.

The model by Watts and Strogatz (1998), while instructive for illustrating the small world effect, is not a detailed representation for real networks. In particular, it is typically observed that the degree distributions of real networks do not agree with those produced by the model. There are various structures that enable a real network to be considered as a small world network. For instance, networks in which nodes are partitioned into cliques or communities, and that are densely connected within each community or clique, whereas edges between these structures are rare, typically show the small world effect. The same holds for the scale-free networks considered in the next section. These are characterized by the presence of 'hub' nodes with a large degree. Routing from one node in the network to another is typically most efficient if moving over a hub node. Similarly, random graphs have short shortest path lengths and small clustering coefficients. Taking into account that edges in a network might have different functions can also lead to a small world property. Revisiting the example of a road network given above, complementing the network with highway roads instead of flight connections has a similar, although less efficient effect that does, however, not require to break the plane embedding. These examples show, that the small world property is not in opposition to other network types, but rather serves to further characterize them.

Scale-free networks Scale-free degree distributions are another universal pattern observed in networks of diverse nature. It was first noticed and explained in a network by de Solla Price (1965) in citation networks of scientific publications and later rediscovered by (Barabási and Albert, 1999). Briefly, scale-free networks are characterized by a heavy-tailed degree distribution, meaning that in these networks there exist nodes of large degree with a non-vanishing probability. This contrasts random networks, whose degree distribution is exponentially bounded for nodes of large degree. Typically, the heavy tail is a power law behavior of the form

$$p_k \propto k^{-\gamma} \ , \tag{2.2.11}$$

where the exponent γ is the central parameter for characterizing the system. Most scale-free networks have an exponent of $2 \leq \gamma \leq 3$, although occasionally larger exponents are measured. For $\gamma > 3$, networks are typically indistinguishable from random networks.

Many of the network examples given earlier in this section have been claimed to show power law behavior. This includes, among many others, the airplane network, the internet, world wide web and email networks, but also biological networks such as the metabolic network or protein interactions, although upon reanalysis with more rigorous statistical methods, several of these of these claims do not hold, hold only in part and in general are now seen more critical (Clauset et al., 2009).

Nevertheless, the existence of large degree nodes has important implications for the structure of a network. In many cases, low degree nodes in scale-free networks are only loosely organized into communities associated with a large degree node. Large degree nodes are connected among each other, thus maintaining the global connectivity of the network. In the first place, networks with this type of structure are characterized by low shortest path lengths, as these paths can be routed through the large 'hub' nodes.

Thus, they classify as small world. However, although some scale-free networks show significant levels of clustering (for instance, in protein interaction networks), this feature is not required for classifying as scale-free, and is not reproduced in the basic form of the most well-known model for scale-free networks.

The existence of hub nodes presents both, the main strength of scale-free networks, but also their weakness, which is best seen in the percolation properties: the percolation threshold in infinite scale-free networks lies at $p = 0$ (Cohen et al., 2000), indicating that these networks are always in possession of a giant component. From a different point of view, this is equivalent to stating that removing randomly any number of nodes from a scale-free network will not cause the network to break down. This behavior is due to two properties. The probability of randomly selecting a hub vanishes in comparison to the probability of selecting a node of small degree. Removing such a small degree node has a vanishing effect on the structure of a scale-free network, whereas the removal of such nodes in a random graph progressively deconstructs the network. Even in the improbable event that a hub is randomly selected, there is likely another hub that can take over. On the other hand, directed and repeated removal of hub nodes is expected to have a drastic effect on the topology of the network.

The mechanism leading to the emergence of scale-free topologies is well understood, and in its basic for has been proposed independently by de Solla Price (1965), and by Barabási and Albert (1999) who coined the name preferential attachment. The rationale behind preferential attachment is that nodes that are added to a network will preferably attach to a node that is already strongly connected. This can be understood in terms of the citation network used by de Solla Price (1965) where it appears reasonable to assume that a publication with high visibility, i.e., with a larger number of citations will be cited preferably in new publications. The same considerations hold for the model by Barabási and Albert (1999) for pages on the world wide web. Preferential attachment is typically defined in terms of the probability of a node i to receive a new link

$$p_i = \frac{k_i}{\sum_j k_j} \ . \tag{2.2.12}$$

In most preferential attachment models, new nodes do not connect to only one existing node, but rather to a number of existing nodes. Preferential attachment, unlike the wiring or rewiring processes detailed in the previous sections, is a mechanism of network evolution and growth. Since the original publication, the preferential attachment model has been diversified in a number of ways, including nonlinear preferential attachment (Krapivsky et al., 2000) and the pairing with other mechanisms modifying network structure (see, e.g., (Albert and Barabási, 2000; Ghoshal et al., 2013)). The mathematical tool of choice for implementing preferential attachment has turned out to be the master equation (Dorogovtsev and Mendes, 2002; Krapivsky and Redner, 2003), which is also used in the present work (see chapters 3 and 7). The model devised in chapter 7 can integrate preferential attachment in both, growth and wiring processes, although the networks considered as part of this work do not fall into the scale-free regime.

2.3 Percolation on graphs

The central question of network percolation can be compactly stated as follows: given a set of local properties characterizing the topology of a graph, with which probability will this graph contain one giant component in which the majority of nodes are connected, rather than being fragmented into small, disconnected components. Although this question generally describes a static relation between local topological properties and a global property of the graph, it is in some cases possible to interpret the transition between fragmented and connected states as a time-dependent process, namely, when the topology is developed or degraded as its local properties change with time.

The relevance of the underlying question becomes immediately clear when considering the following examples. Graphs can be used to represent the (static or dynamic) coupling of a dynamic process. Without any other long-distance interaction other than that conveyed by the graph topology, such processes cannot exhibit global patterns if the topology is fragmented. Both states can be desirable: fragmenting one transportation network may prevent the spreading of an infectious disease, whereas another represents the breakdown of a power grid after a cascade of failures (D'Souza, 2017). Learning about the rules that govern the transition of a network from one state into another, is one step in the direction of being able to exploit the implications that topology imposes on the function of the network. Scale-free networks are a paradigm of this principle: it has been shown that a percolation phase transition in scale-free networks cannot be induced by accumulating random damage, whereas their topology is extremely susceptible to damage targeted at the highly connected hub nodes (Albert et al., 2000), enabling for example control of disease spreading (Pastor-Satorras and Vespignani, 2001).

Many important properties of real-world network models can be tied to the degree distribution p_k. Recapitulating, p_k can be understood as the probability that a randomly selected node in a graph has exactly k neighbors. Similarly, $p_k = N_k/N$ can be defined as the fraction (or population) N_k of nodes of degree k in a graph containing a total of N nodes. The necessary condition that $N = \sum_k N_k$ is equivalent to a normalization condition for the degree distribution:

$$1 = \sum_k p_k \tag{2.3.1}$$

2.3.1 Percolation in the random graph model

In the random graph model, a relation for the existence and expected size of a giant component can be derived from a compact heuristic argument (Newman, 2003). First, the set of nodes in a random graph is divided into two complementing fractions u and S with

$$S = 1 - u \ . \tag{2.3.2}$$

These fractions are equivalent to the probabilities of randomly selecting from the graph a node that is not part of a giant component (in the case of u) or a node that is part of the largest component (in the case of S). Focusing on the former probability, i.e., $u = 1 - S$,

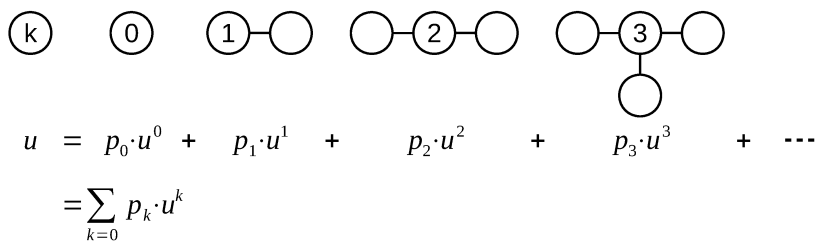


Figure 2.3.1: Illustration of the self-consistency relation (equation 2.3.5) for the probability u of a node not to be part of the giant component: if a node of arbitrary degree k is not part of the giant component, neither are its neighbors. Drawn in the style of Newman et al. (2001).

the following intuitive statement can be made: a randomly selected node can only be part of a non-giant component if none of its k neighbors are part of the giant component. The probability u of selecting such a node consists of two parts. In the first place, the probability of selecting a node of degree k is given by the degree distribution p_k, and in the second place the probability of a given node of degree k having exactly k neighbors that are not part of the giant component is u^k. Combining the two arguments, the joined probability of selecting a node of degree k that has no neighbors that are part of the giant component is thus $p_k u^k$. Summing over all possible node degrees (corresponding to a marginalization of the variable k from the joined probability) leads to

$$u = \sum_k p_k u^k \ . \tag{2.3.3}$$

The arguments leading to this self-consistent relation can be visualized as shown in figure 2.3.1.

Equation 2.3.3 is one useful way of formulating a self-consistent equation for the probability u, from which the size of the largest component can be gleaned via equation 2.3.2, given a solution can be found by some method.

■ 2.3.2 Percolation in the configuration model

In the configuration model, the formulation of a useful expression for the existence of a giant component is not as straightforward, mainly due to the reason that the probability p_k of selecting from the graph a random node of degree k does not equal the probability of reaching a node of degree k by following a randomly selected edge (Newman, 2003).

The above statement deserves special attention, as it is at first glance as unintuitive as it is crucial, and it is easily misunderstood. It especially important to realize that the self-consistency relation given in the previous section (equation 2.3.3) only holds in the case of the random graph model. The reasons are as follows. First, it is necessary to recall that the study of the random graph or configuration models, as represented by the respective degree distributions p_k, is not the study of one graph created by either of the models, but rather the study of the ensemble of graphs that is in accordance with the respective generation rules. These are briefly repeated below.

(i) In the random graph model, randomly selected pairs of nodes are connected with uniform probability. As a consequence, each edge is present with equal probability at each point during the wiring process; each pair of nodes is connected with equal probability. The detailed degree distribution at a that point is dependent on the number of edges that have been brought into place. If the number of nodes in the graph is infinite, the resulting Poisson distribution never changes shape although the parameters of the distribution change with the number of edges in the graph. The degree distribution results from the generation process.

(ii) In the configuration model, nodes are created with a number of k stubs according to the specified degree distribution p_k. In the course of the wiring process, stubs, not nodes, are connected with uniform probability. This leads to an important difference: the probability of a node to receive an new edge is proportional to the number of stubs. The degree distribution governs the generation process.

Summarizing, in the random graph model the probability for a node of degree k to receive a new edge is p_k, whereas in the configuration model this probability is proportional to kp_k in the limit of large graph size. The second crucial realization is that in the ensemble of graphs created by either model, the process of placing a novel edge is identical to the process of following a randomly selected edge that is already present in the graph. By the above rationale it becomes clear that the arguments leading to the formulation of the self-consistency relation 2.3.3 are only valid in the random graph model, and a different approach has to be taken in the configuration model.

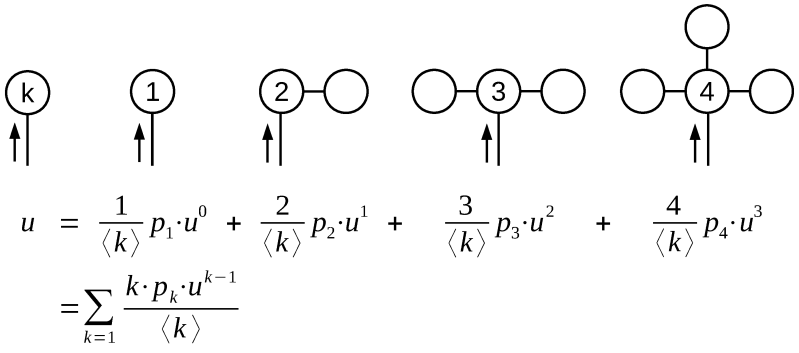


Figure 2.3.2: Illustration of the self-consistency relation (equation 2.3.5) for the probability u of a randomly selected edge to lead into a non-giant component: if randomly arriving at a node of degree k that is not part of the largest component, none of its $k-1$ other neighbors can be part of the giant component. Drawn in the style of Newman et al. (2001).

Provided the reasoning above, it is not difficult to produce a comparable heuristic argument that leads to a self-consistency relation for a quantity u, which now states the probability that when following a randomly selected edge one arrives at a node of any degree k whose $k-1$ outgoing[2] edges to not lead to the largest component. Above, it

[2]The models considered here have been designed for undirected graphs. The terms 'inbound' and 'outgoing' are used to define the edges on which a node is approached, and on which one would leave to explore its other neighbors.

has been stated that the probability to arrive at a vertex of degree k in the configuration model is proportional to kp_k. Normalization and the definition of the average degree in a graph, i.e., $< k >= \sum kp_k$, require that the probability distribution is of the form

$$q_k = \frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{< k >} , \qquad k \geq 1 . \tag{2.3.4}$$

Now, analogously to the arguments leading to equation 2.3.3, the probability of arriving via a random edge at a node that has exactly degree k and thus $k-1$ outgoing edges leading to nodes that are not part of a giant component is $q_k u^{k-1}$. Summing over all possible node degrees[3] and invoking equation 2.3.4 leads to the self-consistency relation for the configuration model:

$$u = \sum_{k=1} \frac{kp_k u^{k-1}}{< k >} = \sum_{k=0} \frac{(k+1)p_{k+1} u^k}{< k >} \tag{2.3.5}$$

The preceding arguments have been illustrated in figure 2.3.2. Dorogovtsev and Mendes (2003) arrive at this result by presenting a similar heuristic argument. It should be pointed out here that the derivation given above is not rigorous, as it implicitly excludes the possibility of loops in the non-giant components or rather, that the non-giant components possess a tree-like structure. However, Newman (2003) points out that the possibility of finding more than a single path between two nodes in the configuration model scales, similar to the number of nodes in the non-giant components as $\mathcal{O}(N^{-1})$, N being the number of nodes in the graph. Thus, in graph of sufficient size the influence of loops can be considered negligible.

Still, equation 2.3.5 cannot be used yet to calculate the size of the giant component. In the random graph model, the size of the largest component was defined as $S = 1 - u$, u being the fraction of nodes in non-giant components. In the configuration model, the quantity u has been defined quite differently and and it is necessary to find a different relation between S and u, which might be formally written as $1 - S = f(u)$. To do so, the following section describes the generating function formalism introduced by Newman et al. (2001), in which such a relation can be established.

■ 2.3.3 The generating function formalism

Although several authors have treated the configuration mathematically and obtained similar results, the probably most recognized treatment has been given by Newman et al. (2001) which will be presented here in an abbreviated form following Newman (2003). Therein, the authors make use of what is referred to as a generating function formalism. Generating functions (Wilf, 1994) have been named in accordance with their characteristic property of generating a probability distribution and the moments thereof as derivatives of the n'th order. Formally, generating functions are given as power series whose coefficients are given by a probability distribution. Generally, an ordinary generating function for a probability distribution in one variable, e.g., $P(k)$ is given as

$$G(x) = \sum_k P(k)x^k . \tag{2.3.6}$$

[3]Note, that this sum excludes $k = 0$ as no edge may arrive at a solitary node.

Selected properties of the generating function (Newman et al., 2001), including normalization

$$G(1) = 1 \ , \tag{2.3.7}$$

generation of the probability distribution

$$\frac{1}{k!}\frac{\partial^k G(x)}{\partial x^k}\bigg|_{x=0} = P(k) \tag{2.3.8}$$

or generation of the n'th moments

$$\left(x\frac{\partial}{\partial x}\right)^n G(x)\bigg|_{x=1} =< k^n > \tag{2.3.9}$$

highlight their usefulness for transforming discrete mathematical problems so that they can be treated by continuous methods. Indeed, in some cases generating functions provide a way to treat equations such as the master equation discussed in chapter 3 analytically by transforming into a (partial) differential equation, see e.g., (Toral and Colet, 2014).

Returning to the problem at hand, Newman et al. (2001); Newman (2003) define generating functions for the probability distributions p_k and q_k and in the case that there is in the graph. Both distributions are, as detailed in the previous section, the probabilities of finding a node of degree k, either when selecting randomly from the set of nodes of a graph, or, respectively, when following a random edge to its end. q_k is referred to as the excess degree distribution, and its exact definition is given in equation 2.3.4.

In accordance with equation 2.3.6, the generating functions for p_k and q_k read

$$G_0(x) = \sum_k p_k x^k \quad \text{and} \quad G_1(x) = \sum_k q_k x^k \ . \tag{2.3.10}$$

In order to find an expression for the self-consistency relation in terms of generating functions, $H_1(x)$ is defined to be the generating function for the probability of non-giant component sizes that may be encountered when selecting a random edge and following it to the node at its end. The properties of $H_1(x)$ differ from the generating functions given above in the sense that it is normalized with regard to the number of edges leading into non-giant components, so $H_1(1) = 1$ only if there is no giant component in the graph. If a giant component exists, $H_1(x)$ is no longer normalized according to equation 2.3.7 but instead takes the value $H_1(1) = u$ (Newman et al., 2001).

A feature necessary for the formulation of generating functions of more complex distributions states that the distribution of the accumulated value of an independent stochastic variable over a number of realizations is generated by a generating function that is equal to the corresponding power of the generating function for the distribution of a single realization. Paraphrasing, this means that if the generating function $G(x)$ for the distribution of a random variable is known, the distribution of the additive value of that variable over n realizations is generated by $G(x)^n$.

Newman et al. (2001) employ this property to formulate a self-consistency relation for $H_1(x)$. Turning to figure 2.3.2 that is a visualization of the self-consistency equation (equation 2.3.5), one may state the following. Given the tree-like structure of the non-giant components, arriving with probability q_k via some inbound edge at a node of degree k means that each outgoing edge leads to a component with a size generated by $H_1(x)$, shown in the figure as a node with no degree specified. If considering at first only nodes that have no connections other than the inbound edge, the distribution of the remaining unity-sized components is uniform and generated by $q_0 x$. Arriving in a similar fashion at a node with a degree larger than one, it will be connected to a corresponding number of components (given by q_k), each with a size distribution generated by $H_1(x)$. In accordance with the powers property given above, realizing k such components leads to a total size distribution generated by $H_1(x)^k$. Now, merging these arguments and summing over the k possible node degrees, one arrives at

$$H_1(x) = x \sum_{k=0} q_k H_1(x)^k \ . \tag{2.3.11}$$

Comparing with the definition of $G_1(x)$ (equation 2.3.10), this result can be rewritten as

$$H_1(x) = x G_1(H_1(x)) \ , \tag{2.3.12}$$

which is the representation of the self-consistency relation of the configuration model (equation 2.3.5) expressed in terms of generating functions. In analogous fashion, defining a generating function $H_0(x)$ for the size distribution of non-giant components when selecting a random node leads to

$$H_0(x) = x G_0(H_1(x)) \ . \tag{2.3.13}$$

The latter expression is not a self-consistency relation but represents, in terms of generating functions, a connection between the size distributions u of finding a component of a given size when arriving via a randomly selected edge, and $1 - S$ of finding a component of a given size when selecting a random node. Due to the analogous definitions, $H_0(x)$ is normalized similar to $H_1(x)$. Substituting the normalization conditions for both generating functions in the case that a giant component is present in the graph, i.e., $H_0(1) = 1 - S$ and $H_1(1) = u$ leads to the sought-for expressions for the size of the giant component

$$1 - S = G_0(u) \ , \tag{2.3.14}$$

and reproduces the self-consistency relation (equation 2.3.5)

$$u = G_1(u) \ . \tag{2.3.15}$$

If a solution of the self-consistency relation can be found, the determination of the size of the giant component via equation 2.3.14 is straightforward. Similarly, the average size of non-giant components can be found by employing the normalization condition and equation 2.3.9 (Newman et al., 2001):

$$< s >= 1 + \frac{< k > u^2}{(1 - S)(1 - G_1'(u))} \quad \text{where} \quad G_1'(u) = \left. \frac{\partial G_1(x)}{\partial x} \right|_{x=u} \tag{2.3.16}$$

Knowing that a giant component exists only in a certain region implies the existence of a transition between these regions. Expressions for this phase boundary can be determined from the self-consistency condition as is demonstrated in 4. The exact condition has first been given by Molloy and Reed (1995) and is therefore referred to as the Molloy-Reed criterion. It states that at the transition,

$$0 =< k^2 > -2 < k >= \sum_k k(k-2)p_k \ . \tag{2.3.17}$$

Notably, not only can the average size of non-giant components be calculated in the scope of the configuration model, but also an explicit expression for the probability distribution of the component size has been given by Newman (2007):

$$n_s = \begin{cases} p_0 & \text{if} \quad s = 1 \\ \frac{<k>}{(s-1)!} \left[\frac{\mathrm{d}^{s-2}}{\mathrm{d}x^{s-2}} G_1(x)^s \Big|_{x=0} \right] & \text{if} \quad s > 1 \end{cases} \tag{2.3.18}$$

n_s represents the probability that a randomly selected node belongs to a component of size s and is also referred to as the cluster size distribution or the concentration of clusters.

Chapter 3

Stochastic dynamics

"When the probabilities of the elementary processes are known, one can write down a continuity equation, from which all other equations can be derived and which we will call therefore the master equation."
— Nordsieck et al. (1940)

Classical deterministic models are of limited use in biological systems where macroscopic behaviors result from stochastic interactions of only a small number of particles. Due to intrinsic noise (and possibly additional extrinsic noise (Hilfinger and Paulsson, 2011)), the outcome of a single experimental trial performed in this regime is highly probabilistic. As a consequence, macroscopic averages are not meaningful unless a large number of trials can be performed. To improve the quality of predictions for single experimental trials, stochastic models have to be employed that can reproduce higher moments of the probability distribution from which the result is drawn.

This chapter presents the master equation as a model for continuous time stochastic processes on a discrete state space that have the Markov property. The master equation details the time evolution of occupation probability of any state that can be realized by a system, and thus contains all necessary information to characterize the full probability distribution at a given time. However, the master equation commonly resists analytical solution, and numerical treatment is limited to complex methods such as the finite state projection method (Munsky and Khammash, 2006; Dinh and Sidje, 2016), or to the simulation of single stochastic trajectories employing the stochastic simulation algorithm and its refinements (Gillespie, 2007).

The remainder of the chapter is organized as follows. The fundamentals of stochastic processes are established in section 3.1, where the Markov property of memory-less processes is introduced in section 3.1.1. From the Markov property a general identity for transitions between states of a stochastic process is formulated in section 3.1.2, which is termed the Chapman-Kolmogorov equation. This identity provides the basis for deriving the master equation, a gain-loss equation for the occupation probability of micro states. This is done in section 3.1.3, with the final result given in section 3.2. There exists more than on way of stating the master equation. An alternative formulation which is particularly useful in conjunction with networks of chemical reactions is given in section

3.3. The chemical master equation builds on the idea that complex chemical reactions can typically be broken down into sequences of elementary reactions. This principle is further outlined in section 3.3.1, leading to the formulation of propensity functions in section 3.3.2, which provide a relation between reaction rate constants and the probability to observe a reaction in a given interval. It is than motivated in section 3.3.3, that the master equation can be simplified to a deterministic rate equation describing the average of the set of stochastic trajectories described by the master equation. Finally, as within the scope of this work the master equation will not be treated analytically, the stochastic simulation algorithm is introduced in section 3.4 as a tool for simulating exact stochastic trajectories from the master equation.

3.1 Markov processes and the Chapman-Kolmogorow equation

Following Gardiner (1985), a stochastic process describes the evolution of probability of a stochastic variable x. Such a time[1] evolution can be formulated in terms of a joint probability

$$P(x_N, t_N;\ x_{N-1}, t_{N-1};\ \dots\ x_0, t_0;\ x_0, t_0)\ , \tag{3.1.1}$$

for the variable x to assume the value or state x_0 at time t_0, x_0 at time t_0, and so on. A conditional probability for the realization of a future state x_{N+1}, t_{N+1}, given the past states $x_N, t_N\ \dots\ x_0, t_0$ have been realized according to the joint probability (equation 3.1.1), can be defined as follows:

$$P(x_{N+1}, t_{N+1} \mid x_N, t_N;\ \dots\ x_0, t_0) = \frac{P(x_{N+1}, t_{N+1};\ x_N, t_N;\ \dots\ x_0, t_0)}{P(x_N, t_N;\ \dots\ x_0, t_0)} \tag{3.1.2}$$

3.1.1 The Markov property

A stochastic process has the Markov property, if the conditional probability for the realization of the future step x_{N+1}, t_{N+1} does not depend on the full history of states, but rather only on the present state x_N, t_N:

$$P(x_{N+1}, t_{N+1} \mid x_N, t_N;\ \dots\ x_0, t_0) = P(x_{N+1}, t_{N+1} \mid x_N, t_N) \tag{3.1.3}$$

A process with the Markov property is said to have no memory of past states other than the most recent state. Paraphrasing, the transition probability into a future state does not depend on how the present state has been realized, and thus can be interpreted as a property of the state rather than the process. Iteratively combining the definition of the joint probability (equation 3.1.2) and the Markov property (equation 3.1.3) leads to

$$\begin{aligned} &P(x_{N+1}, t_{N+1};\ x_N, t_N;\ \dots\ x_0, t_0) \\ &\quad = P(x_{N+1}, t_{N+1} \mid x_N, t_N) P(x_N, t_N \mid x_{N-1}, t_{N-1}) \\ &\qquad \cdot\ \dots\ P(x_1, t_1 \mid x_0, t_0) \cdot P(x_0, t_0) \end{aligned} \tag{3.1.4}$$

[1] A general definition of stochastic processes does not require ordering in time, but the assumption is more intuitive in a physical sense.

This relation highlights, that the probability for the realization of a given final state in a Markov process depends only on the probability of a given initial state, and a chain of transition events between states with probabilities depending only on the states in question.

■ 3.1.2 The Chapman-Kolmogorov equation

It can be gleaned from equation 3.1.4, that joint probability distributions detailing a Markov chain can be decomposed into an initial state and a series of transitions between states, thus making no more than two probability distributions necessary for the description of a Markov process. Decomposing the third order joint probability distribution as outlined above gives

$$P(x'', t''; \; x', t'; \; x, t) = P(x'', t'' \mid x', t')P(x', t' \mid x, t)P(x, t) \; . \tag{3.1.5}$$

Integrating the left hand side of equation 3.1.5 over the intermediate state x' gives

$$\begin{aligned} \int P(x'', t''; \; x', t'; \; x, t) \, \mathrm{d}x' &= P(x'', t''; \; x, t) \\ &= P(x'', t'' \mid x, t)P(x, t) \; . \end{aligned} \tag{3.1.6}$$

This operation is a marginalization of the variable x' and corresponds to combining all possible chains from x to x'', regardless of the intermediate state x'. Hence, the variable is no longer of importance and was therefore removed from the distribution. Integration of the remaining right hand side of equation 3.1.5 and dividing both results by $P(x, t)$ leads to the Chapman-Kolmogorov equation:

$$P(x'', t'' \mid x, t) = \int P(x'', t'' \mid x', t')P(x', t' \mid x, t) \, \mathrm{d}x' \tag{3.1.7}$$

The Chapman-Komolgorov equation is an identity relation that has to be fulfilled by the transition probabilities defining a Markov process. The intuition behind the equation is identical to the reasoning given above: the probability of a transition from x to x'' is equal to the sum of probabilities of all possible paths that have an intermediate state x'.

The central reason for which the Chapman-Komolgorov equation has been given here is that, given appropriate assumptions, differential equations for the time evolution of probability distributions can be derived from it, namely the Fokker-Planck equation and the master equation.

■ 3.1.3 Derivation of the master equation

Following (Mahnke et al., 2009), the Chapman-Komolgorov equation can be transformed into a differential form by substituting the ansatz

$$P(x', t + \tau \mid x, t) = (1 - a(x', t)\tau)\delta(x' - x) + \tau W(x', x, t) \tag{3.1.8}$$

for the transition probability. Equation 3.1.8 is an approximation of the time dependence of the transition probability, stating that the probability for a transition from x to x' to take place within the small interval τ increases linearly with time at a rate $W(x', x, t)$.

The first term denotes the probability of no transition taking place within the interval, i.e., $P(x', t' \mid x, t) = \delta(x' - x)$ with δ being the Dirac delta function. The rate $a(x', t)$ can be related to the transition rate via the normalization condition

$$\begin{aligned} 1 &= \int P(x', t+\tau \mid x, t)\,\mathrm{d}x' \\ &= \underbrace{\int (1 - a(x', t)\tau)\delta(x' - x)\,\mathrm{d}x'}_{1-a(x,t)\tau} + \tau \int W(x', x, t)\,\mathrm{d}x' \,, \end{aligned} \tag{3.1.9}$$

leading to

$$a(x, t) = \int W(x', x, t)\,\mathrm{d}x' \,, \tag{3.1.10}$$

which is the intuitive statement that the the probability to remain at x decreases at the same rate with which the probability to transition from x to any other state x' increases.

Rewriting the ansatz (equation 3.1.8) using the normalization condition with appropriate indices

$$P(x'', t' + \tau \mid x', t') = \delta(x'' - x') \left(1 - \tau \int W(\tilde{x}, x', t')\,\mathrm{d}\tilde{x}\right) + \tau W(x'', x', t') \tag{3.1.11}$$

and substituting into the Chapman-Kolmogorov equation (equation 3.1.7) gives

$$\begin{aligned} P(x'', t' + \tau \mid x, t) = P(x'', t' \mid x, t) &+ \tau \int W(x'', x', t') P(x', t' \mid x, t)\,\mathrm{d}x' \\ &- \tau \int \delta(x'' - x') \int W(\tilde{x}, x', t')\,\mathrm{d}\tilde{x} P(x', t' \mid x, t)\,\mathrm{d}x' \,. \end{aligned} \tag{3.1.12}$$

Rearranging terms and dividing by τ reproduces a difference quotient on the left hand side. After the outer integration in the last term is resolved using the delta function, the integration variable of the remaining integral is renamed to highlight the symmetry of the right hand side terms:

$$\begin{aligned} \frac{P(x'', t' + \tau \mid x, t) - P(x'', t' \mid x, t)}{\tau} = \\ \int W(x'', x', t') P(x', t' \mid x, t)\,\mathrm{d}x' \\ - \int W(x', x'', t') P(x'', t' \mid x, t)\,\mathrm{d}x' \end{aligned} \tag{3.1.13}$$

In the limit of infinitesimally short time intervals, i.e., $\tau \to 0$, the equation can be written in a more compact form. Furthermore suppressing unnecessary indices, the result stands as

$$\frac{\partial P(x, t)}{\partial t} = \int \left[W(x, x', t) P(x', t) - W(x', x, t) P(x, t)\right]\,\mathrm{d}x' \,. \tag{3.1.14}$$

Equation 3.1.14 is the integral form of the master equation for a continuous state space. Replacing the integral with a sum transforms the master equation into a form appropriate for the description of a countable set of states, the details of which will be discussed in the following section.

3.2 The master equation

The master equation is a set of first order differential equations that describe the time evolution of occupation probability of all microstates in a system. It can be interpreted as a balance taken over gains and losses in occupation probability due to transitions from or to other microstates in the system. The master equation can be given in integral form for a continuous system (as derived in section 3.1.3) but will be addressed here in the case of a system with a countable number of discrete microstates.

Most commonly, the discrete master equation is given as

$$\frac{\partial P(x,t)}{\partial t} = \sum_{x'} \left[W_{xx'} P(x',t) - W_{x'x} P(x,t) \right] \tag{3.2.1}$$

where x and x' denote microstates that are occupied with probabilities $P(x,t)$ and $P(x',t)$, respectively. The first term of the sum describes the gain in occupation probability of x due to transitions from x' with rate $W_{xx'} \geq 0$ while the second term quantifies losses due to transitions from x to x' with rate $W_{x'x} \geq 0$.

Unless required by a specific system described by equation 3.2.1, it is not necessary to further restrict the interpretation of x. However, it may help intuition to think of $x = x_0 \dots x_N$ as an enumeration of microstates that can be occupied by a single particle in a multistate system, or as an array of occupation numbers in a multiparticle system.

3.3 The chemical master equation

The chemical master equation (McQuarrie, 1967; Gillespie, 1992) is a particularly useful form of the master equation in the case that the set of transitions in a given system is limited and can be expressed in a way similar to chemical reactions. The chemical master equation gives the time evolution of the probability distribution $P(\vec{N},t)$, where $\vec{N}$ is the vector of (integer) population sizes for all reactant species relevant to a set of reactions. Accordingly, the state space supporting the probability distribution spans all configurations of population numbers and thus tends to be very large for realistic system sizes. This makes the implementation of numerical schemes problematic, as the master equation consists of one differential equation for each individual state. Nevertheless, the chemical master equation provides all necessary information on the kinetics of a chemical system, including not only the concentrations of reactants $c_k = N_k / \sum_k N_k$, but also the population numbers N_k, the dynamics of the total system size $\sum_k N_k$ and the statistics thereof, as described in section 3.2. Notably, the influence of the dynamic system size on the statistics of the system is given by the chemical master equation.

Although the appropriate state space contains all configurations of occupation numbers, transitions between arbitrary states are only possible in a chemical system if they are specified as being part of a set of $r = 1 \dots M$ 'reaction channels'. Each chemical reaction modifies the populations of reactants according to a specific pattern given by the stochiometry of the reaction. This change can be denoted as the stochiometric vector

$$\vec{\nu}_r = \vec{N}(t + \tau_r) - \vec{N}(t) \ , \tag{3.3.1}$$

defined as the difference of population vectors before and after a reaction indexed by r has taken place in an interval τ_r. Interpreting the k'th element of $\vec{\nu}_r$ as a matrix element ν_{kr} that describes how the population of a reactant k is modified in a reaction r, leads to the associated stochiometry matrix, which is of pivotal interest to the field of chemical reaction networks (see e.g. Angeli (2009)).

The chemical master equation stands as

$$\frac{\partial P(\vec{N},t)}{\partial t} = \sum_{r=1}^{M} \left[a_r(\vec{N}-\vec{\nu}_r)P(\vec{N}-\vec{\nu}_r,t) - a_r(\vec{N})P(\vec{N},t) \right] , \tag{3.3.2}$$

where, in accordance with the master equation 3.2.1, the first and second terms denote the gains and, respectively, losses in occupation probability due to the M reaction channels, each individual reaction indexed by r. The function $a_r(\cdot)$ is referred to as the propensity of a chemical reaction (Gillespie, 2007) and is related to the probability of a chemical reaction to occur in a given time interval. A more detailed description will be given in section 3.3.2.

■ 3.3.1 Elementary reactions

In many situations, the jump processes to be described by a master equation are, or can be expressed similar to chemical reactions involving a number of molecular constituents. These reactants are spontaneously transformed or interact stochastically according to a set of uni- or multimolecular processes. Although the reactants will be, for simplicity, referred to as molecules in the scope of this section, it needs to be stressed that the described model is not limited to chemical reactions. Similarly, the account given below is not intended as a comprehensive treatment of the types of chemical reactions, but rather serves to highlight similarities between the types of processes that can be interpreted as reaction schemes, and to build intuition. Therefore, the display will be limited to reactions that are elementary, i.e., they may not involve more than a single step to be completed. More precisely, the reactions of interest are intended to be completed within an (infintesimally) short time interval and shall not possess any stable intermediate states. This is not a general requirement for the chemical master equation and eliminates at first glance most reactions in which stochiometric coefficients greater than one occur. However, as will be briefly discussed below, these reactions can be further broken down into elementary reactions.

Indeed, the discussion in the majority of cases can be limited to the actions of one molecule or the interactions between two molecules by following the ensuing argument: as detailed in section 3.3.2, the probability for a reaction to occur in a given short time interval is limited by the probability of all required constituents to meet, which is bound to decrease substantially if more than two distinct reactant molecules are required. As a consequence, this type of interaction, albeit possible in theory (Gershinowitz and Eyring, 1935), is not expected to have a significant impact due to the low frequency at which it may occur. In chemical reality, however, various reactions are formulated as multimolecular, a paradigm being catalytic reactions where the interaction between two molecules is possible only in the presence of a third reactant. These reactions can usually be decomposed into a sequence of elementary bimolecular interactions, such as the formation

of an intermediate complex of two reactants in an initial reaction, and the subsequent reaction of the complex with a third compound. A prominent example for which this assumption has proven to be well-justified is the kinetic action of enzymes (Henri, 1902; Michaelis and Menten, 1913).

In accordance with the reasoning above, there exist four elementary reactions:

(ia)	$\varnothing \to A$				
(ib)	$A \to \varnothing$	(ii)	$A \rightleftharpoons B$	(iiia)	$A + \varnothing \to B + C$
(iiib)	$A + B \to C + \varnothing$	(iv)	$A + B \rightleftharpoons C + D$		

Reactions (i)-(iv) have been sorted by the number of educts. As will be discussed in section 3.3.2, the rates at which reactions occur depends on the availability of their constituents. Briefly, (ia) and (ib) can be interpreted as the spontaneous creation or destruction of one reactant, equivalent to an elementary birth-death process (Gardiner, 1985), or equivalently, as fluxes into and from an open system. (ia) is bound to occur at a constant rate whereas the rate at which A is removed from the system via (ib) depends on the availability of the remaining A. (ii) corresponds to a spontaneous conversion between two reactants. A possible realization in chemical terms would be an isomerization. Similar to (i), in (iiia) and (iiib) the number of reactants is not conserved, natural interpretations being the formation or decay of chemical complexes, but other interpretations are equally possible: for instance, (iiia) could describe the reproduction of a microorganism if $A = B = C$. (iv) is a rather general reaction scheme with various interpretations within and outside chemistry.

Although generally indispensable for the formulation of chemical reactions, stochiometric coefficients have been omitted, and are required for elementary reactions only in special cases that can be safely neglected here. Indeed, a forward reaction with stochiometric coefficients unequal to unity, e.g., $A + 2B \to X$, cannot take place in a single step as three constituents are required to meet, and thus needs to be further decomposed. This does not apply for a spontaneous decay of the type $A \to B + 2C$. A decomposition of processes of this type is possible in theory, but there is no probabilistic argument prevent a chemical reaction of this type to occur. However, as the processes of interest here are not chemical reactions, there is no reason for a separate treatment.

■ 3.3.2 The propensity function

The final missing ingredient to the formulation of a chemical master equation (equation 3.3.2) from a set of reactions is the propensity function $a_r(\cdot)$. The propensity function is defined such that $a_r(\vec{N})\,\mathrm{d}t$ is the probability for one reaction r to occur in a time interval $\mathrm{d}t$, provided the system is in the state described by $\vec{N}$ (Gillespie, 2007).

Propensity functions as given below are specific in structure for each type of reaction and require the system to be in a perfectly mixed condition. This assumption implies that interactions should not be influenced by the local configuration of molecules. Instead, molecules are assumed to be homogeneously distributed within the system and

interactions depend solely on their availability, i.e., on the current state in terms of population sizes given by $\vec{N}$.

Following (Gillespie, 2007) and limiting the presentation to the uni- and bimolecular cases as detailed in section 3.3.1, the propensity functions read

$$a_r(\vec{N}) = c_r N_i \tag{3.3.3}$$

for a unimolecular reaction with one molecule of the population N_i as an educt and products specific to the reaction channel r,

$$a_r(\vec{N}) = \frac{c_r}{\Omega} N_i N_j \tag{3.3.4}$$

for a bimolecular interaction involving two molecules that originate from different populations N_i and N_j, and finally

$$a_r(\vec{N}) = \frac{c_r}{\Omega} N_i (N_i - 1) \approx \frac{c_r}{\Omega} N_i^2 \tag{3.3.5}$$

in the case of a bimolecular interaction between two molecules of the same species N_i. In all three cases, c_r is a rate constant specific to the reaction r. If more than one constituent is involved in a reaction, the probability for particles to interact depends on the probability of encountering each other, which is inversely proportional to the system size Ω. The approximation given in equation 3.3.5 holds if the population N_i is sufficiently large, i.e., $(N_i + 1)/N_i \approx 1$.

In the case that the interacting particles are indeed molecules, pairwise interactions are understood as collisions between molecules and required to be mass action kinetics, i.e., to be describable by the law of mass action (Voit et al., 2015). Details of the reactant species that have an influence on the reaction probability, such as a potential energy barrier that has to be overcome through the collision, or a dependence on the pairwise orientation, are summarized by the rate constant. The determination of these constants from the physical details of reactants is subject to collision theory (Atkins et al., 2018) or experimental investigation and will not be further described here.

■ 3.3.3 Deterministic simplification: the rate equation

In the thermodynamic limit, i.e., in a large system containing a large number of particles, stochastic fluctuations in global quantities vanish. As this applies to the population sizes N_k, it is expected that trajectories obeying the probability distribution $P(\vec{N}, t)$ described by the (chemical) master equation (equations 3.2.1 and 3.3.2) converge to a deterministic evolution equation of the average when $\sum N_k \to \infty$.

A formalism for the construction of such deterministic rate equations can be borrowed from chemistry. In its general form, a kinetic chemical reaction can be written as (Chen et al., 2010)

$$\sum_k e_{kr} X_k \xrightarrow{c_r} \sum_k p_{kr} X_k \ , \tag{3.3.6}$$

where X_k represents one species of reactants with population size N_k, and $e_{kr},\ p_{kr}$ denote the stochiometric coefficients of educts and products, specific for the reaction r. Based on e_{kr} and p_{kr}, the stochiometric matrix $\nu_{kr} = p_{kr} - e_{kr}$ and the stochimoetric vector $\vec{\nu}_r$ (compare equation 3.3.1) can be defined.

The deterministic rate equations describing the evolution of the population sizes N_k is of the form

$$\frac{\partial N_k}{\partial t} = \sum_{r=1}^{M} \nu_{kr} f_r(\vec{N}) \ . \tag{3.3.7}$$

The functions $f_r(\cdot)$ are reaction rates that follow from the law of mass action (Voit et al., 2015):

$$f_r(\vec{N}) = \hat{c}_r \prod_k N_k^{e_{kr}} \tag{3.3.8}$$

When limited to the uni- and bimolecular cases detailed in section 3.3.2, and with appropriate choices for the rate constants $\hat{c}_r$, the reaction rates are identical to the propensity functions 3.3.3-3.3.5. Indeed, it is possible to show this equivalence by deriving from the chemical master equation a Langevin equation for the population size and noticing that the (additive) noise term vanishes in the thermodynamic limit (see Gillespie (2007)).

■ 3.4 The stochastic simulation algorithm

In many cases it is not feasible to solve the master equation by means of analytical methods. If this is the case, one is required to resort either to approximations of the master equation as outlined in the previous section, to numerical methods, or to simulations. In the present work, the former methods are not disregarded because they are thought to lack accuracy, but rather to exploit a core feature of stochastic simulations. In the basic formulation of the stochastic simulation algorithm, as will be presented in this section, every transition or reaction is carried out explicitly to modify a set of populations or occupation numbers. This makes every trajectory simulated using the stochastic simulation algorithm an exact solution of the associated master equation, and it implies that it is possible to track secondary effects of each transition other than the modification of occupation numbers. The view on networks promoted in this work is that of a stochastic process, where modifications of network structure are identical to conversion of chemicals over the course of a reaction. Simulating these explicitly using the stochastic simulation algorithm and keeping track of all structural changes tied to a reaction facilitates not only to determine how populations of similar nodes evolve, but also how the structure evolves as a whole. This information is not easily accessible if other methods are used to solve the master equation, and it is in many cases not straightforward to define an associated process that can alleviate this problem.

For example, the formation of a giant connected component within a network structure is of salient interest to this work. In principle, it would be possible to model on one hand the evolution of node populations using an an approximation or numerical solution

of the master equation and on the other hand interpret the formation of a giant component as a cluster aggregation process (Krapivsky et al., 2009). However, neither does any of these methods provide a simpler approach, nor is it trivial to relate the two processes.

Although its fundamentals have been known much longer, the stochastic simulation algorithm is generally attributed to Gillespie (1977) who rigorously based in the probability theory for chemical kinetics, which has also led to the formulation of propensity functions in section 3.3.2. The basic concept behind the stochastic simulation algorithm is straightforward: although general statements on the time evolution of a stochastic system as discussed in the previous sections can only be made in terms of probabilities, each realization of such a time evolution (a stochastic trajectory) is uniquely defined by an initial condition, a sequence of transitions and the times at which the transitions occur. In terms of probability theory, a stochastic process that has the Markov property (see section 3.1.1) is determined by the joined probability density function $P(r, \tau \,|\, \vec{N}, t)$, which states the probability density that a system in state $\vec{N}$ at time t changes state by a process indicated by r after a time interval τ. Thus, knowing the statistics of r and τ as a function of the state $\vec{N}$ would enable creating exact trajectories by sequentially drawing sets of random numbers r, τ from the joined probability density and updating time and state accordingly. Following Gillespie (1977), this is facilitated via the ansatz

$$P(r, \tau \,|\, \vec{N}, t)\, \mathrm{d}\tau = P_0(\tau \,|\, \vec{N}, t) a_r(\vec{N})\, \mathrm{d}\tau \ , \tag{3.4.1}$$

where $P(r, \tau \,|\, \vec{N}, t)\, \mathrm{d}\tau$, i.e., the probability that the process r occurs at time $t+\tau$, is interpreted as a conditional probability consisting of the two parts $P_0(\tau \,|\, \vec{N}, t)$ and $a_r(\vec{N})\, \mathrm{d}\tau$. On one hand, $a_r(\vec{N})\, \mathrm{d}\tau$ equals the definition of the propensity function $a_r(\vec{N})$ (compare section 3.3.2), stating the probability that the process r occurs within the interval $[t + \tau, t + \tau + \mathrm{d}\tau)$. $P_0(\tau \,|\, \vec{N}, t)$ on the other hand is defined as the probability that no process occurs within $[t, t + \tau)$. If $a_r(\vec{N})\, \mathrm{d}\tau$ is the probability that a process r occurs in $[\tau, \tau + \mathrm{d}\tau)$, then by normalization $1 - \sum_r a_r(\vec{N})\, \mathrm{d}\tau$ is the probability that no reaction occurs in the same interval. The conditional probability that no processes occur in $[t, t + \tau)$ and $[\tau, \tau + \mathrm{d}\tau)$ is

$$P_0(\tau + \mathrm{d}\tau \,|\, \vec{N}, t) = P_0(\tau \,|\, \vec{N}, t) \left(1 - \sum_r a_r(\vec{N})\, \mathrm{d}\tau \right) \tag{3.4.2}$$

with $P_0(\cdot)$ defined as above. In the limit $\mathrm{d}\tau \to 0$, this is a differential equation with the solution

$$P_0(\tau \,|\, \vec{N}, t) = e^{-a_0(\vec{N})\tau} \ , \tag{3.4.3}$$

where $a_0(\vec{N}) = \sum_r a_r(\vec{N})$. Combining the result with equation 3.4.1 leads to the probability distribution for r and τ, i.e.,

$$P(r, \tau \,|\, \vec{N}, t) = a_r(\vec{N}) e^{-a_0(\vec{N})\tau} \ , \tag{3.4.4}$$

indicating that the time step size τ is distributed exponentially.

Having obtained the distribution 3.4.4, in theory everything is in place to outline how trajectories of a stochastic process can be generated. Prior to describing the algorithm in

greater detail, however, it is necessary to examine how random values for r and τ may be generated from equation 3.4.4. Most software platforms on which stochastic simulations might be implemented provide random number generators that generate uniformly distributed pseudo-random numbers in some interval. In every time step of the stochastic simulation algorithm, two such random numbers, i.e., $\{n_1,\ n_2 \,|\, n_i \in [0,1)\}$, are generated. In the wake of the emergence of Monte-Carlo methods, various procedures have been developed to create maps between uniform random numbers and specific probability distributions, such as equation 3.4.4. The procedure that lends itself to application in the stochastic simulation algorithm is inversion transform sampling (Lemieux, 2009), a method that dates back to von Neumann.

By inversion transform sampling, random numbers can be sampled from any probability distribution provided that its cumulative distribution function is available and can be inverted by some method. The method is briefly stated as follows: given uniform random numbers $u \in [0,1)$, values of a stochastic variable x distributed according to the probability distribution P are obtained from

$$x = F^{-1}(u) \ , \tag{3.4.5}$$

where $F^{-1}(\cdot)$ is the inverse of the cumulative distribution $F(x)$ of $P(x)$. Naturally, the method is especially convenient if $F^{-1}(\cdot)$ can be calculated analytically and does not have to be reevaluated for each time step. This is the case for the exponential distribution. In abbreviated notation, the cumulative distribution of equation 3.4.4 after marginalization of the variable r is given as

$$F(\tau) = 1 - e^{-a_0 \tau} \ . \tag{3.4.6}$$

Solving for τ and substituting the random number n_1 for $F(\tau)$ leads to

$$\tau = \frac{1}{a_0} \ln\left(\frac{1}{1-n_1}\right) \ , \tag{3.4.7}$$

where it is readily realized that $1 - n_1$ is just another uniform random number on $[0,1)$, and thus τ can be given as

$$\tau = \frac{1}{a_0} \ln\left(\frac{1}{n_1}\right) \ . \tag{3.4.8}$$

In the case of r, the situation is somewhat more complicated when using inverse transform sampling as r is a discrete random variable rather than a continuous one. Marginalization of $\tau \in [0,\infty)$ shows that r is distributed according to

$$P(r) = \frac{a_r}{a_0} \ , \tag{3.4.9}$$

meaning that r follows from a weighted random selection. Formally, r is given as the smallest integer satisfying

$$\frac{1}{a_0} \sum_{i=1}^{r} a_i > n_2 \ , \tag{3.4.10}$$

Algorithm 3.4.1: Stochastic simulation algorithm

Data: set of reactions R, rate constants $\{c_r \,|\, r \in R\}$, stochiometric vectors $\{\vec{\nu}_r \,|\, r \in R\}$, initial populations $\vec{N}_0$, start/end time $(t_0,\ t_N)$
Result: time steps t, populations over time $\vec{N}(t)$
begin
  $\vec{N} \longleftarrow \vec{N}_0$
  $t \longleftarrow t_0$
  while $t \leq t_N$ **do**
    $\{a_r \,|\, r \in R\} \longleftarrow$ calculate propensities $(\vec{N},\ \{c_r\})$
    $\{n_1,\ n_2 \,|\, n_i \in [0,1)\} \longleftarrow$ random numbers ()
    $\tau \longleftarrow$ calculate step size $(n_1,\ \{a_r\})$
    $r \longleftarrow$ select reaction channel $(n_2,\ \{a_r\})$

    $\vec{N} \longleftarrow \vec{N} + \vec{\nu}_r$
    $t \longleftarrow t + \tau$
    store t, $\vec{N}$

where $\sum_{i=1}^{r} a_i/a_0$ is the cumulative distribution function of the distribution $P(r)$ and n_2 is a uniform random variable on $[0, 1)$.

The stochastic simulation algorithm is specified in pseudo-code as algorithm 3.4.1. Summarizing, the steps required for the simulation of exact stochastic trajectories are the specification of initial conditions, given by an initial time t_0 and populations $\vec{N}_0$, and by specification of the set of reactions to be simulated, given by the rate constants c_r and the stochiometric changes associated with each reaction, given by $\vec{\nu}_r$. After initialization, time steps are repeatedly executed until an exit condition is reached, as is usually given by a maximal time t_N. Time steps are structured as follows. First, propensities $a_r(\vec{N})$ and $a_0(\vec{N}) = \sum_r a_r(\vec{N})$ are determined based on the current state of the system. Second, random numbers are generated and used to determine step size and the type of the next reaction in accordance with equations 3.4.8 and 3.4.10. Finally, time and populations are updated accordingly and stored, and then the procedure is repeated for the next time step.

Part II

Results

Chapter 4

Analytical models for percolation in small-degree networks

The study of network models in the scope of statistical physics has attracted significant amounts of attention in the recent years. However, the majority of works focus on large or highly-connected systems, and in particular on those evolving towards selected classes of degree distributions. In response to early successes, evolution rules known to lead to precisely those degree distributions are integrated into a wide range of network models as ad hoc assumptions. Still, the considerable interest in these systems has proven to be fruitful, as the concepts emerging in the process paved the way to understanding a large number of real-world networks.

This development has to some degree diverted attention from networks that occupy the far end of the spectrum: small networks where only a small number of node degrees is present. Although they are in some cases covered as limiting cases of more general descriptions, these networks are usually not the focus of the discussion, and simplifications associated with their specific properties are not applied. Nevertheless, small degree networks have numerous relevant real world applications, including those networks that are embedded in space (Barthélemy, 2010) and in particular those, where the number of edges that connected to a node are restricted by their physical dimensions.

Here, the small-degree property is exploited to derive a number of exactly solvable analytical models for the percolation probability and, equivalently, size of the largest component of an appropriate network. In detail, the governing equations of the random graph and configuration models – polynomials of an order corresponding to the largest degree in a graph – are truncated at a small degree for which they can be treated by analytical methods.

The random graph model (Erdös and Rényi, 1959) is generally considered as one of the most basic models for network generation, and it is known that the underlying rules lead to a Poisson-type degree distribution. Still, for networks with a degree distribution that does not deviate too far from the Poisson case, this description is found to produce solid results, as published by Fessel et al. (2012, 2015*a*,*b*) for describing percolation in the small degree network of *P. polycephalum*. If no a priori knowledge on the specific class

of degree distribution of a network is available, the configuration model of graph theory (Molloy and Reed, 1995; Newman et al., 2001) provides an appropriate framework for formulating the conditions under which a giant component exists in the network. Both models are briefly reviewed in sections 2.2.2, section 2.3.1 and section 2.3.2.

Briefly, this chapter is organized into three parts. First, in section 4.1, the random graph model introduced in section 2.2.2 is further addressed. In addition to giving closed form solutions, the role of the average degree as a driving parameter is discussed. Similarly, solutions of the configuration model introduced in section 2.2.2 are derived and presented in section 4.2. Solutions of both models make use of considerations on the percolation on graphs described in section 2.3. The chapter is concluded by a brief comparative analysis of the analytical solutions derived here, which is presented in section 4.3. In particular, focus is laid on the location of the phase transition as a function of the maximal degree in the network.

■ 4.1 Small degree solutions in the random graph model

In the random graph model, the size of the giant component S can be calculated from equation 2.3.2, i.e.,

$$S = 1 - u \ , \tag{4.1.1}$$

where u represents the probability for a randomly selected node not to be part of the giant component. Generally, u is a solution of the self-consistency relation (equation 2.3.3),

$$u = \sum_{k=0}^{n_k-1} p_k u^k \ , \tag{4.1.2}$$

which, on the right hand side, is a polynomial with one term for each node degree. The largest degree in the system is given by $n_k - 1$, where n_k is the number of distinct node degrees. In the most general case, $n_k = \infty$. Here, the observation that spatially embedded transportation systems are quite often limited to a small degree (or feature higher degree nodes with vanishing probability) is used to select for n_k a small number for which equation 2.3.3 is analytically solvable. Brief consideration of equation 2.3.3 suggests analytical solutions should be possible for $n_k \leq 5$, as there exist explicit expressions for the solutions of polynomial equations up to a degree of four. Indeed, this is the case if searching for a solution that covers all possible values of u. However, in the regime where connectivity is local, in the thermodynamic limit no giant component exists in the graph and thus one finds $u = 1$. By making use of this knowledge, and hence searching for an expression for u only in the percolated regime where $0 \leq u < 1$ makes it possible to reduce the degree of the polynomial by one. Thus, solutions are possible to a maximal degree of $n_k - 1 = 5$. To do so, the normalization condition for the degree distribution

$$1 = \sum_{k=0}^{n_k-1} p_k \tag{4.1.3}$$

is multiplied by u and subtracted from the self-consistency relation (equation 2.3.3), leading to

$$0 = \sum_{k=0}^{n_k-1} p_k(u^k - u) \ , \tag{4.1.4}$$

which is conveniently rewritten as

$$0 = p_0(1-u) - u \sum_{k=1}^{n_k-2} p_{k+1}(1-u^k) \ . \tag{4.1.5}$$

It is apparent that each individual term in the sum is zero if $u = 1$. Furthermore, the term $k = 0$ has been omitted from the sum, as it is zero independent of the the value assumed by u. Searching now as outlined for a solution in the case that $0 \leq u < 1$ and making use of the identity

$$1 - u^k = (1-u)\sum_{i=0}^{k-1} u^i \tag{4.1.6}$$

one finds, after some algebra

$$0 = p_0 - \sum_{k=2}^{n_k-1} p_k \sum_{i=1}^{k-1} u^i = p_0 - \sum_{i=1}^{n_k-2} u^i \sum_{k=i+1}^{n_k-1} p_k \ . \tag{4.1.7}$$

It is immediately clear from the limits of the sums that the maximum polynomial degree is $n_k - 2$. The two expressions given in 4.1.7 hold identical information. However, it has been deemed sensible to state both, as it can be helpful to work with polynomials that are either sorted for coefficients or powers, depending on the application.

■ 4.1.1 Condition for the phase transition & geometric significance of the driving parameter

The implication that $0 \leq u < 1$ is only searched for in the peroclated regime makes it possible to derive the condition for the phase transition from equation 4.1.7 by taking the limit $u \to 1$. It needs to be emphasized, that this is only possible after having applied the aforementioned assumption, which can be easily shown by taking the same limit in the unmodified self-consistency relation, i.e., equation 2.3.3. There, in the limit $u \to 1$, the normalization condition for the p_k is reproduced. Now, taking this limit in the modified self-consistency relation leads to

$$0 = \sum_{k=0}^{n_k-1} (k-1)p_k \ , \tag{4.1.8}$$

which only holds at the phase transition, as shown below. It is well known that the condition for the phase transition in the Erdös-Renyi random graph model can be alternatively stated as $< k > -1 = 0$, i.e., if the average degree is equal to unity. Indeed,

writing out $< k > -1$ by substituting $< k >= \sum_k k p_k$ and $1 = \sum_k p_k$ leads to an identical expression as given in equation 4.1.8:

$$< k > -1 = \sum_{k=0}^{n_k-1} (k-1) p_k \tag{4.1.9}$$

Thus, equation 4.1.8 is indeed a valid expression for the phase transition.

Now, whereas equation 4.1.8 only holds at the phase transition, there is no mathematical reason why equation 4.1.9 should not be computable on the entire domain. The term $< k > -1$ is zero only at the phase boundary and can be used as measure which indicates the signed distance of the system from the phase transition. Below, solutions of the self-consistency relation will be given in terms of ratios of node degree probabilities p_k/p_j, where $j = n_k - 1$, as

$$0 = n_k - 2 + \sum_{k=0}^{n_k-2} (k-1) \frac{p_k}{p_{n_k-1}} \ . \tag{4.1.10}$$

In each case, the topological state is found to depend on $n_k - 2$ such ratios. Thus, $n_k - 2$ can be interpreted as the dimension of the topological phase space spanned by the ratios p_k/p_{n_k-1}, in which the phase boundary is a $n_k - 3$ dimensional hyper-plane that is described by a normal form given by equation 4.1.10. Similarly, rearranging equation 4.1.9 also leads to the normal form of a hyper-plane with the same dimension, which is identical to the plane describing the phase boundary on the right hand side

$$\frac{1}{p_{n_k-1}} (< k > -1) = n_k - 2 + \sum_{k=0}^{n_k-2} (k-1) \frac{p_k}{p_{n_k-1}} \ . \tag{4.1.11}$$

the term on the left hand side is related to the distance from the origin of the embedding space. Thus, as the hyper-planes are identical except for their distance from the origin, they are bound to be parallel, and their perpendicular distance is proportional to $(< k > -1) / p_{n_k-1}$. This realization is important, as it shows that a graph topology that is characterized by some particular $< k >$ will always have the same distance from the phase transition, regardless of the detailed configuration p_k that gives rise to the average degree.

■ 4.1.2 Closed form solutions

With equations 4.1.7 and 4.2.6 at hand, everything is in place to formulate analytical solutions for u (and S). For a more compact notation, the quotient

$$q_{ij} = \frac{p_i}{p_j} \tag{4.1.12}$$

is introduced. In all cases, nodes of maximal degree $j = n_k - 1$ have been selected as denominator.

In the scope of this work, closed form solutions in the random graph model have been obtained with maximal degrees $n_k - 1 = 3$ and $n_k - 1 = 4$. In both cases, solutions have

been constructed for the cases $p_2 = 0$ and $p_2 \geq 0$. The solutions omitting second degree nodes are the product of collaborations and have been published in (Fessel et al., 2012, 2015*a*,*b*) in conjunction with data obtained from network growth experiments performed utilizing either the slime mold *P. polycephalum*, see chapter 1.1, or human endothelial cells (Gamba et al., 2003; Serini, 2003). Results of these studies are discussed as part of chapter 6.

The $p_2 = 0$ cases are presented below as published. However, it needs to be emphasized here, that differentiating between the cases $u = 1$ and $0 \leq u < 1$ instead of searching for a solution for all values of u greatly simplifies the calculation as the degree of the polynomial consistency equation is reduced by one. This case differentiation corresponds to searching for u only in the percolated regime, as it is by normalization clear that $u = 1$ when no giant component exists. The giant component for all values of u then just reads

$$S = \begin{cases} 0 & u = 1 \\ 1 - u & 0 \leq u < 1 \end{cases} \tag{4.1.13}$$

and hold exactly the same information as the more complicated closed form solutions for $0 \leq u \leq 1$, as highlighted for the case $n_k = 4$, $p_2 = 0$ in figure 4.1.1. The simplified solutions are not stated here explicitly, but can be readily obtained from the $p_2 \geq 0$ cases by taking the limit $p_2 \to 0$. This realization also makes it possible to solve the random graph model with $n_k - 1 = 5$ analytically as a function of three node degree ratios.

Case $\mathbf{0 \leq u \leq 1}$, $\mathbf{p_2 = 0}$, $\mathbf{n_k = 4}$ In the present case, which has been published in part in (Fessel et al., 2012), the self-consistency relation 2.3.3 simplifies to a polynomial of third order that can be expressed in terms of a single variable $q_{03} = p_0/p_3$, i.e.,

$$0 = u^3 - (1 + q_{03})u + q_{03} \ , \tag{4.1.14}$$

whose physically significant solution is its smallest real root

$$u = -\sqrt{\frac{4}{3}(q_{03} + 1)} \cos\left(\frac{1}{3} \arccos\left(-\frac{1}{2} q_{03} \sqrt{\frac{27}{(q_{03} + 1)^3}}\right) + \frac{\pi}{3}\right) \ , \tag{4.1.15}$$

can be obtained by utilizing one of the existing methods for third order polynomials. In the present case, Cardano's method has been selected. The size of the giant component, given by $S = 1 - u$ is shown in figure 4.1.1. It has been stated above that it is significantly simpler to search for a solution only in the percolated regime. To illustrate this, the corresponding solution that is found when excluding $u = 1$ has been inscribed into the figure as well. For the sake of completeness, the algebraic expression reads

$$u = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4q_{03}} \tag{4.1.16}$$

Indicated in the figure as a dashed line is the position of the percolation transition, which in terms of the ratio p_0/p_3 is located at

$$\frac{p_0}{p_3} = 2 \ . \tag{4.1.17}$$

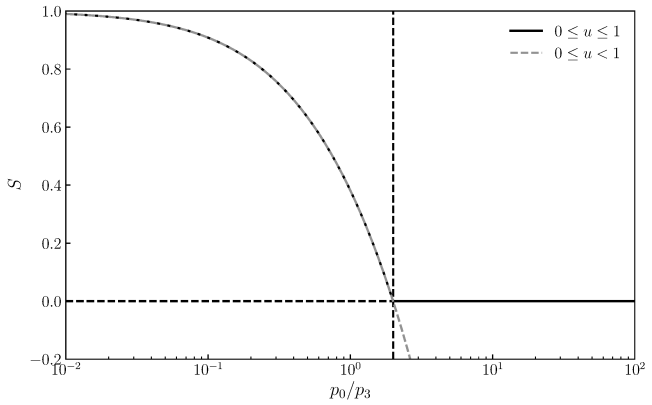


Figure 4.1.1: Giant component size in the random graph model for $n_k = 4$, $p_2 = 0$. A giant component spontaneously emerges when the ratio p_0/p_3 takes a value smaller than 2 (indicated by the dashed line. At the critical point, all node degrees are fixed to the following values: $(p_0, p_1, p_3) = (1/3, 1/2, 1/6)$, corresponding to the average degree $< k >= 1$.

The position of the percolation transition is obtained from the condition for the phase transition as stated in equation 4.1.9. At the transition the system is strongly constrained in a topological sense. In the limit $u \to 1$, the degree probabilities have to satisfy normalization, the condition for the phase transition as stated above or in equation 4.1.9, and also the Molloy-Reed criterion, equation 2.3.17. Although this equation is derived from the configuration model, it has to hold in the random graph model as well, as the latter represents a special case thereof. In accordance with the number of aforementioned conditions, the number of degrees of freedom at the transition is thus given as $n_k - 3$ in the random graph model, or as $n_k - 4$ when imposing the constraint that $p_2 = 0$. In the present case, there are exactly zero degrees of freedom and all degree probabilities are fixed to numerical values as

$$(p_0, p_1, p_3) = \left(\frac{1}{3}, \frac{1}{2}, \frac{1}{6}\right) \ . \tag{4.1.18}$$

Expressing the average node degree near the transition in terms of the ratio p_0/p_3 indicates that the system passes through the critical point $< k_c >$ with a slope of 1/6:

$$< k >= \sum_{k}^{3} p_k k = 1 - \frac{1}{6}\left(\frac{p_0}{p_3} - 2\right) \tag{4.1.19}$$

Series expansion of the order parameter S at $p_0/p_3 = 2$ leads to

$$S = -\frac{1}{3}\left(\frac{p_0}{p_3} - 2\right) = \frac{1}{2}(< k > -1) \ , \tag{4.1.20}$$

indicating that S scales with a critical exponent $\beta = 1$ near the transition as is expected from a solution of the random graph model (Newman, 2003).

Case $0 \leq u \leq 1$, $p_2 = 0$, $n_k = 5$ Choosing not to neglect nodes of degree four from the model comes at the cost of having to solve an equation of fourth degree, which is compactly stated as

$$0 = u^4 + q_{34}u^3 - (1 + q_{34} + q_{04})u + q_{04} \tag{4.1.21}$$

employing the node degree fraction $q_{04} = p_0/p_4$ and $q_{34} = p_3/p_4$. As for polynomials of third degree, there exist explicit solution formulas for equations of the present type. The solution presented below has been obtained by employing Euler's method for quartic equations (Euler and Hewlett, 2009; Tignol, 2016) and has been published in (Fessel et al., 2015*a*,*b*). Stating only the physically significant smallest real root of equation 4.1.21, the solution reads

$$u = \sqrt{z_1} - \sqrt{z_2} - \sqrt{z_3} - \frac{1}{4}q_{34} \ . \tag{4.1.22}$$

Methods for fourth order polynomials construct solutions of the quartic equation from the solutions z_{1-3} of the resolvent cubic, a third order polynomial associated with the quartic equation (Tignol, 2016). These solutions are given as

$$\begin{cases} z_1 = \left(\frac{q_{34}}{4}\right)^2 + \sqrt{-\frac{4}{3}a}\cos\left(\frac{1}{3}\arccos\left(-\frac{b}{2}\sqrt{-\frac{27}{a^3}}\right)\right) \\ z_2 = \left(\frac{q_{34}}{4}\right)^2 - \sqrt{-\frac{4}{3}a}\cos\left(\frac{1}{3}\arccos\left(-\frac{b}{2}\sqrt{-\frac{27}{a^3}}\right) + \frac{\pi}{3}\right) \\ z_3 = -\frac{c}{8z_1z_2} \end{cases} \tag{4.1.23}$$

with the substitutions

$$\begin{cases} a = -\frac{1}{4}q_{04} - \frac{1}{16}q_{34}\left(1 + q_{34} + q_{04}\right) \\ b = -\frac{1}{64}\left(1 + q_{04}\right)\left(1 + 2q_{34} + q_{04} + q_{34}^2\right) \\ c = -\left(1 + q_{34} + q_{04}\right) + \frac{1}{8}q_{34}^3 \end{cases} \ . \tag{4.1.24}$$

In the present case, the size of the giant component is expressed as a function of two

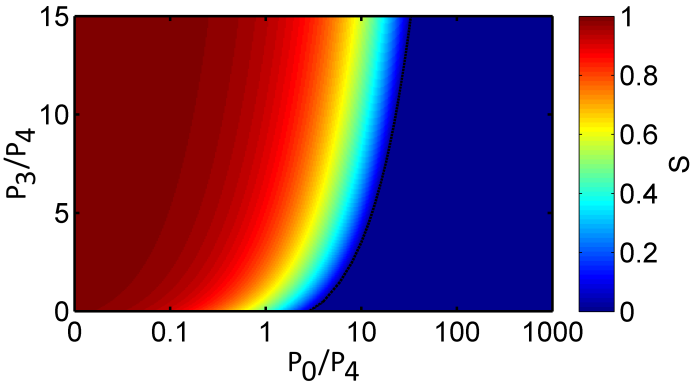


Figure 4.1.2: Giant component size in the random graph model for $n_k = 5$, $p_2 = 0$.

ratios as detailed above, and thus can be displayed in a two-dimensional phase diagram

as shown in figure 4.1.2. Due to the additional degree of freedom introduced by the nodes of fourth degree, the phase transition is line rather than a point. Examining the model in the limit $u \to 1$ as has been done in the $n_k = 4$ case leads to a set of relations between the node degree probabilities p_k:

$$(p_0, p_1, p_3) = \left(\frac{1-3p_4}{3}, \frac{1+4p_4}{2}, \frac{1-12p_4}{6} \right) \tag{4.1.25}$$

It is immediately clear that if $p_4 \to 0$, the node degrees take the same fixed values as for the previous case. Furthermore, from these relations (or from equation 4.1.9) it can be inferred that the line marking the phase transition in figure 4.1.2 is given as

$$\frac{p_3}{p_4} = \frac{1}{2} \left(\frac{p_0}{p_4} - 3 \right) \ . \tag{4.1.26}$$

Cases $\mathbf{0 \leq u < 1}$, $\mathbf{p_2 \neq 0}$, $\mathbf{n_k = 4\,\&\,5}$ It is not always feasible to work with a topology without nodes of second degree. Solutions for this case are conveniently derived from equation 4.1.7. Analytical solutions in this case are similarly possible up to a maximal degree of $n_k - 1 = 5$. However, in comparison to the case $p_2 = 0$, an additional free parameter characterizes the phase transition. The solution for $n_k = 4$ reads

$$u = -\frac{1}{2} \left(1 + q_{23} - \sqrt{q_{23}^2 + 4q_{03} + 2q_{23} + 1} \right) \ . \tag{4.1.27}$$

Inspection of the solution shows that in the limit $p_2 \to 0$ it is identical to the $p_2 = 0$ solution presented in equation 4.1.16. This behavior is further highlighted by the condition for the phase transition which is obtained in the limit $u \to 1$, i.e.,

$$\frac{p_2}{p_3} = \frac{p_0}{p_3} - 2 \ , \tag{4.1.28}$$

indicating that if $p_2 = 0$, the transition is one-dimensional and occurs at $p_0/p_3 = 2$ as discussed above. For $p_2 > 0$, the phase transition is a line in the p_0/p_3, p_2/p_3 plane as displayed in the phase diagram shown in figure 4.1.3. Similarly, the degree fractions are constrained to a single degree of freedom at the transition. Relating these fractions to p_2 gives

$$(p_0, p_1, p_3) = \left(\frac{1+p_2}{3}, \frac{1-2p_2}{2}, \frac{1-2p_2}{6} \right) \ . \tag{4.1.29}$$

For $p_2 = 0$, the constant fractions in the one-dimensional case are reproduced. The present example is exceptionally well-suited to highlight the significance of the average node degree $< k >$ as a driving parameter through a geometric interpretation, as developed in a general fashion in the previous section. Writing the average node degree in terms of p_0/p_3, p_2/p_3 (and employing the normalization condition for the p_k),

$$< k > = \sum_k k p_k = 1 - p_3 \left(\frac{p_0}{p_3} - \frac{p_2}{p_3} - 2 \right) \ , \tag{4.1.30}$$

and rearranging for p_2/p_3 gives

$$\frac{p_2}{p_3} = \frac{p_0}{p_3} - 2 + \frac{1}{p_3} \left(< k > -1 \right) \ , \tag{4.1.31}$$

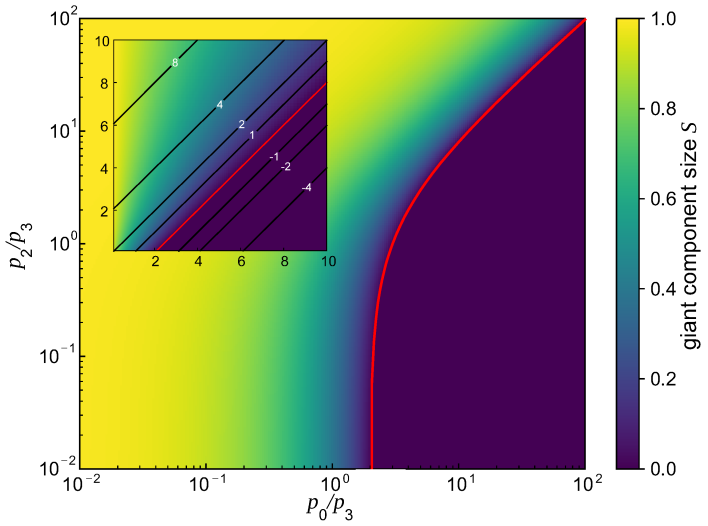


Figure 4.1.3: Giant component size in the random graph model for $n_k = 4$, $p_2 \neq 0$. The model can be conveniently displayed as a two-dimensional phase diagram. The phase transition is shown as a red line. At, the phase transition, $< k >= 1$ is required. The driving parameter $< k >$ is shown for selected values as black lines in the inset. Notably, the driving parameter can be interpreted as a measure for the distance from the phase transition, regardless of the detailed configuration of the p_k. The one-dimensional phase transition is reproduced for $p_2/p_3 \to 0$.

which is a line in the phase plane as well, and it is parallel to the line indicating the phase transition. Indeed, for $< k >= 1$, the condition for the phase transition is reproduced. By this geometric interpretation, it is intuitive to think of $< k >$ as a measure for distance from the transition, regardless of the actual p_k configuration at the transition if a graph is similar to the Erdös-Renyi random graph.

Introducing again nodes of fourth degree into the description requires solution of a cubic function, which is found in the form of

$$u = \frac{1}{6}\left[12\sqrt{3\Delta} - 4\Delta_1\right]^{1/3} - \frac{2}{3}\frac{\Delta_0}{\left[12\sqrt{3\Delta} - 4\Delta_1\right]^{1/3}} - \frac{1}{3}(q_{34}+1) \tag{4.1.32}$$

with

$$\begin{cases} \Delta = 18bcd - 4b^3d + b^2c^2 - 4c^3 - 27^2d^2 \\ \Delta_0 = 3c - b^2 \\ \Delta_1 = 2b^3 - 9bc + 27^2d \end{cases} \tag{4.1.33}$$

and

$$\begin{cases} b = 1 + q_{34} \\ c = 1 + q_{24} + q_{34} \\ d = -q_{04} \end{cases} . \tag{4.1.34}$$

As expected, the solution can no longer be visualized in two dimensions as it depends on three ratios. The phase transition is given as a plane described by

$$\frac{p_2}{p_4} = \frac{p_0}{p_4} - 2\frac{p_3}{p_4} - 3 \; . \tag{4.1.35}$$

As above, relations between the degree probabilities can be established. These are sensibly stated as

$$(p_0, p_1, p_3) = \left(\frac{1 + p_2 - 3p_4}{3}, \frac{1 - 2p_2 + 4p_4}{2}, \frac{1 - 2p_2 - 12p_4}{6} \right) \tag{4.1.36}$$

where it can be readily assessed that taking the limits $p_4 \to 0$ or $p_2 \to 0$ reproduces the relations associated with the previous solutions.

■ 4.2 Small degree solutions in the configuration model

It has been discussed in section 2.3.2 that computation of the size of the giant component in the configuration model is less straightforward since the self-consistency relation cannot be given in terms of the probability that a randomly selected node is part of the giant component. Instead, using the generating function formalism by Newman et al. (2001) it has been shown in section 2.3.3 that the size of the largest component follows from

$$S = 1 - G_0(u) \quad \text{where} \quad G_0(u) = \sum_{k=0}^{n_k - 1} p_k u^k \tag{4.2.1}$$

(equations 2.3.14 and 2.3.10). u, the probability to arrive in a non-giant component via a randomly selected edge is a solution of the self-consistency relation

$$u = G_1(u) \quad \text{where} \quad G_1(u) = \frac{\sum_{k=0}^{n_k - 1} (k+1) p_{k+1} u^k}{\sum_{k=0}^{n_k - 1} k p_k} \; , \tag{4.2.2}$$

which has been given in equation 2.3.5. A more useful and transparent form can be found by carrying out the same steps as in the previous section. First, the self-consistency equation is rewritten:

$$u \sum_{k=0}^{n_k - 1} k p_k = \sum_{k=0}^{n_k - 1} (k+1) p_{k+1} u^k \tag{4.2.3}$$

It is apparent that p_0 and p_2 do not enter the equation. Furthermore, the largest polynomial degree $u^{n_k - 1}$ is associated with the degree probability p_{n_k}, which exceeds the maximum degree and is thus assumed to be negligible. Extracting terms associated with the smallest node degree p_1, eliminating the largest degree as noted above and adjusting the summation index on the right hand sight leads to

$$u p_1 + u \sum_{k=3}^{n_k - 1} k p_k = p_1 + \sum_{k=3}^{n_k - 1} k p_k u^{k-1} \; . \tag{4.2.4}$$

subtracting the left hand side gives

$$0 = p_1(u-1) - u\sum_{k=3}^{n_k-1} kp_k(1-u^{k-2}) \ . \tag{4.2.5}$$

Implying again that a solution is sufficient in the case that $0 \leq u < 1$, identity 4.1.6 is applied and after some rearranging one obtains:

$$0 = p_1 - \sum_{k=3}^{n_k-1} kp_k \sum_{i=1}^{k-2} u^i = p_1 - \sum_{i=1}^{n_k-3} u^i \sum_{k=i+2}^{n_k-1} kp_k \ . \tag{4.2.6}$$

In the previous section, it has been demonstrated that it is possible to obtain an expression for the condition for the phase transition from the modified self-consistency relation by taking the limit $u \to 1$. Indeed, the same is true for the configuration model. Taking the limit $u \to 1$ in equation 4.2.6 leads to

$$0 = \sum_{k=0}^{n_k-1} k(k-2)p_k \ , \tag{4.2.7}$$

which is precisely the result obtained by Molloy and Reed (1995) (equation 2.3.17). It is emphasized again, that this procedure is only possible after limiting to the percolated region, i.e., by assuming $0 \leq u < 1$. As solutions to the configuration model are stated below in terms of degree fractions $q_{ij} = p_1/p_j$, $j = n_k - 1$, this can be usefully rewritten as

$$0 = (n_k-1)(n_k-3) + \sum_{k=0}^{n_k-2} k(k-2)\frac{p_k}{p_{n_k-1}} \ . \tag{4.2.8}$$

From the previous condition it can be gleaned that neither nodes of degree zero nor two have an influence on the position of the phase transition, indicating that the configuration model is only accurate if the relative number of these nodes is small in vicinity of the transition. This shortcoming of the model, amongst others that apply to specific degree distributions that do not qualify as 'well behaved', has already been noted by Molloy and Reed (1995) and has since then been further addressed, e.g., by Joos et al. (2017) who propose a refined model .

■ 4.2.1 Closed form solutions

Equations 4.1.7 and 4.2.6 are very similar in structure and can be easily compared. Generally, the order of the polynomial is smaller by one in the configuration model, and the degrees probabilities p_0 and p_2 naturally do not occur in the description. However, it is necessary to keep in mind that the probabilities u are defined in a different manner in the two models. Calculating the size of the giant component from a solution of equation 4.2.6 via equation 2.3.14 factors in all probabilities p_k, $k \leq n_k - 1$ except one, which may be eliminated via the normalization condition that has not yet been employed in the configuration model. Thus, on one hand solutions in the configuration model can be computed up to a larger maximal degree, but are on the other hand less accessible for investigation in the framework of degree probabilities.

Case $0 \leq \mathbf{u} < 1$, $\mathbf{p_2} \geq 0$, $\mathbf{n_k} = 4$ The solution for u in the configuration model derived from equation 4.2.6 is very simple in structure, as the largest polynomial degree in the configuration model is $n_k - 3$. The linear solution, dependent on the single variable $q_{13} = p_1/p_3$ reads

$$u = \frac{1}{3} q_{13} \ . \tag{4.2.9}$$

Due to the simplicity of the solution, it lends itself to demonstrating how the giant component size in the configuration model S is obtained when u is known explicitly. Following equation 2.3.14, the size of the giant component is given as

$$S = 1 - \sum_{k=0}^{n_k - 1} p_k u^k \ , \tag{4.2.10}$$

indicating that the computation of S is in any straightforward once u is available. In the present case, this leads to

$$S = 1 - p_0 - \frac{1}{3}\frac{p_1^2}{p_3} - \frac{1}{9}\frac{p_2 p_1^2}{p_3^2} - \frac{1}{27}\frac{p_1^3}{p_3^2} \ , \tag{4.2.11}$$

a form, which is impractical to work with unless all relevant p_k are known. With regard to this work, solutions as given above are still useful. In the subsequent chapter, models will be devised that describe the evolution of all node degrees in a system as a function of time either stochastically, or in a deterministic fashion. Taken together with the theory worked up here, these models pose a comprehensive theoretical model for the time evolution of degree-related measures and expected component sizes in random graphs with arbitrary, truncated degree distributions.

Similar to the solutions found in the random graph model, the condition for the phase transition as stated generally in equations 2.3.17 or 4.2.8, which can be adopted for each individual solution. In the present case, one finds $p_1/p_3 = 3$, indicating that although the ratio p_1/p_3 is strongly constrained, the only constraint in place for the remaining degrees p_0, p_2 is set by normalization.

In the random graph model, the degree fractions at the transition are constrained by three (or four, if $p_2 = 0$) equations. In the configuration model, there is one more degree of freedom. The Molloy-Reed criterion, which could be adopted from the configuration model as an additional constraint for the random graph model, is identical to the condition for the phase transition here, and thus provides no new information. Therefore, the number of degrees of freedom at the transition is equal to $n_k - 2$ or $n_k - 3$ if $p_2 = 0$. This increased flexibility of the model at the transition has to be attributed to applicability of the configuration model to arbitrary degree distributions. Stating relations between the node degree probabilities is nevertheless possible, albeit less informative, and thus no explicit expressions are given here.

Cases $0 \leq \mathbf{u} < 1$, $\mathbf{p_2} \neq 0$, $\mathbf{n_k} = 5\,\&, 6$ Analytical solutions to the configuration model are possible up to $n_k = 7$. However, in agreement with the types of networks analyzed

within the scope of this work and with regard to the larger number of degrees of freedom at the transition, only the solutions up to $n_k = 6$ are stated. Rather than discussing the features of these solutions in greater detail here, they are compared to the $n_k = 4$ and random graph cases in the following section.

The solution for $n_k = 5$ depends on the two ratios p_1/p_4, p_3/p_4 and reads

$$u = -\frac{1}{8}\left(4 + 3q_{34} - \sqrt{9q_{34}^2 + 16q_{14} + 24q_{34} + 16}\right) \tag{4.2.12}$$

with the phase transition line given by

$$\frac{p_3}{p_4} = \frac{1}{3}\left(\frac{p_1}{p_4} - 8\right) . \tag{4.2.13}$$

Similar to the third-order solutions given above, the solution for $n_k = 6$ is given as

$$u = \frac{1}{6}\left[12\sqrt{3\Delta} - 4\Delta_1\right]^{1/3} - \frac{2}{3}\frac{\Delta_0}{\left[12\sqrt{3\Delta} - 4\Delta_1\right]^{1/3}} - \frac{1}{3}(\frac{4}{5}q_{45} + 1) . \tag{4.2.14}$$

For a compact formulation, the substitutions

$$\begin{cases} \Delta = 18bcd - 4b^3d + b^2c^2 - 4c^3 - 27^2d^2 \\ \Delta_0 = 3c - b^2 \\ \Delta_1 = 2b^3 - 9bc + 27^2d \end{cases} \tag{4.2.15}$$

have been made, where

$$\begin{cases} b = 1 + \frac{4}{5}q_{45} \\ c = 1 + \frac{3}{5}q_{35} + \frac{4}{5}q_{45} \\ d = -\frac{1}{5}q_{15} \end{cases} . \tag{4.2.16}$$

The constraints at the phase transition can be summarized by the condition for the phase transition, i.e., by

$$\frac{p_4}{p_5} = \frac{1}{8}\left(\frac{p_1}{p_5} - 3\frac{p_3}{p_5} - 15\right) . \tag{4.2.17}$$

Again, the remaining degrees are constrained only by normalization.

■ 4.3 Comparison of closed form solutions

This section compares the small degree solutions in the random graph and configuration models. The discussion focuses on the position of the phase transition, which is the most influential difference between the models. In figure 4.3.1, all solutions devised in the previous sections are shown as functions of the average degree. In addition to the giant component S, which functions as the order parameter of the system, the average size $< s >$ of non-giant components is displayed, which . For comparison and in order to

better assess the influence of the cutoff parameter n_k, a numerical solution of the random graph model with no bound is shown as well.

It has been shown above for the random graph model, that the average degree is a measure for the distance of a system from the phase transition, regardless of the exact configuration of node degrees that is giving rise to the average degree. Although this highlights that the average degree is well-suited as a driving parameter, no algebraic relation is available between the driving parameter and the size of the giant component. For the plots shown in figure 4.3.1, the rate equation model introduced in 7 has been utilized to produce a set of $p_k(t)$ mimicking the construction process of an Erdös-Renyi random graph (see chapter 2) with a maximal degree $n_k - 1$ according to the respective solution S as selected from the previous sections. This case is similar to the bounded degree graphs studied by Ben-Naim and Krapivsky (2011).

Focusing first on the solutions of the random graph model, it is observed that the prior to the transition at $< k >= 1$ there is no visible influence of the cutoff parameter. In this regime, all the size of non-giant components grows identical to the unbounded case. After the transition, there are subtle differences that can be traced back to the maximal degree. In the unbounded case, the size of the giant component approaches $S = 1$ asymptotically, whereas in the bounded cases the the giant component approaches unity as the average degree approaches the maximal degree. If plotted as a function of time (not shown), $S = 1$ and the maximal degree are approached asymptotically in all cases. The larger the maximal degree, the closer solutions match with the unbounded case.

In the configuration model, the behavior with which the maximal degree is approached is identical to the random graph model. However, it is observed that the position of the phase transition appears to be shifted (see also figure 4.3.2). In all cases, the transition is shifted to larger values and the shift is more pronounced if the cutoff parameter is smaller. This effect is discussed in greater detail below. At the transition, the critical exponents associated with S and $< s >$ remain in agreement with the values expected for percolation in random graphs, i.e., $\beta = \gamma = 1$.

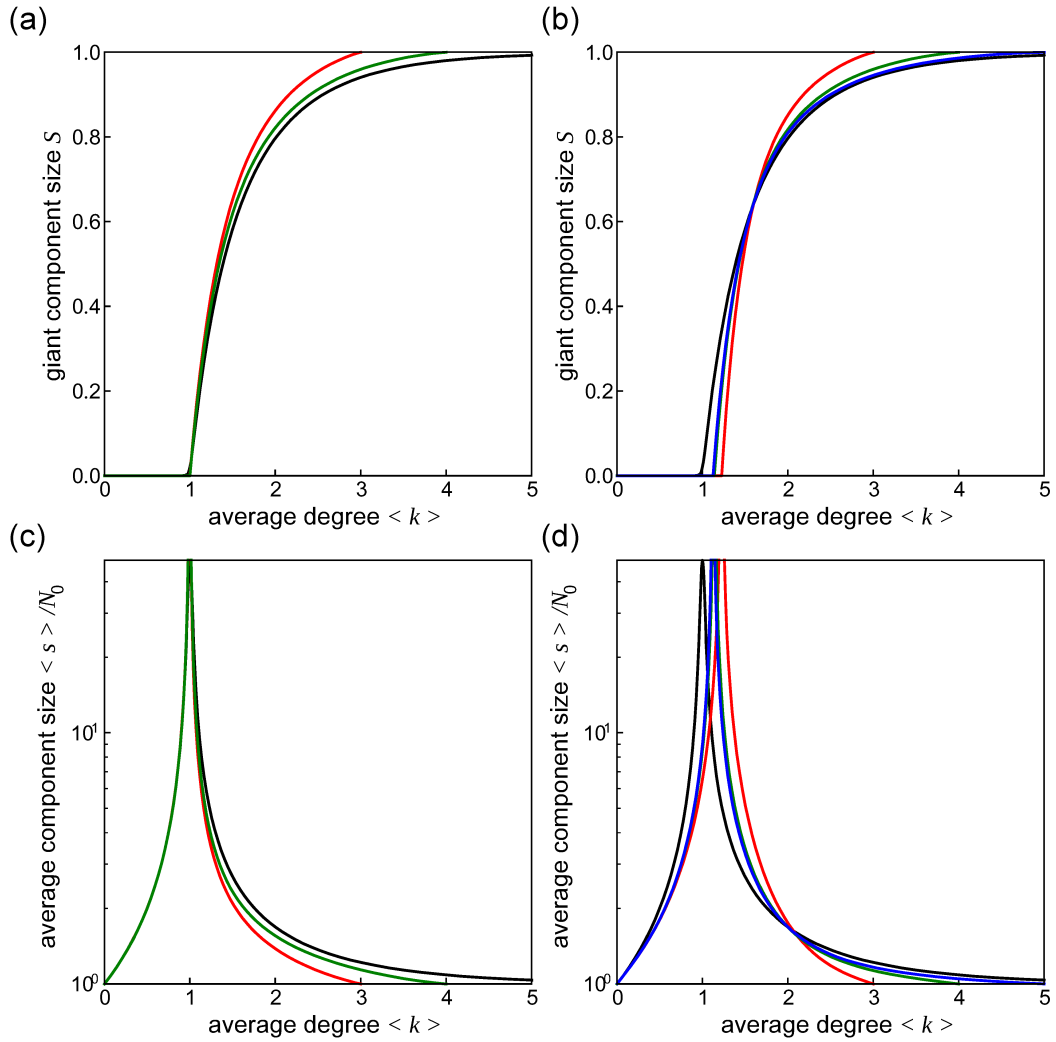


Figure 4.3.1: Overview of closed form solutions. The random graph model with no bound on the degree distribution is shown as a black line in all panels. (a and c) Giant component size and average non-giant component size in the random graph model, truncated either at $n_k = 4$ or 5. Regardless of the maximal degree, the critical point is at $< k_c >= 1$. (b and d) Giant component size and average non giant-component size in the in the configuration model bounded at $n_k = 4$, 5 or 6. Unlike the random graph model, configuration model exhibits a shifted percolation transition if bounded at a smaller degree.

■ 4.3.1 Position of the phase transition

The solutions derived in this chapter describe the probability for the existence of a giant component either accurately or by approximation. This behavior is elucidated in the following. First is necessary to illustrate, how the cutoff parameter characterizing the

solutions has to be understood. To do so, it is helpful to consider the case of bounded degree graphs as investigated by Ben-Naim and Krapivsky (2011), and briefly discussed in chapter 7. Although similar at first glance, there is a crucial difference between the cutoff parameter used to create bounded degree, and the cutoff parameter employed in this chapter.

In bounded degree graphs, nodes cannot evolve to a higher degree after having obtained the maximal number of links. Unless links are lost by some process, nodes of maximal degree will remain at that state. If no such process exists in the system, this state functions as an absorbing state in the evolution of each single node. The evolution of a bounded degree graph where the rate at which connections are established exceeds the rate at which they are removed, will eventually tend to a state that is dominated by nodes that have the maximal degree.

In contrast, the solutions presented in this chapter are not models describing the evolution of graphs, but rather present a way to assess from a suitable degree distribution whether the associated graph has a giant component. The cutoff parameter in this case represents the approximation that the fraction of nodes larger than the bond given by the parameter will be negligible. Rather than representing exclusively situations where nodes accumulate at the maximal degree similar to the case of bounded degree graphs, the small degree solutions in the random graph and configuration models cut off all nodes that have a larger degree.

Now, this is not a problem if the system to which the model is applied possess no nodes that exceed the cutoff parameter, as is the case, for instance, in a bounded degree graph with the same cutoff parameter. It is expected that in those cases, the solutions presented here accurately describe the existence of a giant component. From this point of view, it is also clear in which cases the solutions are only an approximation, and whether the quality of the approximation is sufficient as will be discussed momentarily.

The Molloy-Reed criterion can be visualized as a balance between node classes that drive a system into, or further away from the percolated state. Nodes with a degree larger than two drive the system into the percolated regime, whereas nodes of degree one keep the system in an un-percolated state. Nodes of degree zero and two assume a neutral role. Furthermore, the larger their degree, the more a node contributes to the connectivity of the network and, by extension, to driving the system into the percolated regime. It is clear, that if a too large fraction of large degree nodes is cut off in this fashion, the connectivity of the graph will be greatly underestimated.

To understand how this influences the position of the critical point, it is useful to consider instead of a static degree distribution the evolution of a random graph as defined by Erdös and Renyi in the presence of a cutoff parameter, compared to the same model with no cutoff parameter as a null model. It is known, that at any point in time, the degree distribution of such a graph is given by a Poisson distribution centered at the average degree. The distribution is shifted to larger degrees as the average degree increases. If not bounded, the distribution maintains a broadened but otherwise constant

shape, which is centered at the position at which the Molloy-Reed criterion holds. It is clear, that this process will be well approximated by the solutions presented here as long as the center of the distribution is far from the cutoff parameter, or if no such cutoff is in place.

If a bound is brought into place, nodes in the large-degree tail of the distribution that would drift above the bound are either kept fixed at the maximal degree, or are cut off entirely. The nodes influenced by the bound contribute less to the Molloy-Reed balance, and they do so in a non-linear fashion. Due to this non-linearity, the position indicated by the Molloy-Reed criterion no longer coincides with the center of the distribution. Similarly, the center of the distribution no longer coincides with the average degree, but in this case, the influence of large degree nodes is linear. As the distribution is lacking these large degree nodes, it needs to be shifted to a larger average degree compared to the Erdös-Renyi model, until the Molloy-Reed criterion can be satisfied.

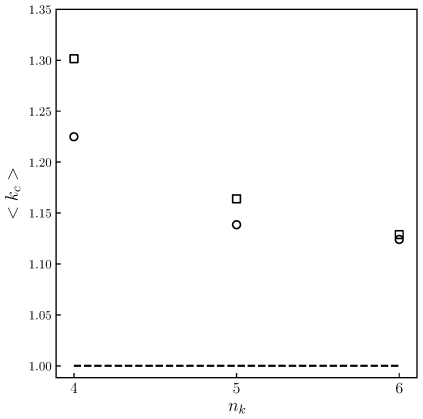


Figure 4.3.2: Critical points of bounded and unbounded random graphs. Bounding the evolution of a random graph at some degree $n_k - 1$ leads to a shift of the transition. This situation (open circles) is expected to be exactly described by the analytical small-degree solutions of the configuration model. In contrast, approximating the evolution of a random graph that is not bounded (open squares) with the same solutions and neglecting large degree nodes in the process, leads to a systematic miscalculation of the critical point that exceeds the shift in the bounded case.

Figure 4.3.2 shows the transition point (determined from the location of the maximum of $< s >$) versus the cutoff parameter in both cases discussed above. If the underlying process is the construction of an unbounded random graph, the degree distribution is simply cut off at the maximal degree and thus the critical point is strongly shifted to larger degrees. If the underlying process is the construction of a bounded random graph, nodes that would have exceeded the maximal degree accumulate there instead. Again, the transition is shifted to the mechanism outlined above, although the shift is less significant.

Chapter 5

A graph model for planar *P. polycephalum* networks

The ability of the slime mold *Physarum polycephalum* to create intricate networks on a centimeter scale has incited many exciting projects, as has been reviewed in part in chapter 1. Although the objectives of these works targeting network structure are diverse, they share at least one common bottom line. By some method, information contained in the structure of the mold, which is usually documented as an image or a series of images, needs to be extracted and has to be brought into a format in which it is suitable for further analysis. This problem classifies in part as an image processing and segmentation task. However, as will be specified below, obtaining graph information on a network that is embedded in an image requires a suitable graph model. In the present chapter, a graph model will be presented which has been designed in the scope of this work and is intended as a way to discretize network information in the plane, given the edges of the network have a measurable thickness.

The focus of this work is on the process in which a *P. polycephalum* network forms by growth and fusion of scattered microplasmodia, slime mold fragments with sizes ranging on the order of $100-500\,\mu m$. Typically, this process is observed for $24-36\,h$. In the process, $2880-4320$ gray scale images are taken that require processing, making an efficient image evaluation and network extraction routine a crucial part of this work. This will be examined more closely in the following section. However, efficient data evaluation is not the only contributing factor to the need for a suitable graph model. Besides data analysis, modeling and simulation of network formation are addressed in chapter 7. Network growth and formation in *P. polycephalum* is a process that combines continuous changes, such as the extension of growth fronts, with discrete topological changes, such as the fusion or fission of links. Due to the presence of discrete events, it has been found useful to regard both portions of the network formation process, continuous modifications as well as events that are discrete by nature, as sequences of events that are discrete in time and space. This approximation lays the foundation for the stochastic modeling and simulation of network formation presented in chapter 7, where only a minimal number of elementary processes gives rise to a flexible model for network formation.

Moreover, in future work a spatial discretization of the *P. polycephalum* network can

be regarded as a framework for the agent-based modeling of dynamic processes that occur on the network topology, and potentially interact with it. There exist numerous works that propose continuous differential equation models in order describe, for instance, contraction waves or the chemical oscillation dynamics observed in *P. polycephalum* networks or fragments (see, e.g., Smith and Saldana (1992); Radszuweit et al. (2014), or Teplov (2017) for a review). Such models are already complicated when considered on regular domains with well-defined boundary conditions, and implementing these on a complex, arbitrary shapes such as a realistic *P. polycephalum* topology, for instance via the phase field method (see, e.g., Löber et al. (2015)), holds numerous challenges to both, the modeling and to the computational implementation. Given a suitable spatial discretization of a network, it may become possible to solve approximate models, where differential equations are solved only on single nodes that interact only with their neighbors.

The chapter is organized as follows. Section 5.1 is targeted at outlining the needs and necessity of a graph model as developed in this chapter for analyzing the network formed by *P. polycephalum*. This is done with reference to the topological skeleton, which has been employed for the same task in earlier works (Fessel et al., 2012, 2015*b*), and the shortcomings thereof are discussed. Building on the requirements outlined in section 5.1, section 5.2 describes and critically discusses the model and method employed for network approximation in this work in greater detail. Finally, section 5.3 reports on a collaborative project (Hillmich, 2018), which has been initiated in order to investigate incremental changes in time resolved *P. polycephalum* graph data. This has led to a methodology that enables tracking of the elementary topological processes as considered in chapter 7 during the evolution of a *P. polycephalum* network.

■ 5.1 Requirements for a suitable graph model

Omitting at this point any details of the routine, the first step in extracting network information from an image in the scope of this work has been selected to be the conversion of a gray scale image into a binary image, where the only differentiation that is made is between the foreground and the background. This process is a paradigmatic example of image segmentation: when concerned only with the connectivity of a network, it is not important whether there is variation in the foreground. Having obtained suitable binary images by some method, the image processing step is completed and it is necessary to decide how graph information is to be extracted from the binary image. From the binary image, such data can be obtained by various methods that are all valid for their respective optimal use cases. In accordance with the objectives of this work, a novel method has been designed, whose final result constitutes a graph model that is specifically suited for the representation of *P. polycephalum* and comparable planar networks. The considerations making the presented method favorable the topological skeleton that has been used in earlier work (Fessel et al., 2012, 2015*b*) are elaborated in this section. The process in which graphs are abstracted from network images is detailed in section 5.2.

Among the tools available for extracting a graph-like structure from a binary image, the skeleton representation stands out, and has been successfully applied to image

data showing the evolution of *P. polycephalum* during early stages of this work, as is published in Fessel et al. (2012, 2015*b*). Briefly, the skeleton of a binary image can be created by progressively thinning binary structures until only lines remain that are one single pixel in width, or by computing the medial axis transform (Blum, 1967), which detects all points in a structure that are closest to at least two other points that lie on the edge of the structure. This process removes potentially unnecessary information such as the original width of the structure and preserves only the topology. A skeleton is easy to transform into a graph, as branching points and end points can be easily identified through a search on the skeleton. However, the skeleton has a number of shortcomings with regard to application in this work.

From a technical point of view, the skeleton proves problematic because it is extremely sensitive to imperfections that line the edge of the structure. Without extensive post-processing, these imperfections tend to lead to unwanted dangling ends, graph structures that do not reflect the shape of the object. Although sophisticated methods exist to prune such ends (Bai et al., 2007), it is generally difficult to create a procedure that works well for all cases. The networks under scrutiny in this work tend to have a significant number of open ends that are, unlike the dangling ends, important features of the network as they represent protrusions that are created as part of the motility apparatus of the slime mold. Distinction between these and dangling ends is an important requirement for a suitable graph model and presents a point where the skeleton was found to produce unsatisfactory results.

Another technical objection against the skeleton as a network representation arises when taking into account the temporal continuity of the graph. Identifying changes of a network topology that changes with time requires as is intended in the scope of this work requires that it is possible to associate partial structures of the graphs that represent adjacent time steps. Due to the sensitivity of the skeleton to tiny details of the binary image and the slight changes thereof, identifying such related structures is difficult with the topological skeleton.

Focusing on the properties of the *P. polycephalum* network, there exist further requirements to a suitable graph model that the skeleton fails to fulfill. First, it is observed, that the veins of a *P. polycephalum* network are very inhomogeneous in thickness due to the hierarchical structure of the network. Second, there exist spacious, sheet-like regions near the rim of a network. Inspecting these so-called growth fronts under the microscope reveals that they are internally highly structured into multiple channels, referred to as internal veins (Oettmeier et al., 2018). Due to limitations to the imaging hardware, this internal structure is not resolved during the imaging process underlying the date processed in this work. However, the graph model designed here is set up to account for the heterogeneity of the network structure by assigning to each node a radius of influence. How such a radius may be defined for the individual node will be detailed below. Suffice to say here, that the associated disk centered at the location of the node should neither overlap with the disks belonging to other nodes, nor with the background. This way, spacious growth fronts can be covered by large nodes, whereas thin veins are reflected only by nodes of small radius.

Assigning nodes a radius has multiple advantages in addition to the correct representation of growth fronts. Defining a minimal radius where nodes cannot fall below functions as a topological filter that eliminates the pruning problem detailed above, and similarly ensures that potentially erroneous details are not represented in the final graph. To illustrate this, one might consider the following example: a typical structure observed in planar networks is a the crossing point of two edges, which is in the present framework referred to as a node of degree four. However, if such a structure is abstracted by a topological skeleton, it is a common error that the node of fourth degree is separated into two adjacent nodes that are separated by a small distance distance.

Finally, the requirement, which is most influential for accurately representing the *P. polycephalum* network in the scope of this work is the clear definition of nodes of second degree. This requirement arises primarily from considerations targeting the structural dynamics of the network and thus needs to be explained with regard to the model devised in chapter 4. There, the construction of a graph is studied as a sequence of basal modifications, i.e., interactions between individual nodes. As implemented there, the set of edges exclusively describes the connectivity between nodes, rather than representing any other information such as physical dimension. In contrast, in a graph representing the branching and end points of a topological skeleton, edges measure the distance between these points. In order to represent the same information on a graph that assigns no further attributes to its edges, it is necessary to discretize this distance into a set of nodes of degree two.

Given such a discretization, nodes of degree two assume an important role in the system. First, the number of degree two nodes in a graph functions similar to a density. A graph with no degree two nodes connected by unit-length edges is topologically identical to a graph where a number of degree two nodes is distributed with equal probability. Although identical in a topological sense, the mesh formed by the latter graph will be more coarse-grained. A second important implication comes from the dynamics perspective. A typical process in the evolution of a *P. polycephalum* network is a scenario, where a node that is moved by some mechanism meets with a vein, leading to fusion and the subsequent emergence of a new branching point. In the light of the probabilistic approach detailed in section 3.3.2, the probability of such a process can be correctly represented in a setting that keeps edges at unit length only if the physical dimensions of the graph are accounted for by second degree nodes.

Summarizing the statements made above, a graph representation is searched for that discretizes a planar network structure into a set of nodes and edges, where edges do not represent the physical dimensions of the network but rather only indicate the connectivity of nodes. In turn, nodes possess individual radii that indicate their region of influence on the underlying network. In addition to nodes that represent junctions or tips of veins, nodes of degree two exist. These need to be defined such that the physical dimensions of the network are correctly represented if all nodes are non-overlapping disks. A minimal radius is required for correct pruning of redundant protrusion-like structures and other erroneous structures as detailed above. An example of such a graph is shown in figure

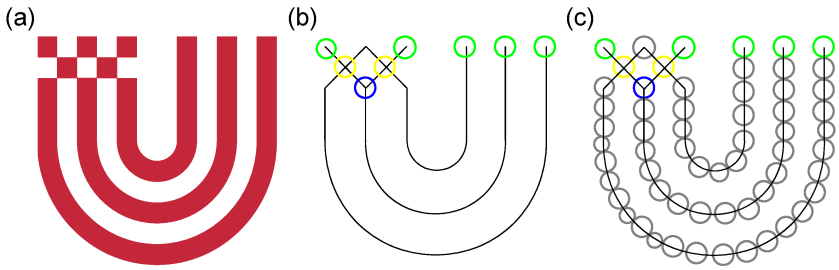


Figure 5.1.1: Nodes of second degree are an important feature for the exact representation of spatial networks. (a) A shape that is embedded in the 2-dimensional plane can be represented by several graph models. (b) If only end- and branching points of the shape are considered as nodes, it becomes necessary to treat edges of varying length. (c) Incorporating nodes of second degree into the model enables treating edges of unit length.

5.1.1 panel (c), which has been obtained from a binary u-shaped object as displayed in panel (a) using the algorithm proposed below. For comparison, the topological skeleton is shown in panel (b). Prominently, the length of edges is distributed evenly in the graph containing second degree nodes, whereas the distribution is very uneven in the topological skeleton. The coloring of nodes used in the figure indicates the degree of the node in question and is displayed in figure 5.1.2. The same colorcode will be used throughout the remainder of this work whenever it is deemed helpful.

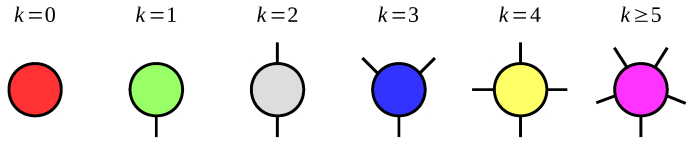


Figure 5.1.2: Color code for node degrees in small degree networks. Unless irrelevant or distracting, the present color code will be used throughout this work to indicate the degree of a node or observations associated with a given degree.

■ 5.2 The Apollonian graph approximation

Within the scope of this work, a method has been devised to approximate a graph satisfying the requirements stated above from a binary image. In essence, two steps require completion in order to obtain the graph representation. First, the foreground area in the binary image is approximated by placing a number of disks whose centers will later give rise to the nodes of the graph approximation. In the second step, the connectivity of the graph, i.e., the set of edges is established by comparing the result of the circle packing procedure to the skeleton of the binary image. Both procedures are described in greater detail below and are given in an abbreviated algorithmic form in algorithm 5.2.1. Furthermore, crucial steps in the procedure are displayed in figure 5.2.2. The algorithms

proposed here have been further improved and were optimized in a collaborative project (Hillmich, 2018). Unless indicated otherwise, all *P. polycephalum* graph data analyzed in this work have been obtained using the improved procedures. Details on the improvements are given below.

Named with reference to the Apollonian Gasket fractal (Mandelbrot, 1982), the Apollonian graph approximation presented here should not be confused with Apollonian networks, a type of planar graph that forms when all adjacent circles in the original Apollonian Gasket are connected by edges. Although Apollonian networks possess a number of interesting features, namely and among others, a scale-free degree distribution, and clustering and shortest-path characteristics that indicate presence of the small-world effect (Andrade et al., 2005), they are dissimilar from the type of network obtained by the mechanism proposed here due to the regularity of their construction.

Algorithm 5.2.1: Apollonian graph approximation

set of nodes
Data: binary image I, minimal radius r_0
Result: set of coordinates $\{(x, y)\}$, set of radii $\{r\}$
begin
  $r \longleftarrow \infty$
  $E \longleftarrow$ Euclidean distance transform (I)
  while $r \geq r_0$ **do**
    $r \longleftarrow$ max (E)
    $(x, y) \longleftarrow$ coordinates max (E)
    $I \longleftarrow$ remove disk from foreground $(I, (x, y), r)$
    $E \longleftarrow$ local Euclidean distance transform $(E, I, (x, y), r)$
    store r, (x, y)

set of edges
Data: binary image I, set of coordinates $\{(x, y)\}$, set of radii $\{r\}$
Result: graph G
begin
  $S \longleftarrow$ compute skeleton (I)
  $P \longleftarrow$ end & branching points (S)
  $G \longleftarrow$ initialize graph $(N \longleftarrow (\{(x, y)\}, \{r\}), E \longleftarrow \varnothing)$
  $G \longleftarrow$ clean non-overlapping nodes (S, G)
  $S \longleftarrow$ color skeleton by closest node (S, G)
  foreach P_i *in* P **do**
    $\{P_i'\} \longleftarrow$ neighboring points (S, P_i)
    foreach P_i' *in* $\{P_i'\}$ **do**
      $(x_j, y_j) \longleftarrow$ current coordinates (P_i')
      while $S(x_j, y_j) = N_i$ **do**
        $(x_j, y_j) \longleftarrow$ trace skeleton $(S, (x_j, y_j))$
      else
        $N_j \longleftarrow S(x_j, y_j)$
        $G \longleftarrow$ add edge (N_i, N_j)
  $G \longleftarrow$ clean self edges & duplicate edges (G)
  store G

■ 5.2.1 Obtaining the set of nodes

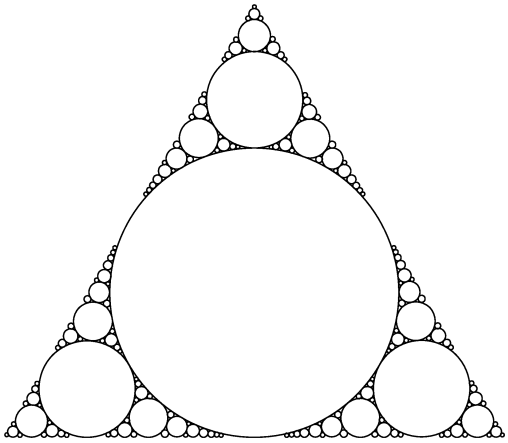


Figure 5.2.1: Apollonian Gasket in a triangular shape. The fractal can be constructed by recursively placing the largest possible circle into some shape, such that it does not overlap with any other circle. The present example has been generated using algorithm 5.2.1.

The graph approximation devised here is based on the concept that each point on the network is associated to the closest node. The circular area, in which points are associated to a node, defines that nodes radius of influence. Furthermore, these circular areas should neither overlap with the areas of other nodes, nor should they cover parts of the background that are not part of the network. In order to minimize the number of redundant nodes on the graph, the number of nodes needs to be minimized while the area covered is maximized. This defines a circle-packing problem on an arbitrary planar shape.

To solve this problem, a procedure based on the construction of the Apollonian Gasket fractal (Mandelbrot, 1982) is employed. The Apollonian Gasket is typically constructed from an initial configuration where a circular arena is inscribed with a smaller circle that shares a tangent with the former. From this initial configuration, the fractal is constructed by placing recursively the largest possible circle within the free space such that it does not overlap with any other circle that is already in place. Now, this construction procedure is not limited to the case outlined above, but can be applied to arbitrary shapes as is demonstrated in figure 5.2.1, where a triangular shape has been packed with circles. In a similar fashion, a binary shape representing a network can be packed with circles. The center points of all circles that have been placed up to that point can be used as precursor nodes for a planar graph. The construction procedure of the fractal does not terminate naturally but continues indefinitely due to the scale-invariant nature of the fractal. However, as the procedure is computationally costly, it is sensible to terminate when the radius of the next circle to be placed falls under a threshold value. For *P. polycephalum* networks, a sensible cutoff radius is related to the average thickness of a vein. If the threshold is select too large, thin veins may not be correctly covered with nodes, whereas if the threshold is too small the procedure will have to iterate for a long

time before completing, only to create a large set of redundant nodes. However, this is only a problem with regard to computation time and not with regard to the final result. As will be detailed below, only those nodes whose radius of influence covers a portion of the topological skeleton will be kept.

Turning to the details of the computational procedure presented as algorithm 5.2.1, the starting point are binary representations of *P. polycephalum* networks as seen in figure 5.2.2 panel (b). These are obtained from gray-scale images similar to the example shown in figure 5.2.2 panel (a), via a custom local adaptive binarization routine based on Otsu's method for histogram-based threshold selection (Otsu, 1979). From there on, the algorithm follows the construction procedure outlined above in a straightforward fashion. First, the Euclidean distance transform (Borgefors, 1986) of the shape is computed, which assigns to each point in the foreground of a binary shape the distance from that specific point to the closest point in the background. This makes it particularly useful for the present application, as a circle with a radius according to that distance, which is centered around the location of the maximal value of the transformed image touches the edge separating back- and foreground in one or more locations without overlapping with the background. The coordinates and radius found this way define the first node of the graph, as is indicated in figure 5.2.2, panel (c). There, the euclidean distance transform of the binary image shown in panel (b) is displayed, and the circle found during the first iteration has been highlighted as a red line. In a second step, all points that lie within this circle are removed from the foreground and assigned to the background. Following this, the procedure could in principle be repeated by computing once more the distance transformation of the (now modified) shape. However, this is computationally costly if dealing with an extensive image. In addition, it is noticed that the image changes only in the vicinity of the modified region. It has thus been found to be more efficient to calculate instead the distance transformation in a window centered at the coordinates of the removed circular area, with a window size measuring twice the radius of the removed circle. Repeating this process until the maximal value of the distance transform falls below the threshold value leads to the result shown in figure 5.2.2 panel (d). If the threshold is chosen sensibly (in the present case, half the radius of an average vein has been selected), the surface of the network is tightly packed with circles with only little redundancy, and few veins that are not correctly filled because their thickness undercuts the threshold.

In any case, the result is a graph that does not yet have any edges. The process during which these are obtained, and during which redundant nodes are removed, employs the topological skeleton and is described in the following section.

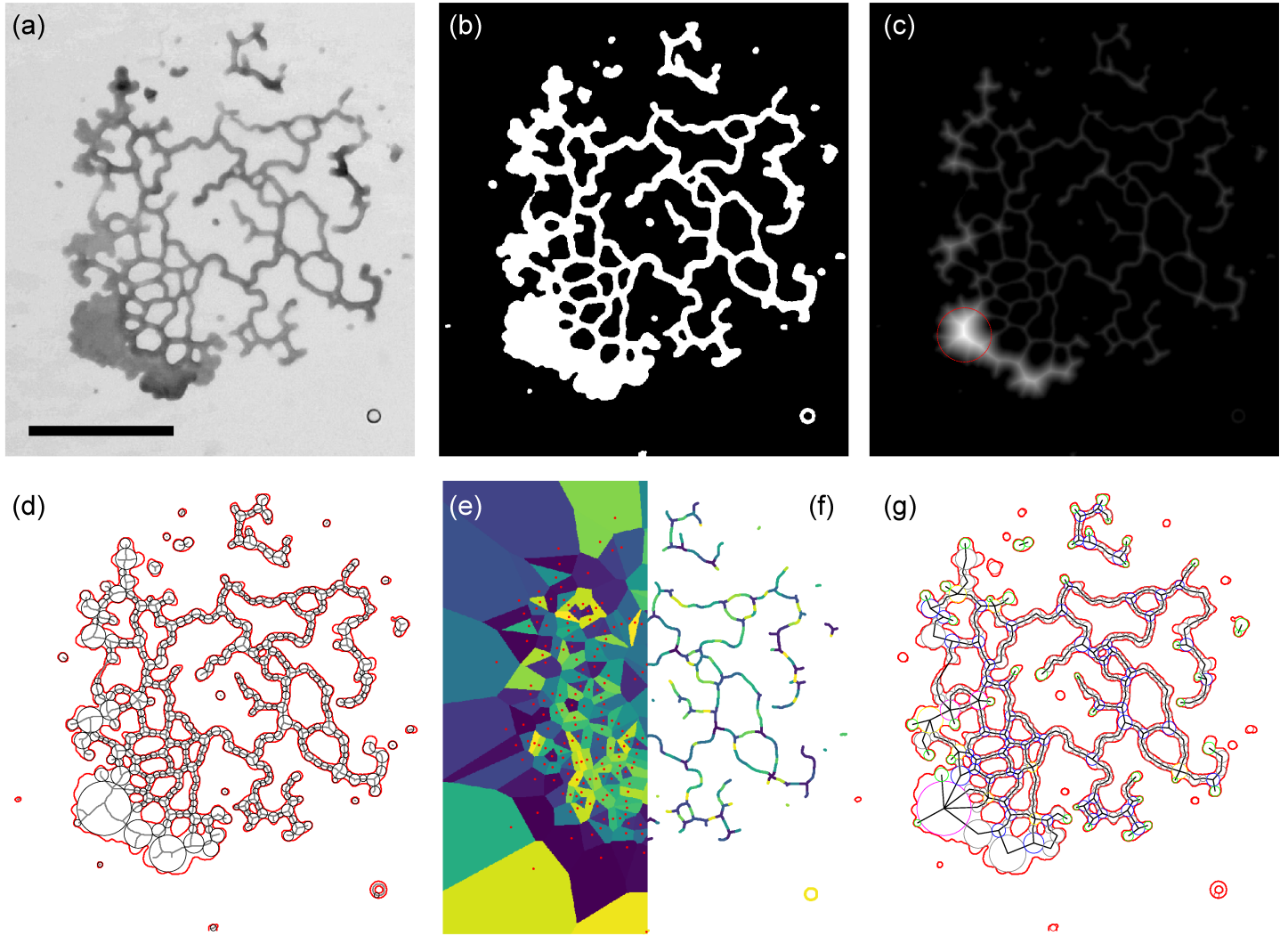


Figure 5.2.2: Apollonian graph approximation. (a) Gray-scale image of a *P. polycephalum network*. The scale bar represents 4 mm. (b) Binarized image obtained by adaptive Otsu thresholding. (c) Euclidean distance transform and circle centered at the maximal value, with the maximal value as radius. (d) Apollonian circle packing and topological skeleton. (e) Voronoi tesselation of circle centers (red dots). (f) Skeleton colored for node influence. (g) Final Apollonian graph approximation. Nodes are colored according to their degree.

5.2.2 Obtaining the set of edges

In section 5.1, a number of shortcomings of the traditional skeleton for representing planar graph data have been named. Although these objections remain valid, the skeleton can be used to establish the connectivity between the nodes whilst avoiding the problems associated with the skeleton. The placement of nodes detailed in the previous section constitutes a natural definition of second degree nodes, thus overcoming the one of the main problems of the skeleton representation. As the surface of the network is almost completely covered with nodes, details of the skeleton that occur on a scale that is smaller than the minimal radius of nodes are effectively filtered out. These features can be seen in figure 5.2.2 panel (d), where the skeleton is drawn in addition to the nodes discovered in the previous section. First, all veins of the network a discretized into a number of circles that are almost homogeneous in size, with their diameter reflecting the thickness of the vein. Vein length can be given in units of the average node radius or equivalently, by the number of second degree nodes between two branching or end points. Second, turning to the largest circles in the same panel, it is noticed that intricate skeleton structures are, in a sense, summarized by a single node.

Omitting algorithmic details, the procedure, which establishes a planar connected graph from the skeleton and the set of nodes can be briefly outlined as follows. First, nodes whose radius of influence does not overlap with the skeleton are removed from the graph. Second, the connectivity between the remaining nodes is established by starting at an end or branching point of the skeleton, identifying the node that is closest, and tracing the adjacent skeleton edges until other nodes are encountered. Once this step is completed for all end or branching points, duplicate edges (edges that are present twice) and self edges (edges that attach to the same node twice) are removed from the graph in a final step.

The procedure is given in algorithmic form in algorithm 5.2.1. Detailing the steps outlined above, the algorithm commences by computing the topological skeleton through a thinning operation, and by identifying end and branching points based on the number of neighboring points. It should be emphasized at this point, that the skeleton as used here does not contain any redundant points, i.e., removing any point from the skeleton changes the connectivity it represents. Next, a graph structure is initialized from the set of nodes, and each node is assigned its radius as an additional information. At this level of detail, it is not important, by which data structure the graph is represented exactly. Suffice to say, that additional information needs to be defined on the nodes as detailed before, whereas edges are undirected and have no other function but reflecting the connectivity between nodes. Prior to establishing this connectivity, it is checked for all nodes whether their radius overlaps with the skeleton. If this is not the case, the node is seen as redundant and thus removed from the graph.

Not all points of the skeleton are overlapped by some node, which is not a problem unless the point in question is an end or branching point. It is intended to explore the skeleton from one end or branching point to another, and attributing this connection to the graph. To be able to do this, each end or branching point needs to be associated with a node, as can be illustrated as follows. Following an edge of the skeleton from

one point that is correctly associated to a node to one branching point where this is not the case raises a problem. From this point, at least two other points can be found by exploration. If these are correctly assigned to other nodes, the problem can be solved by simply connecting them to the last node that was encountered. However, as soon as such a problem extends to the second neighbors, serious distortions from the spatial network structure become possible. To avoid this problem, it is ensured that each point on the skeleton is associated to a node. The simplest measure to base this on is spatial distance. The tool that lends itself to this task is Voronoi tesselation (Okabe et al., 2008), technique which assigns to each point of a set of points in the plane the region in which the distance to that point is shorter than the distance to any other point in the set. Figure 5.2.2 panel (e) shows the Voronoi diagram for a portion of the set of nodes, where each node is characterized by its center. Based on this diagram, each point on the skeleton can be associated to a node. In a sense, the skeleton is colored based on distance, as shown in panel (f) in the same figure. Now, having ensured connectivity of the graph, edges between nodes are established by starting at end and branching points and exploring all adjacent branches until discovering a portion of the skeleton that is assigned to a different node. Then an edge is established. After completion of the exploration process and unless these cases were suppressed in the process or by the selected graph model, there is likely a number of duplicate edges and possibly self edges, leading from one node to the same node. These need to be removed from the graph before returning the final result.

The resulting graph is shown in figure 5.2.2 panel (g), and represents a typical example of the graph representation of an adult *P. polycephalum* network, as will be analyzed in greater detail in chapter 6.

■ 5.2.3 Criticism, improvements and future directions

Before moving on, it should be emphasized again that the resulting graph is one valid first approximation of a network, which has an number of caveats. Primarily, as the spatial discretization follows from the sequential placement of circles, some unfavorable configurations are possible. For example, due to the bounded size of nodes, it is unlikely that a vein is covered without gaps, i.e., nodes that are neighbors are not necessarily in direct contact. Similarly, due to the sequential placement, it occurs that in the neighborhood of average sized nodes gaps remain that are finally filled with nodes only slightly above the threshold, leading to inhomogeneous node size distributions in otherwise homogeneously structured veins. This problem is in part relieved by an optimization introduced by Hillmich (2018), which allows nodes to shift slightly after having been placed. However, the constraint that nodes cannot overlap with each other or the background is still in place. Weakening this constraint slightly, and allowing nodes to grow, shrink and shift in favor of other nodes, could lead to more even configurations. Further concerns with the present procedure arise when taking into account the temporal continuity of the graph representation. These are the focus of the following section.

Nevertheless, by visual inspection results obtained from by the proposed procedure appear to be very well suited, and although there is room for improvement, the graph model that is a approximated is found to lead to good results in chapter 6.

5.3 Tracking topological modifications

In chapter 7, network construction processes are modeled as a sequence of topological modifications. Therein, these sequences are regarded as realizations of a stochastic process that is characterized by a number of parameters. These parameters are rate constants that control the frequency at which the associated modification process occurs. In principle, obtaining experimental values for these constants from temporally resolved graph data is possible in two ways. First, using regression it is in principle possible to fit a model to experimental data. However, the number of parameters is significant and grows in a non-linear fashion with the largest node degree in the graph data, thus leading to an overfitting problem. One is thus required to turn to the second method, which is counting the number of modification that occur during the graph construction process and calculating the rate constants from the result.

Although the described method appears straightforward, there is a number of technical difficulties associated with it. First and foremost, graphs obtained from image evaluation as described in the previous section are calculated for each image individually. Now, in order to determine the modifications that lead to the change of a graph between two time steps it is necessary to monitor how the connectivity of each individual node evolves during that time step. This makes it necessary to track nodes through time, or, more specifically, determine for each node the predecessor and successor in adjacent time steps. In an ideal scenario, this is not a difficult task as the interval at which images of networks are taken are selected small enough to resolve topological changes well. However, due to noise stemming from the image segmentation process and due to the dynamic oscillation of the slime mold, veins constantly change slightly in diameter, or, in less ideal cases may even get severed between images as they fall below the threshold. These effects lead to the erroneous temporary disappearance of nodes, to spontaneous changes in diameter or to shifts of the center, making tracking a complicated task.

This issue has been addressed by Hillmich (2018) in a collaborative project. Therein, the original algorithms in the previous section have been reevaluated and improved where necessary, in order to obtain graph approximations that are more stable in time. Furthermore, a tracking algorithm has been implemented as part of the same project. The improvements aimed at enabling the tracking of nodes made there can be summarized as follows.

First, building on the assumption that networks do not change too much between two time steps, it appears sensible to initialize the set of nodes of a time step $t + 1$ from the set of nodes at time t. In greater detail, instead of obtaining all nodes of $t + 1$ from the circle packing procedure, the centers of nodes found in the preceding image are used as initial guesses. As it is unlikely that a node does not change position at all between two time steps, it is then allowed to move by up to a certain distance, such that the final position maximizes the radius of the node while maintaining the constraint that nodes cannot overlap with the background. This process improves the temporal continuity of the set of nodes and suppresses cases where different configurations of nodes that approximate an area almost equally well alternate between time steps, but comes at the cost of a new parameter that requires to be adjusted. It has been observed, that to large

movement distances lead to a directed movement of nodes along a vein over multiple time steps, as a product of the oscillatory behavior of the slime mold network. Setting up the parameter with care is thus unavoidable. Nodes that cannot be adjusted using the described procedure are added to the graph by Apollonian circle packing.

Second, it is observed in structures that have extensions near the threshold value, that nodes in these regions that are stably detected for a number of time steps cannot be placed when the region drops below the threshold value for a small number of time steps before reappearing. By introducing a maximal gap size as an additional parameter, these gaps can be detected and bridged by inserting a node whose radius, position and connectivity are derived from its predecessor and successor. Again, this optimization, which comes at the cost of adding one more parameter, contributes to improving the temporal continuity of the graph. Unless selecting excessive values for the allowed gap size, this optimization has not been observed to lead to erroneous behaviors.

The previous optimizations make use the concept of the predecessor and successor of a node, and thus make it necessary to introduce the tracking procedure. Omitting here the technical details, which are found in Hillmich (2018), it is possible to associate a node at one time step with a node at a later time, based on a number of criteria. These criteria include differences in radius, spatial distance and temporal distance, the latter being identical to the gap size introduced above. By definition, a node can only have one predecessor and one successor.

Based on the tracking, topological modifications of the graph can be detected by monitoring changes in the set of edges. Briefly, using the symmetric difference of the sets of edges of two adjacent graphs, all those edges can be detected that were either not present in the first graph and have thus been established between the time steps, or those, that were present in the first, but can no longer be detected in the subsequent time step. In an ideal scenario, the tracking would be completed here and one might identify elementary topological modifications[1] from the set difference as defined above. However, due to imperfections in the image processing and due to the size of the time steps, it is observed in many cases that one node does not only participate in one single event, but rather interacts with a number of neighbors simultaneously as indicated by changes in the set of edges. Similarly, it is observed in some cases, that multiple nodes appear or disappear in one time step as a connected complex.

Based on the concept that exactly simultaneous changes are extremely unlikely, it appears reasonable to assume that complex modifications represent a sequence of elementary changes. However, the order in which these may appear is ambiguous. To solve this problem, Hillmich (2018) proposes a method to decompose these higher-order modifications into elementary modifications. In essence, a stability score is introduced that allows assigning a temporal order to all simultaneously occurring changes that constitute one higher-order modification by assuming that changes with the highest stability

[1]These include fusion (adding one edge between two nodes), separation (removing one edge between two nodes), growth (adding a new node and connecting it to one existing node) and retraction (removing one node that has one single neighbor, and removing the associated edge). See section 7.2 for more detail.

score are carried out first. The stability score is based on the consideration that the *P. polycephalum* graph structures are locally tree-like, and henceforth, that changes of this graph structure resemble the (de-)construction of such trees. In accordance with the necessity that nodes cannot appear out of thin air, and neither vanish into thin air, it is sensible to assume that growth starts at the root node of such a tree, whereas removal of nodes is more likely in the lesser connected branches. On the technical side, the stability score is calculated for each changing edge as obtained from the set difference. There are two contributions that are based on changes to the number of nodes in the graph and on changes to the set degrees of participating nodes. Briefly, changes are favored that minimize changes to the number of nodes, and changes are favored that minimize the change of node degrees. The stability score is recalculated whenever the most favorable modification has been selected and applied.

Chapter 6

Dynamics of *P. polycephalum* network formation

As detailed in chapter 1, *P. polycephalum* and its network have been a subject of interest for some time. However, the literature on *P. polycephalum* at large focuses on the study of microplasmodia, for instance as a model system for cell mechanics (Fessel et al., 2017) or cell motility (Lewis et al., 2015; Zhang et al., 2017), or on the behaviors of the adult network, including migration or adaptation of the network structure in response to external stimuli or constraints (Ueda et al., 1975; Häder and Schreckenbach, 1984; Takamatsu et al., 2009). This leaves ample room in the middle ground, where the dynamics of microplasmodia lead to the self-organized formation of an extended vein network.

The present chapter focuses on the evaluation of experiments performed on the network formation of the slime mold *P. polycephalum*. The focus lies on the formation of a *P. polycephalum* network from disconnected fragments. These disconnected fragments are microplasmodia (Bernitt et al., 2010), vital units of *P. polycephalum* that can be created by manually cutting or scraping, or by submerging macroplasmodia in liquid shaking cultures containing growth medium (Gawlitta et al., 1980).

The chapter is organized as follows. First, in section 6.1 the experimental procedure used for data acquisition is introduced and a closer look is taken at the type of result obtained in this fashion. Furthermore, in order to sensitize the reader for the type of network formed by *P. polycephalum*, some structural properties of an adult network are investigated. In section 6.2, the topological evolution of *P. polycephalum* is described. Most prominently, it is found that *P. polycephalum* evolves towards a topological steady state in which the network expands and migrates while maintaining a constant degree distribution. Furthermore, in the light of percolation, the evolution of a giant component is investigated and compared to analytical solutions prepared in chapter 4. A shift of the percolation transition with regard to the models is noticed and possible sources for the shift are discussed. In section 6.3, an attempt is made to distinguish between functionally different phases of growth observed in the slime mold. In total, four clearly distinguishable phases are identified and evaluated with regard to the structural properties of the network. Building on the tracking procedure for topological modifications described in section 5.3, rates characterizing the development of *P. polycephalum* net-

works are obtained from experimental data in section 6.4, and it is concluded that the phases of growth can be distinguished based on these rates. Furthermore, it is shown that the rate equation model proposed in section 7.3.1 appropriately recreates the topological evolution of *P. polycephalum* using the experimentally determined rates as parameters.

6.1 Experimental procedure & available data

6.1.1 Experimental procedure

Data analyzed within this work have been exclusively obtained using an identical routine and an identical imaging system. The experimental procedures is similar to that used for data acquisition in Fessel et al. (2012, 2015*b*) and can be summarized as follows. Microplasmodia of the slime mold *P. polycephalum* are maintained as axenic shaking cultures containing a volume of 100 ml growth medium, plus microplasmodia. The growth medium is a semi-defined medium that is kept at an acidic pH of 4.6. The contents (5 g/l glucose, 5 g/l enzymatic hydrolysate of soybean meal and 0.01 g/l hemin) have been modified after Daniel and Rusch (1961); McCullough and Dee (1976). Liquid cultures are kept in the dark at an ambient temperature of 24 °C. Cultures are passaged every three days up to a total of 30 passages. For inoculation of a new passage, the microplasmodia contained in 2 ml volume of the previous culture are used. All experiments contributing to the dataset have been performed with 3 days old cultures, when the microplasmodium biomass in the culture has regrown to a reasonable amount and when microplasmodia are still well fed. Glucose depletion in the culture is expected at a similar time (Lee et al., 2018). Performing experiments on later days leads to substantially different results, including a prominent 'mesoplasmodium' pattern, where microplasmodia no longer percolate into one network but evacuate the site of inoculation as extended, globular units (Lee et al., 2018) when plated on agar with a low glucose content.

In order to perform experiments, microplasmodia of the slime mold *Physarum polycephalum*, strain WT33×LU898, are harvested by centrifugation of a 2 ml volume of the liquid culture at 331 g for 3 min. After removing the supernatant, microplasmodia are resuspended with 2 ml of growth medium and henceforth plated on a 9 cm plate containing semi-defined medium and variable amounts of glucose, ranging from 2.5 g/l to 20g/l. The volume of droplets in which microplasmodia are transferred to the agar plate is subject to experimental control. In the present work, the volume of inoculation was set to be either 25 µl or 100 µl. In case of the smaller volume, $15 - 25$ network growth sequences can be monitored in one dish, in case of the larger volume this number is limited to four. This discrepancy in experimental efficiency is reflected by the respective populations in the dataset. In total, 522 network growth sequences have been analyzed for this work, consisting of 380 sequences inoculated at 25 µl, and the remaining 142 experiments inoculated at 100 µl.

Adequate imaging of *P. polycephalum* network development requires a versatile imaging system. The diameter of individual veins varies from below 100 µm to 500 µm, while developed networks may span several centimeters. Development of such a network usually requires several days. Given the stochastic nature of network growth and the finite

size of the system, a large number of experiments is necessary to obtain good statistics and explore the influence of system size. A suitable imaging system needs to provide, alongside resolution and field of view as specified above, the means to control temperature and humidity over an extended time frame.

This need is met by a custom made setup, which has been slightly modified from the setup used for imaging in (Fessel et al., 2012, 2015*b*). For image acquisition, a Canon EOS 5DS R digital camera, in conjunction with a Canon EF 50 mm/2:5 macro lens provides excellent resolution at a field of view that allows capturing of the entire dish. The camera is mounted above the agar, while suitable illumination at low intensity is provided from below. The atmosphere around the plate is kept at a constant humidity of 84 percent by contact to a supersaturated potassium chloride solution. The entire system is mounted in an incubator capable of heating and cooling, in order to maintain a constant temperature of 24 °C and in order to avoid exposure to other light sources.

In all experiments, images were taken every 30 s for $24 - 36$ h. This time frame allows the slime mold to regrow into one network containing all inoculated microplasmodia. Imaging was usually terminated when networks started migrating from the site of inoculation. For image processing and evaluation, custom automated routines implemented in Python 3.6.4 (van Rossum, 1995) were used. Processing steps include pre-processing, namely cropping and filtering, subsequent binarization using a custom method based on the adaptive Otsu's method (Otsu, 1979), and further post-processing including volumetric filtering and the removal of foreground and background regions containing only few pixels. Graph extraction from binary images has been performed using the procedures presented in chapter 5, in an improved version implemented by Hillmich (2018). The procedures listed above, and procedures implemented for further analysis rely on the following packages: NumPy (Oliphant, 2006), SciPy (Jones et al., 2001), NetworkX (Hagberg et al., 2008), Scikit-Image (van der Walt et al., 2014), Scikit-Learn (Pedregosa et al., 2011), Matplotlib (Hunter, 2007).

6.1.2 Available data

As described in the previous section, two parameters were subject to experimental control. First, the nutrient content of the agar was varied, and second, different volumes were inoculated onto the agar. Furthermore, unlike other works concentrating on the influence of the placement of food sources, the network growth observed here is not constrained in the sense that nutrients are homogeneously distributed in the agar layer. In the present work, only effects associated with different volumes are analyzed systematically. Populations of different glucose concentrations are uneven in size, with the majority of experiments performed at an intermediate concentration of 10 g/l. It has been found that in many cases further dividing the dataset into subsets by both parameters leads to an insufficient sample size for the remaining concentrations. Furthermore, even with both parameters under control, the observed patterns show a high variability. Evaluating the topological observables of interest to this work with regard to nutrient content showed little to no significant effects in the present data. Nevertheless, effects of the glucose concentration can be seen by visual inspection. In particular, it is observed that networks on high-glucose agar do not expand as quickly as those growing on agar with

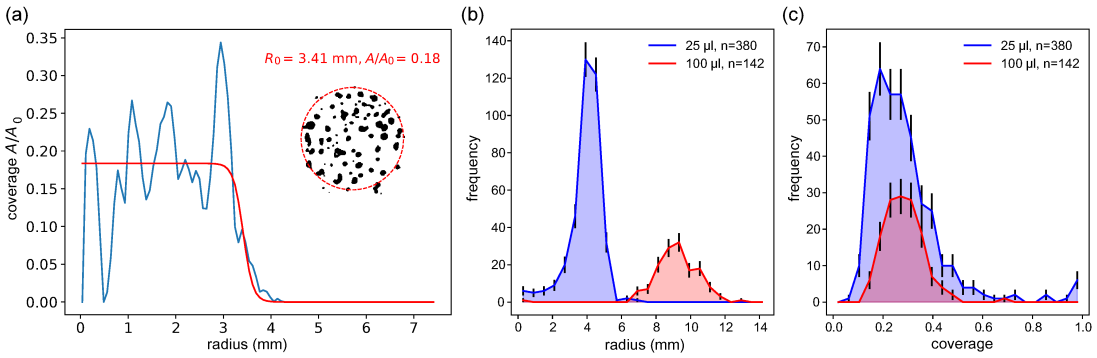


Figure 6.1.1: Patch radius and density. Experiments were conducted with two different inoculated volumes. (a) Procedure for determining initial conditions. The site of inoculation is integrated over the angle and fitted with a tangens hyperbolic. (b) Different volumes lead to distinct radii. Due to technical restrictions applying only to the larger volume, the sample size is in this regime significantly smaller. (c) Area coverage is comparably distributed in both samples, and thus the number of fragments has to scale with patch radius.

little to no glucose content. This effect is accompanied by a lack of pronounced growth fronts in networks growing on high-glucose agar. It is conjectured, that the influence of glucose is most prominently seen during the migration and exploration stages of *P. polycephalum* growth (Vogel et al., 2016) that occur after the structuring of the network is complete. For more information on the influence of attractants and repellents on structure formation during migration and expansion, the reader is referred to the literature (Nakagaki, Yamada and Ueda, 2000; Takamatsu et al., 2009). In the present work, data acquisition is stopped at the onset of these phases, and thus only few effects are seen. In order to show the influence of glucose on the present data, two exemplary analyses are shown in the next section.

The area that is covered when placing a droplet of microplasmodial suspension on an agar plate is referred to as area of inoculation or simply as patch. Inoculating different volumes onto an agar plate results in different patch sizes, and thus to a different number of microplasmodia that may participate in network formation, given equal concentrations in the inoculum. Here, experiments were performed for two different volumes (25 µl and 100 µl) as detailed in the previous section. Figure 6.1.1 highlights the dependence of the radii and coverage, of the patches on the different inoculated volumes. The coverage is defined as the fraction A/A_0 of patch area A_0 covered by microplasmodia. In order to estimate these values, a binary representation of the patch is converted to radial coordinates centered at the centroid of the patch. Next, an intergation is performed over the angle, leading to an estimated radial distribution of microplasmodia as shown in figure 6.1.1 panel (a) after proper normalization. In a spherical patch perfectly covered by microplasmodia, this procedure would lead to a step function. Due to imperfections of the spherical boundary in real patches, and due to the stochastic distribution of microplasmodia in the patch, the step function is smeared out at the boundary, and it is

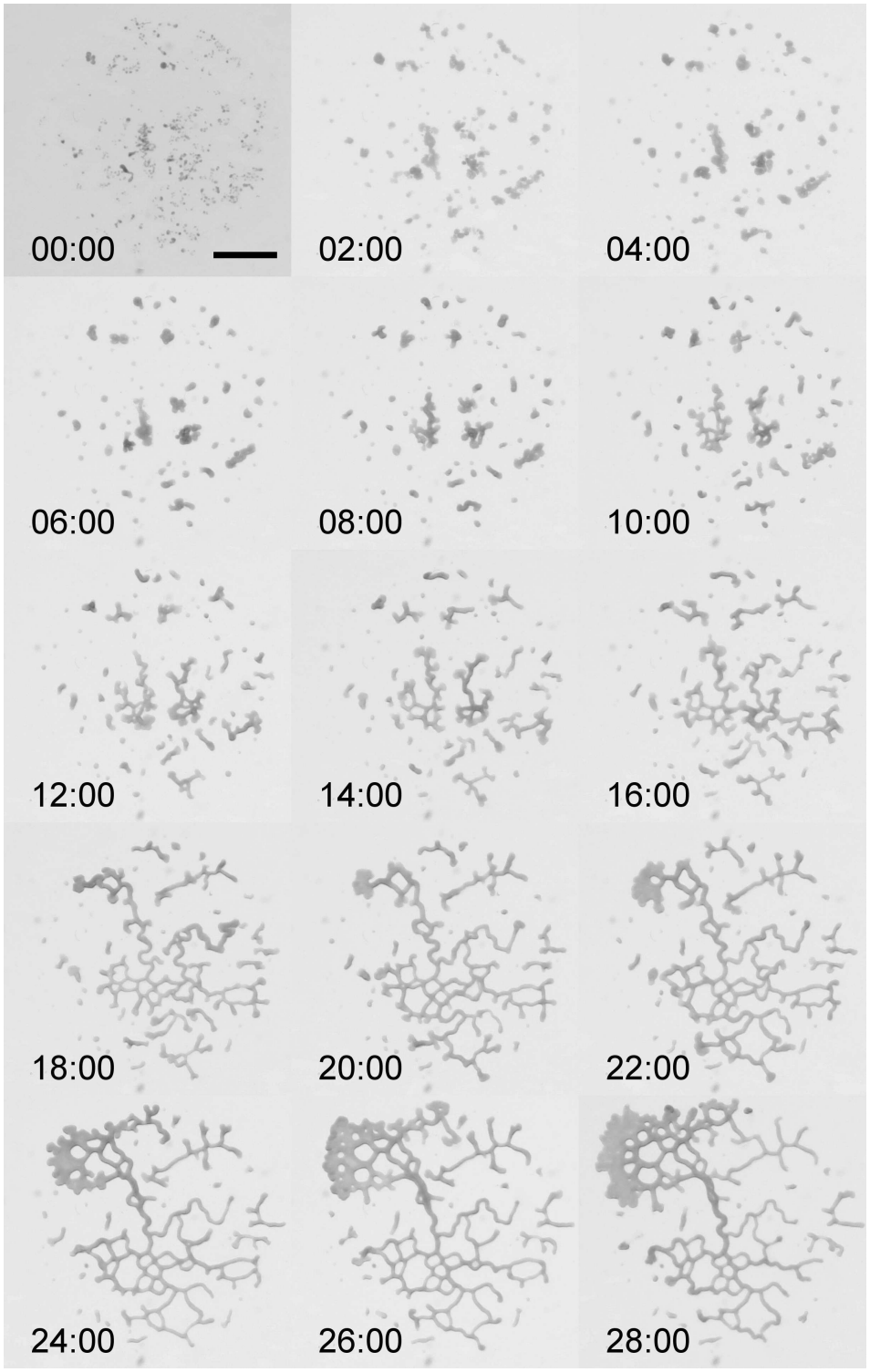


Figure 6.1.2: Typical formation sequence of a *P. polycephalum network*. Time is in hh:mm. The scale bare represents 3 mm. Please refer to the text for a detailed description.

inhomogeneous within the patch. In order to estimate the patch radius r_0 and coverage A/A_0 from the distribution, a tangens hyperbolic is used as an empirical fit model:

$$c(r) = -\frac{c_0}{2}\left(\tanh(d(r - r_0)) + 1\right) \tag{6.1.1}$$

Here, $c(r)$ is the fraction by which a ring of radius r centered around the patch centroid is covered by microplasmodia, and $c_0 = A/A_0$ is the coverage of the patch. r_0 is the patch radius and d is a constant indicating how strongly the step function is smeared out. All three parameters are found by fitting. The results are shown in figure 6.1.1 panels (b) and (c). The radii for both inoculum volumes are displayed in panel (b) and indicate that for the selected inoculum sizes, there is barely any overlap between the populations. In contrast, coverages as shown in panel (c) are centered at a very similar value indicating that the number of microplasmodia is independent of the patch size. As indicated by the histograms, the number of experiments performed at small volume is much larger, which is due to the possibility of fitting a much larger number of patches into one single dish if the patches are smaller in radius.

Having assessed the initial configuration, a closer look can be taken at the typical evolution of a *P. polycephalum* network. A typical example inoculated with small volume (25 µl) is shown in figure 6.1.2. Selected graphs depicting stages of the same series are shown in figure 6.2.1. Over the course of 28 h, it is observed that the initially plated microplasmodia fuse and expand, eventually developing into one connected structure with a pronounced vein that develops growth fronts when expanding over the patch borders.

Directly after inoculation (00:00), the movement and fusion of individual microplasmodia is governed by the hydrodynamics of the evaporating droplet, rather than by movement induced by the slime mold. Microplasmodia are moved due to drying effects and move towards each other in a directed motion due to capillary forces. Therefore, the second frame (02:00) represents the initial configuration from which the initial conditions are retrieved, and from which on the slime mold moves and grows actively rather than passively. The initial configuration of the present example is characterized by a patch size of roughly 4.3 mm at a density of 0.19, placing it in the center of the radius distribution and on the lower end of the density distribution. Furthermore, it is noticed that the distribution of microplasmodia in the patch is not homogeneous. This is frequently observed and can be attributed to slime residues in the liquid culture that lead to a locally variant viscosity. In larger patch sizes, this effect is mitigated by the larger number of microplasmodia, leading to an overall more homogeneous distribution. The drying off usually complete within the first hour, leading to a configuration as seen in the second frame of the time series, where microplasmodia have aggregated to larger structures and are in the process of fusing. Within the next hours (02:00-04:00), there are only few changes visible in the time series. During this time, aggregates containing multiple microplasmodia complete internal fusion events and reorganize, before commencing growth and movement at roughly 08:00 hours. From 08:00-14:00, objects move, fuse and assume local network configurations. In the frames leading up to 14:00, the component size distribution is perceived to be rather spread out. In contrast, when approaching the configuration in 14:00 it is noticed that the majority plasmodial mass is combined in components that are large and possess tubular network structures. Between 14:00 and

16:00, the most prominent topological transition occurs: average-sized components fuse, and shortly after, the majority of plasmodial mass is connected in one giant component. This is a percolation transition discussed in chapters 2 and 4 in a finite-sized system. After 14:00, further fusion and growth events occur and the size of the giant component continues to grow. From 18:00 on, it is first observed that veins become heterogeneous in thickness and height, resulting in a darker color. The last stage of growth observed in the present example covers all frames from 18:00 to 28:00. There, it is on one hand observed that a growth front begins to form, and on the other hand the network coarsens and begins to retract from some regions, as plasmodial mass is moved into the growth region, which eventually expands beyond the boarder of the patch. This is consistent with observations in Fessel et al. (2015*b*), where it is reported that no area is explored by the slime mold before the network is structured to some degree. Typically, point is reached at or shortly after percolation. The vein that connects the growth front with the remainder of the network has to transport a significant amount of plasmodial mass and thickens in the process, in order to accommodate the increased need. This has been first described by the current-reinforcement model by Tero et al. (2007), in which the thickness of veins increases if there is sufficient flow through the vein, and decreases otherwise.

Focusing on the developing growth front, it is apparent that in the uroid region of the fan-like structures, network veins form by hole formation. The number of growth fronts is not necessarily limited to one, as seen in the present case. In other networks, especially in the larger system size, growth fronts protrude in several directions almost simultaneously, and in some cases with exceptionally high coverage the entire perimeter of the patch may be covered by one single connected front. Not shown in the present example is the expansion behavior after the protrusion of growth fronts. It will be discussed later on, that growth fronts move with constant direction and speed for an extended time span in direct succession to the protrusion stage. Upon halting, a wave with a wave length scaling with the size of the network and a period of $60 - 120\,\mathrm{min}$ develops. This 'growth wave' drives the expansion of the network in this phase, leading to a pulsatile forward-forward-backward movement observed in the frontal region, with the expansion velocity equal to the phase velocity of the standing wave. Although the number of growth fronts and other details, such as the fraction of plasmodial mass that is not part of the largest component in the final state, differ, the topological evolution detailed above encompasses the growth of all slime mold networks considered in this work. A second time series, this time for a network grown from the high volume initial condition, is shown later in this chapter in figure 6.3.4 for comparison. In contrast to the present time series, only a selected number of prominent time points, such as the the percolation transition, are shown.

The final state of the time series in figure 6.1.2 is the state towards all networks progress, and in which does not change substantially in the following hours. This state, which is here referred to as the adult or developed network is more closely examined in figure 6.1.3. There, the final frame of the time series in figure 6.3.4 is displayed in panel (b). The network has developed a large number of growth fronts protruding from an intricate, hierarchical network structure (Baumgarten and Hauser, 2013). In the given example, the hierarchical structure of the network can be seen very clearly: a collection of thinner veins constituting a mesh-like structure is connected to two major veins that are

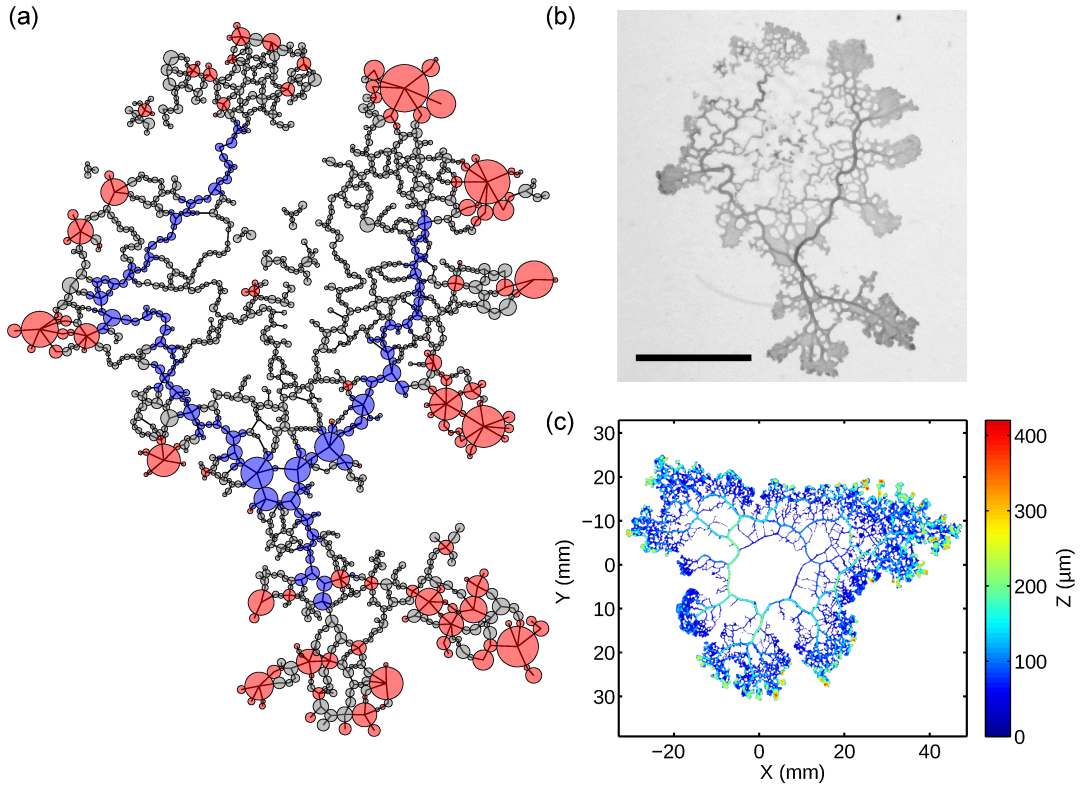


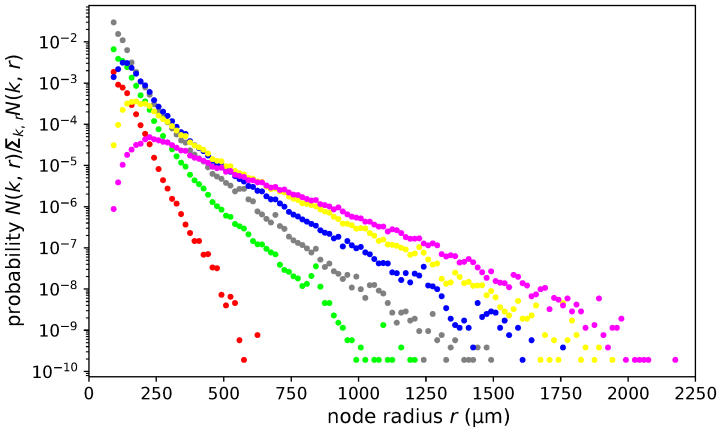
Figure 6.1.3: Anatomy of an adult *P. polycephalum* network aged 30 h. Nodes and veins in *P. polycephalum* networks are heterogeneous in physical dimension, connectivity and function. (a) Cluster analysis of node properties and global connectivity enables sectioning the network into functionally and structurally distinct regions. Nodes labeled in red mark growth front regions and are commonly large and highly connected (or are found in the vicinity of such nodes), whereas blue nodes have high betweenness centrality and form the central backbone of the network. The hierarchical structure emerges only after the network has reached its topological steady state. (b) Original network image, for a time series, refer to figure 6.3.4. (c) Height profile of a *P. polycephalum* network obtained by micro CT (Fessel et al., 2017). Nodes with large height are found in growth fronts, where cellular matter is moved by the network in order to expand, or in the central backbone where nodes have been established over a long time.

clearly discernible due to their darker color. As pointed out above, these veins form, as they are responsible for transporting a the largest amount of endoplasm streaming within the slime mold. These backbone-like veins typically persist on length scales on the order of the size of the network, and it is therefore conjectured that their presence provides the network with improved long-distance transport properties similar to the small-world effect (Watts and Strogatz, 1998), which is typically not observed in mesh-like structures. The darker color of these veins in a gray scale image resulting from transmitted light is associated with their height. This is confirmed by panel (c) of the same figure, showing a

height map of a different *P. polycephalum* network, which possesses a similar structure. The height map has been obtained as part of a collaborative project via fixation, casting and subsequent micro-CT scanning (Fessel et al., 2017), and provides further insight into the general structure of a network. As indicated by gray values in panel (b), veins that constitute the backbone of the network are not only more extended in the plane but possess an increased height. Furthermore, it can be gleaned from the CT-scan that spacious growth-frontal regions are indeed not not homogeneous in height. At the very tips of growth fronts, maximal height values in comparison to the remainder of the network are observed. These are followed by thin, sheet-like regions in which holes may form. For a more detailed view on the expansion dynamics of *P. polycephalum* growth fronts, refer to Baumgarten and Hauser (2014); Rodiek et al. (2015).

A different way to dissect the structure of the network in figure 6.1.3 panel (b) is shown in panel (a), where the graph obtained by the method presented in chapter 5 is shown. As can be seen there, nodes possessing a larger radius of influence (indicate by circles) are predominantly found in the growth fronts, and to some degree in the backbone vein, whereas the remaining nodes are almost homogeneous in size. Typically, large nodes found in growth fronts also have an above-average node degree due to a number of protruding, finger-like structures that are attached to them. These formation of finger-like structures constitutes an important part of the expansion mechanism of growth fronts (Baumgarten and Hauser, 2014). Correlations between node radius and degree will be discussed below.

The coloring of the graph structure in figure 6.1.3 clearly distinguishes between nodes found in the frontal region (colored in red), nodes constituting the backbone of the network (colored in blue) and the remaining nodes that a part of the mesh-like network structure (colored in gray). This coloring has been obtained by an exemplary clustering analysis, an unsupervised learning technique (Duda et al., 2001) that attempts to group unlabeled data into groups of similar features. Here, nodes are classified by measures such as their degree and radius, by the degree and radius of their neighbors, and finally by a measure for the centrality of a node within the network. The betweenness centrality (da F. Costa et al., 2007) is determined for each node individually and depends on how many shortest paths between all pairs of nodes in a graph run through that node. The present result has been obtained by k-means clustering of all nodes in the network, with features as detailed above. K-means clustering (Lloyd, 1982) partitions nodes into a number of clusters such that the squared distance between all points an the respective cluster centers is minimized in feature space. The result highlights that the qualitative observations made above can be quantified based on measures that take into account the functional and structural properties of the network. The observation, that nodes with a large betweenness centrality fall constitute the backbone of the network has been made before (Baumgarten and Hauser, 2013; Fessel et al., 2015*b*) and highlights that the ability of *P. polycephalum* to adapt its structure for the optimal solution of transport problems (Tero et al., 2007; Bonifaci et al., 2012) can also be observed in the unconstrained growth, where no food sources are present. Instead of converging to the shortest path between food sources, in the unconstrained case the hierarchical structure of the slime mold follows from the optimization principle.



0 1 2 3 4 5+

Figure 6.1.4: Node size is correlated with degree. Degree-specific histograms of node sizes normalized by the total number of nodes. The tail of the distributions suggests an exponential decay. The distribution is related to the joined probability of obtaining a node of degree k and radius r when selecting a random node from a randomly selected *P. polycephalum* network.

As pointed out above, figure 6.1.3 indicates that nodes of large radius typically are connected to a large number of neighbors, whereas the number of neighbors connecting to smaller nodes is limited. This observation is consistent with the notion that the physical dimensions of nodes present an upper limit to the possible number of neighbors due to crowding. However, inspecting graphs such as the example given in 6.1.3 panel (a) shows that the the crowding condition is rarely reached. The correlation between node degree, i.e., the number of direct neighbors, and node radius is shown in figure 6.1.4. In the figure, the distribution of node sizes is displayed for all 522 datasets analyzed in this work. A distinction has been made between node degrees, including nodes of degree 0 to 5+, the latter summarizing all node with a degree greater than five.

For nodes of small degree ($k = 0, 1, 2$) the minimal radius allowed in the graph is the radius that is found with the largest probability. Larger radii are increasingly improbable, the decline in probability being most significant in solitary nodes and decreasing with the number of neighbors. The distributions appear to be of an exponential type for larger radii, which is consistent with observations made by Baumgarten and Hauser (2010) on the distribution of edge lengths, i.e., the distance between adjacent nodes on a skeleton, in *P. polycephalum*. However, Baumgarten and Hauser (2010) employs a graph model omitting nodes of second degree, whereas these are included in the present analysis. The distance between adjacent nodes in the present model is approximated by the sum of the radii of both nodes.

Nodes of larger degree decay exponentially as well, but the decay does not start at the minimal radius. Instead, the maximum of the distribution is shifted to a degree larger than the minimal degree. The magnitude of the shift appears to correlated with the degree, i.e., the most probable radius for nodes of degree four is larger than the most

probable degree for nodes of degree three. Similarly, nodes with a radius smaller than the respective most probable radius are found less often if the degree is larger. The probability to find a node with large radius (larger than $\approx 600\,\mu m$) is clearly sorted by degree. In this regime, it is most probable to find a node of degree ≥ 5 and least probable to find a node of degree zero or one, with solitary nodes of this size being virtually inexistent. The strong decline of solitary nodes with radius and their absence at larger radii highlights that it is extremely improbable to find spherical, microplasmodium-like structures beyond a certain radius. Beyond that radius, all structures have either developed holes or have broken symmetry and developed into a tubular shape.

■ 6.1.3 Influence of glucose

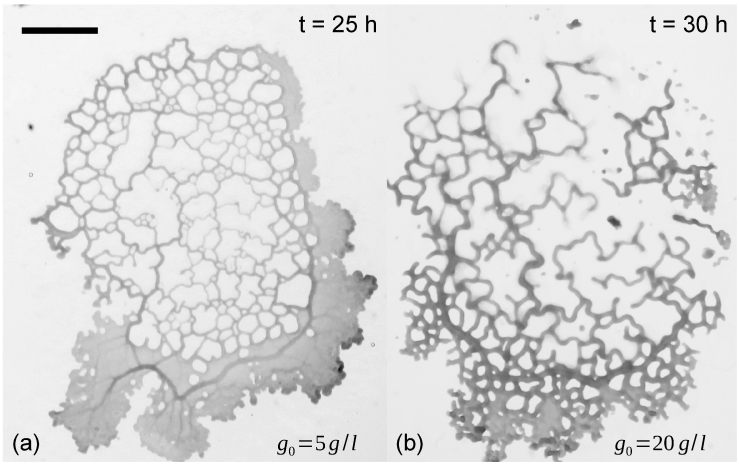


Figure 6.1.5: Adult networks at different nutrient concentrations. (a) On agar with a low glucose content, networks grow into a characteristic structure. Most prominently, first expansion across the border of the are of inoculation occurs via spacious growth fronts. (b) Networks with a high glucose content are visually distinct. Veins appear more pronounced and growth fronts either are not formed or are prone to hole formation. The scale bar represents 5 mm.

It has been pointed out in the previous section, that the control of glucose in the agar has an influence on the morphology and expansion dynamics of the *P. polycephalum* network. Here, effects associated with varying the glucose concentration are illustrated based on two exemplary analyses. In order to provide the reader with a visual cue to how networks formed on different substrates, networks formed on low (panel (a), 5 g/l) and high (panel (b), 20 g/l) glucose agar are displayed in figure 6.1.5. On in inspection, one notices that on low glucose agar, the network appears to be more intricate and mesh-like with a large number of thin veins. In contrast, the network formed on high glucose agar only possess veins of large diameter that appear more pronounced in the image. In addition, is not as meshed as in the previous case. Turning attention to the growth fronts, one notices that these are significantly more spacious and sheet-like in the low-glucose case, whereas growth fronts in the high glucose case visually appear thicker, but with a larger number of holes. In addition, the structure of the leading edge differs. Although finger-like structures are seen in both cases, the border appears to be more smooth in

the low glucose case, and fingers are more pronounced in the high glucose case. Finally, indicated by the time stamps in the figure, networks on high glucose agar tend to protrude across the patch border more slowly.

The observations discussed above hint that *P. polycephalum* attempts to minimize the contact area with the agar substrate. One possible explanation for the observed behavior is that, although *P. polycephalum* remains vital at the high glucose concentration, the glucose is no longer acting solely as a nutrient, but has a toxic effect on the slime mold (Ueda et al., 1976; Dussutour et al., 2010), potentially due to osmosis. The second possible explanation is that with increasing nutrient availability, *P. polycephalum* no longer needs to forage with the same efficiency as in a region where nutrients are scarce, and thus minimizes the energy expenditure for building growth fronts. Although it is likely that both effects play a role in the observed process, the lack of thin veins in addition to the lack of growth fronts makes it seem more likely that the surface is minimized due to contact with a toxic substance. Not analyzed in this work is the influence of starvation conditions before plating. Very briefly, what is observed if *P. polycephalum* is plated later than on the third day after inoculation of the culture, is that the ratio of plasmodial mass contained in the growth front to that contained in the network leans toward the former with increasing age. Eventually, after six days growth fronts no longer are no longer trailed by a notable network and move away from the patch as disconnected units. This effect is referred to as the mesoplasmodium stage (Lee et al., 2018), and it is not yet entirely clear whether the observed effect is due to unfavorable conditions near the site of inoculation or whether it is as a foraging strategy. This effect has so far only been observed on low glucose agar. In any case, the fact that starvation favors growth fronts suggests that their size is maximized for nutrient uptake if nutrients are scarce, indicating again that both effects play a role.

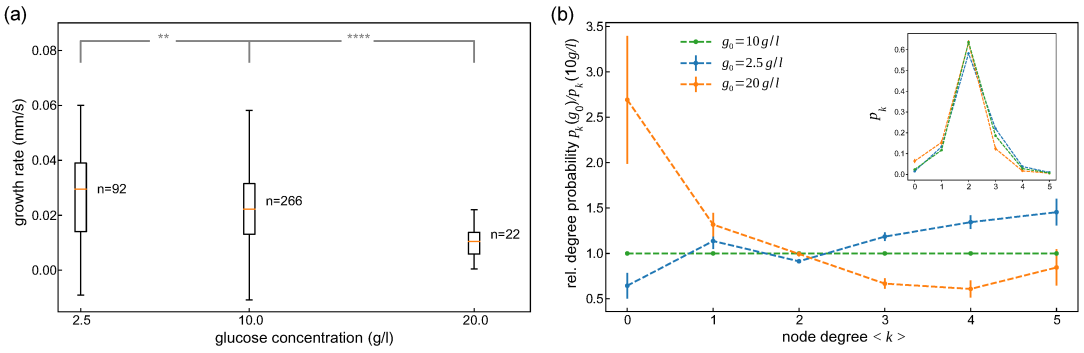


Figure 6.1.6: Influence of the nutrient concentration. (a) Expansion rate versus glucose concentration. Higher glucose concentrations lead to decreased expansion rates. (b) Influence of glucose on the degree distribution. Growth fronts of networks grown on high-glucose agar are less spacious. As a consequence, there are fewer highly-connected nodes. The original degree distributions are shown in the inset. Error bars represent one standard error.

In figure 6.1.6, the effects described above are quantified in set of experiments inocu-

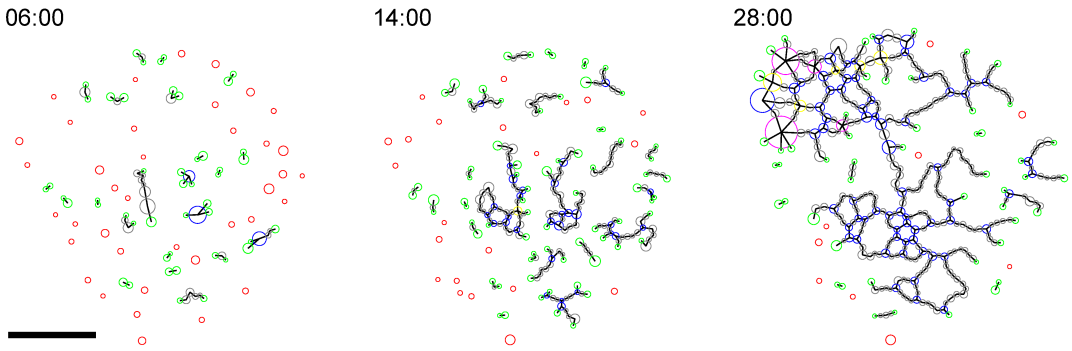
lated at small volume, i.e., at 25 µl. In order to demonstrate the effect of faster expansion and increased growth front size, the rate of area growth is shown in panel (a) as a function of the glucose concentration. The result shows the expected decline in growth rates at higher concentrations, but the accuracy of the analysis suffers from the small sample number at 20 g/l. Although significance levels are high, a rigorous analysis would benefit from a larger number of trials.

Panel (b) of figure 6.1.6 shows the influence of the glucose concentration on the degree distribution in the developed state. The distributions were averaged over all networks sharing the same initial conditions and have been normalized towards the distribution at intermediate glucose concentration (10 g/l) for easier visual inspection. The non-normalized distributions are shown as an inset. Focusing first on the inset, it is realized that the degree distributions do not differ substantially on an absolute scale, which validates the choice to include all data regardless of glucose concentration in the topological analyses presented in this chapter. The normalized distributions show systematic effects. Large degree nodes are found more frequently at low glucose concentrations, and nodes of small degree are present more frequently in high glucose concentrations. Taken to together with observations made above and in the previous section, large degree nodes are commonly found in growth fronts. The increased number of small degree nodes can be explained via the less mesh-like character of networks on high glucose agar, and in the case of solitary nodes that are not part of any connected components by an inability to move and fuse of the particular units, potentially due to a toxic environment. Similar effects can be observed when comparing the node size distributions for different glucose contents. However, due to the tendency of large degree nodes to have a large radius, the node size distribution holds no new information here and is therefore not analyzed separately.

■ 6.2 Topological evolution of *P. polycephalum*

The previous section has familiarized the reader with the network formation process and the structure of typical developed networks of *Physarum polycephalum*. In this section, the same process is evaluated with regard to the graph topological properties and dynamics of the evolving network. To determine these properties, graph representations of 522 image series have been obtained using the algorithms presented in chapter 5.

In principle, the data evaluated here are quite similar to the data set employed in Fessel et al. (2012), reflecting that both studies pursue similar objectives. The present data set is not only larger by a factor of five, but has been conducted with better control of the initial conditions, and reproducibility has been improved at all steps. For instance, in Fessel et al. (2012), the day time point at which experiments are started after inoculation of the culture is not well controlled and varies between day two and day six. It has since then been realized (Oettmeier et al., 2018; Lee et al., 2018), that the age of the culture has a crucial influence on the growth and exploration behavior of *P. polycephalum*. However, the most central improvements have not been made to the data acquisition, but rather to the data evaluation.



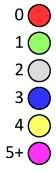


Figure 6.2.1: Selected graphs associated with the network evolution depicted in figure 6.1.2. The shown stages are the initial configuration after aggregation (06:00), average sized components before percolation (14:00) and the final state with a pronounced growth front protruding from the patch (28:00). Time is in hh:mm, the scale bar represents 3 mm.

Very briefly, the routines presented in chapter 5 approximate spatial networks with extended veins by a planar graph whose nodes are described by a position in the plane and by a radius. Due to the planarity of the network, nodes can only possess edges that connect them to their direct neighbors. In the ideal case, the circles given by the radii of connected nodes share a tangent. In reality, due to the approximative nature of the procedure used for graph extraction, there are gaps between nodes. However, as the connectivity of the graph is derived from the topological skeleton (Blum, 1967) of the underlying network rather than from geometrical considerations addressing circles and their neighbors, the graph is not disrupted. In contrast to the topological skeleton, the graph model employed here fixes the number of degree two nodes. The number of degree two nodes in a vein is defined as the number of disks with a diameter approximately equal to the thickness of the vein. In figure 6.2.1, a number of graphs are displayed for selected time steps in the time series in figure 6.1.2.

■ 6.2.1 *P. polycephalum* evolves towards a steady state

The set of graphs representing the evolution of a network can be analyzed with regard to various measures (da F. Costa et al., 2007). Here, focus is first put on the degree distribution, which presents the basis for numerous network applications and hold important implications for the structure of a graph. In chapter 7, modifications of the degree distribution provide the basis for stochastic modeling of network evolution. The degree distribution, and the associated average degree are intrinsic properties of a graph, and therefor do not scale with the system size. Thus, all data sets have been taken into consideration for this section.

The degree dynamics of the example time series shown in figure 6.1.2 are shown in figure 6.2.2 panel (a), starting after the aggregation step has completed[1]. The dynamics

[1]Microplasmodia move due to drying effects and capillary forces during the first hours. This step is

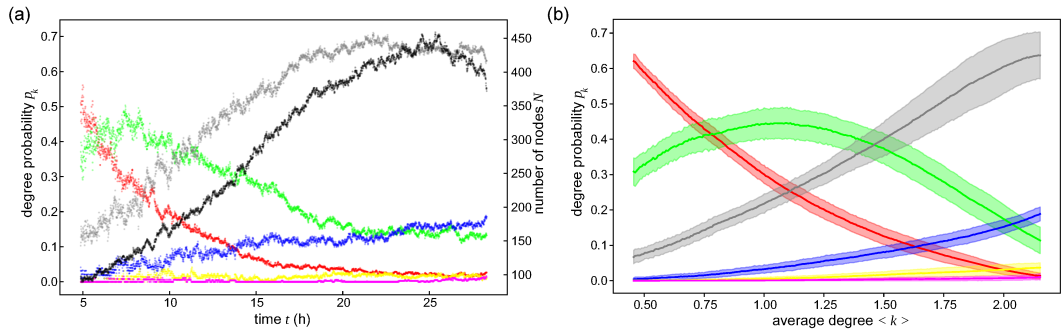


Figure 6.2.2: Evolution of topological properties of the network formation time series shown in figure 6.1.2. Dynamics of the degree distribution. Initially, the network consists of solitary nodes and locally connected components. As time progresses, components fuse and extend, causing the degree distribution to shift to larger node degrees. (a) Degree distribution as a function of time for the time series shown in figure 6.1.2. At around 20 h, the degree distribution no longer changes as the network has reached the adult stage. The number of nodes changes with network growth and is shown in black. (b) Degree distribution and one standard deviation averaged over all 522 time series as function of the average degree.

agree with the qualitative description of network growth given above. Dominated initially by solitary nodes and locally connected components, the network evolves to larger degrees as time progresses. The ratio at which nodes are present in the initial configuration fluctuates between different experiments due to the preparation procedure and the shape of microplasmodia. Generally, this state is dominated by nodes of degree zero and one with few nodes of larger degrees if elongated objects are present. In the following hours, plasmodial fragments evolve either by moving until another fragment is encountered and fusion may occur, or by growth, i.e., elongation of spherical objects and elongation of tubular structures.

In any case, the system evolves first into a configuration dominated by components that have only a small number of holes, and therefore resemble trees. During this stage, the number of degree one nodes that mark the end of all branches is maximal. Due to growth and fusion, tree-like components develop holes and are joined with other components, eventually leading to the percolation transition. This will be discussed in the following section. Once the disconnected network has fused into one single component, structuring is typically almost complete. Node degree probabilities continue to change until *P. polycephalum* has reached a final structure, which it maintains during the onset of migration and during the migration period. This final structure is characterized by a constant relation between node degree probabilities, as can be seen in the figure after ≈ 20 h. Although growth and network remodeling continues, and even though fractions of the network are continuously degraded, no changes of the degree distribution other

generally left out during the data analysis as it does not reflect the behavior of the organism. See section 6.1.2.

than fluctuations due to the finite size of the organism are observed.

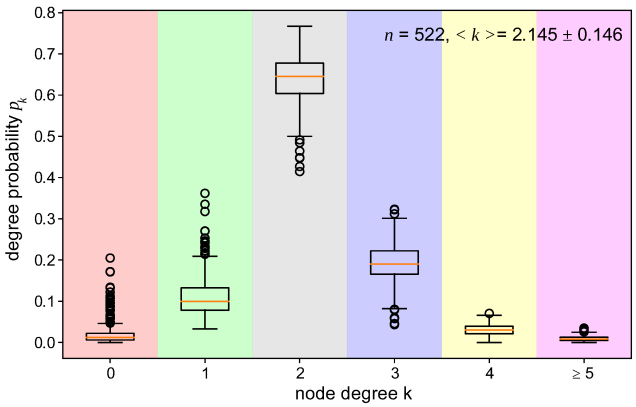


Figure 6.2.3: Topological steady state of *P. polycephalum*. The adult network formed by the slime mold has a fixed degree distribution, which is dominate by nodes of degree two. Nodes with a degrees larger than three are become increasingly rare, validating the assumption that the network formed by *P. polycephalum* can be viewed as a small degree network. Average values and standard deviations in numerical form are summarized in table 6.3.1.

This topological steady state is shared across the final states of all time series analyzed during this work. The statistics of the degree distribution at the topological steady state are shown in figure 6.2.3. As seen there, nodes of degree zero are typically depleted ($2 \pm 2.4\,\%$). As they are not functional units of the graph, they can be neglected at this point. Furthermore, nodes with degree four or five are detected, but occur with rather small frequencies of $3 \pm 1.3\,\%$ in the case of degree four nodes, and only $0.9 \pm 0.7\,\%$ nodes in the network having a degree that is five or greater.

The network is dominated by nodes of second degree ($61.8 \pm 5.6\,\%$) that connect the set of nodes of degree one ($10.8 \pm 4.4\,\%$), which represent fingers of a growth front or tips of veins, with the set of nodes of degree three ($18.5 \pm 4.5\,\%$) that are junctions of veins. These nodes all have a unique function for the structure of the network, and therefore, the ratios between their respective fractions are meaningful. For instance, the ratio between nodes of degree one and three indicates whether the graph has a large number of branches, i.e., whether it is similar to a tree, or whether the graph has a large number of holes or faces, and could therefore be described as mesh-like. The ratio of nodes that have degree two towards the the remaining nodes (excluding nodes of degree zero) relates the expected vein length to the average thickness of all veins across the network.

At first glance, the degree distribution found here disagrees with results obtained by Baumgarten et al. (2010). There, the authors find that *P. polycephalum* develops towards a regular graph with a degree of three. However, the discrepancy between the studies can be solved upon closer examination of the material and assumptions taken into account for the respective analyses. First, in Baumgarten et al. (2010) a graph model is

employed that does not systematically defined nodes of second degree, unlike the graph model employed here. Second, the study utilizes a section of a *P. polycephalum* network grown as an extended sheet. As the network is not grown from fragments, there exist no solitary plasmodial fragments. Furthermore, growth fronts and the perimeter of the network are excluded from the analysis, greatly reducing the probability of encountering nodes of degree one or nodes with a degree larger than three. Still, in the present work nodes of degree one or four are encountered even if disregarding growth fronts. It is believed, that in a space-filling, adult network of *P. polycephalum* the number of spurious veins not connecting to any other part of the network and thus ending in a degree one node will be reduced due to the remodeling of the network. Nodes of degree four, albeit encountered only at a low frequency are still possible. Baumgarten et al. (2010) argue that nodes of degree four are a connection of two nodes of third degree, or junction of overlapping veins where no flow is exchanged due to a membrane being present between the channels. From the present data, it is not possible to validate or disprove this claim. Doing so would require microscopic imaging of degree four nodes in the forming network, which has not been performed as part of the study. In any case, the result of Baumgarten et al. (2010) holds for the tubular, vein-like portion of the *P. polycephalum*, which in theory could be extended to infinity as a globally homogeneous graph structure, but does not translate well to the case described here, where heterogeneously structured, newly formed networks are under consideration.

The average degree associated with the degree distribution at the topological steady state is $< k >= 2.145 \pm 0.146$. Generally, the average degree of a graph increases if the graph obtains nodes of a larger degree, and can thus function as an indicator for the structuring state. As a consequence of the topological steady state, the average degree cannot grow unbounded as is the case in the random graph model, but eventually saturates when the steady state is approached. Above, the evolution of the node degree distribution has been discussed as a function of time. Unless the time dependent behavior is of special interest, time is typically not the ideal variable to relate topological measures to, as it is more sensitive to fluctuations in certain parameters such as the initial density. Furthermore, it is more difficult to compare structuring processes that are, in principle, identical series of structuring steps occurring at different speed. Comparing these as a function of a parameter that summarizes the topological state, such trajectories can be collapsed onto a single curve. To illustrate this, the degree distributions of all 522 sequences have been combined by averaging in figure 6.2.2 panel (b) as functions of the average degree. As can be seen there, the resulting curves have a rather narrow standard deviation, indicating that for one average degree only a limited number of degree configurations of the *P. polycephalum* network are possible. Although experiments were set up carefully in order to warrant similar initial conditions, such a collapse is not possible as a function of time.

■ 6.2.2 Graph-based percolation in *P. polycephalum*

In the previous sections, stages during the structure formation in *P. polycephalum* have been investigated. The most prominent change of the structure between the initial and final state is the development of a global connectivity. Solitary fragments of plasmodial mass connect as a result of fusion processes, eventually forming one network that exhibits

globally coordinated dynamics and migrates. This topological phase transition between the disconnected and connected phases can be analyzed in the scope of percolation on a graph topology (see section 2.3 and chapters 2, 4). In the present section, the evolution of component sizes is investigated as a function of the average degree, or of a completion parameter as detailed below. They serve as indicators for the the topological state. Findings on the topological steady state in the previous section validate the assumption, that *P. polycephalum* as a small degree network can be compared to the analytical solutions for percolation on graphs devised in chapter 4. However, agreement between the present data and models is found to be poor, raising the question why such a discrepancy has not been observed in the study published in Fessel et al. (2012). Possible reasons and the circumstances under which the mismatch between data and theory can be corrected are discussed. Furthermore, estimates for the position of the phase transition are given and the influence of system size during all analyses is elucidated.

A suitable driving parameter In chapter 4, the average degree has been employed as the parameter that needs to be adjusted to drive a system into the percolated regime. As done there, it is entirely possible to discuss the location of the percolation threshold in terms of the average degree. However, in many cases percolation is discussed with regard to a driving parameter limited to the range $[0, 1]$, where zero indicates a system that is entirely disconnected, and one reflects the opposite state where structuring in the system is complete. The fact that *P. polycephalum* tends to a topological steady state provides the opportunity to set up such a parameter.

As percolation in *P. polycephalum* occurs in the plane, it appears sensible to compare the process to the standard models representing site or bond percolation on a lattice, where the number of maximal sites or bonds makes it easy to set up a driving parameter that is limited as described above. However, *P. polycephalum* is different with regard to several features. First, there is no regular lattice to which the network conforms. In principle, it is possible to approach this problem as done by Dirnberger and Mehlhorn (2017), regarding the final graph as a lattice and successively removing edges. Although this approach offers insights into the tolerance of the network to random attacks, the reversed process does not resemble the network formation considered here, leading up to the second point. Even though some nodes may be followed through the entire network formation process, these rarely remain in one location, as the plasmodium moves, oscillates and expands. Furthermore, the arena in which the percolation process takes place does not have fixed boundaries. In the present study, the slime mold is not constrained and extends across the border of the site of inoculation. This makes the definition of a final state problematic, as there is no natural limit to the process unless the topological steady state is considered. The previous consideration implies the final difference. Components do not only grow by random addition of edges, but may grow by adding new nodes at their perimeter in an active process. Fusion only occurs due to moving fragments, or if growing units come into contact.

The following consideration leads to a suitable driving parameter $p \in [0, 1]$ for percolation in *P. polycephalum*. During the evolution, it is not known, which edges and nodes that are currently present in the graph will be part of the final configuration, nor

is the final configuration known in detail. It is possible, however, to answer at any point the question whether the graph is percolated given the current configuration. Thus, it appears sensible at any point, to define the driving parameter such that it indicates the completion of the graph with regard to the configuration of nodes rather than the final configuration. The completion is indicated by the number of edges E, that is present in the graph in comparison to the number of edges E_0 that would be in the graph, if the graph was complete. Taken together with the relation

$$E = \frac{N}{2} \sum_k k p_k = \frac{N}{2} < k > , \tag{6.2.1}$$

it follows immediately that

$$p = \frac{E}{E_0} = \frac{< k >}{< k_0 >} , \tag{6.2.2}$$

where k_0 is the average degree in a *P. polycephalum* network that has completed the structuring process. It is similarly possible to substitute for E_0 the number of edges in a different known graph type. For instance, in a complete graph where all possible edges are present, $E_0 = N(N-1)/2$. However, complete graphs are not planar and the number of edges exceeds the possible number of edges in a *P. polycephalum* by far. A more reasonable approach stems from the consideration, that the number of edges in a graph that can be embedded in the plane is bounded. The reason behind this is summarized by the Euler characteristic for planar graphs (Barthelemy, 2018), which relates the number of edges, nodes, faces and components. The advantage of this approach is that the limit approached in the process stems from theoretical reasoning rather than from an experimental value. Still, as the planar graph spanned by *P. polycephalum* does not approach that limit, it appears more sensible to work instead with the parameter suggested by the topological steady state of *P. polycephalum*. Due to the simple structure of the driving parameter, working with the average degree or with equation 6.2.2 is equivalent, and both are used in this work.

Evolution of component sizes Unlike the degree distribution investigated in the previous section, component sizes are extrinsic quantities that are influenced by the size of the system. Therefore, in the present section a distinction is made between the experiments performed at small or large volumes.

Figure 6.2.4 compiles an overview of the evolution of component sizes obtained experimentally and from the theoretical solutions devised in chapter 4. The typical time series used as an example in the previous sections is shown as a point cloud. Two measures associated with this quantity are displayed in the figure. First, the giant component size S, which increases as the system becomes more structured, and second the average size of non-giant components $< s >$ is shown in red. Unlike the latter, the giant component size is normalized towards the total number of nodes in the graph at each point of the evolution, implying that it is bounded between zero and one, eventually evolving towards $S = 1$ if all plasmodial matter is combined in one giant component. The giant component size is typically regarded as the order parameter indicating whether the system has

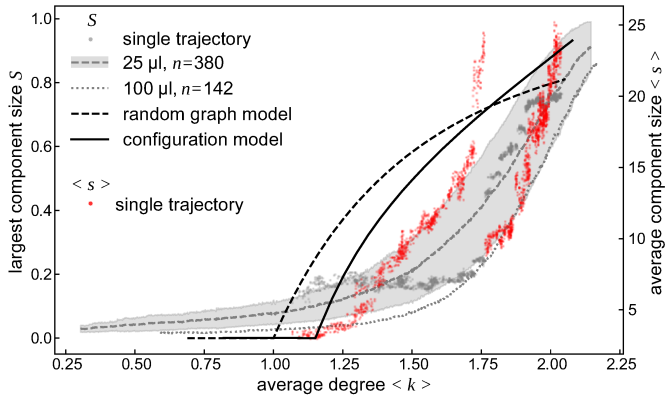


Figure 6.2.4: Percolation in *P. polycephalum*. The normalized giant component size S (grey and black) and the size of average, non-giant components $< s >$ (red) indicate a substantial change in global connectivity. Data do not match well with the random graph or configuration models, and agreement is found to be worse if system size is increased, indicating that the discrepancy is not due to finite-size effects. The location of the transition, approximated by the first peak of the average component size, is considerably shifted with regard to the models. The cascade-like peak structure observed for the average component size is frequently observed and is attributable to the secluded growth of non-giant components.

transitioned into the percolated (ordered) phase.

The average size of non giant components takes the role of the fluctuations of the order parameter, which are maximal at the phase transition. It is small when there are only few nodes connected in each component, approaches its maximum value as components grow large shortly before percolation and declines afterwards as more components fuse to the giant component that is excluded from the average. It eventually tends to zero when all components have merged into the giant component. The average component size is computed by evaluating

$$< s >= \frac{\sum_s' s^2 n_s}{\sum_s' s n_s} \tag{6.2.3}$$

(Stauffer and Aharony, 1992; Binder, 1997). The sum runs over all component sizes s found in the graph excluding the largest component, which is indicated by the dashed sum. n_s is the distribution or concentration of clusters at each point of the evolution. In an infinite system, cluster sizes diverge at the phase transition, and so does the average cluster size. In a finite-sized system as in the slime mold network under consideration, the cluster size cannot evolve past the system size and therefore, the sum is cut off at a finite value, causing the average cluster size to be rounded off at the transition. It is thus maximal and peaked as observed here, but does not diverge.

In the present case, a peculiar structure with two peaks is observed. Similar to the

rounding of the divergence, this is a finite size effect associated with the discrete nature of fusion events between components. It typically occurs if the leading non-giant component is merged into the giant component, and if only few non-giant components remain thereafter. As these components tend to expand locally due to growth events, it is in this situation possible that one of the remaining non-giant components grows to become the second largest component, thus leading to the observed behavior. This is observed quite often in the data set, although the magnitude, frequency and number of secondary peaks varies. The effect is not systematically studied in the scope of this work, but could be investigated systematically by employing the simulations devised in chapter 7. Such an analysis appears worthwhile, as the shifted onset of percolation (discussed in the following), as well as the post-transitional cascades are reminiscent of explosive percolation (Nagler et al., 2012; Achlioptas et al., 2009), where the growth of the largest cluster is suppressed by favoring merging events between smaller components. Similarly, the growth of non-giant components is, in a sense, favored here as these cannot merge globally but instead have to structure and expand locally first, until gaps to the largest cluster are successfully closed. In order to maintain readability of the figure, the non-giant component sizes are only shown for the example trajectory.

In addition to the example trajectory, giant component data is shown in figure 6.2.4 as averages over the fractional data sets with small or large inoculation volume. The average over the small volume population of data sets is shown with one standard deviation as a shaded region. The example trajectory described above, which is part of this portion of the data, remains well within the shaded region. This is not always the case, especially if the structuring is almost complete. In a low density situation, multiple equally sized large components may persist until the end of the experiment and never come into contact due to the stochasticity of the experiments. In this case, in essence two, not one adult networks form from the initial patch, which has a profound effect on the measured size of the giant component. As it is impossible to find data points where the size of the giant component exceeds the maximal number of nodes in the system, the size of the giant component is thus systematically underestimated by a small amount. Similar to this, fluctuations of individual trajectories before the described point can be attributed to the finite size of the system. However, due to the large data set fluctuations are smoothed out well.

For better readability of the figure, the average of the large volume fraction is shown without standard deviation, as it holds little new information. The average, however, reveals that in large systems the size of the giant component is typically smaller in comparison to small systems at the same level of local structure as determined by the average degree. This results in a shift of the percolation transition to larger $< k_c >$ in larger systems, as quantified below (small: $< k_c >= 1.72 \pm 0.15$, large: $< k_c >= 1.93 \pm 0.15$, see figure 6.2.7). This effect becomes especially notable when comparing to the theoretical solutions devised in chapter 4, which are shown in the figure as black lines. Working here with a maximal node degree of $n_k - 1 = 5$, the appropriate solution in the configuration model is given by equation 4.2.14, whereas the solution to the random graph model has been determined numerically.

It is immediately clear from figure 6.2.4 that neither model matches well with the experimental data, although the configuration model appears to fit better. In both models, the transition occurs at significantly lower values of $< k >$ than in the experimental data, and the slope near the transition is very different. Moreover, data and models are curved in opposing directions, with the models approaching the limiting value $S = 1$ immediately after the transition, whereas the data increase comparable to an exponential up to a point of inflection. A first candidate for the source of error would be finite size effects. However, the effect described above is even more significant in the large volume data set, leaving only the conclusion that the models do not agree well with the processes leading to structure formation in *P. polycephalum*, or cannot be applied to the data in the present form. As will be discussed below, both cases provide a possible explanation for the observed mismatch.

Candidates that might be suspected of leading to the observed behavior, in addition to finite size effects that have effectively been ruled out, need to be searched for. They are expected among the assumptions underlying the models that differ from the structure formation process detailed in this chapter. Possible candidates therefore are as follow, (i) the planarity of the *P. polycephalum* network, which implies that it is not possible to add edges between arbitrarily selected pairs of nodes, and (ii) the fact that both models only evolve by the addition of edges, disregarding the possibility of growth processes and thus remaining at a constant system size, whereas *P. polycephalum* graphs change size during their evolution.

Influence of growth This question provides one of the core motivations for devising a network growth model that allows to study the observed effects in detail by varying the relevant parameters systematically. This is done in chapter 7, where a model is constructed that allows four elementary processes (growth, retraction, fusion and separation) to alter the structure of a graph. The model is realized as a rate equation describing the deterministic evolution of the degree distribution, and a stochastic master equation providing on one hand the stochasticity of the degree distribution, and on the other hand the possibility to study the dynamics of component sizes as a result of the elementary processes. As the model is not planar, it does not allow direct study of the influence of planarity. What can be tested is non-planar structure formation resulting from a set of processes that is limited in terms of rates of occurrence similar to a process similar to a process occurring in two dimensions. However, this does not permit the possibility of long range connections.

Based on this premise, it is found that shifting the ratio between the rates at which a system exhibits either growth or fusion processes towards growth, is sufficient to reproduce the effect observed here. If there is no growth in the system, the configuration model is found to be in exact agreement with simulations except for finite size effects, whereas systems characterized by significant growth rates deviate similarly to figure 6.2.4 (see figure 7.5.7). Unlike in a fusion-driven process, where the existence of a giant component is a property of a static graph characterized by its average degree as a driving parameter, growth of a network is an active process (Dorogovtsev and Mendes, 2002). It is thus not too surprising that deviations from the static random graph and configuration models

are observed in the data (see also chapter 2).

The influence of spatial embedding is not systematically studied here. However, by the following argument it appears plausible that space in conjunction with growth also leads to a shift of the transition. Above, the effect of density has been touched upon. In a low density system that evolves primarily by growth, components only fuse if they come into contact. There, components will advance in structure locally, albeit global connectivity can only be achieved once components have extended so far that the density-dependent distance between them is overcome. Although such a system may still be globally disconnected, components can be highly structured and possess an average degree that would have induced a percolation transition in a denser system.

Influence of second degree nodes The disagreement between data and theory in the present data set raises the question how it is possible that a good agreement between *P. polycephalum* network growth experiments and the random graph mode has been reported in Fessel et al. (2012). Briefly recapitulating the study, 48 network growth experiments were conducted with a setup and experimental procedure similar to the one employed here, albeit with less well-controlled initial conditions as discussed above. Data processing steps also compare to the present study, with the significant difference that nodes and edges were obtained from the topological skeleton, rather than from the graph model devised here, leading to the absence of nodes of degree two in the data. For theoretical modeling of the transition, the random graph model was employed, leading to the closed form solution with a maximum degree of $n_k - 1 = 3$ given in equation 4.1.15, which depends on one single ratio of node degree probabilities p_0/p_3 as a driving parameter. The key results are shown in figure 6.2.5. Panel (a) displays the evolution of the giant component as a function of p_0/p_3. The driving parameter starts at large values in an initial condition given by a set of solitary nodes and moves to small values as solitary nodes are depleted and the network structures towards nodes of third degree, and therefore the graph appears to be oriented in opposing direction when compared to figure 6.2.4.

As can be seen in the figure, the theoretical solution is in good agreement with the data, given by a point cloud and the average. The shape of the average is reminiscent of the data analyzed in figure 6.2.4, but agrees with the theoretical solution significantly better. Also shown in the figure are data obtained from percolating human endothelial cells as published in the form of an image sequence by (Gamba et al., 2003; Serini, 2003). Analyzed with the same procedure used for *P. polycephalum* networks, the data are found to be in excellent agreement with the close form solution. The relation between the parameter p_0/p_3 and the average degree employed as a driving parameter in the present study is shown in figure 6.2.5. It is found, that at the transition the graph topology is strongly constrained, causing all experimental trajectories to pass through the phase transition point $(p_0/p_3) = 2, < k >= 1)$ with identical slope. Further features of the theoretical solution employed there are discussed in Fessel et al. (2012) and in chapter 4.

Upon further investigating the observed discrepancy between current data and the date from Fessel et al. (2012) it is first noticed that employing p_0/p_3 as a driving parameter does not solve the problem. However, eliminating nodes of second degree from

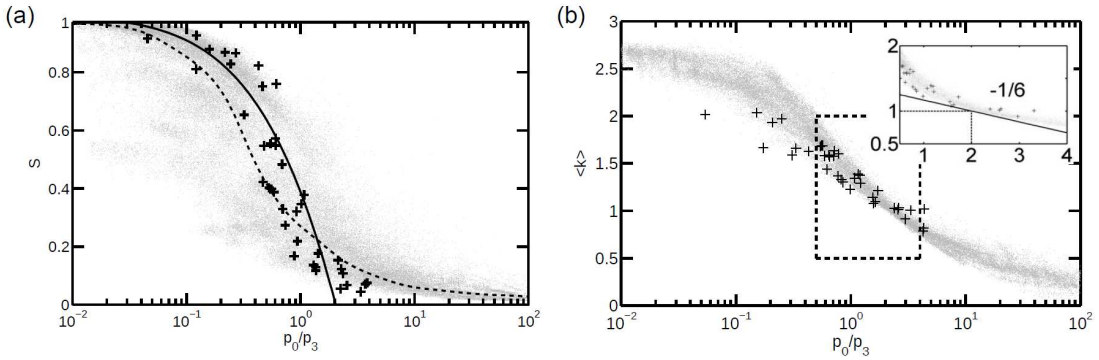


Figure 6.2.5: Percolation of *P. polycephalum* and human endothelial cells as published in Fessel et al. (2012). (a) Giant component size of *P. polycephalum* (point cloud, moving average given by the dashed line) and vascularizing human endothelial cells (crosses) as a function of the node degree fraction p_0/p_3. The solid line indicates the random graph model in the case $n_k = 4$, $p_2 = 0$ as calculated in chapter 4. (b) Average degree as a function of p_0/p_3. At the transition (indicated by $p_0/p_3 = 2$), the topology is strongly constrained: all datasets pass through the point $< k >= 1$ with almost identical slope (solid line, inset). *P. polycephalum* obtained from 48 experiments; endothelial cell data is from reanalysis of image data published in Serini (2003).

the degree distribution of the present data and renormalizing accordingly leads to the situation depicted in figure 6.2.6. There, the giant component averaged over all available data regardless of system size is shown in comparison to the theoretical models devised in chapter 4. In panel (a), nodes of degree two are still part of the degree distribution, and therefore agreement between theory and experiment is comparable to figure 6.2.4. Removing nodes of degree two results in a significantly better agreement between data and theory. In the case of the random graph model, agreement is comparable to the result in Fessel et al. (2012). The configuration model is in even closer agreement, describing the growth of the giant component almost exactly after percolation. Prior to percolation, there is still a mismatch, which resembles the smearing-out of the transition due to finite size effects. In section 7.5.2 it is shown via simulations, that the remaining mismatch is indeed a finite size effect and can be reduced by increasing the system size, unlike the mismatch shown in panel (a) or in figure 6.2.4.

The procedure of eliminating nodes of degree two from the degree distribution in order to obtain better agreement between data and the configuration model works very well for the present data and in systems characterized by similar growth and fusion rates, but does not hold generally. In figure 6.2.6, the average of simulated trajectories obtained from 1024 simulations of network growth generated using the procedure described in section 7.3.3 are shown in addition to the non-synthetic slime mold data. Initial conditions and the rates at which topological modifications occur were set up to match with the values obtained from *P. polycephalum* data in this chapter. In both panels of figure 6.2.6, simulated data approximate the evolution of the slime mold reasonably well. Similar to the experimental data, agreement between the synthetic data and the theoretical model

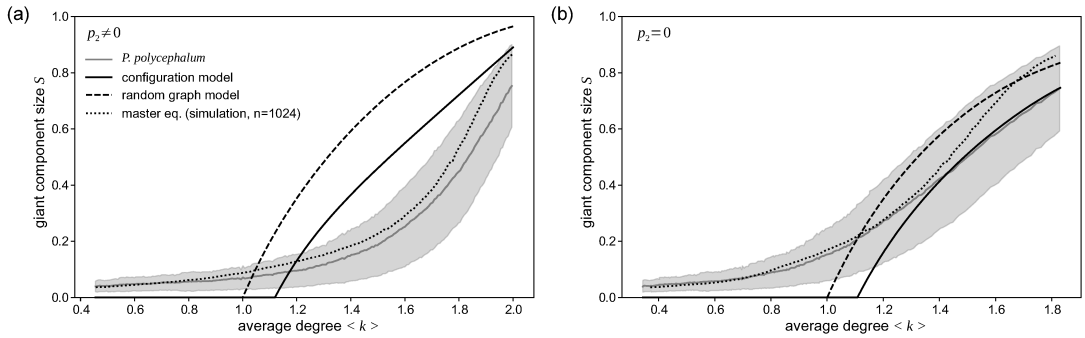


Figure 6.2.6: Eliminating second degree nodes significantly improves the agreement between *P. polycephalum* percolation data and the configuration model. (a) Simulations performed with experimental parameters agree well with data for the size of a giant component in *P. polycephalum* networks. However, the theoretical random graph and configuration models do not quantify the situation well, as there is growth present in the system. (b) This discrepancy has not been observed in (Fessel et al., 2012, 2015*b*). Indeed, this effect is explained by the lack of second degree nodes in the models applied there. Artificially eliminating nodes of second degree from the data displayed in (a) alleviates this problem.

is significantly improved if nodes of degree two are removed, as can be seen in panel (b) of the same figure. In contrast, it has been observed that if the rates characterizing the simulations are not set in accordance with experimental data but rather employ significant growth rates, the effect of removing second degree nodes is diminished and agreement between theory and simulation cannot be considerably improved.

Estimation of the critical point It has been observed in this section that the location of the critical point is shifted with regard to the random graph or configuration models as null models. Determining accurately the location of the percolation transition from experimental or synthetic data is a difficult task that typically requires the study of numerous system sizes. Various techniques have been utilized in the literature that make use of the finite-size scaling hypothesis (Binder, 1997) as described in section 2.1.3.

Although experiments have been performed for two approximate system sizes, and although the number of realizations is on the order of 10^2, one cannot expect the methods outlined there to lead to accurate results in the present data. Instead, the following route is taken. Here, the critical point is first estimated for both system sizes using a phenomenological approach. In a second step, simulations mimicking the evolution of the system are performed, leading to results similar to those seen in figure 6.2.6 for the giant component. Computationally, the system size can be much more freely varied at the cost of computational efficiency.

Finite-size effects lead to a smearing out of the percolation transition. The giant component size is not zero before the transition as in a finite system there will always be one largest component. Similarly, the average non-giant component size is rounded

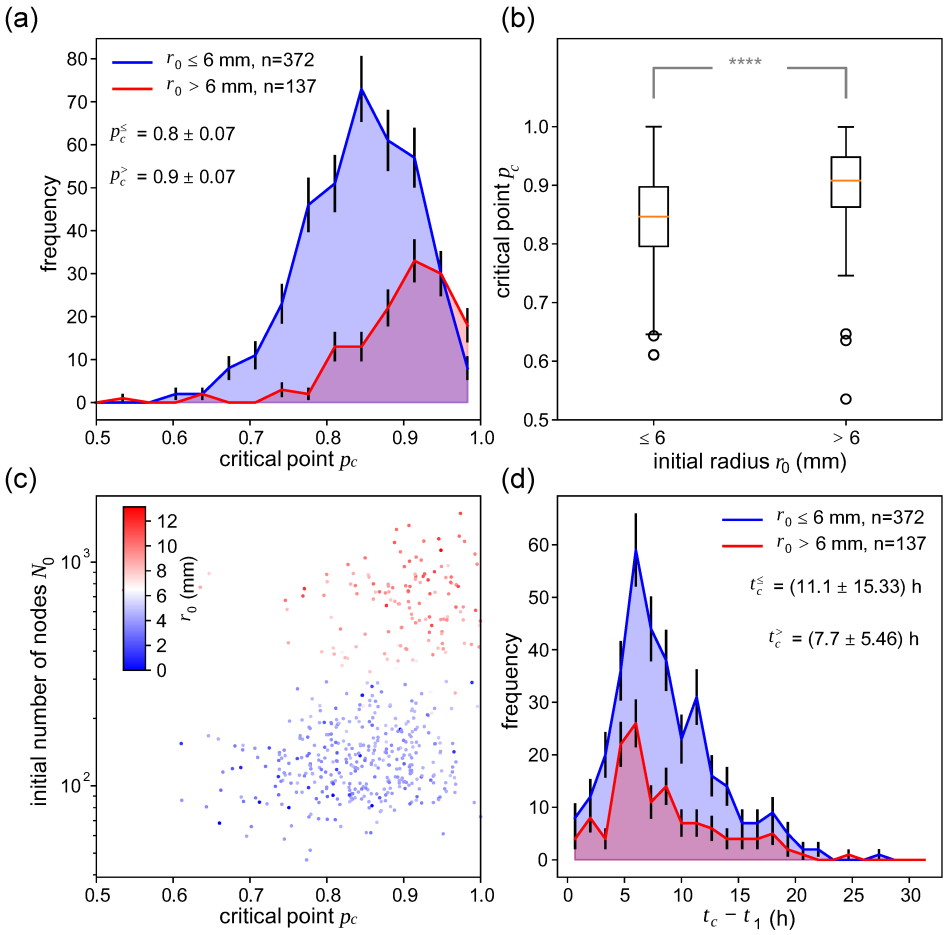


Figure 6.2.7: In a finite system, the giant and non-giant components share similarity with a logistic function and its derivative. Fitting this heuristic model to experimental trajectories leads to an approximation of the critical point. (a) Histograms for both system sizes. Due to finite size effects, the critical point is not identical for both system sizes. (b) Although histograms overlap significantly, statistical testing shows that the shift is significant ($p < 0.0001$). (c) Data shown as point clouds separated by number of nodes. Within the populations, there are no further trends visible. (d) Time at which the transition occurs, measured after local aggregation of microplasmodia (see figure 6.3.3). Values are given as mean ± standard deviation.

off as the correlation length cannot diverge. As a consequence, the giant component size shares similarity with a logistic function

$$S \approx \frac{S_{\max}}{1 + e^{(-k_S(p-p_c))}} \tag{6.2.4}$$

and the average component size resembles its rescaled derivative,

$$< s > \approx \frac{4 < s >_{\max} e^{(-k_{<s>}(p-p_c))}}{\left(1 + e^{(-k_{<s>}(p-p_c))}\right)^2} \ , \tag{6.2.5}$$

which is peaked at the point of inflection of S. Further, $S_{\max} = S(p = 1)$ and $< s >_{\max} =< s(p = p_c) >$ are the maximal values taken by either function, and k_S, $k_{<s>}$ are free parameters characterizing the steepness of the logistic function and the peak width of its derivative, respectively.

Simultaneously optimizing the parameters (including p_c) of both functions enables rather robust determination of the peak and therefore allows estimation of the system-size dependent critical point. Results are shown in figure 6.2.7. In panel (a), histograms for the location of the critical point are shown, separated by patch radius in the initial configuration. It is depicted that there is a rather large variance due to fluctuations of the individual trajectories and due to possible inaccuracy or missdetection during fitting. However, it is seen that the estimated critical point characterizing the population inoculated at large volume is shifted against the small-volume population. Although there is a large overlap, the shift is statistically significant ($p < 0.0001$) as determined by the robust two-sided Mann-Whitney U test (Mann and Whitney, 1947). The data is shown as box plots and annotated with the significance level in panel (b). Panel (c) shows the data as point clouds, with points labeled by initial patch radius and positions given by system size and critical point. The clear separation of populations in terms of number of nodes reflects the initial conditions shown in figure 6.1.1 with the coloring indicating that the number of nodes correlates well with the initial condition.

As discussed above, time is usually not a good measure for the structuring process. Nevertheless, the time at which percolation occurs will be of interest below, where characteristic time points of the growth process are investigated, indicating that there is significant correlation with time. Panel (d) of figure 6.2.7 therefore shows the histograms of percolation times for both experimental populations. Here, the picture is inverted. Although the histograms are quite broad with a substantial standard deviation, large volume experiments appear to percolate earlier, although the critical point is shifted to larger values as described above.

■ 6.3 Identifying phases of growth

In this chapter, the formation of the *P. polycephalum* network from fragments has been discussed first in terms of the time evolution of one example time series in section 6.1.2. Second, it has been established in section 6.2, that all networks undergo an identical topological evolution. This raises the question, whether the high similarity in the structuring process leads to correlations in the time-dependent behavior.

This is further motivated by findings reported in Fessel et al. (2015*b*), where it is observed that the slime mold only commences to forage when the network has surpassed a certain degree of structure, which is typically at or shortly after percolation. Moreover, the exploration phase of the slime mold occurs in several steps as discussed above, where it has been stated that a phase of rapid, expansion at almost constant velocity is followed by the onset of a pulsatile growth, where a global wave forms on the network topology, moving the concentration maximum of plasmodial matter to different regions and therefore leading to growth in those regions. This transition is likely identical to

the transition from a smooth, isotropic growth phase into a digitated, directed growth phase as described by Vogel et al. (2016) in a rather different setting. Lee et al. (2018) report on the exploration behavior of starved *P. polycephalum* fragments from initial conditions comparable to these employed in this work, except for the age of the culture, and typically higher coverage values. There, the authors find average-sized plasmodial units evacuating the site of inoculation prior to global percolation in a directed motion that is temporally correlated for all units inside a patch. Units are found to emerge and stop to form networks at similar times and after a similar distance has been covered.

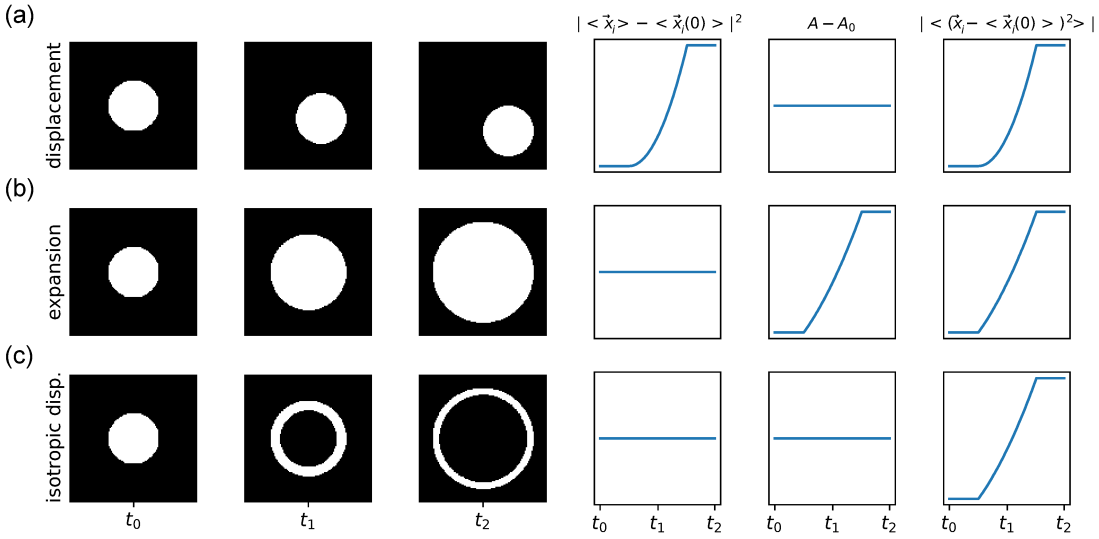


Figure 6.3.1: After structuring, *P. polycephalum* rarely resides near the site of inoculation. Exploration patterns that are commonly observed combine a number of the following idealized cases: (a) directed displacement, where the entire cell mass is shifted in one or more directions, (b) isotropic expansion, where a substantial portion of the network remains at the site of inoculation, or (c) isotropic displacement, where the area covered by the network remains constant. Inspection of test data suggests that the mean square displacement is the most robust measure for identifying exploration phases.

This raises the question whether similar correlations can be observed in the present data, or whether the behavior is less correlated in the absence of starvation-induced signals and possibly negative chemotaxis due to unfavorable conditions near the site of inoculation. By observation (see section 6.1.2 for details), the following phases can be clearly distinguished:

(i) **Aggregation.** Microplasmodia aggregate to small connected components. This process has a passive and an active component, as microplasmodia are in part driven by capillary forces due to drying of the plated droplet, but also exhibit self-propelled movement after. As the passive process completes significantly faster, the phase could likely be further subdivided. The end of the phase can be determined through the number of nodes, which is minimal after the aggregation process is

completed and before components start to extend. This is possible at constant plasmodial mass as nodes describing extended aggregates have a larger radius than those describing single microplasmodia.

(ii) Network formation. This phase starts when local aggregation is complete. Aggregates structure into local, network-like components and start to extend until they come into contact with other fragments. Eventually during this phase, the percolation transition will occur. The network continues to become more structured until growth fronts appear and extend across the patch border. The end of this phase is defined as the point in time where a substantial portion of the slime mold begins to evacuate the patch.

(iii) Rapid expansion. During this phase, one or more growth fronts formed by the slime mold move into the unexplored area around the site of inoculation. This happens in a directed, smooth motion. Size and number of growth fronts, as well as the directions in which they move vary significantly between time series. Similarly, the portion of plasmodial mass remaining in the patch differs, and it has been observed that the glucose concentration contained in the agar has an influence on this fraction. The end of this phase is defined as the point in time where growth fronts stop their directed motion and begin to grow in pulses.

(iv) Oscillating expansion. As soon as growth fronts stop directed motion, plasmodial height above the agar layer oscillates with a period between one and two hours. The oscillations constitute a global wave that drives plasmodial mass from one growth front to the next. If the slime mold has grown in an almost isotropic pattern where the entire perimeter is covered by growth fronts, circular motion of the wave is observed frequently. Growth occurs according to the phase velocity of the wave, and is localized to the regions where plasmodia are currently thickest (as indicated by darker gray values). During the observation period, this phase typically does not end. It constitutes the typical motion pattern of adult slime mold networks.

Modes of expansion In order to quantify the start and end times of the growth phases observed in individual experiments, two measures are used. First, the aggregation phase starts at the beginning of the experiment and terminates when the number of nodes is minimal. This requires determination of the minimum of the number of nodes, which is usually straightforward as it is supposed to be a global minimum. However, due to artifacts erroneous global minimums may occur in single frames. To overcome this, the minimum of a spline interpolation with a coarse support of the number of nodes is searched. The typical time evolution of the number of nodes is shown in figure 6.3.2. The time series giving rise to the data evaluated in figure 6.3.2 is shown in figure 6.3.4 below.

Determination of start and end points of the remaining phases is more problematic, as these are tied to the expansion of the network across the patch border. Expansion can occur in a variety of patterns. It is observed, that the entire network displaces in one single direction, leaving little to no mass behind in the patch. Similarly, it is observed that the slime mold expands almost isotropically while maintaining a network that covers the

entire area including the patch, or without leaving such a network behind and expanding as a ring-like structure. Moreover, these patterns are frequently observed to occur in superposition or only in a partial form. This leads to the need for a robust measure.

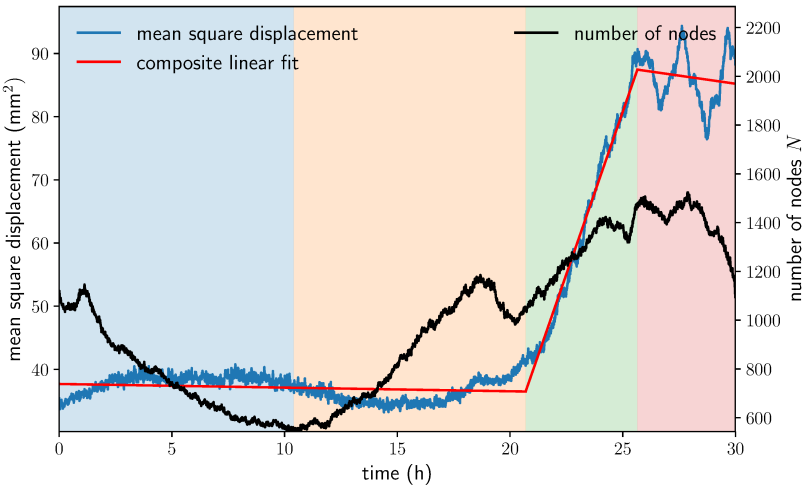


Figure 6.3.2: Inspection of the number of nodes and the mean-square displacement of newly-forming *P. polycephalum* networks enables distinction of growth phases.

The problem is illustrated using test data in figure 6.3.1. Displayed in the figure are three possible measures, i.e., displacement of the centroid of a structure, covered area, and the mean square displacement from the initial centroid, given as

$$R^2 = | < (\vec{x}_i(t) - < \vec{x_i}(0) >) > | \,, \tag{6.3.1}$$

here, $< \vec{x_i}(0) >$ is the centroid in the initial configuration and $\vec{x}_i(t)$ are the coordinates of all points in the structure at some point in time. The average runs over all these points. The first two measures are only sensitive to one of the processes illustrated in the figure. A displaced object does not necessarily change area when leaving the patch, and an expanding object changes area without being displaced. Formation of an expanding, ring-like or partially ring-like structure does neither imply a substantial change in area, nor is the centroid shifted. However, all three processes are correctly recognized by the mean square displacement. Another appropriate measure is given by the searched area as employed in Fessel et al. (2015*b*), which collects all locations visited by the slime mold at least once during its evolution. However, this measure requires processing of data from all states up to the current state, unlike the measures given above, which requires only the initial and current states to be processed.

Figure 6.3.2 shows an example trajectory of the mean square displacement obtained from one *P. polycephalum* network growth experiment, together with the evolution of the number of nodes. The phases outline above are clearly distinguishable by eye. First, the mean square displacement shows no substantial changes and the number of node drops to its global minimum, thus concluding the aggregation phase and initiating the

network growth phase in which percolation occurs. Upon completion, the network leaves the patch in order to expand and a steady increase is visible in the mean square displacement, which abruptly halts after approximately 26 h. At that point, an oscillatory pattern becomes visible as the slime mold enters the oscillatory phase.

The transitions between the phases are analyzed computationally by fitting with a piecewise linear function of the form

$$R^2 = \begin{cases} m_1(t-t_0) + R_0^2 & t_0 < t \leq t_2 \\ m_2(t-t_2) + m_1(t_2-t_0) + R_0^2 & t_2 < t \leq t_3 \\ m_3(t-t_3) + m_1(t_2-t_0) + m_2(t_3-t_2) + R_0^2 & t_3 < t \leq t_N \end{cases}, \tag{6.3.2}$$

where t_0 and t_N are fixed to the start and end of the time series. Fitting was performed using the basin-hopping global optimization method by Wales and Doye (1997), and although there are six free parameters, phases could be identified correctly with only few exceptions. As it is not expected that the mean square displacement changes substantially during the first phase, the aggregation and network formation phases were treated as a single phase by the fitting procedure. The essential fitting results are the transition times t_2 and t_3, which mark the transitions from network formation to expansion, and from expansion to oscillatory growth.

The results for all 522 growth experiments are summarized in figure 6.3.3 as histograms. All histograms are clearly peaked, and although the spread is substantial, the resulting structure is in agreement with the phenomenological description of phases given above. The aggregation phase on average lasts for 7.0 ± 2.4 h, network growth completes in 9.9 ± 4.4 h and expansion is completed after 7.4 ± 5.0 h. The comparison to percolation times given in figure 6.2.7 indicates that the distibutions shown there are peaked near the end of the network formation phase as expected.

	t_0		t_1		t_c		t_2		t_3		t_N	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
p_0	0.269	0.116	0.221	0.125	0.061	0.048	0.049	0.051	0.032	0.070	0.020	0.024
p_1	0.412	0.047	0.399	0.066	0.234	0.063	0.191	0.085	0.125	0.072	0.108	0.044
p_2	0.257	0.091	0.305	0.108	0.543	0.088	0.585	0.097	0.618	0.098	0.618	0.056
p_3	0.048	0.030	0.059	0.034	0.130	0.036	0.141	0.045	0.185	0.056	0.185	0.048
p_4	0.009	0.009	0.013	0.012	0.027	0.015	0.027	0.016	0.030	0.014	0.030	0.013
p_5	0.003	0.007	0.003	0.004	0.006	0.006	0.006	0.006	0.009	0.006	0.009	0.007
$< k >$	1.126	0.293	1.252	0.324	1.845	0.184	1.923	0.236	2.084	0.265	2.145	0.146

Table 6.3.1: Topological features of notable time points. *P. polycephalum* develops from fragments into a well-defined topological steady state, passing through a number of characteristic points in time. Data shown are average μ and standard deviation σ, collected from 522 experimental trajectories. Individual trajectories were averaged in 30 min windows around the respective points. Refer to figures 6.3.3 and 6.3.2 for a definition of the times t_0–t_3. t_c marks the percolation transition and t_N indicates the end of the observation period.

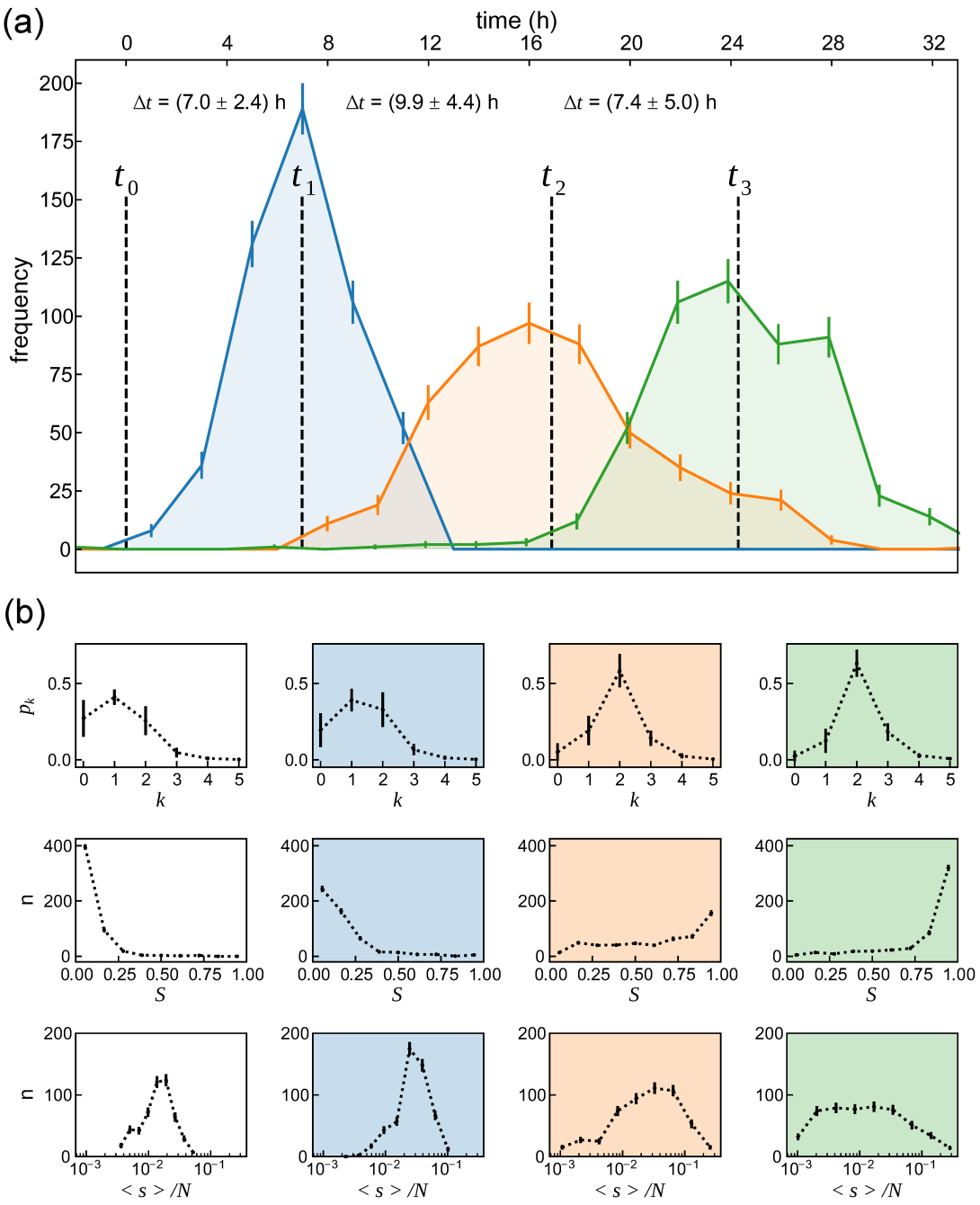


Figure 6.3.3: The dynamics of distinct growth experiments show similar phases of growth. Panel (a) highlights the dwell times of a forming network in either phase. (b) Graph measures at the transition times t_0–t_3. Refer to the text for further details.

The correlation between the critical point and the onset of the expansion phase illustrates the connection between topological state and the observed behavior. This is

detailed further in figure 6.3.3, where the degree distribution an histograms for the size of the giant and average non-giant components are shown at the onset of all phases, averaged over all data sets. Focusing first on the degree distribution, the evolution seen here is identical to the evolution of topology described in the previous section. The distribution shifts to the right as the network obtains a larger number of connections. The most substantial change is observed in the network formation phase, where percolation takes place. During the expansion phase, there is barely any change in the degree distribution, indicating that the system is at the topological steady state before commencing expansion. Numerical values characterizing the degree distribution at all transitions are given in table 6.3.1.

The giant component size shows a similar picture. First, the distribution leans to small giant component sizes but becomes more homogeneous as time progresses. After percolation and with the start of expansion, the distribution leans to larger component sizes. This process continues during the expansion phase, as it is possible that networks consisting of a sufficiently large component and a several non-giant components to expand while continuing the fusion process. Turning to the distribution of average non-giant components, it is noticed that it is rather sharply peaked at the start of the experiment when the system is dominated by solitary microplasmodia, and shifts to the right during the aggregation step. During network formation, the distribution broadens and ceases to shift. During expansion, the broadening continues until the component sizes are almost homogeneously distributed across the dataset. The probability to encounter large non-giant components decreases.

In order to conclude the study of growth phases, stages during the development of a *P. polycephalum* network and the associated graphs have been compiled into figure 6.3.4. Selected time points mark the start and end of all phases, the critical point, and the end of the observation. The experiment has been inoculated with a volume of 100 µl. The development observed in the figure agrees with the typical pattern described in this chapter. The first four images show the initial condition prior to aggregation, the end of the aggregation phase prior to network formation, the percolation transition at which a giant component is first detected and the start of outward growth, where a first growth front protrudes from the patch. Following a drastic change during the expansion phase, the network halts and starts to oscillate with a frequency of $1 - 2\,\mathrm{h}$. This is indicated by different gray level distributions on the otherwise similar configuration in the last two images. Data from the time series depicted here has been used in figure 6.3.2, where the oscillation is clearly visible in the mean square displacement.

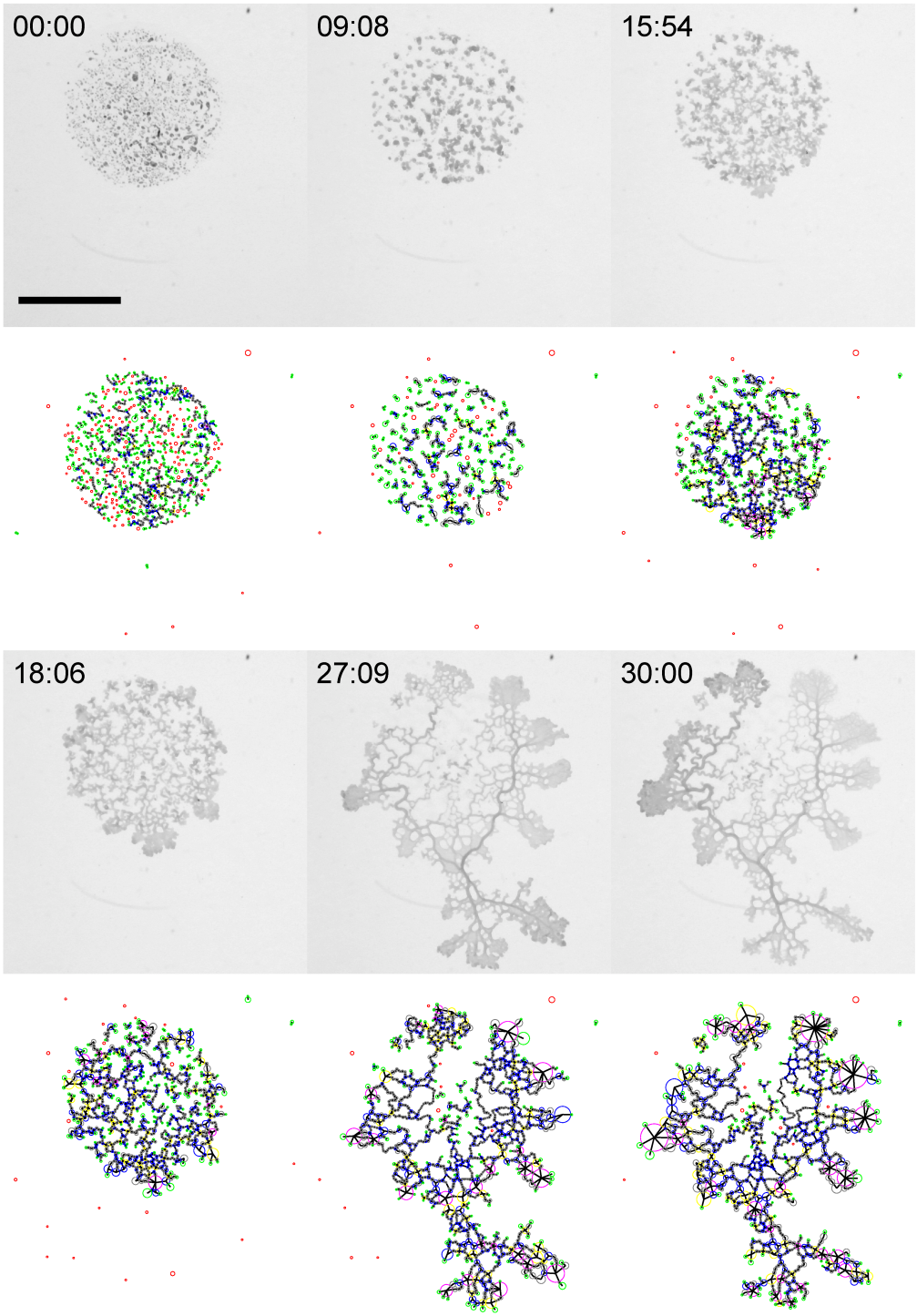


Figure 6.3.4: *P. polycephalum* network and graph representation at the time points detailed in table 6.3.1. Given in temporal order, these points are t_0: inoculation, t_1: completion of local fusion, t_c: percolation, t_2: start of expansion, t_3: start of oscillating growth, t_N: end of observation period. The scale bare represents 10 mm.

6.4 Tracking of topological changes

Section 5.3 introduces the concept that the topological progression of a graph evolving over time can be understood as a sequence of elementary topological changes. This concept is picked up in chapter 7 in order to devise a theoretical model of network formation consisting of a rate equation and a stochastic master equation. Simulations referred to earlier in this section are stochastic trajectories of the aforementioned models, which are realized by explicitly carrying out a succession of elementary topological modifications.

Background The idea behind this is based on the concept that complex networks of chemical reactions can be broken down into elementary steps in which no more than two reactants interact. The reasoning behind this is introduced in section 3.3.1. Here, this concept is extended to graph theory, making use of the discrete nature of changes in a graph topology, that are bound along the lines of either the addition and removal of edges or the addition and removal of nodes. Here, these changes are formulated as four elementary processes:

(i) Growth. One novel node is added to the network and connected to one existing node.

(ii) Retraction. One existing node with exactly one neighbor is removed from the graph.

(iii) Fusion. Two unconnected nodes of the graph are connected by an edge.

(iv) Separation. An edge connecting two nodes on the graph is removed.

The proposed set of processes constitutes a model that enables a system to restructure by adding or removing edges, and to grow or shrink by adding and removing single nodes. The way in which nodes are added or removed is motivated by the growth of the slime mold network, where the addition of a new node happens through the extension of a finger-like protrusion as detailed above. A closer look on the elementary processes seen as the basis of slime mold network formation is taken in section 7.2.

Omitting technical details here, the parameters describing the model are rate constants, similar to the reaction rate constants of chemical reactions (see section 3.3.1) that determine how often per unit time a selected process may occur, based on the availability of nodes required for the process.

Although there exist only four elementary processes, rate constants can be different for different node types as characterized by their degree. For instance, considering growth rates during the network formation phase of *P. polycephalum*, it much more frequently observed that growth occurs at small degree nodes (i.e., degree zero or one) rather than at nodes of larger degree that are already stably integrated into part of the network. However, there is still a crucial difference between the rate at which an event occurs and the rate constant describing it: an topological modification with a high rate constant may occur very infrequently, if the nodes required for that modification are only rarely encountered on the graph. The relation between rate and rate constant that is employed

in the present model is, again, borrowed from chemistry, see section 3.3.2. There, reactions (such as an radioactive decay) that only require one reagent to be present occur with a rate proportional to their availability, whereas a the rate of a reaction with two reactants is proportional to the availability of both, and further modified by the system size as reactants need not only be available, but are also required to meet in the system volume.

Considering again interactions between nodes in a network, this leads to the following dependence of a rate constant c on the reaction rate R, in the case of a single-node process

$$c_i = \frac{R_i}{N p_i} \tag{6.4.1}$$

where N represents the number of nodes in the graph and i indicates the degree of the acting node. p_i is, as above, the probability to select a node of degree i if randomly selecting from the graph. In the case of a process involving two nodes, the relation reads

$$c_{ij} = \frac{R_{ij}}{N p_i p_j} \ . \tag{6.4.2}$$

Further, section 5.3 reports on a collaborative project (Hillmich, 2018) as part of which a method has been devised to track the occurrence of elementary processes during time series of graphs as analyzed in this chapter. As a result, it has become possible to determine the topological modifications that have taken place between any two consecutive graphs representing states of a network.

Rate constants reflect the formation of *P. polycephalum* networks Figure 6.4.1 shows rate constants for all elementary processes that may occur in a graph with a negligible fraction of nodes that have degree six or higher. Rate constants were determined for each time series individually by counting elementary processes for each time step, converting into a rate by taking into account time step size and by subsequent application of the relations given above. The resulting data are time-resolved. In order to state the result more compactly, averages were computed within the phases defined in the previous section, and over all time series.

Results are shown in figure 6.4.1, where rate constants are displayed separately for all four elementary process classes as functions of the degrees of nodes required for the process in question. Processes have been sorted such that, in the first place, the sum of degrees increases, and, in the second place, by the smallest participating node degree.

For all process classes, trends are clearly visible. For instance, growth rate constants are largest for small and large nodes and decrease towards nodes of second degree. Rate constants associated with fusion decrease as degrees grow larger, whereas retraction and separation increase with the node degree. It needs to be emphasized again at this point that rate constants do not reflect how frequently a process is observed, but rather state the tendency for one set of constituent nodes to undergo a specific process. Similarly, a low rate constants indicates higher stability of a node type.

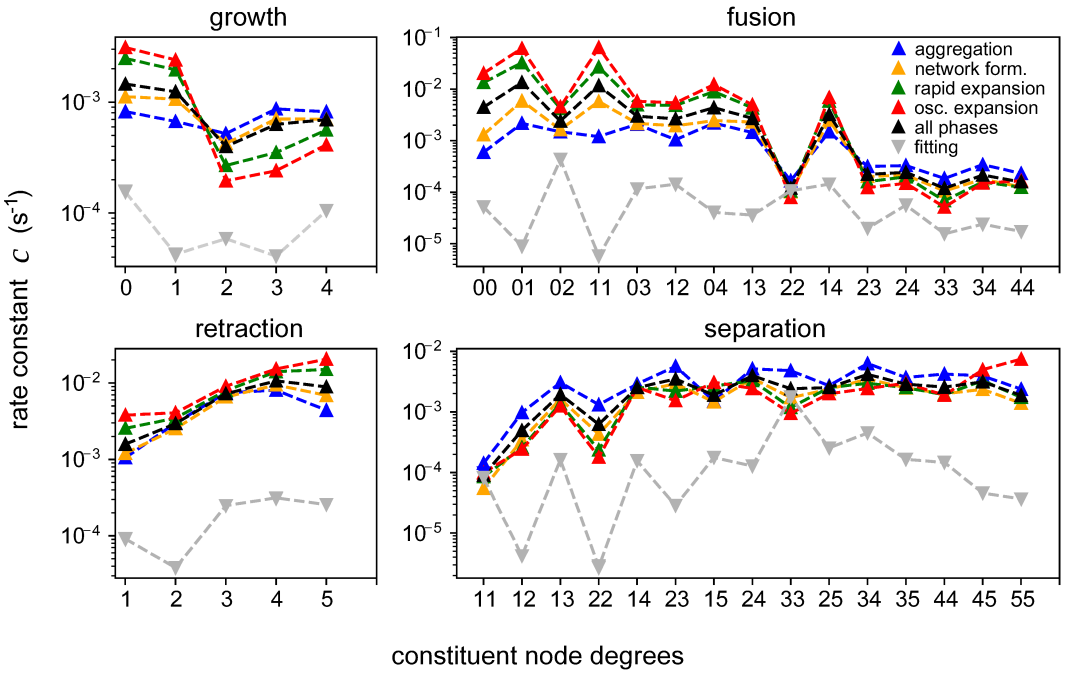


Figure 6.4.1: Rate constants characterizing the topological evolution of *P. polycephalum*, displayed against their constituent node degrees. Upward-pointing triangles indicate rate constants obtained by tracking of topological changes, downward-pointing triangles indicate rate constants obtained by fitting. For rate constants obtained by counting, a distinction is made between the phases of growth, giving first indication that the phases can be distinguished by their constants. In the case of fitting, fitting results for different n_k are shown, indicating the order of the fit does not substantially change the result.

The trends are in accordance with insights gained into the structure formation in *P. polycephalum* as discussed in this section. Focusing on growth, it is observed that rate constant values are large for nodes of degree zero and one, indicating the tendency of the slime mold to grow at finger-like tips, and for microplasmodia to break symmetry and elongate. The increase of the rate constants towards nodes of larger degrees stems from the expansion mechanism observed in growth fronts: spacious are described by large degree nodes that produce finger-like protrusions in a growth process. Nodes of second degree are the most stable and therefore dominant node species in the graph, and therefore their tendency to grow is smallest. The same is true for most events involving second degree nodes that are associated with other process classes.

Turning to fusion, nodes of small degree are more motile and generally predisposed to undergo fusion process when compared to large nodes that are already tightly embedded in the network. Retraction and separation as reverse processes to growth and fusion in principle show, as expected, reversed trends. Both classes have large values for large node degrees, indicating that these are typically rather unstable and exist only for short times

in growth front regions where one observes rapid remodeling of the network structure.

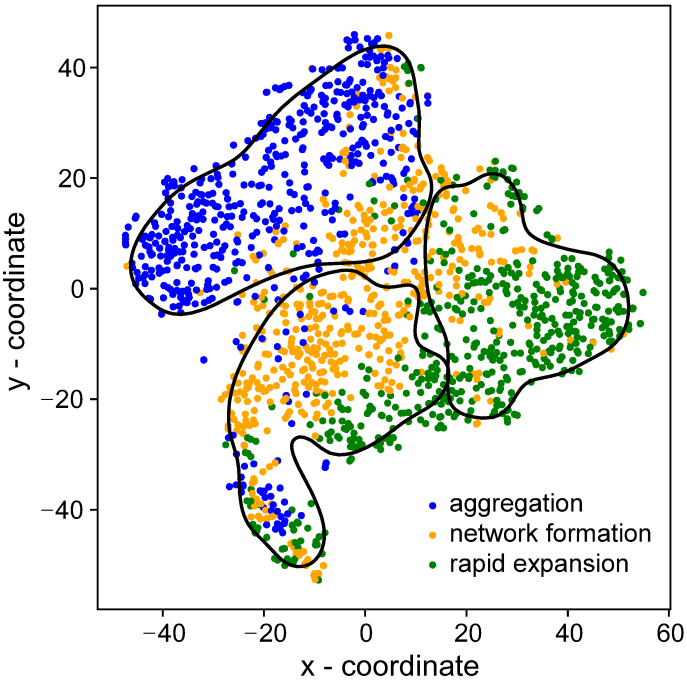


Figure 6.4.2: Clustering & multidimensional scaling analysis of rate constants. Each point of a color represents one of 522 data sets, characterized by the rate constants in one of the growth phases. Projected into a two-dimensional space that spatial positioning such that distance represents the similarity between data points, data obtained from the same growth phase cluster together. Cluster detection via the Gaussian mixture model works in principle, but fails at the cluster boundaries where there is significant overlap.

Growth phases can be distinguished through rate constants Rate constants systematically differ between growth phases, although the pattern may appear counter-intuitive at first. Focusing again on growth, it is observed that the rate constants characterizing small degree nodes are smallest in the aggregation phase and largest in the expansion phases. For large degree nodes, exactly the opposite is the case. This finding seems puzzling at first, as for example the aggregation phase is associated with the presence interaction of small degree nodes. However, the result simply means that if a rare large degree node exists in the aggregation phase, or similarly, if a rare small degree node exists in the expansion phase, it is typically not stable and expected to interact soon. In contrast, small degree nodes are available in abundance throughout the aggregation phase and allowed to interact in numerous different ways.

The difference between phases as reflected by the rate constants can be used for phase identification. Figure 6.4.2 shows a multidimensional scaling plot (Borg and Groenen, 2005) of all data sets, characterized by their rate constants in the first three phases,

i.e., aggregation, network formation and rapid expansion. Multidimensional scaling is a method of dimensionality reduction that attempts to position a number of high-dimensional measurements in a lower-dimensional space, such that the distance in that space approximates the dissimilarity between the measurements. Unless the selected case is particularly simple, the axes of the lower-dimensional space are, by themselves, meaningless. In the present case, measurements are phases of *P. polycephalum* network formation characterized by a set of rate constants, and the lower-dimensional space has been selected to have two dimensions.

Each data point in figure 6.4.2 represents one of the first three phases for each of the 522 data sets. The growth phase associated with each data point is indicated by color, in accordance with figures shown above (figures 6.3.2 and 6.3.3). The data points are divided into three clusters, which are clearly distinguishable by eye but have a significant overlap. The overlap has to be attributed to missdetection of phases, as this has been automated using the routine described in section 6.3, or to phases that are correctly distinguished by the method but share similar rate constants nevertheless. The latter is entirely possible, as there is significant variation in the observed growth modes and initial conditions. Furthermore, for the present visualization, no distinction has been made between glucose levels, or density, which are expected to influence the relative rate constant values. However, an analysis of this effect needs further data and therefore has not been performed here.

Multidimensional scaling is a versatile tool for assisting in the exploration of a data set, visually, or as a method of dimensionality reduction. However, if in the present case the data were unlabeled (i.e., if the association to a specific growth phase was unknown), the result above would not provide much insight. Machine learning provides numerous tools that can be employed in order to detect unknown structures in a data set. One such method is clustering (Duda et al., 2001), which has been briefly employed in section 6.1.2 to distinguish nodes based on their role in a graph structure. Here, clustering can be applied to the same data that has been used for the multidimensional scaling analysis to detect clusters of similar data with no labels.

Shown in figure 6.4.2 as black lines are approximate cluster boundaries, obtained after clustering of the data using the Gaussian mixture model (Everitt and Hand, 1981), which attempts to detect sub-populations of data in a data set that agree with a Gaussian probability distribution. In the present case, the number of clusters was selected to be three. As can be seen in the figure, the clusters approximate the structure observed in the multidimensional scaling plot but fail in the regions where there is significant overlap[2]. Although the accuracy of the method is lacking, the overlap of clusters detected here and the labeled data is an indication that it is in principle possible to distinguish between phases based on the set of rate constants. This would be possible, for instance, through supervised learning. As rate constants are measured in a time-resolved fashion, one possible application would be to assign phases to all graphs in a time series using a machine-learned algorithm trained on the present, labeled data set, rather than by using

[2]Note, that the clustering procedure has been applied to the raw data, not to the projected data retrieved from multidimensional scaling.

the fitting procedures described in section 6.3.

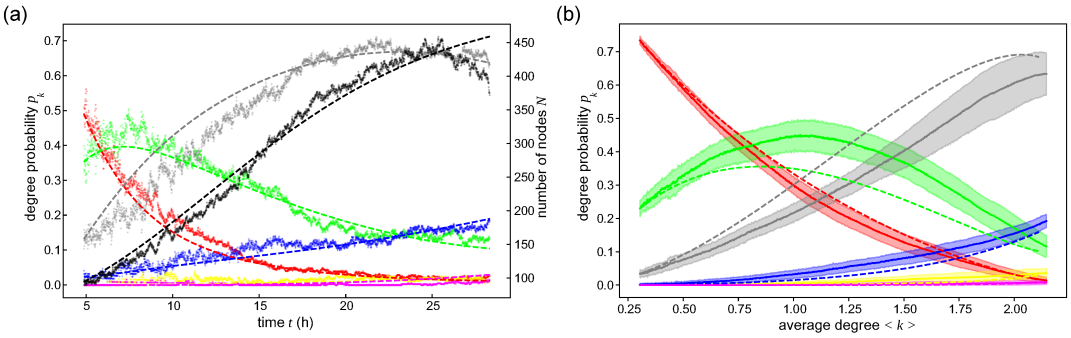


Figure 6.4.3: The rate equation model applied to experimental data. The rate equation model devised in chapter 7 is well suited to describe the dynamics of the *P. polycephalum* degree distribution. (a) Fit of the rate equation model (dashed lines) to the time series exemplified in this chapter. Quality of fit is high with an adjusted coefficient of determination of $\bar{R}^2 = 0.967$. (b) Rate equation model with parameters obtained by tracking topological modifications. Due to the effect of noise, the dynamics are not correctly reflected, but the topological evolution is reproduced with acceptable accuracy. Experimental data are mean ± standard deviation, obtained over all data sets. Rate constants are averages over all data sets and all phases.

Rate constant magnitudes are overestimated due to noise A second method capable for estimation of rate constants from the experimental data is through fitting of the rate equation model (equations 7.3.7 and 7.3.8) developed in chapter 7 by following the same premises stated at the outset of this section. A fitting routine, which simultaneously fits equation 7.3.7 to the time evolution of the degree distribution, and equation 7.3.8 to the evolving number of nodes. Although fitting results are typically excellent, this approach is problematic as the number of parameters is large. The number of rate constants required to describe processes up to some maximum degree is calculated in section 7.2, and is found there to increase almost with the square of the maximum degree. The fitting therefore does not necessarily reflect the correct rate constants, as different combinations of processes may lead to the same evolution.

More accurate approaches would be to either restrict the set of process to those that are most frequently observed in *P. polycephalum*, or to construct phenomenological relations between rate constants and the constituent degrees from the trends observed in figure 6.4.1. The former approach has been implemented successfully during early stages of this work as outlined in Fessel and Döbereiner (2016), but was later foregone in order to pursue the more general model presented in chapter 7, which has possible applications beyond slime mold research. The latter approach, i.e., relating rate constants to node degree through phenomenological relations will be pursued in the future.

The evolution of the degree distribution of the exemplary time series displayed in

figure 6.2.2 panel (b) is reproduced in figure 6.4.3 panel (b) in conjunction with the rate equation devised in chapter 7, with parameters obtained by fitting. The quality of fit is high by either visual inspection or through the coefficient of determination $\bar{R}^2 = 0.967$, which has been adjusted to consider the large number of parameters. In the shown case, fitting has been performed with all possible processes for nodes with a maximal degree of five enabled. Using a smaller maximal degree reduces the number of parameters. It has been confirmed, that the parameters shared between rate equation model fits for multiple maximal degrees are close to identical regardless of the maximal degree selected.

Fitting has been performed for all 522 data sets and averaged rate constant results are given in figure 6.4.1. Trends are roughly shared with rate constants obtained by tracking, but the magnitude of rate constants obtained by fitting is up to two orders of magnitude smaller.

The offset between the two measurements can be explained by findings reported in sections 7.4.2 and 7.5.2. Briefly, it is observed there that increasing rate constant magnitudes while maintaining a constant ratio between rate constants causes the dynamics of the system to be sped up while the stages through which the topology evolves remain identical. This is confirmed in figure 6.4.3 panel (b), where the topological evolution of all data sets is shown as a function of the the average degree, similar to figure 6.2.2 panel (b). The experimental data are complemented by a numerical realization of the rate equation with parameters set according to the rate constants obtained by tracking, as shown in figure 6.4.1. Rate constants were averaged over all data sets and over all phases.

Although there are deviations, the result reproduces qualitative aspects of the experimentally observed evolution remarkably well. However, as the time scale of the rate constants obtained by tracking is off, the evolution of the number of nodes is not correctly reproduced.

This immediately leads to the question why the rate constants obtained by tracking are off. The answer is partly given in section 7.4.2. Working with the example of a stochastic system that is returning to equilibrium, it is discussed there, that the system which evolves by stochastic transitions between microstates will return to equilibrium faster if at a high temperature, in contrast to a system at a low temperature. This can be understood by interpreting temperature as a measure for transitional noise: at higher temperature, the frequency with which stochastic transitions are activated is increased.

In principle, noise in the network formation process has three possible sources. First, errors in the graph abstraction procedure and second, errors related with dynamics on short time scales, such as the characteristic oscillation of *P. polycephalum*. These processes can be understood as an 'on/off' noise, where edges or nodes disappear momentarily, only to reappear shortly after. In terms of the two-dimensional structure formation, this noise is insignificant, as fluctuations only occur in place. However, by the tracking procedure and the non-local nature of the mean field model, each such fluctuation is recognized as a topological change that is not differentiable from those that contribute to structure formation. Therefore, with sufficient noise, rate constant magnitudes will

be significantly overestimated.

It should in principle be possible to correct for the described effect by treating the differences between processes and reversed processes. However, even in the described 'on/off' noise, it cannot be ensured that one erroneous process is locally corrected by the reversed process, since it is possible that there are larger cycles. These occur frequently in growth fronts if large nodes break down, for instance through erroneously detected hole formation. Furthermore, the algorithm presented by Hillmich (2018) (see also section 5.3) breaks down complex topological modifications in to sequences or cascades of events, making it difficult to make any statement on the reverse process that would balance such a cascade.

Chapter 7

A probabilistic model for topological dynamics

The present chapter incorporates concepts from chapter 3 for the modeling of stochastic dynamical processes in order to devise a model for network evolution that is based on a set of elementary processes modifying the topology of a network. In detail, a master equation describing the dynamics and fluctuations of the degree distribution and size of a network is formulated based on the chemical master equation (equation 3.3.2). Instead of treating the master equation analytically, its characteristics are studied in a twofold approach. First, a deterministic rate equation describing the mean-field behavior of the master equation is derived and analyzed. Second, exact stochastic trajectories of the master equation are simulated using the stochastic simulation algorithm (see section 3.4), a process which can be exploited to generate network topologies suitable for studying percolation.

The remainder of the chapter is organized as follows. First, in sections 7.1 and 7.2 the model devised in this chapter is motivated and the foundation is laid by presenting the set of elementary processes to which any topological modification, including those required for system growth, can be reduced. The elementary processes are formulated as chemical reactions and the concept of reaction propensities (see section 3.3.2) is invoked as a bridge to formulating the master and rate equations. The model is then introduced in section 7.3, where in section 7.3.1 the deterministic rate equation is derived based on the elementary processes, which is extended in section 7.3.2 to the master equation incorporating fluctuations of the system due to finite system size. As outlined above, instead of attempting to treat the master equation by analytical methods, in section 7.3.3 an extension of the stochastic simulation algorithm is presented, which enables simulation of network growth in accordance with the models devised in this chapter. The algorithm represents the growth of networks that are not embedded in space. First steps towards implementing planar network growth in the algorithm are then taken in section 7.3.4.

Section 7.4 collects analyses performed on the rate equation and simulation models. As the proposed models are characterized by a large number of parameters, a simplified model is proposed in section 7.4.1, which assigns the same rate constant to all processes of one type. Analyses of the model are presented in the following order. First, section

7.4.2 focuses on the topological dynamics of the rate equation model, commencing with an overview over behaviors that are observed in the simplified model before characterizing the steady state towards which the model evolves for large times. Finally, the time required to reach the steady state is investigated based on the parameters of the model, indicating that identical ratios of rate constants lead to the same steady state, whereas their magnitudes control the time required for the evolution. Second, the analysis is extended to stochastic simulations, demonstrating first the excellent agreement between simulations and the deterministic model in the planar and two-dimensional settings (section 7.5.1), before turning to percolation of the non-spatial model in section 7.5.2. Therein, after presenting and discussing qualitatively typical features of simulated percolation data, the influence of model parameters is tested. In particular, it is found that system growth leads to a shift of the critical point. The analysis is continued more systematically in section 7.6, where the location of the critical point and the critical exponents are determined via finite size scaling techniques (see section 2.1.3) upon variation of growth and the maximal node degree in the network. Although both parameters shift the critical point, no systematic influence on the critical exponents is found. Critical exponents indicate that the non-spatial model shares a universality class with mean-field percolation as is expected, for instance, in random graphs. Finally it is shown that critical exponents in the two-dimensional model agree well with standard values for two-dimensional percolation.

■ 7.1 Topological modifications & the relevance system growth

Every change in a network topology is a change in its set of nodes and edges. For instance, the addition of new edges between random nodes in the random graph model is arguably the most basic topological modification a network may undergo. Modifying the set of edges based on selected rules has proven to be a powerful tool for understanding how and why certain network topologies emerge: rewiring a small number of edges in a regular graph gives rise to the small world effect (Watts and Strogatz, 1998), preferentially attaching new edges to nodes that have already accumulated a number of connections characteristically leads to the emergence of scale-free networks (Barabási and Albert, 1999), and selecting the pair of nodes that will be connected by an edge based on global rather than local properties can produce behaviors such as explosive percolation (D'Souza and Nagler, 2015). However, no matter how the rules governing the placement of edges are selected, the size of the network, defined as the number of nodes, will remain unchanged. Processes such as the examples given above produce topologies that closely resemble the developed state of real-world networks, but these generation processes rarely correspond to the actual time evolution of a system. The assumption that all nodes were, in some form, in place for the entire duration of the generation process in many cases corresponds to a wrong initial condition.

The time evolution of real networks needs to incorporate growth, i.e., modifications to the set of nodes. As a consequence, the set of possible modifications is bound to be much more diverse – and more specific. A new web page that is added to the world wide web represents the introduction of a new node into the system. This process does not only modify the set of nodes, as the web page is typically only useful if it is connected

to one or more other websites that are already part of the web. Indeed, the described process requires the introduction of a number of edges in addition to the novel node. While this process and the rules describing its occurrence have been found to provide a good description for the growth of the world wide web (Barabási and Albert, 1999), it is easy to find a system that appears to evolve by a different rule. This immediately raises the question whether such an observation implies that systems are entirely different, i.e., they require individual, separate treatments to describe their time evolution, or whether a common framework can be found[1].

The principle underlying the models for network growth that are developed in this chapter presents an attempt of solving this problem. Similar to the concept of reducing complicated chemical reactions to a set of elementary reactions (see section 3.3.1), it is assumed that more complex changes to a network topology do not occur in an infinitesimally short time, but instead can be broken down into sequentially occurring elementary modifications. Based on the reasoning above, one will conjecture that the most basic mechanistic changes to a topology are the addition and removal of edges, and the addition and removal of solitary nodes, i.e., nodes that do not have any connections. Revisiting the world wide web as an example, this picture proves to be correct. The process of introducing a new web page that will have a certain number of edges can be easily broken down: first, a web page without any connections is introduced. Then, one after another, the edges are added until the desired number is reached.

From a mechanistic point of view, this picture is most certainly correct, but in a physical sense it is problematic. In the probabilistic framework employed below (see also section 3.3.2), a process in which a node appears out of thin air is bound to occur at a rate that is independent of the state of the system. Such a limitation to the influx of nodes into the open system also poses a limitation to the growth of the network, which may not be intended. This scenario can be described as passive growth and is similar to diffusion-limited aggregation. The significance of the opposing process, and the distinction to passive growth can be best highlighted using an simplified example: the size of a colony of yeast cells in starvation can be increased by two possible methods. First, corresponding to passive growth, new cells may be introduced into the system that will eventually aggregate to the colony and increase its size, regardless of the number of cells that is already present. Instead, one may also present nutrients to the colony, enabling each individual cell to divide at an average rate that is limited by the supply of nutrients. If nutrients are presented in abundance, the number of cell divisions per unit time will be proportional to the number of individuals, and the colony grows exponentially until size or density restrictions imposed by other mechanisms dominate. This active growth process, which may be controlled via the time constant for the average proliferation frequency shares similarities with reaction-limited aggregation and appears to be much more applicable to a self-organized network growth process.

Either system is an open system where particle number is not conserved, similar

[1]It shall not be claimed here that it is always feasible to forgo a highly specialized yet simple description for one that is more general in its essentials, albeit more complicated to handle in the long run. Nevertheless, knowing that very different systems share a common underlying framework may pave the way for discovering universal features that emerge from the properties of the underlying description.

to a thermodynamic grand-canonical ensemble, and growth in either system requires some type of flow into the system. Controlling this flow equals controlling the growth of the system in either case. The active case is given preference here, as the mechanism by which the system grows should be a part of the underlying framework and not an external parameter. The most simple, yet physically meaningful growth process that takes into account the reasoning given above is the addition of a new edge-node complex to a node that is already present in the network. In summary, the set of elementary processes modifying a network topology are, as above, the addition and removal of edges and the addition and removal of a node-edge complex. The former processes modifies only the set of edges whereas the latter processes modifies the sets of nodes and edges alike.

■ 7.2 Elementary topological modifications

In the previous section, four processes have been established that present a basis into which any topological modification can be decomposed. Two 'forward' processes increase connectivity or size of the network, whereas the inverse 'backward' processes decrease connectivity by removing edges or decrease the size of the network by removing nodes. Briefly, the four processes are

(i) Growth. A complex consisting of one degree one node and one edge is added to an existing node of arbitrary degree.

(ii) Retraction. A complex consisting of one degree one node and one edge that was attached to a node of arbitrary degree larger than zero is removed from the graph.

(iii) Fusion. A pair of nodes of arbitrary degrees is connected by adding one edge.

(iv) Separation. One edge that connects a pair of nodes with arbitrary degrees larger than zero is removed from the graph.

In the following, theses classes of the elementary processes (i-iv) will be indicated by subscripting the letters G, R, F, S.

Each process modifies the set of edges. This requires that during each occurrence of one of the elementary processes, at least one node changes its degree. Since in each process exactly one edge is added or removed, the change of individual node degrees is limited to $\Delta k = \pm 1$. Furthermore, the elementary processes have been set up such that at most two distinct nodes are involved. Given these two observations, it is possible to formulate the four processes similar to chemical reactions, where nodes of a given degree are viewed as a reactant species. All process can thus be given by a chemical equation of the form

$$(k_1,\, k_2) \to (k_1',\, k_2') \;. \tag{7.2.1}$$

Here, k_i and k_i', $i = 1, 2$ are the degrees of one interacting node before and after the process takes place. k_i, k_i' are node degrees defined as usual as the number of edges that is connected to a given node, with the exception that nodes that do not exist prior to, or after the elementary process are denoted by $\varnothing$. As the nodes of the degrees k_i are

required as constituents for a process to occur, they will be referred to as constituent degrees. With these definitions, the classes of elementary processes can be written as follows. A growth process following rule (i) is given as

$$(\varnothing,\, k) \to (1,\, k+1) \tag{7.2.2}$$

and the corresponding inverse (ii) follows

$$(1,\, k) \to (\varnothing,\, k-1) \quad \text{where} \quad k \geq 1\ . \tag{7.2.3}$$

Processes involving fusion or separation involve two arbitrary node degrees k_1 and k_2. Setting up the chemical equations is straightforward. First, one has

$$(k_1,\, k_2) \to (k_1+1,\, k_2+1) \tag{7.2.4}$$

in the case (iii) describing fusion and

$$(k_1,\, k_2) \to (k_1-1,\, k_2-1) \quad \text{where} \quad k_1 \geq 1 \quad \text{and} \quad k_2 \geq 1 \tag{7.2.5}$$

in case (iv) that denotes the separation of a connected pair of nodes. It is necessary to emphasize that the class of a process and the exact constituents unambiguously define a process and its outcome. Furthermore, in the set of processes that involve two constituent node types, $k_1 = k_2$ is a valid interaction. No notational distinction is made between $(k_1,\, k_2)$ and $(k_2,\, k_1)$. None of the reactions are catalytic in the sense that an educt remains unchanged over the course of a reaction, but for appropriate choices of k_1 and k_2 the chemical equations resemble the structure of autokatalytic reactions when particles are indistinguishable. For example, a fusion process will not alter the number of k_2 nodes in a graph if $k_2 = k_1 + 1$. This type of interaction can be intuitively understood as a treadmill.

As given in equations 7.2.2-7.2.5, only four distinct processes are required for manipulation of a network topology, and they can be given in a form similar to chemical reactions. As described in section 3.3.2, each chemical reaction is characterized by a constant that holds information on how frequently a reaction is bound to occur when a given amount of educts is present. Viewing nodes of different degrees as distinctly different chemical species raises the question whether the constants characterizing each process should be dependent on the type of node, or type of nodes required by the process as educts. This question cannot be answered in a general fashion and will be of interest later on. At this point, with a description in mind that is as flexible and general as possible, it will be assumed that each type of interaction of nodes is characterized by one constant. This implies that the maximum number of possible reactions depends on the number of different node degrees that can be occupied in a system. In the following chapters, graphs will be considered, in which the maximal degree is bounded. The number of node degrees that can be assumed in such a system is denoted by n_k. As this includes nodes of degree zero, the maximal degree is given by $n_k - 1$.

Comparable considerations lead to the model published by Ghoshal et al. (2013). Therein, the authors propose a model for network growth based on similar processes. These include the addition and removal of links in the network, two processes identical

to fusion and separation, and the addition or deletion of nodes. Although the latter processes appear similar to growth and retraction, they are different in the sense that the nodes that are added or removed in the process come with an arbitrary number of novel edges, or are removed regardless of the number of edges connected to them. This makes the model more appropriate to scenarios, where, for instance, web pages are added to the world wide web and immediately obtain a large number of connections, a process that in the model developed here is broken down into a number of elementary steps occurring in swift succession. Ghoshal et al. (2013) do not define single rate constants for elementary processes, but instead view processes as driven by a superposition of random and preferential attachment. Therefore, the model is characterized by a smaller number of constants, at a lesser degree of flexibility compared to the model proposed here.

The model by Ghoshal et al. (2013) has proven to be well suited for studying the influence of the processes above, and the influence of random or preferential attachment on the degree distribution. The authors determine on the parameter phase space the regions where the degree distribution is either exponential, or dominated by a power law behavior. The latter is analyzed analytically and by numerical methods for selected parameters. However, the authors do not study the formation of components, which is possible here due to the close relation of the model to the network simulation algorithm (algorithm 7.3.1). Furthermore, the modular construction of the present model enables studying attachment processes that are neither preferential attachment, nor entirely random, and the set of processes can be easily limited so that only selected processes drive the formation of the network.
The models constructed in the subsequent sections are deterministic or stochastic realizations of systems in which processes of the types given above can occur, and as such the parameters characterizing these models are the constants associated with these processes. The number of parameters characterizing a model is crucial to the analysis, and hence it is of advantage to determine the maximum number of interaction types for a given number of degrees $n_k \geq 2$. If there are less than two degrees allowed in a system, no interactions can take place as all interactions have been defined such that they change the degree of the interacting nodes. With the limitation in mind that node degrees are bounded in the range $0, \ldots, n_k - 1$, which cannot be exceeded as result of a process, one finds

$$N_{\mathrm{G}} = \sum_{k=0}^{n_k-2} 1 = n_k - 1 \qquad N_{\mathrm{R}} = \sum_{k=1}^{n_k-1} 1 = n_k - 1 \tag{7.2.6}$$

in the cases of growth and retraction, and

$$N_{\mathrm{F}} = \sum_{k=0}^{n_k-2} k + 1 = \frac{n_k(n_k-1)}{2} \qquad N_{\mathrm{S}} = \sum_{k=1}^{n_k-1} k = \frac{n_k(n_k-1)}{2} \tag{7.2.7}$$

for fusion and separation by considering simple combinatorial arguments. The total number of processes when taking all types into account is thus given by

$$N_{\mathrm{GRFS}} = (n_k - 1)(n_k + 2) \ . \tag{7.2.8}$$

■ 7.3 Modeling topological dynamics

In the previous section, it has been implicitly made use of the concept that the numbers of nodes that share a degree can be regarded as populations that interact through and are modified by the elementary processes as defined in equations 7.2.2-7.2.5. Here, this concept is further elaborated in order to describe the dynamics of a system of such populations in two ways. A deterministic rate equation is derived in section 7.3.1, and a master equation modeling the time evolution of the probability distribution is given in section 7.3.2 as a stochastic approach based on chapter 3.

The models developed here are based on two assumptions. First, it is assumed that in a first approximation all populations are homogeneously distributed in space, justifying a mean-field formulation. Second, following up on the first assumption, the probability or rate of occurrence of a process is assumed to depend only on the availability of its constituents at a given time, rather than on their configuration in space, local effects, or other mechanisms. More specifically, the occurrence of a growth process is assumed to be entirely stochastic (governed by a rate constant), and the occurrence of a process involving two constituents is assumed to be governed by a rate constant and the probability or rate at which the constituents are expected to meet in a the system[2].

The similarity of these conditions to the perfectly mixed condition in a system of reacting chemicals (see section 3.3.2), and the formulation of the elementary processes as chemical reactions highlights that the tools laid out in section 3.3 can be employed for modeling the time evolution of a network modified by the elementary processes.

■ 7.3.1 Rate equation

Let a system of large system size consist of a number of populations N_k of nodes of degree k, which are modified by a set of elementary processes that are consistent with those given in equations 7.2.2-7.2.5. Based on the assumptions given in the previous section, the average rate at which each process occurs is regulated by one or two populations of nodes, depending on the constituents required by the specific process. This leads to the following two rate equations:

$$\left(\frac{\partial N_k}{\partial t}\right)^{(r)} = \sum_i c^{(r)} l_{ki}^{(r)} N_i \tag{7.3.1}$$

describes the rate of change of the k'th population due to the growth process r, which is dependent on a single population N_i. The process is defined by the rate constant $c^{(r)}$ and the interaction matrix $l^{(r)}$, whose matrix elements $l_{ki}^{(r)}$ are identical to the change of the k'th population when the process r occurs exactly once. Similarly, retraction, fusion and separation depend on two constituent populations. In these cases, the rate of change

[2]This assumption can be alleviated later on, when it is discussed that the rate constants employed in the formulation of the model can be modified in order to be able to incorporate rules other than those given in this section.

is given as

$$\left(\frac{\partial N_k}{\partial t}\right)^{(r)} = \sum_{ij} \frac{c^{(r)}}{\Omega} q^{(r)}_{kij} N_i N_j \ , \tag{7.3.2}$$

where $q^{(r)}$ is defined as $l^{(r)}$, but with an additional dimension. Ω is the size of the system and corresponds to the fact that the probability of an encounter between two nodes is smaller in a larger system, as briefly discussed in section 3.3.2.

Knowing the rate of change with respect to a single process r, it is straightforward to obtain the rate of change when the system is modified by a number of processes. This is calculated by summing equations 7.3.1 and 7.3.2 over the set of possible processes:

$$\begin{aligned}
\frac{\partial N_k}{\partial t} &= \sum_r \left(\frac{\partial N_k}{\partial t}\right)^{(r)} \\
&= \sum_r \sum_i c^{(r)} l^{(r)}_{ki} N_i + \frac{1}{\Omega} \sum_r \sum_{ij} c^{(r)} q^{(r)}_{kij} N_i N_j \\
&= \sum_i \underbrace{\sum_r c^{(r)} l^{(r)}_{ki}}_{L_{ki}} N_i + \frac{1}{\Omega} \sum_{ij} \underbrace{\sum_r c^{(r)} q^{(r)}_{kij}}_{Q_{ki}} N_i N_j
\end{aligned} \tag{7.3.3}$$

Equation 7.3.3 is a set of k first order ordinary differential equations with a quadratic non-linear term. With the interaction matrices L_{ki} and Q_{kij} left unspecified, it can be regarded as a second dynamical system, e.g., a generalized predator prey system. In some case, for instance when devising a numerical implementation, it is helpful to recognize that 7.3.3 can be written in a matrix-vector form that describes the evolution of the population vector $\vec{N}$:

$$\frac{\partial \vec{N}}{\partial t} = L\vec{N} + \frac{\vec{N}^{\mathrm{T}} Q \vec{N}}{\Omega} \tag{7.3.4}$$

Rate equation for the degree distribution Equation 7.3.3 describes the evolution of node populations. Aiming for a framework that is more consistent with the pivotal role of the degree distribution in complex networks, it appears sensible to formulate the system of rate equations accordingly. This is accomplished by treating the differentials of the normalization condition

$$\frac{\partial N_k}{\partial t} = N \frac{\partial p_k}{\partial t} + p_k \frac{\partial N}{\partial t} \ , \tag{7.3.5}$$

and by introducing the total number of nodes as a variable, which obeys

$$\frac{\partial N}{\partial t} = \sum_k \frac{\partial N_k}{\partial t} \ . \tag{7.3.6}$$

The resulting system is equivalent to the rate equations describing the population sizes, but now consists of $k+1$ equations, given by the time evolution of the degree distribution

$$\frac{\partial p_k}{\partial t} = -p_k \frac{1}{N} \frac{\partial N}{\partial t} + \sum_i L_{ki} p_i + \frac{N}{\Omega} \sum_{ij} Q_{kij} p_i p_j \ , \tag{7.3.7}$$

where the normalization condition is reflected by the first term, and an equation describing the time evolution of the total number of nodes:

$$\frac{\partial N}{\partial t} = N \sum_{ki} L_{ki} p_i + \frac{N^2}{\Omega} \sum_{kij} Q_{kij} p_i p_j \tag{7.3.8}$$

In a graph that is not explicitly embedded in space, the system size Ω cannot be given in terms of spatial dimensions. Instead, it is sensible to select a system size expressed by the number of nodes in the network. If the graph is not expected to change size excessively, $\Omega = N_0$, N_0 being the initial number of nodes, appears to be a reasonable choice. The behavior of the system changes when Ω is defined to be the (dynamically changing) number of nodes, i.e., $\Omega = N$. In the latter case, the system reads

$$\frac{\partial p_k}{\partial t} = -p_k \frac{1}{N} \frac{\partial N}{\partial t} + \sum_{i} L_{ki} p_i + \sum_{ij} Q_{kij} p_i p_j \tag{7.3.9}$$

$$\frac{\partial N}{\partial t} = N \left(\sum_{ki} L_{ki} p_i + \sum_{kij} Q_{kij} p_i p_j \right) \ . \tag{7.3.10}$$

Briefly examining the clear structure of the system reveals that behavior for large times is dominated by the term $\Theta_k = \sum_i L_{ki} p_i + \sum_{ij} Q_{kij} p_i p_j$: omitting the trivial case that $N = 0$, a constant number of nodes $\dot{N} = 0$ requires $\sum_k \Theta_k = 0$. Furthermore, if the system neither changes size nor restructures its topology, i.e., if $\dot{N} = \dot{p}_k = 0$, $\Theta_k = 0$ is required. If the system has reached a topological steady state defined by $\dot{p}_k = 0$, the size of the system can only change if all populations grow as $\dot{N}_k = N\Theta_k$ and the degree distribution is required to satisfy $p_k = \Theta_k / \sum_k \Theta_k$ or $\Theta_k = 0$, the latter corresponding to $\dot{N} = 0$.

Generation rules for interaction matrices Up this point, no further mention has been made on the structure of the interaction matrices l_{ki} and q_{kij}, other than that they describe how the population N_k changes if the corresponding process is realized exactly once[3]. To set up the combined interaction matrices

$$L_{ki} = \sum_r c^{(r)} l_{ki}^{(r)} \quad \text{and} \quad Q_{kij} = \sum_r c^{(r)} q_{kij}^{(r)} \ . \tag{7.3.11}$$

in a general fashion, it is necessary to find a rule that describes the generation of l_{ki} and q_{kij} for an intended process. In the previous section, it has been emphasized that the class of a process, i.e., growth, retraction, fusion or separation, and the constituent node degrees unambiguously define a process and its outcome. It is thus possible to search for the matrices l and q as functionals of the constituent node degrees, with a structure defined by the class of the process as given in equations 7.2.2-7.2.5. Indeed, such generation rules can be expressed in terms of the Kronecker delta, which is defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} \ . \tag{7.3.12}$$

[3]It is worth noting, that these matrices are closely related to the stochiometric matrices given in section 3.3.3, although this similarity is not employed here.

In the case of growth, given by equation 7.2.2, the interaction matrix for a process that involves a constituent node of degree k is given as

$$l_{ij}^{(\mathrm{G})}(k) = (\delta_{i1} + \delta_{ik+1} - \delta_{ik})\delta_{jk} \ . \tag{7.3.13}$$

The structure can be understood as follows. The three bracketed terms correspond to, in order, the addition of a degree one node to the graph, the increase of the population N_{k+1} and the decrease of the population N_k, both because one node of degree k obtained a new edge in the process. The multiplicative term ensures that the process is only applied to the populations if the second index corresponds to the degree k of the constituent node. By very similar arguments, expressions can be given for the other classes. Starting with retraction, where one constituent node has a fixed degree $k = 1$, one finds

$$q_{ijh}^{(\mathrm{R})}(k) = \frac{1}{2}(\delta_{ik-1} - \delta_{ik} - \delta_{i1})(\delta_{j1}\delta_{hk} + \delta_{jk}\delta_{h1}) \ . \tag{7.3.14}$$

The numerical factor corresponds to the symmetry of the matrix. In the cases of fusion and separation, there are two constituent degrees k_1 and k_2. For fusion, the generation rule reads

$$q_{ijh}^{(\mathrm{F})}(k_1, k_2) = \frac{1}{2}(\delta_{ik_1+1} + \delta_{ik_2+1} - \delta_{ik_1} - \delta_{ik_2})(\delta_{jk_1}\delta_{hk_2} + \delta_{jk_2}\delta_{hk_1}) \ , \tag{7.3.15}$$

and for the inverse process, separation:

$$q_{ijh}^{(\mathrm{S})}(k_1, k_2) = \frac{1}{2}(\delta_{ik_1-1} + \delta_{ik_2-1} - \delta_{ik_1} - \delta_{ik_2})(\delta_{jk_1}\delta_{hk_2} + \delta_{jk_2}\delta_{hk_1}) \tag{7.3.16}$$

Implications and generalization In summary, the rate equation model, given by equations 7.3.7 and 7.3.8, combined with the rules for generation of interaction matrices, equations 7.3.13-7.3.16 presents a general framework for the average time evolution of a randomly grown graph, based on only for reversible elementary mechanisms. The evolution is uniquely defined by an initial condition and the set of transition rate constants. Given appropriate choices for the set of parameters, the model is conjectured to be capable of producing a wide range of degree distributions. As discussed in section 7.2 the number of transition rates grows faster than the square number of possible node degrees. This large number of parameters highlights the flexibility of the model, but also presents its major weakness, as information on the rates at which all processes occur needs to be available. However, this drawback can be easily alleviated whenever it is reasonable to assume that the value of a rate constant associated with a process is a function of the constituent node degrees. For instance, in the most simple scenario one might select $c_{\mathrm{F}}(k_1, k_2) = \text{const.}$, i.e., any pair of nodes will be connected with equal probability. If all other constants are equal to zero, the model reduces to the case of an Erdös-Rényi random graph. Similarly, a model in which $c_{\mathrm{G}} \propto k$ is selected and all other constants are set to zero, exhibits a form of preferential attachment and will result in a scale-free degree distribution. These approaches can be easily combined. A composite model in which both previous examples are realized grows via a preferential attachment mechanism while being randomly connected. In any case, whenever it is possible to find functions that relate the constants of the four individual process classes to their constituent nodes, the number of parameters required to define the resulting model is

greatly reduced. Furthermore, it is imaginable to integrate dependencies on the node populations (other than those suggested by the kinetic picture outlined in section 3.3.2, which are already implemented in the model) via the rate constants that in this case can no longer be treated as constants. Similarly, via the rate constants, a coupling to a co-evolving secondary quantity can be implemented. For example, the growth rate can be tied to a time-dependent nutrient concentration, which in turn can be influenced by the total number of nodes.

In section 7.4, a special choice for the rate constants will be consulted that lends itself to the discussion of the general features of the rate equation model. In detail, a single constant will be assigned to each process class, and accordingly, all processes belonging to that class are described by the same rate constant, whereas the rate at which they are bound to occur still follows the kinetic picture outlined in section 3.3.2.

■ 7.3.2 Master equation

The model derived in section 7.3.1 described the mean-field behavior of an evolving network. However, as outlined in chapter 3, is only accurate in the thermodynamic limit. When interested in systems of finite size, it is necessary to consider stochastic methods such as the master equation. The networks considered in the scope of this work are exclusively of small size, and finite-size effects obscure the percolation transition the evolution from a fragmented into a connected state. To study these effects and characterize the transition more closely, a master equation for the topological dynamics is formulated based on the chemical master equation

Repeating very briefly the details given in section 3.3, the chemical master equation given in equation 3.3.2, as

$$\frac{\partial P(\vec{N},t)}{\partial t} = \sum_{r=1}^{M} \Big[a_r(\vec{N}-\vec{\nu}_r)P(\vec{N}-\vec{\nu}_r,t) - a_r(\vec{N})P(\vec{N},t) \Big] \tag{7.3.17}$$

describes the time evolution of the occupation probability $P(\vec{N},t)$ of all states characterized by the vector of population numbers $\vec{N}$. In terms of the previous section, each entry in this vector corresponds to the number of nodes with a certain degree. The master equation is a gain loss equation, that bases the gain (or loss) in occupation probability of one state on all possible transitions to (or from) that state. These transitions, enumerated by r, are each described by the stochiometric vector $\vec{\nu}_r$ that details the changes in all populations due to one single occurrence of a transition r (see equation 3.3.1). The transitions that are allowed within a network evolving in accordance with the model presented in this chapter have been given in section 7.2. Below, the transition vectors for all process classes will be given in a notation based on the Kronecker delta, similar to the transition matrices employed in the previous section. In the case of growth, a node of degree k is removed in exchange for one node of degree one that is connected to a node of degree $k+1$, giving rise to the population vector

$$\nu_i^{(\mathrm{G})}(k) = \delta_{i1} + \delta_{ik+1} - \delta_{ik} \ . \tag{7.3.18}$$

Similar considerations enable formulating the transition vectors for the cases of retraction, where a degree one node is removed alongside a node of degree k, leading to the emergence of a node of degree $k-1$:

$$\nu_i^{(\mathrm{R})}(k) = \delta_{ik-1} - \delta ik - \delta_{i1} \tag{7.3.19}$$

In the cases of fusion and separation, edges are added or removed between two nodes of degree k_1 and k_2, which are subsequently transformed into nodes with degrees that are either larger or smaller by one than the original nodes. The associated transition vectors read

$$\nu_i^{(\mathrm{F})}(k_1, k_2) = \delta_{ik_1+1} + \delta_{ik_2+1} - \delta_{ik_1} - \delta_{ik_2} \tag{7.3.20}$$

in the case of fusion and

$$\nu_i^{(\mathrm{S})}(k_1, k_2) = \delta_{ik_1-1} + \delta_{ik_2-1} - \delta_{ik_1} - \delta_{ik_2} \tag{7.3.21}$$

in the case of separation. Comparing the transition vectors to the generation rules for transition matrices as given in equations 7.3.13-7.3.16 shows that the non-zero columns of these matrices are indeed identical to the transition vectors.

Finally, the quantities $a_r(\vec{N})$ and $a_r(\vec{N} - \vec{\nu}_r)$ have been described in section 3.3.2 as the propensities of a process, detailing the tendency to leave from, or respectively, arrive at a certain state via the transition r. The propensities can be given as

$$a_i(\vec{N}) = c_i^{(\mathrm{G})} N_i \tag{7.3.22}$$

in the case of growth where there is just one constituent node and

$$a_i(\vec{N}) = \frac{c_{ij}^{(\mathrm{R,S,F})}}{\Omega} N_i N_j \tag{7.3.23}$$

for all other cases. In all cases, c^{r} is a rate constant that is unique to an elementary process given by the process class and the constituent node degrees.

The advantage of having expressed the transition vectors in terms of the Kronecker delta is that the sum over the set of transitions in equation 3.3.2 can be transformed into sums over the constituent node degrees. Making use of this and substituting the propensity functions into equation 3.3.2, the master equation describing the system can be written down:

$$\begin{aligned}
\frac{\mathrm{d}P(\vec{N})}{\mathrm{d}t} = & \sum_{i=0}^{n_k-2} c_i^{(\mathrm{G})} \left[(N_i - \nu_i^{(\mathrm{G})}) P(\vec{N} - \vec{\nu}^{(\mathrm{G})}) - N_i P(\vec{N}) \right] \\
& + \sum_{i=1}^{n_k-1} \sum_{j=1}^{n_k-1} \frac{c_{ij}^{(\mathrm{R})}}{\Omega} \left[(N_i - \nu_i^{(\mathrm{R})})(N_j - \nu_j^{(\mathrm{R})}) P(\vec{N} - \vec{\nu}^{(\mathrm{R})}) - N_i N_j P(\vec{N}) \right] \\
& + \sum_{i=0}^{n_k-2} \sum_{j=0}^{n_k-2} \frac{c_{ij}^{(\mathrm{F})}}{\Omega} \left[(N_i - \nu_i^{(\mathrm{F})})(N_j - \nu_j^{(\mathrm{F})}) P(\vec{N} - \vec{\nu}^{(\mathrm{F})}) - N_i N_j P(\vec{N}) \right] \\
& + \sum_{i=1}^{n_k-1} \sum_{j=1}^{n_k-1} \frac{c_{ij}^{(\mathrm{S})}}{\Omega} \left[(N_i - \nu_i^{(\mathrm{S})})(N_j - \nu_j^{(\mathrm{S})}) P(\vec{N} - \vec{\nu}^{(\mathrm{S})}) - N_i N_j P(\vec{N}) \right]
\end{aligned} \tag{7.3.24}$$

Although it is in some cases possible to treat master equations by means of analytical methods, such as the generating function formalism outlined in section 2.3.3, this is not attempted here. Instead, the dynamics of the system are simulated by a modified stochastic simulation algorithm.

7.3.3 The stochastic network simulation algorithm

The stochastic simulation algorithm detailed in section 3.4 (Gillespie, 2007) is well suited to simulate the dynamics of the node populations $\vec{N}$, as it generates exact stochastic trajectories of the master equation 7.3.24. However, in the scope of this work that focuses on network structure, this would provide only a minor advantage over the deterministic model given in section 7.3.1, as it treats sets of nodes as populations and thus is not suited to produce detailed configurations.

In contrast, graphs are designed to represent, in addition to the populations of nodes, the detailed connectivity between individual nodes. A graph structure thus has to consist of an enumerated set of nodes V and a set of edges E that specifies the pairs of individual nodes possessing an edge. There exist various representations of graphs, the most well know of those likely being the adjacency matrix A_{ij}, whose elements are equal to unity if two nodes $i \neq j$ are connected, and zero otherwise. Regardless of which graph representation is chosen, the elementary processes discussed in the previous sections are represented therein as topological changes.

The key feature characterizing the stochastic simulation algorithm that will be used here is that each elementary process contributing to the stochastic time evolution of a system is selected randomly obeying a distribution that represents the stochastic features of the system in a physically correct fashion, before it is carried out explicitly through modification of the population numbers. Here, the first step is adopted from the stochastic simulation algorithm, whereas the second step, i.e., the modification of population numbers is replaced by a modification of graph structure. The algorithm, which is described more closely below, is shown in pseudo-code as algorithm 7.3.1.

In detail, the algorithm is initialized with data very similar to that required for the stochastic simulation algorithm: the initial state of the system is given by an initial time and by initial population numbers, and a set of processes is passed to the algorithm, characterized by rate constants and a library detailing class and constituent nodes of the process. The first central difference to the stochastic simulation algorithm is that prior to the first time step an initial graph $G = (V, E)$ is created that has a degree distribution specified by the population numbers. This is facilitated by the configuration model (Molloy and Reed, 1995), which enables creation of a random graph with a prescribed degree distribution.

The time steps computed by the algorithm are structured quite similar to the stochastic simulation algorithm: propensities are calculated according to the current node populations, which can be readily obtained from evaluation of the graph G. Based on the propensities and two uniform random numbers, the step size τ is calculated and the next process r is selected, based on equations 3.4.8 and 3.4.10. At this point, in the stochastic

simulation algorithm, populations and time are updated before moving on to the subsequent time step. Operating on graphs, however, updating the system in the present case is somewhat more technical. First, the selected process has to be examined, i.e., its class and constituent node degrees are determined. Then, based on the class of the process, one of four distinct routes is taken.

In the case of growth or fusion, one, or respectively, two nodes are selected uniformly at random from the correct populations. In the case of fusion, care is taken that the two nodes are neither identical nor connect by a direct edge. The graph is then updated by adding a new edge, which in the case of growth leads to a newly introduced node, or connects the pair of selected node in the case of fusion. The scenario, in which fusion is selected and no pair of nodes exists that is allowed to fuse due to the restrictions given above, is conjectured to be extremely rare as it can only occur if the system size is very small (i.e., on the order of 10^1).

When a process is selected that is of either of the classes retraction or separation, the situation is more problematic and highlights immediate limitations of the mean-field assumption with regard to the approach presented here. In these cases, not nodes, but rather connected binary structures have to be selected. For illustration, the following example can be considered: in a graph that has a significant number of degree one nodes, a process that removes the connection between two such nodes will be selected with a non-negligible frequency. However, a pair of two degree one nodes that are selected uniformly at random is unlikely to posses an edge, and thus may not be separable. In principle, the propensity function as defined in section 3.3.2 that bases the probability of a process on the availability of the constituents in the full system cannot be expected to be entirely accurate in this case. A more precise approach for a future iteration of the theory would base the propensities of retraction and separation processes on the random selection of an edge that leads to appropriate nodes. In the present work, an empirical approach is utilized that ensures agreement between simulation and theory with few exceptions as discussed below. Instead of selecting at random two nodes from the constituent populations, one constituent population is selected and evaluated for nodes that posses a neighbor within the second constituent population. From the set of nodes obtained in this fashion, one node and the appropriate neighbor are selected randomly, and the graph G is then modified as prescribed by the process class. If no such pair is found in the graph, the time step is repeated without updating the graph.

It needs to be emphasized that the described problem is a problem of the mean-field theory, rather than of the simulation algorithm presented here. The mean-field assumption is bound to lead to an overestimation of the frequency of processes that require a specific topological configuration of nodes. In principle, this is not different from the simplifications imposed by a mean-field assumption with regard to the influence of spatial heterogeneity, and can be accounted for by adjusting the value of associated rate constants. However, as the described effect exclusive modifies retraction and separation, it is noted separately.

The aforementioned procedures were designed such that the occurrence of processes

is in agreement with the deterministic model. Exceptions are only observed when either retraction or separation are dominant mechanisms, or if the system size is extremely small. In these situations it is possible that time steps frequently need to be repeated because no nodes are found in appropriate configurations. In this case, slight deviations between the stochastic network simulation algorithm and the stochastic simulation algorithm and, respectively, the deterministic solutions are observed, as will be discussed in section 7.5.

Returning to the structure of the stochastic network simulation algorithm, further differences towards the stochastic simulation algorithm are small. Instead of updating only time and populations, also the graph structure is updated according to the selected process before the next time step is attempted. Finally, after completion of each time step, various graph measures including component sizes are computed and stored. In theory, every measure of interest for non-spatial, undirected graphs can be computed at this point, including clustering measures or shortest path lengths (consult da F. Costa et al. (2007) for a comprehensive review of graph measures). Similarly, graphs can be stored for subsequent analysis. However, for practical considerations the number of computations to be performed and the size of outputs is generally kept as small as possible. Nevertheless, this final notion underlines the central advantage of network simulations over the stochastic simulation algorithm in the system at hand: structural features of graphs produced by the model can be efficiently analyzed, as will be demonstrated in the following section.

Algorithm 7.3.1: Stochastic network simulation algorithm

Data: set of processes R, rate constants $\{c_r \,|\, r \in R\}$,
initial populations $\vec{N}_0$, start/end time $(t_0,\ t_N)$
Result: time steps t, node populations over time $\vec{N}(t)$,
graph measures over time $(S(t),\ s(t),\ \chi(t),\ \ldots)$

```
begin
    G = (V, E) ⟵ configuration model (N⃗_0)
    t ⟵ t_0
    while t ≤ t_N do
        {a_r | r ∈ R} ⟵ calculate propensities (V, {c_r})
        {n_1, n_2 | n_i ∈ [0, 1]} ⟵ random numbers ()
        τ ⟵ calculate step size (n_1, {a_r})
        r ⟵ select process (n_2, {a_r})

        (class, k_1, k_2) ⟵ look up process (r)
        switch class do
            case growth do
                i ⟵ random k = k_1 node (V)
                j ⟵ G: new node ()
                G: new edge (i, j)
            case retraction do
                (i, j) ⟵ search k = k_1 node with k = 1 neighbor (V, E)
                if i ≠ ∅ then
                    G: remove node (j)
                else
                    retry time step
            case fusion do
                (i, j) ⟵ random k = k_1 and k = k_2 nodes (V)
                if i ≠ j and E(i, j) = ∅ then
                    G: new edge (i, j)
                else
                    retry time step
            case separation do
                (i, j) ⟵ search k = k_1 node with k = k_2 neighbor (V, E)
                if i ≠ ∅ then
                    G: remove edge (i, j)
                else
                    retry time step

        t ⟵ t + τ
        (S, s, χ, ...) ⟵ calculate graph measures (G)
        store t, N⃗, S, s, χ, ...
```

7.3.4 Towards a planar, stochastic network growth model

The simulation algorithm described in the previous section enables the construction of graphs that evolve in accordance with the rate and master equations devised in this chapter. The average dynamics of these systems are fully defined by the set of initial conditions and by the set of constants governing the rate at which nodes interact with one another. However, the graphs created by the stochastic simulation algorithm do not have a fixed spatial dimension. Assuming that all edges between nodes have unit length, high-dimensional spaces are required if a spatial embedding is searched for a highly structured graph. More intuitively stated, in the mean-field picture pursued in this chapter, interactions are possible between any pair of nodes - a property which is impossible in a graph that is embedded in space. In order to connect distant nodes in a spatial network, it is either necessary to introduce non-local edges, or to increase the dimension of the embedding space.

Thus, in a spatial graph that is in accordance with the graph model proposed in chapter 5, interactions can only occur locally between nodes and their immediate spatial neighbors. As it is expected that this has a substantial influence on an evolving network, first steps are taken here to devise a two-dimensional implementation of the stochastic growth model. As briefly discussed in chapter 5, such a two-dimensional model is not only useful to study the influence of dimensionality on stochastic network formation, but can also serve as a discretization of a growing spatial structure, where dynamical processes such as the regulation of a pressure or the exchange of fluid flow are computed on nodes that interact with their connected neighbors. These processes on the other hand can be coupled to the topological evolution by locally altering the stochasticity of topological modifications, and therefore a suitable model may serve as a bridge between stochastic dynamics and self-organizing network growth.

In order to fully define a simulation model in two dimensions, rate constants and initial conditions as employed in the previous sections are not sufficient. Various spatial aspects need to be considered, such as node and edge dimensions, growth angles or threshold and movement distances. The most substantial change, however, is the way in which fusion works. In the mean-field picture, any two nodes can in principle be connected if the appropriate process is selected. In two dimensions, as outlined above, this is in many cases impossible. Instead, nodes undergo fusion if they are in close proximity, making fusion a passive process in reaction to growth or movement of nodes.

In principle, the algorithm implementing the two-dimensional simulation is identical in structure to algorithm 7.3.1, in the sense that the time step and class of the next process are selected according to the stochastic simulation algorithm (section 3.4). The most important changes affect the graph structure, which now assigns a set of coordinates and a (unit) radius to all nodes, and to the way in which processes are executed once selected.

Furthermore, the initial configuration is also set up differently and no longer utilizes the configuration model. In order to mimic the situation observed in *P. polycephalum* growth experiments (see section 6.1.2), an initial patch area is defined via an initial patch radius R_0. Subsequently, nodes N_i are randomly placed within the patch such that the

disks defined by their radii r_i do not overlap. In greater detail, nodes are placed sequentially, and after each node that is placed, distances to all other nodes are calculated. If an overlap is detected in this way, the step is repeated. The entire procedure is repeated until an area is covered according to the initial coverage c_0. This process has been studied in the literature under the name 'random sequential adsorption' (Feder, 1980). Following the placement of initial nodes, non-overlapping nodes whose centers are less than a fusion threshold distance d_F apart are connected by an edge.

Following successful construction of the initial configuration, the first time step is attempted in the usual way, by calculating propensities based on rate constants. At this point, the only difference is that fusion no longer has a rate constant and therefore cannot be selected as the next process. Indeed, retraction and separation function identical to the non-dimensional case. Thus, the only difference lies within the growth process.

It needs to be emphasized here, that the selected method presents one possible attempt at establishing the growth process in two dimensions. It is not intended to replicate growth as observed in the *P. polycephalum* network exactly. The selected method builds on an exact geometric construction, which is then distorted by a noise term. The construction functions as follows. A node is selected at random from the population of nodes that can constitute the growth process. For the selected node, all connected neighbors are discovered an the centroid of the set is determined. Next, a straight line is constructed through that centroid, and through the center of the primary node. A novel node is than placed on that line, such that its position with reference to the primary node opposes the centroid of the neighboring nodes. The distance to the primary node is selected such that the nodes are in contact but do not overlap, and in order to mimic the randomness during growth in *P. polycephalum*, the axial position with reference to the primary node is distorted by randomly selecting from a Gaussian distribution centered at the exact constructed location. The width of the distribution is a parameter of the model.

The procedure described above is only carried out if the overlap of the novel node with any other node in the graph is small. If this is not the case, a different primary node is selected from the same population and the procedure is repeated. If placed successfully, the novel node is connected by an edge with all nodes that are are close than d_F. If the primary node has no neighbors, growth is carried out in a random direction. If growth leads to fusion one or more nodes that are not identical to the primary node, the time step size is increased accordingly.

It has been selected not to display the detailed algorithm here, as it is in a preliminary form and in principle identical to algorithm 7.3.1 with changes as outlined above. Furthermore, no systematic analysis of simulated two-dimensional network growth is carried out in the scope of this work. Instead, exemplary data are shown in section 7.5.1. Differences in percolation the percolation properties of the two-dimensional model in comparison to the non-spatial model are addressed in section 7.6.

7.4 Analysis of the model

The present section aims to analyze the models devised in this chapter in a reduced but flexible framework that will be introduced subsequently. The analysis of the model is separated into the same two parts into which the model can be broken down. First, the numerical solutions of the deterministic rate equation are analyzed with regard to its general dynamics and features, and with regard to the degree distributions and system sizes that are approached for large times. Second, stochastic simulations of the master equation are considered. This analysis encompasses the influence of stochasticity on the dynamics of the model, and the influence of finite size effects. In particular, in contrast to the traditional stochastic simulation algorithm, the modified stochastic simulation algorithm (section 7.3.3) enables the simulation of network growth, so that structural quantities like component sizes can be studied as a result of the underlying model. In the light of percolation (section 2.1 and chapter 2), the influence of finite size effects on the size of the giant component and on average component sizes are investigated by finite size scaling, leading to the determination of critical exponents. Furthermore, the influence of system growth on the critical exponents is examined.

7.4.1 The GRFS model

The model for network growth presented in this chapter provides a flexible framework that can be adjusted to describe a variety of processes. However, this flexibility comes at the cost of a large number of rate constants that need to be specified making the general analysis of the model difficult. In order to investigate the features of the model in a reduced framework that retains central characteristics of the model without sacrificing too much of its flexibility, the 'GRFS' model is proposed as a first simplification. Instead of assigning one unique constant to each process, all processes of one class are assigned the same constant. Namely, all growth processes are assigned the growth constant G, whereas R characterizes retraction. Similarly, F and S belong to the process classes of fusion and separation. Paraphrasing, when all constituent node degrees are equally available, all processes of a class may occur with equal probability. It is the major drawback of this simplification that it cannot be expected to appropriately described the dynamics of scale-free networks or other network types that base the probability of an elementary process on the constituent node degree rather than solely on the availability of the constituents. However, the growth processes of interest to this work are expected to be more closely related to random graphs than to scale-free networks: node degrees are exclusively small, and thus any dependence on the degrees themselves is expected to be less significant due to the relative homogeneity of the degree distribution.

7.4.2 Analysis of the deterministic solution

Being a set of ordinary differential equations, the rate equation model can be efficiently analyzed through explicit numerical integration employing the fifth order Runge-Kutta method (Dormand and Prince, 1980). A first glance at the behavior of the model can be obtained from numerical realizations for selected combinations of rate constants as displayed in figure 7.4.1. Starting with a set of unconnected nodes and omitting cases where the system is driven only by separation and retraction, the following behaviors

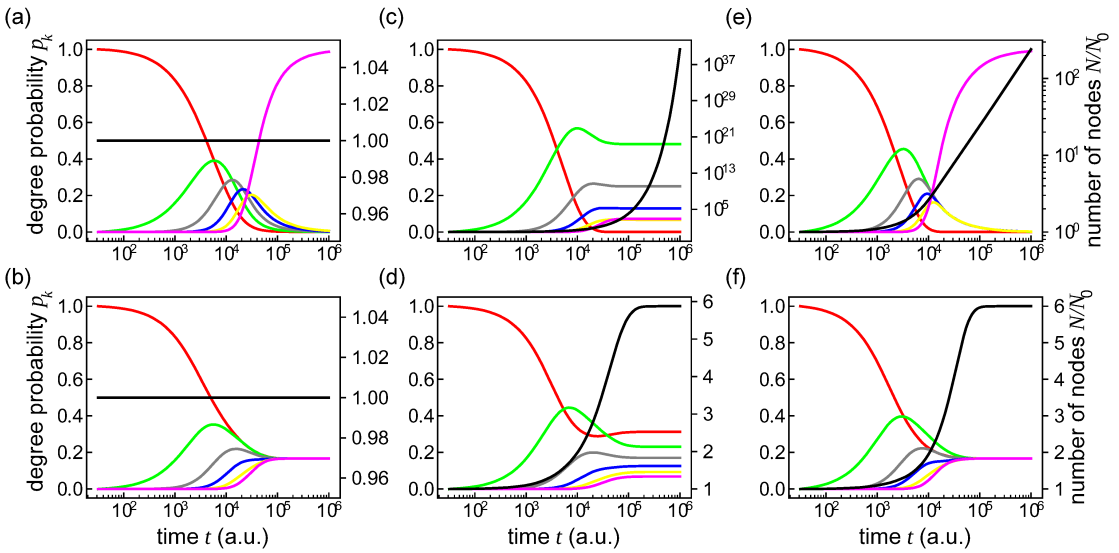


Figure 7.4.1: Numerical solutions of the rate equation model for selected rate constant values (zero unless noted otherwise) and $n_k = 6$. As initial condition, N_0 nodes of degree zero are present. Black lines correspond to number of nodes, all other lines indicate one population of nodes in accordance with the color code presented in figure 5.1.2. In all cases, the degree distribution becomes stationary for large times. (a) $F \neq 0$. If the system evolves only by fusion, the number of nodes remains unchanged and the graph evolves towards a regular graph with degree $n_k - 1$. (b) $F = S \neq 0$. When separation is enabled in addition to fusion, the system does not evolve towards a regular graph. Possible degree distributions depend on the ratio F/S and are shown in figure 7.4.4. In the selected case, node populations equilibrate as fusion and separation occur at identical rates. (c) $G \neq 0$. Each initial zero node evolves into a tree. The number of nodes grows exponentially. (d) $G = R \neq 0$. When retraction is present, the number of nodes evolves into a stationary state dependent on the ratios G/R and F/S as detailed in figure 7.4.3. (e) $G = F \neq 0$. With growth and fusion present, the system evolves towards a growing regular graph. (f) $G = R = F = S \neq 0$. With all processes enabled, degree distribution and number of nodes become stationary for large times. Details of the stationary state agree with panels (b) and (d).

can be identified.

Phenotypes of the GRFS model A system evolving exclusively by fusion (see figure 7.4.1 panel (a)) is identical to the Erdös-Rényi random graph model unless there is a limiting degree $n_k - 1$. If such a limit exists, the system initially resembles the Erdös-Rényi case but evolves towards a regular graph with degree $n_k - 1$. The degree distribution is initially peaked at $k = 0$ and is shifted to larger degrees due to fusion processes. Similar to a random graph, the degree distribution is comparable to the binomial distribution unless a large number of nodes has obtained the limiting degree. The

similarity to a binomial distribution under the described circumstances, especially for a small number of completed process steps is shared by all cases that involve fusion. The limiting degree can be interpreted as an absorbing state in the evolution of the degree of single nodes. Graphs of this type have been studied by (Ben-Naim and Krapivsky, 2011).

As soon as retraction is enabled, there always remains a fraction of nodes with a degree smaller than the maximal degree. In the case shown in figure 7.4.1 panel (b), the rate constants of fusion and retraction are identical, and in the steady state all node are present with equal probability. From a physics point of view, this case is interesting for two reasons: first, although the rate constants are identical, the system undergoes a topological evolution. Rather than being driven by external control, this process can be understood as a consequence of thermodynamic principles. The system tends towards thermal equilibrium, assuming the macrostate (characterized by the degree probabilities p_k) that maximizes the number of microstates (characterized by all possible configurations of nodes that lead to a given degree distribution). When the rate constants are identical, it is conjectured that at this state, all types of nodes are present with equal probability. Indeed, in the present case it is implied that at equilibrium all processes occur with equal probability, allowing characterization of this state as a detailed balance at which each process is balanced by its reverse process.

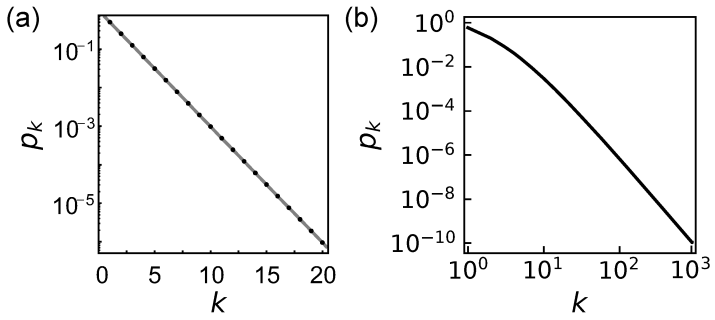


Figure 7.4.2: Degree distribution of trees produced by the rate equation model. If only growth is present in the rate equation model, the resulting structure is bound to be tree-like. (a) In the GRFS model, the degree distribution of the resulting tree is exponential and well-described by equation 7.4.1 (solid line). Black dots are numerical data. (b) In contrast to the GRFS model, growth following a preferential attachment rule leads to a scale-free degree distribution (numerical data obtained using the rule $c^{(\mathrm{G})}(k) = Gk$).

In systems that involve growth, the number of nodes does not remain constant unless the system is at a stationary point. If a system evolves exclusively by growth as shown in figure 7.4.1 panel (c), the resulting structure is bound to be a set of N_0 trees. For large n_k, the system is at a steady state when the degree distribution approaches

$$p_k = \begin{cases} 0 & k = 0 \\ (1/2)^k & 1 \leq k < n_k - 1 \end{cases} \tag{7.4.1}$$

indicating that there exist twice as many nodes of degree k than there are of degree $k+1$. This distribution follows from the argument that a node of degree k has to undergo k equally probable growth events whereas a node of degree $k+1$ has to undergo $k+1$ such events. Nodes that reach the maximal degree do not influence nodes of smaller degree, as they do not participate in any process. However, they have an influence on the degree distribution through normalization. For sufficiently large n_k, their fraction can be approximated as

$$p_{n_k-1} = \sum_{i=n_k-1}^{\infty} (1/2)^i = 4 \cdot (1/2)^{n_k} \quad . \tag{7.4.2}$$

At the steady state, trees continue to grow exponentially while preserving the degree distribution detailed above. The degree distribution detailed in equation 7.4.1 and numerical data are shown in figure 7.4.2 panel (a). For comparison, and in order to highlight the limitations of the GRFS model, the scale-free degree distribution of a tree produced by the rate equation using the preferential attachment rule $c^{(\mathrm{G})}(k) = Gk$ is shown in panel (b). Following this rule, the rate constant of each growth process is based on the degree of the constituent node rather than being constant as assumed in the GRFS model. This seemingly subtle change to a basic process drastically changes the degree distribution.

When retraction is enabled in addition to growth, the picture changes to a degree as shown in figure 7.4.1 panel (d). Although structures are still tree-like (as it is impossible for them to fuse) with an exponential degree distribution that is cut off at $n_k - 1$. However, unlike the previous case the system does not grow exponentially once the steady state is reached. Instead, the number of nodes assumes a maximal value. Notably, this balance of growth and fusion is not limited to the case where the rate constants are identical. This effect is further analyzed below. In particular, figure 7.4.3 details the number of nodes at the steady state for several combinations of rate constants.

Figure 7.4.1 panels (e) and (f) combine the selections made in panels (a) and (c), or panels (b) and (d), respectively. Combining growth with fusion, panel (e) documents the evolution of an almost regular graph in a system that grows linearly for large times. The linear growth may appear counter-intuitive in the first place, but has to be attributed to the vanishing fraction of nodes that has not reached the maximal degree. As long as such nodes exist, nodes of degree one are produced at some rate. In the present example, this rate is identical to the fusion rate. Hence, there is a constant production of nodes of small degree that balances the immobilization of nodes at the maximal degree. Nevertheless, at large the steady state assumed by the degree distribution is identical to the scenario that involves only fusion, as shown in panel (a). Similarly, in panel (f) where all rate constants have identical values, the dynamics of the node degrees appear to be governed by fusion and separation, whereas the dynamics and steady state of the system size agree with the case where only growth and retraction are present.

In summary, the rate equation model is capable of producing a variety of unique behaviors even with strong restrictions in place for the rate constants. The behavior for small times appears to be quite similar in all cases due to the initial condition, and due to the fact that no process allows degree modifications other than $\Delta k = \pm 1$ for

individual nodes, i.e., an initial population of zero nodes has to develop into one nodes before it can reach higher degrees, and so forth. Due to this, if fusion is present and if the system is far from the limiting degree, the degree distribution resembles a binomial distribution as observed in evolving random graphs. In contrast, the behavior for large times is distinctly different based on the choice of constants. Regardless of the choice of constants, the system evolves towards steady state characterized by a degree distribution and the number of nodes that can either stabilize at some value or settle into a state of linear or exponential growth. Although not detailed above, it is similarly possible that the number of nodes in a network decreases as will be discussed below, or even vanishes if $G = 0$ and $R > 0$. It is worth emphasizing that if $R > 0$, i.e., if there is retraction present in the system, the number of nodes always approaches a stationary point. Similarly, the degree distribution only stabilizes at the maximal node degree if $S = 0$, i.e., if separation is not partaking in the evolution of the system. In the following, the analysis focuses on the large time behavior in the case that all constants are larger than zero but do not necessarily have identical values.

Number of nodes at the steady state The first important point to notice about the steady state of the GRFS model is that the number of nodes, as well as the degree distribution at the steady state depend only on the ratios F/S and G/R rather than on the constants themselves. Figure 7.4.3 displays the number of nodes N_S at the steady state obtained through numerical simulation as described above, in relation to the initial number of nodes N_0 for numerous combinations of the aforementioned ratios. As suggested by figure 7.4.1, it is expected that the ratio of growth to retraction has a dominating effect on the number of nodes. Indeed, as can be gleaned from panels (b-d), there is a linear dependence on G/R for different values of F/S. However, the slope characterizing this relation is found to depend on F/S. Furthermore, for different numbers of node degrees n_k the numerical data are found to reside on different parallel lines if $F/S \geq 1$. It is intuitive that for systems that are dominated by separation there is no dependence on this cutoff parameters as node degrees are bound to be exclusively small.

The described dependencies can be clarified by inspection of the number of nodes as a function of F/S for one single value of G/R as displayed in figure 7.4.1, panel (a). It is noted that for small values of F/S the number of nodes scales as a power law F/S^{α} with exponent $\alpha = -1$. When fusion is the dominant mechanism, i.e., if $F/S \gg 1$, the number of nodes scales as a power law as well but the exponent is observed to depend on the number of node degrees n_k as $\alpha = (n_k - 3)/2$. Summarizing the observations as outlined above, the number of nodes at the steady state is given by the scaling relations

$$N_S = \begin{cases} N_0 \cdot \frac{G}{R} \cdot \left(\frac{F}{S}\right)^{-1} & \frac{F}{S} \ll 1 \\ N_0 \cdot \frac{G}{R} \cdot \left(\frac{F}{S}\right)^{\frac{n_k-3}{2}} & \frac{F}{S} \gg 1 \end{cases} . \tag{7.4.3}$$

The scaling is accurate for large and small values of F/S but fails as soon as F and S are of an order and especially if $F = S$. Here, the scaling relations predict $N_S/N_0 = 1$, but inspection of numerical data immediately suggests a dependence on n_k that has not yet been accounted for. It is found that a phenomenological correction factor n_k^{β} that tends

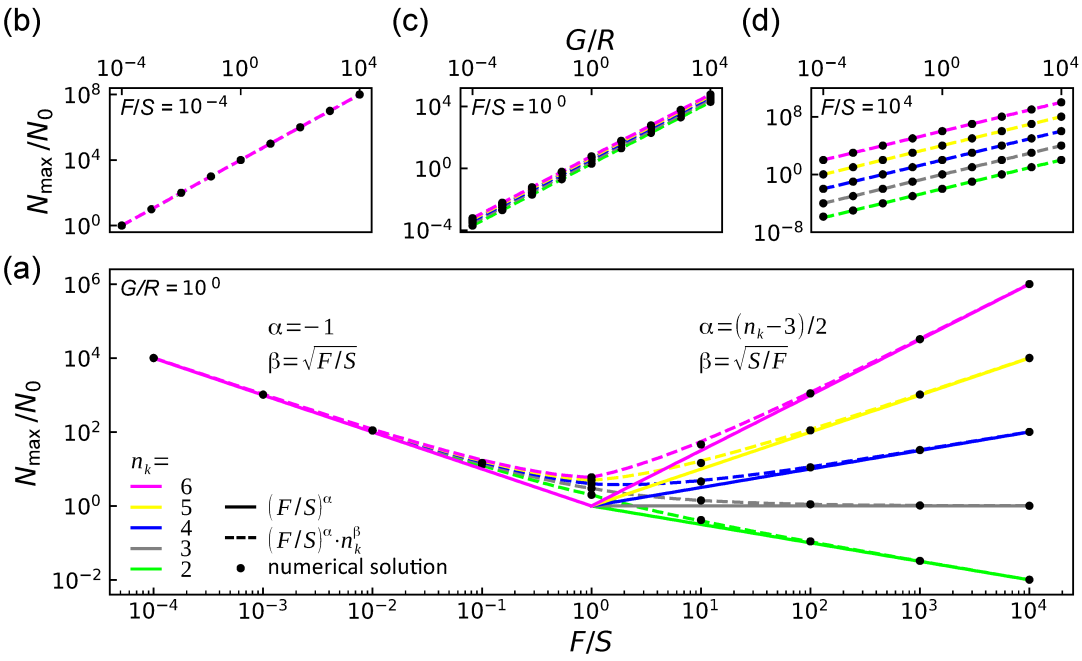


Figure 7.4.3: The number of nodes at the steady state of the rate equation model depends on the ratios F/S and G/R, and on the number of node degrees n_k. Numerical data (black dots) are in excellent agreement with equation 7.4.4, indicated as dashed lines. Solid lines indicate scaling behavior for small or large values of F/S. (a) For small and large values of F/S, the number of nodes scales as a power law with exponents -1 and $(n_k - 3)/2$, respectively. If F and S are of an order, it is necessary to introduce a correction factor n_k^β. (b-d) Modifying the ratio G/R linearly influences the number of nodes. The scaling with F/S is preserved regardless of G/R. The number of node degrees n_k only has an influence when the system is dominated by fusion.

to unity for large and small F/S greatly improves the quality of the scaling relations. The corrected relations read

$$N_\mathrm{S} = \begin{cases} N_0 \cdot \frac{G}{R} \cdot \left(\frac{F}{S}\right)^{-1} \cdot n_k^{\sqrt{F/S}} & \frac{F}{S} \leq 1 \\ N_0 \cdot \frac{G}{R} \cdot \left(\frac{F}{S}\right)^{\frac{n_k-3}{2}} \cdot n_k^{\sqrt{S/F}} & \frac{F}{S} \geq 1 \end{cases} \tag{7.4.4}$$

and are displayed in figure 7.4.3. The excellent agreement for $F = S$ is highlighted by panel (c). Being able to estimate the number of nodes that is approached by a process is of a major advantage for setting up simulations in correct manner. Although not the only influencing factor, excessive system sizes or growth rates will lead to long run-times required for completion.

Degree distribution at the steady state The second quantity of interest to the investigation of the large time behavior of the rate equation model, and arguably the most

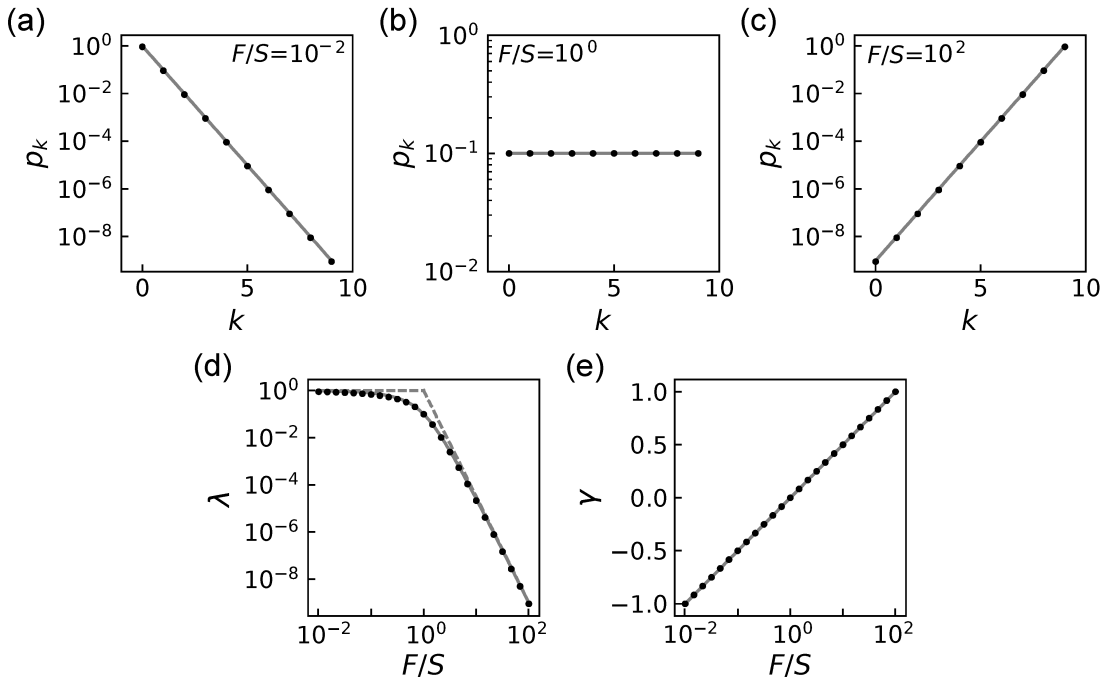


Figure 7.4.4: Degree distribution at the steady state of the rate equation model. At the steady state, the node degrees predicted by the rate equation model are distributed exponentially. Numerical data (black dots) are in excellent agreement with equation 7.4.9, indicated by a solid line. (a) If separation is the dominant mechanism nodes are almost exclusively solitary with a degree zero of zero. (b) If fusion and separation are balanced, all node degrees are represented with equal probability, resembling equipartition in a thermodynamic system. (c) Systems dominated by fusion approach regular graphs with a degree close to $n_k - 1$. (d & e) Parameters of the exponential distribution. Dashed lines indicate asymptotes for small and large values of F/S, solid lines represent the corrected model.

important, is the degree distribution to which the evolving graph converges. Referring to figure 7.4.1, the degree distributions for some selected choices of rate constants have been discussed qualitatively above. In the more realistic case where all processes are possible in a system, the degree distribution at the steady state is more variable. First, it is realized that the degree distribution at the steady state is not influenced by the constants describing growth or retraction as soon as fusion and separation are enabled, and instead depends exclusively on the ratio F/S and the cutoff parameter n_k. It is therefore possible to learn about the structure of the degree distribution by investigation for selected values of F/S, as displayed in figure 7.4.4 panels (a-c). As already noted above, in the detailed balance case $F = S$, the degree distribution is given by equipartition as can be seen in panel (b). For values of F/S larger or smaller than unity, the degree distribution is of an exponential type, leaning either towards nodes of small degree if separation dominates (panel (a)), or towards nodes of large degree if fusion governs the graph evolution (panel (c)). As the dependence of the parameters characterizing the distribution on F/S cannot

be extracted from single values, a generic exponential distribution

$$p_k = \lambda\, 10^{\gamma k} \tag{7.4.5}$$

is fitted to the distribution for a wide range of F/S. The resulting parameters λ and γ are shown in figure 7.4.4 panels (d) and (e). Inspection of the parameter γ suggests a logarithmic relation, and indeed the function

$$\gamma = \frac{\log_{10}(F/S)}{2} \tag{7.4.6}$$

is found to be in excellent agreement with the data for all values of F/S and enables writing the degree distribution as

$$p_k = \lambda \left(\frac{F}{S}\right)^{\frac{k}{2}} \quad . \tag{7.4.7}$$

Obtaining the relation for the parameter λ proves to be less straightforward. However, from previous discussion of the phenotypes of the rate equation model, asymptotes of the parameters can be obtained. First, if F/S tends to zero, it is expected that the network consists almost exclusively of solitary nodes of degree zeros, and therefore the degree distribution has to be peaked at $k = 1$, implying $\lambda = 1$ as the parameter functions as an intercept. For $F = S$, equipartition is expected and therefore it is required that $\lambda = 1/n_k$ as all degrees need to be present with equal probability. The asymptote for large values of F/S stems from inspection of the parameter λ in figure 7.4.4 panel (d). Here, power law scaling of the parameter with an exponent given by $-(n_k - 1)/2$ is observed, and thus the asymptotic behavior can be summarized as

$$\lambda = \begin{cases} 1 & \frac{F}{S} \ll 1 \\ \frac{1}{n_k} & \frac{F}{S} = 1 \\ \left(\frac{F}{S}\right)^{-\frac{n_k-1}{2}} & \frac{F}{S} \gg 1 \end{cases} \quad . \tag{7.4.8}$$

Although the asymptotes correctly reflect the behavior for large and small values of F/S, they fail near $F = S$. Similar to the procedure applied during investigation of the number of nodes, this problem can also be alleviated by multiplication of a correction factor that tends to unity for large and small values. This factor can be given as $n_k^{-F/S}$ for $F/S \leq 1$, and $n_k^{-S/F}$ for $F/S \geq 1$. The corrected parameters as well as the asymptotes are shown in figure 7.4.4 panel (d).

In summary, the degree distribution for arbitrary values of F/S and arbitrary values of G/R reads

$$p_k = \begin{cases} n_k^{-F/S} \cdot \left(\frac{F}{S}\right)^{\frac{k}{2}} & F/S \leq 1 \\ n_k^{-S/F} \cdot \left(\frac{F}{S}\right)^{\frac{k-n_k+1}{2}} & F/S \geq 1 \end{cases} \tag{7.4.9}$$

and is found to be in excellent agreement with numerical data as can be seen in figure 7.4.4 panels (a-c).

Influence of rate constant magnitudes In the previous analyses it has been established that the behavior of the GRFS model for large times does not depend on the individual values of the constants G, R, F, S, but instead changes only with the parameter ratios F/S or G/R, leaving open the question which role is taken by the magnitude of the rate constants. Indeed, the most significant cue to answering this question has been addressed in part upon discussion of figure 7.4.1 panels (b) and (f), where it has been noted that the uniform degree distribution that is observed whenever fusion and separation have identical values. There, it has been pointed out that the tendency of a system that is not driven to evolve into a uniform distribution is a process that is similar to relaxation into thermodynamic equilibrium. This analogy can be expanded.

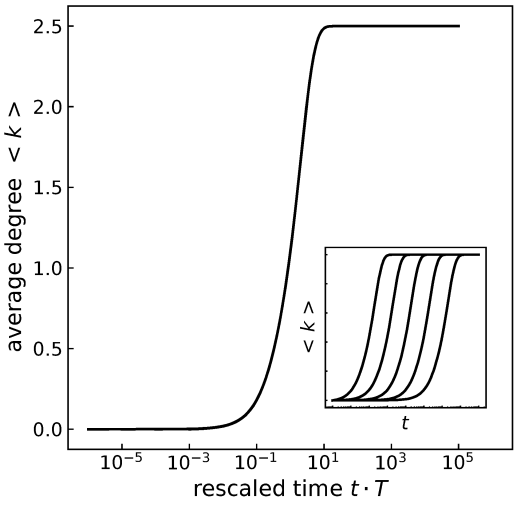


Figure 7.4.5: Unlike the details of the steady state, the dynamics of the model depend on the magnitude of the rate constants. The average degree as an indicator for the topological state develops into the steady state significantly faster when larger absolute values are selected for the rate constants F and S, although their ratio remains constant. The inset shows the dynamics of the average degree for five distinct values of $F = S = T$ for $n_k = 6$. The data collapse onto a single curve when time is rescaled by T.

The mechanism by which a perturbed thermodynamic system returns to equilibrium is driven by thermal noise. Transitions that are allowed in the system occur randomly due to thermal energy fluctuations. If the system is not in equilibrium, it is more likely for those transitions to occur that bring the system closer to the equilibrium state. A thermodynamic system will always return to equilibrium unless it is at a temperature that is too small to allow elementary transitions. At larger temperatures, the relaxation process occurs faster than at small temperatures, as the rate at which transitions occur is increased. Extending this logic to the graph evolution model presented in this chapter, the magnitude of the rate constants controls the number of random transitions per unit time and thus has a role comparable to the thermodynamic temperature. Indeed, as can be seen in the inset of figure 7.4.5, a system in which fusion and separation returns to

equilibrium faster if the magnitude $T = F = S$ is larger. Further, it is observed that the topological dynamics can be collapsed on a single curve when regarded as a function of the rescaled time $T \cdot t$.

■ 7.5 Analysis of stochastic simulations

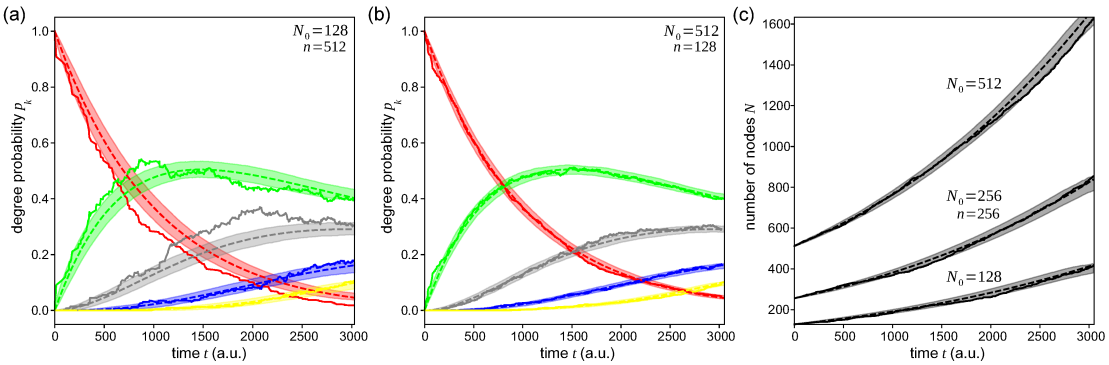


Figure 7.5.1: Typical trajectories obtained by stochastic simulation. Panels (a) and (b) show the time evolution of the degree distribution of a random graph evolving by growth and fusion, i.e., $G \neq 0$, $F \neq 0$, $R = S = 0$ for two initial system sizes N_0. Panel (c) shows the evolution of the system size for three different N_0. In all panels, dashed line indicates the rate equation, the solid trajectory indicates one selected simulation and the shaded region indicates the mean ± standard deviation obtained by realizing n simulations.

Stochastic methods as introduced in chapter 3 can provide new insights into the dynamics of a system. The method employed in this work builds on the stochastic simulation algorithm (as presented in section 3.4) and has been presented in section 7.3.3. Due to the implementation similar to the stochastic simulation algorithm, it is expected that the stochastic network simulation algorithm produces trajectories that are exact samples of the master equation given in section 7.3.2. Furthermore, it is expected that the average behavior of these trajectories is described by the deterministic rate equation that is presented in section 7.3.1 and analyzed in the previous section in a reduced model that requires fewer constants to be specified. In the present section, the analysis is extended to the stochastic model.

Example trajectories of simulations obtained via the stochastic network simulation algorithm (algorithm 7.3.1) are displayed in figure 7.5.1. Panels (a) and (b) show the evolution of the node degree distribution for two different system sizes. As the simulated process is driven by growth and fusion, it is worthwhile to inspect also the evolution of the system size, which is done in panel (c). Although single trajectories fluctuate significantly, averaging them over a large number of realizations leads to an average behavior that is in excellent agreement with the mean-field solution predicted by the rate equation model. The latter is indicated in the figure as a dashed line, whereas the average ± one standard deviation over a representative number of simulations is shown as a shaded region. Furthermore, it is observed that the magnitude of fluctuations as indicated by

the standard deviation depends on the system size. As expected, fluctuations of the node degree probabilities are significantly larger for small system sizes. The evolution of the number of nodes equals the evolution of the system size. In this case, it is the relative magnitude of fluctuations that decreases as system size increases.

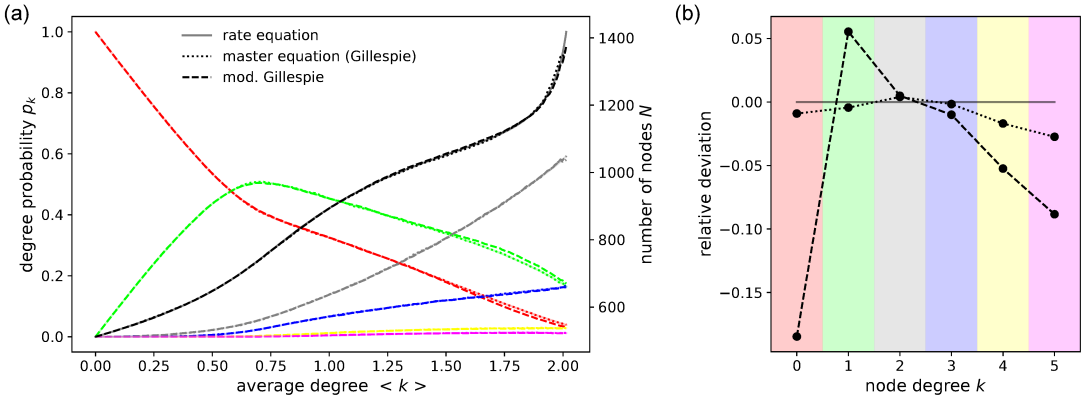


Figure 7.5.2: Comparison of rate equation, Gillespie stochastic simulation algorithm and modified SSA for graph generation. (a) Stochastic simulation of the master equation (averaged over 1024 realizations, $N_0 = 162 \pm 60$, parameters obtained from section 5.3) and the deterministic rate equation match accurately. Slight deviations are observed when comparing with the modified SSA (identical parameters), especially when rare nodes with a large degree first emerge. This effect is attributable to situations in which an event is randomly selected that would require a configuration of connected nodes that is not found in the present graph, leading to an effective modification of the propensity function. (b) Deviation of the stochastic models relative to the deterministic solution near $< k >= 2$.

In the selected case, the stochastic simulation algorithm (algorithm 3.4.1) and the stochastic network simulation algorithm (algorithm 7.3.1) are virtually identical. Driven by growth and fusion, the probability that a selected process cannot be carried out on the present graph is extremely small. However, as soon as retraction and separation are active in a system, there are several possible scenarios where deviations may occur. Possible cases have been briefly discussed in section 7.3.3. Instead of performing an in-depth analysis of possible deviations between the respective algorithms, the following practical approach has been chosen here. In all cases throughout this work, where simulations are performed and where retraction or separation are significant, the results have been compared to the rate equation model with the result, that deviations due to the approximative nature of the algorithm are not significant. To highlight this, in figure 7.5.2 the numerical solution of the rate equation is shown alongside results obtained from both algorithms, averaged over 1024 realizations. The parameters were selected such that the evolution of node degrees matches with the evolution observed in the experimental data investigated in chapter 6, in order to create a representative example of deviations in a realistic scenario.

Notably, in the shown example deviations remain absent unless the average degree grows beyond $< k > \approx 1.5$. In this regime, it is observed that the stochastic network simulation algorithm slightly differs from the rate equation and the stochastic simulation algorithm, which remain in excellent agreement throughout the entire process as expected. The deviations, which do not surpass 20 % in the case of the vanishing p_0-fraction and remain much smaller for all others, are systematic alterations of the propensity function. The rationale presented in section 7.3.3 can be briefly recapitulated as follows. If a process is selected due to the availability of its constituents, but cannot be carried out because the process requires a certain configuration of nodes that is at the present not available on the graph topology, the problematic time step will be repeated until a suitable process is selected. The propensity of the processes that were skipped are thus effectively zero, even though in theory they should take a nonzero value. However, the nature of this systematic deviation makes it a problem of the theory rather than of the algorithm, and due to its low impact on the results, it will not be considered in the following.

■ 7.5.1 Simulating the GRFS model in two dimensions

In section 7.3.4, necessary steps for a two-dimensional implementation of the network growth model have been outlined. A prototype of the algorithm has been implemented and has been used to generate a small number of exemplary analyses. A systematic analysis of the model has not been performed. In the present section, typical data generated by the algorithm is presented. The most significant contribution of the two-dimensional model to this work is presented and discussed in section 7.6, where differences in percolation between the spatial and non-spatial settings are considered.

In figure 7.5.3, a sequence of graphs is shown that has been created using the prototype implementation of the proposed algorithm. The simulation was performed with an initial area coverage of $c_0 = 0.1$ and $N_0 = 128$ initial nodes. Rate constants were selected in accordance with the GRFS model presented in section 7.4.1. In the present example, only growth has been enabled.

As can be seen in the figure, the two-dimensional simulation recreates typical steps also observed in the formation of the *P. polycephalum* network. Initially, there exist, with few exceptions based on the selected area coverage, only solitary nodes. As these grow, larger components are generated locally and fuse as they come into contact. At first, these components are tubular but soon resemble locally tree-like structures. These structures rapidly become more intricately meshed upon further growth and fusion, and finally merge into a single, giant component in a percolation transition.

Comparing to *P. polycephalum*, there are, of course, obvious differences. For instance, nodes grow stochastically based on their rate constant, rather than through their functional role in the network. This implies the absence of growth fronts and the absence of the typical hierarchical structure. As all growth rates are equal in the present example, the mesh-structure of the graph is locally very intricate, favoring nodes of large degree. In *P. polycephalum*, these are only observed in growth fronts.

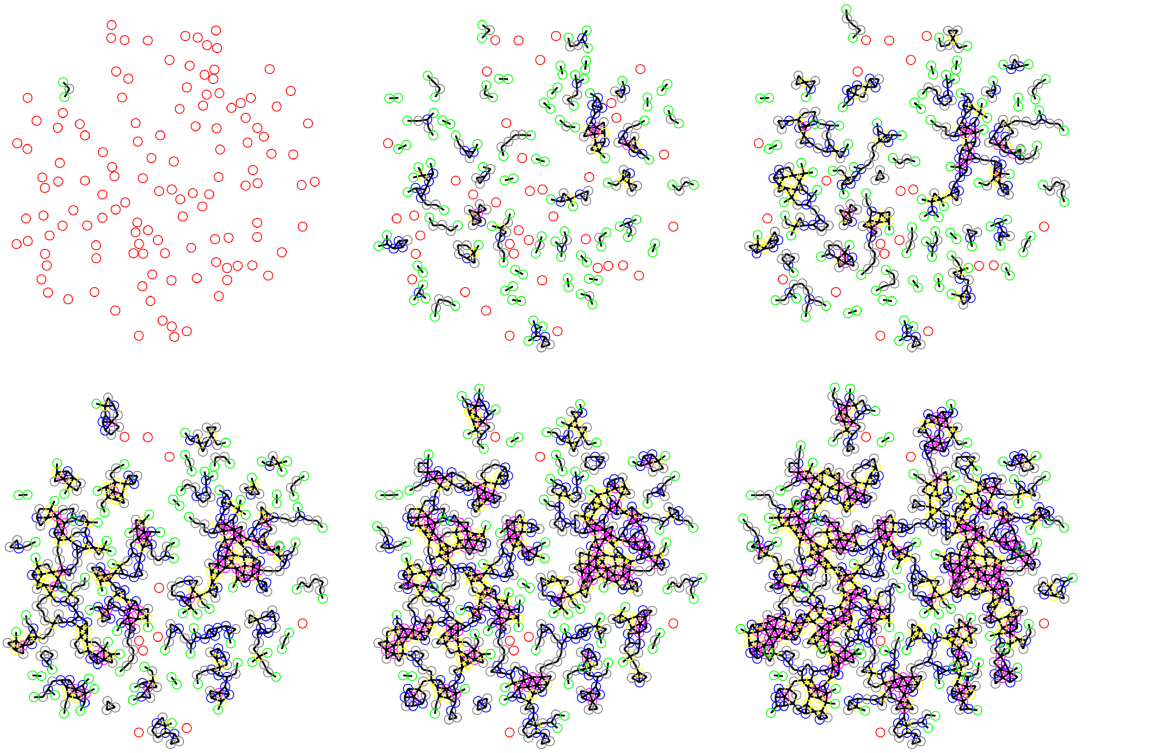
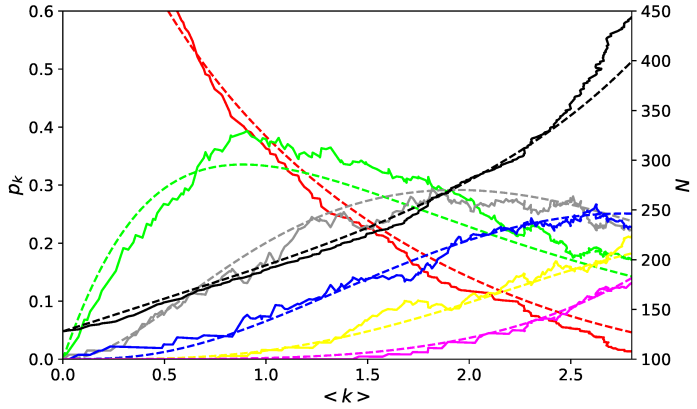


Figure 7.5.3: Network growth simulated in two dimensions ($n_k = 6$, $c_0 = 0.1$, $N_0 = 128$). The simulated system is driven exclusively by growth; fusion occurs only when nodes come into contact. In the present example, all nodes have identical growth rates, leading to a significantly larger fraction of high-degree nodes compared to *P. polycephalum* networks.

The evolution of the degree distribution of the simulated example is shown in figure 7.5.4, and resembles the process described above. Comparing qualitatively to similar data in *P. polycephalum* or in the non-spatial, simulated cases seen in the previous sections, there are no substantial differences: solitary nodes are incorporated into the graph structure, intermediately giving rise to degree one nodes as a dominant species. After this point, nodes of larger degrees become more significant. Although the system is driven only by growth and fusion, it is observed after passing an average degree of $< k >= 2$ that there are fluctuations associated with an effective decrease in average degree. This is due to the reason that a growth event, which adds one node and one edge to the graph, effectively decreases the average degree if $< k >= 2$. This is addressed in greater detail in the next section.

It is not straightforward to set up parameters of the rate equation to match the two-dimensional simulation, as the two-dimensional simulation is controlled purely by growth and spatial aspects. However, tracking of fusion events during simulation makes it possible to estimate the fusion rate constant. The associated growth to fusion ratio in the present case was found to be $G/F = 0.78$. Numerically solving the rate equation



0 1 2 3 4 5+

Figure 7.5.4: Topological evolution of the two-dimensional simulation shown in figure 7.5.3. The simulated trajectory is shown as solid lines, dashed lines represent a numerical solution of the rate equation for $G \neq 0$, $F \neq 0$. Fusion rates were obtained by tracking of fusion events in the two-dimensional simulation.

with these values recreates the topological evolution, as shown in figure 7.5.4 as dashed lines.

The percolation transition observed in figure 7.5.3 is discussed in greater detail in section 7.6, where it is confirmed that the scaling behavior of the system at the transition is consistent with expected values for two-dimensional percolation.

■ 7.5.2 Percolation in the GRFS model

The major advantage of the network simulation algorithm over the rate or master equations and their respective numerical realizations is the opportunity to derive graph-related measures from the results. The graph-structural aspect that is of central interest to this work is the emergence of a giant component in a percolation transition as detailed in chapter 2. This process, which can be reproduced by simulations as demonstrated in this section, has been studied in section 6.2.2 on the example of reforming *P. polycephalum* networks. Moreover, in chapter 4 analytical solutions to the percolation problem are devised for random graphs described by the Erdös-Rényi model (Erdös and Rényi, 1959) and in the configuration model (Molloy and Reed, 1995), which treats graphs of arbitrary degree distribution. Such solutions are possible only if nodes of a degree larger than a certain maximum degree occur at negligible frequency, which is shown to be the case in section 6.2. *P. polycephalum* develops into a topological steady state characterized by a fixed degree distribution with only a minor fraction of nodes of large degree.

However, it is observed in section 6.2.2 that theory and experiment do not agree well. Among the possible reasons discussed there are the influence of growth on the network, and finite size effects. In the present section, percolation will be studied in the GRFS model (section 7.4.1) upon variation of the growth to fusion ratio G/F, leading to the

result that growth indeed causes deviations as observed in the data. Furthermore, the influence of the system size is analyzed in two cases, upon variation of the maximum allowed degree, and in case of the growth to fusion ratio.

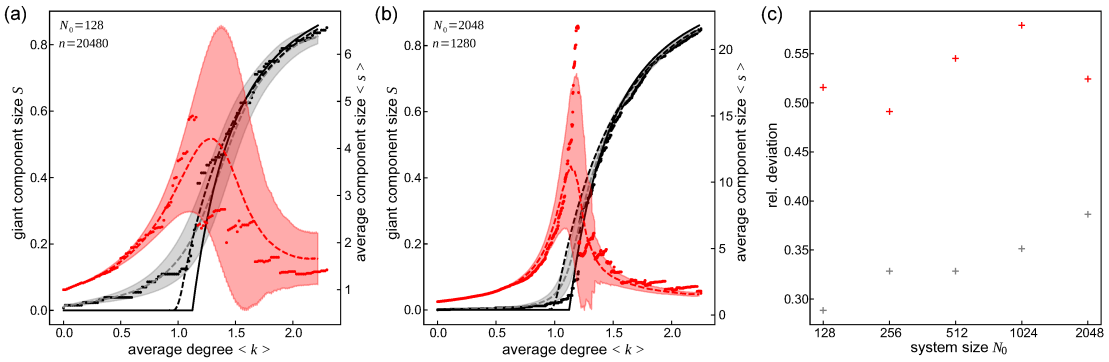


Figure 7.5.5: Component sizes in the GRFS model for $F \neq 0$ and $G = R = S = 0$. For the selected set of parameters, the GRFS model is identical to the Erdös-Rényi model except for the cutoff parameter n_k limiting the maximum degree, which is expected to cause a shift of the transition. Panels (a) and (b) show the giant component size (gray) and the average non-giant component size (red) for two different system sizes. Dashed lines average all simulations, shaded regions are one standard deviation. Dots indicate one trajectory. The critical region is more narrow when system size is large, and the critical point appears shifted for smaller sizes. Theoretical solutions are shown for the configuration (solid) and random graph (dashed) as black lines. (b) Relative fluctuations near the critical point. Relative fluctuations of the average component size remain constant. Absolute fluctuations of the giant component size decrease, whereas relative fluctuations increase.

Typical data Figure 7.5.5 panel (a) and (b) show typical results obtained from simulating the growth of the giant and average non-giant components using the network simulation procedures (algorithm 7.3.1) with two different system sizes. In the present examples, only fusion drives the process and for the cutoff parameter $n_k = 6$ is used. Data shown are typical simulated trajectories at either system size, and the average over all simulations together with the standard deviation. Furthermore, theoretical solutions in the configuration and random graph models (see chapter 4) are shown.

Both component size measures show the expected behavior. As nodes fuse, the giant component grows continuously until it covers the entire system, whereas the average component size grows until percolation, and decreases afterwards as more average sized components are fused to the giant component. In a single trajectory, these post-transitional fusion events can be clearly seen as discrete jump events that occur whenever a component is taken out of the average.

Comparing the data in both panels indicates that in smaller system sizes the critical region is not as well defined as in the large system size. The giant component size appears

more smeared out, and the peak of the average component size is rather broad. As both systems are finite, the average component size does not diverge. Instead, it is noticed that at the critical point the standard deviation of the average component size takes its maximal value. Relative deviations σ_S/S and $\sigma_{<s>}/< s >$ at the location of the maximum of $< s >$ are shown in panel (c) for five different system sizes, including the data shown in panels (a) and (b). The standard deviation $\sigma_{<s>}$ of the average component size maintains an approximately constant relation to the height of the peak, although both quantities grow with system size. The giant component size becomes more narrowly defined as system size increases, indicating that for the entire curve, absolute deviation is reduced. In contrast, relative deviations near the critical point increase as seen in panel (c), highlighting the divergence of order parameter fluctuations at the critical point.

It is immediately seen that in both cases the average agrees well with the theoretical models in the case that only fusion is present in the system. Predictions made by the configuration model are more accurate as the random graph model cannot account for deviations of the degree distribution from the Poisson distribution. As discussed in section 4.3.1, this is attributable to the cutoff parameter n_k, which does not allow nodes in the tail of the distribution to evolve to larger degrees. The most prominent effect related to this is a shift of the expected critical point in relation to the random graph model.

Deviations from the theoretical models are smallest in the percolated regime and largest shortly before percolation. This is a typical finite size effect. In a finite system, one component in the graph will always be largest, even prior to the phase transition. The second prominent influence of finite system size is a shift of the peak of the average component size. In an infinite system, this peak is located at the critical point, but in finite systems it is shifted into the percolated regime. The influence of finite size effects are visible in figure 7.5.5, but are better resolved in figure 7.6.1 below.

So far, only fusion has driven the development of the simulated system. Before covering the influence of growth in the following section, it is sensible to consider briefly the influence of the process opposing fusion, i.e., separation. If separation is enabled in addition to fusion, the configuration model still describes the evolution of the giant component correctly (not shown). In section 7.4.2 it has been described that the degree distribution at the steady state of the model is determined solely by the parameter F/S, i.e., the ratio of fusion and separation rates. The configuration model, which is applicable in the current situation, predicts the size of a giant component based on the degree distribution. Now, if the degree distribution is determined by the parameter F/S, so is the size of the giant component.

This is verified in figure 7.5.6, where the size of the giant component S at the steady state, as obtained from simulations, is shown as a function of F and S. It can be seen there, that no variation of S can be detected along lines of constant ratio F/S. If the system is dominated by fusion, the giant component at the steady state almost surely contains all nodes in the system, whereas for separation-dominated systems there exists no giant component. The transition between the two phases is smooth. In particular, if $F = S$, the giant component typically connects half of the nodes in the system when

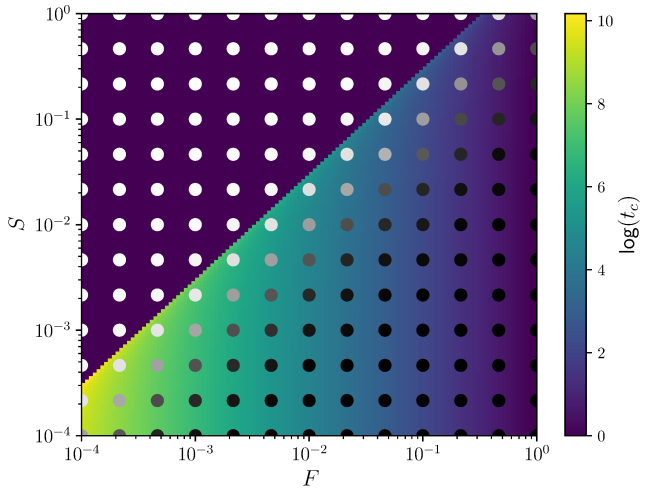


Figure 7.5.6: Giant component size & percolation time. In the absence of growth, the configuration model predicts the emergence and size of a giant component correctly. The ratio of fusion to separation determines whether a giant component develops in a system, and the magnitude of the parameters determines the time required by the system to percolate. Percolation time t_c is determined from the configuration model. Colored dots represent giant component size in the topological steady state obtained via simulations.

at the steady state. Notably, the location of the phase transition as indicated by the configuration model does not coincide with the line $F = S$. The predicted transition is shifted slightly and marks the line above which no giant component exists at all.

It has been discussed in section 7.4.2 that, although the topological details of the steady state are exclusively defined by the ratio F/S, the time required to reach the steady state depends on the magnitude of the individual constants. If both constants have a large value, this can be interpreted as a large amount of noise being present in the system. The same effect is observed when monitoring the time required by the system to percolate, as can be seen in figure 7.5.6. Although times shown there have been obtained via the configuration model, the same effect is observed in simulations.

A percolation transition, which in good approximation is driven by the ratio between fusion and separation, is prominently observed in mitochondrial networks (Sukhorukov et al., 2012; Zamponi et al., 2018). Mitochondria are organelles of eukaryotic cells that are best known for providing cells with the energy-rich molecule adenosine triphosphate through the respiratory chain. Opposing textbook presentation, mitochondria typically are not encountered as the small, kidney-shaped units as which they are commonly pictured. Instead, it is observed that mitochondria form tubular networks that span distances comparable to the size of the cell (Rafelski, 2013). These networks are not static but instead undergo a constant cycle of fusion and fission (separation), in which parts of the network are cut off, and are possibly reattached later on. Whether a mitochondrial

network is in a fragmented state, or whether it is connected throughout the cell depends on the balance between fusion and fission processes. The function of the remodeling of the network has not yet been entirely clarified, but a deficiency thereof has been linked to numerous diseases. It has therefore been suggested as a control mechanism target at maintaining mitochondrial quality. Indeed, computational studies suggest, that during aging, optimal quality is maintained through shifting of the balance in direction of fusion Figge et al. (2012). The process has been addressed in the light of graph percolation by Sukhorukov et al. (2012), who propose a model of two events – tip-to-tip and tip-to-side fusion – and their inverse processes. The model is characterized by rate constants, and thus is of a very similar concept as the model proposed here. Indeed, limiting the set of fusion and separation events in the present model is conjectured to recreate the situation in mitochondrial networks.

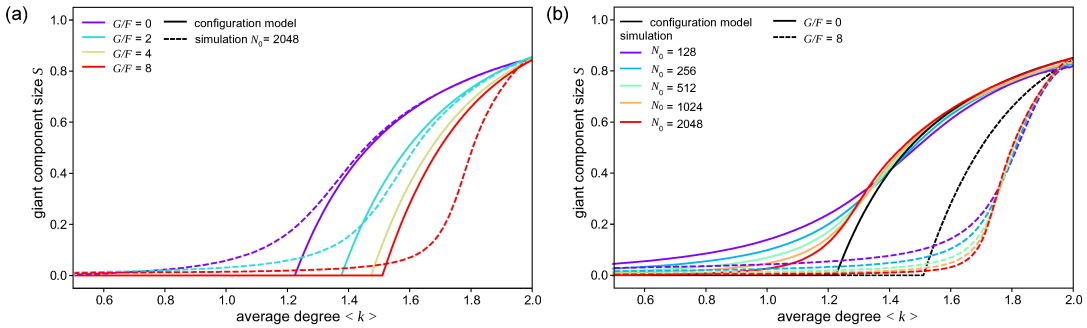


Figure 7.5.7: Influence of growth and system size on the emergence and size of a giant component. (a) When no growth is present, simulated trajectories are well-described by the configuration model except for deviations due to random fluctuations and finite size effects. When growth is present, the percolation transition is shifted and trajectories no longer match with the configuration model. (b) When no growth is present, increasing system size leads to better agreement between simulations and the configuration model; when growth is present, deviations are more pronounced in larger systems.

Influence of growth In order to elucidate whether deviations of experimental data from the theoretical models are associated with system growth, simulations have been performed upon variation of the growth to fusion ration G/F. It has already been demonstrated above that when $G/F = 0$, only finite size effects cause deviations from the theoretical solutions. This picture changes drastically as soon as growth is enabled.

In figure 7.5.7 panel (a) simulation averages at one system size ($N_0 = 2048$, $n = 896$) are displayed for $G/F = 0 \ldots 8$, compared to the configuration model. For $G/F = 0$, the result is identical to the cases described above, whereas for $G/F > 0$ there are strong deviations from the theory. The giant component remains small until the graph has structured to a significantly larger degree. The shift is analyzed quantitatively below.

Introducing finite size effects to the picture leads to the situations depicted in panel (b). If there is no growth, finite size effects cause the simulation average to converge

towards the theoretical prediction. If, however, growth is present the finite size effects intensify the distortion due to growth. The picture created here via simulations resembles closely the situation observed in *P. polycephalum* data.

One argument for why there is a shift comes from the consideration of the limiting case $F/G = 0$, i.e., when the growth rate exceeds the fusion rate greatly. It is conjectured that in an infinite system, which evolves exclusively by growth all components are necessarily trees. Due to the infinite size of the system it is impossible for the system to develop a giant component and therefore, the critical point needs to be shifted to infinity.

From a different point of view, the deviating shape of S in the presence of growth can be motivated. In the GRFS model, all nodes have an identical probability of undergoing a growth process, except for a fraction of nodes that has the maximal allowed degree. This fraction, however, is small until the graph is highly structured which is the case only after percolation. Focusing on the stage in which a graph is still in the process of developing its structure, the following reasoning appears plausible.

In a finite system, there exists at each point during the evolution a component which is largest, and therefore the graph can always be partitioned into nodes that are part of the giant component and those that are not. Since all nodes in the graph may add one connected novel node through a growth process with identical probability, components grow with a rate that is proportional to their size. However, each growth process progresses the parameter $< k >$, regardless of the component in which the growth process takes place.

It is informative to take a closer look at how $< k >$ changes upon single growth events. From $N < k >= 2E$, where N and E represent the number of nodes and edges, it follows that growth (E+1, N+1) increases the average degree if $N > E$ or equivalently if $< k >< 2$, and decreases the average degree otherwise. The amount, by which the average degree is progressed due to single one single growth event is

$$\Delta < k >= \frac{2- < k >}{N+1} \ , \tag{7.5.1}$$

which tends to zero as $< k > \rightarrow 2$. A network advancing only by growth cannot obtain an average degree larger than two.

As long as the giant component size is small, growth happens predominantly in the portion of the graph that is not part of the giant component. Thus, for each such process, the graph obtains a higher degree of structure without an increase in giant component size. In a sense, the giant component is left behind, and keeps a smaller value at a higher level of structure, compared to the case with no growth. This effect is diminished at higher average degrees, as the influence of growth processes on $< k >$ progressively tends to zero when approaching an average degree of two, as seen in equation 7.5.1. At the same time, graphs with a larger degree tend to have a larger giant component due to fusion processes. Thus, a larger fraction of growth takes place in the giant component, and the effect described above is even further diminished.

The influence of growth can be summarized as follows. Comparing to a model with no growth, growth causes the giant component to be smaller at a similar average degree as long as the giant component is small. As the graph approaches larger degrees, the influence of growth is diminished by two factors. First, at larger degrees the giant component is larger and therefore growth effects are more prone to take place in the giant component. Second, the rate at which growth progresses the average degree decreases if the graph is more structured. Taken together, both effects are conjectured to lead to the effect observed in figure 7.5.7, where increased levels of growth cause the giant component to remain small first, but to rise sharply at larger average degrees.

Another study investigating the influence of growth has been performed by Callaway et al. (2001). There, a network forms by the same rules governing the formation of a random graph, while one single, solitary node is added per unit time. The authors observe two effects. First, the phase transition is shifted in comparison to the configuration model, but in the direction of smaller average degrees. Second, the giant component is smoothed out at the transition.

The first point raises the question why Callaway et al. (2001) observe formation of a giant component at smaller average degrees, when it is observed here that growth rather inhibits the growth of the giant component. The authors reason that degree correlations due to the different age of nodes are key to understanding the process. Nodes added in their model are exclusively of degree zero, meaning that at the point at which they are inserted all preexisting nodes have had a higher probability to have obtained an edge. This leads to the emergence of a core of more highly connected nodes, which leads to the early emergence of a giant component.

The difference between the two processes can also be understood following the logic presented above. In the model developed here, the nodes that are added due to growth increase the average degree of the graph, unless the graph surpasses an average degree of two. Therefore, the average degree progresses faster than the average component as growth may occur in all components of the graph, but is more likely in the fraction that does not belong to the largest component. In contrast, in the model presented by Callaway et al. (2001), growth is limited to the fraction of the graph that is not part of the giant component. Although this should, in principle, inhibit the growth of the giant component, it also decreases the average degree and therefore reverses the effect observed here.

Callaway et al. (2001) observe numerically that the order parameter is smooth the transition between the non-percolated and percolated regimes, implying that the phase transition is of infinite order (Kosterlitz and Thouless, 1973), rather than of second order as in percolating non-growing networks. Heuristic arguments that an infinite order phase transition is a feature of all growing networks have subsequently been provided by Dorogovtsev et al. (2001), and have since then been supported by rigorous works such as Bollobás et al. (2004) focusing on select systems and making results translatable between different models. Although the works referenced here indicate that an infinite order phase transition is shared between growing systems, it is unclear whether the results translate

to the present model.

Should this be the case, it is to be expected that there are influences on the finite size scaling presented in the following section, as in infinite order phase transitions critical quantities, such as the susceptibility or non-giant component size, are not guaranteed to diverge (Keesman et al., 2016).

■ 7.6 Finite size scaling & critical exponents

Throughout this work, the location of the critical point of several percolation processes has been addressed. In chapter 4, analytical solutions of the configuration model have been devised, and it has been found that limiting the degree distribution to small degrees, either by neglecting nodes of larger degree or by disallowing nodes to surpass a certain degree, leads to a shift of the critical point. In chapter 6, it has been observed that the critical point in *P. polycephalum* percolation is shifted in comparison to the configuration model. This effect can be attributed to system growth and finite size effects, and has been further studied in the previous section qualitatively through the analysis of simulations.

In the present section, this analysis is further expanded. Building on a more quantitative approach for determination of the critical point, the behavior of the system in the critical region is analyzed by determining the critical exponents that characterize the scaling of the giant and average non-giant component sizes near the critical point. The determination of critical exponents in simulated systems is typically performed via the analysis of finite size effects, as their scaling behavior with the finite system size is governed by the same critical exponents describing the scaling behavior in the critical region. This is further outlined in section 2.1.3.

Three data sets are considered as part of the analysis. First, critical point and exponents are studied in the GRFS model upon variation of the parameter n_k when the system evolves exclusively by fusion. The value $n_k - 1$ is the limiting degree that cannot be exceeded in a small degree graph. In chapter 4, it has been described that such a cutoff shifts the location of the phase transition in comparison to the unconstrained random graph model. Second, the influence of growth is considered, and the critical point and exponents are investigated as functions of the growth to fusion ratio G/F. Finally, using the rate constants obtained from experimental *P. polycephalum* data, simulations mimicking the development of the slime mold network are performed and analyzed.

The methods employed here are described in section 2.1.3. Two alternative procedures are employed to estimate the parameters of interest. In the first procedure, the critical point is determined using the low-sampling method by Bastas et al. (2014), and subsequently a data collapse is performed using the method by Houdayer and Hartmann (2004) with only the critical exponents as parameters. The second procedure uses the critical point provided by the low-sampling method as a guess, and the data collapse is performed with the critical exponents and the critical point as optimization parameters. It is found that there exist discrepancies in the critical point and exponents between

both procedures. In order to provide a consistent picture, results are reported for both methods.

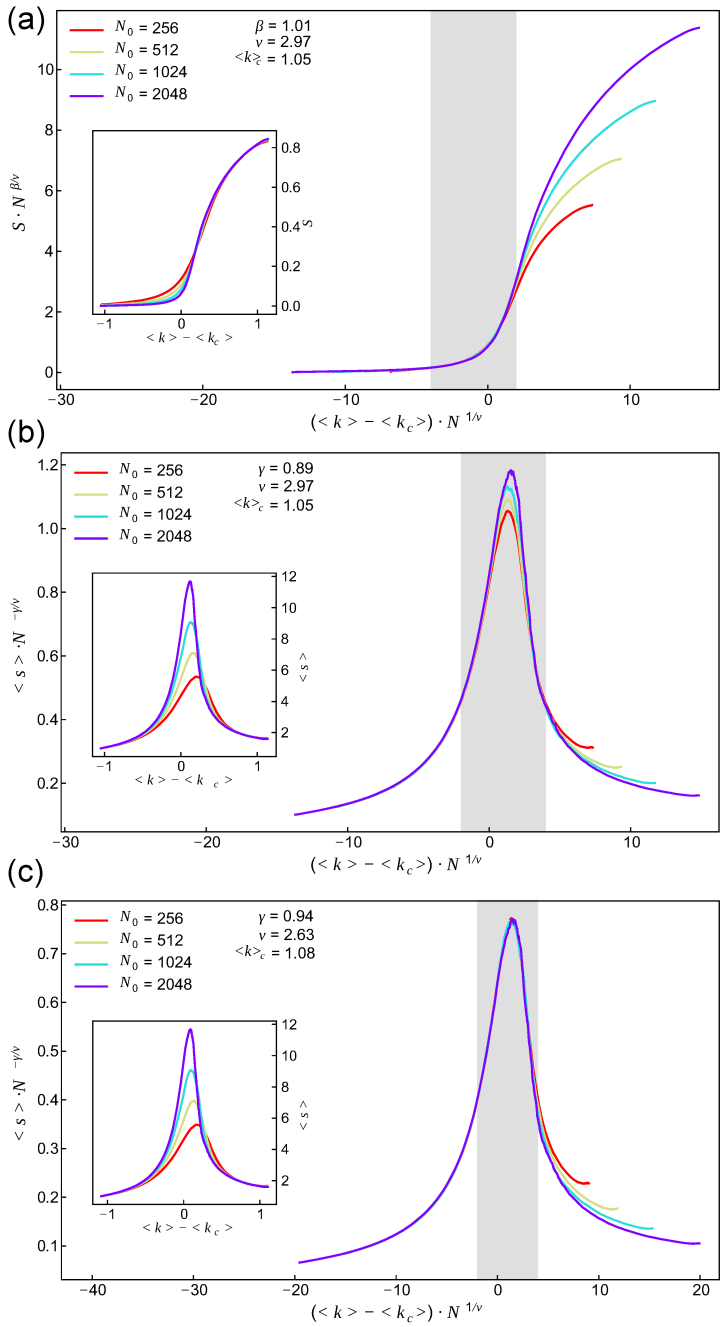


Figure 7.6.1: Data collapse of S, $< s >$ in the GRFS model with $n_k = 5$ when the system is driven only by fusion. All collapses are computed in the shaded region. Insets show the unscaled data. (a and b) use a predetermined critical point. (a) Collapse of the giant component size S. Exponents β and ν are in excellent agreement with the values expect for bond percolation on random graphs. The critical point is shifted in comparison to the random graph model. (b) Collapse of the average non-giant component size. The data collapse is not accurate, and the critical exponent deviates moderately from the random graph case. (c) A much better collapse is achieved when determining the critical point through the data collapse. However, in this case ν deviates from the random graph case.

7.6.1 Influence of a bounded degree

Focusing, as outlined, first on the case where the cutoff parameter n_k is varied, simulations have been performed for five system sizes ranging from $128 - 2048$ nodes such that each system size is represented by exactly twice the number of nodes used in the next smaller system. Critical quantities were averaged over $20480 - 1280$ realizations, halving the number of simulations for each smaller system size due to increased runtime. Working in the GRFS model, the set of parameters is given as $F = 1$, $S = G = R = 0$.

Figure 7.6.1 shows a typical result for one value of the cutoff parameter ($n_k = 5$). Shown there as insets are the averaged of the critical quantities S and $< s >$. The observed behavior is identical to the behavior reported earlier in this section: in the giant component size, finite size effects are most significant shortly before percolation, with a decreased effect for larger system sizes. The average non-giant component differs between system sizes mainly due to peak height and position, where larger systems are peaked closer to the expected critical point.

The figure focuses on the data collapse for the critical quantities. Data collapses shown in panels (a) and (b) have been obtained using the first procedure, where the critical point $< k_c >$ is estimated using the procedure by Bastas et al. (2014), and the collapse adjusts only the critical exponents γ, β, ν. The data collapse procedure has been adjusted such that both collapses are performed simultaneously with a joined quality function $(Q_S + Q_{<s>}) \cdot |Q_S + Q_{<s>}|$ that is minimized with regard to the critical exponents. Here, Q_S and $Q_{<s>}$ are quality functions as defined by Houdayer and Hartmann (2004) (see also section 2.1.3) for both critical quantities. As the scaling relations given in section 2.1.3 only hold in the critical region, the data collapse was computed only in that region, indicated in the figures as a shaded bar. In units of the rescaled axes, the following ranges were found to lead to the most consistent results: $S : [-4, 2]$ and $< s >: [-2, 4]$.

The model that is most closely related to the case discussed here is the random graph model, as the construction mechanism in both models is identical, and the only difference is posed by the degree bound implemented here. The critical point as determined via the method by Bastas et al. (2014) ($< k_c >= 1.05$) is slightly shifted in comparison to the expected value from the random graph model ($< k_c^{\mathrm{RG}} >= 1$). Qualitatively, this is consistent with the reasoning proved in section 4, where it is argued that limiting the maximal degree of a graph is bound to cause the critical point to shift to larger values. Turning to the data collapse, data could be successfully collapsed for both critical quantities, although the collapse of the giant component size is, by visual inspection, more accurate.

Critical exponents in the presented example are $\gamma = 0.89$ for the average component size exponent, $\beta = 1.01$ for the scaling of the giant component size and $\nu = 2.97$ for the scaling of the correlation length. In order to assess the accuracy of these values, it is necessary to take a brief look at the expectations set by the random graph model. Values for the critical exponents γ, β have been given by Erdös and Rényi (1959) as $\gamma = \beta = 1$. The exponent β found here is in excellent agreement, whereas the exponent γ deviates,

which is likely due to the poor quality of the data collapse.

Determining the expected value for the correlation length exponent is less straightforward. It is typically stated in the literature, that in the random graph model the critical exponents assume their mean-field values, which in the case of percolation read

$$\gamma = 1 \qquad \beta = 1 \qquad \bar{\nu} = \frac{1}{2} \; . \tag{7.6.1}$$

As stated above, this is immediately true for the first exponents, whereas the measured value for ν, at first glance, differs substantially. This is, however, only a problem of definitions.

For systems of low dimensionality ($d < 4$), the critical exponents have to fulfill the following hyperscaling relation (Widom, 1965):

$$d\nu = \gamma + 2\beta \tag{7.6.2}$$

However, this relation breaks down if $d \geq 4$, and for large-dimensional systems, d takes the value of the upper critical dimension, which is $d_c = 6$ (see, e.g., Saberi (2015)), where all critical exponents assume their mean-field values. Turning again to the hyperscaling relation, it is noticed that in this case the equation takes the value $\gamma + 2\beta = d_c\nu = 3$, which is close to the value measured here for the correlation length exponent.

Finite size scaling relations on networks are typically expressed in terms of the number of nodes N, as there is no lattice with fixed sites. On a lattice, scaling is expressed in terms of the lattice dimension L. However, comparing the two quantities, it is clear that N is more closely related to the lattice volume $V = L^d$, as these two quantities express the number of sites (or bonds) in the system. By this logic, it appears sensible that scaling with graph size should be expressed as

$$N^{1/\nu} = \left(N^{1/d}\right)^{1/\bar{\nu}} \tag{7.6.3}$$

with $\nu = d\bar{\nu}$ and $N^{1/d}$ resembling the linear dimension of a graph, similar to $L = V^{1/d}$ on a lattice. Thus, since the critical exponent mean-field values given above are typically derived with reference to a lattice, they should correspond to the linear dimension, rather than to N. Now, since in percolation $d = d_c = 6$ if $d \geq 6$, and since the dimension of a large random graphs is typically greater than d_c, a mean-field exponent $\bar{\nu} = 1/2$ leads to $\nu = d_c\bar{\nu} = 3$, which is the value measured here, indicating correctness of the result under the assumption that the graphs simulated here cannot be embedded in spaces with fewer than 6 dimensions.

Values of ν close to three have been observed in other networks, for instance in scale free networks if the exponent characterizing the degree distribution is larger than three (Cohen et al., 2002; Radicchi and Castellano, 2015), which also places them in the random network regime. Although the degree distribution is still governed by a power law, there are fewer hub-nodes, reducing their influence on the structure of the network.

In summary, inspection of the example confirms, that, regardless of accuracy, the GRFS model as studied here shares an universality class with the random graph model, with both models having critical exponents corresponding to the percolation mean-field case.

Returning to figure 7.6.1, panel (c) shows the data collapse of $< s >$, which has been performed via the second method, i.e., by using $< k_c >$, γ, β, ν as optimization parameters. As can be seen in the figure, the quality of the data collapse is much better, and the critical exponent $\gamma = 0.94$ is closer to the mean-field case. However, the percolation transition is shifted to a slightly larger value ($< k_c >= 1.08$) and the critical exponent $\nu = 2.63$ now lies considerably lower. Not shown here is the data collapse for S employing the same method. The quality of the collapse is high, indistinguishable from the collapse in panel (a), but the associated critical exponent takes a significantly smaller value ($\beta = 0.75$).

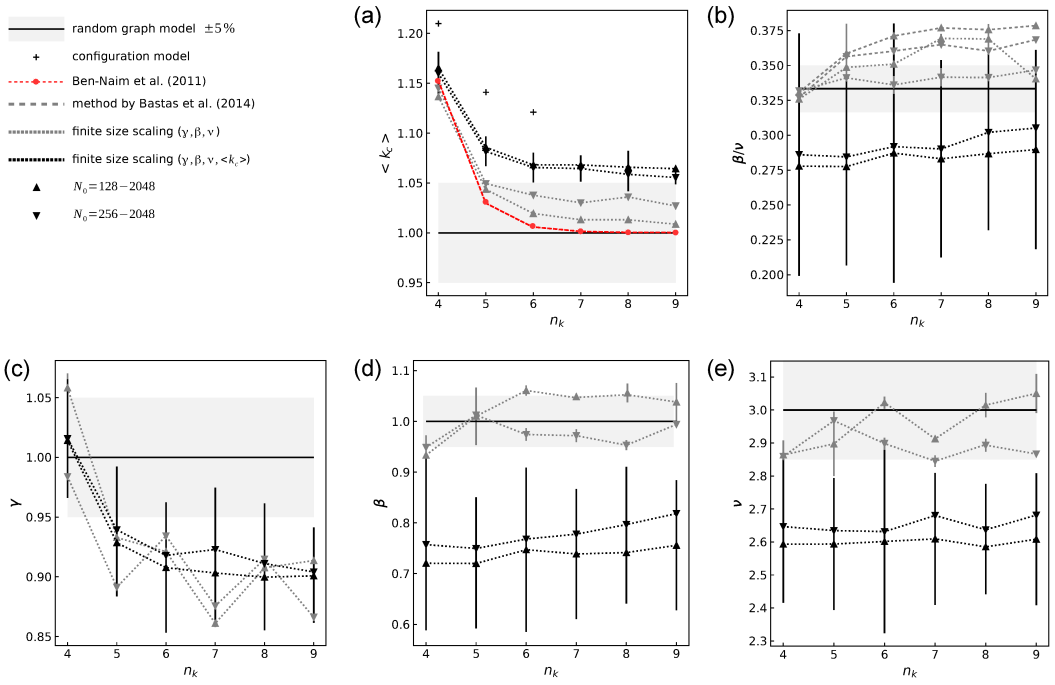


Figure 7.6.2: Influence of the degree bound on critical point and exponents. Determining the critical point prior to performing the data collapse (gray lines) leads to results that are in closer agreement with the mean-field critical exponents, but results in a poorer quality for the data collapse. Determining all parameters as part of the data collapse (black lines) leads to deviations for β and ν, although the quality of the data collapse is higher (compare figure 7.6.1).

In figure 7.6.2, the critical point and critical exponents are given for $n_k = 4 - 9$. Results are shown for both procedures outlined above and compared to the random graph model. It is typically feasible to exclude too small system sizes from the analysis. In order

to the effect, data are shown for all system sizes, and for all system sizes excluding the smallest size. Panel (a) shows the dependence of the critical point on the degree bound. On the qualitative level, all curves show the same trend: if bounded at a smaller degree, the critical point is shifted to larger degrees. This effect is discussed in greater detail in section 4.3.1. Results obtained via the method by (Bastas et al., 2014) are typically less shifted in comparison to results obtained via the data collapse method. Excluding the smallest system size from the analysis leads to larger shifts. The configuration model predicts a shift that is larger than measured by the described methods in computational data, whereas critical points obtained numerically by Ben-Naim and Krapivsky (2011) are shifted less strongly. At present, no explanation for the observed discrepancy is available.

Panels (c-e) display values obtained for the critical exponents, and panel (b) shows the ratio β/ν. It has been chosen to show this ratio separately, as the method by Bastas et al. (2014) returns an estimate of this measure in addition to the critical point. Critical exponent measurements share a pattern. Exponents obtained via the first procedure, i.e., with $< k_c >$ being fixed prior to the data collapse, are typically very close (rel. deviations of up to 10 %, but typically below 5 %) to the mean-field values, whereas exponents where all parameters have been determined via the data collapse deviate systematically. Focusing first on results obtained with $< k_c >$ fixed, it is unclear, in particular for β and ν, whether there are any systematic effects. As the random graph model represents the mean-field case, and as larger n_k imply a higher degree of similarity between the present case and the random graph model, one would expect exponents to be closest to the mean-field value for large n_k. For β and ν, there are no systematic effects visible, and their ratio β/ν deviates most significantly from the mean-field value for large n_k. The same is observed for γ, which lies systematically below the mean-field value and is closest for $n_k = 4$. Quality of the data collapses in the different methods is similar to the respective examples shown above.

Focusing on the case, where all parameters are determined by the data collapse, the exponents β, ν and the ratio β/ν are found to be below the mean-field value and below values obtained via the first method. For the aforementioned values, only very slight changes are observed upon variation of n_k, which are, in all cases, below the expected error. The exponent γ behaves roughly similar to results for fixed $< k_c >$ and deviates most significantly for large n_k.

Comparing results with all system sizes included to those, where the smallest system size has been excluded hints at a possible explanation: in the second method, results move closer to the mean-field case if only large systems are included. This suggests that the deviation is a finite size effect that would require even larger system sizes to be considered for results to be more accurate in the second method.

■ 7.6.2 Influence of growth

The influence of growth is studied in figure 7.6.3 in simulations of growing networks with $n_k = 4$. Again, system sizes range from 128 to 2048 nodes, albeit fewer simulations were performed for each system size, ranging from 14336 simulations in the smallest system to 896 simulations of the largest system size. The analyses included in the figure are iden-

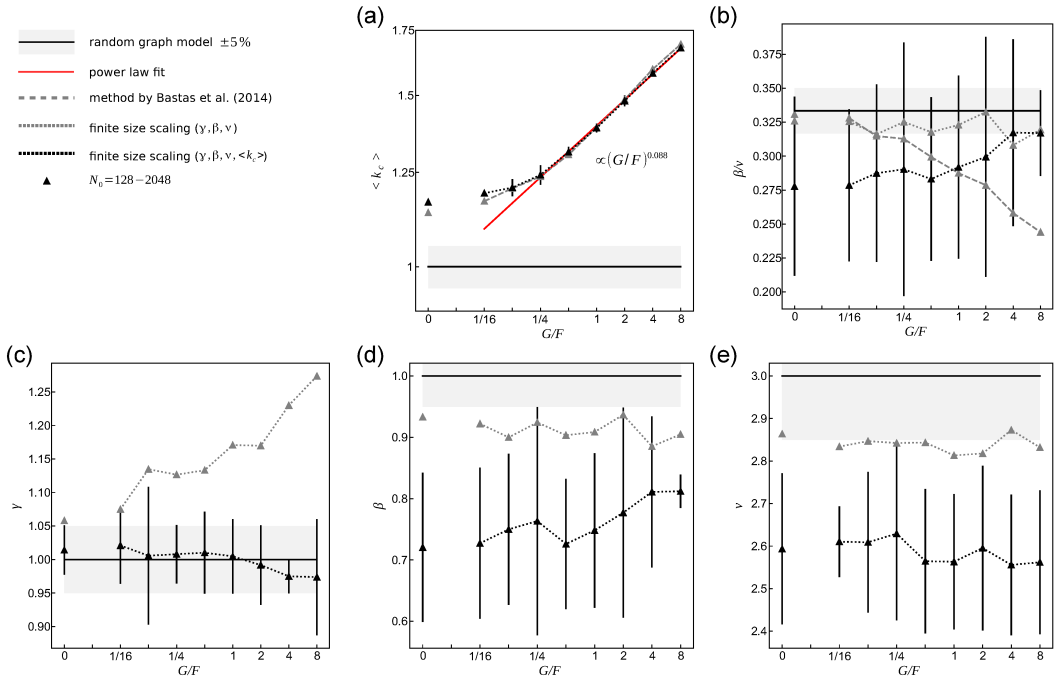


Figure 7.6.3: Influence of growth on the critical point and exponents. The shift of the critical point as a result of growth processes appears to follow a power law. If the critical point is fixed prior to the data collapse, exponents excluding γ tend to the mean-field case. Not fixing the critical point leads to deviations similar to those seen in figure 7.6.2. In this case, however, γ is in excellent agreement with the mean-field value.

tical to the previous section, with the exception that only results for one set of system sizes are shown. Excluding the smallest system size has a similar effect as in the previous section.

Panel (a) shows the influence of growth on the critical point in a double-logarithmic plot. When there is no growth present, i.e., if $G/F = 0$, results are identical to those obtained for $n_k = 4$ in the previous. Gradually increasing the growth rate leads to a shift of the transition as discussed in greater detail in section 7.5.2. Quantitative results obtained here suggest that there exists a power-law dependence of the location of the critical point on the growth by fusion ratio:

$$< k_c > \propto (G/F)^{\alpha} \tag{7.6.4}$$

The exponent as obtained by fitting is $\alpha = 0.088$. Furthermore, variation between critical point location obtained via the Bastas et al. (2014) method and the data collapse method are small. However, as will be seen below, these small variations still lead to substantial differences in the critical exponents.

Panels (b-e) are set up as in the previous section. Focusing first on β and ν, it is observed that the structure of the results is, again, similar. Fixing the critical point

through the method by Bastas et al. (2014) and obtaining exponents through the data collapse leads to exponents that are below, but within 10 % of the mean-field value with no systematic behavior attributable to variation of G/F. Determining all parameters via the data collapse has the same effect as above. β and ν are determined to be smaller than the mean-field value and the respective values obtained by the first method. Comparing to the small variations between critical points determined by either method, the observed effect highlights the sensitivity of critical exponent measurements to the correct determination of the critical point.

The ratio β/ν shows dissimilar trends for all methods. The estimate obtained by the Bastas et al. (2014) method decreases with increasing growth rates, whereas the value obtained by a subsequent data collapse is approximately constant and agrees well with the mean-field value. Values obtained purely through a data collapse show a slightly increasing trend. Judging from the accuracy of the methods, it is unclear which trend is correct. In order to provide a definitive answer, it will be necessary to include data from system sizes that are much larger than the considered data.

The same hold for results obtained for γ. In this case, the method where the critical point is determined by the data collapse leads to substantially better data collapses, accompanied by exponent values that are in excellent results for the exponent. The first method, on the other hand, shows a strong increase with growth. However, as pointed out above, in this case data collapses are of increasingly poor quality as growth increases.

In summary, the influences of a bounded degree and of system growth on the critical exponent have been further quantified. For bounded degree networks, computational results are in good agreement with results obtained by (Ben-Naim and Krapivsky, 2011), and agree qualitatively with results obtained in the configuration model, as discussed in chapter 4. Further, it has been found that the dependence of the critical point on the growth rate takes the form of a power law dependence for larger growth rates.

Overall, the critical exponents found by alternative methods are in good agreement with the mean-field values that are expected for the random graph model, indicating a shared universality class. Some trends are observed, although it is not entirely clear which method can be trusted. For more reliable results, it will become necessary to consider systems with a significantly larger number of nodes.

It is important to point out that not all graphs share the same universality class. For instance, the behavior of scale free networks is characterized by critical exponents that depend on the exponent describing the degree distribution (Cohen et al., 2002). Critical exponents primarily depend on the dimensionality of the system. In graphs, the dimension of an embedding space is typically very large if the graph is highly connected, indicating that critical exponents tend to their mean-field values. However, in graphs that are not highly connected, or in those that have been constructed specifically in some embedding space (for instance, planar graphs), it is expected that the critical exponents take different values. Furthermore, it is unclear how the possibility of an infinite order phase transition induced by system growth (see section 7.5.2) influences the

critical exponents, and, in particular, the accuracy with which they can be determined.

7.6.3 Finite size scaling with *P. polycephalum* rate constants and in two dimensions

In the previous section, influences of a degree bound and of growth on the critical point and exponents have been investigated in the GRFS model (see section 7.4.1). Critical exponents are in good agreement with the mean-field values expected for percolation on high-dimensional graphs. The critical point is shifted to larger values systematically, either if the degree distribution is bounded at smaller maximal node degrees or if there the growth rate of the system is increased. To complement the study of critical exponents, in the present section two exemplary analyses are presented, in which critical exponents are determined for simulations with rate constants as obtained from *P. polycephalum* network growth, and for two-dimensional simulations of the GRFS model.

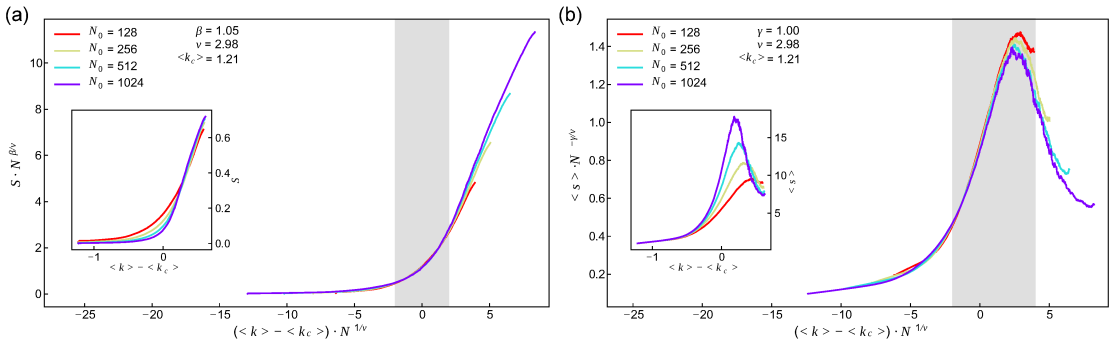


Figure 7.6.4: Finite size scaling with rate constants obtained from experimental data (section 5.3). Simulating the growth of *P. polycephalum* in the non-spatial network model leads to a percolation transition characterized by the mean-field critical exponents. Insets show the original data. The data collapse has been computed in the shaded region.

Simulations of *P. polycephalum* growth were performed using the stochastic network simulation algorithm (algorithm 7.3.1) with system sizes ranging from $128 - 1024$ nodes, and data were averaged over $12288 - 2048$ realizations, with more realizations performed for smaller system sizes. Rate constants were obtained as described in section 5.3 by tracking of topological changes. Rate constants are averages over full experiments and over the entire data set of growth experiments as detailed in section 6.1.2. Therefore, the simulations correspond to the topological evolution shown in figure 6.4.3.

Results are shown in figure 7.6.4, which is set up similar to figure 7.6.1. Shown are data collapses for the giant component size and for the size of average, non-giant components. Original data are shown as insets. The data collapses were obtained by fixing the critical point using the method by Bastas et al. (2014), and subsequent collapsing of the date using the method outlined in section 2.1.3. The critical exponents ($\gamma = 1.00$, $\beta = 1.05$, $\nu = 2.98$) are in excellent agreement with the mean-field critical exponents found in high-dimensional graphs ($\gamma = \beta = 1$, $\nu = 3$, see previous section).

This result is not too surprising, as similar results have been obtained in the GRFS model for very different growth rates.

Determined via the method of Bastas et al. (2014), the critical point is located at $< k_c >= 1.21$, which is significantly smaller than values found in experimental data (see section 6.2.2). However, it is shifted to larger values in comparison to the random graph model due to the effect of growth present in the slime mold network. Possible reasons for the observed discrepancy include systematic deviations due to finite size effects, as simulations were performed only for a limited range of system sizes. Observing the original data in figure 7.6.4 panel (b), this appears unlikely, as for larger system sizes the peaks of the average component size tend towards the determined critical point. It more likely that the shift is due to the non-spatial implementation of the simulations, as addressed in sections 6.2.2 and 7.3.4. Briefly, a network that may develop global connections between any pair of nodes is bound to have a giant component at a much smaller degree of local structure when compared to a network that can only grow locally.

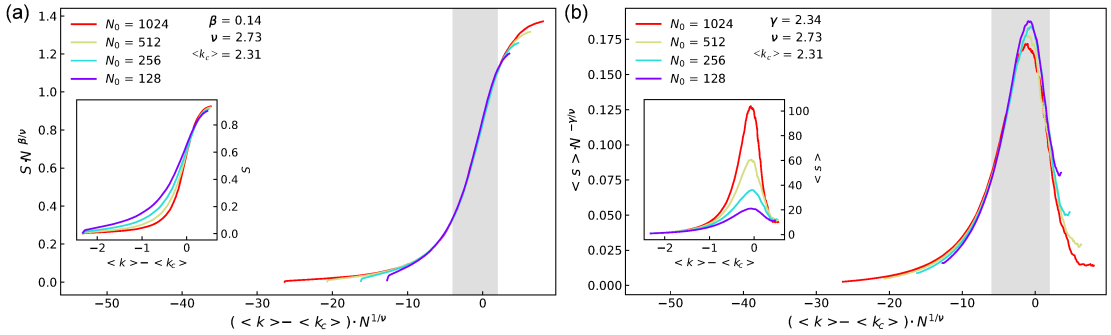


Figure 7.6.5: Finite size scaling of the spatial GRFS model. The two-dimensional implementation of the GRFS model exhibits a percolation transition characterized by the critical exponents for two-dimensional percolation. Notably, the percolation threshold is shifted considerably in comparison to the non-spatial implementation.

This hypothesis is confirmed qualitatively by the second exemplary study shown in figure 7.6.5. In section 7.3.4, first steps have been taken toward implementing the network growth model devised in this work in two dimensions. Although the model is not studied here systematically, it has been possible to confirm using a small set of simulations that the critical exponents in this model agree well with the standard exponents for percolation in two dimensions. Briefly, simulations were performed for system sizes ranging from $128 - 1024$ nodes, with small repetition numbers ranging from $4096 - 512$, again, with the larger number of simulations performed at small system size. Simulations have not been performed with parameters obtained from experimental data, but within the GRFS model (see section 7.4.1) with only growth enabled. As in the two-dimensional implementation nodes fuse when encountering one another rather than by a stochastic process described by a rate, fusion is also present in the system. Monitoring fusion rates during simulations with the present set of parameters indicates that there are approximately as many fusion events as there are growth events, making the results in figure

7.6.5 comparable to those in figure 7.6.3 for $G/F = 1$.

The data collapse shown in figure 7.6.5 has been obtained by the same method as in the slime mold case, with the critical point fixed prior to the data collapse. The critical point is determined by the method described by Bastas et al. (2014) to be at $< k_c >= 2.32$, which is substantially larger than observed in the non-spatial case at a similar growth to fusion ratio. Turning to the critical exponents, the values determined by the data collapse ($\gamma = 2.34$, $\beta = 0.14$, $\nu = 2.73$) agree well with the critical exponents of two dimensional percolation ($\gamma = 43/18 \approx 2.39$, $\beta = 5/36 \approx 0.139$, $d\bar{\nu} = 2{\cdot}4/3 \approx 2.66$) (see, e.g., Isichenko (1992) for tabulated values of critical exponents in percolation).

In principle, this result is not too surprising. In other two dimensional graphs (e.g., Melchert (2013); Norrenbrock (2016)) critical exponents are found to agree well with the exponents for two-dimensional percolation. However, percolation on these graphs works very differently. Similar to a lattice, all nodes are present throughout the entire process, and the driving parameter reflects the fraction of edges present at a given state. In the present example, neither the number, nor the position of nodes is fixed. Furthermore, in planar graphs similar to the examples given above and in lattices, any edge that will be present in the final configuration may be placed at any time, whereas in the present case clusters may only grow at their perimeter. This leads to a delay of the percolation transition as discussed in section 6.2.2.

Although no two-dimensional simulations have been performed with *P. polycephalum* rate constant data, it is expected that the critical exponents governing the growth process are identical to those found in the planar GRFS model.

Chapter 8

Summary

In this work, network formation in the slime mold *P. polycephalum* after fragmentation has been studied in the light of percolation. Unlike most processes exhibiting standard percolation, *P. polycephalum*'s network formation is an active process, driven by growth, movement and deformation of fragments. Standard models for percolation on graphs, such as the random graph and configuration models describe percolation as a property of static graphs whose structure is characterized by a driving parameter. Comparison of experimental data with percolation in the aforementioned models validates this concern: development of the giant component as described in the model does not match well with experimental results.

This observation has guided the present study. Recapitulating the steps taken towards the elucidation of the network formation process, three interrelated core topics have been pursued here that will be briefly summarized below. First, however, it appears reasonable to further exemplify the relation between these topics. These were, on one hand, data analysis and the analytical description of percolation within the configuration and random graph models, and, on the other hand, modeling of network formation and percolation as a stochastic process, in order to investigate the role of elementary processes such as growth for percolation.

Underlying data analysis, a novel method and graph model for abstracting spatial networks has been proposed. Although being more elegant, accurate and extending the set of possible directions during data analysis, the graph model is not an essential requirement for studying graph-based percolation in *P. polycephalum* and comparing to theoretical models. This has been shown in early stages of this project (Fessel et al., 2012, 2015*a*,*b*). Instead, the need for this method is prompted by the third core topic, which is modeling of network formation as a stochastic process. Network formation can be understood as a sequence of elementary processes that modify the topology of a graph. Possible processes include changes in connectivity as indicated by the set of edges, and growth of the network implies changes in the set of nodes. It has turned out that correctly representing these processes in a stochastic framework makes it necessary to account for nodes of degree two, which are typically poorly defined in spatial network representations.

The graph model proposed as part of this work brings a number of advantages over

the topological skeleton widely used for object representation. First, it is more robust as it effectively filters out local variations in object shape, and second, it represents not only the topology but also lateral extensions of the object. Third, it provides a clear definition of second degree nodes, which henceforth function as a measure for the density of the spatial network. Finally, as nodes are extended structures, tracking topological changes becomes less problematic. In *P. polycephalum* these advantages make it possible to distinguish between functionally different regions such as growth fronts and veins, and to discretize the slime molds spatio-temporal development.

Focal aspects pursued during data analysis encompassed, on one hand, the description and quantification of growth phases, and, on the other hand, the investigation of topological evolution. Four phases of growth have been distinguished. Briefly, during the first phase termed aggregation, microplasmodia move, driven by active and passive mechanisms and form extended but compact structures. The second phase is network formation, and it is the phase in which the percolation transition occurs. Aggregates of microplasmodia extend and fuse, leading to the emergence of a connected network structure. Following this, in the third phase the connected structure explores its surroundings in a rapid, typically directed motion. Finally, upon halting the slime mold develops a global wave pattern with a long period that drives further expansion of the network. The times at which *P. polycephalum* transitions between phases are shared between experiments.

Studying the topological evolution has led to the discovery of two prominent points that are closely related to the phases detailed above. Prior to the critical point of percolation, the system is fragmented and changes only locally. Following percolation, the slime mold soon develops a global pattern, leading to expansion and foraging. During this phase, the degree distribution tends to a steady state, in which it remains, although the slime mold constantly remodels its network. The final degree distribution is dominated by second and third degree nodes, with a small number of higher degree nodes localized to growth regions where the slime mold forms sheet-like structures. The lack of large degree nodes during the network formation process justifies treatment of the system as a small degree network in which these nodes are considered negligible.

Making use of the small degree property, it has been possible to solve the random graph and configuration models analytically. Solving the random graph model is possible as long as nodes of degree six and higher are negligible, and the configuration model can be solved if the majority of node degrees remains below degree seven. Limiting or cutting of the degree distribution at some maximal degree leads to a shift of the location of the percolation transition in either model. However, comparing the analytical solutions to experimental data led to the observation stated above. Percolating *P. polycephalum* networks require a significantly higher degree of structure to enter the critical region. Furthermore, finite size effects, although present in the system and a frequent candidate for deviations from ideal percolation behavior, could be ruled out as main source of the deviation, indicating that other mechanisms are responsible.

All further efforts have been pursued with the goal of clarifying the observed behav-

ior. In order to be able to distinguish the effects of fusion, growth and their reversals, a model has been constructed, building on the concept that interactions between nodes resemble chemical reactions, with one rate constant characterizing each possible reaction. The model has been implemented as a stochastic master equation and as a rate equation describing the behavior of the system in the thermodynamical limit. Whereas treatment and analysis of the rate equation has been possible by explicit numerical methods, the master equation was simulated using a proposed algorithm for simulating network development, which is based on the stochastic simulation algorithm. Therefore, it generates exact trajectories of the master equation and simultaneously constructs graphs in accordance with this development, from which, for instance, network measures of interest to the study of percolation can be obtained.

The strength of the model lies in its versatility stemming from flexible choice of the rate constants. Reducing the system to one single rate constant reproduces the random graph model, whereas a selection of constants representing preferential attachment leads to the emergence of scale free topologies. Comparison to experimental data shows that for an appropriate selection of rate constants, the topological dynamics of the model are in agreement with those observed in experimental data. Rate constants have been obtained from experimental data by tracking topological modifications. It has been found via a clustering method that it is possible, based on the measured rate constants, to distinguish between the phases of growth described above.

Furthermore, studying the role of elementary processes in a simplified model led to a number of conclusions. First, if no growth is present in the system, percolation in the configuration model describes percolation in computational data generated from the model. In this case, the solution of the rate equation in conjunction with the configuration model provides a deterministic model describing the behavior of the simulated master equation in the thermodynamical limit.

Employing the computational model, it has been possible to study the effect of growth on the system. Increasing the rate at which the network grows in comparison to the rate at which nodes fuse, leads to a shift of the critical point according to a power law. The shift indicates that in the presence of growth a higher degree of structure is required for the system to percolate. This is attributable mainly to local growth of components, which increases their degree of structure without undergoing fusion events with other components.

The percolation transition in computational data has been further characterized using finite size scaling. Critical exponents characterizing the system have been determined, indicating that the universality class of the simulated system is in agreement with that of mean-field percolation. Within the available data, no systematic influence on the critical exponents has been found upon variation of growth rates, or due to limitations to the degree distribution. However, preliminary simulations in a two-dimensional implementation of the simulation algorithm show that spatial embedding significantly influences the critical exponents. In this model, the critical exponents have been found to agree well with the expectation for two-dimensional percolation. As the spatial simulation shares

important characteristics with percolation in *P. polycephalum*, it is conjectured that the critical exponents governing *P. polycephalum* percolation are identical to those found by the computational method.

Bibliography

Achlioptas, D., D'Souza, R. M. and Spencer, J. (2009), 'Explosive percolation in random networks', *Science* **323**(5920), 1453–1455.

Adamatzky, A., ed. (2016), *Advances in Physarum Machines*, Springer International Publishing.

Adamatzky, A. and Schubert, T. (2014), 'Slime mold microfluidic logical gates', *Mater. Today* **17**(2), 86–91.

Akita, D., Kunita, I., Fricker, M. D., Kuroda, S., Sato, K. and Nakagaki, T. (2016), 'Experimental models for murray's law', *J. Phys. D: Appl. Phys.* **50**(2), 024001.

Akita, D., Schenz, D., Kuroda, S., Sato, K., ichi Ueda, K. and Nakagaki, T. (2017), 'Current reinforcement model reproduces center-in-center vein trajectory of physarum polycephalum', *Dev. Growth. Differ.* **59**(5), 465–470.

Albert, Jeong and Barabasi (2000), 'Error and attack tolerance of complex networks', *Nature* **406**, 378–382.

Albert, R. and Barabási, A.-L. (2000), 'Topology of evolving networks: Local events and universality', *Phys. Rev. Lett.* **85**(24), 5234–5237.

Albert, R. and Barabasi, A.-L. (2002), 'Statistical mechanics of complex networks', *Rev. Mod. Phys.* **74**(1), 47–97.

Alim, K. (2018), 'Fluid flows shaping organism morphology', *Philos. Trans. Royal Soc. B* **373**(1747), 20170112.

Alim, K., Amselem, G., Peaudecerf, F., Brenner, M. P. and Pringle, A. (2013), 'Random network peristalsis in physarum polycephalum organizes fluid flows across an individual', *Proc. Natl. Acad. Sci. U.S.A.* **110**(33), 13306–13311.

Alim, K., Andrew, N., Pringle, A. and Brenner, M. P. (2017), 'Mechanism of signal propagation in physarum polycephalum', *Proc. Natl. Acad. Sci. U.S.A.* **114**(20), 5136–5141.

Alonso, S., Radszuweit, M., Engel, H. and Bär, M. (2017), 'Mechanochemical pattern formation in simple models of active viscoelastic fluids and solids', *J. Phys. D: Appl. Phys.* **50**(43), 434004.

Alonso, S., Strachauer, U., Radszuweit, M., Bär, M. and Hauser, M. J. (2016), 'Oscillations and uniaxial mechanochemical waves in a model of an active poroelastic medium: Application to deformation patterns in protoplasmic droplets of physarum polycephalum', *Physica D* **318-319**, 58–69.

Andrade, J. S., Herrmann, H. J., Andrade, R. F. S. and da Silva, L. R. (2005), 'Apollonian networks: Simultaneously scale-free, small world, euclidean, space filling, and with matching graphs', *Phys. Rev. Lett.* **94**(1).

Angeli, D. (2009), 'A tutorial on chemical reaction network dynamics', *Eur. J. Control* **15**(3-4), 398–406.

Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y. and Zhou, C. (2008), 'Synchronization in complex networks', *Phys. Rep.* **469**(3), 93–153.

Atkins, P., De Paula, J. and Keeler, J. (2018), *Atkins' Physical Chemistry*, Oxford University Press.

Bai, X., Latecki, L. and yu Liu, W. (2007), 'Skeleton pruning by contour partitioning with discrete curve evolution', *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(3), 449–462.

Barabási, A.-L. and Albert, R. (1999), 'Emergence of scaling in random networks', *Science* **286**(5439), 509–512.

Barabási, A.-L. and Oltvai, Z. N. (2004), 'Network biology: understanding the cell's functional organization', *Nat. Rev. Genet.* **5**(2), 101–113.

Barthelemy, M. (2018), *Morphogenesis of Spatial Networks*, Springer International Publishing.

Barthelemy, M., Bordin, P., Berestycki, H. and Gribaudi, M. (2013), 'Self-organization versus top-down planning in the evolution of a city', *Sci. Rep.* **3**(1).

Barthélemy, M. (2010), 'Spatial networks', *Phys. Rep.* **499**(1–3), 1–101.

Bastas, N., Kosmidis, K., Giazitzidis, P. and Maragakis, M. (2014), 'Method for estimating critical exponents in percolation processes with low sampling', *Phys. Rev. E* **90**(6).

Baumgarten, W. and Hauser, M. J. (2010), 'Detection, extraction, and analysis of the vein network of the slime mould physarum polycephalum', *J. Comput. Interdiscip. Sci.* **1**(3).

Baumgarten, W. and Hauser, M. J. B. (2013), 'Functional organization of the vascular network of physarum polycephalum', *Phys. Biol.* **10**(2), 026003.

Baumgarten, W. and Hauser, M. J. B. (2014), 'Dynamics of frontal extension of an amoeboid cell', *Europhys. Lett.* **108**(5), 50010.

Baumgarten, W., Ueda, T. and Hauser, M. J. B. (2010), 'Plasmodial vein networks of the slime moldPhysarum polycephalumform regular graphs', *Phys. Rev. E* **82**(4).

Beekman, M. and Latty, T. (2015), 'Brainless but multi-headed: Decision making by the acellular slime mould physarum polycephalum', *J. Mol. Biol.* **427**(23), 3734–3743.

Ben-Naim, E. and Krapivsky, P. L. (2011), 'Dynamics of random graphs with bounded degrees', *J. Stat. Mech: Theory Exp.* **2011**(11), P11008.

Bernitt, E., Oettmeier, C. and Döbereiner, H. G. (2010), Microplasmodium dynamics of physarum polycephalum, *in* 'IFMBE Proceedings', Springer Berlin Heidelberg, pp. 1133–1136.

Binder, K. (1997), 'Applications of monte carlo methods to statistical physics', *Rep. Prog. Phys.* **60**(5), 487–559.

Blum, H. (1967), *Models for the Perception of Speech and Visual Form*, MIT Press, Cambridge, chapter A Transformation for Extracting New Descriptors of Shape, pp. 362–380.

Boccaletti, S., Latora, V., Chavez, M. and Hwang, D.-U. (2006), 'Complex networks: structure and dynamics', *Phys. Rep.* **424**, 175–308.

Boisseau, R. P., Vogel, D. and Dussutour, A. (2016), 'Habituation in non-neural organisms: evidence from slime moulds', *Proc. Royal Soc. B* **283**(1829), 20160446.

Bollobás, B., Janson, S. and Riordan, O. (2004), 'The phase transition in the uniformly grown random graph has infinite order', *Random Struct. Algor.* **26**(1-2), 1–36.

Bonifaci, V., Mehlhorn, K. and Varma, G. (2012), 'Physarum can compute shortest paths', *J. Theor. Biol.* **309**, 121–133.

Borg, I. and Groenen, P. J. F. (2005), *Modern Multidimensional Scaling*, Springer New York.

Borgefors, G. (1986), 'Distance transformations in digital images', *Computer Vision, Graphics, and Image Processing* **34**(3), 344–371.

Broadbent, S. R. and Hammersley, J. M. (1957), 'Percolation processes', *Math. Proc. Cambridge Philos. Soc.* **53**(03), 629.

Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E. J. and Strogatz, S. H. (2001), 'Are randomly grown graphs really random?', *Phys. Rev. E* **64**(4).

Charras, G. and Paluch, E. (2008), 'Blebs lead the way: how to migrate without lamellipodia', *Nat. Rev. Mol. Cell Biol.* **9**(9), 730–736.

Chen, L., Wang, R., Li, C. and Aihara, K. (2010), *Modeling biomolecular networks in cells: structures and dynamics*, Springer.

Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009), 'Power-law distributions in empirical data', *SIAM Rev.* **51**(4), 661–703.

Cohen, R., ben Avraham, D. and Havlin, S. (2002), 'Percolation critical exponents in scale-free networks', *Phys. Rev. E* **66**(3).

Cohen, R., Erez, K., ben Avraham, D. and Havlin, S. (2000), 'Resilience of the internet to random breakdowns', *Phys. Rev. Lett.* **85**(21), 4626–4628.

da F. Costa, L., Rodrigues, F. A., Travieso, G. and Boas, P. R. V. (2007), 'Characterization of complex networks: A survey of measurements', *Adv. Phys.* **56**(1), 167–242.

Daniel, J. W. and Rusch, H. P. (1961), 'The pure culture of physarum polycephalum on a partially defined soluble medium', *J. Gen. Microbiol.* **25**(1), 47–59.

Davidsen, J., Ebel, H. and Bornholdt, S. (2002), 'Emergence of a small world from local interactions: Modeling acquaintance networks', *Phys. Rev. Lett.* **88**(12).

de Solla Price, D. J. (1965), 'Networks of scientific papers', *Science* **149**(3683), 510–515.

Dinh, K. N. and Sidje, R. B. (2016), 'Understanding the finite state projection and related methods for solving the chemical master equation', *Phys. Biol.* **13**(3), 035003.

Dirnberger, M. and Mehlhorn, K. (2017), 'Characterizing networks formed by p. polycephalum', *J. Phys. D: Appl. Phys.* **50**(22), 224002.

Dormand, J. R. and Prince, P. J. (1980), 'A family of embedded runge-kutta formulae', *J. Comput. Appl. Math.* **6**(1), 19–26.

Dorogovtsev, S. N., Goltsev, A. V. and Mendes, J. F. F. (2008), 'Critical phenomena in complex networks', *Rev. Mod. Phys.* **80**(4), 1275–1335.

Dorogovtsev, S. N. and Mendes, J. F. F. (2002), 'Evolution of networks', *Adv. Phys.* **51**(4), 1079–1187.

Dorogovtsev, S. N. and Mendes, J. F. F. (2003), *Evolution of networks: from biological nets to the internet and www*, Oxford University Press, Oxford.

Dorogovtsev, S. N., Mendes, J. F. F. and Samukhin, A. N. (2001), 'Anomalous percolation properties of growing networks', *Phys. Rev. E* **64**(6).

Dousse, O., Mannersalo, P. and Thiran, P. (2004), Latency of wireless sensor networks with uncoordinated power saving mechanisms, *in* 'Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '04', ACM Press.

D'Souza, R. M. (2017), 'Curtailing cascading failures', *Science* **358**, 860–861.

D'Souza, R. M. and Nagler, J. (2015), 'Anomalous critical and supercritical phenomena in explosive percolation', *Nat. Phys.* **11**(7), 531–538.

Duda, R. O., Hart, P. E. and G., S. D. (2001), *Pattern classification*, John Wiley & Sons, New York, chapter Unsupervised learning and clustering.

Durham, A. C. (1976), 'Control of chemotaxis in physarum polycephalum', *J. Cell Biol.* **69**(1), 218–223.

Dussutour, A., Latty, T., Beekman, M. and Simpson, S. J. (2010), 'Amoeboid organism solves complex nutritional challenges', *Proc. Natl. Acad. Sci. U.S.A.* **107**(10), 4607–4611.

Ebel, H., Mielsch, L.-I. and Bornholdt, S. (2002), 'Scale-free topology of e-mail networks', *Phys. Rev. E* **66**(3).

Erdös, P. and Rényi, A. (1959), 'On random graphs', *Publ. Math. Debrecen* **6**, 290–297.

Euler, L. and Hewlett, J. (2009), *Elements of Algebra*, Cambridge University Press.

Everitt, B. S. and Hand, D. J. (1981), *Finite Mixture Distributions*, Springer Netherlands.

Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999), On power-law relationships of the internet topology, *in* 'Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication - SIGCOMM '99', ACM Press.

Feder, J. (1980), 'Random sequential adsorption', *J. Theor. Biol.* **87**(2), 237–254.

Fessel, A. and Döbereiner, H.-G. (2016), Motifs of growth and fusion govern physarum polycephalum network formation, *in* 'Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies, BICT'15', ACM.

Fessel, A. and Döbereiner, H.-G. (2017), 'Nonlinear compliance of elastic layers to indentation', *Biomech. Model. Mechanobiol.* **17**(2), 419–438.

Fessel, A., Oettmeier, C., Bernitt, E., Gauthier, N. C. and Döbereiner, H.-G. (2012), 'Physarum polycephalum percolation as a paradigm for topological phase transitions in transportation networks.', *Phys. Rev. Lett.* **109**, 078103.

Fessel, A., Oettmeier, C. and Döbereiner, H.-G. (2015*a*), An analytical model for percolation in small link degree transportation networks, *in* 'Proceedings of the 8th International Conference on Bio-Inspired Information and Communications Technologies, BICT'14', ACM.

Fessel, A., Oettmeier, C. and Döbereiner, H.-G. (2015*b*), 'Structuring precedes extension in percolating physarum polycephalum networks', *Nano Commun. Networks* **6**, 87–95.

Fessel, A., Oettmeier, C., Wechsler, K. and Döbereiner, H.-G. (2017), 'Indentation analysis of active viscoelastic microplasmodia of p. polycephalum', *J. Phys. D: Appl. Phys.* **51**(2), 024005.

Figge, M. T., Reichert, A. S., Meyer-Hermann, M. and Osiewacz, H. D. (2012), 'Deceleration of fusion–fission cycles improves mitochondrial quality control during aging', *PLoS Comput. Biol.* **8**(6), e1002576.

Fisher, M. E. and Barber, M. N. (1972), 'Scaling theory for finite-size effects in the critical region', *Phys. Rev. Lett.* **28**(23), 1516–1519.

Fisher, M. E. and Essam, J. W. (1961), 'Some cluster size and percolation problems', *J. Math. Phys.* **2**(4), 609–619.

Flory, P. J. (1941), 'Molecular size distribution in three dimensional polymers. i. gelation', *J. Am. Chem. Soc.* **63**(11), 3083–3090.

Gamba, A., Ambrosi, D., Coniglio, A., de Candia, A., Talia, S. D., Giraudo, E., Serini, G., Preziosi, L. and Bussolino, F. (2003), 'Percolation, morphogenesis, and burgers dynamics in blood vessels formation', *Phys. Rev. Lett.* **90**(11).

Gardiner, C. W. (1985), *Handbook of stochastic methods*, Springer series in synergetics, Springer.

Gawlitta, W., Wolf, K. V., Hoffmann, H.-U. and Stockem, W. (1980), 'Studies on microplasmodia of physarum polycephalum', *Cell Tissue Res.* **209**(1).

Gershinowitz, H. and Eyring, H. (1935), 'The theory of trimolecular reactions', *J. Am. Chem. Soc.* **57**(6), 985–991.

Ghoshal, G., Chi, L. and Barabási, A.-L. (2013), 'Uncovering the role of elementary processes in network evolution', *Sci. Rep.* **3**(1).

Gillespie, D. T. (1977), 'Exact stochastic simulation of coupled chemical reactions', *J. Phys. Chem. A* **81**(25), 2340–2361.

Gillespie, D. T. (1992), 'A rigorous derivation of the chemical master equation', *Physica A* **188**(1-3), 404–425.

Gillespie, D. T. (2007), 'Stochastic simulation of chemical kinetics', *Annu. Rev. Phys. Chem.* **58**, 35–55.

Grassberger, P. (1983), 'On the critical behavior of the general epidemic process and dynamical percolation', *Math. Biosci.* **63**(2), 157–172.

Guy, R. D., Nakagaki, T. and Wright, G. B. (2011), 'Flow-induced channel formation in the cytoplasm of motile cells', *Phys. Rev. E* **84**(1).

Hagberg, A., Swart, P. and Schult, D. (2008), Exploring network structure, dynamics, and function using networkx, Technical report, Los Alamos National Lab. (LANL), Los Alamos, NM (United States).

Häder, D.-P. and Schreckenbach, T. (1984), 'Phototactic orientation in plasmodia of the acellular slime mold, physarum polycephalum', *Plant Cell Physiol.* .

Henley, C. L. (1989), 'Self-organized percolation: a simpler model', *Bull. Am. Phys* **34**, 838.

Henri, V. (1902), 'Théorie générale de l'action de quelques diastases', *C. R. Acad. Sci.* **135**, 916–919.

Hilfinger, A. and Paulsson, J. (2011), 'Separating intrinsic from extrinsic fluctuations in dynamic biological systems', *Proc. Natl. Acad. Sci. U.S.A.* **108**(29), 12167–12172.

Hillmich, S. (2018), Graph-based analysis of growth patterns in the true slime mold physarum polycephalum, Master's thesis, Universität Bremen.

Houdayer, J. and Hartmann, A. (2004), 'Low-temperature behavior of two-dimensional gaussian ising spin glasses', *Phys. Rev. B* **70**(1).

Hunter, J. D. (2007), 'Matplotlib: A 2d graphics environment', *Comput. Sci. Eng.* **9**(3), 90–95.

Isenberg, G. and Wohlfarth-Bottermann, K. (1976), 'Transformation of cytoplasmic actin importance for the organization of the contractile gel reticulnm and the contraction relaxation cycle of cytoplasmic actomyosin', *Cell Tissue Res.* **173**(4).

Isichenko, M. B. (1992), 'Percolation, statistical topography, and transport in random media', *Rev. Mod. Phys.* **64**(4), 961–1043.

Jeong, H., Mason, S. P., Barabási, A.-L. and Oltvai, Z. N. (2001), 'Lethality and centrality in protein networks', *Nature* **411**(6833), 41–42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. and Barabási, A.-L. (2000), 'The large-scale organization of metabolic networks', *Nature* **407**(6804), 651–654.

Jones, E., Oliphant, T., Peterson, P. and et al. (2001), 'Scipy: Open source scientic tools for python'. Online; accessed 2018-11-20.

Jones, J., Tsuda, S. and Adamatzky, A. (2011), Towards physarum robots, *in* 'Bio-Inspired Self-Organizing Robotic Systems', Springer Berlin Heidelberg, pp. 215–251.

Joos, F., Perarnau, G., Rautenbach, D. and Reed, B. (2017), 'How to determine if a random graph with a fixed degree sequence has a giant component', *Probab. Theory Relat. Fields* **170**(1-2), 263–310.

Jump, J. A. (1954), 'Studies on sclerotization in physarum polycephalum', *Am. J. Bot.* **41**(7), 561–567.

Kadanoff, L. P. (1966), 'Scaling laws for ising models near tc', *Physics Physique Fizika* **2**(6), 263–272.

Keesman, R., Lamers, J., Duine, R. A. and Barkema, G. T. (2016), 'Finite-size scaling at infinite-order phase transitions', *J. Stat. Mech: Theory Exp.* **2016**(9), 093201.

Kesten, H. (1982), *Percolation Theory for Mathematicians*, Birkhäuser Boston.

Kirkpatrick, S. (1973), 'Percolation and conduction', *Rev. Mod. Phys.* **45**(4), 574–588.

Kosterlitz, J. M. and Thouless, D. J. (1973), 'Ordering, metastability and phase transitions in two-dimensional systems', *J. Phys. C: Solid State Phys.* **6**(7), 1181–1203.

Kozma, R., Puljic, M. and Perlovsky, L. (2009), 'Modeling goal-oriented decison making through cognitive phase transitions', *New Math. Nat. Comput.* **05**(01), 143–157.

Krapivsky, P. L., Redner, S. and Ben-Naim, E. (2009), *A Kinetic View of Statistical Physics*, Cambridge University Press.

Krapivsky, P. L., Redner, S. and Leyvraz, F. (2000), 'Connectivity of growing random networks', *Phys. Rev. Lett.* **85**(21), 4629–4632.

Krapivsky, P. and Redner, S. (2003), Rate equation approach for growing networks, *in* 'Statistical Mechanics of Complex Networks', Springer Berlin Heidelberg, pp. 3–22.

Kuroda, S., Takagi, S., Nakagaki, T. and Ueda, T. (2015), 'Allometry in physarum plasmodium during free locomotion: size versus shape, speed and rhythm', *J. Exp. Biol.* **218**(23), 3729–3738.

Latty, T. and Beekman, M. (2010*a*), 'Food quality and the risk of light exposure affect patch-choice decisions in the slime mold physarum polycephalum', *Ecology* **91**(1), 22–27.

Latty, T. and Beekman, M. (2010*b*), 'Speed-accuracy trade-offs during foraging decisions in the acellular slime mould physarum polycephalum', *Proc. Royal Soc. B* **278**(1705), 539–545.

Löber, J., Ziebert, F. and Aranson, I. S. (2015), 'Collisions of deformable cells lead to collective migration', *Sci. Rep.* **5**(1).

Lee, J., Oettmeier, C. and Döbereiner, H.-G. (2018), 'A novel growth mode of physarum polycephalum during starvation', *J. Phys. D: Appl. Phys.* **51**(24), 244002.

Lemieux, C. (2009), *Monte Carlo and Quasi-Monte Carlo Sampling*, 1 edn, Springer New York.

Lewis, O. L. and Guy, R. D. (2017), 'Analysis of peristaltic waves and their role in migrating physarum plasmodia', *J. Phys. D: Appl. Phys.* **50**(28), 284001.

Lewis, O. L., Zhang, S., Guy, R. D. and del Alamo, J. C. (2015), 'Coordination of contractility, adhesion and flow in migrating physarum amoebae', *J. R. Soc. Interface* **12**(106), 20141359–20141359.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. and Åberg, Y. (2001), 'The web of human sexual contacts', *Nature* **411**(6840), 907–908.

Lloyd, S. (1982), 'Least squares quantization in PCM', *IEEE Trans. Inf. Theory* **28**(2), 129–137.

Lämmermann, T. and Sixt, M. (2009), 'Mechanical modes of 'amoeboid' cell migration', *Curr. Opin. Cell Biol.* **21**(5), 636–644.

Mahnke, R., Kaupužs, J. and Lubashevsky, I. (2009), *Physics of stochstic processes*, Wiley-VCH.

Mandelbrot, B. B. (1982), *The fractal geometry of nature*, W. H. Freman and Co., New York.

Mann, H. B. and Whitney, D. R. (1947), 'On a test of whether one of two random variables is stochastically larger than the other', *Ann. Math. Stat.* **18**(1), 50–60.

Marbach, S., Alim, K., Andrew, N., Pringle, A. and Brenner, M. P. (2016), 'Pruning to increase taylor dispersion in physarum polycephalum networks', *Phys. Rev. Lett.* **117**(17).

McCullough, C. H. R. and Dee, J. (1976), 'Defined and semi-defined media for the growth of amoebae of physarum polycephalum', *J. Gen. Microbiol.* **95**(1), 151–158.

McQuarrie, D. A. (1967), 'Stochastic approach to chemical kinetics', *J. Appl. Probab.* **4**(3), 413–478.

Melchert, O. (2009), 'autoscale.py - a program for automatic finite-size scaling analyses: A user's guide', *ArXiv* .

Melchert, O. (2013), 'Percolation thresholds on planar euclidean relative-neighborhood graphs', *Phys. Rev. E* **87**(4).

Michaelis, L. and Menten, M. L. (1913), 'Die kinetik der invertinwirkung', *Biochem. Z.* **49**, 333–369.

Milgram, S. (1967), 'The small world problem', *Psychology today* **2**, 60–67.

Molloy, M. and Reed, B. (1995), 'A critical point for random graphs with a given degree sequence', *Random Struct. Algor.* **6**, 161–180.

Munsky, B. and Khammash, M. (2006), 'The finite state projection algorithm for the solution of the chemical master equation', *J. Chem. Phys.* **124**(4), 044104.

Nagler, J., Tiessen, T. and Gutch, H. W. (2012), 'Continuous percolation with discontinuities', *Phys. Rev. X* **2**(3).

Nakagaki, T., Kobayashi, R., Nishiura, Y. and Ueda, T. (2004), 'Obtaining multiple separate food sources: behavioural intelligence in the physarum plasmodium', *Proc. Royal Soc. B* **271**(1554), 2305–2310.

Nakagaki, T., Yamada, H. and Ito, M. (1999), 'Reaction diffusion advection model for pattern formation of rhythmic contraction in a giant amoeboid cell of the physarum plasmodium', *J. Theor. Biol.* **197**(4), 497–506.

Nakagaki, T., Yamada, H. and Tóth, Á. (2000), 'Maze-solving by an amoeboid organism', *Nature* **407**(6803), 470–470.

Nakagaki, T., Yamada, H. and Ueda, T. (2000), 'Interaction between cell shape and contraction pattern in the physarum plasmodium', *Biophys. Chem.* **84**(3), 195–204.

Newman, M. E. J. (2003), 'The structure and function of complex networks', *SIAM Rev.* **45**(2), 167–256.

Newman, M. E. J. (2007), 'Component sizes in networks with arbitrary degree distributions.', *Phys. Rev. E* **76**, 045101.

Newman, M. E. J., Strogatz, S. H. and Watts, D. J. (2001), 'Random graphs with arbitrary degree distributions and their applications', *Phys. Rev. E* **64**(2), 026118.

Newman, M. E. J. and Ziff, R. M. (2000), 'Efficient monte carlo algorithm and high-precision results for percolation', *Phys. Rev. Lett.* **85**(19), 4104–4107.

Nordsieck, A., Lamb Jr, W. and Uhlenbeck, G. (1940), 'On the theory of cosmic-ray showers i the furry model and the fluctuation problem', *Physica* **7**(4), 344–360.

Norrenbrock, C. (2016), 'Percolation threshold on planar euclidean gabriel graphs', *Eur. Phys. J. B* **89**(5).

Ntinas, V., Vourkas, I., Sirakoulis, G. C. and Adamatzky, A. I. (2017), 'Modeling physarum space exploration using memristors', *J. Phys. D: Appl. Phys.* **50**(17), 174004.

Oettmeier, C., Brix, K. and Döbereiner, H.-G. (2017), 'Physarum polycephalum—a new take on a classic model system', *J. Phys. D: Appl. Phys.* **50**(41), 413001.

Oettmeier, C., Lee, J. and Döbereiner, H.-G. (2018), 'Form follows function: ultrastructure of different morphotypes of physarum polycephalum', *J. Phys. D: Appl. Phys.* **51**(13), 134006.

Oh, K. W., Lee, K., Ahn, B. and Furlani, E. P. (2012), 'Design of pressure-driven microfluidic networks using electric circuit analogy', *Lab. Chip* **12**(3), 515–545.

Okabe, A., Boots, B., Sugihara, K., Chiu, S. N. and Kendall, D. G. (2008), *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2. edn, John Wiley & Sons, Inc.

Oliphant, T. E. (2006), *A guide to NumPy*, Trelgol Publishing.

Otsu, N. (1979), 'A threshold selection method from gray-level histograms', *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66.

Pastor-Satorras, R. and Vespignani, A. (2001), 'Epidemic spreading in scale-free networks.', *Phys. Rev. Lett.* **86**, 3200–3203.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011), 'scikit-learn: machine learning in python', *J. Mach. Learn. Res.* **12**, 2825–2830.

Pershin, Y. V., Fontaine, S. L. and Ventra, M. D. (2009), 'Memristive model of amoeba learning', *Phys. Rev. E* **80**(2).

Pershin, Y. V. and Ventra, M. D. (2013), 'Self-organization and solution of shortest-path optimization problems with memristive networks', *Phys. Rev. E* **88**(1).

Prehofer, C. and Bettstetter, C. (2005), 'Self-organization in communication networks: principles and design paradigms', *IEEE Commun. Mag.* **43**(7), 78–85.

Prost, J., Jülicher, F. and Joanny, J.-F. (2015), 'Active gel physics', *Nat. Phys.* **11**(2), 111–117.

Radicchi, F. and Castellano, C. (2015), 'Breaking of the site-bond percolation universality in networks', *Nat. Commun.* **6**(1).

Radszuweit, M., Engel, H. and Bär, M. (2014), 'An active poroelastic model for mechanochemical patterns in protoplasmic droplets of physarum polycephalum', *PLoS ONE* **9**(6), e99220.

Rafelski, S. M. (2013), 'Mitochondrial network morphology: building an integrative, geometrical view', *BMC Biol.* **11**(1), 71.

Redner, S. (1998), 'How popular is your paper? an empirical study of the citation distribution', *Eur. Phys. J. B* **4**(2), 131–134.

Reid, C. R., Beekman, M., Latty, T. and Dussutour, A. (2013), 'Amoeboid organism uses extracellular secretions to make smart foraging decisions', *Behav. Ecol.* **24**(4), 812–818.

Reid, C. R., Latty, T., Dussutour, A. and Beekman, M. (2012), 'Slime mold uses an externalized spatial "memory" to navigate in complex environments', *Proc. Natl. Acad. Sci. U.S.A.* **109**(43), 17490–17494.

Reid, C. R., MacDonald, H., Mann, R. P., Marshall, J. A. R., Latty, T. and Garnier, S. (2016), 'Decision-making without a brain: how an amoeboid organism solves the two-armed bandit', *J. R. Soc. Interface* **13**(119), 20160030.

Rodiek, B. and Hauser, M. J. B. (2015), 'Migratory behaviour of physarum polycephalum microplasmodia', *Eur. Phys. J. Spec. Top.* **224**(7), 1199–1214.

Rodiek, B., Takagi, S., Ueda, T. and Hauser, M. J. B. (2015), 'Patterns of cell thickness oscillations during directional migration of physarum polycephalum', *Eur. Biophys. J.* **44**(5), 349–358.

Romeralo, M., Escalante, R. and Baldauf, S. L. (2012), 'Evolution and diversity of dictyostelid social amoebae', *Protist* **163**(3), 327–343.

Saberi, A. A. (2015), 'Recent advances in percolation theory and its applications', *Phys. Rep.* **578**, 1–32.

Saigusa, T., Tero, A., Nakagaki, T. and Kuramoto, Y. (2008), 'Amoebae anticipate periodic events', *Phys. Rev. Lett.* **100**(1).

Sauer, H. W. (1986), Introduction to physarum, *in* 'The Molecular Biology of Physarum polycephalum', Springer New York, pp. 1–17.

Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G. and Manna, S. S. (2003), 'Small-world properties of the indian railway network', *Phys. Rev. E* **67**(3).

Serini, G. (2003), 'Modeling the early stages of vascular network assembly', *EMBO J.* **22**(8), 1771–1779.

Shirakawa, T. and Gunji, Y.-P. (2007), 'Emergence of morphological order in the network formation of physarum polycephalum', *Biophys. Chem.* **128**(2-3), 253–260.

Smith, D. and Saldana, R. (1992), 'Model of the ca2+ oscillator for shuttle streaming in physarum polycephalum', *Biophys. J.* **61**(2), 368–380.

Sole, R. V. and Montoya, M. (2001), 'Complexity and fragility in ecological networks', *Proc. Royal Soc. B* **268**(1480), 2039–2045.

Sorge, A. (2015), 'Pyfssa 0.7.6'.

Stauffer, D. and Aharony, A. (1992), *Introduction To Percolation Theory*, Taylor & Francis, London.

Stockmayer, W. H. (1944), 'Theory of molecular size distribution and gel formation in branched polymers II. general cross linking', *J. Chem. Phys.* **12**(4), 125–131.

Sukhorukov, V. M., Dikov, D., Reichert, A. S. and Meyer-Hermann, M. (2012), 'Emergence of the mitochondrial reticulum from fission and fusion dynamics', *PLoS Comput. Biol.* **8**(10), e1002745.

Sun, Y. (2017), 'Physarum-inspired network optimization: A review', *ArXiv* .

Takagi, S. and Ueda, T. (2008), 'Emergence and transitions of dynamic patterns of thickness oscillation of the plasmodium of the true slime mold physarum polycephalum', *Physica D* **237**(3), 420–427.

Takagi, S. and Ueda, T. (2010), 'Annihilation and creation of rotating waves by a local light pulse in a protoplasmic droplet of the physarum plasmodium', *Physica D* **239**(11), 873–878.

Takamatsu, A., Takaba, E. and Takizawa, G. (2009), 'Environment-dependent morphology in plasmodium of true slime mold physarum polycephalum and a network growth model', *J. Theor. Biol.* **256**(1), 29–44.

Teplov, V. A. (2017), 'Role of mechanics in the appearance of oscillatory instability and standing waves of the mechanochemical activity in the physarum polycephalum plasmodium', *J. Phys. D: Appl. Phys.* **50**(21), 213002.

Tero, A., Kobayashi, R. and Nakagaki, T. (2007), 'A mathematical model for adaptive transport network in path finding by true slime mold', *J. Theor. Biol.* **244**(4), 553–564.

Tero, A., Takagi, S., Saigusa, T., Ito, K., Bebber, D. P., Fricker, M. D., Yumiki, K., Kobayashi, R. and Nakagaki, T. (2010), 'Rules for biologically inspired adaptive network design', *Science* **327**(5964), 439–442.

Tignol, J.-P. (2016), *Galois' Theory of Algebraic Equations*, 2. edn, World Scientific.

Toral, R. and Colet, P. (2014), *Stochastic numerical methods*, Wiley-VCH.

Tsuda, S., Aono, M. and Gunji, Y.-P. (2004), 'Robust and emergent physarum logical-computing', *Biosystems* **73**(1), 45–55.

Tsuda, S., Zauner, K.-P. and Gunji, Y.-P. (2007), 'Robot control with biological cells', *Biosystems* **87**(2-3), 215–223.

Ueda, T., Muratsugu, M., Kurihara, K. and Kobatake, Y. (1976), 'Chemotaxis in physarum polycephalum', *Exp. Cell. Res.* **100**(2), 337–344.

Ueda, T., Terayama, K., Kurihara, K. and Kobatake, Y. (1975), 'Threshold phenomena in chemoreception and taxis in slime mold physarum polycephalum', *J. Gen. Physiol.* **65**(2), 223–234.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E. and Yu, T. (2014), 'scikit-image: image processing in python', *PeerJ* **2**, e453.

van Rossum, G. (1995), Python tutorial, technical report cs-r9526, Technical report, Centrum voor Wiskunde en Informatica, Amsterdam.

Verma, T., Araújo, N. A. M. and Herrmann, H. J. (2014), 'Revealing the structure of the world airline network', *Sci. Rep.* **4**(1).

Vogel, D., Gautrais, J., Perna, A., Sumpter, D. J. T., Deneubourg, J.-L. and Dussutour, A. (2016), 'Transition from isotropic to digitated growth modulates network formation inPhysarum polycephalum', *J. Phys. D: Appl. Phys.* **50**(1), 014002.

Voit, E. O., Martens, H. A. and Omholt, S. W. (2015), '150 years of the mass action law', *PLoS Comput. Biol.* **11**, e1004012.

Wales, D. J. and Doye, J. P. K. (1997), 'Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms', *J. Phys. Chem. A* **101**(28), 5111–5116.

Watts, D. J. and Strogatz, S. H. (1998), 'Collective dynamics of 'small-world' networks', *Nature* **393**(6684), 440–442.

Werthmann, B. and Marwan, W. (2017), 'Developmental switching in physarum polycephalum: Petri net analysis of single cell trajectories of gene expression indicates responsiveness and genetic plasticity of the waddington quasipotential landscape', *J. Phys. D: Appl. Phys.* **50**(46), 464003.

Widom, B. (1965), 'Surface tension and molecular correlations near the critical point', *J. Chem. Phys.* **43**(11), 3892–3897.

Wilf, H. S. (1994), *Generatingfunctionology*, 2. edn, Academic Press, London.

Wilson, K. G. (1975), 'The renormalization group: Critical phenomena and the kondo problem', *Rev. Mod. Phys.* **47**(4), 773–840.

Wohlfahrth-Bottermann, K. E. (1979), 'Oscillatory contraction activity in physarum', *J. Exp. Biol.* **40**, 15–32.

Yi, S. D., Jo, W. S., Kim, B. J. and Son, S.-W. (2013), 'Percolation properties of growing networks under an achlioptas process', *Europhys. Lett.* **103**(2), 26004.

Zamponi, N., Zamponi, E., Cannas, S. A., Billoni, O. V., Helguera, P. R. and Chialvo, D. R. (2018), 'Mitochondrial network complexity emerges from fission/fusion dynamics', *Sci. Rep.* **8**(1).

Zhang, S., Guy, R. D., Lasheras, J. C. and del Álamo, J. C. (2017), 'Self-organized mechano-chemical dynamics in amoeboid locomotion of physarum fragments', *J. Phys. D: Appl. Phys.* **50**(20), 204004.

Zhu, L., Aono, M., Kim, S.-J. and Hara, M. (2013), 'Amoeba-based computing for traveling salesman problem: Long-term correlations between spatially separated individual cells of physarum polycephalum', *Biosystems* **112**(1), 1–10.