
Evolution Boolescher Netzwerke

Evolution of Boolean Networks

Zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation von Dipl.-Phys. Agnes Szejka aus Złotów

Februar 2010 — Darmstadt — D 17



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Physik
Institut für Festkörperphysik
AG Drossel

Evolution Boolescher Netzwerke
Evolution of Boolean Networks

Genehmigte Dissertation von Dipl.-Phys. Agnes Szejka aus Złotów

1. Gutachten: Prof. Dr. Barbara Drossel
2. Gutachten: Prof. Dr. Markus Porto

Tag der Einreichung: 07. Januar 2010
Tag der Prüfung: 15. Februar 2010

Darmstadt – D 17

„We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.“

T. S. Eliot



Abstract

Boolean networks have been used as models for gene regulatory networks since 1969, and since the 1980s they have also become the object of evolutionary studies. Many dynamical properties of Boolean networks were not properly understood until the last decade, and many interesting evolutionary questions are still open. The random Boolean networks that constitute the starting points of the simulated network evolution in this thesis, consist of N nodes, each of which gets K inputs from randomly chosen nodes. The nodes represent genes that are either expressed and therefore have a value of 1 in the model or not expressed and thus have a value of 0. The directed links between nodes represent regulatory interactions between genes. The state of a node in the network is a Boolean function of the states of its input nodes. Depending on the value of K and on the update rules chosen, Boolean networks can show different dynamical behaviour, that can be characterised as frozen, critical or chaotic.

In this thesis, an adaptive walk algorithm is employed to simulate network evolution and to gain insight into the fitness landscapes of the networks. The fitness criteria used depend on the dynamical robustness of the attractors against small perturbations and on the robust response of the networks to external stimuli, both of which are important properties of biological systems. The majority of the simulations are carried out with only a subset of all possible update rules for the nodes, namely canalizing and threshold functions. They have both been shown to be relevant in a biological context. The properties of the simulated evolution, of the evolved networks and of the fitness landscapes are studied and discussed.

Zusammenfassung

Boolesche Netzwerke werden seit 1969 zur Modellierung von Genregulationsnetzwerken verwendet, seit den 80er Jahren sind sie auch das Objekt evolutionärer Studien. Während viele dynamische Eigenschaften Boolescher Netzwerke erst während des letzten Jahrzehnts wirklich verstanden wurden, sind auch viele interessante evolutionäre Fragestellungen noch offen. Die zufälligen Booleschen Netzwerke, mit denen die Simulationen in dieser Arbeit beginnen, bestehen aus N Knoten, von denen jeder K Eingänge von zufällig ausgewählten Knoten erhält. Dabei repräsentieren die Knoten Gene, die entweder exprimiert werden und damit im Modell den Wert 1 haben, oder nicht exprimiert werden und im Modell den Wert 0 annehmen. Die gerichteten Verbindungen zwischen den Knoten repräsentieren regulatorische Einflüsse auf die Expression eines Gens durch andere Gene. Der Zustand eines Knotens im Netzwerk ist eine Boolesche Funktion der Zustände seiner Eingangsknoten. Je nach der Wahl von K und den Aktualisierungsfunktionen können Boolesche Netzwerke verschiedenes dynamisches Verhalten zeigen, das als gefroren, kritisch oder chaotisch charakterisiert wird.

In dieser Arbeit wird der Algorithmus der Adaptiven Wanderung (*engl.: adaptive walk*) dazu verwendet, Evolution zu simulieren und Einblicke in die Fitnesslandschaften der Netzwerke zu gewinnen. Die angewendeten Fitnesskriterien sind die dynamische Robustheit der Attraktoren gegenüber kleinen Störungen und die robuste Antwort der Netzwerke auf externe Stimuli, beides wichtige Eigenschaften von biologischen Systemen. Die meisten Simulationen werden mit einer Untermenge aller möglichen Aktualisierungsregeln der Knoten durchgeführt, mit kanalisierenden und Schwellenwertfunktionen, die im biologischen Kontext relevant sind. Die Eigenschaften der simulierten Evolution, der evolvierten Netzwerke und der Fitnesslandschaften werden untersucht und diskutiert.

Inhaltsverzeichnis

1	Einleitung	7
1.1	Netzwerkforschung	7
1.2	Netzwerke in der Zelle	8
1.3	Aufbau der Arbeit	10
2	Boolesche Zufallsnetzwerke	13
2.1	Definition des Modells	13
2.2	Kanalierende und Schwellenwertfunktionen	15
2.3	Dynamische Regime	17
2.3.1	Der Parameter λ	17
2.3.2	Eigenschaften der Netzwerke in den verschiedenen dynamischen Regimen	18
2.3.3	Sind Zellen kritisch?	23
2.4	<i>Saccharomyces cerevisiae</i> – Boolesches Modell eines realen Netzwerks	24
3	Modelle der Evolution	27
3.1	Fitnesslandschaften	27
3.1.1	Neutrale Evolution am Beispiel von RNA- und Proteinmodellen	29
3.2	Computersimulationen von Netzwerkevolution	32
4	Der Phasenübergang in Schwellenwertnetzwerken	35
4.1	Das Phasendiagramm	35
4.1.1	Die Entwicklung von b_t	36
4.1.2	Die Sensitivität λ	38
4.2	Numerische Simulationen	39
4.2.1	Der Anteil von Nullen und Einsen	40
4.2.2	Das Phasendiagramm	45
4.3	Die Bäckerhefe	47
5	Simulierte Evolution mittels der „Adaptiven Wanderung“	49
5.1	Erzeugung eines Netzwerkes	50
5.2	Fitnessbestimmung	50
5.3	Mutationen	51
5.4	Abbruchbedingungen der Simulation	52
5.5	Bemerkungen zu biologischen Grundlagen	52
6	Evolution Boolescher Netzwerke unter Selektion für robuste Dynamik	57
6.1	Generelle Eigenschaften evolvierter Boolescher Zufallsnetzwerke	57
6.2	Der Derrida-Plot und seine Modifikationen	59
6.2.1	Anwendung auf evolvierte Netzwerke	61
6.2.2	Anwendung auf das Zellzyklusregulationsnetzwerk der Bäckerhefe	63
6.2.3	Fazit	63



6.3	Auswirkungen der Netzwerkgröße auf die Evolution von kanalisierenden und Schwellenwertnetzwerken	68
6.3.1	Fazit	71
7	Evolution von kanalisierenden und Schwellenwertnetzwerken unter Selektion für eine stabile Antwort auf externe Stimuli	73
7.1	Der Weg zur maximalen Fitness	74
7.2	Neutrale Evolution auf dem Plateau maximaler Fitness	77
7.3	Eigenschaften der Netzwerke mit maximaler Fitness	81
7.4	Zusammenfassung	84
7.5	Betrachtungen zu neutralen Fitnesslandschaften	84
8	Zusammenfassung und Ausblick	87
	Literaturverzeichnis	89

1 Einleitung

1.1 Netzwerkforschung

Viele komplexe Systeme können als Netzwerke modelliert werden. Mathematisch spricht man von einem Graph, der sich aus Knoten zusammensetzt, die mit Kanten untereinander verbunden sind. Die Kanten repräsentieren Wechselwirkungen zwischen den Knoten.

Als eine Anwendung der mathematischen Graphentheorie auf Fragestellungen der verschiedensten Fachgebiete, lebte die Netzwerkforschung in den späten 1990ern als eine „neue Wissenschaft“ auf [56]. Die Graphentheorie wurde allerdings schon seit ihren Anfängen 1736 auf praktische Probleme angewendet. In diesem Jahr löste der Schweizer Mathematiker Leonhard Euler das sehr konkrete Königsberger Brückenproblem mit Mitteln, die wir heute der Graphentheorie zurechnen. Dessen Gegenstand war die Frage nach der Möglichkeit eines Rundweges, der jede der sieben betrachteten Königsberger Brücken nur einmal verwendet. Er bewies, dass dies nicht möglich war, da zu allen durch den Fluss voneinander getrennten Gebieten eine ungerade Anzahl von Brücken führte. Um aber einen solchen Weg realisierbar zu machen, müssten alle, bis auf höchstens zwei der Gebiete, die dann als Ausgangs- bzw. Endpunkt dienen würden, eine gerade Anzahl von Brücken haben, um sie auch wieder verlassen zu können (s. Abb. 1.1).

Als nächstes sind die Arbeiten des Ungarischen Mathematikers Paul Erdős zu erwähnen, der die Graphentheorie in den 1950ern in Veröffentlichungen zu Zufallsgraphen wieder aufnahm [27, 28].

In den späten 1960er und 70er Jahren begannen Sozialwissenschaftler soziale Netzwerke und das Verhalten von Gruppen mit Hilfe der Graphentheorie zu untersuchen. In dieser Zeit prägte Stanley Milgram den Begriff der Kleine-Welt-Netzwerke und entfachte das Interesse an der Fragestellung, wie sich die Netzwerktopologie auf menschliches Verhalten auswirken könnte und umgekehrt. Milgram führte ein Experiment durch, bei dem Freiwillige aus zwei Bundesstaaten der USA beauftragt wurden, einen Brief an unbekannte Personen durch circa das halbe Staatsgebiet der Vereinigten Staaten weiterzuleiten, wobei die Briefe nur an solche Personen weitergegeben werden durften, mit denen die Freiwilligen persönlich bekannt

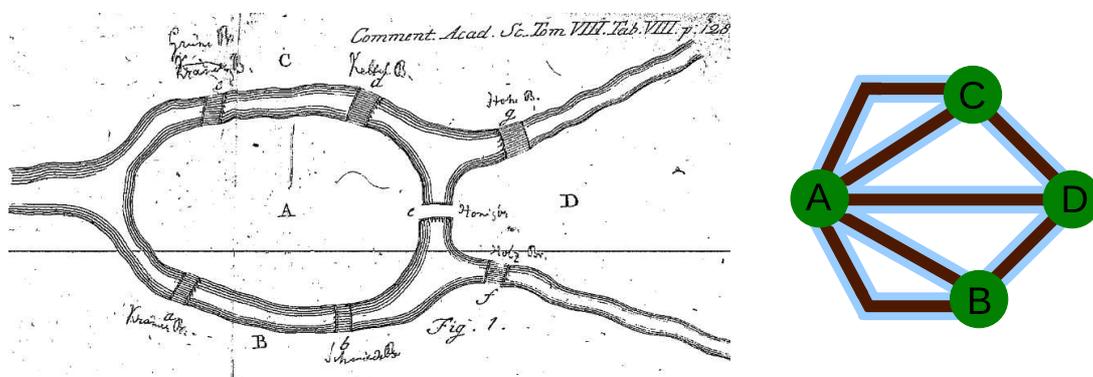


Abbildung 1.1: Das Königsberger Brückenproblem aus der Originalpublikation von Euler und seine Darstellung als Graph. Die durch den Fluss getrennten Stadtteile sind als Knoten, die Brücken als Kanten dargestellt.

waren. Bei den Briefen, die ihre Ziele erreichten, reichten die Etappen entlang der Kette von Übermittlern von zwei bis zu zehn, mit einem Mittelwert von 5,5. So entstand die Vorstellung von der „kleinen Welt“ und der Begriff „six degrees of separation“ [56].

Von besonderem Interesse für die vorliegende Arbeit ist die Einführung von Booleschen Zufallsnetzwerken in den biologischen Kontext von Genregulation im Jahr 1969 durch Stuart A. Kaufmann. Auf diese wird in Kapitel 2 näher eingegangen.

In den 1990ern benutzten Duncan Watts, Steven Strogatz und Albert-László Barabási Netzwerke, um physikalische und biologische Phänomene zu modellieren. Watts brachte die Struktur von wenig verknüpften Kleine-Welt-Netzwerken mit verschiedenartigen Phänomenen wie Phasenübergängen in Materialien, der Funktionalität von biologischen Organismen und dem Verhalten von Stromnetzen in Verbindung [93]. Strogatz untersuchte den Einfluss der Netzwerkstruktur auf komplexe adaptive Systeme in der Physik und erklärte, warum Herzschläge von Säugetieren regelmäßige synchronisierte Muster aufweisen, sowie, warum bestimmte Spezies von Glühwürmchen ohne zentralisierte Kontrolle unisono rhythmisch zirpen [83]. Es stellte sich heraus, dass lebende Organismen dazu neigen ihr Verhalten zu synchronisieren [56].

Barabási und Mitarbeiter prägten den Begriff der skalenfreien Netzwerke [7]. Deren Gradverteilung skaliert mit einem Potenzgesetz, das heißt, es gibt viele Knoten mit wenigen Verknüpfungen und wenige Knoten mit vielen Verknüpfungen. Diese Art Verknüpfungsmuster ist von besonderem Interesse, da viele empirische Netzwerke wie Zitationsgraphen von wissenschaftlichen Artikeln, das World Wide Web, das Internet sowie bestimmte soziale und zelluläre Netzwerke eine solche Topologie aufweisen [2]. Skalenfreie Netzwerke können unter anderem durch einen Mechanismus genannt *preferential attachment* entstehen [7], bei dem die Wahrscheinlichkeit eines Knotens eine weitere Kante zu erhalten proportional zur Anzahl der Kanten ist, die er schon hat. Dieser Mechanismus ist auch unter dem Namen *rich-get-richer*-Prinzip bekannt.

1.2 Netzwerke in der Zelle

Die Eigenschaften einer Zelle werden durch die vielschichtigen Wechselwirkungen zwischen ihren Bestandteilen wie DNA, RNA und Proteine bestimmt. Mit Hilfe von Signalübertragung und regulatorischen Mechanismen sind Zellen in der Lage auf eine sich ändernde Umwelt zu reagieren. Es wird zunehmend erkannt, dass das Verständnis von Eigenschaften, die aus der Funktion der gesamten Zelle erwachsen, integrierte theoretische Beschreibungen der Beziehungen zwischen den verschiedenen Bestandteilen der Zelle benötigt [1].

DNA, RNA, Proteine

DNA ist der Träger der Erbinformation, die in der Abfolge der vier Basen Adenin (A), Guanin (G), Cytosin (C) und Thymin (T) gespeichert ist. Im Normalzustand liegt die DNA als Doppelstrang vor. Dabei besteht ein Strang aus einem Zucker-Phosphat-Rückgrat und den Basen, die durch Wasserstoffbindungen mit den Basen des komplementären DNA-Strangs verbunden sind. Dabei sind fast immer Adenin und Thymin gepaart sowie Guanin und Cytosin. Teile der DNA kodieren für Proteine. Die DNA-Abschnitte, die für Proteine oder anderes kodieren, werden Gene genannt. Proteine sind aus Aminosäuren zusammengesetzt. Da es 20 verschiedene

Aminosäuren gibt, aus denen Proteine aufgebaut werden und die in der DNA kodiert sind, werden drei Basen gebraucht um eine Aminosäure zu kodieren. Eine solche Kombination von drei Basen wird als Codon bezeichnet. Da es 64 mögliche Drei-Basen-Kombinationen gibt, ergibt sich eine gewisse Redundanz. Drei dieser Kombinationen kodieren nicht für Aminosäuren, sondern sind sogenannte Stopcodons, die das Ende der Translation signalisieren.

Die Erzeugung eines Proteins geschieht in zwei Stufen. Die erste von ihnen besteht in der Transkription der DNA in RNA. RNA ist ein der DNA verwandtes Molekül, bei dem die Base Thymin durch Uracil ersetzt ist und der Rückgrat-Zucker Desoxyribose durch Ribose. Bei Eukaryoten werden nun noch die Introns (nicht-proteinkodierende DNA-Abschnitte) herausgeschnitten. Das kann auf verschiedene Arten geschehen, die zu verschiedenen Proteinen führen. Das Resultat ist die Boten-RNA (auch mRNA von engl. messenger RNA). In Prokaryoten kodiert ein mRNA-Strang meist für mehrere Proteine, die zusammenwirken. Die zweite Stufe des Prozesses ist die Translation von mRNA in Proteine. Dafür wird zum einen die Transfer-RNA (tRNA) benötigt, welche die Aminosäuren mit den entsprechenden Codons zusammenbringt, und die Ribosome, die die Aminosäuren miteinander verbinden.

Da nicht alle Proteine immer benötigt werden, sondern nur in bestimmten Situationen oder bei multizellulären Organismen nur in bestimmten Zellen, ist ihre Herstellung auf unterschiedliche Arten und auf unterschiedlichen Ebenen reguliert. Dabei können auch externe Stimuli eine Rolle spielen [25].

Genregulationsnetzwerke

Die Transkription von Genen wird durch die Wirkung von Transkriptionsfaktoren (Proteine, die allein oder in Komplexen an die regulatorischen Sequenzen der DNA binden) beeinflusst. Ihr Binden kann die Transkription eines Gens inhibieren oder aktivieren. Da Transkriptionsfaktoren von Genen kodiert werden, bedeutet das letztendlich, dass Gene gegenseitig ihre Expression regulieren und so Genregulationsnetzwerke bilden. Detaillierte Netzwerkmodelle enthalten zwei Arten von Knoten und zwei Arten von gerichteten Kanten. Die Knoten repräsentieren die Transkriptionsfaktoren und mRNAs und die Kanten Transkriptionsregulation und Translation. Zur Vereinfachung werden die Produkte meist mit den entsprechenden Genen, die für sie kodieren, zusammengelegt, so dass es nur noch eine Art von Knoten und eine Art von Kanten gibt. Die Knoten des Netzwerks entsprechen dann den Genen und eine gerichtete Kante von Gen i zu Gen j existiert, wenn Gen i die Expression von Gen j reguliert [1].

Protein-Protein-Interaktionsnetzwerke

Ebenso können Proteine miteinander interagieren, zum Beispiel kann ein Protein ein anderes modifizieren und damit auch dessen Interaktionsmöglichkeiten verändern. Es können auch mehrere Proteine miteinander wechselwirken und Proteinkomplexe bilden, die wiederum neue Funktionen haben. Die Gesamtheit dieser Prozesse wird als Protein-Protein-Interaktionsnetzwerk bezeichnet. Hierbei sind die Proteine die Knoten. Eine ungerichtete Verbindung zwischen ihnen besteht, wenn zwei Proteine binden [1].

Metabolische Netzwerke

Die biochemischen Reaktionen des Stoffwechsels einer Zelle können in einem metabolischen Netzwerk zusammengefasst werden. Die Stoffflüsse in diesem Netzwerk werden durch Enzyme reguliert, die die Stoffwechselreaktionen katalysieren. Das detaillierteste Modell eines metabolischen Netzwerks hat drei Arten von Knoten: Metabolite, Reaktionen und Enzyme, und zwei Arten von gerichteten Kanten: die einen stehen für den Massenfluss und die anderen für katalytische Regulation. Es existieren auch vereinfachte Repräsentationen. In Substratgraphen sind die Knoten Reaktanten, die durch Kanten verbunden sind, wenn sie in der gleichen Reaktion vorkommen. In Reaktionsgraphen repräsentieren die Knoten Reaktionen, die miteinander verbunden sind, wenn sie mindestens einen Metabolit gemeinsam haben.

Oft sind die verschiedenen Wechselwirkungsebenen verknüpft. Zum Beispiel kann ein externes Signal eine Kaskade von Reaktionen auslösen, die sowohl biochemische Wechselwirkungen als auch die Transkriptionsregulation beeinflussen [1].

Signalübermittlungsnetzwerke

Signalübermittlungsnetzwerke wurden eine lange Zeit als lineare Ketten von biochemischen Reaktionen und Protein-Protein-Wechselwirkungen angesehen, die bei einem Sensormolekül für extra- oder intrazelluläre Signale starten und bei ihren intrazellulären Zielen enden. Mehr und mehr wird jedoch deutlich, dass viele Komponenten an verschiedenen Signalwegen teilnehmen, sodass sich ein verbundenes Signalübermittlungsnetzwerk ergibt [1].

1.3 Aufbau der Arbeit

In dieser Arbeit werden Boolesche Netzwerke zur Modellierung von Genregulation benutzt. Im Vordergrund der Untersuchungen steht die evolutionäre Entwicklung der Netzwerke. Welche Einflüsse formen welche Art von Netzwerk? Was sind die topologischen und dynamischen Eigenschaften der Netzwerke, nachdem Mutationen und Selektion auf sie eingewirkt haben? Was lässt sich daraus für die Eigenschaften realer Genregulationsnetzwerke schließen? Für welche Eigenschaften des evolutionären Weges finden sich auch Hinweise in der Natur und in anderen theoretischen Betrachtungen zur Evolution?

Im nächsten Kapitel wird zunächst das Boolesche Netzwerkmodell eingeführt. Insbesondere wird auf die dynamischen Eigenschaften der zufälligen Netzwerke eingegangen, welche die Ausgangspunkte der Modellevolution sind. Die zwei vornehmlich verwendeten Aktualisierungsfunktionen der Dynamik – kanalisierende und Schwellenwertfunktionen – werden eingeführt und schließlich das Boolesche Modell eines realen Regulationsnetzwerkes anhand des Beispiels der Zellzyklusregulation der Bäckerhefe vorgestellt.

Kapitel 3 beschäftigt sich mit einigen aktuellen Grundlagen der Theorie der Evolution. Dazu gehört die Diskussion der Metapher der Fitnesslandschaften, die in dieser Arbeit für Boolesche Netzwerke untersucht werden. Besondere Beachtung gilt der Theorie der ausgedehnten neutralen Pfade auf Fitnesslandschaften, die 1969 als wichtiger Mechanismus molekularer Evolution in die Diskussion eingebracht wurde, und Gegenstand aktueller Untersuchungen insbesondere für RNA und Proteine ist. Wie sich zeigen wird, sind die Fitnesslandschaften des in dieser Arbeit untersuchten Modells durch ausgedehnte neutrale Plateaus gekennzeichnet.

In Kapitel 4 wird eine Mean-Field-Theorie der Netzwerke dazu benutzt dynamische Eigenschaften von Schwellenwertnetzwerken, vornehmlich den Übergang zwischen den dynamischen Phasen, zu berechnen. Es ist wichtig, das Verhalten der unveränderten zufälligen Netzwerke zu kennen, bevor mit deren Veränderung durch simulierte Evolution begonnen werden kann. In der Tat zeigen sich einige Eigenschaften des ursprünglichen Modells, die bei der Parameterwahl für die Evolution beachtet werden müssen. Ferner werden Implikationen für das Zellzyklusregulationsnetzwerk der Bäckerhefe diskutiert, das mit dem selben Satz von Funktionen modelliert wurde. Die Ergebnisse dieses Teils der Arbeit wurden in Zusammenarbeit mit Tamara Mihaljev und Barbara Drossel erhalten und im *New Journal of Physics* veröffentlicht [87].

In Kapitel 5 wird der evolutionäre Algorithmus der „Adaptiven Wanderung“ vorgestellt, der zur Simulation von Evolution in dieser Arbeit benutzt wird. Nachfolgend werden die biologischen Grundlagen der verwendeten Fitnesskriterien und Mutationsarten diskutiert.

In Kapitel 6 wird zunächst Evolution mit einfacheren Selektionskriterien betrachtet, die noch keinen Einfluss von externen Stimuli beinhalten.

Die ersten beiden Abschnitte des Kapitels beschäftigen sich mit den Eigenschaften von evolvierten Booleschen Zufallsnetzwerken mit allen Booleschen Funktionen. Der Derrida-Plot, der dazu verwendet werden kann die Netzwerke nach den dynamischen Phasen, in denen sie sich befinden können, zu unterteilen, und Modifikationen dieser Abbildung werden für evolvierte und zufällige Netzwerke verglichen. Es ergeben sich Hinweise darauf, dass evolvierte Netzwerke nicht nach dem einfachen Schema, das für zufällige Netzwerke gültig ist, einem der dynamischen Regime zugeordnet werden können. Es lässt sich vermuten, dass dies auch für reale biologische Netzwerke, wie es teilweise versucht wurde (s. Kap. 2.3.3), nicht sinnvoll ist. Die Ergebnisse aus diesem Teil der Arbeit entstanden in Zusammenarbeit mit Christoph Fretter und Barbara Drossel und wurden in der Zeitschrift *New Journal of Physics* veröffentlicht [34].

Im dritten Abschnitt des Kapitels wird der evolutionäre Verlauf bis zum Erreichen des Fitnessmaximums für kanalisierende und Schwellenwertnetzwerke verglichen. Der Einfluss der Netzwerkgröße auf die Eigenschaften der simulierten Evolution einerseits und der evolvierten Netzwerke andererseits wird für die beiden Netzwerkarten näher betrachtet. Die Netzwerksimulationen aus diesem Abschnitt wurden zum Teil in Zusammenarbeit mit Christopher Priester durchgeführt.

In Kapitel 7 wird die Evolution von kanalisierenden und Schwellenwertnetzwerken unter verschiedenen Selektionskriterien, die eine stabile Antwort auf externe Stimuli beinhalten, untersucht. Die verschiedenen Selektionskriterien führen zu Netzwerken mit unterschiedlichen Eigenschaften, doch in allen Fällen finden sich Wege zum Fitnessmaximum, das kein isolierter Gipfel in der Fitnesslandschaft ist, sondern ein ausgedehntes Plateau darstellt, auf dem sich die Netzwerke weit durch den Raum aller möglichen Netzwerke bewegen können. Es finden sich Netzwerke mit sehr unterschiedlicher Topologie und Dynamik, welche die gleiche, durch das Selektionskriterium vorgegebene, Funktion ausführen, und in der Fitnesslandschaft durch neutrale Pfade miteinander verbunden sind. Während diese Netzwerke also durch selektiv neutrale Mutationen verknüpft sind, die eine bestimmte Funktion bewahren, können sich andere Funktionen entwickeln. Die Netzwerke sind robust bei der Ausführung ihrer Funktion, auch unter Störungen, und gleichzeitig evolvierbar. Die Ergebnisse dieses Kapitels wurden in Zusammenarbeit mit Barbara Drossel erhalten und in der Zeitschrift *Physical Review E* veröffentlicht [86].

Im letzten Kapitel sind die wichtigsten Ergebnisse und Schlussfolgerungen der Arbeit noch einmal zusammengefasst. Anschließend wird ein kurzer Ausblick gegeben.



2 Boolesche Zufallsnetzwerke

2.1 Definition des Modells

Boolesche Zufallsnetzwerke (**R**andom **B**oolean **N**etworks) wurden im Jahr 1969 von Stuart A. Kauffman zur Modellierung von Genregulation eingeführt [44]. In ihnen repräsentieren die Knoten Gene und eine gerichtete Verbindung zwischen zwei Knoten den Einfluss eines Gens auf die Transkription des anderen. Das Expressionslevel eines Gens σ wird mit einem von zwei möglichen Werten angegeben, entweder das Gen wird nicht transkribiert – es ist ausgeschaltet (0) – oder es wird transkribiert – es ist eingeschaltet (1). Bei dem einfachsten der Booleschen Modelle wird der Unkenntnis des genauen Verknüpfungsmusters zwischen den Genen und der genauen Art ihrer gegenseitigen Beeinflussung Rechnung getragen, indem Verknüpfungen und Aktualisierungsregeln der Knoten völlig zufällig gewählt werden. Konkret bedeutet das, dass jedem Knoten i K_i zufällige Knoten als Eingangsknoten zugewiesen werden. Im klassischen Booleschen Zufallsmodell hat jeder Knoten die gleiche Anzahl von Eingängen $K_1 = K_2 = \dots = K_N = K$, im Folgenden wird mit K die mittlere Anzahl von Eingängen pro Knoten in einem Netzwerk bezeichnet.

$$K = \frac{1}{N} \sum_{i=1}^N K_i$$

Der Aktivitätszustand jedes Knotens ist eine Funktion der Zustände seiner Eingangsknoten im vorhergehenden Zeitschritt, die wiederum zufällig aus allen Booleschen Funktionen gewählt wird. Das heißt, der Wert jedes Knotens i ist in jedem Zeitschritt gegeben durch

$$\sigma_i(t) = F_i[\sigma_{i_1}(t-1), \sigma_{i_2}(t-1), \dots, \sigma_{i_{K_i}}(t-1)], \quad (2.1)$$

wobei F eine zufällige Boolesche Funktion ist. Die Aktualisierung aller N Knoten erfolgt synchron.

Eine Boolesche Funktion kann als eine Zeichenkette von Nullen und Einsen dargestellt werden, die jeder möglichen Kombination von Eingangswerten einen Ausgangswert zuordnet, also die Länge 2^{K_i} hat. Realisiert wird die zufällige Zuordnung einer Funktion zu einem Knoten, indem für jede Eingangswertkombination mit der Wahrscheinlichkeit p der Ausgangswert 1 und mit der Wahrscheinlichkeit $1-p$ der Wert 0 gewählt wird. Die Anzahl der möglichen Booleschen Funktionen, aus denen gewählt werden kann, ist $2^{2^{K_i}}$. Im einfachsten Fall sind beide Knotenzustände gleichwahrscheinlich und $p = 0,5$. Eine bestimmte Wahl von Eingangsverbindungen und Aktualisierungsfunktionen bestimmt die *Realisation* eines Netzwerkes.

In Abbildung 2.1 ist ein Boolesches Zufallsnetzwerk der Netzwerkgröße $N = 4$ und der Konnektivität $K = 2$ zu sehen.

Als *Netzwerkzustand* wird der Vektor aller Zustandsvariablen der einzelnen Knoten eines Netzwerkes bezeichnet $\vec{\sigma}(t) = \{\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t)\}$. Für gewöhnlich wird zu Beginn einer Netzwerksimulation einer dieser Zustände zufällig als Anfangszustand ausgewählt. Da jeder Knoten nur zwei Werte annehmen kann, gibt es 2^N verschiedene solche Zustände, die durch

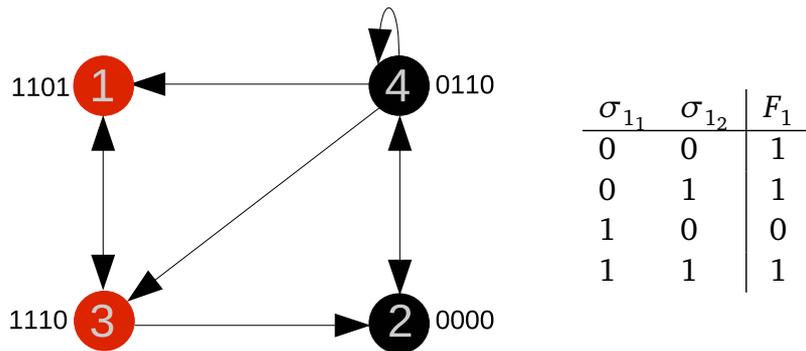


Abbildung 2.1: Boolesches Zufallsnetzwerk mit $N = 4$, $K = 2$. Die Anzahl der Eingänge ist für alle Knoten gleich, die Anzahl der ausgehenden Verbindungen variiert von Knoten zu Knoten, sie ist in Booleschen Zufallsnetzwerken Poisson-verteilt. Die Zahlen in den Knoten geben ihre Indizes an. Rote Knoten sind eingeschaltet, schwarze Knoten ausgeschaltet. Die Zeichenketten neben den Knoten zeigen ihre Booleschen Funktionen. Die Tabelle führt die Aktualisierungsregel für Knoten 1 beispielhaft aus.

ihre zeitliche Abfolge miteinander verbunden sind und den Zustandsraum bilden. Da die Aktualisierung der Zustände deterministisch erfolgt (Gl. 2.1), gibt es für jeden Netzwerkzustand nur einen Nachfolger. Die Endlichkeit des Netzwerkes führt zudem dazu, dass die Trajektorien im Zustandsraum irgendwann periodisch werden. Die periodischen Teile der Trajektorien werden als *Attraktoren* bezeichnet, die Zustandsabfolgen, die zu ihnen hinführen als *Transienten*. Die Anzahl der Zustände, die zu einem Attraktor gehören, heißt Attraktorlänge. Verschiedene Netzwerkrealisationen unterscheiden sich in der Anzahl und der Länge der Attraktoren. In Abbildung 2.2 ist der Zustandsraum des Netzwerkes aus Abbildung 2.1 zu sehen. Die Attraktorzustände sind in türkis gezeigt, die transienten Zustände in orange. Es gibt zwei Attraktoren, einen der Länge 1 und einen der Länge 4.

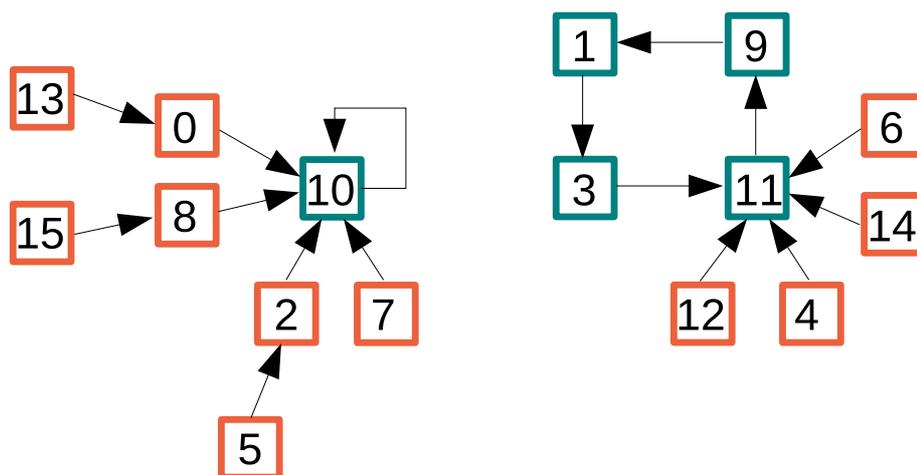


Abbildung 2.2: Zustandsraum des Netzwerkes mit $N = 4$, $K = 2$ aus Abb. 2.1. Attraktorzustände sind in türkis dargestellt, transiente Zustände in orange. Die Bezeichnung der Netzwerkzustände ergibt sich aus der Umrechnung der Binärzahl $\sigma_1\sigma_2\sigma_3\sigma_4$ in eine Dezimalzahl.

2.2 Kanalisierende und Schwellenwertfunktionen

Schwellenwertnetzwerke und kanalisierende Netzwerke sind Unterklassen von zufälligen Booleschen Netzwerken, die nur einen Teil aller möglichen Booleschen Aktualisierungsfunktionen nutzen. Kanalisierende Funktionen treten gehäuft in biologischen Genregulationsnetzwerken auf. Harris et al. fanden bei der Analyse von über 150 Genregulationssystemen eine starke Tendenz zu kanalisierenden Funktionen [39]. Schwellenwertfunktionen wurden erfolgreich bei der Modellierung eines biologischen Regulationsnetzwerkes verwendet, dessen der Zellzyklusregulation der Bäckerhefe (s. Kapitel 2.4) [57], und werden auch zur Simulation von Netzwerkevolution benutzt [15, 79].

Kanalisierende Funktionen

Kanalisierende Funktionen zeichnen sich dadurch aus, dass der Wert eines einzelnen Eingangsknotens, unabhängig von den anderen Eingängen, die Ausgabe der Funktion bestimmen kann. Nimmt der kanalisierende Eingangsknoten diesen kanalisierenden Wert nicht an, hängt der Ausgang auch von den restlichen Knotenwerten ab. In dieser Arbeit wird der folgende Satz von kanalisierenden Funktionen verwendet (nach [63])

$$F(\sigma_1, \sigma_2, \dots) = \sigma_1 \text{ OR } G(\sigma_2, \dots) \quad (2.2)$$

$$F(\sigma_1, \sigma_2, \dots) = (\text{NOT } \sigma_1) \text{ AND } G(\sigma_2, \dots) \quad (2.3)$$

$$F(\sigma_1, \sigma_2, \dots) = (\text{NOT } \sigma_1) \text{ OR } G(\sigma_2, \dots) \quad (2.4)$$

$$F(\sigma_1, \sigma_2, \dots) = \sigma_1 \text{ AND } G(\sigma_2, \dots). \quad (2.5)$$

Hier bezeichnet σ_1 den *kanalisierenden* Eingang und G eine Boolesche Zufallsfunktion, die die Abhängigkeit von den restlichen Knotenwerten enthält. Durch sie können auch noch andere Eingänge kanalisierend wirken. Für die ersten beiden Klassen von kanalisierenden Funktionen (2.2 und 2.3) ist der kanalisierende Wert 1, für die anderen beiden Klassen (2.4 und 2.5) ist er 0. In beiden Fällen kann der *kanalisierte* Wert, also der Wert, den die Funktion ausgibt, wenn Kanalisation auftritt, 0 oder 1 sein. So gilt für die Funktionsklassen 2.2 und 2.4, dass ihr kanalisierter Wert 1 ist, für die Klassen 2.3 und 2.5 ist er 0. Funktionen mit nur einem Eingang sind per Definition kanalisierend. Für sie muss unter den folgenden Funktionen gewählt werden:

$$F(\sigma_1) = \sigma_1 \quad (2.6)$$

$$F(\sigma_1) = (\text{NOT } \sigma_1) \quad (2.7)$$

$$F(\sigma_1) = 1 \quad (2.8)$$

$$F(\sigma_1) = 0. \quad (2.9)$$

Schwellenwertfunktionen

Schwellenwertfunktionen tragen der Tatsache Rechnung, dass Gene positiv oder negativ reguliert werden können, d.h. ihre Transkription kann durch die Produkte anderer Gene unterdrückt oder gesteigert werden. Die Verbindungen in einem Schwellenwertnetzwerkmodell von

Genregulation sind also gewichtet, in dieser Arbeit mit ± 1 . Der Wert eines Knotens im nächsten Zeitschritt ist gegeben durch

$$\sigma_i(t+1) = \begin{cases} 1, & \sum_j c_{ij} \sigma_j(t) - h > 0 \\ 0, & \sum_j c_{ij} \sigma_j(t) - h < 0 \\ \sigma_i(t), & \sum_j c_{ij} \sigma_j(t) - h = 0. \end{cases} \quad (2.10)$$

Hierbei ist h der Schwellenwert, der in dieser Arbeit für alle Knoten gleich ist und sich auch während der Evolution nicht ändert. Die Verknüpfungen c_{ij} sind ± 1 (aktivierend/inhibierend) mit gleicher Wahrscheinlichkeit und $c_{ij} = 0$, wenn Knoten j keinen Eingangsknoten von Knoten i ist. Somit kann der Eingang $s_j = c_{ij} \sigma_j$ von Knoten j zu Knoten i drei verschiedene Werte annehmen: 0, +1 oder -1 . Ein Knoten wird aktiviert (auf „1“ geschaltet), wenn die Summe seiner Eingänge über dem Schwellenwert liegt und inaktiviert (auf „0“ geschaltet), wenn die Summe seiner Eingänge unter dem Schwellenwert liegt. Ergibt die Summe der Eingänge genau den Schwellenwert h , ändert der Knoten seinen Wert im nächsten Zeitschritt nicht. Für die Evolutionssimulationen in dieser Arbeit wurde ein nicht-ganzzahliger Schwellenwert gewählt, so dass der zuletzt besprochene Fall nicht auftreten kann. Dies geschieht aufgrund des besonderen Verhaltens der Netzwerke im Fall von ganzzahligen Schwellenwerten, das in Kapitel 4 gefunden und näher untersucht wird, und um eine gute Vergleichbarkeit zwischen kanalisierenden und Schwellenwertnetzwerken zu erreichen (s. Kapitel 7).

Es gibt verschiedene Möglichkeiten eine Aktualisierungsfunktion für Schwellenwertnetzwerke zu definieren. Dabei können die c_{ij} -Werte kontinuierlich sein und zufällig aus einer bestimmten Wahrscheinlichkeitsverteilung gewählt werden; die Knoten können die Werte ± 1 haben anstatt 1 und 0; die Schwellenwerte können von Knoten zu Knoten variieren; und die Aktualisierungsregel kann sich unterscheiden in dem Fall, in dem die Summe der Eingänge genau den Schwellenwert ergibt (s. z.B. [79, 53, 74, 73]). Den meisten Untersuchungen liegt die folgende Regel zugrunde

$$\sigma_i(t+1) = \begin{cases} 1, & \sum_j c_{ij} (2\sigma_j(t) - 1) - h \geq 0 \\ -1, & \sum_j c_{ij} (2\sigma_j(t) - 1) - h < 0. \end{cases} \quad (2.11)$$

Ein Vergleich der beiden Aktualisierungsregeln 2.10 und 2.11 findet sich in Kapitel 4.

Während inzwischen klar ist, dass zufällig konstruierte Netzwerke nicht ausreichen, um die essentiellen Eigenschaften von Genregulationsnetzwerken zu erfassen, sind Boolesche Netzwerke, die auf biologischen Daten basierend konstruiert wurden, durchaus erfolgreich in der Modellierung der Dynamik von realen Netzwerken. So wurden zum Beispiel das Segment-Polaritäts-Netzwerk von *Drosophila melanogaster* [3] und das Zellzyklusnetzwerk der Bäckerhefe *Saccharomyces cerevisiae* [57] (s. Kapitel 2.4) mit Boolescher Dynamik der Gene modelliert und beide Modelle zeigen Attraktoren, die mit den biologischen Abfolgen der Ereignisse übereinstimmen. Das dynamische Verhalten von zufälligen Booleschen Netzwerken unterscheidet sich in signifikanter Weise von den erwähnten realistischen Modellen.

2.3 Dynamische Regime

Im Grenzfall unendlicher Systemgröße zeigen alle Booleschen Modelle bei Variation der Parameter einen Phasenübergang zwischen zwei Phasen, in denen sich das dynamische Verhalten der Netzwerke qualitativ unterscheidet. So können Boolesche Zufallsnetzwerke in drei verschiedenen dynamischen Regimen sein, abhängig von ihrem K -Wert und der statistischen Gewichtung der einzelnen Booleschen Funktionen. Die Dynamik kann als gefroren, kritisch oder chaotisch charakterisiert werden.

Um Phasenübergänge theoretisch zu beschreiben, werden unter anderem Mean-Field-Theorien (deutsch auch: Molekularfeldtheorien) benutzt. Die von Derrida und Pomeau eingeführte „Annealed Approximation“ ist eine solche [24] und die in den folgenden Abschnitt vorgestellten Eigenschaften Boolescher Netzwerke wurden in ihrem Rahmen ermittelt. Mean-Field-Theorien gehen davon aus, dass jeder Knoten nur einen mittleren Effekt der anderen Knoten spürt, räumliche Fluktuationen werden ignoriert. So vernachlässigt die Annealed Approximation Korrelationen zwischen Knoten und kommt der Betrachtung eines unendlich großen System gleich, in dem die Knoten in jedem Zeitschritt zufällig neu verknüpft werden.

2.3.1 Der Parameter λ

Der Parameter λ , auch Sensitivität genannt [61, 81], ist K mal die Wahrscheinlichkeit, dass sich der Wert eines Knotens ändert, wenn sich einer seiner Eingangswerte ändert. Mit seiner Hilfe kann zwischen den drei Regimen unterschieden werden: Ist $\lambda < 1$, befindet sich das Netzwerkensemble in der gefrorenen Phase. (Eine Störung an einem Knoten breitet sich in einem Zeitschritt im Mittel zu weniger als einem weiteren Knoten aus.) Gilt $\lambda > 1$, ist das Netzwerkensemble in der chaotischen Phase. (Eine Störung breitet sich in einem Zeitschritt im Mittel zu mehr als einem Knoten aus.) Für $\lambda = 1$, auf der Grenze zwischen den beiden Phasen, sind die Netzwerke kritisch.

Bei Netzwerken mit allen Booleschen Funktionen hängt der Phasenübergang von den Parametern K und p ab

$$\lambda = 2Kp(1 - p) \quad [26].$$

Das heißt, sind alle Booleschen Funktionen gleichwahrscheinlich ($p = 0,5$), liegt die Anzahl von Eingängen pro Knoten, für die die Netzwerke kritisch sind, bei $K_c = 2$.

Für die hier verwendeten kanalisierenden Netzwerke gilt nach [63]

$$\lambda = \rho_1 + \rho_2 - 2\rho_1\rho_2 + 2\rho_2(1 - \rho_2)[\eta + q(1 - 2\eta)](K - 1),$$

wobei ρ_1 der Anteil an Funktionen ist, die 1 als *kanalisierten* Wert haben (2.2, 2.4), η der Anteil der Funktionen, die 1 als *kanalisierenden* Wert haben (2.2, 2.3), ρ_2 der Anteil an Einsen in den Funktionen G (entspricht also dem Parameter p , wie er für klassische Boolesche Zufallsnetzwerke eingeführt wurde) und q der Anteil von Knoten mit dem Wert 1 (entspricht b_t in Kapitel 4). Da in dieser Arbeit jede der kanalisierenden Funktionen mit gleicher Wahrschein-

lichkeit gewählt wird, ist der Wert von q im stationären Zustand (s. [26]) sowie die Werte der anderen Parameter gleich 0,5 und die Gleichung reduziert sich zu

$$\lambda = \frac{K + 1}{4}.$$

Die kritische Anzahl von Eingängen pro Knoten für die in dieser Arbeit betrachteten zufälligen kanalisierenden Netzwerke ist also $K_c = 3$.

Das Phasendiagramm der Schwellenwertnetzwerke, wie sie in dieser Arbeit verwendet werden, wird in Kapitel 4 berechnet. Für einen h -Wert von $-0,5$ findet der Übergang zwischen den K -Werten 4 und 5 statt. Bis zu einem Wert von $K = 4$ sind die Netzwerke gefroren, für $K \geq 5$ sind sie chaotisch.

2.3.2 Eigenschaften der Netzwerke in den verschiedenen dynamischen Regimen

Das dynamische Verhalten der Netzwerke kann durch einige Messgrößen charakterisiert werden, mit deren Hilfe man auch die verschiedenen dynamischen Regime unterscheiden kann.

Attraktorlänge und -anzahl

Wie oben beschrieben, muss jeder Anfangszustand irgendwann auf einem Attraktor enden. Der Zustandsraum besteht gewöhnlich aus mehreren Attraktoren mit unterschiedlich großen Einzugsgebieten (Basins) und Längen. Als Einzugsgebiet eines Attraktors wird die Menge aller Zustände bezeichnet, die auf den Attraktor führen, einschließlich der Attraktorzustände selbst. Die Anzahl und die Länge der Attraktoren unterscheidet sich von Realisation zu Realisation, dennoch lassen sich einem bestimmten dynamischen Regime bestimmte Eigenschaften der Attraktoren zuordnen.

In der gefrorenen Phase ist die Länge der Attraktoren typischerweise Eins, das heißt, sie sind Fixpunkte, oder zumindest sehr kurz. Ihre mittlere Länge ist von N unabhängig. Realisationen mit langen Attraktoren sind exponentiell unwahrscheinlich. Auch die mittlere Anzahl von Attraktoren ist klein und unabhängig von N . Im unwahrscheinlichen Fall, dass in einer Realisation ein langer Attraktor gefunden wird, kann aber wiederum erwartet werden, dass es sehr viele von ihnen gibt [5].

In chaotischen Netzwerken sind die Attraktoren im Mittel sehr lang, sowohl die mittlere Länge der Attraktoren als auch die der Transienten wächst exponentiell mit N [5]. Die Anzahl der Attraktoren ist wie für gefrorene Netzwerke gering. Sie steigt linear mit der Anzahl an relevanten Knoten (das sind die Knoten, die die Attraktoren bestimmen) [9].

Bei kritischen Netzwerken wächst die Anzahl an relevanten Knoten nur wie $N^{1/3}$ mit der Netzwerkgröße, während die mittlere Anzahl und Länge der Attraktoren exponentiell mit der Anzahl der relevanten Knoten ansteigt [26].

In Kauffmans Interpretation der Booleschen Netzwerke als Genregulationsnetzwerke stellt ein Attraktor einen Zellzyklus dar. Somit entsprechen verschiedene Attraktoren verschiedenen Zelltypen und das Umschalten zu anderen Attraktoren kann als Zelldifferentiation interpretiert werden [45].

Transiente Zustände

Als transient bezeichnet man diejenigen Zustände, die ein Netzwerk durchläuft, bevor es einen Attraktor erreicht. Dieser Teil einer Trajektorie wird als Transiente bezeichnet. Die Anzahl der transienten Zustände in einer Trajektorie ist die Transientenlänge.

In der chaotischen Phase sind die Transienten ähnlich lang und komplex wie die Attraktoren, in der gefrorenen Phase ist die mittlere Transientenlänge durch eine Konstante mal $\ln N$ nach unten beschränkt [5].

Gefrorene und aktive Knoten

Knoten, die ihren Wert im stationären Zustand (auf einem Attraktor) nicht mehr ändern, werden als gefroren bezeichnet. Sie können auf einem bis hin zu allen Attraktoren eines Netzwerkes gefroren sein. Die meisten sind auf allen Attraktoren gefroren und bilden damit einen gefrorenen Kern von Knoten. Demgegenüber spricht man von aktiven beziehungsweise blinkenden Knoten, wenn sie ihren Zustand auf einem Attraktor ändern.

Wie bereits der Name verrät, sind in der gefrorenen Phase die meisten Knoten gefroren. Sie enden bei festen Werten auf allen Attraktoren. Das bedeutet auch, dass die meisten Knoten einen Wert annehmen, der unabhängig vom Anfangszustand ist.

Bei kritischen Netzwerken skaliert die mittlere Anzahl von aktiven Knoten mit der Netzwerkgröße wie $N^{2/3}$ [26].

In chaotischen Netzwerken gibt es noch mehr aktive Knoten. Für maximal chaotische Boolesche Netzwerke ($K = N$ und alle Booleschen Funktionen gleichwahrscheinlich) ist die mittlere Anzahl an Knoten, zu der sich eine Störung des Zustandes eines Knotens in einem Zeitschritt ausbreitet, $N/2$. Die Hälfte der Knoten ändern ihren Wert.

Hamming-Abstand und Derrida-Plot

Der Hamming-Abstand, benannt nach Richard Wesley Hamming, ist die Distanz zwischen zwei Zeichenketten gleicher Länge, das heißt, die Anzahl an Stellen in den Ketten, an denen sich die Zeichen unterscheiden. Bei Booleschen Netzwerken wird der Hamming-Abstand von zwei Netzwerkzuständen

$$\vec{\sigma}_t = \{\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t)\} \quad \text{und} \quad \vec{\tilde{\sigma}}_t = \{\tilde{\sigma}_1(t), \tilde{\sigma}_2(t), \dots, \tilde{\sigma}_N(t)\}$$

betrachtet. Seine Bestimmung kann man sich so vorstellen, dass von jedem Netzwerk im Ensemble eine Kopie erzeugt wird und die zwei Kopien eines Netzwerkes von verschiedenen Anfangszuständen aus gestartet werden. Der Hamming-Abstand zum Zeitpunkt t ist dann gegeben durch die Anzahl der Knoten, die in den beiden Netzwerken verschiedene Zustände annehmen

$$H(t) = \sum_{i=1}^N (\sigma_i(t) - \tilde{\sigma}_i(t))^2 [26].$$

Meist wird der normierte Hamming-Abstand betrachtet, der Hamming-Abstand geteilt durch N , also der Anteil an Knoten in verschiedenen Zuständen. Ist dieser Anteil sehr klein, kann die

Wahrscheinlichkeit dafür, dass sich mehr als ein Eingang eines Knotens in den beiden Kopien unterscheidet, vernachlässigt werden, und die Änderung von $H(t)$ ist durch

$$H(t + 1) = \lambda H(t)$$

gegeben [26]. Hierbei entspricht λ der weiter oben eingeführten Sensitivität (s. Kapitel 2.3.1). Damit erhält man

$$H(t) = H_0 e^{\ln(\lambda)t}.$$

In der chaotischen Phase ($\lambda > 1$) wächst der Hamming-Abstand anfänglich exponentiell, in der gefrorenen Phase ($\lambda < 1$) nimmt der Abstand exponentiell ab mit der Zeit. Für $\lambda = 1$ ist die zeitliche Entwicklung des Hamming-Abstandes hauptsächlich von Fluktuationen bestimmt [5].

Für lange Zeiten geht der Hamming-Abstand in der gefrorenen Phase gegen Null. Die Knoten der zwei Netzwerkkopien nehmen identische Werte an, abgesehen von möglicherweise einer endlichen Anzahl an Knoten, die zusammen einen Beitrag von 0 am normierten Hamming-Abstand haben [26]. Für chaotische Netzwerke erreicht der Hamming-Abstand einen von Null verschiedenen stationären Wert (s. auch Abb. 2.5).

Um zwischen den Phasen zu unterscheiden, kann man also zum einen die anfängliche Änderung der Hamming-Distanz betrachten oder zum anderen ihr Langzeitverhalten. Für beide Fälle kann man einen Parameter λ berechnen. Zu beachten dabei ist, dass λ sich mit der Zeit ändern kann und erst seinen Fixpunkt erreicht, wenn der Anteil an Knoten mit dem Wert 1 im Netzwerk, b_t genannt, einen Fixpunkt erreicht hat. Für klassische Boolesche Netzwerke erreicht b_t nach einem Zeitschritt seinen festen Wert, der durch p gegeben ist. Für kanalisierende und Schwellenwertnetzwerke erreicht b_t seinen Fixpunkt langsamer, das heißt, auch λ ändert seinen Wert über einen längeren Zeitraum. Das bedeutet, dass der Hamming-Abstand anfänglich einen anderen Trend zeigen kann als zu späteren Zeiten. Für andere Wahlen von Aktualisierungsfunktionen zeigen sich noch deutlichere Schwierigkeiten bei der Bestimmung der Phase mit Hilfe von λ . In [37] werden Boolesche Schwellenwertnetzwerke, beschrieben durch Gleichung 2.11, mit zwei Eingängen pro Knoten untersucht. Bei Variation der Gewichte der einzelnen Funktionen, wird nicht nur eine gefrorene und eine chaotische Phase gefunden, sondern auch ein Regime, in dem b_t mit einer Periode von zwei oszilliert. Dieses Regime ist unterteilt in eine Phase mit Attraktoren der Periode zwei und eine Phase mit sehr langen Attraktoren.

Der Derrida-Plot bildet den anfänglichen Hamming-Abstand zwischen den Zuständen zweier identischer Netzwerke auf denjenigen ab, der nach einem Zeitschritt erhalten wird. In den Abbildungen 2.3, 2.4 und 2.5 sind Derrida-Plots für ein gefrorenes, ein kritisches und ein chaotisches Boolesches Zufallsnetzwerk mit einer festen Anzahl K von zufällig ausgewählten Eingängen pro Knoten und mit Aktualisierungsfunktionen, die zufällig aus allen möglichen Booleschen Funktionen bestimmt wurden, zu sehen. Zur Ermittlung der Plots werden zwei identische Netzwerke erzeugt und mit dem selben zufälligen Anfangszustand initialisiert. In einer der Kopien werden die Zustände von H Knoten invertiert. Der Hamming-Abstand ist damit H . Nun werden die Netzwerke einmal aktualisiert, um den nächsten Zeitschritt zu erhalten, und der Hamming-Abstand wird erneut ausgewertet. Für ein gegebenes Netzwerk wird diese Prozedur mehrfach für jeden H -Wert von 1 bis N wiederholt und der Mittelwert zu je-

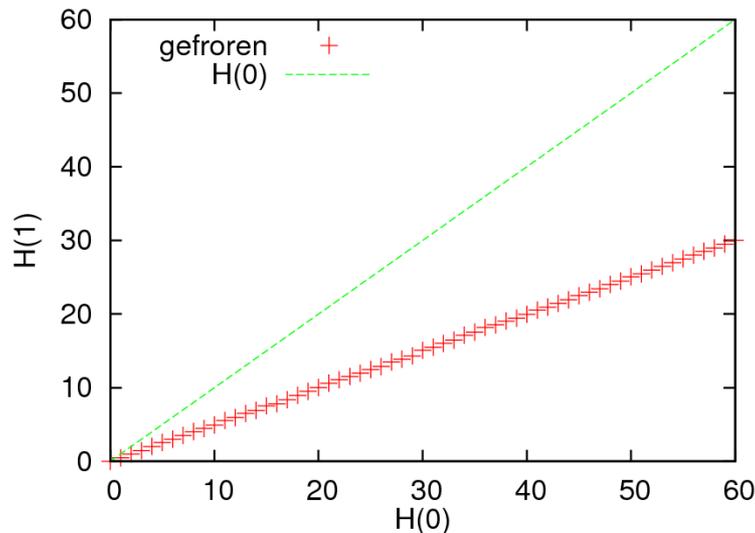


Abbildung 2.3: Derrida-Plot eines gefrorenen Booleschen Zufallsnetzwerkes mit $N = 60$, $K = 1$.

dem Wert von H bestimmt. Jeder Punkt in den drei Abbildungen 2.3, 2.4, 2.5 ist gemittelt über mindestens 200 verschiedene Störungen und Anfangszustände.

Die Derrida-Abbildung hat die folgenden Eigenschaften, die mühelos innerhalb der „Annealed Approximation“ erklärt werden können [24, 49]:

1. In der gefrorenen Phase ($K = 1$) ist der Derrida-Plot linear und hat die Steigung $1/2$. Für Netzwerke mit $K = 1$ gibt es vier verschiedene Aktualisierungsfunktionen: die zwei gefrorenen Funktionen, die unabhängig vom Eingangswert immer den selben Ausgang haben, eine Funktion, die den Eingangswert kopiert und eine Funktion, die den Eingangswert invertiert. Diese werden den Knoten mit gleicher Wahrscheinlichkeit zugeordnet, was bedeutet, dass die Hälfte aller Knoten eine konstante Funktion hat und damit bei $t = 1$ in den beiden Netzwerkkopien mit Sicherheit im selben Zustand ist. Die Wahrscheinlichkeit, dass ein Knoten zum Zeitpunkt $t = 1$ in den beiden Kopien nicht den selben Zustand annimmt ist dann $1/2$ mal der Wahrscheinlichkeit H/N , dass der Zustand seines Eingangsknotens sich zum Zeitpunkt $t = 0$ in den beiden Netzwerken unterscheidet.
2. Für $K = 2$ hat die Anfangssteigung des Derrida-Plots den kritischen Wert Eins. Die Wahrscheinlichkeit, dass der Zustand eines Knotens sich zum Zeitpunkt $t = 1$ in den beiden Netzwerkkopien unterscheidet ist

$$\frac{H(1)}{N} = r \left(1 - (1 - H(0)/N)^K \right). \quad (2.12)$$

Das ist die Wahrscheinlichkeit, dass sich mindestens der Zustand eines seiner Eingangsknoten in den beide Kopien unterscheidet mal der Wahrscheinlichkeit r , dass ein Unterschied in einem oder mehreren Eingängen zu einem Unterschied im Ausgang führt. Die Steigung der Funktion $H(1)$ von $H(0)$ ist rK für kleine $H(0)$. Wenn alle Aktualisierungsfunktionen mit gleicher Wahrscheinlichkeit ausgewählt werden (wie es hier geschehen ist), gilt $r = 1/2$.

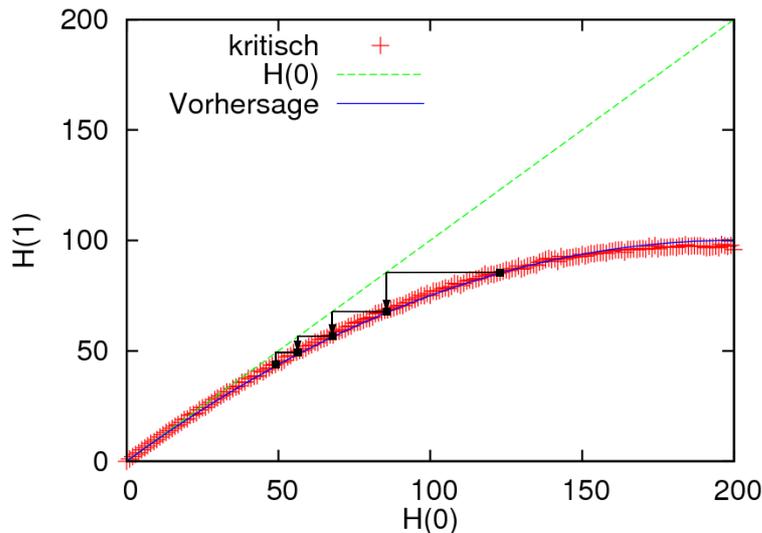


Abbildung 2.4: Derrida-Plot eines kritischen Booleschen Zufallsnetzwerkes mit $N = 200$, $K = 2$. Quadrate (■) und Pfeile sollen die zeitliche Entwicklung des Hamming-Abstandes verdeutlichen.

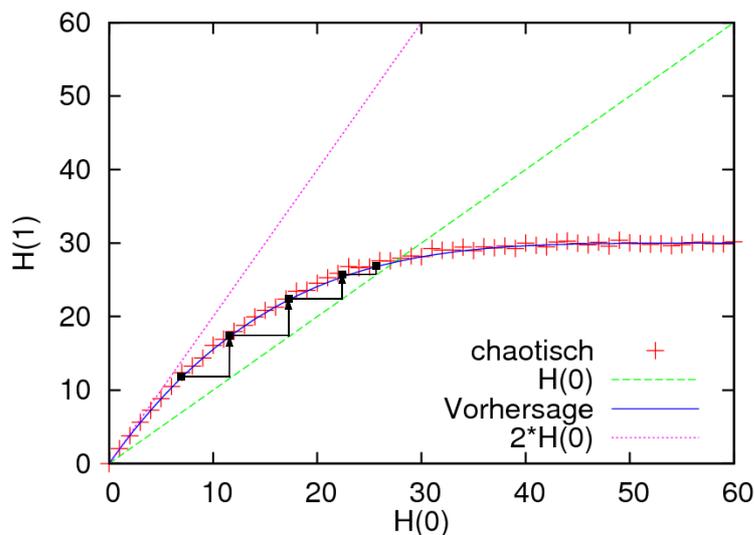


Abbildung 2.5: Derrida-Plot eines chaotischen Booleschen Zufallsnetzwerkes mit $N = 60$, $K = 4$. Quadrate (■) und Pfeile sollen die zeitliche Entwicklung des Hamming-Abstandes verdeutlichen.

3. In der chaotischen Phase, mit $K > 2$ und $N \gg K$, ist die Anfangssteigung der Derrida-Abbildung $Kr = K/2$. Die Wahrscheinlichkeit, dass sich der Zustand eines Knotens zum Zeitpunkt $t = 1$ in den beiden Kopien unterscheidet, ist wieder durch Gleichung 2.12 mit $r = 1/2$ gegeben.

Obwohl die simulierten Netzwerke klein sind, zeigt sich eine gute Übereinstimmung ihrer Derrida-Plots mit den Vorhersagen der Annealed Approximation. Das Verhältnis von $H(1)$ und $H(0)$, das man mit Gleichung 2.12 erhält, stimmt sehr gut mit den Ergebnissen der Simulationen von kritischen und chaotischen Netzwerken überein (s. Abb. 2.4 und 2.5).

Wie schon erwähnt wird einerseits die Anfangssteigung und andererseits der Wert des Hamming-Abstandes im stationären Zustand herangezogen, um zwischen den Phasen un-

terscheiden zu können. Die Fixpunkte der Abbildung kann man dort ablesen, wo sich $H(1)$ und die Gerade $H(0)$ schneiden. In den Abbildungen 2.4 und 2.5 zeigen die Quadrate (■) zeitlich aufeinanderfolgende Werte des Hamming-Abstands an, der sich mit jeder Iteration dem stabilen (im Falle des kritischen Netzwerkes dem marginal stabilen) Fixpunkt nähert. Die Stabilität eines Fixpunktes ist durch die Steigung der Funktion an diesem Punkt gegeben. Ist die Steigung kleiner als Eins, ist der Fixpunkt stabil, ist sie größer als Eins, ist er instabil. In Abbildung 2.5 sind zwei Fixpunkte zu sehen, ein instabiler bei $H = 0$ und ein stabiler bei einem endlichen Wert von H , auf den die zeitliche Entwicklung des Hamming-Abstandes führt.

Während die Steigung der Derrida-Abbildung im Ursprung gleich ist für gleiche mittlere Eingangswerte, hängt die exakte Form der Derrida-Abbildung auch von der Eingangsgradverteilung ab [49].

2.3.3 Sind Zellen kritisch?

Kauffman argumentierte, ein reales Genregulationsnetzwerk müsse in der kritischen Phase oder zumindest nahe an der Grenze in der gefrorenen Phase liegen, um sowohl kontrolliert komplexe Aufgaben erfüllen zu können als auch evolvierbar zu sein [45]. Obwohl inzwischen gezeigt wurde, dass kritische Boolesche Netzwerke biologisch unrealistische Attraktorlängen und -anzahlen haben [75, 62], ist diese Vorstellung immer noch lebendig. Hierbei wird ein Konzept aus der Gleichgewichtsphysik auf biologische Systeme angewendet, die weit entfernt vom Gleichgewicht sind. Es gibt einige Studien, die anhand von Daten zur Genexpression versuchen, Zellen einem der dynamischen Regime zuzuordnen. Für die Bäckerhefe wird mit Hilfe von Lawinengrößenverteilungen nach so genannten „Gen-Knockout“-Experimenten ein Wert für λ abgeschätzt, der zwischen 0,85 und 1 liegt [70, 78]. Shmulevich et al. [80] betrachten die Informationsmenge, die in einer Zeitreihe von Genexpressionsdaten von HeLa-Zellen (einer Krebs-Zelllinie) gespeichert ist, indem sie die Anzahl an neuen Wörtern in dieser Reihe bestimmen und sie mit den entsprechenden Daten für Boolesche Zufallsnetzwerke aus allen drei Regimen vergleichen. Auch sie erhalten das Ergebnis, dass die Zellen kritisch oder gefroren sein müssen. Nykter et al. [65] schließlich messen die Informationsdistanz für zwei aufeinanderfolgende Zeitpunkte aus Transkriptionsmessungen in Mäusemakrophagen und finden kritisches dynamisches Verhalten.

Entsprechend der Vorstellung, dass Zellen kritisch seien oder eine Dynamik knapp unter der kritischen Grenze im gefrorenen Bereich aufweisen müssten, sind viele evolutionäre Studien bemüht einen Mechanismus zu finden, der Netzwerke hin zu Kritikalität evolviert [66, 8, 60, 13, 58].

In Kapitel 6 dieser Arbeit werden einige der in Kapitel 2.3.2 besprochenen Unterscheidungsmerkmale der dynamischen Regime für evolvierte Boolesche Netzwerke betrachtet. Insbesondere werden der Derrida-Plot und einige von ihm abgeleitete Abbildungen für evolvierte und zufällige Ausgangsnetzwerke verglichen. Am Ende stellt sich die Frage, ob diese einfache Klassifizierung in gefroren, kritisch und chaotisch sinnvoll ist, wenn die Netzwerke nicht zufällig zusammengesetzt sind, sondern z.B. durch ihre evolutionäre Geschichte geformt sind, wie es auch bei biologischen Netzwerken der Fall ist.

2.4 *Saccharomyces cerevisiae* – Boolesches Modell eines realen Netzwerks

Das wohl berühmteste Beispiel, bei dem die Funktion eines realen regulatorischen Netzwerks mit Hilfe einer Booleschen Idealisierung seiner dynamischen Eigenschaften nachgestellt werden konnte, ist das Zellzyklusregulationsnetzwerk der Bäckerhefe *Saccharomyces cerevisiae* [57].

Der Zellzyklusprozess, den eine Zelle zwischen einer Zellteilung und der nächsten durchläuft, besteht aus vier Phasen. In der G_1 -Phase wächst die Zelle und geht im Anschluss unter geeigneten Bedingungen zur Zellteilung über. In der S -Phase wird die DNA repliziert. In der G_2 -Phase bereitet sich die Zelle auf die Mitose vor und in der M -Phase trennen sich die Chromosomen und die Zelle teilt sich. Damit ist ein Zyklus vollendet und die zwei Tochterzellen befinden sich wieder in der G_1 -Phase. Am Zellzyklusprozess der Bäckerhefe sind ca. 800 Gene beteiligt, doch die Anzahl an Schlüsselgenen, die den Prozess kontrollieren und regulieren, ist viel kleiner. Basierend auf extensiven Literaturstudien, konstruieren Li et al. ein Netzwerk der Schlüsselregulatoren (s. Abb. 2.6). Die Dynamik auf dem Netzwerk wird durch Schwellenwertfunktionen wie in Gleichung 2.10 bestimmt mit einem Schwellenwert von 0. Für die Knoten gibt es also nur zwei Zustände: 1 oder 0. Die grünen Pfeile in Abb. 2.6 sind aktivierende Verbindungen, die roten repräsentieren Deaktivierung. Die gelben Pfeile bedeuten „Selbstdegradation“. Sie werden bei Knoten eingeführt, die keine negative Regulation durch andere erfahren, und stellen eine Vereinfachung der tatsächlichen Prozesse dar. Der Zellzyklusprozess beginnt mit einem „Zellgrößensignal“, wenn die Zelle die richtige Größe für die Teilung erreicht hat, und führt über eine wohldefinierte Folge von Zuständen wieder zurück zum stationären G_1 -Zustand. Der Zustandsraum dieses Booleschen Netzwerkes besteht aus sieben Fixpunkten, wobei das Basin des stationären G_1 -Zustands ca. 86% aller Zustände beinhaltet. Das bedeutet eine hohe Stabilität dieses Zustands. Startet man von diesem Zustand und aktiviert das „Zellgrößensignal“, folgt das Netzwerk der biologischen Zellzyklussequenz. Alle Zustände, die zum G_1 Zustand führen, konvergieren zudem zum biologischen Pfad, was ihn zu einer anziehenden Trajektorie der Dynamik macht (s. Abb. 2.7). Überdies ist das Netzwerk stabil gegen Mutationen in den Verbindungen. Li et al. betrachten das Löschen, das Hinzufügen und das Ändern des Vorzeichens einer Verbindung und finden, dass zwischen 41% und 65%

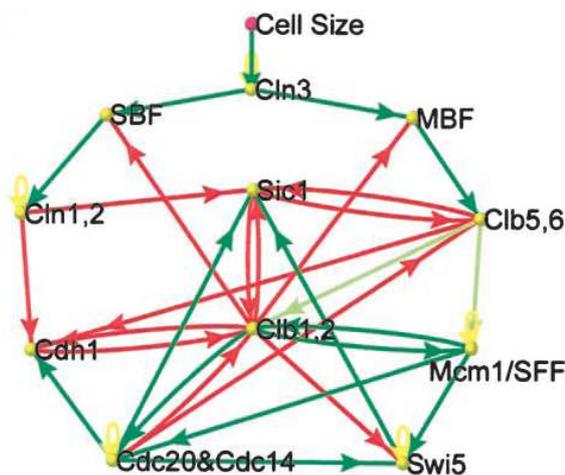


Abbildung 2.6: Zellzyklusregulationsnetzwerk der Bäckerhefe nach Li et al. [57].

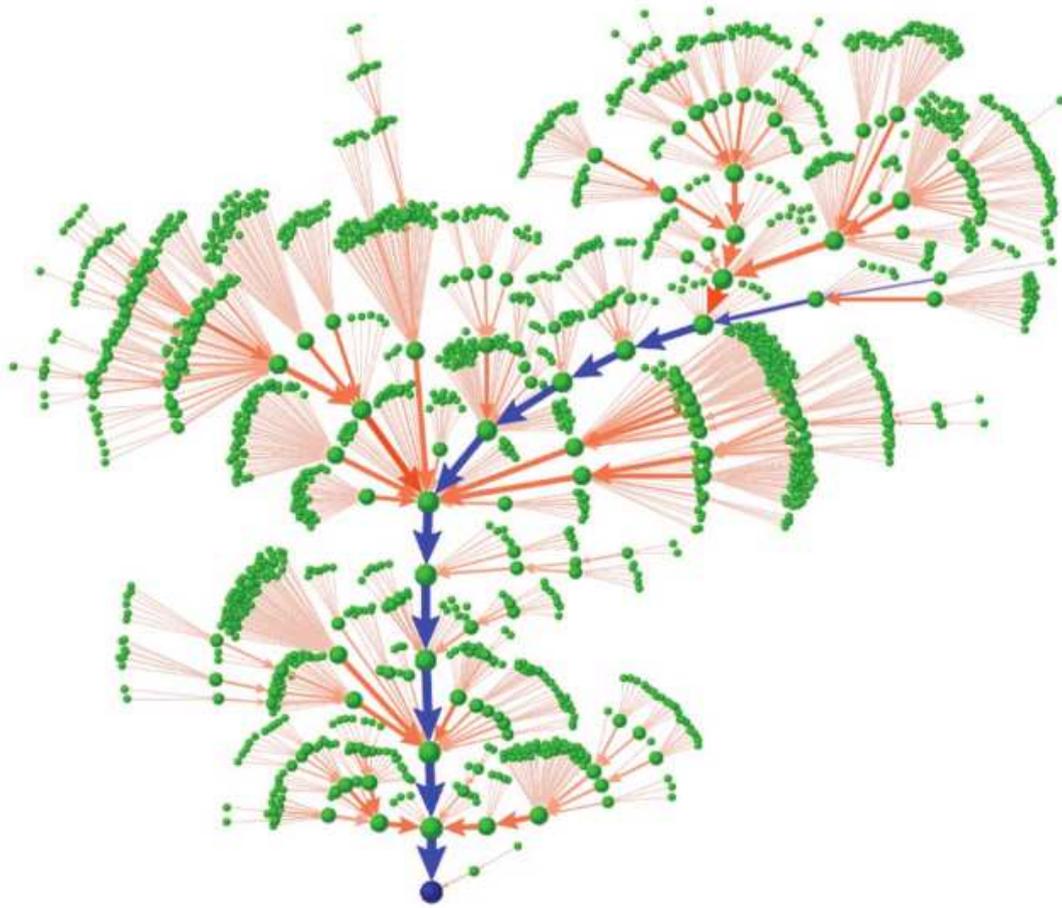


Abbildung 2.7: Ausschnitt des Zustandsraums des Netzwerks aus Abb. 2.6, das Basin des G_1 -Fixpunkts (blauer Knoten). Die Zellzyklussequenz ist mit blauen Pfeilen gekennzeichnet. Die Größe eines Knotens und die Dicke eines Pfeiles sind proportional zum Logarithmus des Flusses durch sie [57].

der Netzwerke für die verschiedenen Mutationen immer noch den G_1 -Fixpunkt erreichen und dass die biologische Trajektorie die wahrscheinlichste ist [57].

Seit dem Erscheinen der Arbeit haben sich zahlreiche weitere Publikationen angeschlossen, die das System weitergehend untersuchen. Sie reichen von zusätzlichen Robustheitsstudien [16, 96] über Analysen von strukturell verschiedenen Netzwerken, die die Zellzyklussequenz reproduzieren [55], bis zu evolutionären Betrachtungen, die den neutralen Raum dieser funktionalen Netzwerke untersuchen [12], oder die gar von einem zufälligen Netzwerk mit elf Knoten hin zu einem evolvieren, das die biologisch relevante Trajektorie durchläuft [38].

Das Modell der Zellzyklusregulation der Bäckerhefe wird an zwei Stellen in dieser Arbeit noch einmal näher betrachtet. In Kapitel 4.3 werden die aus der Untersuchung der Netzwerke mit Schwellenwertfunktionen, mit denen auch das Netzwerk der Bäckerhefe modelliert wurde, gewonnen Erkenntnisse auf dieses Netzwerk angewendet. In Kapitel 6.2.2 dient es bei der Analyse von Derrida-Plots von evolvierten und zufälligen Booleschen Netzwerken als Beispiel eines durch Evolution geformten biologischen Netzwerkes.



3 Modelle der Evolution

Evolution bedeutet Veränderung in Form und Verhalten von Organismen zwischen Generationen innerhalb einer Population einer Spezies [72]. Das Material, das dabei von Generation zu Generation weitergegeben wird, ist die genetische Information. Ihre Veränderung kann eine Veränderung der Merkmale eines Organismus bewirken. Eine Änderung in der DNA-Sequenz, die vererbt wird, heißt Mutation.

Es ist ein weit verbreiteter Glaube, dass die sehr seltenen vorteilhaften, zufälligen Punktmutationen während der DNA-Replikation (Substitution einer Nukleinsäure durch eine andere oder Löschung/Einfügung einer Nukleinsäure) die Hauptquelle der genetischen Variationen sind, die den Prozess der Evolution antreiben. Es gibt jedoch auch Mechanismen, durch die das Genom auf weniger zufällige Art verändert wird. Zu ihnen gehören die Veränderungen der Erbinformation durch Transposonen. Das sind DNA-Abschnitte, die ihre Anzahl und Position im Genom verändern können (s. Kap. 5.5) [25].

Die genetische Information einer Zelle oder eines Organismus wird als Genotyp bezeichnet. Demgegenüber steht der Phänotyp, die Ausprägung dieser Information, die Merkmale eines Organismus, die sowohl von seinen Genen als auch von seiner Umwelt abhängen.

Die Basis der Evolution bilden Mutationen (und eventuell sexuelle Rekombination) und natürliche Selektion. Dabei wirkt die Selektion nicht direkt auf den Genotyp, sondern auf den Phänotyp. Ist dieser gut angepasst, hat also eine hohe Fitness, wird er überleben, sich reproduzieren und Teile seines genetischen Materials an nachfolgende Generationen weitergeben. Dabei kann es für einen Genotyp von Vorteil sein mehrere verschiedene Phänotypen zu produzieren, die an verschiedene Umweltbedingungen angepasst sind [54].

Der Begriff der Genotyp-Phänotyp-Abbildung bezeichnet die Art und Weise wie sich genetische Variation auf phänotypische Variation abbildet. Die Abbildung von RNA-Sequenzen auf RNA-Strukturen stellt dabei das einfachste und das einzig handhabbare Beispiel einer Genotyp-Phänotyp-Abbildung dar (s. Abschnitt 3.1.1). Auch die in dieser Arbeit betrachteten Netzwerke liefern ein simples Modell einer Genotyp-Phänotyp-Abbildung, wobei die Netzwerkstruktur und die Funktionen den Genotyp bilden und die Attraktoren, die ein Netzwerk erreichen kann, die möglichen phänotypischen Ausprägungen.

Ein zentrales Konzept der biologischen Evolution ist, dass sich die Individuen in einer Population in der Angepasstheit an ihre Umwelt (Fitness) unterscheiden. Interessant sind diejenigen Unterschiede, die genetische Grundlagen haben, da nur Veränderungen des Erbmaterials Anpassung und Neuerungen dauerhaft machen. Die Metapher der Fitnesslandschaft bietet die Möglichkeit sich diese Beziehung zwischen Genom und Fitness zu veranschaulichen, sei es als eine direkte Abbildung oder durch den Phänotyp vermittelt [36].

3.1 Fitnesslandschaften

Im klassischen Sinne wird die Fitness in Fitnesslandschaften als eine Eigenschaft eines Individuums gesehen, die in Form der Überlebens- und Reproduktionsfähigkeit gemessen werden kann und durch seine Gene bestimmt wird.

Die Metapher der Fitness- oder der adaptiven Landschaft wurde von Sewall Wright in seinem Artikel von 1932 eingeführt [95]. Bei der Visualisierung des Problems, dass nur ein kleiner

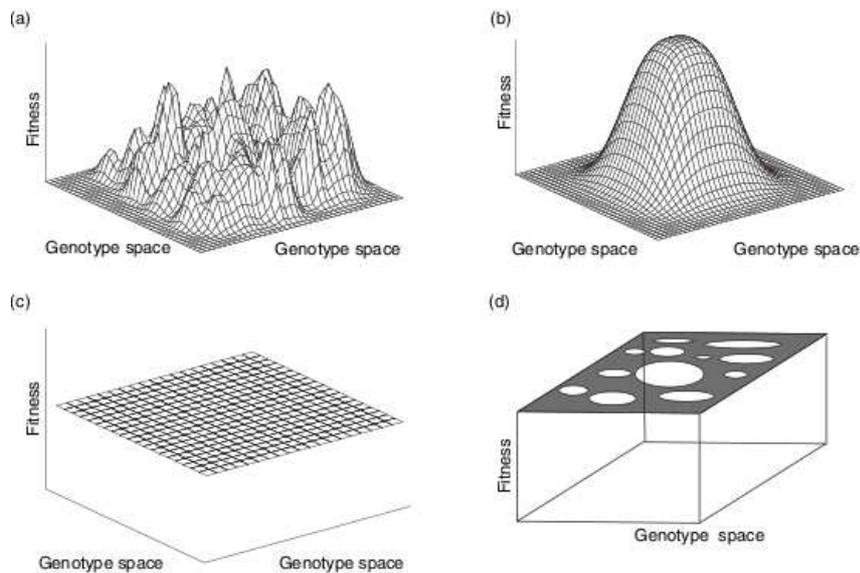


Abbildung 3.1: Vier Arten von Fitnesslandschaften: (a) raue Landschaft, (b) Landschaft mit einem einzelnen Gipfel, (c) flache Landschaft, (d) löchrige Landschaft [36]

Bruchteil der möglichen genetischen Kombinationen für eine Population zu einem bestimmten Zeitpunkt zugänglich ist, führte er auch die Idee der schrittweisen Bewegung auf einer genetischen Landschaft ein, die später von Kauffman und Levin [46] formalisiert wurde [67]. Wright formulierte auch das Problem der rauen Landschaften (s. Abb. 3.1 (a)): Selektion würde eine Spezies zwar leicht zum nächsten Gipfel führen, aber es könnte unzählige andere höhere Gipfel geben, die durch Täler voneinander getrennt wären. Welcher Art sollte also der Mechanismus sein, durch den eine Spezies in einer solchen Landschaft kontinuierlich Wege von kleineren zu größeren Gipfeln finden kann? Dieser Frage liegt die Ansicht der Zwangsläufigkeit der Evolution zugrunde, dass es immer einen Fortschritt geben muss. Es gibt aber auch eine andere mögliche Betrachtungsweise, die die Selektion nicht als das „Überleben des Fittesten“ sondern als das „Nicht-Überleben des Nicht-Fitten“ sieht. Das heißt, wenn etwas gut genug ist um zu überleben (und sich zu reproduzieren) wird die Wirkung der natürlichen Selektion nicht darin bestehen die Fitness zu verbessern, auch wenn die Lösung suboptimal ist im Hinblick auf alle möglichen Lösungen, sondern die Population auf dem aktuellen lokalen Gipfel zu halten [67]. Die zweite Frage, die sich ergibt, wenn man raue Fitnesslandschaften betrachtet, ist, wie neue Spezies entstehen können. Ein Ansatz zur Lösung dieser beiden Fragestellungen ist die genetische Drift, die die zufällige Veränderung der Häufigkeit eines Allels (Ausprägung eines Gens) in einer Population bezeichnet.

Fishers Theorie zur natürlichen Selektion führt zum Bild einer Population, die sich immer nur bergauf bewegen kann, ohne durch Fitnessstäler gehen zu müssen, und so schließlich das globale Fitnessmaximum erreicht [25]. Dies impliziert eine Fitnesslandschaft mit nur einem Gipfel (s. Abb. 3.1 (b)), obwohl Fisher selbst die Idee der adaptiven Landschaften ablehnte [68].

Die neutrale Theorie der Evolution, zuerst von Kimura vorgeschlagen, besagt, dass die meisten Mutationen auf dem molekularen Level neutral sind, also die Fitness nicht ändern [50, 51]. In einer Fitnesslandschaft würde die Bewegung mit Hilfe von neutralen Mutationen in einer

flachen Ebene resultieren (s. Abb. 3.1(c)). Natürlich treten in der neutralen Theorie auch viele negative (fitnesssenkende) Mutationen und wenige positive Mutationen auf.

Gavrilets führte den Begriff der „löchrigen Fitnesslandschaft“ ein. Ausgehend von der Feststellung, dass einem hochdimensionalen Genotypraum eine kleine Anzahl von Fitnesswerten gegenübersteht und damit verschiedene Genotypen sehr ähnliche Fitnesswerte haben müssen, stellt er fest, dass gut angepasste Genotypen verbundene Cluster bilden könnten [35]. Aus einigen simplen Modellen schließt er, dass die Existenz von perkolierenden fast-neutralen Netzwerken von Genotypen mit hoher Fitness eine generelle Eigenschaft von adaptiven Landschaften mit einer großen Anzahl von Dimensionen zu sein scheint. Eine Darstellung einer solchen adaptiven Landschaft wäre eine flache Oberfläche mit Löchern, die die Genotypen repräsentieren, die nicht zu der perkolierenden Menge dazugehören (s. Abb. 3.1(d)). Dabei werden Fitnessunterschiede zwischen den gut angepassten Genotypen vernachlässigt. Populationen könnten sich auf einer solchen Landschaft entlang der Cluster bewegen und sich genetisch auseinanderentwickeln, ohne durch Täler gehen zu müssen. Speziation tritt auf, wenn eine Population bei einem Genotyp angelangt, der durch ein Loch von ihrem Ausgangsgenotyp getrennt ist [36].

Jede der vier vorgestellten Metaphern betont bestimmte Aspekte von Fitnesslandschaften und evolutionärer Dynamik, während sie andere vernachlässigt. Fitnesslandschaften von realen biologischen Systemen sind im Allgemeinen unbekannt. Erst in letzter Zeit haben Untersuchungen von konkreten Landschaften wie denen von RNA oder Proteinen begonnen. Empirische Unterstützung gibt es für alle vier vorgestellten Metaphern von Fitnesslandschaften [36].

Auch die Diskussion über das Verständnis, die Art und die Zweckmäßigkeit von Fitness- bzw. adaptiven Landschaften ist immer noch aktuell [42, 69, 18].

3.1.1 Neutrale Evolution am Beispiel von RNA- und Proteinmodellen

Wie schon zuvor angesprochen, ist die Abbildung von RNA-Sequenzen auf die räumlichen Strukturen, in die sich die Moleküle falten, die einfachste Genotyp-Phänotyp-Abbildung. Es gibt einige gängige RNA-Strukturvorhersage-Algorithmen, die verwendet werden, um zu einer Sequenz die Sekundärstruktur mit der minimalen freien Energie zu bestimmen. Eine Sekundärstruktur ist gegeben durch eine Liste der Watson-Crick- (A-U, G-C) und G-U-Basenpaare (s. Abb. 3.2). Sekundärstrukturen decken den dominierenden Part der dreidimensionalen Faltungsenergien ab und können oft erfolgreich verwendet werden bei der Interpretation der Funktion und der Reaktivität der Moleküle [77]. RNA-Sequenzen mit der selben Faltung, die miteinander durch Punktmutationen verbunden sind, bilden ein sogenanntes neutrales Netzwerk.

Berechnet man eine obere Grenze für die Anzahl an möglichen planaren Sekundärstrukturen mit minimaler freier Energie für eine Kettenlänge n (unter Annahme von sterischen Beschränkungen und einer Mindestanzahl von aufeinanderfolgenden Basenpaaren), ergibt sich eine Zahl, die konsistent kleiner ist, als die Zahl der Sequenzen [77].

Schuster et al. betrachten in ihrem Artikel von 1994 [77] die Faltung von zufälligen Sequenzen mit fester Kettenlänge und bestimmen damit zunächst wie häufig die einzelnen Strukturen im Sequenzraum sind. Sie finden einige wenige Formen, die häufig vorkommen und viele, die selten sind. In ihren Computorexperimenten finden sie außerdem, dass 21,7% aller neutralen Wege (Wege, die die Sekundärstruktur erhalten, bestimmt mit monoton anwachsendem

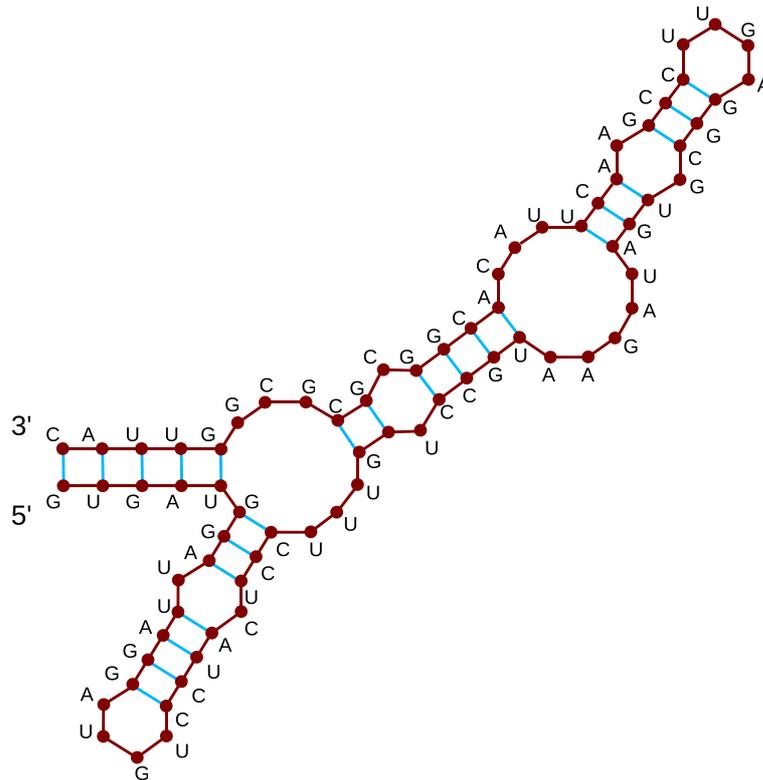


Abbildung 3.2: Typische Visualisierung einer RNA-Sekundärstruktur nach Fontana [32]. Eine RNA-Sekundärstruktur ist ein Kontaktgraph zwischen Nukleotiden an den Positionen $i = 1, \dots, n$ entlang der Sequenz. Position 1 ist das 5'-Ende. Der Graph hat zwei Arten von Kanten: das Rückgrat, das Nukleotid i mit Nukleotid $i + 1$ verbindet (rot) und Wasserstoffbrückenbasenpaarungen zwischen nicht-benachbarten Positionen (blau).

Abstand von einer Referenzsequenz) durch den gesamten Sequenzraum perkolieren und bei Sequenzen enden, die vollkommen verschieden sind von der Referenzsequenz.

Würde man eine neutrale Änderung so definieren, dass sie neutral in jeder Umgebung, unter jeder physiologischen Bedingung oder jedem genetischem Hintergrund sein muss, könnte sich eine neutrale Evolutionstheorie nicht halten. Eine neutrale Mutation ist so zu sehen, dass sie einen bestimmten Aspekt der Funktion eines biologischen Systems unverändert lässt, in einer bestimmten Umgebung und einem bestimmten genetischen Hintergrund. Andere Aspekte können dabei sehr wohl Neuerungen unterliegen und eine neutrale Veränderung kann zu einer nicht-neutralen werden, wenn sich die umgebenden Bedingungen ändern [90].

Im Falle der RNA entspricht eine neutrale Mutation einer Basensubstitution, die die Struktur mit der minimalen freien Energie erhält. Sie kann aber vieles andere beeinflussen wie die freie Energie selbst oder die Strukturen, die unter bestimmten Temperaturbedingungen zugänglich sind. Evolution entlang eines neutralen Netzwerkes führt zur Anhäufung von zunächst stillen Mutationen und kann so schließlich phänotypische Neuerungen herbeiführen [32].

Betrachtet man zwei zufällige Sequenzen, die Teil eines neutralen Netzwerkes sind, findet man, dass die meisten Strukturen in ihren Nachbarschaften sich voneinander unterscheiden. Mit anwachsender Größe des neutralen Netzwerkes verringert sich zwar die Anzahl verschiedener Strukturen in der 1-Nachbarschaft jeder Sequenz doch dieser Effekt wird mehr als kompensiert.

siert durch die viel höhere Anzahl von verschiedenen Strukturen, die mit Hilfe eines größeren neutralen Netzwerkes erreicht werden können. So haben Populationen, die auf größeren neutralen Netzwerken evolvieren, Zugang zu mehr phänotypischen Neuerungen [91].

Auch Untersuchungen biologischer RNA-Strukturen haben ergeben, dass es eine sehr große Anzahl von Sequenzen gibt, die sich in die gleichen Strukturen falten. Diese Anzahl ist größer als die von den meisten zufälligen Phänotypen [40]. In einer weiteren Studie zur Abschätzung von realen neutralen Netzwerken betrachten Schultes et al. zwei Ribozyme (RNA-Moleküle, die wie Enzyme chemische Reaktionen katalysieren) mit sehr unterschiedlichen Funktionen: zum einen eine synthetische Ligase (katalysiert eine chemische Verbindung zweier Moleküle), zum anderen ein sich selbst spaltendes Ribozym des Hepatitis-D-Virus (HDV) [76]. Sie konstruieren eine Sequenz, die gleichzeitig die erforderlichen Basenpaarungen beider Ribozyme erfüllen kann und dabei die meisten der Residuen erhält, die für die Aktivität jedes der Ribozyme wichtig sind. Obwohl die Aktivitäten verlangsamt sind, kann die eine Sequenz beide Reaktionen katalysieren. Durch Punktsubstitutionen von Nukleotiden in dieser gemeinsamen Sequenz finden Schuster et al., dass nur vier Mutationen zwei Ribozymsequenzen separieren, die zwei völlig verschiedene Faltungen annehmen und nahezu Prototyp-Aktivitäten aufweisen. Anschließend konstruieren sie neutrale Wege von diesen beiden Sequenzen zu dem jeweiligen Prototyp. Die Leichtigkeit, mit der diese beiden Wege gefunden werden konnten, bestärkt die Idee, dass es viele neutrale Wege gibt für jede Ribozymfaltung. Da die beiden Ribozyme weder eine evolutionäre Geschichte noch strukturelle Eigenschaften gemeinsam haben, ist es nicht unwahrscheinlich, dass auch die neutralen Netzwerke anderer Paare von Ribozymen einander nahe kommen. Weiterhin schließen Schultes et al., dass es für Ribozyme ähnlicher Größe viele neutrale Netzwerke geben könnte, die einander nahe kommen, und dass sogar die Netzwerke aller Ribozyme nicht zu weit voneinander entfernt sein könnten.

Die Proteinfaltung ist ein vielfach komplexeres Problem, das auch die Untersuchung der Verteilung und Konnektivität von Sequenzen mit der selben Faltung im Sequenzraum ungleich schwieriger macht. Die Anzahl von theoretisch möglichen Proteinen ist gewaltig (Für Kettenlängen von $n = 100$ Aminosäuren gibt es 20^{100} Sequenzen), hingegen scheint die Zahl der stabilen nativen Tertiärstrukturen limitiert zu sein [6]. Üblicherweise geht man davon aus, dass es nur einige 1000 verschiedene gibt [11]. Hinzu kommt, dass scheinbar nicht miteinander in Beziehung stehende Sequenzen im Wesentlichen die gleiche Struktur annehmen [6].

Babajide et al. [6] untersuchen die neutralen Netzwerke für einige Proteine mit bekannter dreidimensionaler Struktur mit Hilfe von Techniken zur inversen Faltung. Das heißt, es wird bestimmt, welche Aminosäuresequenzen sich in eine gegebene Struktur falten. (Sie gehen davon aus, dass die Genauigkeit bei den von ihnen benutzten Methoden bei ca. 65% liegt.) Als Qualitätsmessung einer Struktur in einer bestimmten Faltung, verwenden sie den z -Wert. Dieser misst die Differenz zwischen der Energie der Sequenz x in einer bestimmten Faltung ψ und der mittleren Energie in einer Menge von alternativen Faltungen. Die Faltung von x in die Struktur ψ wird dann akzeptiert, wenn der z -Wert mindestens so gut ist wie der Wert der nativen Sequenz. Eine Adaptive Wanderung (s. auch Kap. 5) bestimmt einen neutralen Weg durch den Sequenzraum, indem zufällige Punktmutationen durchgeführt werden (Austausch einer Aminosäure), wobei der Hamming-Abstand zur Anfangssequenz ansteigen muss. Eine Mutation wird akzeptiert, wenn sich die neue Sequenz in ψ faltet und der z -Wert abnimmt und unter einem Schwellenwert z^* liegt. Die gefundenen neutralen Netzwerke erstrecken sich über den gesamten Sequenzraum.

Auch Bastolla et al. [11] finden ausgedehnte neutrale Netzwerke im Sequenzraum. Sie verwenden für ihre Untersuchung das Gitter-Modell von Proteinen, da damit das Proteinfaltungsproblem handhabbar wird. Dabei wird eine Proteinkonfiguration als „self-avoiding walk“ (ein sich nicht kreuzender Weg) auf einem Gitterwürfel dargestellt. Jeder der N besetzten Gitterplätze repräsentiert eine Aminosäure, ausgewählt aus den 20 möglichen. Bei dem evolutionären Algorithmus wird dann in jedem Zeitschritt eine Aminosäure mutiert und die Mutation akzeptiert, wenn drei Bedingungen erfüllt sind: der Grundzustand bleibt erhalten, ist thermodynamisch stabil und leicht erreichbar.

Für eine weitere Studie [10] werden sieben der meist untersuchten Proteinfaltungen herangezogen. Hier wird ein Protein durch seine Kontaktmatrix beschrieben, die einen Eintrag 1 hat, wenn zwei Aminosäuren in der Kette in Kontakt sind, und 0 andernfalls. Die freie Energie wird aus der Summe der Kontaktwechselwirkungen berechnet. Auch hier wird eine inverse Faltungstechnik verwendet und der z -Wert sowie ein weiterer Parameter, der die Stabilität einer Faltung beschreibt, als Selektionsparameter benutzt. Für jede der sieben Proteinfaltungen können ausgedehnte neutrale Netzwerke im Sequenzraum gefunden werden.

Die neutrale Theorie der Evolution erfährt durch Computersimulationen und Experimente zunehmend Aufschwung. Die ausführlichsten Studien gibt es an RNA- und Proteinmodellen, doch auch bei der Untersuchung von Netzwerkmodellen der Genregulation finden sich ausgedehnte neutrale Netzwerke [21, 20, 29] (s. Kapitel 7). Die in dieser Arbeit durchgeführten Evolutionssimulationen liefern ein weiteres Beispiel. Alle Selektionskriterien führen auf weitläufige neutrale Ebenen in der Fitnesslandschaft, auf denen sich die Netzwerke weit durch den Raum der möglichen Realisationen bewegen können (s. Kapitel 7).

3.2 Computersimulationen von Netzwerkevolution

Evolutionäre Studien an Booleschen Netzwerken haben bereits eine lange Tradition. Die einfachsten möglichen Mutationen in einem solchen Netzwerk betreffen die Verbindungen und die Funktionen der Knoten, wenn die Netzwerkgröße konstant gehalten werden soll, und können andernfalls auch das Löschen oder Hinzufügen von Knoten gemeinsam mit ihren Verbindungen bedeuten. Kompliziertere Simulationen mit Populationen von Netzwerken können auch die Rekombination zwischen zwei Netzwerken beinhalten. Bisherige Veröffentlichungen nutzen nur eine Untermenge der möglichen Mutationen. In [47] sind dies das Umverknüpfen von Verbindungen oder das Ändern von Bits der Booleschen Funktionen. Das Maß für die Fitness ist dort der Hamming-Abstand eines Attraktors zu einem vorher festgelegten Zielzustand. Diese Art von simulierter Evolution ist nicht in der Lage Netzwerke mit 100% Fitness hervorzubringen. In [14, 15] bestehen die möglichen Mutationen im Hinzufügen und/oder Entfernen einer Verbindung. Das Selektionskriterium ist die Robustheit gegenüber Mutationen, das heißt, das Tochternetzwerk muss den selben Attraktor erreichen wie das Mutternetzwerk. Die evolvierten Netzwerke haben kürzere Attraktoren und größere gefrorene Komponenten als zufällige Netzwerke. In [66, 8] repräsentieren die Knoten des Netzwerks konkurrierende Agenten und eine Vorschrift, die auf der Spieltheorie beruht, bestimmt den Knoten mit der geringsten Fitness. Die Aktualisierungsfunktion dieses Knoten wird geändert. Auf diese Weise evolviert das Netzwerk zu einem Zustand „am Rande des Chaos“ mit vielen kanalisierenden Funktionen. In [17] sind die Mutationen Umverknüpfungen von Verbindungen und die Netzwerke werden nach der Robustheit der Attraktoren gegenüber kleinen Störungen der Aktualisierungszeitpunkte selektiert. Nur wenige Mutationen werden benötigt, um einen Satz von vollkommen

stabilen Attraktoren zu erhalten. Während der Evolution sinkt die mittlere Anzahl an Attraktoren und die Größe des größten Einzugsgebietes nimmt zu. In [79] wird eine Population von Netzwerken betrachtet. Eine Mutation kann eine Verbindung hinzufügen, löschen oder die Gewichtung der Verbindung ändern. Die Fitness wird mit Hilfe des Hamming-Abstands zu einem am Anfang bestimmten Fixpunktzustand berechnet. Ein zufällig gewähltes Netzwerk wird mit einer Wahrscheinlichkeit die proportional zu seiner Fitness ist in die nächste Generation übernommen. Zusätzlich zu den Mutationen findet Rekombination zwischen zwei Netzwerken der nächsten Generation statt, indem Reihen der Konnektivitätsmatrizen vertauscht werden. Während der Evolution erhöht sich die Robustheit der Netzwerke gegen Mutationen, das heißt, sie kehren nach der Mutation einer Verbindung wieder auf den ursprünglichen Attraktor zurück. Netzwerke mit höherer Konnektivität erreichen dabei viel größere Robustheit. Das dynamische Verhalten der Netzwerke ändert sich hingegen kaum während der Evolution, die evolvierten Netzwerke zeigen sich nur ein wenig mehr geordnet. Die Verteilungen der Attraktoranzahl, der Attraktorlängen und der Transienzzzeiten zeigen eine Verschiebung hin zu solchen Verteilungen, die man für geordnetere Netzwerke mit geringerem Verknüpfungsgrad erwarten würde.

Der computersimulierten Evolution in dieser Arbeit wird kein Ziel vorgegeben, weder ein Operationsregime noch eine konkrete Funktion in Form eines Attraktors (oder allgemeiner einer Sequenz von Zuständen), um auszuloten, welche Mechanismen in der Lage sind zu solchen Netzwerken zu führen. Vielmehr sollen die Fitnessdefinitionen und die Mutationen biologisch relevant sein, um zu sehen welcher Art die Netzwerke sind, die so entstehen.



4 Der Phasenübergang in Schwellenwertnetzwerken

Bevor mit der evolutionären Untersuchung der Booleschen Netzwerke begonnen werden kann, ist es nötig das dynamische Verhalten des betrachteten Modells an sich zu kennen. Zu Zufallsnetzwerken mit allen Booleschen Funktionen und auch zu Booleschen Netzwerken mit kanalisierenden Funktionen gibt es schon zahlreiche Untersuchungen. Insbesondere sind die Phasendiagramme der Netzwerkmodelle bekannt (s. Kapitel 2.3). Zwar weiß man seit langem, dass Schwellenwertnetzwerke ein ähnliches Verhalten zeigen wie klassische Boolesche Zufallsnetzwerke [53], Studien zu zufälligen Booleschen Netzwerken mit den hier betrachteten Schwellenwertfunktionen existierten bislang allerdings noch nicht. Früheren Untersuchungen zu Phasendiagrammen von Schwellenwertnetzwerken liegen Netzwerkmodelle zugrunde, die sich von dem hier behandelten Netzwerkmodell in der Wahl der Knotenwerte σ_i , der Kopplungsparameter c_{ij} , der Schwellenwerte h oder auch in der Aktualisierungsfunktion selbst unterscheiden [53, 74, 73].

In dem hier behandelten Modell sind die Knoten zufällig miteinander verbunden, es gibt eine feste Anzahl von Eingängen pro Knoten und die Schwellen sind reellwertig. Die Zeitentwicklung der Zustände der Knoten ist durch Gleichung 2.10 gegeben. Mit den selben dynamischen Regeln, einem Schwellenwert von $h = 0$ und K -Werten, die von Knoten zu Knoten variieren, wurde auch das Zellzyklusregulationsnetzwerk der Bäckerhefe modelliert (s. Kapitel 2.4). Es existieren neben dieser, wie schon zuvor erwähnt, auch noch andere Varianten des Schwellenwertmodells. Modelle, die Knoten-Werte $\sigma_i = 1$ und 0 benutzen, können auf Modelle mit den Werten $r_i = \pm 1$ abgebildet werden, indem die Substitution $\sigma_i = (r_i + 1)/2$ durchgeführt wird. Für die in dieser Arbeit genutzte Aktualisierungsfunktion führt das zu

$$r_i(t+1) = \begin{cases} 1, & \sum_j c_{ij} r_j(t) > 2h - \sum_j c_{ij} \\ -1, & \sum_j c_{ij} r_j(t) < 2h - \sum_j c_{ij} \\ r_i(t), & \sum_j c_{ij} r_j(t) = 2h - \sum_j c_{ij}. \end{cases} \quad (4.1)$$

Das bedeutet, dass jeder Knoten i seinen eigenen Schwellenwert h_i hat, der von Werten der c_{ij} abhängt und die hier untersuchte Dynamik damit verschieden ist von der der ± 1 -Modelle, die weiter verbreitet sind.

Im Folgenden wird in diesem Kapitel die Annealed Approximation dazu benutzt, Ausdrücke für die zeitliche Entwicklung der Knoten in den Zuständen „an“ und „aus“ abzuleiten, um damit im Weiteren die Sensitivität λ zu bestimmen. Die Ergebnisse werden anschließend mit Simulationen von Netzwerken verglichen.

4.1 Das Phasendiagramm

Mit Hilfe der Annealed Approximation kann man die Parameterwerte für h und K bestimmen, für die die Netzwerke in der chaotischen oder in der gefrorenen Phase sind. Die Unterscheidung

erfolgt anhand der Sensitivität λ der Netzwerke. Um aber λ bestimmen zu können, muss b_t bekannt sein, der Anteil an Knoten im Zustand 1 zum betrachteten Zeitpunkt t . λ ist eine Funktion von b_t und wird nur konstant, wenn b_t einen Fixpunkt erreicht hat.

Wie schon angedeutet, wurde die Annealed Approximation bereits erfolgreich genutzt um das Phasendiagramm von verschiedenen Klassen von Booleschen Netzwerken vorherzusagen. In diesen Netzwerken sind die Korrelationen zwischen Knoten offensichtlich irrelevant für die Auswertung von b_t und λ . Weiter unten wird sich zeigen, dass das nicht für alle Schwellenwertnetzwerke gilt.

4.1.1 Die Entwicklung von b_t

Zunächst soll b_{t+1} als Funktion von b_t bestimmt werden. Dafür sind folgende Überlegungen vonnöten. Für nicht-ganzzahlige Werte von h wird ein Knoten im nächsten Zeitschritt den Wert 1 annehmen, wenn die Summe seiner Eingänge größer als h ist. Das heißt,

$$b_{t+1}^{(1)} = \sum_{m=\lfloor h \rfloor + 1}^K \binom{K}{m} \left[\sum_{l=\lfloor \frac{m+h}{2} \rfloor + 1}^m \binom{m}{l} \right] \left(\frac{b_t}{2} \right)^m (1 - b_t)^{K-m}. \quad (4.2)$$

Hierbei ist m die Anzahl der Eingänge mit dem Wert 1. Für positive Schwellenwerte müssen mindestens $\lfloor h \rfloor + 1$ der Eingangsknoten eingeschaltet sein (den Wert 1 haben), wenn die Summe der Eingänge über dem Schwellenwert liegen soll. Aus dem gleichen Grund muss die Anzahl l von positiven (aktivierenden) Kopplungen von diesen eingeschalteten Eingangsknoten mindestens $\lfloor \frac{m+h}{2} \rfloor + 1$ sein. Für negative h -Werte tragen auch alle Konfigurationen mit $m < |h|$ zu der Summe bei. Es gibt $\binom{K}{m}$ verschiedene Möglichkeiten m eingeschaltete Knoten aus den K Eingangsknoten auszuwählen und $\binom{m}{l}$ verschiedene Möglichkeiten l aktivierende Verbindungen unter den Verbindungen dieser eingeschalteten Knoten auszuwählen. Schließlich ist $b_t^m (1 - b_t)^{K-m}$ die Wahrscheinlichkeit, dass m Eingangsknoten im Zustand $\sigma_j = 1$ sind und die anderen im Zustand $\sigma_j = 0$, und $\left(\frac{1}{2}\right)^m$ die Wahrscheinlichkeit dafür, dass die positiven und negativen Kopplungen genauso verteilt sind wie sie es sind.

Für ganzzahlige h kann die Summe der Eingänge genau gleich dem Schwellenwert sein, was für nichtganzzahlige h nicht möglich ist. Im Rahmen der Annealed Approximation wird ein Knoten mit einer Eingangssumme von h im nächsten Zeitschritt mit einer Wahrscheinlichkeit b_t eingeschaltet sein. Gleichung 4.2 erhält deshalb noch einen zweiten Term für den Fall, dass h ganzzahlig ist und wird zu

$$b_{t+1}^{(2)} = b_{t+1}^{(1)} + \sum_{l=h}^{\lfloor \frac{K+h}{2} \rfloor} \binom{K}{2l-h} \binom{2l-h}{l} (1 - b_t)^{K-2l+h} \left(\frac{b_t}{2} \right)^{2l-h} \cdot b_t. \quad (4.3)$$

Hierbei ist l wieder die Anzahl der eingeschalteten Eingangsknoten mit positiven Kopplungen. Die Anzahl der eingeschalteten Knoten mit negativen Kopplungen muss $l - h$ sein, damit die Summe der Eingänge auf dem Schwellenwert liegen kann.

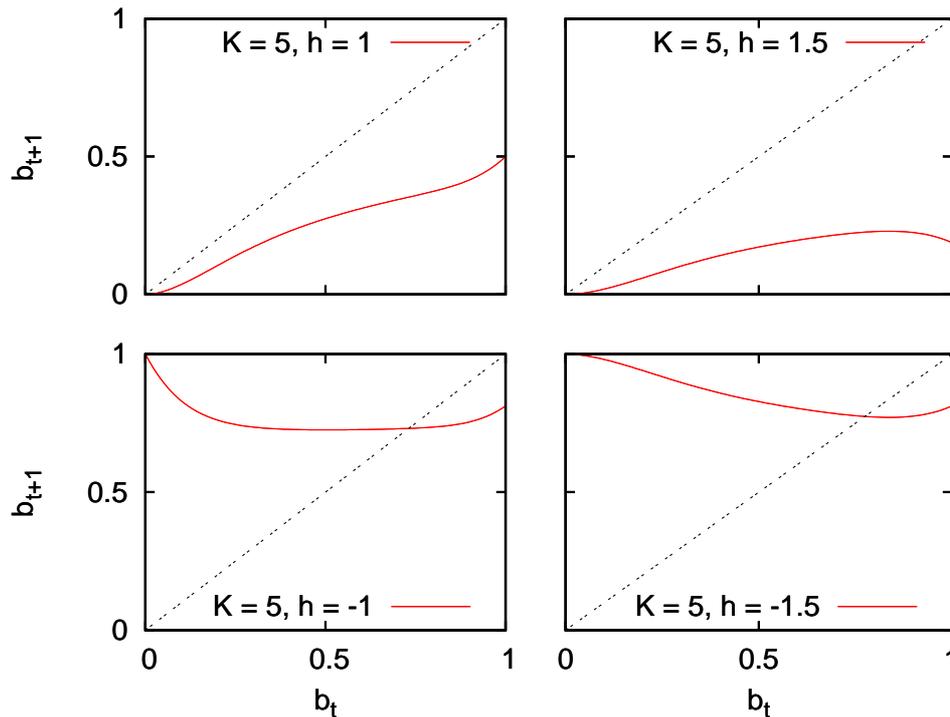


Abbildung 4.1: Die Abbildungen b_{t+1} vs. b_t für $K = 5$ und verschiedene h . Die gestrichelte Linie ist die Winkelhalbierende.

Mit der Rekursionsrelation für b_t kann man nun Abbildungen b_{t+1} vs. b_t für verschiedene h und K plotten. Die Fixpunkte b^* der Abbildung (s. Abb. 4.1) sind stationäre Lösungen der Annealed Approximation.

Ein Fixpunkt $b^* = 0$ existiert immer dann, wenn der kleinste m -Wert, der zu der Summe in Gleichung 4.2 beiträgt, größer als Null ist. Das ist der Fall für alle $h \geq 0$. Der Fixpunkt $b^* = 0$ ist stabil, wenn die Steigung der Abbildung an der Stelle $b_t = 0$ kleiner als Eins ist, was der Fall ist für alle $h \geq 1$. Für $0 < h < 1$, ist die Abbildung durch $b_{t+1} = Kb_t/2$ zur führenden Ordnung in b_t gegeben, und der Fixpunkt $b^* = 0$ ist damit instabil für $K > 2$. Für $K = 2$ muss die nächste Ordnung in b_t mitgenommen werden und man erhält $b_{t+1} = b_t(1 - 3b_t/4)$. Damit ist der Fixpunkt $b^* = 0$ stabil. Für $h = 0$ hat man $b_{t+1} = b_t(1 + K/2)$ zur führenden Ordnung und der Fixpunkt $b^* = 0$ ist instabil.

Um Informationen über die anderen Fixpunkte zu erhalten, werden die Rekursionsrelationen für b_t numerisch iteriert und die Abbildungen geplottet. Man findet, dass für $h \geq 1$ und genügend kleine K der einzige stabile Fixpunkt $b_1^* = 0$ ist. Mit wachsendem K erscheint aber ein zweiter stabiler Fixpunkt. Abbildung 4.2 zeigt für $h = 1$ wie sich die Abbildung mit wachsendem K verändert. Ein zweiter stabiler Fixpunkt mit $b_2^* = 0,200$ erscheint bei $K = 12$. Er bewegt sich mit anwachsendem K langsam hin zum Wert 0,5, was der asymptotische Wert für $K \rightarrow \infty$ ist. Für $0 < h < 1$ und $K = 2$ ist der einzige Fixpunkt $b^* = 0$. Wie zuvor bemerkt, ist dieser Fixpunkt für $K > 2$ instabil, und es existiert ein stabiler Fixpunkt mit dem Wert $0 < b^* \leq 0,5$. Er bewegt sich in Richtung 0,5 mit wachsendem K . Für $h = 0$ gibt es nur den stabilen Fixpunkt $b^* = 0,5$ für alle Werte von K . Für $h < 0$ liegt der stabile Fixpunkt zwischen 0,5 und 1 und bewegt sich in Richtung 0,5 mit wachsendem K .

Für $K = 1$ können alle Fixpunkte analytisch bestimmt werden. Für $h > 1$, kann kein Eingang zu einem Knoten über dem Schwellenwert liegen und es gibt keinen Fixpunkt außer $b^* = 0$.

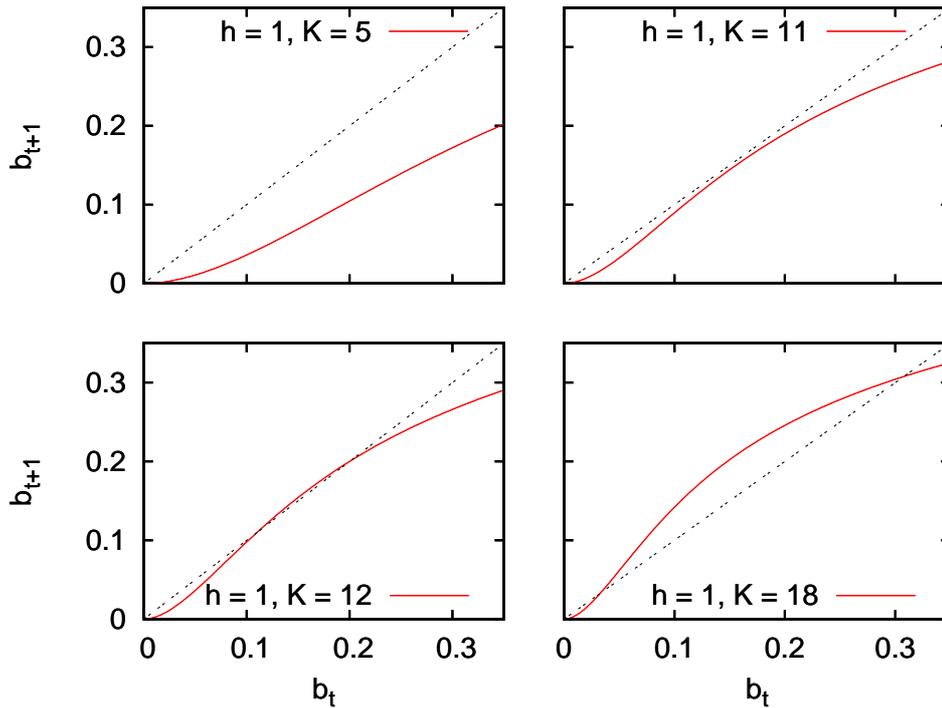


Abbildung 4.2: Abbildung b_{t+1} vs. b_t für $h = 1$ und verschiedene K . Die gestrichelte Linie ist die Winkelhalbierende.

Genauso sind für $h < -1$ die Eingänge aller Knoten über dem Schwellenwert und daher gilt $b^* = 1$. Auswertung der Rekursionsrelation für die verbleibenden Werte von h ergibt $b^* = 1$ für $h = -1$; $b^* = 2/3$ für $-1 < h < 0$; $b^* = 1/2$ für $h = 0$ und $b^* = 0$ für $h > 0$.

4.1.2 Die Sensitivität λ

Hat man einen Fixpunkt b^* für ein Paar von Parameterwerten h und K gefunden, kann man den zugehörigen Wert für λ bestimmen. Für nicht-ganzzahlige h kann sich die Änderung eines Eingangs nur auf den Knotenwert auswirken, wenn alle anderen $K - 1$ Eingänge sich so aufsummieren, dass sie entweder direkt unter oder direkt über dem Schwellenwert liegen (s. Abb. 4.3). Das führt auf

$$\lambda = \frac{K}{2} \cdot \sum_{m=\lfloor h \rfloor}^{K-1} \binom{K-1}{m} \binom{m}{\lfloor \frac{m+1+h}{2} \rfloor} (1 - b^*)^{K-1-m} \left(\frac{b^*}{2}\right)^m. \quad (4.4)$$

Hier ist m wieder die Anzahl von eingeschalteten Eingangsknoten. Die Anzahl von eingeschalteten Eingangsknoten mit positiven Kopplungen $l = \lfloor \frac{m+1+h}{2} \rfloor$ ist so gewählt, dass $2l - m$ nahe beim Schwellenwert liegt, das heißt, direkt darunter oder direkt darüber. Der Faktor $1/2$ ist die Wahrscheinlichkeit, dass die K te Verbindung das richtige Vorzeichen hat.

Für ganzzahlige Schwellenwerte h ist die Situation wieder eine andere. Die Änderung eines Eingangs kann den Knotenwert nur ändern, wenn die Summe aller Eingänge vor der Änderung h war. In der entgegengesetzten Situation, wenn eine Änderung eines Eingangs die Summe aller Eingänge genau auf den Schwellenwert bringt, ändert sich die Ausgabe nicht. (Würde man

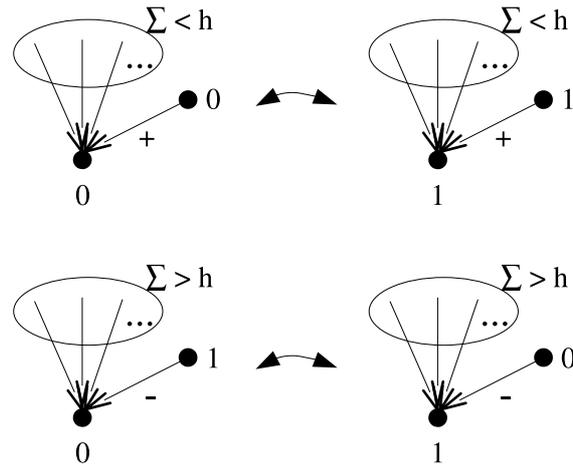


Abbildung 4.3: Für nicht ganzzahlige h kann sich die Änderung eines Eingangs nur auf den Knotenwert auswirken, wenn alle anderen $K - 1$ Eingänge sich so aufsummieren, dass sie entweder direkt unter oder direkt über dem Schwellenwert liegen.

die Annealed Approximation bis ins Extrem treiben und ignorieren, dass es eine Korrelation zwischen dem Zustand eines Knotens und den Zuständen seiner Eingänge gibt, müsste man auch den Fall berücksichtigen, dass die Summe der Eingänge $h \pm 1$ vor der Änderung eines Eingangsknotens ist.)

Für ganzzahlige h ergibt sich demzufolge

$$\lambda = \frac{K}{2} \cdot \sum_{l=h}^{\lfloor \frac{K+h}{2} \rfloor} \binom{K}{2l-h} \binom{2l-h}{l} (1-b^*)^{K-2l+h} \left(\frac{b^*}{2}\right)^{2l-h}. \quad (4.5)$$

Unter Verwendung der letzten beiden Gleichungen, kann λ für jede Kombination von h und K ausgewertet werden. Das resultierende Phasendiagramm ist in Abbildung 4.4 zu sehen. Nur Netzwerke mit $K = 2$ und einem Schwellenwert von $0 < h < 1$ sind kritisch. Netzwerke mit $h > 1$ sind in dem gezeigten K -Bereich gefroren. Wenn es zwei stabile Fixpunkte für $h \geq 1$ gibt, sind die Netzwerke gefroren bei $b^* = 0$ und chaotisch bei $b^* > 0$. Für $h \leq 1$ sind Netzwerke mit ganzzahligem h geordneter als die mit nicht-ganzzahligem h .

4.2 Numerische Simulationen

Die Ergebnisse, die mit Hilfe der Annealed Approximation gewonnen wurden, sollen nun mit denen aus Computersimulationen von Netzwerken mit festen Verknüpfungsmustern verglichen werden. Die Anzahl der Knoten war $N = 10^5$ in allen Simulationen. Die Schwellenwerte wurden aus dem Bereich $-3 < h < +3$ gewählt mit nicht-ganzzahligen Werten von $h = -2,5; -1,5; \dots; 1,5; 2,5$. Alle nicht-ganzzahligen Schwellen, die zwischen den selben ganzen Zahlen liegen, führen selbstverständlich zum gleichen dynamischen Verhalten. Somit ist es ausreichend diese Werte zu betrachten.

Zunächst werden die Werte b_f betrachtet, die nach genügend langer Zeit gefunden werden, wenn mit einem anfänglichen Anteil b_0 von Einsen im System begonnen wird.

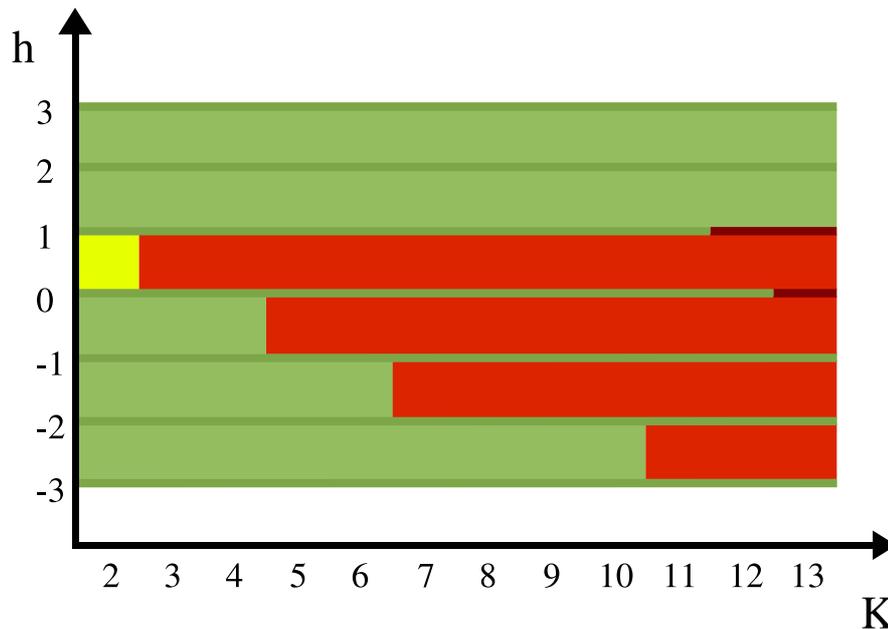


Abbildung 4.4: Mit Hilfe der Annealed Approximation bestimmtes Phasendiagramm. Die chaotische Phase ($\lambda > 1$) ist in rot dargestellt, die gefrorene Phase ($\lambda < 1$) in grün und die kritische Region ($\lambda = 1$) in gelb. Für ganzzahlige h sind die Phasen in dunkleren Schattierungen der Farben gezeigt.

4.2.1 Der Anteil von Nullen und Einsen

$$h > 0$$

Für nicht-ganzzahlige Schwellenwerte h stimmen die Simulationen gut mit den Vorhersagen der Annealed Approximation überein.

$$h = 0,5:$$

Der einzige Fixpunkt für $K = 2$, $b^* = 0$, ist nur marginal stabil, weil die Abbildung an dieser Stelle eine Steigung von Eins hat. Die meisten simulierten Netzwerke erreichen diesen Fixpunkt nicht, sondern laufen in Attraktoren von unterschiedlicher Länge mit einem kleinen b_f von der Größenordnung von 10^{-2} . Das liegt daran, dass die Rekursionsformel $b_{t+1} = b_t(1 - 3b_t/4)$ nur angewendet werden kann, so lange b_t^2 größer als von der Ordnung $1/N$ ist. Der negative quadratische Term beschreibt den inhibitorischen Effekt eines zweiten eingeschalteten Eingangs mit einer negativen Verbindung. Die Abnahme von b_t kommt zum Stillstand, wenn b_t so klein ist, dass es keine Knoten mit zwei eingeschalteten Eingängen mehr gibt, was für immer kleinere Werte von b_t der Fall ist, wenn man die Systemgröße erhöht. Die Abweichung von Simulationen und Annealed Approximation ist somit klar ein Effekt der endlichen Systemgröße.

Für $K = 3$ und 4 sind die mittleren b -Werte aus den Simulationen in guter Übereinstimmung mit den berechneten Werten (dies wurde auf drei Nachkommastellen genau überprüft).

$$h = 1,5:$$

Für nicht-ganzzahlige h erscheint ein zweiter Fixpunkt mit dem Wert b^* bei K -Werten, die sich mit den Vorhersagen der Annealed Approximation decken. Für $h = 1,5$ passiert das bei

$K = 15$. Der mittlere b -Wert aus den Simulationen zeigt allerdings eine kleine Abweichung von ca. 1% vom berechneten Wert. Dies kann wieder der endlichen Systemgröße zugeschrieben werden, da die Steigung der Abbildung in der Nähe des Fixpunktes für den K -Wert, bei dem dieser Fixpunkt zuerst auftritt, nahe bei Eins liegt. Für $K = 16$ und 17 stimmen die Werte wieder gut überein.

$h = 2,5$:

Für $h = 2,5$ erscheint ein stabiler Fixpunkt mit dem Wert $b^* > 0$ bei $K = 40$, genau wie von den Berechnungen vorhergesagt. Der über den Attraktor gemittelte Wert von b^* aus den Simulationen stimmt mit dem aus der Annealed Approximation erhaltenen in drei Dezimalstellen überein. Eine solch gute Übereinstimmung wird auch für $K = 41$ und 42 gefunden.

Der Fall von ganzzahligen Schwellenwerten h ist besonders und es zeigen sich Abweichungen zwischen Simulationen und Annealed Approximation.

$h = 1$:

Für Netzwerke mit einem K -Wert zwischen zwei und elf, sagt die Annealed Approximation einen einzelnen stabilen Fixpunkt $b^* = 0$ vorher. Im Gegensatz dazu zeigen die Simulationen, dass bereits ab einem Wert von $K = 5$ stationäre Zustände mit einer größeren Anzahl von Knoten mit dem Wert 1 existieren. Der Wert b_f , dem die Netzwerke sich für lange Zeiten annähern, hängt vom Anfangswert b_0 ab (s. Abb. 4.5). Jede Kurve in Abbildung 4.5 entspricht einer Netzwerkrealisation. Jeder Punkt in der Abbildung ist ein Fixpunkt der Dynamik, also ein Attraktor der Länge Eins. Wie zu sehen ist, zeigen zwei verschiedene Netzwerke mit den selben Werten für K und h annähernd das gleiche Verhalten. Das bedeutet, dass die Funktion $b_f(b_0)$ nicht von der genauen Realisation der Netzwerke abhängt. Für Netzwerke mit $K = 5$, erscheint der von Null verschiedene Fixpunkt, wenn der anfängliche Anteil an Einsen in der Nähe von 50% liegt.

Erst für $K \geq 12$ sagt die Annealed Approximation einen stabilen Fixpunkt $b^* > 0$ voraus. Für diese K -Werte ergeben die Simulationen einen b_f -Wert, der unabhängig von b_0 ist, wenn b_0 nicht zu klein ist. Vergleicht man die Werte b^* , die man durch Iteration von Gleichung 4.3 bekommt mit den Werten, die man erhält, wenn man über mehrere Netzwerksimulationen mittelt, findet man, dass die simulierten Netzwerke ca. 41% mehr eingeschaltete Knoten haben, als es von der Annealed Approximation vorhergesagt wird. Mit wachsendem K nimmt diese Abweichung von der Annealed Approximation ab; sie liegt bei ca. 25% für $K = 13$ und bei ca. 20% für $K = 14$.

Zum Verständnis der Entstehung einer solchen breiten Menge von dynamischen Fixpunkten in diesen Systemen, soll Folgendes angemerkt werden:

(i) An jedem Fixpunkt gibt es Knoten, deren Eingangssumme genau den Schwellenwert ergibt. Wenn die Änderung des Zustands eines solchen Knoten keinen Zustand eines anderen Knoten ändert, hat man einen weiteren Fixpunkt mit einer anderen Anzahl von eingeschalteten Knoten gefunden.

(ii) Alle Fixpunkte eines Netzwerkes mit $h = 1,5$ oder $h = 0,5$ sind auch Fixpunkte dieses Netzwerkes wenn $h = 1$ ist. Es kann erwartet werden, dass alle beobachteten Fixpunkte b -Werte zwischen diesen beiden Grenzwerten haben und das wird in der Tat gefunden.

(iii) Wenn es eine Menge von Fixpunkten mit verschiedenen b -Werten gibt, sollte der Endwert von b , der in einer Simulation erreicht wird, vom Anfangswert abhängen. Das wiederum

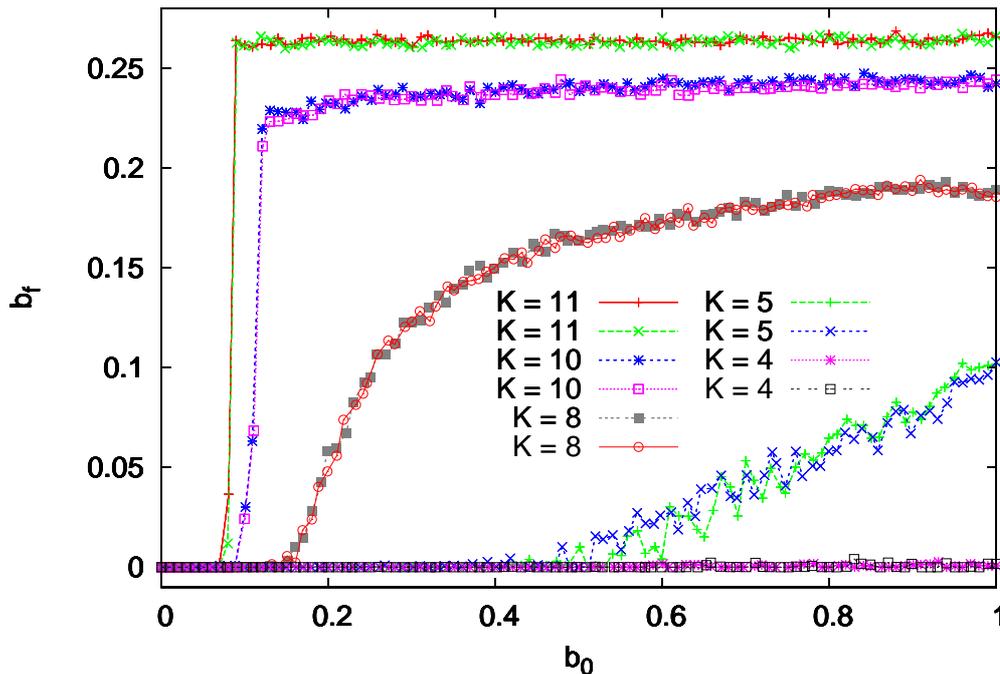
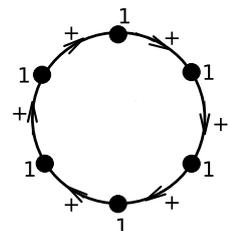


Abbildung 4.5: Der Anteil an Einsen am Ende der Simulation in Abhängigkeit vom anfänglichen Anteil an Einsen in Netzwerken mit $h = 1$ und verschiedenen Werten von K . Jeder Datensatz entspricht einer Netzwerkrealisation.

bedeutet, dass der Zustand eines Knotens auf einem Fixpunkt von der dynamischen Geschichte dieses Knotens und seiner Eingänge abhängt. Solche Korrelationen zwischen Zuständen eines Knotens zu verschiedenen Zeitpunkten werden von der Annealed Approximation nicht berücksichtigt. Diese stellt somit keine gute Näherung in dieser Situation dar. Mit anwachsendem K werden die Kurven in Abbildung 4.5 zusehends unabhängig von b_0 . Dies kann anwachsenden Transienzzzeiten zugeschrieben werden, die das „Gedächtnis“ für den Anfangszustand der Knoten abschwächen.

(iv) Es kann ein Satz von eingeschalteten Knoten existieren, die untereinander mit positiven gerichteten Verknüpfungen verbunden sind, so dass eine Schleife entsteht. Wenn die Summen aller anderen Eingänge zu diesen Knoten Null sind, befinden sie sich auf der Schwelle und bleiben auf dem Wert 1, wenn sie einmal eingeschaltet wurden. Korrelationen dieser Art zwischen Knoten sind nicht in der Annealed Approximation enthalten.



$h = 2$ und 3 :

Für Netzwerke mit einem Schwellenwert von $h = 2$ erscheint in der Annealed Approximation ein zweiter stabiler Fixpunkt bei $K = 35$, wird in simulierten Netzwerken aber bereits ab einem Verknüpfungsgrad von $K = 31$ gefunden. Bei $K = 35$ hat der zweite stabile Fixpunkt b^* einen ca. 36% höheren Wert, als von der Annealed Approximation vorhergesagt. Dieser Unterschied reduziert sich zu 26% bei $K = 36$ und zu 21% bei $K = 37$. Auch für $h = 3$ erscheint ein zweiter stabiler Fixpunkt bei einem kleineren K (bei $K = 65$) als es nach der Annealed Approximation zu erwarten wäre (bei $K = 70$). Der Wert von b^* , der bei $K = 70$ erreicht wird, liegt ca. 29% über dem erwarteten Wert. Man findet eine Abweichung von ca. 23% bei $K = 71$ und von ca.

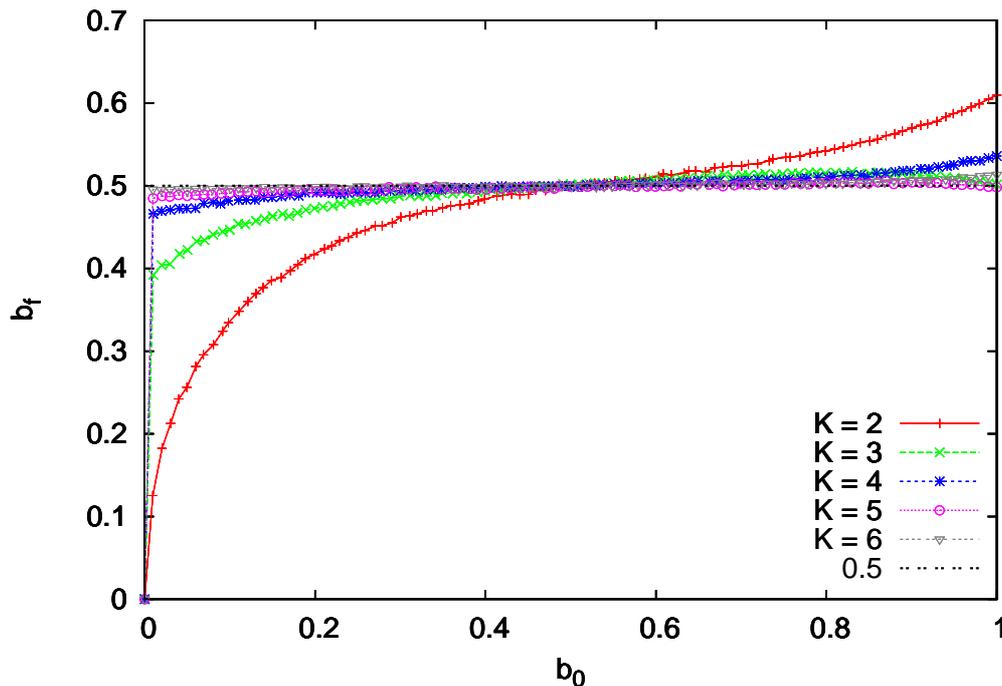


Abbildung 4.6: Der Anteil an Einsen am Ende der Simulation in Abhängigkeit vom anfänglichen Anteil an Einsen in Netzwerken mit $h = 0$ und verschiedenen K -Werten. Jeder Datensatz entspricht einer Netzwerkrealisation.

20% bei $K = 72$. Für $h = 2$ liegen alle beobachteten b -Werte zwischen denen, die man für $h = 1,5$ und $h = 2,5$ bekommt, für $h = 3$ zwischen denen, die man für $h = 2,5$ und $h = 3,5$ bekommt.

$h = 0$

Für $h = 0$ und Konnektivitäten von $K = 2$ bis 5 finden sich wieder dynamische Fixpunkte mit b_f -Werten, die sich von dem mit Hilfe der Annealed Approximation berechneten Fixpunkt-Wert unterscheiden (s. Abb. 4.6). Mit zunehmendem K jedoch nähern sich die b_f -Werte dem erwarteten Wert von 0,5.

In Abbildung 4.6 kann man noch einen anderen interessanten Effekt beobachten, der auch für andere ganzzahlige h und nicht zu große K auftritt. Abhängig von der Summe von h und K weisen die Kurven verschiedenes Verhalten bei hohen b_0 -Werten auf. Ist $(h + K)$ gerade, steigt b_f am Ende an, da es in diesem Fall wahrscheinlicher ist, dass die Eingänge eines Knotens genau den Schwellenwert ergeben als für ungerade $(h + K)$.

Den trivialen Fall $K = 1$ zu betrachten ist lehrreich: Jeder Knoten hat genau einen Eingang. Beginnend von einem zufällig ausgewählten Knoten, kann man der Kette von Eingängen folgen, die diesem Knoten vorangehen. Für große Systeme ist die mittlere Länge solcher Ketten lang (von der Größenordnung \sqrt{N} [31]). Jede Kette endet schließlich in einer Schleife. Entlang der Kette schließen sich positive und negative Verbindungen in einer zufälligen Reihenfolge an. Die Fixpunkte eines solchen Systems von Ketten kann man leicht bestimmen: Offensichtlich hat man einen Fixpunkt, wenn alle Knoten im Zustand Null sind. Wenn man dann einen Knoten anschaltet, der nur negative Ausgangsverbindungen hat, erhält man den nächsten Fixpunkt.

Schaltet man einen Knoten an, der eine positive Ausgangsverbindung hat, wird auch der Knoten angeschaltet, der durch diese Verbindung beeinflusst wird, und so weiter bis das Ende der Reihe von positiven Verbindungen erreicht ist. Den Fixpunkt mit dem Maximum an eingeschalteten Knoten erhält man, wenn man allen Knoten mit einer positiven Eingangsverbindung und allen Knoten mit einer negativen Eingangsverbindung, denen eine ungerade Anzahl von Knoten mit negativen Verbindungen vorangeht, den Wert Eins zuweist. Auf diesem Wege sind alle Knoten, die eine negative Eingangsverbindung und einen Eingangsknoten mit dem Wert Eins haben, im Zustand Null. Der zu diesem Fixpunkt gehörige b -Wert ist

$$b \simeq \frac{1}{2} + \frac{1}{8} \sum_{n=0}^{\infty} \left(\frac{1}{4}\right)^n = \frac{2}{3}.$$

Für $-1 < h < 0$ wäre das der einzige Fixpunkt. Das Beispiel $K = 1$ zeigt also, dass der maximale Fixpunktwert von b für ganzzahlige h identisch ist mit dem, den man erhält, wenn man den Schwellenwert h geringfügig senkt. In gleicher Weise erhält man den minimalen Wert 0 als Fixpunktwert, wenn h geringfügig größer als 0 ist. Alle dazwischenliegenden Werte von b für $h = 0$ erhält man, wenn man Teile der eingeschalteten Knoten in dem Zustand mit maximalem b -Wert ausschaltet.

Dieses Beispiel verdeutlicht auch, dass nicht alle Fixpunktwerte von b von einem zufälligen Zustand aus erreicht werden können. Zum Beispiel kann man zum Maximalwert $2/3$ von b nicht von einer zufälligen Anfangskonfiguration aus hingelangen. Wenn anfänglich alle Knoten im Zustand Eins sind, das heißt, $b_0 = 1$, hat man nach einem Zeitschritt einen Zustand mit $b = 1/2$, bei dem alle Knoten mit einer positiven Eingangverbindung im Zustand Eins sind. Gibt es anfänglich weniger Einsen, kann es im Endzustand nicht mehr Einsen geben. Werte zwischen $1/2$ und $2/3$ können daher nur erreicht werden, wenn man von speziell präparierten Anfangszuständen ausgeht. Diese Beobachtung für $K = 1$ erklärt, warum die Simulationen für größere Konnektivitäten K (und auch für andere ganzzahlige Schwellenwerte h) b -Werte erreichen, die zwischen denen mit der Annealed Approximation bestimmten Werten für benachbarte nicht-ganzzahlige h -Werte liegen, aber nie das gesamte b -Intervall zwischen diesen Grenzen von einem zufälligen Anfangszustand mit festem b aus erreicht wird.

Die Tatsache, dass die Ergebnisse der Annealed Approximation und die der Simulationen für größere K besser übereinstimmen, kann dem Kleinerwerden des Intervalls zwischen diesen Grenzen mit wachsendem K zugeschrieben werden.

$h < 0$

Abbildung 4.7 zeigt die finalen b -Werte als Funktion der anfänglichen b -Werte für Netzwerke mit $K = 4$ und verschiedene negative h . (Für die gezeigten Parameterwerte haben alle Netzwerke $\lambda < 0$, das heißt, sie sind gefroren.) Wie man sehen kann, hängt für nicht-ganzzahlige Werte von h der finale Anteil von Einsen im Netzwerk nicht von b_0 ab. Auch hier stimmen, wie für $h > 0$ die Werte von b_f aus den Simulationen gut mit den Werten von b^* überein, die mit Hilfe der Annealed Approximation erhalten wurden. Für ganzzahlige h findet man erneut Fixpunkte der Dynamik mit b_f -Werten, die von b_0 abhängen. Die Mittelwerte für b_f sind stets kleiner, als die Werte, die von der Annealed Approximation vorhergesagt werden. Die Werte von b_f für Netzwerke mit ganzzahligen h liegen immer zwischen denen von Netzwer-

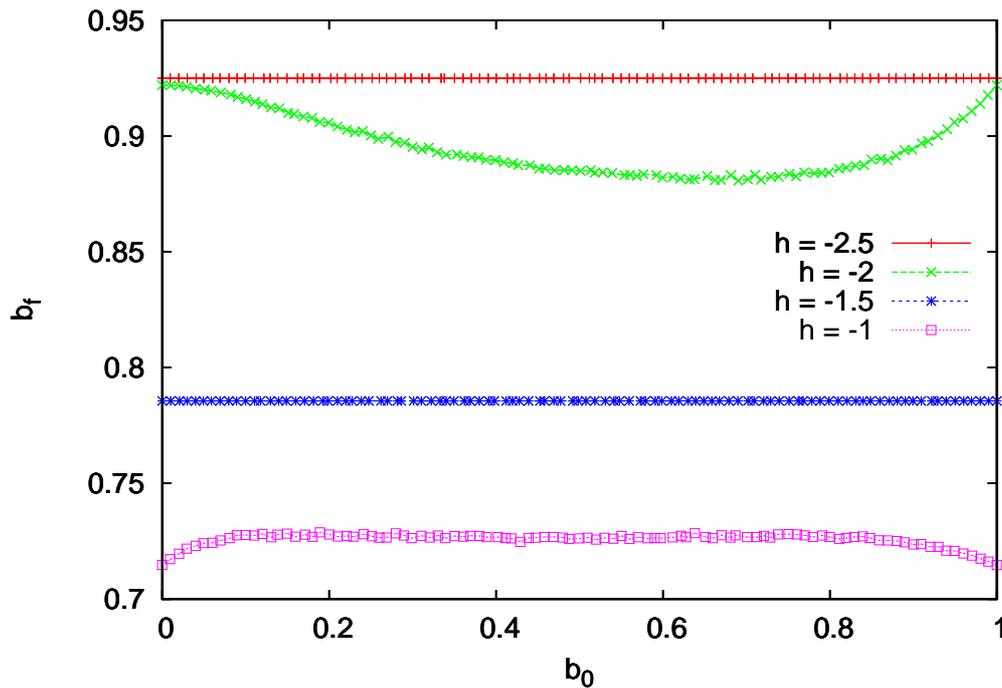


Abbildung 4.7: Der Anteil an Einsen am Ende der Simulation in Abhängigkeit vom anfänglichen Anteil von Einsen in Netzwerken mit einer Konnektivität von $K = 4$ und verschiedenen negativen Schwellenwerten h . Jeder Datensatz entspricht einer Netzwerkrealisation.

ken mit benachbarten nicht-ganzzahligen h -Werten mit dem selben K , wie es zuvor auch für nicht-negative Werte von h beobachtet wurde.

4.2.2 Das Phasendiagramm

Nun sollen die dynamischen Eigenschaften der simulierten Netzwerke betrachtet werden, das sind zum einen die Längen ihrer Attraktoren und zum anderen die Anzahl von Knoten, die ihren Zustand ändern, während das Netzwerk den Attraktor durchläuft. Zu diesem Zeitpunkt ändern sich die b -Werte der betrachteten Netzwerke nicht mehr (abgesehen von kleinen Fluktuationen, aufgrund der endlichen Systemgröße). Das bedeutet aber nicht zwangsläufig, dass auch ein Fixpunkt im Zustandsraum erreicht wurde (das heißt, ein Fixpunkt der Dynamik). Ist der Attraktor länger als Eins bedeutet das, dass der Anteil an Knoten, die ihren Zustand von Eins zu Null ändern zu jedem Zeitschritt annähernd gleich dem Anteil an Knoten ist, die ihren Zustand von Null zu Eins ändern. Mit dem Wissen über typische Attraktorlängen und Anzahlen von blinkenden Knoten in den verschiedenen Phasen (s. Kap. 2.3.2) können die Simulationen mit den λ -Werten verglichen werden, die im Rahmen der Annealed Approximation berechnet wurden.

$$h > 0$$

Die einzige Parameterkombination, für die die betrachteten Netzwerke kritisch sein können, ist $h = 0,5$ und $K = 2$. Die in den Simulationen gefundenen Attraktoren haben Längen von einer

bis zu vier Stellen. Die Anzahl der Knoten, die ihren Zustand auf dem Attraktor ändern ist von der Größenordnung 10^3 , was vereinbar ist mit der erwarteten Anzahl von der Größenordnung $N^{2/3}$ [48].

Für $K > 2$ and $h = 0,5$ sollten die Netzwerke der Annealed Approximation zufolge chaotisch sein. In den Simulationen sind die Attraktoren länger als der betrachtete Suchbereich (maximale Länge der Transienten: 10^5 , maximale Attraktorlänge: 10^4), was konsistent ist mit der Erwartung eines chaotischen Netzwerkes. Des Weiteren sagt die Annealed Approximation voraus, dass Netzwerke mit größeren h -Werten gefroren sein sollten, wenn der einzig stabile Fixpunkt $b^* = 0$ ist. Netzwerke mit K -Werten, für die ein zweiter stabiler Fixpunkt $b^* > 0$ existiert, haben $\lambda > 1$ an diesem Fixpunkt und sollten also chaotisch sein. Simulationen von Netzwerken mit nicht-ganzzahligem Schwellenwert $h = 1,5$ und $2,5$ zeigen Ergebnisse, die mit dieser Vorhersage konsistent sind. Netzwerke mit ganzzahligem h zeigen wieder Abweichungen. Für $h = 1$ und $K = 12$ sind auch die Netzwerke am zweiten Fixpunkt $b^* > 0$ gefroren und nicht chaotisch. Für $h = 1$ und $K = 13$ hingegen werden innerhalb des Suchbereichs keine Attraktoren gefunden, was auf chaotische Dynamik hinweist. Für $h = 2$ ist die Situation eine andere. Wie im vorangegangenen Abschnitt dargelegt, erscheint ein zweiter stabiler Fixpunkt $b^* > 0$ schon bei $K = 31$ und die Netzwerke zeigen chaotisches Verhalten an diesem Fixpunkt. Ausgehend von der Annealed Approximation sollte das erste chaotische Netzwerk eine Konnektivität von $K = 35$ haben. Ähnliches gilt auch für Netzwerke mit $h = 3$: sie sind chaotisch, wenn sie den zweiten stabilen Fixpunkt b^* erreichen, dieser Fixpunkt erscheint jedoch bei K -Werten, die kleiner sind als durch die Annealed Approximation vorhergesagt.

$h = 0$

Für $h = 0$ ist die Situation ähnlich der für $h = 1$. Ausgehend von der Annealed Approximation sollten Netzwerke ab einer Konnektivität von $K = 13$ chaotisch sein, aber in allen Simulationen von Netzwerken bis zu Konnektivitäten von $K = 16$ findet man nur Fixpunktattraktoren. Das bedeutet, dass die Netzwerke in der gefrorenen Phase sind.

$h < 0$

Für $h < 0$ werden zunächst solche Werte für h und K betrachtet, für die erwartet wird, die Netzwerke in der gefrorenen Phase zu finden. In der Tat werden in allen Simulationen nur Fixpunktattraktoren gefunden. Als nächstes sollen Parameterkombinationen betrachtet werden, die nahe an den Übergängen zwischen gefrorener und chaotischer Dynamik liegen (vgl. Abb. 4.4). Für ganzzahlige h findet man wie zu erwarten Abweichungen von der Annealed Approximation. So wie für $h = 0$ und 1 findet man die gefrorene Phase ausgedehnt hin zu höheren K -Werten. Das bedeutet, dass Netzwerke mit nur Fixpunktattraktoren in Regionen des Parameterraumes gefunden werden, in denen sie laut Annealed Approximation chaotisch sein sollten. Für nicht-ganzzahlige $h = -0,5$ und $-2,5$ sollten chaotische Netzwerke für $K \geq 5$ beziehungsweise $K \geq 11$ gefunden werden. In allen Simulationen mit diesen Parametern gehen die Attraktorlängen über den Suchbereich hinaus. Für $h = -0,5$ und $K = 4$ auf der anderen Seite der Phasengrenze, wo die Netzwerke gefroren sein sollten, findet man hauptsächlich kurze Attraktoren auf denen weniger als 1% der Knoten ihren Wert ändern. In Übereinstimmung mit dem berechneten Wert $\lambda \approx 0,995$, ist der Hauptteil dieser Netzwerke gefroren. Für $h = -2,5$ und $K = 10$ stellt sich die Situation ähnlich dar. Für $h = -1,5$ sagt die Annealed Approximation

chaotische Dynamik für $K \geq 7$ vorher. Aber da $\lambda \approx 1,008$ bei $K = 7$, ist man bei diesen Parameterwerten sehr Nahe an der Phasengrenze. Die simulierten Netzwerke haben Attraktorlängen, die von einstelligen Werten bis hin zu solchen reichen, die länger als der Suchbereich sind. Für $K = 6$ hat man $\lambda \approx 0,903$ und die Attraktoren, die gefunden werden sind kurz mit weniger als 1% an Knoten, die ihre Werte ändern.

Zusammenfassend ist das Phasendiagramm in Abbildung 4.4, das mit Hilfe der Annealed Approximation bestimmt wurde, gültig für Boolesche Schwellenwertnetzwerke mit nicht-ganzzahligen Schwellenwerten h . Für ganzzahlige $h \leq 1$ findet der Übergang von der gefrorenen in die chaotische Phase bei größeren Werten von K statt als von der Annealed Approximation vorhergesagt. Für ganzzahlige $h > 1$ geschieht das bei einem kleineren Wert als vorhergesagt. Da die meisten dieser Übergänge nicht in dem in Abbildung 4.4 gezeigten Fenster von K -Werten liegen, sieht die zugehörige Abbildung, die mit Hilfe der Simulationen gewonnen wurde, kaum anders aus und wird daher hier nicht noch einmal gezeigt.

Der Grund, warum die Übergänge von der gefrorenen zur chaotischen Phase nicht bei den K -Werten auftreten, die von der Annealed Approximation vorhergesagt werden, wenn h ganzzahlig ist, ist der gleiche, weshalb die stationären Werte b_f nicht mit den berechneten Fixpunkten b^* übereinstimmen. Für ganzzahlige h kann die Annealed Approximation die dynamischen Eigenschaften des Systems nicht korrekt erfassen, weil die Näherung Gedächtniseffekte vernachlässigt. Sogar wenn bei der Auswertung der λ -Werte im Rahmen der Annealed Approximation die Werte b_f aus den Simulationen verwendet werden, tritt der berechnete Phasenübergang nicht bei den gleichen K -Werten auf, wie sie aus den Computersimulationen gewonnen werden. In den zu Anfang des Kapitels erwähnten Studien zu Phasenübergängen in Schwellenwertnetzwerken [53, 74, 73] trat diese Situation nicht auf. Die Annealed Approximation war in diesen Fällen also ausreichend um die Phasenübergänge korrekt zu berechnen.

4.3 Die Bäckerhefe

Nun soll das Zellzyklusregulationsnetzwerk der Bäckerhefe (s. Kap. 2.4) im Licht der gewonnenen Erkenntnisse noch einmal betrachtet werden. Obwohl in diesem Netzwerk die Trajektorie, die dem Zellzyklus entspricht, erstaunlich stabil ist, ist der dominante Fixpunkt sehr empfindlich gegenüber Störungen an bestimmten Knoten [33]. Wenn auf diesem Fixpunkt der Zustand von jeweils einem der elf Knoten geändert wird, kehrt das Netzwerk nur in sechs von elf Fällen wieder zu dem Fixpunkt zurück. Sechs der insgesamt sieben Fixpunkte können von anderen Fixpunkten aus erreicht werden, indem nur der Zustand eines Knotens geändert wird, und es gibt eine Gruppe von nur drei Knoten unter denen alle diese Änderungen stattfinden. Die Eingangssummen von allen drei Knoten befinden sich an der Schwelle und das Ändern des Zustandes eines dieser Knoten ändert nichts am Zustand irgendeines anderen Knoten. Wie in diesem Kapitel gesehen, ist das Auftreten von solchen Sätzen von Fixpunkten, die sich nur durch den Zustand eines Knotens unterscheiden, charakteristisch für Schwellenwertnetzwerke mit ganzzahligem Schwellenwert. Das ist bedingt durch die Aktualisierungsregel, die bewirkt, dass ein Knoten seinen Zustand beibehält, wenn die Summe seiner Eingänge genau den Schwellenwert ergibt. Dies wirft die Frage auf, ob die nicht-dominanten Fixpunkte des Zellzyklusregulationsnetzwerkes der Bäckerhefe eine biologische Bedeutung haben oder nur Artefakte der Aktualisierungsregel des Modells sind.



5 Simulierte Evolution mittels der „Adaptiven Wanderung“

Die Adaptive Wanderung (adaptive walk) ist eine Methode Evolution zu simulieren, bei der die Fitnesslandschaft erkundet wird. Der ihr zugrundeliegende Algorithmus führt in einer Fitnesslandschaft immer bergauf, d.h. es werden nur Mutationen akzeptiert, die die Fitness erhöhen oder beibehalten. Fitnesssenkende Mutationen treten nie auf.

Das Konzept der Adaptive Wanderung wurde von Kauffman eingeführt [46]. Die Grundannahme besteht dabei darin, dass alle Mitglieder einer Population die meiste Zeit über den selben Genotyp haben. Tritt eine fitnesssteigernde Mutation auf, wird sie schnell in die gesamte Population übernommen, fitnesssenkende Mutationen hingegen werden schnell entfernt. Dies führt zu dem Bild einer Population, die in der Fitnesslandschaft immer bergauf geht, bis sie ein lokales Maximum erreicht. Dort verharrt sie für immer [25].

Der in dieser Arbeit verwendete Algorithmus zur Simulation von Evolution ist in Abbildung 5.1 schematisch dargestellt. Die fettgedruckten Punkte werden im Folgenden näher erläutert. Im Anschluss daran werden die biologischen Grundlagen der verwendeten Fitnesskriterien und Mutationen diskutiert. Dieses Kapitel bildet die Grundlage für die folgenden zwei Kapitel.

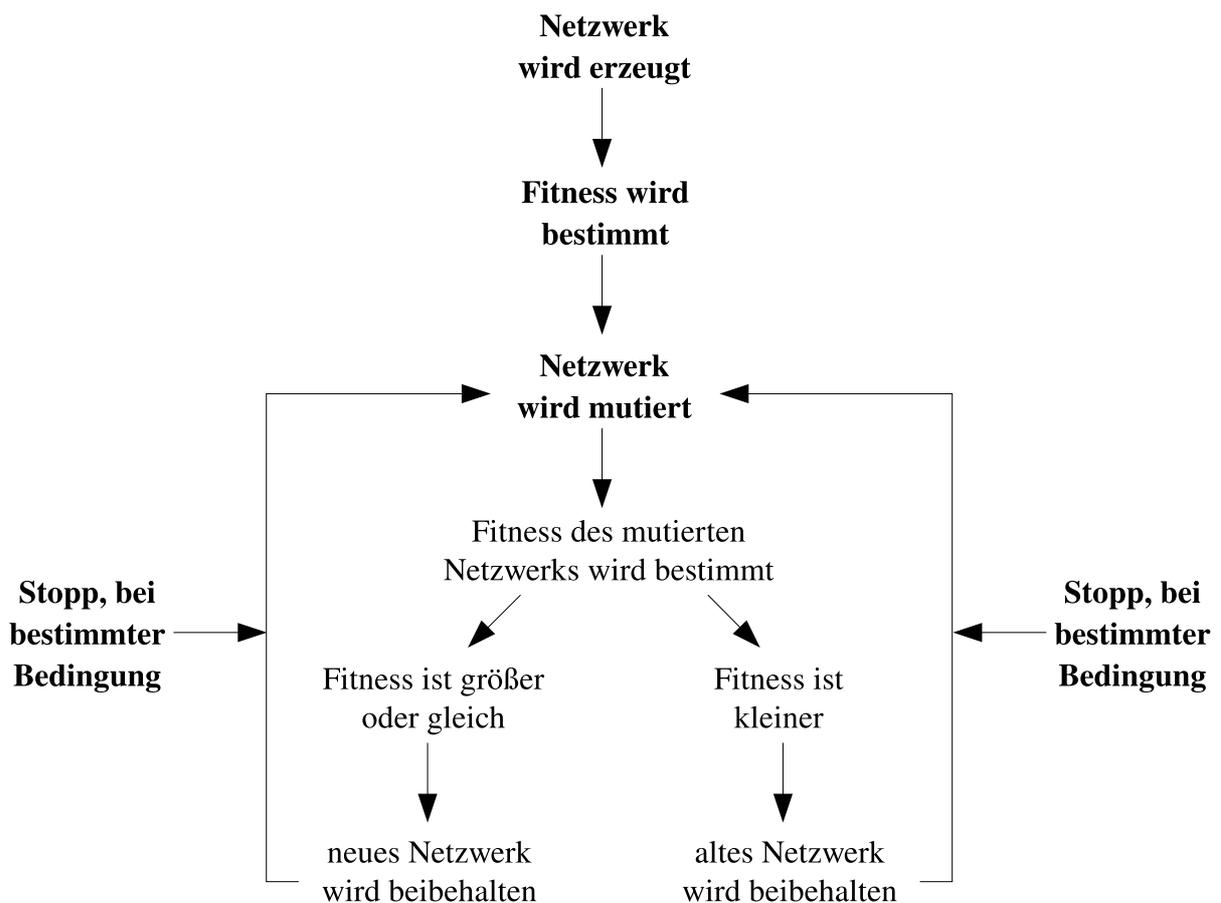


Abbildung 5.1: Schematische Darstellung des Algorithmus für die Adaptive Wanderung.

5.1 Erzeugung eines Netzwerkes

Am Anfang jeder Simulation werden N Knoten erzeugt. Jeder von ihnen bekommt einen zufälligen Anfangswert 0 oder 1, eine zufällige Aktualisierungsfunktion und K zufällig ausgewählte Eingangsknoten zugewiesen. (Zu Beginn haben alle Knoten die gleiche Anzahl von Eingängen, dies ändert sich jedoch im Laufe der Evolution.) Für kanalisierende Netzwerke gilt der erste Eingang als der kanalisierende. Bei Netzwerken mit allen Booleschen Funktionen wird eine zufällige Boolesche Funktion generiert, indem für jede Eingangskombination mit gleicher Wahrscheinlichkeit eine Eins oder eine Null als Ausgabewert gewählt wird. Bei kanalisierenden Netzwerken wird jedem Knoten zufällig eine Funktion F aus den Funktionsklassen 2.2-2.5 zugewiesen. Die Funktion G der restlichen $K - 1$ nicht (zwangsläufig) kanalisierenden Eingänge wird so bestimmt, wie die Funktionen im Fall der Netzwerke mit allen Booleschen Funktionen. Natürlich können durch die Wahl der Funktion G weitere kanalisierende Eingänge entstehen. Bei Schwellenwertnetzwerken wird jedem Eingang eines Knotens mit gleicher Wahrscheinlichkeit eine aktivierende (+1) oder inhibierende (-1) Wirkung zugewiesen.

5.2 Fitnessbestimmung

Im Rahmen dieser Arbeit werden drei verschiedene Fitnesskriterien verwendet. Das simpelste von ihnen wurde bereits in [84, 85] benutzt und ist ein Bestandteil der anderen beiden:

1. Das Netzwerk wird zunächst gemäß Gleichung 2.1 aktualisiert, bis es einen Attraktor erreicht. Dann wird nacheinander der Wert jedes Knotens geändert und wiederum gewartet, bis das Netzwerk bei einem Attraktor angelangt ist. Die Robustheit des Attraktors, und damit die Fitness, ist der Anteil an Umschaltvorgängen, die auf den ursprünglichen Attraktor zurückführen.

Ein verschärftes Fitnesskriterium erweitert das erste und bezieht die transienten Zustände bis zum Erreichen des Attraktors mit ein:

2. Ist der ursprüngliche Attraktor nach durchschnittlich zwei Zeitschritten nach dem Umschalten eines Knotens (gemittelt über alle Umschaltvorgänge) nicht wieder erreicht, wird der Fitnesswert auf Null gesetzt.

Eine tiefgreifendere Erweiterung bezieht die Reaktion auf einen externen Stimulus mit ein. Hierfür wird einer der Knoten als externer Eingang ausgewiesen.

3. Zunächst wird die Fitness des ersten gefundenen Attraktors wie unter Punkt 1 bestimmt (außer dass der als externer Eingang ausgewiesene Knoten hier nicht umgeschaltet wird). Dann wird der externe Eingangsknoten umgeschaltet und gewartet, bis das Netzwerk erneut auf einen Attraktor läuft. Dieser muss verschieden vom ersten Attraktor sein. Ist es nicht, wird der Fitnesswert auf Null gesetzt. Erreicht das Netzwerk nach Umschalten des externen Eingangsknotens einen zweiten Attraktor, wird auch dessen Robustheit wie unter Punkt 1 (unter Ausnahme des externen Eingangsknotens) bestimmt. Nun wird der externe Eingangsknoten wieder auf seinen Ursprungswert zurückgesetzt und der Attraktor bestimmt, auf den das Netzwerk nun läuft. An dieser Stelle werden wiederum zwei Kriterien unterschieden:

- Das „Return“-Kriterium: Das Netzwerk muss zum ersten Attraktor zurückkehren (*engl.: return*), wenn der externe Eingangsknoten zurückgeschaltet wird. Passiert das nicht, wird der Fitnesswert um Eins erniedrigt (Version 1) oder auf Null gesetzt (Version 2).
- Das „Stay“-Kriterium: Das Netzwerk muss auf dem zweiten Attraktor bleiben (*engl.: stay*), wenn der externe Eingangsknoten zurückgeschaltet wird. Ist das Kriterium nicht erfüllt, wird der Fitnesswert um Eins erniedrigt (Version 1) oder auf Null gesetzt (Version 2).

5.3 Mutationen

Es sind vier Arten von Mutationen möglich, die mit gleicher Wahrscheinlichkeit ausgewählt werden und an einem ebenfalls zufällig gewählten Knoten stattfinden:

- *Hinzufügen einer Verbindung*: Der Knoten bekommt eine Verbindung von einem gleichermaßen zufällig ausgewählten Eingangsknoten. Bei Schwellenwertnetzwerken wird der Verbindung mit einer Wahrscheinlichkeit von jeweils einhalb eine aktivierende oder inhibierende Wirkung zugeordnet. Diese Mutation kann nicht stattfinden, wenn der zu mutierende Knoten bereits die maximal mögliche Anzahl an Eingängen erreicht hat. Die maximale Anzahl von Eingängen liegt für kanalisierende Netzwerke und Netzwerke mit allen Booleschen Funktionen bei $K_{\max} = 10$, da die Simulationen sehr langsam werden für große K . Bei Schwellenwertnetzwerken ist die maximale Anzahl von Eingängen nur durch die Knotenzahl beschränkt $K_{\max} = N$.
- *Entfernen einer Verbindung*: Eine zufällige Eingangsverbindung des Knotens wird gelöscht. Diese Mutation kann selbstverständlich nicht stattfinden, wenn der ausgewählte Knoten keine Eingangsverbindungen mehr hat.
- *Umverknüpfen einer Verbindung*: Für eine zufällig ausgewählte Verbindung wird ein neuer Ausgangsknoten mit gleicher Wahrscheinlichkeit unter allen Knoten bestimmt. Diese Mutation ist nicht durchführbar, wenn der zu mutierende Knoten entweder keine Verbindungen mehr besitzt oder bereits Verbindungen mit allen anderen Knoten des Netzwerkes hat.
- *Ändern der Aktualisierungsfunktion*: Für den ausgewählten Knoten wird eine neue Aktualisierungsfunktion bestimmt. Bei Netzwerken mit allen Booleschen Funktionen und bei Schwellenwertnetzwerken geschieht das wie bei der Initialisierung des Netzwerks am Anfang der Adaptiven Wanderung. Für kanalisierende Netzwerke wird mit gleicher Wahrscheinlichkeit eine der vier kanalisierenden Funktionen ausgewählt. Auch diese Mutation wird nicht durchgeführt, wenn der ausgewählte Knoten keine Eingänge hat.

Es ist zu beachten, dass beim Hinzufügen oder Entfernen von Knoten die Aktualisierungsfunktion von Netzwerken mit allen Booleschen Funktionen sowie die Funktion G im Fall von kanalisierenden Funktionen neu ausgewürfelt werden muss. Verliert bei kanalisierenden Netzwerken ein Knoten die Verbindung zu seinem kanalisierenden Eingang, wird der nachfolgende Eingang zum kanalisierenden für ihn.

Kann eine Mutation an dem ausgewählten Knoten nicht durchgeführt werden, so wird ein neuer Knoten ausgewählt. Dies geschieht höchstens N^2 mal, danach gilt die Mutation als endgültig gescheitert und eine neue Mutation, die wieder mit einer Wahrscheinlichkeit von $1/4$ aus den oben genannten Mutationen gewählt wird, wird versucht.

5.4 Abbruchbedingungen der Simulation

Der Algorithmus enthält Bedingungen, die die simulierte Evolution abbrechen sollen, wenn sie nicht mehr oder zu langsam vorankommt. Diese sind:

- Eine bestimmte Anzahl von Mutationen wird durchgeführt und keine von ihnen wird akzeptiert. Das würde z.B. passieren, wenn das Netzwerk auf einem lokalen Maximum in der Fitnesslandschaft gefangen wäre und seine Fitness reduzieren müsste, um die Wanderung fortsetzen zu können. Da die Simulationen zeigen, dass die Netzwerke nie auf lokale Maxima laufen, kann dieser Wert so angepasst werden, dass dieses Abbruchkriterium nie erfüllt wird.
- Der Versuch das Netzwerk zu mutieren scheitert N Mal hintereinander. Auch diese Bedingung ist für die hier gezeigten Simulationsergebnisse nicht von Bedeutung.

Somit ist das Ende einer Adaptiven Wanderung dadurch festgelegt, dass entweder 100% Fitness oder eine bestimmte vorgegebene Anzahl von akzeptierten Mutationen erreicht wird. Diese Zahl reicht von 15.000 bis zu 800.000 Mutationen, je nach durchgeführter Untersuchung, die in den nächsten Kapiteln näher erläutert werden.

5.5 Bemerkungen zu biologischen Grundlagen

Fitnesskriterien

Robustheit als die Fähigkeit eines Systems eine bestimmte Funktion aufrechtzuerhalten unter verschiedenen Bedingungen, ist in der Biologie weit verbreitet [52]. Die robuste Funktion ist dabei wichtig für das Überleben bzw. die Fitness des Systems. Biologische Systeme müssen robust sein gegenüber internen und externen Störungen, die sich z.B. als Stochastizität in der Genexpression oder Veränderungen der Umweltbedingungen manifestieren. Zu beachten dabei ist, dass Robustheit eine relative, keine absolute Eigenschaft ist. Kein System kann in all seinen Funktionen gegen jede Art von Störung robust sein. Somit muss, wenn von Robustheit die Rede ist, angegeben werden, welches Verhalten unter welcher Art von Störung unverändert bleibt. Neben der Robustheit der Hauptfunktion gegen eine Vielzahl von Störungen, können biologische Systeme sehr empfindlich auf andere Störungen oder auf Störungen in anderen Funktionen reagieren [82]. Die Robustheit einer überlebenswichtigen Funktion gegenüber Rauschen (in der Genexpression), wie sie in dieser Arbeit als das simpelste Fitnesskriterium angenommen wird, erscheint somit als eine Minimalanforderung an ein biologisches System (in diesem Fall Genregulationsnetzwerk). Als Erweiterung scheint es plausibel, dass diese Funktion nach einer Störung schnell wieder aufgenommen wird, was die schnelle Rückkehr der Netzwerke zum Attraktor, nach einer kurzen transienten Zeit, impliziert. Das zusätzliche Einbeziehen einer sich ändernden Umwelt, macht es nötig, dass ein System auf externe Stimuli reagiert um z.B. auch unter veränderten Bedingungen überleben zu können oder eine hohe Fitness zu bewahren.

So kann das E. coli Bakterium zwischen verschiedenen Genexpressionsmustern umschalten, je nachdem, ob Lactose oder Glucose in seiner Umgebung vorhanden sind und so seinen Stoffwechsel auf den Abbau des entsprechenden Zuckers einstellen. Die Lambda-Phage wiederum, ein Virus, das E. coli befällt, kann als Reaktion auf Stress zwischen dem lysogenen und dem lytischen Zustand umschalten. Im lysogenen Zustand ist das Virusgenom im Wirtsgenom integriert, wird mit ihm dupliziert, seine Gene aber nicht exprimiert. Im lytischen Zyklus werden die Virusgene exprimiert, das Virus vermehrt sich so und führt am Ende zur Auflösung (Lyse) der Wirtszelle. Bei der Embryonalentwicklung führt eine Spezialisierung von einer Zelle, die noch alle anderen Zelltypen hervorbringen kann, bis hin zu den ausdifferenzierten Körperzellen zu irreversiblen Veränderungen in der Genexpression. Die verschiedenen Körperzelltypen unterscheiden sich darin, welche der Gene eines Genoms in ihnen exprimiert werden. Welche Entwicklung eine Zelle nimmt, hängt zum einen von externen Faktoren ab (von Signalen, die zwischen den Zellen vermitteln und von der Umgebung) und zum anderen von der bisherigen Entwicklung der Zelle und internen Regulatoren [59]. Die ersten beiden Beispiele der Reaktion von Zellen auf ihre Umwelt entsprechen dem „Return“-Kriterium als Selektionskriterium, das letzte Beispiel entspricht dem „Stay“-Kriterium.

Weiterhin ist bekannt, dass Robustheit als Ziel, zelluläre Komplexität zur Folge haben kann. Das Aufrechterhalten bestimmter Funktionen kann zusätzliche Kontrollschleifen erforderlich machen, die das System komplexer und damit auf der anderen Seite wieder anfälliger gegenüber Störungen machen [82]. Ein solcher Effekt zeigt sich bei den Evolutionssimulationen in Kapitel 7.

Mutationen

Eine wichtige Grundlage biologischer Evolution kann in der in dieser Arbeit verwendeten Art künstlicher Evolution eines Netzwerkes erfasst werden: Mutationen verändern den Genotyp (in diesem Fall durch das Netzwerk repräsentiert), Selektion wirkt auf den Phänotyp (in diesem Fall die Attraktoren).

Mutationen finden auf der Ebene der DNA-Sequenz statt. Eine Möglichkeit die Effekte von Mutationen in der DNA-Sequenz auf die Netzwerkebene zu untersuchen, ist die Modellierung eines künstlichen Genoms („Artificial Genome model“ eingeführt von Reil [71]). Dafür wird ein zufälliger String von Basen (mit 0,1,2,3 bezeichnet) erzeugt. Auf diesem String werden bestimmte Abschnitte als Gene und andere als Bindungsstellen für Transkriptionsfaktoren identifiziert. Mit Hilfe einer Vorschrift, die aus den Gensequenzen Transkriptionsfaktorsequenzen ableitet, wird Genexpression simuliert. Nun können die Bindungsstellen auf dem String gefunden (Übereinstimmung mit der Transkriptionsfaktorsequenz) und mit einer weiteren Vorschrift auch ihre regulatorische Wirkung auf das dazugehörige Gen bestimmt werden. So erhält man das Netzwerk von sich gegenseitig regulierenden Genen. In [92] werden fünf Arten von Sequenzmutationen (Punktmutationen, Transpositionen, Inversionen, Löschungen und Genduplikation) auf ihre Wirkung auf die Netzwerkebene hin untersucht. Die meisten Punktmutationen wirken sich nicht auf das Netzwerk aus. Weniger als vier Prozent der Punktmutationen haben das Hinzufügen oder Entfernen eines Gens zur Folge (durch das Modell bedingt, resultiert das Ändern eines Wertes in einer Gensequenz gleichzeitig im Löschen und im Hinzufügen eines Gens). Das Löschen und Hinzufügen einer Verbindung trat in unter zehn Prozent der Fälle auf. Wobei Verbindungen über die Gene definiert sind, die sie regulieren. Das bedeutet, gibt es eine Umverknüpfung (ein Gen reguliert nun ein anderes als zuvor), wird das als eine Löschung und als eine Addition einer Verbindung gezählt. Umverknüpfungsmutationen wie sie im evo-

lutionären Algorithmus dieser Arbeit vorkommen, werden daher gar nicht registriert. Viele der Punktmutationen ziehen dabei mehrere Veränderungen auf der Netzwerkebene nach sich, z.B. das gleichzeitige Hinzufügen eines Knotens und eine Umverknüpfung. Für Transpositionen ergab sich eine Häufigkeit für das Hinzufügen und Entfernen von Genen von unter zehn Prozent, während die Verknüpfungen in 90% der Fälle betroffen waren. Löschungen von Sequenzsegmenten betrafen am häufigsten Verknüpfungen (94%), in 61% der Fälle wurden Gene entfernt. Oft hatten diese Mutationen aber auch das Hinzufügen einer Verbindung zur Folge, selten auch die eines Gens. Diese Effekte sind auf die Methode zurückzuführen wie Verbindungen und Gene im Rahmen des Modells definiert werden. Aus dem selben Grund haben Sequenzinversionen große Auswirkungen auf das Netzwerk. Während das Modell des künstlichen Genoms einige Hinweise darauf liefert, wie sich biologische Mutationen auf der Netzwerkebene auswirken, sind auch diesem Modell Grenzen gesetzt und einige Auswirkungen erweisen sich als unrealistisch. Was sich im Vergleich mit den in der vorliegenden Arbeit verwendeten Mutationsarten zeigt, ist, dass hier zum einen das gemeinsame Auftreten von mehreren Mutationen auf der Netzwerkebene nicht berücksichtigt ist, zum anderen sind Duplikationen und Löschungen von Genen nicht vorgesehen. Die Veränderung von Funktionen wurde im Modell des künstlichen Genoms nicht untersucht. Man kann die Mutationsarten in dieser Arbeit als von Transposonen verursacht sehen, die einen wichtigen Mutationsmechanismus in Zellen darstellen.

Transposonen sind mobile DNA-Segmente, die in der Lage sind sich im Genom zu bewegen und sich zu replizieren. Es werden zwei Klassen von Transposonen unterschieden: Die Klasse-1-Transposonen oder Retrotransposonen haben RNA-Sequenzen als mobile Zwischenprodukte, benutzen also einen sogenannten „copy-and-paste“-Mechanismus, bei dem das Transposon dupliziert wird. Bei Klasse-2-Transposonen oder DNA-Transposonen wird in den meisten Fällen die DNA-Sequenz an sich bewegt („cut-and-paste“-Mechanismus), es ist aber auch replikative Transposition möglich. Transposonen machen ca. 50% des menschlichen Genoms aus. Im Vergleich dazu sind nur etwa 3% des Genoms proteinkodierende Gene. Transposonen können am besten als genomische Parasiten beschrieben werden, die ihr Überleben im Wesentlichen ihrer Fähigkeit verdanken sich schneller zu replizieren als der Wirt, der sie trägt [30]. Es gibt viele Hinweise darauf, dass Transposonen eine wichtige Rolle in der Evolution der eukaryotischen Genregulation hatten. Transposonen können Genregulationsnetzwerke auf vielfältige Arten beeinflussen. In Abbildung 5.2 sind verschiedene Insertionsorte von Transposonen und ihre möglichen Wirkungen dargestellt. Transposonen können Verbindungen zwischen Genen löschen, wenn sie die Bindungsstellen der Regulatoren mit ihrer Insertation zerstören (Abb. 5.2 (b)). Sie können Verbindungen hinzufügen, wenn sie selbst Bindungsstellen tragen und so auch Umverknüpfungen verursachen (Abb. 5.2 (c)). Sie können auch die Funktionen verändern, wenn sie eine alternative Startsequenz für die Transkription einfügen oder die Transkription eines Gens stilllegen (Abb. 5.2 (a),(e)). Transposonen können zudem die DNA-Abschnitte ihrer Wirte mobilisieren, reorganisieren und duplizieren [30].

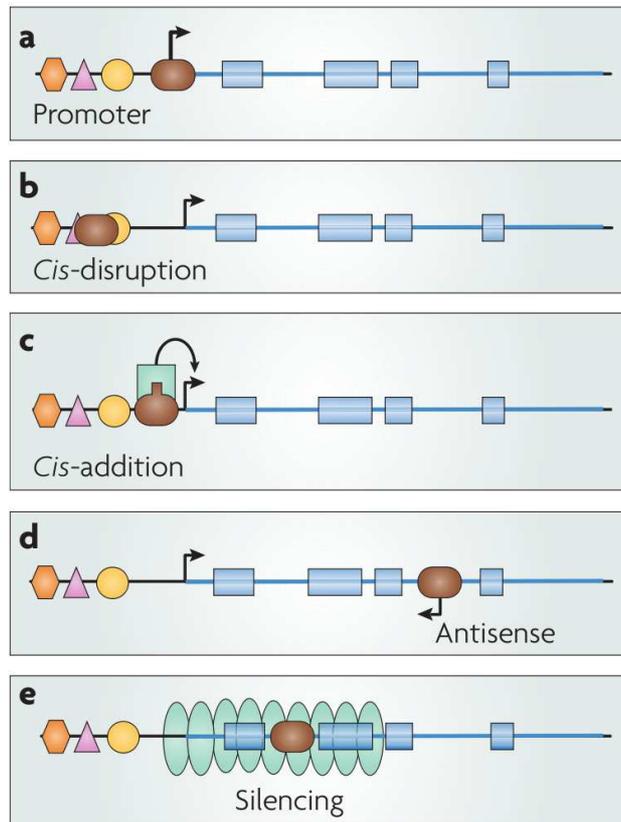


Abbildung 5.2: Transposonen können die Genexpression auf verschiedene Arten beeinflussen (nach [30]). (a) Ein Transposon (braun), das sich stromaufwärts (entgegen der Ableserichtung bei der Transkription) eines Gens (blau) eingefügt hat, kann Promoterregionen enthalten und damit einen alternativen Transkriptionsstartort einführen, (b) ein oder mehrere cis-Elemente (zu einem Gen gehörende Abschnitte, an die die regulierenden Moleküle binden) zerstören, (c) oder ein neues cis-Element einfügen. (d) Zusätzlich kann ein Transposon, das sich in ein Intron eingefügt hat, zu einer Transkription in Gegenrichtung führen, die die reguläre Transkription stören kann. (e) Schließlich kann ein Transposon auch als Keimzentrum für die Formation von Heterochromatin (dicht gepackte Bereiche der DNA, die dadurch nicht abgelesen werden können) dienen und damit die Transkription eines oder mehrerer benachbarter Gene stilllegen.



6 Evolution Boolescher Netzwerke unter Selektion für robuste Dynamik

In diesem Kapitel wird die Evolution mit den einfacheren Versionen des Fitnesskriteriums behandelt, die nur die Robustheit eines Attraktors messen und noch nicht die Reaktion auf externe Eingänge beinhalten (s. Kapitel 5).

In Kapitel 6.1 und 6.2 werden zunächst evolvierte Netzwerke mit allen Booleschen Funktionen betrachtet. Beginnend mit der Zusammenstellung einiger genereller Eigenschaften der evolvierten Netzwerke, sollen Modifikationen des Derrida-Plots zuerst vorgestellt und dann auf diese Netzwerke angewendet werden, um sie tiefergehend zu charakterisieren und mit zufälligen Booleschen Netzwerken zu vergleichen. Schließlich werden der Derrida-Plot und die mit ihm verwandten Darstellungen auch für das Zellzyklusregulationsnetzwerk der Bäckerhefe (s. Kapitel 2.4) ausgewertet.

Kapitel 6.3 beschäftigt sich mit der Evolution von kanalisierenden und Schwellenwertnetzwerken. Der Weg der beiden Netzwerkarten bis zum Erreichen von 100% Fitness soll verglichen werden. Da übereinstimmendes Verhalten für unterschiedliche Netzwerkgrößen der beiden Netzwerktypen gefunden wird, wird die Abhängigkeit der Evolution von der Netzwerkgröße untersucht.

6.1 Generelle Eigenschaften evolvierter Boolescher Zufallsnetzwerke

Die in diesem Kapitel untersuchten Netzwerke wurden mit Hilfe der Adaptiven Wanderung und dem Fitness-Kriterium, wie unter Punkt 2 in Kapitel 5 beschrieben, erhalten. Die Simulationen wurden mit verschiedenen Netzwerkgrößen bis zu $N = 90$ und Anfangskonnektivitäten von $K_{\text{ini}} = 1$ und 2 durchgeführt. Die Adaptive Wanderung wurde nach 15.000 Schritten, d.h. nach 15.000 akzeptierten Mutationen, beendet. Da die Netzwerke nach nur wenigen Schritten eine Fitness von 100% erreicht haben, sind die meisten dieser Mutationen neutral, d.h. sie ändern die Fitness nicht, und die Netzwerke vollführen Zufallsbewegungen auf dem Plateau von 100% Fitness. In Kapitel 6.2.1 werden repräsentative Netzwerke aus diesen Simulationen mit den Größen $N = 11, 50$ und 80 betrachtet, um eine gute Vergleichbarkeit zu haben sowohl mit den Derrida-Plots von zufälligen Netzwerken als auch mit denen des biologischen Beispiels der Bäckerhefe (Kapitel 6.2.2).

Auf dem Plateau von 100% Fitness fluktuiert die mittlere Konnektivität für Netzwerke mit $N = 11$ um einen Wert von $K = 3,71$, für Netzwerke mit $N = 50$ und 80 um einen Wert von $K = 3,56$. Dies ist in Abbildung 6.1 für längere Simulationsläufe gezeigt. Für Zufallsnetzwerke mit allen Booleschen Funktionen ist eine Konnektivität $K > 2$ ein Charakteristikum des chaotischen Regimes. Die Verteilung der Anzahl an Eingängen pro Knoten ändert sich während der Adaptiven Wanderung allerdings von einer konstanten Verteilung am Anfang der Simulation, bei der jeder Knoten die gleiche Anzahl an Eingängen hat, zu einer breiten Verteilung der Eingangsverbindungen, die ein Maximum bei dem Wert 1 und einen exponentiellen Abfall für größere Anzahlen von Eingangsverbindungen zeigt. In Abbildung 6.2 ist beispielhaft die Eingangsgradverteilung nach 15.000 Mutationen für ein $N = 80$ -Knoten-Netzwerk gezeigt, gemittelt über 45 Simulationsläufe. In der kleinen Abbildung ist zu sehen, dass sich die Eingangsgradverteilung nicht mehr stark verändert, wenn die Netzwerke auf dem Plateau von 100% Fitness sind. Die Verteilung der Ausgangsverbindungen ändert sich während des

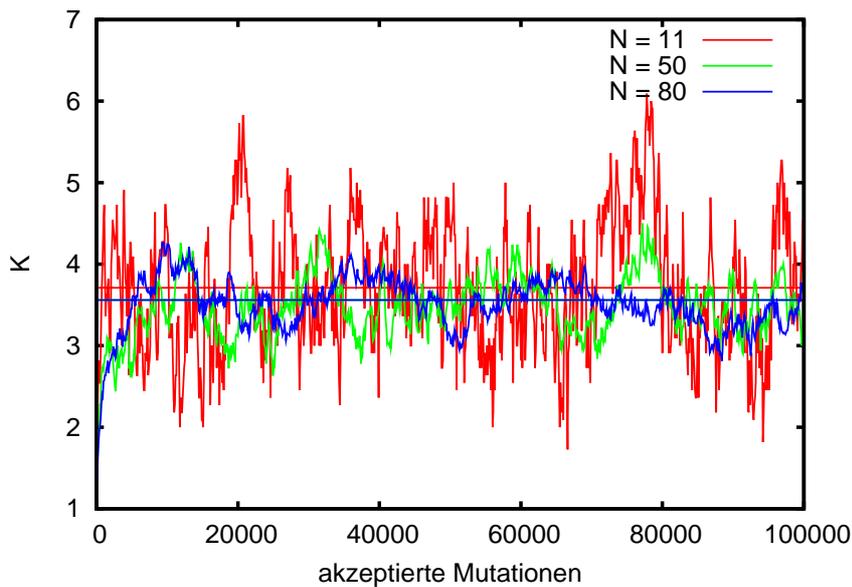


Abbildung 6.1: Entwicklung des K -Werts über 100.000 Mutationsschritte für drei verschiedene Netzwerkrealisationen mit $N = 11, 50$ und 80 . Die horizontalen Linien entsprechen dem Mittel der K -Werte über die Evolutionszeit für die verschiedenen Netzwerkrealisationen. (Für $N = 50$ und 80 liegen sie übereinander.)

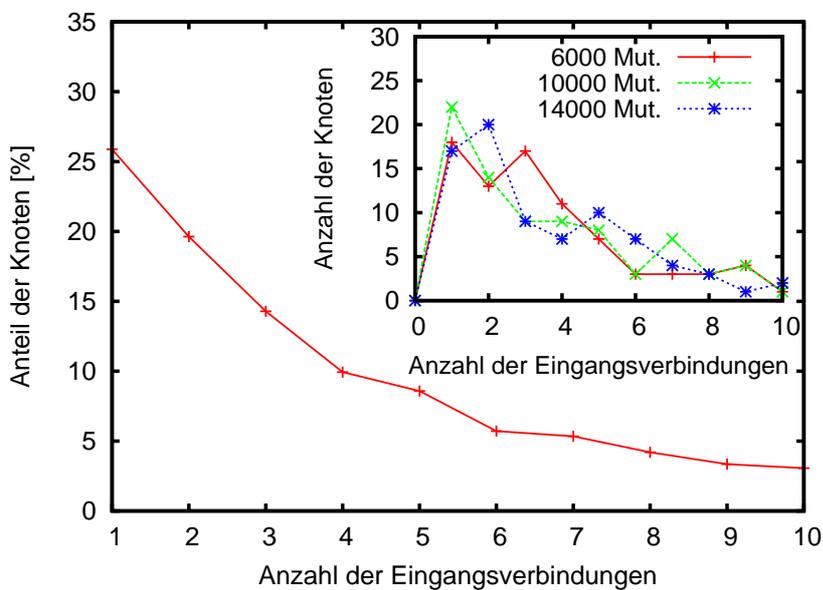


Abbildung 6.2: Verteilung der Anzahl an Eingangsknoten für ein $N = 80$ -Knoten-Netzwerk nach 15.000 Mutationen gemittelt über 45 Simulationsläufe. Bild im Bild: Eingangsknotenverteilung zu drei verschiedenen Zeitpunkten (nach 6.000, 10.000 und 14.000 Mutationen) für einen Simulationslauf.

Evolutionslaufs hingegen überhaupt nicht, sie bleiben Poisson-verteilt wie bei den zufälligen Anfangsnetzwerken. Die Auswertung der Sensitivität anhand der Funktionen, die in den einzelnen Netzwerken auftreten (vgl. [43]), ergibt $\lambda > 1$, ein Kennzeichen des chaotischen Regimes. Die hohe Robustheit der evolvierten Netzwerke kann einem Zustandsraum zugeschrieben werden, der einen Attraktor mit einem riesigen Einzugsbereich enthält, der im Allgemeinen mehr als 90% aller Zustände umfasst. Trotz der hohen λ -Werte sind die Attraktoren der evolvierten Netzwerke kurz, meist einstellig, und nur wenige Knoten verändern ihren Wert auf den Attraktoren. Dies sind Eigenschaften, die für die gefrorene Phase charakteristisch sind (vgl. Kapitel 2.3.2).

Zusammenfassend kann man sehen, dass nach der Evolution Merkmale der chaotischen und der gefrorenen Phase in einem einzigen Netzwerk gefunden werden.

6.2 Der Derrida-Plot und seine Modifikationen

Wie schon in Kapitel 2.3.3 dargestellt, gibt es Bemühungen zelluläre Netzwerke zu klassifizieren und ihre Dynamik einem der Regime gefroren, kritisch, chaotisch zuzuordnen, die man aus der Untersuchung von Booleschen Zufallsnetzwerken kennt. Neben verschiedenen anderen Eigenschaften der Netzwerke, die charakteristisch für ein bestimmtes Regime sind, kann der Parameter λ bzw. der Derrida-Plot benutzt werden, um zwischen den Regimen zu unterscheiden (s. Kapitel 2.3). Wie in Kapitel 2.3.2 beschrieben, bildet der Derrida-Plot den anfänglichen Hamming-Abstand zwischen den Zuständen zweier identischer Netzwerke auf denjenigen ab, der nach einem Zeitschritt erhalten wird.

Der generalisierte Derrida-Plot

Für den generalisierten Derrida-Plot wird der Hamming-Abstand $H(t > 1)$ für Zeitschritte größer als Eins als Funktion von $H(0)$ aufgetragen. In Abbildung 6.3 ist das für $t = 1, \dots, 5$ und ein chaotisches Boolesches Zufallsnetzwerk mit $N = 60$ und $K = 4$ zu sehen. Jeder Datenpunkt in dieser und in den folgenden Abbildung ist gemittelt über 1000 Kombinationen von einem zufällig gewählten Anfangsnetzwerkzustand und zufällig gewählten gestörten Knoten. Die Linien, die die Punkte verbinden, dienen der Veranschaulichung. In Abbildung 6.4 wird der generalisierte Derrida-Plot für $t = 2$ mit der zweifach iterierten Derrida-Abbildung verglichen. Bezeichnet $H(1) = f(H(0))$ die Derrida-Abbildung, dann ist die Abbildung nach zweifacher Iteration $f(f(H(0)))$ (und entsprechendes für höhere Iterationen). Im Folgenden wird die zweifach iterierte Abbildung mit H^2 abgekürzt. Im Rahmen der Annealed Approximation erwartet man, dass $H(2) = H^2$ und Entsprechendes für größere Zeiten. Die Annealed Approximation vernachlässigt Korrelationen zwischen Knoten oder zwischen verschiedenen Zeiten. Daher hängt der Hamming-Abstand zu einem Zeitpunkt nur vom Hamming-Abstand im vorangegangenen Zeitschritt ab, aber nicht vom genauen Zustand des Systems. Ist die Annealed Approximation für den Derrida-Plot von Booleschen Zufallsnetzwerken gültig, muss zudem gelten, dass alle Plots für $t \geq 1$ den selben Fixpunkt haben, d.h. sie müssen die Winkelhalbierende am selben Punkt schneiden. Beides sieht man in den Abbildungen 6.3 und 6.4 als erfüllt. Dies ist umso mehr erstaunlich, als dass zu erwarten ist, dass die Annealed Approximation erst im thermodynamischen Limes unendlich großer Netzwerke exakt wird und die hier betrachteten Netzwerke klein sind. Man findet gleiche Qualität der Übereinstimmung

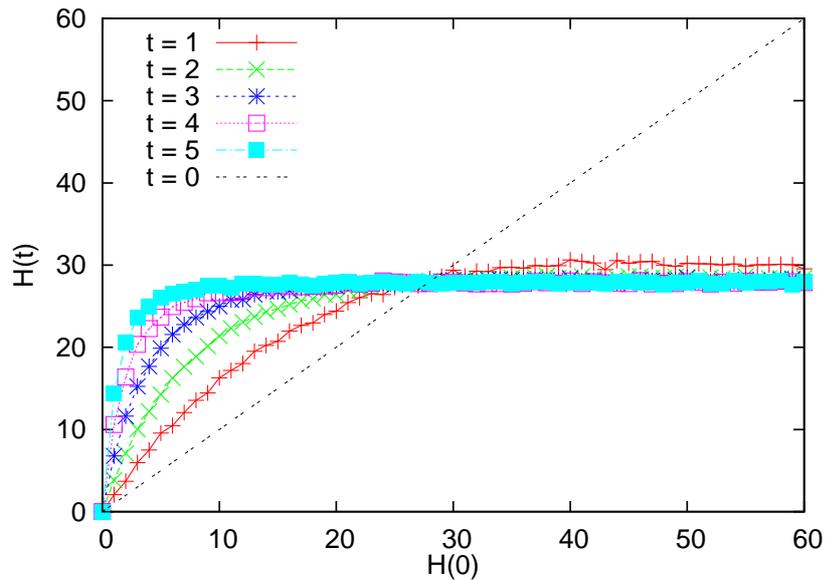


Abbildung 6.3: Generalisierter Derrida-Plot für ein Netzwerk mit $K = 4$ und $N = 60$ und Zeitschritte $t = 1, \dots, 5$. Die gestrichelte schwarze Linie ist die Winkelhalbierende.

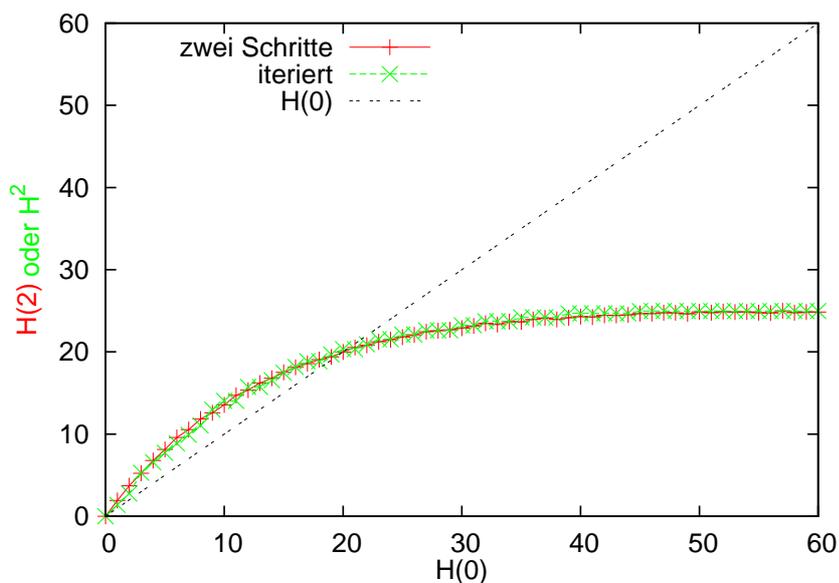


Abbildung 6.4: Generalisierter Derrida-Plot $H(t = 2)$ („zwei Schritte“) und iterierte Derrida-Abbildung H^2 („iteriert“) von einem Netzwerk mit $N = 60$ und $K = 3$. Die gestrichelte schwarze Linie ist die Winkelhalbierende.

für chaotische, kritische und gefrorene Boolesche Zufallsnetzwerke, während im Folgenden zu sehen sein wird, dass dies für evolvierte Netzwerke nicht gilt.

Der modifizierte Derrida-Plot

Der modifizierte Derrida-Plot unterscheidet sich vom ursprünglichen Derrida-Plot in der Wahl des Anfangszustandes. Während der eigentliche Derrida-Plot zufällige Zustände betrachtet, geht der modifizierte Derrida-Plot nur von Zuständen aus, die sich auf Attraktoren befinden. Konkret wurde das im zweiten Fall so realisiert, dass man das Netzwerk zunächst

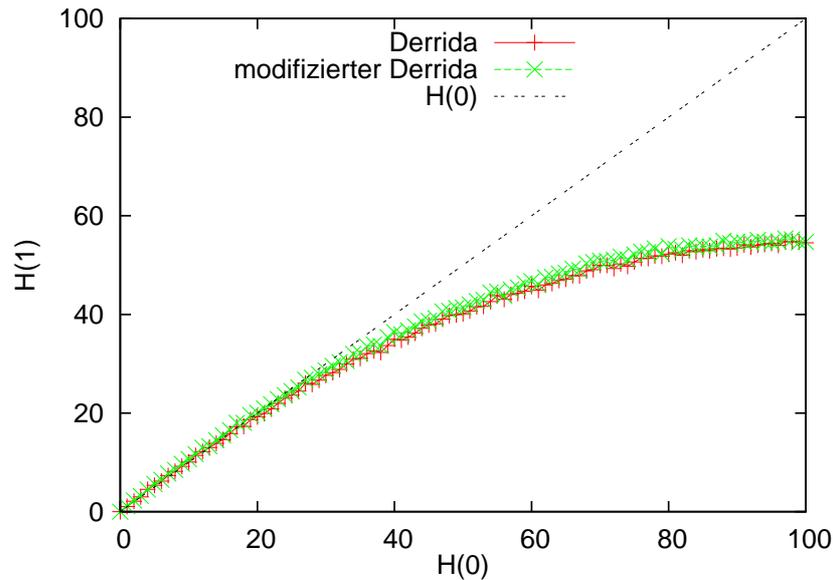


Abbildung 6.5: Derrida-Plot und modifizierter Derrida-Plot von einem Netzwerk mit $N = 100$ und $K = 2$. Die gestrichelte schwarze Linie ist die Winkelhalbierende.

von einem zufälligen Zustand aus auf einen Attraktor laufen lässt. Auf diese Weise werden die Attraktoren nach Basinggröße gewichtet.

In der Annealed Approximation existiert der Begriff des Attraktors nicht. Ein Attraktor ist eine periodische Abfolge von Zuständen und kann nur auftreten, wenn die Verbindungen zwischen den Knoten sich mit der Zeit nicht ändern. Da ein Attraktor nur einen Bruchteil aller möglichen Zustände enthält, könnte man davon ausgehen, dass sich der Derrida-Plot ändert, wenn man nur Zustände auf einem Attraktor betrachtet. In Abbildung 6.5 sind der ursprüngliche und der modifizierte Derrida-Plot für ein kritisches Boolesches Zufallsnetzwerk der Größe $N = 100$ aufgetragen. Die beiden Kurven stimmen sehr gut überein. Eine gleichermaßen gute Übereinstimmung findet sich auch für chaotische und gefrorene Netzwerke. Dieses Ergebnis impliziert, dass Attraktorzustände „typische“ Zustände sind, die nicht von transienten Zuständen unterschieden werden können, wenn es um die Auswertung der zeitlichen Veränderung des Hamming-Abstandes geht. In [49] wurden diese zwei Arten von Derrida-Plots für Netzwerke mit einer Untermenge von kanalisierenden Funktionen betrachtet. Es wurde gefunden, dass sich die beiden Abbildungen unterscheiden. Der Grund ist, dass in diesem Fall der Anteil an Nullen und Einsen in zufälligen Anfangszuständen und in Attraktorzuständen verschieden sind und sich λ während der zeitlichen Entwicklung von einem zufälligen Anfangszustand zum stationären Zustand des Netzwerkes verändert, wie in Kapitel 2.3.2 dargestellt.

6.2.1 Anwendung auf evolvierte Netzwerke

Die eben vorgestellten Abwandlungen des Derrida-Plots sollen nun auf die evolvierten Netzwerke angewendet werden. Die Abbildungen 6.6 – 6.8 zeigen die Derrida-Abbildungen und ihre generalisierten, iterierten und modifizierten Versionen für drei evolvierte Netzwerke mit $N = 11, K = 3,63$; $N = 50, K = 3,12$ und $N = 80, K = 3,39$. Abbildung 6.6 zeigt die generalisierten Derrida-Plots. Sie unterscheiden sich deutlich von denen der zufälligen Booleschen Netzwerke. Für Boolesche Zufallsnetzwerke schneiden sich die Kurven alle am selben Punkt,

der ein Fixpunkt der iterierten Derrida-Abbildungen ist. Der Fixpunktwert muss gleich 0,5 mal der Anzahl der nichtgefrorenen Knoten N_{nf} im stationären Zustand sein, da jeder dieser Knoten während eines Zeitschritts seinen Zustand mit einer Wahrscheinlichkeit von 0,5 ändert. Da die iterierte Abbildung zu diesem Fixpunkt konvergieren muss, nähern sich die generalisierten Derrida-Plots der Booleschen Zufallsnetzwerke mit steigendem t einer Stufenfunktion an mit dem Wert 0 bei $H = 0$ und dem Fixpunktwert $0,5 N_{\text{nf}}$ für alle $H > 0$ (s. Abbildung 6.3). Im Gegensatz dazu nimmt der maximal erreichte Wert bei den evolvierten Netzwerken mit wachsendem t ab und genauso reduziert sich der Wert des Fixpunktes der Abbildung. Obwohl die Anfangszustände weit weg von einem Attraktor sein können, erfahren die Trajektorien im Zustandsraum Einschränkungen, weil sie zu dem selben kleinen Teil des Zustandsraums hingezogen werden. Abbildung 6.6 zeigt wie mit wachsendem t mehr und mehr Knoten einfrieren, während sich das System dem Attraktor nähert.

Abbildung 6.7 zeigt die Kurven des zwei-Schritt-Derrida-Plots und der iterierten Derrida-Abbildung für die selben drei Netzwerke wie in Abbildung 6.6. Im Unterschied zu zufälligen Netzwerken (s. Abbildung 6.4) stimmen die beiden Kurven für die evolvierten Netzwerke nicht überein. Für größere Werte von H liegen die zwei-Schritt-Derrida-Plots unterhalb der iterierten Derrida-Abbildung. Dies deutet erneut darauf hin, dass sich Trajektorien nicht so schnell voneinander entfernen wie es von der ersten Iteration suggeriert wird.

In Abbildung 6.8 schließlich sieht man den Vergleich von Derrida-Plot und modifiziertem Derrida-Plot der evolvierten Netzwerke. Die anfänglichen Steigungen der modifizierten Plots sind nahe Eins, was bedeutet, dass diese Netzwerke kritisch erscheinen bei Störungen die auf Attraktoren erfolgen. Im Gegensatz dazu sind die Anfangssteigungen der ursprünglichen Derrida-Plots größer als Eins, was bedeutet, dass die Netzwerke zunächst chaotisch erscheinen, wenn man von einem beliebigen Anfangszustand startet. Ein solcher Unterschied in der Anfangssteigung erscheint plausibel, da diese Netzwerke evolviert wurden, um stabile Attraktoren zu haben. In [49] wurden diese zwei Arten von Plots für den Datensatz von 139 kanalisierenden Regulationsfunktionen, wie sie von Harris et al. [39] aus diversen biologischen Publikationen zusammengetragen wurden, (mit zufälliger Topologie) untersucht und ein qualitativ ähnliches Verhalten wie für die in dieser Arbeit betrachteten evolvierten Netzwerke gefunden: Für Störungen von Knotenwerten auf Attraktoren ist die Steigung des Derrida-Plots flacher als wenn von völlig zufälligen Zuständen ausgegangen wird und die beiden Plots kreuzen sich, wenn ein großer Teil der Knoten gestört wird.

Hier sind nur die Ergebnisse für drei evolvierte Netzwerke aufgeführt. Für andere evolvierte Netzwerke ist das qualitative Verhalten der Kurven, die soeben besprochen wurden, das gleiche. Die genaue Form der Kurven allerdings variiert bei verschiedenen Netzwerken der selben Größe.

All diese Ergebnisse deuten darauf hin, dass sich die Zustandsräume der evolvierten Netzwerke fundamental von denen von zufälligen Booleschen Netzwerken unterscheiden. Sie enthalten einen dominanten Attraktor, der den Hamming-Abstand zwischen zwei Trajektorien bereits während der ersten paar Schritte beeinflusst. Die evolvierten Netzwerke verbinden Eigenschaften, die charakteristisch sind für die gefrorene Phase (die sehr kurzen Attraktoren und ihre geringe Anzahl), für die chaotische Phase (die anfängliche Steigung des Derrida-Plots) und für das kritische Regime (die Anfangssteigung des modifizierten Derrida-Plots und die große Anzahl an gefrorenen Knoten).

6.2.2 Anwendung auf das Zellzyklusregulationsnetzwerk der Bäckerhefe

Nun sollen die verschiedenen Plots auch auf ein Modell eines realen biologischen Netzwerkes angewendet werden, auf das Zellzyklusregulationsnetzwerk der Bäckerhefe. Wie bereits aus Kapitel 2.4 bekannt, zeigt auch dieses Netzwerk ein sehr robustes Verhalten. Dieses robuste Verhalten kann der Größe des Einzugsgebiets des biologisch relevanten Fixpunktes zugeschrieben werden und ist keine Folge einer bestimmten Struktur des Zustandsraums [94]. Der generalisierte, der iterierte und der modifizierte Derrida-Plot für dieses Modell sind in Abbildung 6.9 zu sehen. Alle drei Abbildungen zeigen wieder deutliche Abweichungen von denen für zufällige Boolesche Netzwerke. Der generalisierte Plot ähnelt dem eines chaotischen Booleschen Zufallsnetzwerkes, wenn man von der Anfangssteigung und dem fast stationären Fixpunkt ausgeht. Für größere H und t wird die Kurve allerdings nicht horizontal wie für Zufallsnetzwerke. Dies muss mit dem großen Einzugsgebiet des Hauptfixpunktes und den langen Transienzzeiten zusammenhängen, die verhindern, dass das System ein stationäres Verhalten zeigt. Im Gegensatz zu den Netzwerken, die auf Robustheit evolviert wurden, liegt der zwei-Schritt-Derrida-Plot des Hefezellzyklusregulationsnetzwerkes über der Kurve der iterierten Derrida-Abbildung (s. Abb. 6.9 (b)). Dies deutet darauf hin, dass sich die Trajektorien schneller voneinander entfernen, als es durch die erste Iteration suggeriert wird. Genauso liegt die Kurve des modifizierten Derrida-Plots über der des ursprünglichen (s. Abb. 6.9 (c)), was bedeutet, dass Störungen der Attraktorzustände größere Auswirkungen haben als Störungen von zufälligen Zuständen. Dies stimmt gut mit der Beobachtung überein, dass der Fixpunkt des Zellzyklus empfindlicher gegenüber Störungen ist als andere Zustände des Modells [33].

6.2.3 Fazit

Die Simulationsergebnisse deuten darauf hin, dass die übliche Klassifizierung der Dynamik Boolescher Netzwerke in gefrorenes, kritisches und chaotisches Verhalten nicht auf evolvierte Netzwerke angewendet werden kann, da diese Netzwerke Merkmale aller drei Verhaltensweisen zeigen. Dennoch erlauben die in diesem Kapitel vorgestellten Arten von Derrida-Plots einen detaillierten Einblick in den Zustandsraum und damit in das tatsächliche dynamische Verhalten der Netzwerke. Es lässt sich zudem schließen, dass die einfache Klassifizierung auch für biologische Netzwerke nicht geeignet ist, da auch sie durch ihre evolutionäre Geschichte geformt sind. Diese Arbeit bestätigt und ergänzt damit die Arbeit von Sevim und Rikvold [79], die zu ähnlichen Schlussfolgerungen gelangen.

Schon für nicht völlig zufällige Boolesche Netzwerke, nämlich solche mit einer anderen Funktionenverteilung, fanden Kesseli et al. [49], dass nicht eine der Derrida-Abbildung alleine alle Fragen beantworten kann, die im Rahmen der Annealed Approximation von Interesse sind. Sie verglichen drei Derrida-Abbildungen miteinander: eine, die vernachlässigt, dass b_t einen anderen stationären Wert annehmen kann als $1/2$, eine, die den stationären Zustand b^* berücksichtigt und damit in Simulationen von einem Attraktorzustand aus startet und eine dritte, die b^* mit einer Störung nicht verändert.

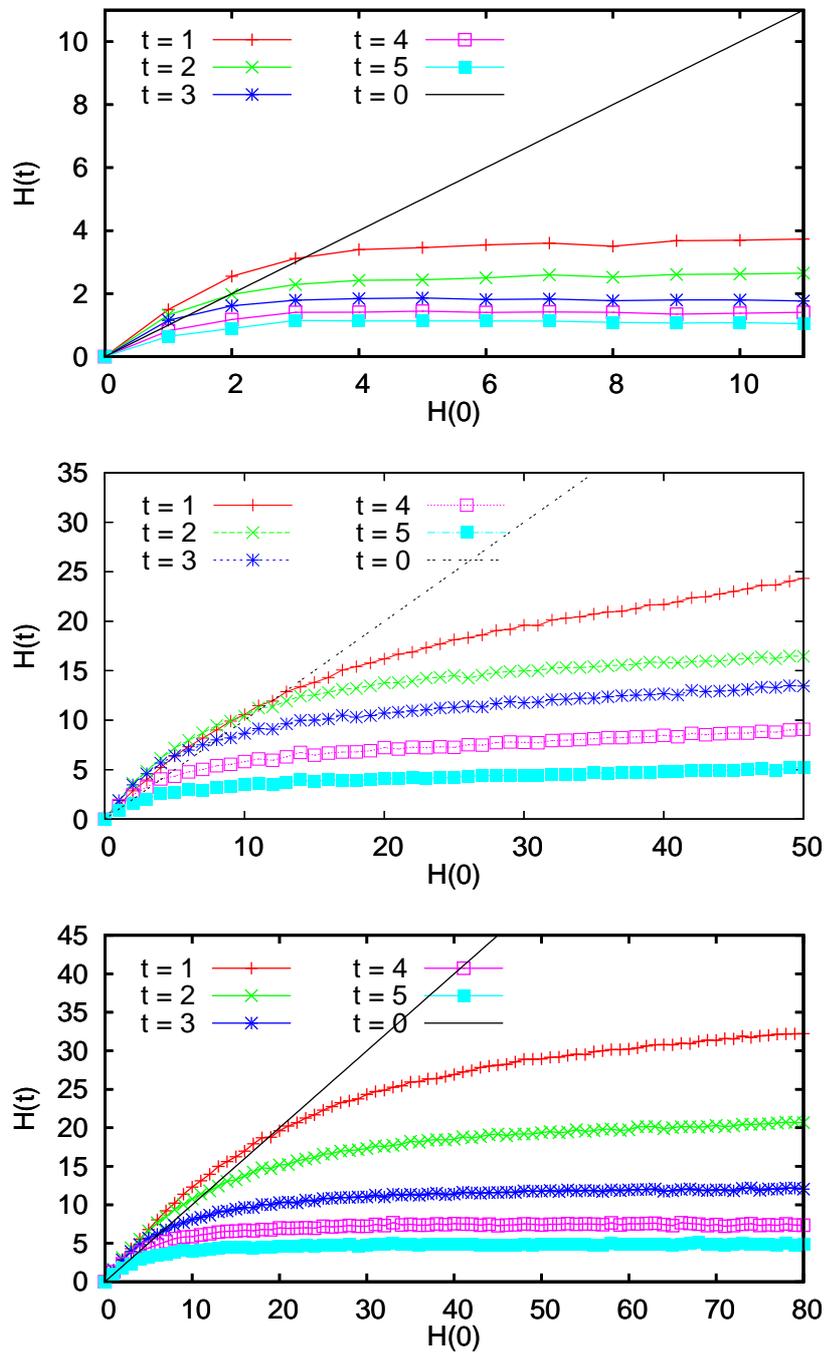


Abbildung 6.6: Generalisierte Derrida-Plots mit $t = 1, \dots, 5$ von drei evolvierten Netzwerken mit $N = 11, K = 3,63$; $N = 50, K = 3,12$; $N = 80, K = 3,39$.

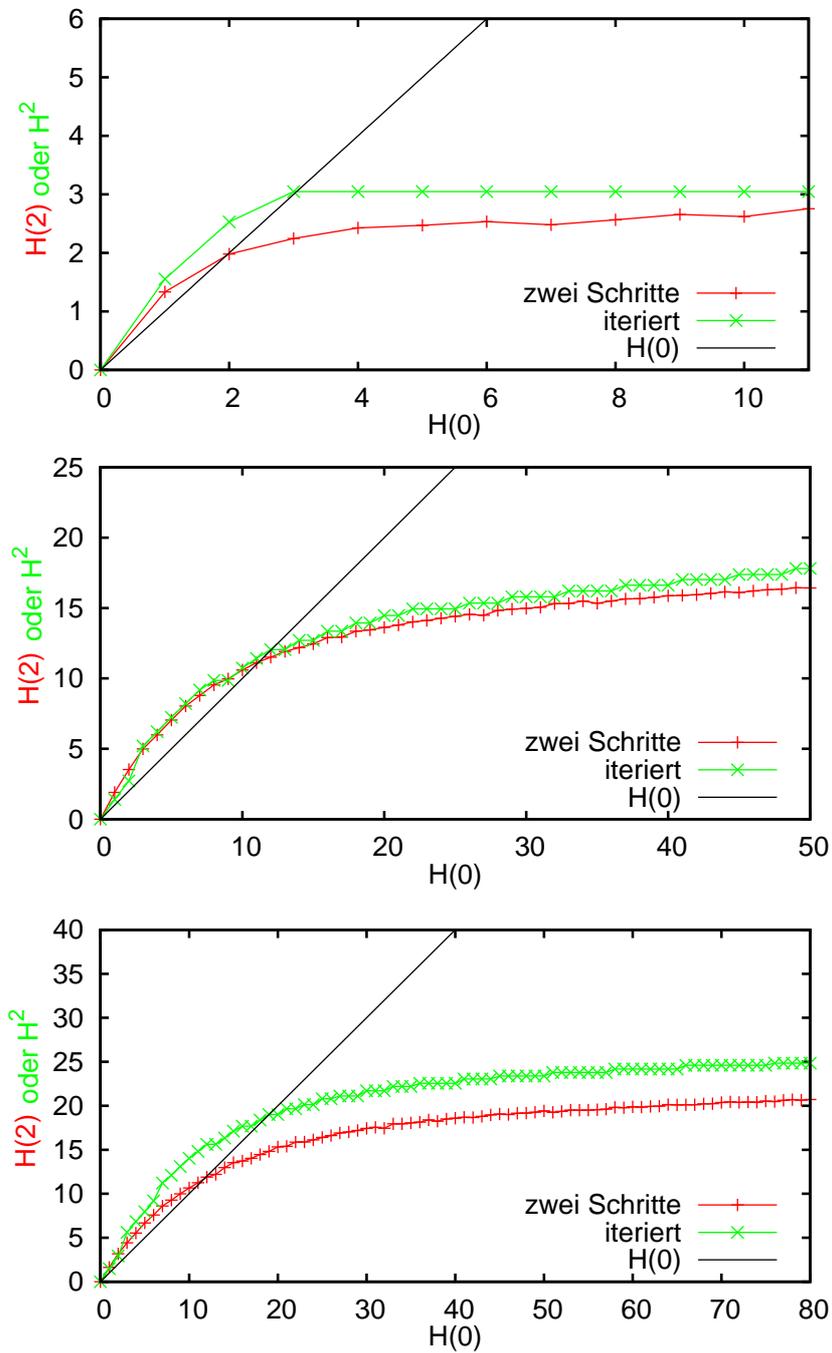


Abbildung 6.7: Generalisierter Derrida-Plot $H(t = 2)$ („zwei Schritte“) und iterierte Derrida-Abbildung H^2 („iteriert“) von den drei evolvierten Netzwerken mit $N = 11$, $K = 3,63$; $N = 50$, $K = 3,12$; $N = 80$, $K = 3,39$.

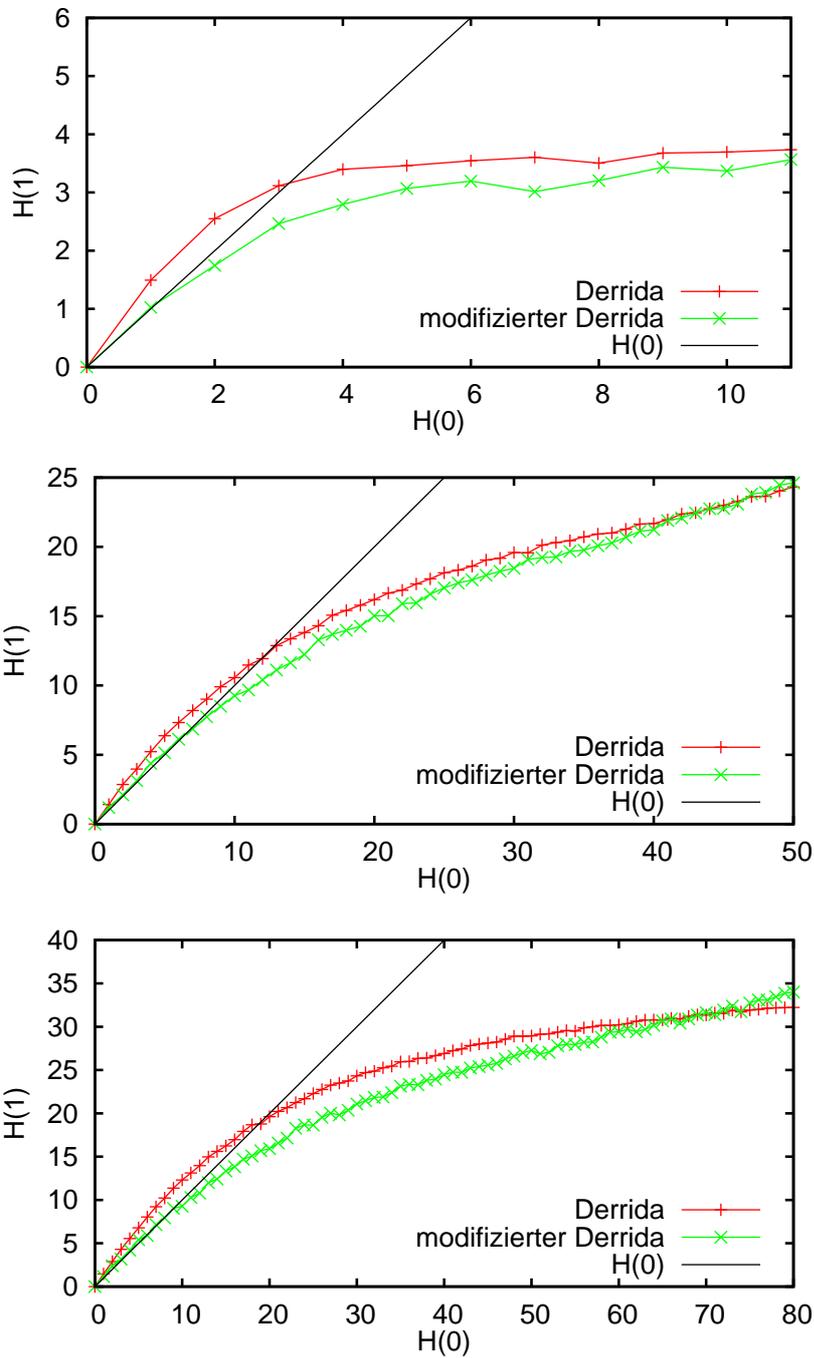


Abbildung 6.8: Derrida-Plot und modifizierter Derrida-Plot von den drei evolvierten Netzwerken mit $N = 11, K = 3,63$; $N = 50, K = 3,12$; $N = 80, K = 3,39$.

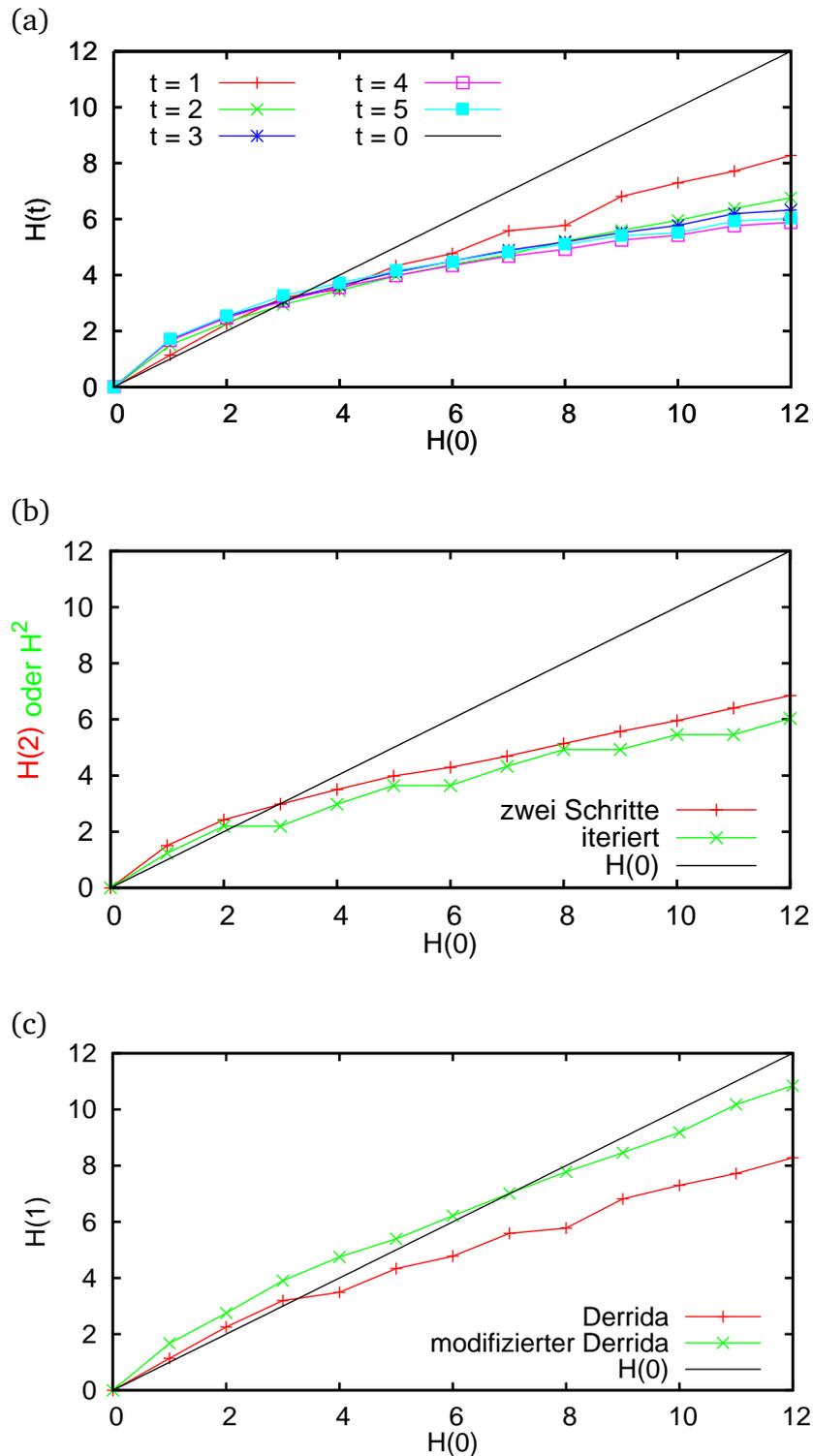


Abbildung 6.9: Generalisierter, iterierter und modifizierter Derrida-Plot für das Hefezellzyklusnetzwerk von Li et al. [57]

6.3 Auswirkungen der Netzwerkgröße auf die Evolution von kanalisierenden und Schwellenwertnetzwerken

Die Evolution von kanalisierenden Booleschen Netzwerken mit dem einfachsten Fitnesskriterium, das nur die Robustheit eines Attraktors misst, wurde bereits in einer früheren Arbeit behandelt [85]. Bevor im nächsten Kapitel striktere Fitnesskriterien, die die Reaktion auf einen externen Eingang beinhalten, auf kanalisierende und Schwellenwertnetzwerke angewendet werden, sollen die Wege der Netzwerke bis zum Fitnessmaximum für das einfache Kriterium miteinander verglichen werden. Die Simulationen werden für verschiedene Netzwerkgrößen und Anfangskonnektivitäten durchgeführt. Die Anfangskonnektivitäten reichen für kanalisierende Netzwerke von $K_{\text{ini}} = 1$ bis 5 und für Schwellenwertnetzwerke von 1 bis 7. Um die gewonnenen Daten für die beiden Netzwerktypen vergleichen zu können, werden diese nicht über K_{ini} sondern über λ_{ini} geplottet. Für kanalisierende Netzwerke wird λ gemäß Gleichung 2.3.1 berechnet, für Schwellenwertnetzwerke mit Hilfe der Gleichungen 4.2 und 4.4. Nach Auswertung des Phasendiagramms für Schwellenwertnetzwerke (vgl. Abb. 4.4) erscheint $K_{\text{ini}} = 1$ bis 7 und $h = -0,5$ als einzig sinnvolle Parameterwahl, um vergleichbar viele Werte von λ_{ini} im Bereich von 0,5 bis 1,5 wie für kanalisierende Netzwerke zu haben. So deckt λ den gefrorenen, den kritischen und den chaotischen Bereich ab. Die Abhängigkeiten zwischen K und λ für kanalisierende und Schwellenwertnetzwerke sind in Abbildung 6.10 für den in dieser Arbeit relevanten Bereich zu sehen.

In Abbildung 6.11 sind die Anfangsfitness, die Anteile der akzeptierten und der neutralen Mutationen, sowie das Verhältnis der Attraktorlängen am Anfang und bei Erreichen von 100% Fitness für kanalisierende Netzwerke der Größe $N = 100$ und für Schwellenwertnetzwerke der Größen $N = 100$ und 200 zu sehen. Jeder der Punkte in dieser und in den folgenden Abbildungen ist eine Mittelung über mindestens 1000 Simulationsläufe. Man sieht, dass die Kurven

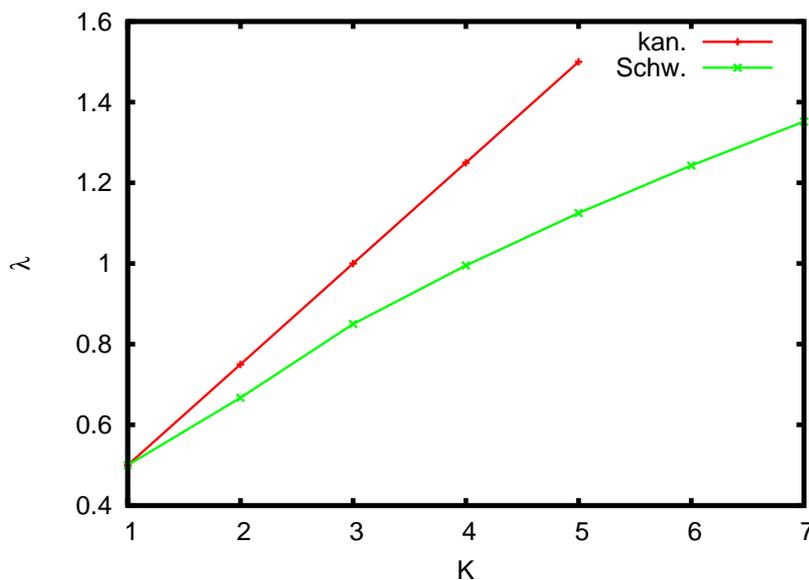


Abbildung 6.10: λ in Abhängigkeit von K für kanalisierende und Schwellenwertnetzwerke, für den in dieser Arbeit genutzten Wertebereich. Die durchgezogenen Linien zwischen den Punkten dienen der besseren Anschauung, λ -Werte sind nur für ganzzahlige K berechnet.

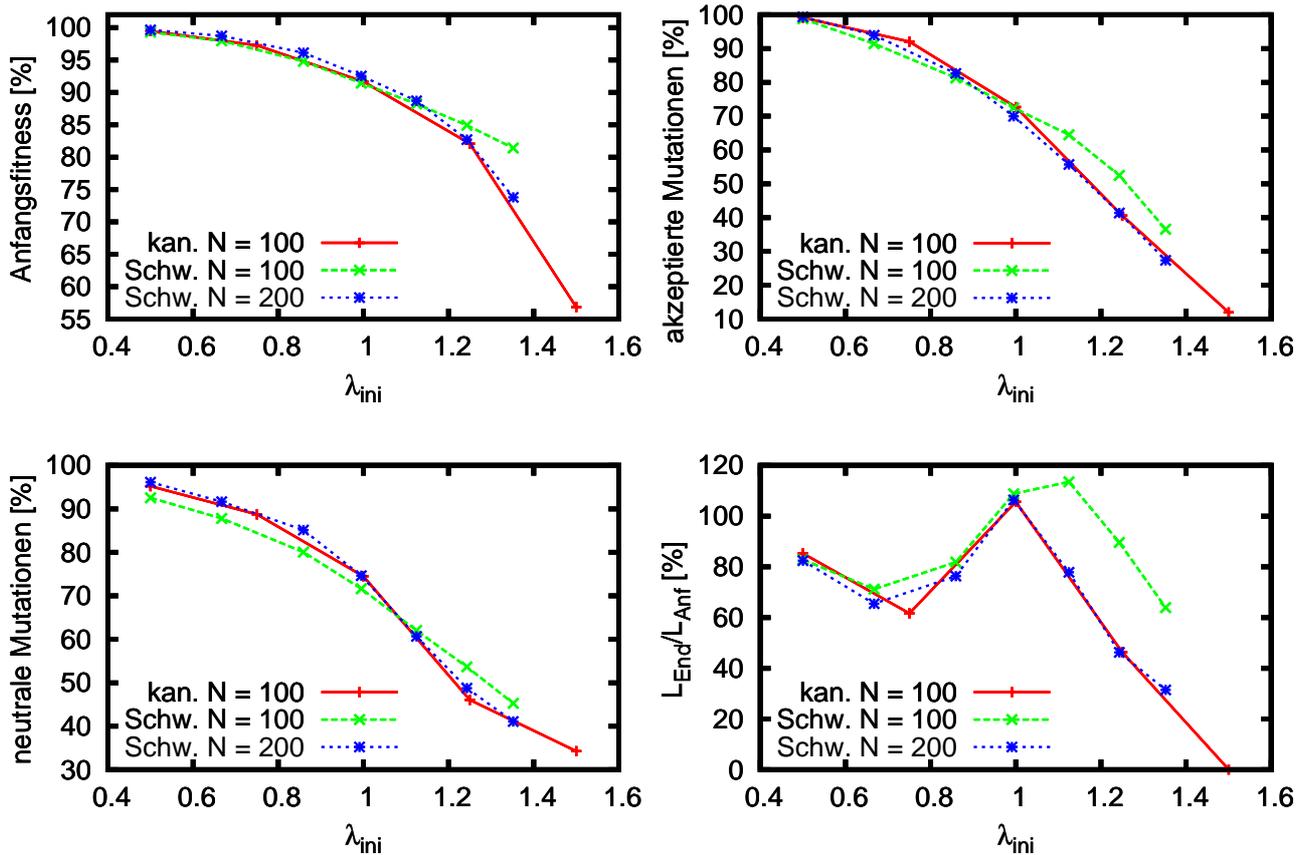


Abbildung 6.11: Anfängliche Fitness, Anteil der akzeptierten Mutationen, Anteil der neutralen Mutationen und das Verhältnis der Attraktorlängen am Anfang der Simulation und nach Erreichen von 100% Fitness (L_{End}/L_{Anf}) für kanalisierende und Schwellenwertnetzwerke. Das Verhalten bis die Maximalfitness erreicht ist, stimmt für kanalisierende und Schwellenwertnetzwerke gut überein, wenn die Schwellenwertnetzwerke größer sind als die kanalisierenden. Die Linien zwischen den Punkten dienen der Veranschaulichung.

für kanalisierende und Schwellenwertnetzwerke nur gut übereinstimmen, wenn die Schwellenwertnetzwerke größer sind. Im Folgenden soll deshalb die Abhängigkeit der Evolutionsverläufe von der Netzwerkgröße näher betrachtet werden.

Abbildung 6.12 zeigt die Fitness der zufälligen Netzwerke vor der Evolution für kanalisierende Netzwerke der Größen $N = 50, 80, 100$ und 200 (links) und für Schwellenwertnetzwerke der Größen $N = 50, 80, 100, 200$ und 220 (rechts). Schwellenwertnetzwerke zeigen für gleiche Netzwerkgrößen in der gefrorenen Phase ($\lambda < 1$) eine geringere Fitness als kanalisierende Netzwerke, die sich mit wachsendem N bei Schwellennetzwerken auch stärker erhöht. Hingegen hat in der chaotischen Phase ($\lambda > 1$) die Netzwerkgröße stärkere Auswirkungen auf kanalisierende Netzwerke: Die Fitness ist beträchtlich niedriger für größere Netzwerke. Da das hier angewendete Fitnessmaß auf der Robustheit eines Attraktors gegenüber Störungen beruht, ist es ein direktes Unterscheidungsmerkmal zwischen den dynamischen Phasen der Netzwerke, die sich durch die Ausbreitung von Störungen in den Netzwerken voneinander unterscheiden. Das heißt, für $\lambda < 1$ ist eine hohe Fitness zu erwarten, wogegen sie für $\lambda > 1$ niedrig sein sollte. Für kanalisierende Netzwerke wird dies mit anwachsendem N immer deutlicher, die

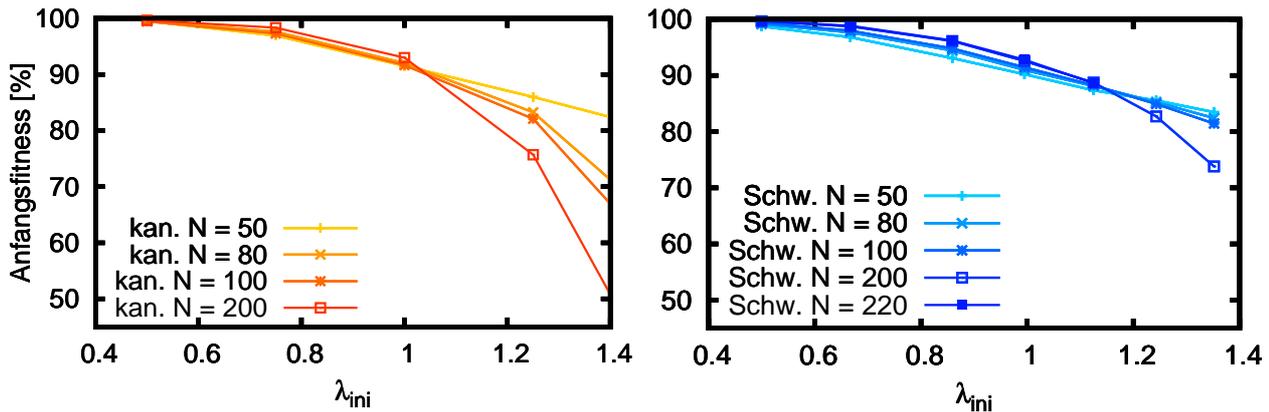


Abbildung 6.12: Die anfängliche Fitness der zufällig erzeugten Netzwerke für kanalisierende (links) und Schwellenwertnetzwerke (rechts) für verschiedene Netzwerkgrößen. Die Linien zwischen den Punkten dienen der Veranschaulichung. Die letzten Punkte für $\lambda = 1,5$ für kanalisierende Netzwerke sind in der Abbildung nicht mehr zu sehen, um in der Darstellung die bestmögliche Vergleichbarkeit zwischen den Netzwerktypen zu haben. Die Linien zeigen die Trends.

Kurven in Abbildung 6.12 rechts schneiden sich in einem Punkt sehr nahe am kritischen Wert $\lambda = 1$. Schwellenwertnetzwerke hingegen nähern sich mit wachsender Größe viel langsamer diesem Verhalten an, die Kurven schneiden sich für die gezeigten Netzwerkgrößen an einem Punkt mit deutlich höherem λ -Wert. Es scheint als würden Schwellenwertnetzwerke erst für höhere N als kanalisierende Netzwerke ein Verhalten zeigen, dass sich gut mit der Annealed Approximation beschreiben lässt.

In Abbildung 6.13 ist die Anzahl der Mutationen sowie der akzeptierten Mutationen zu sehen, die die Netzwerke im Mittel benötigen, um bis zur maximalen Fitness zu gelangen. Für kanalisierende Netzwerke gibt es bei der Anzahl der Mutationen ein lokales Maximum bei $\lambda = 1$, für kritische Netzwerke scheinen mehr Versuche notwendig zu sein, um bis zum Fitnessmaximum zu gelangen als für Netzwerke mit leicht höherem oder leicht niedrigerem λ . Auch bei der Weglänge der Adaptiven Wanderung, d.h. der Anzahl der akzeptierten Mutationen, gibt es ein lokales Maximum bei $\lambda = 1$ für kleine Netzwerkgrößen, das verschwindet, sobald die Netzwerke $N \geq 100$ haben. Dann sind die Wege für Netzwerke aus der gefrorenen Phase am längsten, chaotische Netzwerke sind dem Fitnessmaximum am nächsten. Dividiert man den Wert für akzeptierte Mutationen durch den der insgesamt versuchten Mutationen, ergibt sich der Anteil der akzeptierten Mutationen, der monoton sinkt für anwachsendes λ (s. Abb. 6.11). Für Schwellenwertnetzwerke scheint sich ebenfalls ein lokales Maximum der Anzahl von Mutationen bei $\lambda = 1$ mit hohem N herauszubilden, ist aber selbst bei einer Netzwerkgröße von $N = 220$ nur schwach zu erkennen. Bei der Anzahl der akzeptierten Mutationen ist für Schwellenwertnetzwerke aller hier betrachteten Netzwerkgrößen ein Maximum bei einem λ -Wert zu sehen, der einer anfänglichen Konnektivität von $K = 2$ entspricht. Die Anzahlen der Mutationen steigen bei Schwellenwertnetzwerken unabhängig vom anfänglichen λ -Wert nicht so stark mit der Netzwerkgröße an wie für kanalisierende Netzwerke.

Das Verhalten der Attraktorlängen für verschiedene λ -Werte und anwachsendes N ist für kanalisierende und Schwellenwertnetzwerke ganz ähnlich. Für die kleinsten Netzwerkgrößen gibt es ein Maximum des Verhältnisses von End- zu Anfangslänge der Attraktoren bei einem

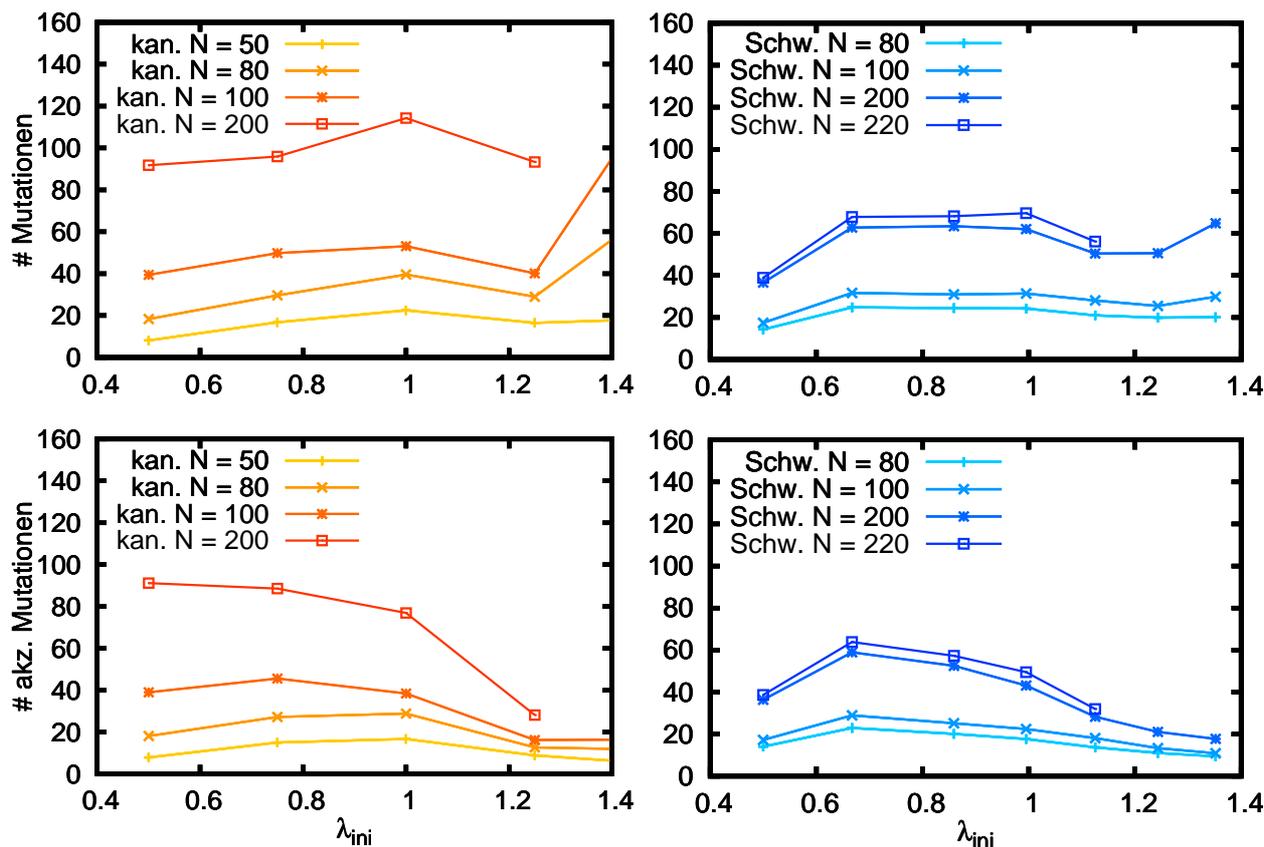


Abbildung 6.13: Anzahl der Mutationen (oben) und der akzeptierten Mutationen (unten) für kanalisierende (links) und Schwellenwertnetzwerke (rechts) und verschiedene Netzwerkgrößen. Die Linien zwischen den Punkten dienen der Veranschaulichung.

Wert $\lambda > 1$ für beide Netzwerktypen (Abb. 6.14), das sich für größere N zu $\lambda = 1$ verschiebt. Bei kanalisierenden Netzwerken passiert diese Verschiebung schon zwischen den Netzwerkgrößen $N = 50$ und $N = 80$, für Schwellenwertnetzwerke erst zwischen den Netzwerkgrößen $N = 100$ und $N = 200$. Für noch größere N wird das Maximum bei $\lambda = 1$ zu einem lokalen Extrempunkt für kanalisierende Netzwerke. Für genügend kleine Netzwerke und $\lambda = 1$ wachsen die Attraktoren während der Adaptiven Wanderung, genauso für die kleinsten Netzwerke mit dem Maximum bei höheren λ . In allen anderen Fällen nimmt die Attraktorlänge während der Evolution ab. Bei der Betrachtung der Attraktorlängen zeigt sich genauso wie bei der Anzahl der Mutationen das hervortretende Verhalten der Netzwerke, die mit dem kritischen Wert $\lambda = 1$ gestartet sind.

6.3.1 Fazit

Wie sich anhand einiger Eigenschaften der Adaptiven Wanderung und der Netzwerke bis zum Erreichen von maximaler Fitness zeigt, unterscheiden sich kanalisierende und Schwellenwertnetzwerke für gleiche Netzwerkgrößen in ihrem Verhalten. Wählt man hingegen größere Schwellenwertnetzwerke zum Vergleich, verhalten sich die beiden Netzwerktypen sehr ähnlich. Schwellenwertnetzwerke scheinen vor der Evolution höhere N als kanalisierende

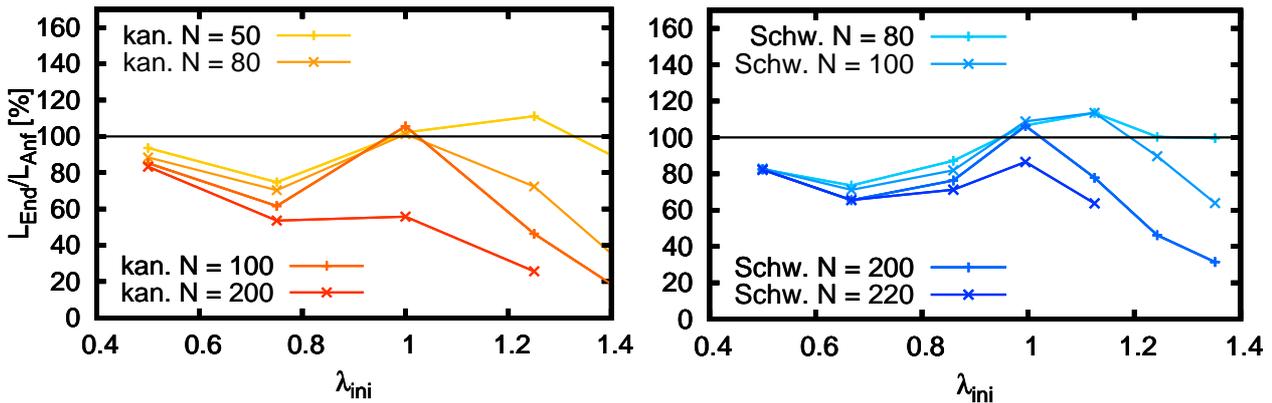


Abbildung 6.14: Das Verhältnis der Attraktorlängen am Anfang der Simulation und nach Erreichen von 100% Fitness für kanalisierende und Schwellenwertnetzwerke verschiedener Größen. Die Linien zwischen den Punkten dienen der Veranschaulichung.

Netzwerke zu benötigen, um ein Verhalten zu zeigen, das sich gut innerhalb der Annealed Approximation verstehen lässt. Während der Evolution sieht man für die beiden Netzwerktypen ähnliche Veränderungen der Eigenschaften mit wachsendem N , für Schwellenwertnetzwerke aber wiederum zu höheren N verschoben. Wie in Kapitel 7 zu sehen, bewirkt ein stärkeres Selektionskriterium, dass die Evolution der beiden Netzwerkarten für gleiche Netzwerkgrößen und relativ kleines N sehr ähnlich ist. Das Langzeitverhalten der Netzwerke für das einfachste Selektionskriterium, beim Fortführen der Adaptiven Wanderung über das Erreichen von 100% Fitness hinaus, wird in einer späteren Publikation mit Christopher Priester und Barbara Drossel behandelt werden.

7 Evolution von kanalisierenden und Schwellenwertnetzwerken unter Selektion für eine stabile Antwort auf externe Stimuli

In den bisherigen Evolutionsmodellen für Boolesche Netzwerke wurde eine wichtige Eigenschaft realer Genregulationsnetzwerke nicht berücksichtigt: Reale Regulationsnetzwerke reagieren auf externe Stimuli. In diesem Kapitel soll daher die Evolution von zwei Arten von Booleschen Netzwerken unter Selektion für eine robuste Antwort auf Änderungen eines externen Eingangs untersucht werden. Abhängig von der betrachteten Anwendung, sind zwei Versionen des Selektionskriteriums relevant: Zellen müssen in der Lage sein, reversibel auf Änderungen in ihrer Umwelt zu reagieren, wie z.B. Änderungen der verfügbaren Nahrungsquellen oder auf temporären Stress. In diesem Fall müssen reversible Umschaltvorgänge zwischen Genexpressionsmustern stattfinden. Auf der anderen Seite durchlaufen Zellen während der Embryonalentwicklung irreversible Umschaltvorgänge: die Zellen ändern ihre Genexpressionsmuster, wenn der Organismus in einen neuen Entwicklungsabschnitt übergeht. Die in diesem Kapitel angewendeten Selektionskriterien verlangen somit das reversible bzw. irreversible Umschalten zwischen zwei Attraktoren. Zusätzlich zur Befolgung von einem dieser Umschaltmuster wird auch verlangt, dass beide Attraktoren stabil sind gegenüber Störungen von einzelnen Knoten (wenn dieser Knoten nicht der externe Eingang ist). In beiden Fällen wird für beide Attraktoren die Zahl von Ein-Knoten-Umschaltvorgängen bestimmt, die wieder auf den Attraktor zurückführen. Dies ist ein Maß für die Robustheit der Attraktoren.

Um festzustellen, ob die erhaltenen Ergebnisse allgemeingültig sind, werden die zwei in Kapitel 2.2 vorgestellten und als für Genregulationsnetzwerke realistisch angesehenen Sätze von Aktualisierungsfunktionen benutzt: kanalisierende und Schwellenwertfunktionen.

Die Evolution eines einzelnen Netzwerkes wird mit dem in Kapitel 5 beschriebenen Algorithmus der Adaptiven Wanderung (unter Verwendung des dritten Fitnesskriteriums) durchgeführt. Die Simulationen starten mit der Erzeugung eines zufälligen Netzwerkes mit K Eingängen pro Knoten und mit Aktualisierungsfunktionen, die entweder aus dem Set der kanalisierenden oder aus dem der Schwellenwertfunktionen zufällig gewählt werden. Einer der Knoten wird zum externen Eingangsknoten bestimmt. Er erhält deshalb keine Eingänge von den anderen Knoten. Wie erwähnt, wird zwischen zwei Fitnesskriterien unterschieden: Beim „Return“-Kriterium muss das Netzwerk nach Zurückschalten des externen Eingangs wieder auf den ersten Attraktor zurückkehren, beim „Stay“-Kriterium muss das Netzwerk auf dem zweiten Attraktor bleiben, auch wenn der externe Eingang wieder seinen Ursprungswert annimmt. Zusätzlich werden noch zwei Fitnessfunktionen unterschieden. In einer ersten Version wird die Fitness, die durch die Robustheit der beiden Attraktoren gegeben ist, um Eins erniedrigt, wenn das „Return“- bzw. das „Stay“-Kriterium nicht erfüllt wird. In der zweiten Version wird die Fitness auf Null gesetzt. Während diese beiden Versionen den selben Netzwerken einen Fitness-Wert von 100% zuordnen, unterscheidet sich der Pfad, den die Netzwerke bis zum Fitnessmaximum nehmen. Für die erste Version tendieren die Netzwerke dazu, zuerst die Robustheit ihrer beiden Attraktoren zu erhöhen, um dann eines der beiden Kriterien zu erfüllen, während die Reihenfolge für die zweite Version der Fitnessfunktion umgekehrt ist.

Für kanalisierende Netzwerke ist die maximale Anzahl an Eingängen auf $K_{\max} = 10$ beschränkt, da die Simulationen bei höheren K zu langsam werden. Für Schwellenwertnetzwerke gilt $K_{\max} = N$. Die minimale Anzahl an Eingängen ist in beiden Fällen $K_{\min} = 1$.

Die Simulationen werden mit Netzwerkgrößen von $N = 50$ durchgeführt. Die Anfangskonnektivitäten reichen für kanalisierenden Netzwerke von $K_{\text{ini}} = 1$ bis 5 und für Schwellennetzwerke von 2 bis 7. Die gewonnenen Daten sind der Vergleichbarkeit wegen wieder über λ_{ini} statt über K_{ini} aufgetragen.

Im nächsten Abschnitt werden die Eigenschaften der evolvierten Netzwerke und der Fitnesslandschaft nach kurzen Evolutionszeiten bis zum Erreichen von 100% Fitness präsentiert und diskutiert. Der darauffolgenden Abschnitt behandelt lange Evolutionszeiten. Anschließend werden einige Netzwerke mit maximaler Fitness im Detail untersucht und beschrieben, wie sie die Selektionskriterien erfüllen. Im letzten Teil dieses Kapitels sollen die Ergebnisse zusammengefasst und evaluiert werden.

7.1 Der Weg zur maximalen Fitness

Jeder Punkt in den Abbildungen in diesem Abschnitt ist gemittelt über mindestens 600 Simulationsläufe. Auf der linken Seite jeder Abbildung sind die Ergebnisse für das „Return“-Kriterium für beide Versionen der Fitnessfunktion und für kanalisierende und Schwellenwertnetzwerke geplottet. Auf der rechten Seite der Abbildungen ist das gleiche für das „Stay“-Kriterium zu sehen.

Zunächst wird die Fitness der zufälligen Anfangsnetzwerke betrachtet (Abb. 7.1). Ausgenommen der Netzwerke, die mit der zweiten Version des „Stay“-Kriteriums evolviert wurden (bei dem die Fitness Null gesetzt wird, wenn das Netzwerk nicht auf dem zweiten Attraktor bleibt, wenn der externe Eingang wieder zurückgeschaltet wird), haben die Anfangsnetzwerke eine mittlere Fitness von mehr als 40% der maximal möglichen Fitness. Das bedeutet, dass die Netzwerke eine beachtliche Wahrscheinlichkeit haben, zum selben Attraktor zurückzukehren, nachdem ein Knoten gestört wurde. Für gefrorene und kritische Netzwerke liegt das an der großen Anzahl von gefrorenen Knoten. Chaotische Netzwerke haben auch einen gefrorenen Kern, dessen Größe jedoch mit anwachsendem λ abnimmt. In früheren Untersuchungen [85], bei denen die dynamische Robustheit der Attraktoren das einzige Selektionskriterium war, nahm die Anfangsfitness mit anwachsendem λ und N ab. Abbildung 7.1 zeigt keine solche Abnahme der Fitness mit ansteigendem λ . Offenbar wird die geringere dynamische Robustheit von Netzwerken mit höherem λ durch die größere Flexibilität kompensiert, die es den Netzwerken erlaubt, einfacher zwischen zwei Attraktoren umzuschalten, wenn der externe Eingang umgeschaltet wird. Die Wahrscheinlichkeit zum ersten Attraktor zurückzukehren, nachdem der Wert des externen Eingangsknotens zurückgeschaltet wurde, ist so hoch, dass die mittlere Anfangsfitness zwischen der ersten und der zweiten Version des „Return“-Kriteriums um weniger als 1/3 abnimmt. Nur für die zweite Version des „Stay“-Kriteriums zeigt sich eine signifikant niedrigere Fitness. Da die Fitness als Funktion der Sensitivität λ aufgetragen wurde, hängt sie nicht vom gewählten Satz von Aktualisierungsfunktionen ab. Für die erste Version der beiden Selektionskriterien wächst die Anfangsfitness mit der anfänglichen Konnektivität bzw. mit λ . Für Netzwerke mit einer chaotischeren Dynamik ist es einfacher die hier eingeführten Fitnesskriterien zu erfüllen, weil sie weniger und dafür längere Attraktoren haben.

Abbildung 7.2 zeigt die Länge der Wege, gemessen in akzeptierten Mutationen, bis die Netzwerke maximale Fitness erreichen. Hier sieht man einen großen Unterschied zwischen den beiden Kriterien. Netzwerke, die mit dem „Stay“-Kriterium evolviert wurden, brauchen viel mehr akzeptierte Mutationen, bis sie maximale Fitness erreichen. Das „Stay“-Kriterium ist schwieriger zu erfüllen als das „Return“-Kriterium. Für beide Kriterien und beide Arten von

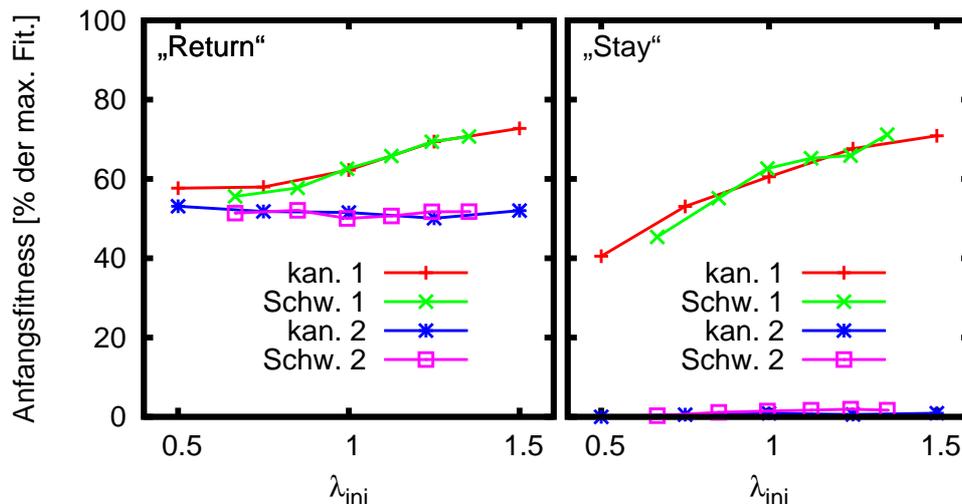


Abbildung 7.1: Fitness der zufällig generierten Anfangsnetzwerke für verschiedene Anfangswerte von K (λ). Abgesehen von der zweiten Version des „Stay“-Kriteriums haben zufällige Anfangsnetzwerke schon eine hohe Fitness.

Aktualisierungsregeln, werden weniger Mutationen benötigt, um das globale Fitnessmaximum zu erreichen, wenn λ_{ini} größer ist, sogar wenn die Netzwerke mit ähnlich hohen Fitnesswerten starten. Chaotischere Netzwerke haben weniger gefrorene Knoten und können mit einer Mutation drastischer verändert werden. Dynamische Robustheit scheint schneller zu evolvieren in Netzwerken mit Schwellenwertfunktionen als in Netzwerken mit kanalisierenden Funktionen, da im ersten Fall die Anzahl an benötigten Mutationen um die maximale Fitness zu erreichen niedriger ist.

Abbildung 7.3 zeigt den Prozentsatz von akzeptierten Mutationen unter den versuchten Mutationen. Für das „Return“-Kriterium (s. Abb 7.3 links) ist der Prozentsatz unabhängig von der benutzten Version des Kriteriums oder von der Art der Aktualisierungsfunktionen. Der Anteil an akzeptierten Mutationen nimmt mit wachsendem λ_{ini} bzw. K_{ini} ab. Die beiden Abbildungen 7.2 links und 7.3 links zusammengenommen lassen erkennen, dass obwohl chaotische Netzwerke näher an der maximalen Fitness sind, mehr versuchte Mutationen benötigt werden, um die erfolgreichen Mutationen zu finden. Für das „Stay“-Kriterium ist die Abnahme des Anteils an akzeptierten Mutationen mit λ nicht so ausgeprägt wie für das „Return“-Kriterium. Für die zweite Version des „Stay“-Kriteriums scheint der Anteil sogar immer bei 20% zu liegen unabhängig von λ_{ini} . Offenbar werden fast alle Mutationen darauf verwendet ein Netzwerk zu finden, das das „Stay“-Kriterium erfüllt. Ist es dann erfüllt, wird sehr schnell hohe Robustheit der Attraktoren erreicht.

Als nächstes wird das Verhältnis der Attraktorlängen am Ende und am Anfang der Adaptiven Wanderung zum Fitnessmaximum betrachtet (Abb. 7.4). Für das „Return“-Kriterium ist die Änderung in der Attraktorlänge gleich für beide Versionen des Kriteriums, für beide Sätze von Aktualisierungsfunktionen und für beide Attraktoren. Attraktoren von Netzwerken, die in der gefrorenen Phase starten, werden etwas kürzer während der Adaptiven Wanderung, Attraktoren von Netzwerken, die mit $\lambda \geq 1$ starten, werden ein wenig länger auf dem Weg zur maximalen Fitness. Da die Anzahl von Mutationen zwischen dem Anfangs- und dem Endnetzwerk gering ist, ist auch die Änderung in den Attraktorlängen klein. Attraktoren von gefrorenen

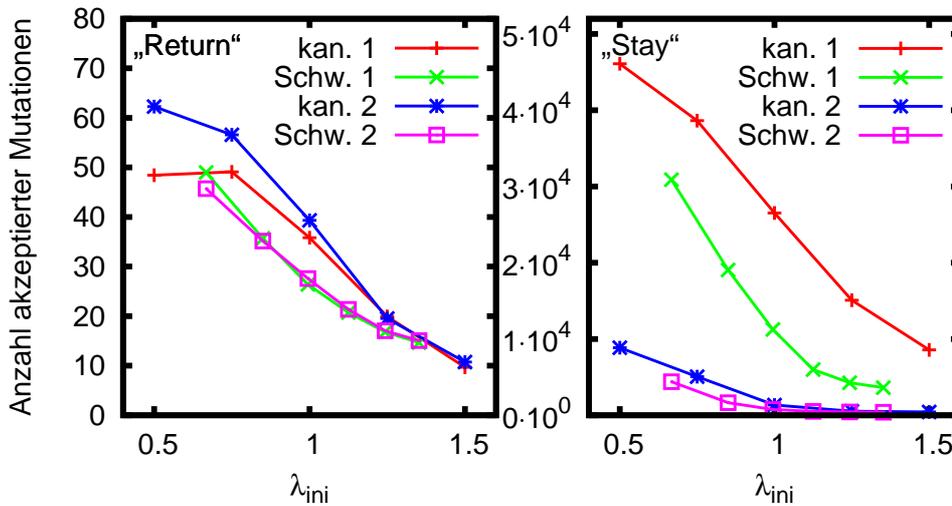


Abbildung 7.2: Anzahl der akzeptierten Mutationen bis die maximale Fitness erreicht ist. Die zwei Versionen des „Stay“-Kriteriums sind viel schwieriger zu erfüllen, viel mehr Mutationen sind nötig um 100% Fitness zu erreichen. Unabhängig von Selektionskriterium oder Art der Aktualisierungsfunktion brauchen Netzwerke, die mit einer höheren Konnektivität starten, weniger Mutationen um 100% Fitness zu erreichen.

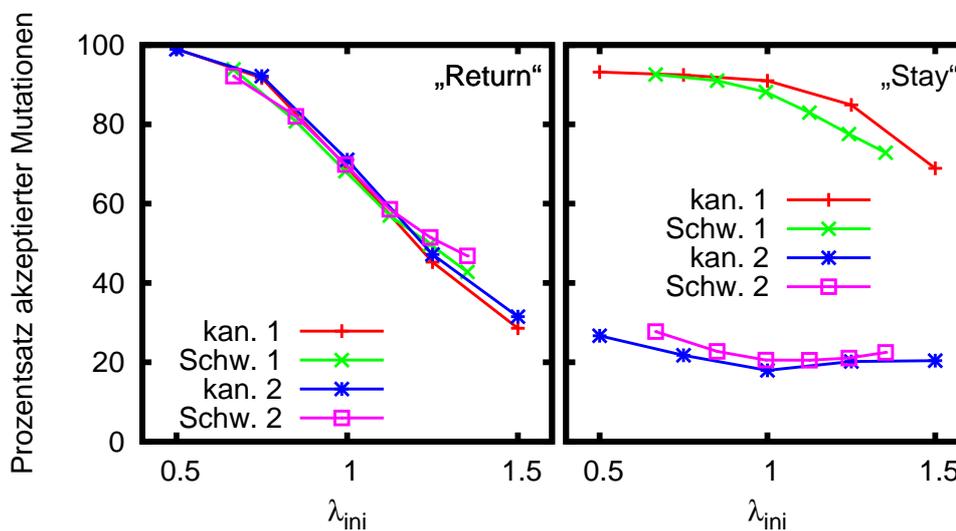


Abbildung 7.3: Prozentsatz der akzeptierten Mutationen bis die maximale Fitness erreicht ist. Der Anteil an akzeptierten Mutationen bis maximale Fitness erreicht ist unterscheidet sich zwischen den beiden Kriterien. Netzwerke mit verschiedenen Arten von Aktualisierungsregeln verhalten sich ähnlich, aber es gibt große Unterschiede zwischen den beiden Versionen des „Stay“-Kriteriums.

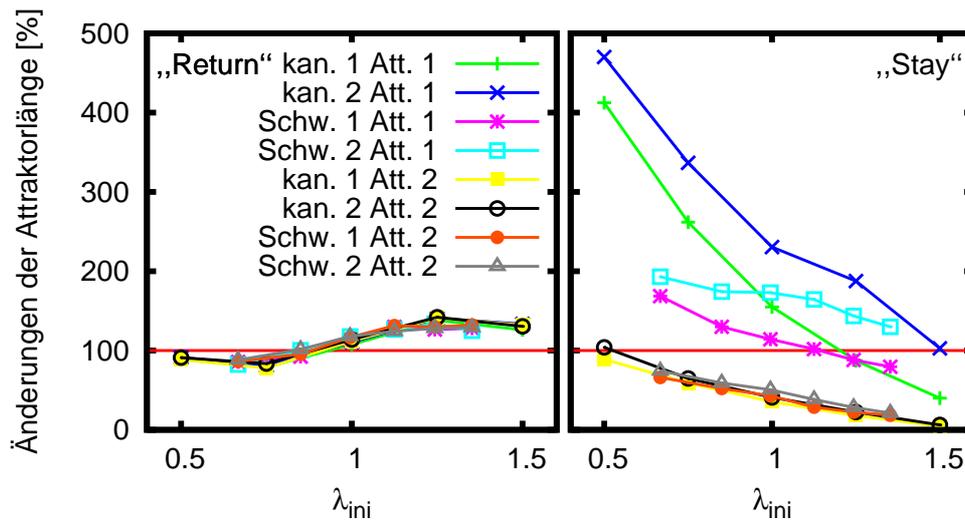


Abbildung 7.4: Änderungen in den Attraktorlängen zwischen Anfangsnetzwerken und 100%-Fitness-Netzwerken. Für das „Return“-Kriterium sind die Änderungen in den Attraktorlängen nicht so ausgeprägt wie für das „Stay“-Kriterium. Für das „Return“-Kriterium ändern sich die Längen der beiden Attraktoren in der gleichen Art. Für die zweite Version des „Stay“-Kriteriums wächst der erste Attraktor immer während der Evolution, für die erste Version des „Stay“-Kriteriums wächst er für kleine λ_{ini} und schrumpft für große λ_{ini} . Der zweite Attraktor schrumpft immer.

Netzwerken können robuster gemacht werden, indem sie kleiner gemacht werden. Dagegen werden Attraktoren von chaotischen Netzwerken robuster, wenn man sie länger macht, da es damit weniger Attraktoren mit größeren Einzugsgebieten gibt.

Für das „Stay“-Kriterium wird der zweite Attraktor immer kleiner auf dem Weg zur maximalen Fitness. Das Kleinerwerden ist ausgeprägter für größere λ_{ini} , weil die Netzwerke mit längeren Attraktoren starten und mit sehr kurzen Attraktoren enden. Es scheint, dass das „Stay“-Kriterium leichter zu erfüllen ist, wenn der zweite Attraktor ein Fixpunkt und das Netzwerk gefroren ist. Im Gegensatz dazu muss der erste Attraktor genügend lang sein, so dass er auf eine Änderung des externen Eingangs reagieren kann. Dieser Effekt ist größer bei kanalisierenden Funktionen. Bei der zweiten Version des „Stay“-Kriteriums wächst der erste Attraktor immer an, unabhängig von den genutzten Aktualisierungsfunktionen (zumindest für die hier betrachteten Werte von λ). Der anfängliche Aufwand, um zunächst ein Netzwerk zu finden, das das „Stay“-Kriterium erfüllt, erhöht die Länge des ersten Attraktors in größerem Maße als wenn das Netzwerk zuerst auf dynamische Robustheit optimiert wird.

In allen Fällen haben die Attraktoren gleiches Gewicht im Zustandsraum, das heißt, ihre Basins haben ähnliche Größen.

7.2 Neutrale Evolution auf dem Plateau maximaler Fitness

Nachdem die Netzwerke maximale Fitness erreicht haben, können sie mit Hilfe von neutralen Mutationen weiter evolvieren. Offenbar gibt es ein riesiges Plateau von maximaler Fitness, das den gesamten Netzwerkraum umspannt. Für jede Variante des Modells wird die Evolution über

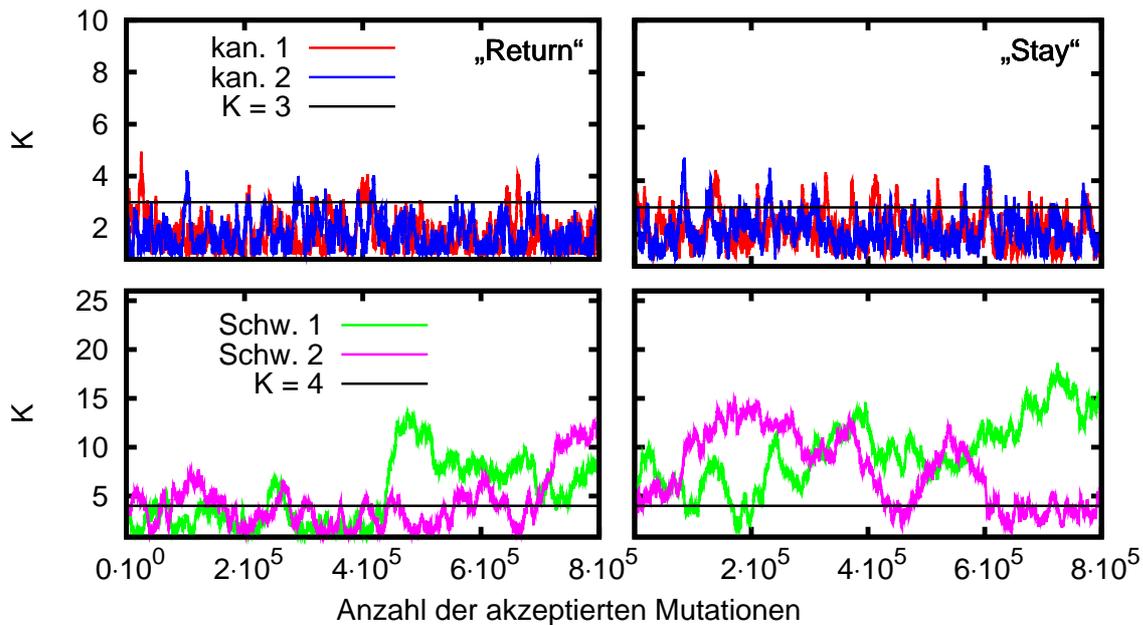


Abbildung 7.5: Mittlere Konnektivität als Funktion der evolutionären Zeit für Netzwerke mit 100% Fitness. Während der Evolution werden viele Netzwerke mit 100% Fitness gefunden, aber die mittlere Konnektivität bleibt gering.

eine lange Zeit weiterverfolgt, bis die Netzwerke 800.000 akzeptierte Mutationen erreichen. Abbildung 7.5 zeigt die Evolution der mittleren Konnektivität K . Die großen Fluktuationen von K deuten darauf hin, dass strukturell sehr unterschiedliche Netzwerke in der Lage sind, die selbe Aufgabe zu erfüllen. Die horizontalen Linien kennzeichnen die K -Werte, über denen $\lambda > 1$ gilt. Sie würden chaotische von gefrorenen Netzwerken trennen, wenn diese zufällige Boolesche Netzwerke wären. Obwohl die evolvierten Netzwerke nicht in dieser Art klassifiziert werden können, gibt es Korrelationen zwischen der Attraktorlänge und der mittleren Konnektivität (s. unten).

Für kanalisierende Netzwerke zeigen die K -Werte für das „Return“- und für das „Stay“-Kriterium gleiches Verhalten. Die meisten Netzwerke mit 100% Fitness haben K -Werte zwischen eins und drei, aber auch Lösungen mit $K > 3$ werden durch den evolutionären Prozess gefunden. Wie zuvor erwähnt, gibt es eine Beschränkung der Eingangsanzahl auf $K_{\max} = 10$, diese ist jedoch nicht von Relevanz, da die Netzwerke nicht in die Nähe dieses Wertes kommen. Genauso kommen die Schwellenwertnetzwerke nicht in die Nähe der maximal möglichen Konnektivität $K_{\max} = 50$. Im Vergleich zu den kanalisierenden Netzwerken zeigen die K -Werte der Schwellenwertnetzwerke über längere Zeiten Korrelationen und sie verbringen mehr Zeit über der horizontalen Linie. Die kanalisierenden Netzwerke scheinen eine stärkere „Kraft“ zu erfahren, die die mittlere Konnektivität senkt, als Schwellenwertnetzwerke, deren K -Wert weniger eingeschränkt ist. Das könnte daran liegen, dass das Hinzufügen einer Verbindung in einem Schwellenwertnetzwerk die Antwort eines Knotens auf seine Eingänge nur wenig komplizierter macht, während die Antwort in einem kanalisierenden Netzwerk viel komplizierter werden kann.

Bei Schwellenwertnetzwerken sieht man einen Unterschied zwischen den beiden Fitnesskriterien. Beim „Stay“-Kriterium scheinen die typischen K -Werte größer zu sein, was bedeutet, dass die oben erwähnte „Kraft“ in diesem Fall schwächer ist.

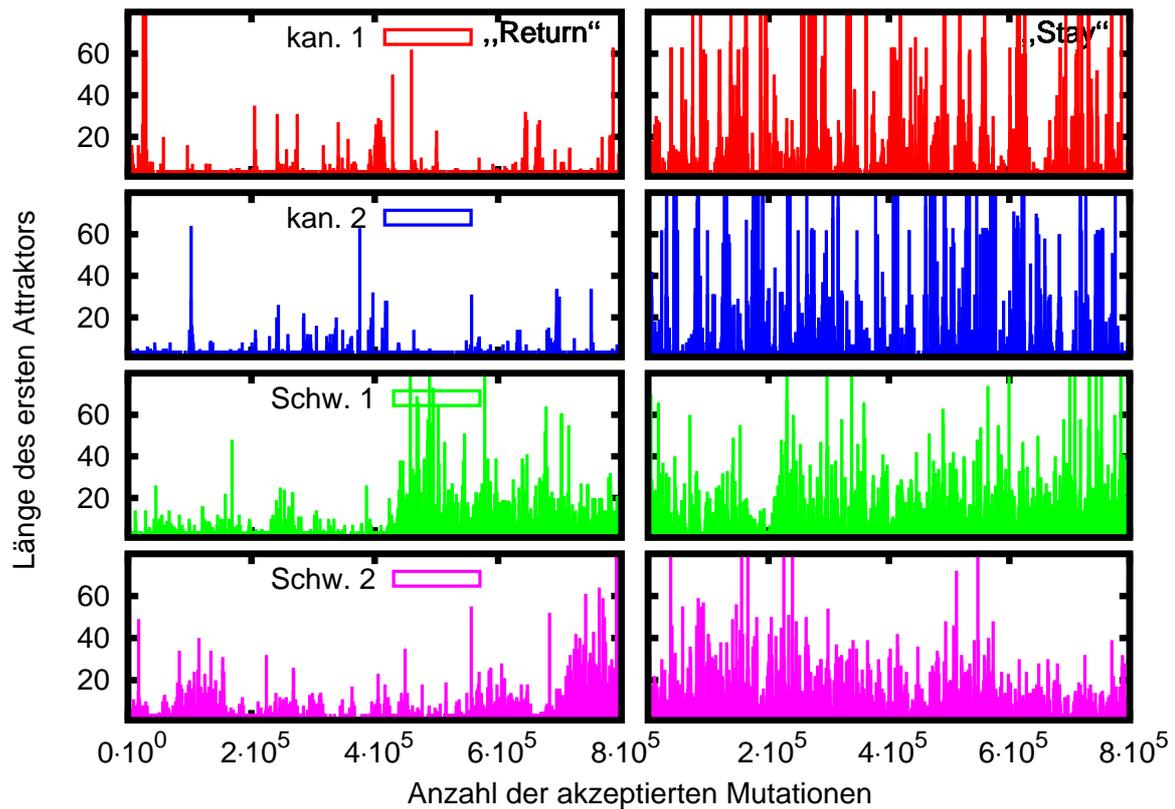


Abbildung 7.6: Länge des ersten Attraktors als Funktion der evolutionären Zeit. Für das „Stay“-Kriterium ist der erste Attraktor länger als für das „Return“-Kriterium.

In Abbildung 7.6 sind die Längen der ersten Attraktoren gegen die evolutionäre Zeit aufgetragen. In allen Fällen produziert das „Stay“-Kriterium längere Attraktoren als das „Return“-Kriterium. Während dies für kanalisierende Netzwerke nicht durch größere K -Werte erklärt werden kann, könnte das für Schwellennetzwerke der dominierende Grund für den Unterschied sein. Die Untersuchung der Mechanismen, mit denen die Netzwerke es bewerkstelligen das „Stay“-Kriterium zu erfüllen, zeigt, dass diese Netzwerke Feedbackschleifen enthalten müssen, die ein Signal von den ersten Knoten, die durch den externen Eingang beeinflusst werden, wieder zurück zu diesen Knoten leiten. Das bedeutet, dass der erste Attraktor weit davon entfernt sein muss gefroren zu sein. Wie in [19] gezeigt, führen Feedbackstrukturen zu komplexerer Dynamik mit längeren Attraktoren. Schwellenwertnetzwerke scheinen in der Lage zu sein, entstehende Feedbackstrukturen im Hinblick auf komplexere Dynamik effizienter abfedern zu können als kanalisierende Netzwerke.

Abbildung 7.7 zeigt den Prozentsatz an negativen Mutationen unter den probierten Mutationen in Schritten von 100 akzeptierten Mutationen. Negative Mutationen sind solche, die die Fitness senken und nicht akzeptiert werden. Die horizontalen Linien in Abbildung 7.7 zeigen die Mittelwerte. Der Anteil an negativen Mutationen ist generell geringer für das „Return“-Kriterium. In diesem Fall kann man keine großen Unterschiede zwischen den beiden Versionen des Kriteriums und zwischen Netzwerken mit unterschiedlichen Arten von Aktualisierungsregeln erkennen. Ein Vergleich mit Abbildung 7.5 zeigt Korrelationen zwischen dem Wert von K und dem Anteil an negativen Mutationen im Falle des „Return“-Kriteriums, aber nicht im Falle des „Stay“-Kriteriums. Das bedeutet, dass mit dem „Return“-Kriterium evolvierte Netz-

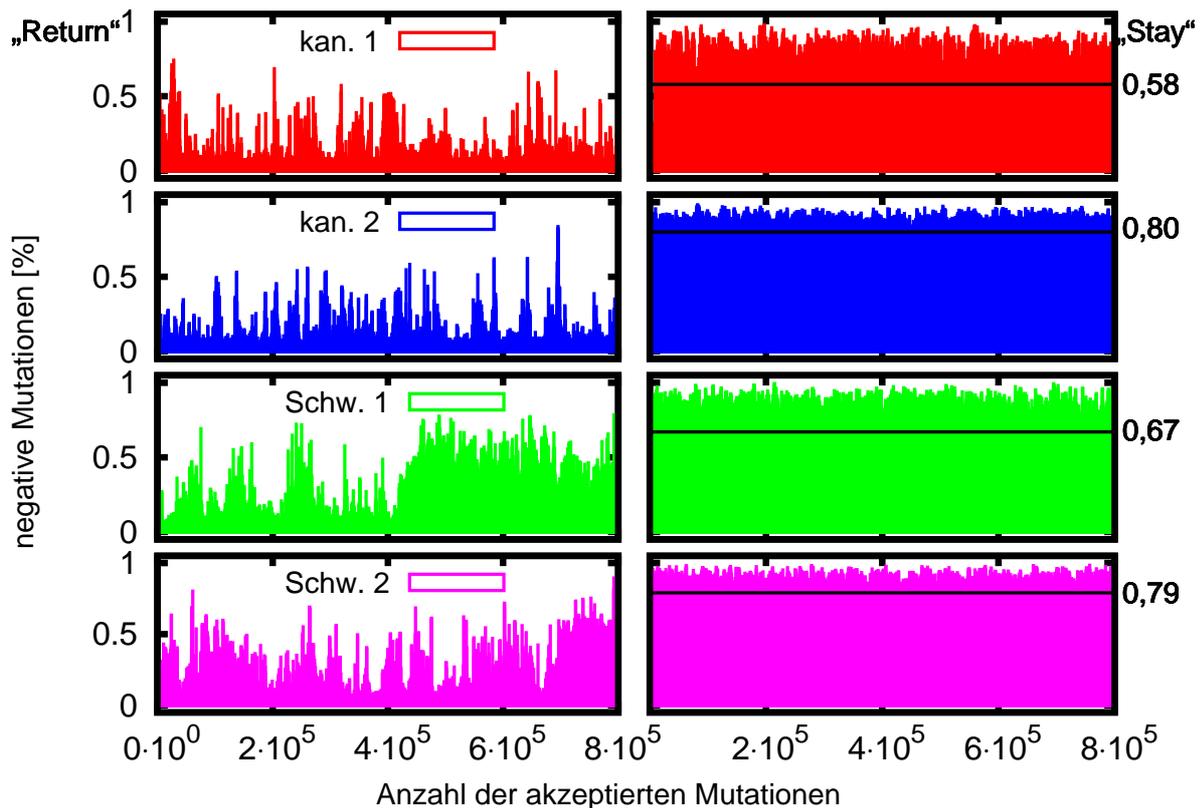


Abbildung 7.7: Anteil an Mutationen, die die Fitness senken, als Funktion der evolutionären Zeit.

werke mit größeren K -Werten (die deswegen wahrscheinlich auch chaotischer sind) sensibler auf Mutationen reagieren. Ihre Robustheit gegenüber Mutationen ist geringer für größere K -Werte. Die Fitnessabnahme kann Attraktoren zugeschrieben werden, die weniger robust gegen Störungen sind. So eine Korrelation zwischen Robustheit gegenüber Rauschen und Robustheit gegenüber Mutationen wurde zuvor schon in evolutionären Modellen von Genregulationsnetzwerken beobachtet [21, 41, 79]. Eine Korrelation zwischen Robustheit gegenüber Rauschen oder Temperatur und Robustheit gegenüber Mutationen kann auch in Micro-RNA gefunden werden [88]. Für das „Stay“-Kriterium fluktuiert der Anteil negativer Mutationen um einen mittleren Wert, der auf der rechten Seite von Abbildung 7.7 gezeigt wird. Die zweite Version des Kriteriums zeigt einen Anstieg in den negativen Mutationen, sowohl für kanalisierende als auch für Schwellenwertnetzwerke. Demnach muss die Fitnessabnahme für Netzwerke, die mit dem „Stay“-Kriterium evolviert wurden, – da sie gleich ist für alle Werte von K – im Unvermögen der Netzwerke auf dem zweiten Attraktor zu bleiben, begründet sein. Erstaunlich ist, dass hier Unterschiede zwischen der ersten und der zweiten Version des Kriteriums zu sehen sind, da die Verschiedenheit der Versionen keine Bedeutung mehr hat, wenn die maximale Fitness einmal erreicht ist. So scheint die anfängliche Evolution zu unterschiedlichen Regionen im Netzwerkraum zu führen, abhängig von der Version des „Stay“-Kriteriums.

Um die „phänotypische“ Ähnlichkeit von zwei benachbarten Netzwerken auf dem Plateau von 100% Fitness abschätzen zu können, wurde der Unterschied zwischen den Attraktoren von zwei aufeinanderfolgenden Netzwerken eines Laufs der Adaptiven Wanderung für 100 Schritte ausgewertet. Der Unterschied zwischen zwei Attraktoren wurde gemessen, indem gezählt wurde, wie viele Knoten auf dem Attraktor in einem Netzwerk gefroren sind, während sie auf dem

Attraktor im anderen Netzwerk blinken, und wie viele von den Knoten, die in beiden Netzwerken gefroren sind, verschiedene Zustände haben. Eine gründliche quantitative Auswertung würde den Rahmen dieser Untersuchung sprengen, jedoch konnten folgende Trends beobachtet werden: Die Ähnlichkeit ist bei Schwellenwertnetzwerken größer als bei kanalisierenden Netzwerken und für Netzwerke, die mit dem „Stay“-Kriterium evolviert wurden, größer als für diejenigen, die mit dem „Return“-Kriterium evolviert wurden. Für Schwellenwertnetzwerke, die mit dem „Stay“-Kriterium evolviert wurden, wird mit wachsendem K eine wachsende Anzahl von Mutationen phänotypisch neutral, wobei sich der zweite Attraktor noch langsamer ändert als der erste.

7.3 Eigenschaften der Netzwerke mit maximaler Fitness

Wie in Abschnitt 7.1 gesehen, ist das „Stay“-Kriterium sehr viel schwieriger zu erfüllen als das „Return“-Kriterium. In Abschnitt 7.2 wurde des Weiteren gezeigt, dass die Erfüllung des „Stay“-Kriteriums auch einfacher wieder zu zerstören ist.

Für jeden der acht untersuchten Fälle (zwei Arten von Aktualisierungsfunktionen, zwei Fitnesskriterien, zwei mögliche Fitnesszuweisungen) wurde das erste Netzwerk, das 100% Fitness erreicht hatte, und acht weitere Netzwerke in Zeitabständen von 100.000 akzeptierten Mutationen ausgewählt, um zu untersuchen, wie genau diese Netzwerke die ihnen auferlegten Aufgaben erfüllen. Die Lösungen, die durch den evolutionären Prozess gefunden wurden, unterscheiden sich klarerweise für die zwei verschiedenen Fitnesskriterien, aber es wurden keine offensichtlichen Unterschiede zwischen den zwei möglichen Versionen der Fitnesszuweisung gesehen. Für kanalisierende und Schwellenwertnetzwerke sind die Lösungen entsprechend der unterschiedlichen Aktualisierungsregeln verwirklicht, ansonsten sind keine qualitativen Unterschiede zwischen den Lösungen für die beiden Sätze von Aktualisierungsfunktionen ersichtlich.

Für das „Stay“-Kriterium propagiert das externe Eingangssignal in der Regel nur durch einen Knoten, der mit dem externen Eingang verbunden ist, unabhängig davon, ob es mehrere solche Knoten gibt. Das liegt daran, dass es ein Feedback im Netzwerk zu den Knoten geben muss, die direkt auf den externen Eingang reagieren, um sie daran zu hindern wieder zu ihrem früheren Verhalten zurückzukehren, wenn der externe Eingang zurückgeschaltet wird. Der externe Eingang selbst darf auf diesen Knoten nicht kanalisierend wirken. Demgegenüber zwingt das „Return“-Kriterium den Netzwerken sehr viel weniger Einschränkungen auf. Das externe Signal kann sich ohne nachteilige Effekte durch mehr als einen Ausgang des externen Knotens ausbreiten und der externe Knoten kann kanalisierend auf diese Knoten wirken oder auch nicht.

Das Umschalten zwischen den beiden Attraktoren passiert generell entweder durch eine Kaskade von Umschaltvorgängen, bei denen gefrorene Knoten ihren Wert ändern, oder dadurch, dass blinkende Knoten eingefroren werden bzw. gefrorene Knoten anfangen zu blinken. Üblicherweise kommt nur eine dieser Möglichkeiten in einem Netzwerk vor. Die Umschaltkaskade tritt auf, wenn beide Attraktoren Fixpunkte sind. Gemischte Formen, wo Umschaltkaskaden und das Einfrieren oder Auftauen von Knoten kombiniert sind, existieren auch, sind aber nur für Schwellenwertnetzwerke häufig, die mit dem „Stay“-Kriterium evolviert wurden. Selten passiert es auch, dass die selben Knoten blinken, aber auf andere Art und Weise.

In Abbildung 7.8 sind die für das Umschalten zwischen den Attraktoren relevanten Netzwerkeile von kanalisierenden Netzwerken zu sehen, die mit dem „Stay“-Kriterium evolviert wurden. Knoten, die als Kreise dargestellt sind, blinken entweder auf einem der Attraktoren

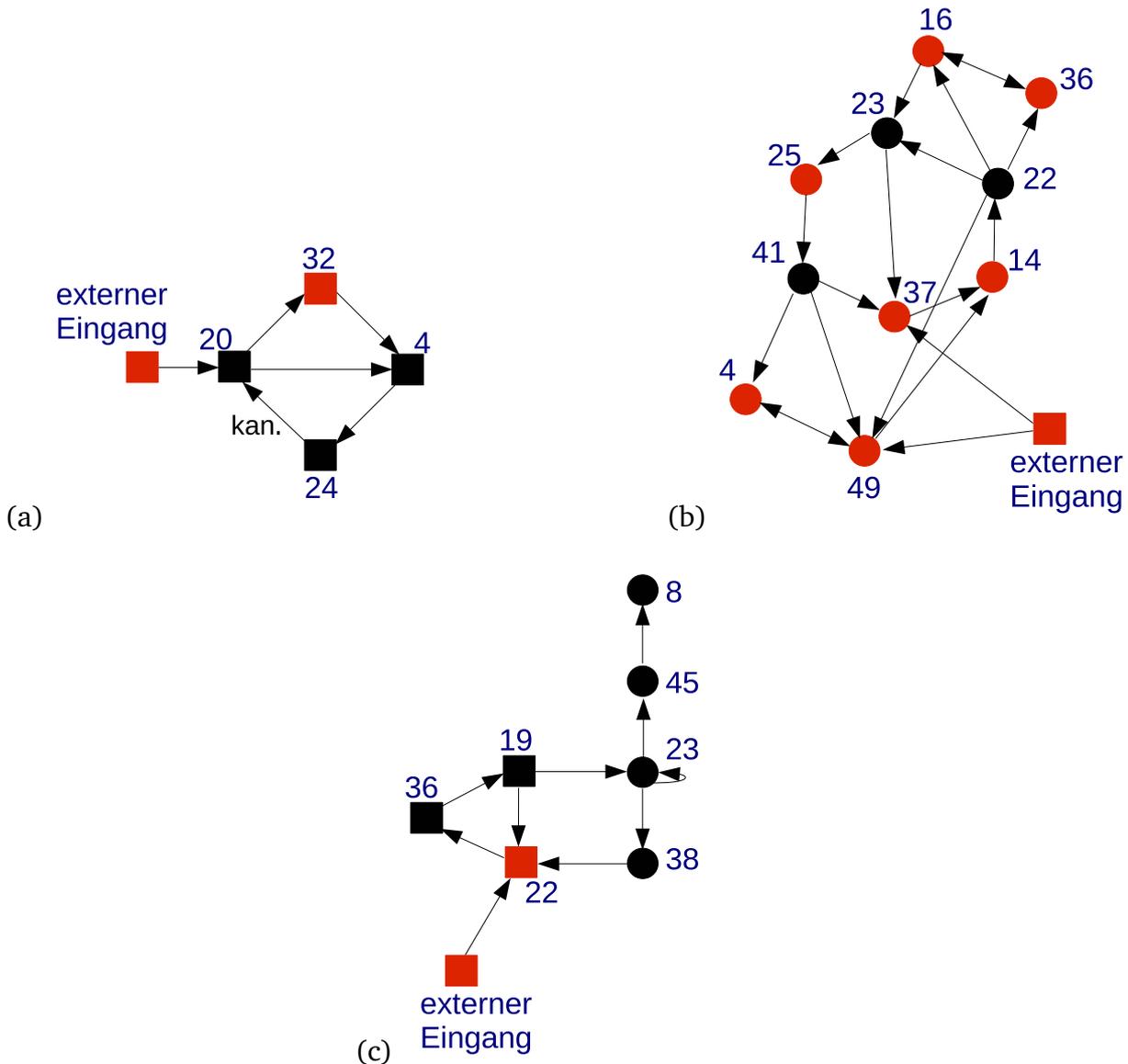


Abbildung 7.8: Relevante Teile von kanalisierenden Netzwerken mit 100% Fitness, die mit dem „Stay“-Kriterium evolviert wurden. Rote Knoten sind eingeschaltet, schwarze ausgeschaltet.

oder auf beiden. Knoten, die als Quadrate dargestellt sind, sind auf beiden Attraktoren gefroren. Im erste Beispiel in Abbildung 7.8 sind beide Attraktoren Fixpunkte. Wenn der externe Eingang eingeschaltet wird, wird zuerst Knoten 20 eingeschaltet, dann Knoten 32 ausgeschaltet, dann Knoten 4 eingeschaltet und schließlich Knoten 24 ausgeschaltet. Knoten 24 kanalisiert Knoten 20 auf den Wert Null, so dass dieser nicht zurückschaltet, wenn der externe Eingang wieder ausgeschaltet wird. Der externe Eingang hat einen zweiten Ausgang zum Netzwerk, der aber nicht relevant für die Attraktoren ist. Das zweite Beispiel ist auch das zweithäufigste in Netzwerken, die mit dem „Stay“-Kriterium evolviert wurden. Wenn nicht beide Attraktoren Fixpunkte sind, ist der erste Attraktor gewöhnlich viel länger als der zweite, der meist ein Fixpunkt ist, zuweilen aber auch die Länge Zwei hat. Das heißt, auf dem ersten Attraktor gibt es blinkende Knoten, die einfrieren, nachdem der externe Eingang umgeschaltet wurde. In Abbildung 7.8 (b) blinken alle gezeigten Knoten außer dem externen Eingang auf dem ers-

ten Attraktor, auf dem zweiten Attraktor sind alle gefroren. Hier breitet sich das Signal durch zwei von vier möglichen Verbindungen zum externen Eingang aus und beide Knoten müssen durch Feedbackschleifen auf ihren gefrorenen Wert fixiert werden. Das dritte Beispiel zeigt den eher unüblichen Fall, bei dem zuerst die drei gefrorenen Knoten 22, 36 und 19 umgeschaltet werden, woraufhin Knoten 23 aufhört zu blinken und mit ihm auch die anderen drei blinkenden Knoten. Diese drei sehr verschiedenen Lösungen des „Stay“-Kriteriums traten während der selben Adaptiven Wanderung auf und sind daher durch neutrale Mutationen miteinander verbunden.

In Abbildung 7.9 sind Beispiele für relevante Teile von Schwellenwertnetzwerken, die mit dem „Stay“-Kriterium evolviert wurden, gezeigt. Wieder stehen Quadrate für Knoten, die auf beiden Attraktoren gefroren sind und Kreise für Knoten, die auf mindestens einem der Attraktoren blinken. Grüne Pfeile stellen aktivierende Verbindungen (+1) dar und rote Linien mit stumpfen Enden inhibierende Verbindungen (-1). In beiden Beispielen sind die blinkenden Knoten auf dem zweiten Attraktor gefroren. Wie in kanalisierenden Netzwerken, lassen sich Feedbackschleifen erkennen, die verhindern, dass das Netzwerk wieder zurück auf den ersten Attraktor läuft. Der deutlichste Unterschied zwischen kanalisierenden und Schwellenwertnetzwerken ist, dass die relevanten Teile der Schwellenwertnetzwerke komplizierter erscheinen als die relevanten Teile von kanalisierenden Netzwerken mit vergleichbarer Attraktorlänge. Der Grund dafür ist, dass die Konnektivität dieser Netzwerke viel größer ist, da der Wert von λ viel langsamer mit der Konnektivität steigt als für kanalisierende Netzwerke. Genauso scheint die Wahrscheinlichkeit Fixpunktattraktoren zu finden in Netzwerken mit großem K für Schwellenwertnetzwerke größer zu sein als für kanalisierende Netzwerke. Da die Konnektivität der Schwellenwertnetzwerke die meiste Zeit über höher ist als die der kanalisierenden Netzwerke,

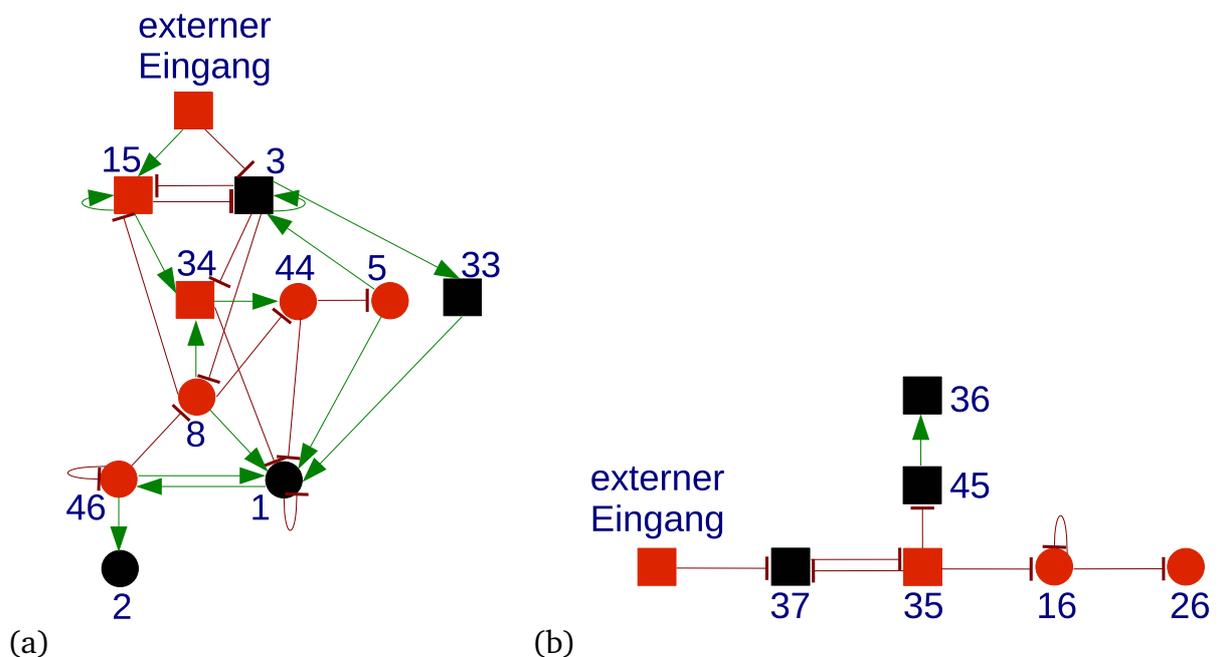


Abbildung 7.9: Relevante Teile von Schwellenwertnetzwerken mit 100% Fitness, die mit dem „Stay“-Kriterium evolviert wurden. Die grünen Pfeile repräsentieren aktivierende Verbindungen, die roten Linien mit stumpfem Ende inhibierende Verbindungen. Rote Knoten sind eingeschaltet, schwarze ausgeschaltet.

ist auch die Anzahl der Ausgänge des externen Eingangs größer in diesen Netzwerken. Das Signal breitet sich aber auch hier nur durch einen oder zwei Ausgänge aus.

7.4 Zusammenfassung

Dieses Kapitel beschäftigt sich mit der Evolution von Booleschen Netzwerken mit zwei verschiedenen Aktualisierungsschemata, unter Selektion für zwei robuste Attraktoren, zwischen denen die Netzwerke als Reaktion auf ein externes Signal reversibel oder irreversibel umschalten.

Kanalisierende und Schwellenwertnetzwerke zeigen ähnliches Verhalten während der Evolution hin zu maximaler Fitness, unterscheiden sich aber zu späteren Zeiten. Es stellte sich heraus, dass zu verlangen, dass die Netzwerke auf dem zweiten Attraktor bleiben, wenn der externe Eingang wieder seinen ursprünglichen Wert annimmt, ein viel strikteres Kriterium darstellt als zu verlangen, dass die Netzwerke wieder zum ersten Attraktor zurückkehren. Es werden mehr Mutationen benötigt, um das erstgenannte Kriterium zu erfüllen und die maximale Fitness kann auch leichter wieder zerstört werden, das heißt, es gibt in diesem Fall mehr Mutationen, die vom Fitnessmaximum wegführen. Es scheint schwierig zu sein, ein Feedback zu erzeugen und aufrechtzuerhalten, das verhindert, dass Knoten wieder zu ihrem früheren Verhalten zurückschalten. Im Gegensatz dazu scheint es einfach zu sein, ein Netzwerk zu erzeugen, das in der Lage ist, robust zwischen zwei verschiedenen Attraktoren hin und her zu schalten. Allerdings erreichten die Netzwerke immer maximale Fitness unabhängig vom Selektionskriterium und konnten außerdem weiter auf dem riesigen Plateau von 100% Fitness evolvieren. Auf diesem Plateau evolvieren Schwellenwertnetzwerke häufiger zu Konnektivitäten, die dem chaotischen Regime zugeordnet werden. Die Netzwerke erreichten sogar noch höhere Konnektivitäten, wenn sie mit dem „Stay“-Kriterium evolviert wurden als wenn sie mit dem „Return“-Kriterium evolviert wurden. Für diese Netzwerke mit hoher Konnektivität sind viele Mutationen auch „phänotypisch neutral“, das heißt, die Attraktoren ändern sich nicht. Betrachtet man die evolvierten Netzwerke im Detail, sieht man, dass strukturell und dynamisch sehr unterschiedliche Netzwerkkomponenten imstande sind, die selbe Aufgabe zu erfüllen.

7.5 Betrachtungen zu neutralen Fitnesslandschaften

Obwohl die hier betrachteten Spielzeugmodelle von Netzwerken viel einfacher sind als reale Genregulationsnetzwerke, können sie dennoch dazu beitragen, etwas über reale evolutionäre Prozesse zu lernen.

- Die Selektion basiert auf dem dynamischem Verhalten, das heißt, auf dem „Phänotyp“ der Netzwerke, während Mutationen auf die Netzwerkstruktur wirken, d.h. also auf den „Genotyp“. Die Fitnesslandschaft, die man auf diesem Weg erhält, ist deutlich verschieden von den rauen Landschaften mit vielen isolierten Fitnessmaxima, die oft ins Auge gefasst werden, wenn Evolution modelliert wird, bei der die Fitnesswerte direkt einem Genotyp zugewiesen werden [46]. In dem hochdimensionalen Genotypraum, der durch Netzwerkmodelle gegeben ist, gibt es viele neutrale Richtungen und perkolierende Cluster von hoher Fitness. Diese Erkenntnis stimmt überein mit den allgemeineren Überlegungen von Gavrillets, der aus ein paar simplen Modellen für Genotyp-auf-Fitness-Abbildungen schloss, dass die Existenz von perkolierenden, annähernd neutralen Netzwerken eine allgemeine Eigenschaft von Fitnesslandschaften mit einer großen Anzahl von Dimensionen

zu sein scheint [35]. Untersuchungen an RNA und Proteinen zeigen, dass deren Fitnesslandschaften diese perkolierenden, neutralen Netzwerke aufweisen (s. Kap. 3.1.1).

- Sehr verschiedene Lösungen mit hoher Fitness werden gefunden, ohne durch Fitnessstärker gehen zu müssen. Neutrale Mutationen sind wichtige Mittel zur Navigation durch den Genotypraum und um neue Möglichkeiten auszuprobieren. In [91] wurde gezeigt, dass Populationen, die auf Fitnesslandschaften mit mehr neutralen Nachbarn pro Knoten evolvieren, weiter ausgebreitet im Genotypraum sind und so eine größere Anzahl von Phänotypen durch nur eine oder zwei Mutationen erreichen können.
- Wenn durch Evolution die Lösung zu einer Aufgabe gefunden werden kann, ist es wahrscheinlich, dass diese nicht nur auf eine Art sondern auf verschiedene Arten gelöst wird. Dieses Verhalten unseres Spielzeugmodells passt mit dem Phänomen der konvergenten Evolution [64] zusammen, das beschreibt, dass sich ähnliche Merkmale bei nicht nahe verwandten Arten unabhängig voneinander entwickeln. Konvergente Evolution wurde bei der Anpassung an ökologische Nischen und bei der Evolution von bestimmten Funktionen wie z.B. die der verschiedenen Sinnesorgane beobachtet. In allen diesen Fällen wurde die gleiche Herausforderung mehrmals unabhängig voneinander bewältigt, indem Lösungen gefunden wurden, die in den allgemeinen Merkmalen übereinstimmen, aber auf dem genetischen Level und in den phänotypischen Details sehr unterschiedlich erscheinen.

Es scheint als ob die Fitnesslandschaften, die man durch evolutionäre Simulationen basierend auf Netzwerkmodellen erhält, viele Eigenschaften von biologischer Evolution reflektieren. Die Simulationen unterstützen die Theorie von Evolution auf ausgedehnten neutralen Netzwerken in der Fitnesslandschaft. Neben Hinweisen auf die Wichtigkeit von neutraler Evolution aus Studien an RNA und Proteinen (s. Kapitel 3.1.1), finden sich auch in anderen Netzwerkmodellen verbundene Cluster von Netzwerken mit der gleichen Fitness in Form des gleichen Phänotyps. In [21] ist der Phänotyp ein bestimmter Gleichgewichtsnetzwerkzustand. Die Nachbarn in diesem Cluster unterscheiden sich durch Mutationen in einer Verbindung voneinander. Weitere Untersuchungen dieses Modells ergaben, dass diese Netzwerke mit dem selben Phänotyp sehr unterschiedliche Genotypen haben können [20]. Entlang eines solchen neutralen Clusters zu evolvieren, erlaubt es den Netzwerken verschiedene neue Phänotypen zu entwickeln und ermöglicht damit Neuerungen. In [29] wird ein Netzwerk ohne Schleifen untersucht. Die Knoten sind dabei in Ebenen angeordnet und ein Knoten kann nur Knoten regulieren, die unter seiner eigenen Ebene liegen. So ergibt sich, dass die oberste Ebene nicht reguliert wird und damit die externen Eingangsknoten repräsentiert. Die unterste Ebene repräsentiert die Ausgangsknoten, die die Reaktion auf den Eingang darstellen. Der „Phänotyp“ des Netzwerkes sind alle Ausgänge, die man als Antwort auf alle möglichen Eingangskombinationen erhält. Eine Mutation besteht aus dem Entfernen einer zufälligen Verbindung mit anschließendem Hinzufügen einer Verbindung zwischen zwei Knoten, die zuvor nicht verbunden waren. Auch hier finden sich neutrale Cluster, die durch den Netzwerkraum perkolieren.



8 Zusammenfassung und Ausblick

Diese Arbeit beschäftigt sich mit der Untersuchung von Booleschen Netzwerken, die als Modelle für Genregulationsnetzwerke verwendet werden. Die Knoten in diesen Netzwerken können nur zwei verschiedene Zustände annehmen, die über Aktualisierungsfunktionen aus den Werten der mit ihnen verbundenen Knoten bestimmt werden. Klassische Boolesche Netzwerke zeichnen sich durch eine zufällige Verknüpfung der Knoten mit konstanter Verteilung der Eingangsverbindungen und eine zufällige Wahl der Booleschen Aktualisierungsfunktionen aus. Abhängig vom Verknüpfungsgrad und der Wahl der Aktualisierungsregeln, können sich Boolesche Netzwerke in drei verschiedenen dynamischen Regimen befinden, die durch qualitativ unterschiedliches Verhalten der Netzwerke gekennzeichnet sind. In der Vergangenheit wurde versucht verschiedene Zellen einem der drei Regime zuzuordnen, doch erscheint es nicht sinnvoll das Konzept von Phasenübergängen in Gleichgewichtssystemen auf Systeme anzuwenden, die weit weg vom Gleichgewicht operieren. Schon für nicht völlig zufällig verknüpfte Netzwerke versagt die einfache Klassifizierung.

In dieser Arbeit wurden die dynamischen Eigenschaften von evolvierten Booleschen Netzwerken untersucht und im Vergleich mit den Eigenschaften von klassischen Booleschen Zufallsnetzwerken gezeigt, dass eine einfache Zuordnung von evolvierten Netzwerken zu einem der dynamischen Regime nicht möglich ist. Am Beispiel eines realen Regulationsnetzwerkes, das mit Hilfe eines Booleschen Netzwerkes modelliert wurde, konnte gezeigt werden, dass eine solche Zuordnung auch für biologische Systeme kritisch zu sehen ist. Trotz allem bietet die Untersuchung der verschiedenen dynamischen Eigenschaften, wie sie für Boolesche Zufallsnetzwerke betrachtet werden, eine gute Möglichkeit, Einblicke in das tatsächliche Verhalten evolvierter (biologischer und nicht-biologischer) Netzwerke zu erhalten.

Die weiteren evolutionären Simulationen dieser Arbeit wurden mit einer Untermenge an Booleschen Zufallsnetzwerken durchgeführt, mit solchen die kanalisierende und solchen die Schwellenwertfunktionen als Aktualisierungsregeln für die Knoten haben. Diese zwei Unterklassen von Booleschen Funktionen haben sich in der Vergangenheit als sinnvolle Annahmen für Betrachtungen im biologischen Kontext herausgestellt. Bevor mit den evolutionären Studien begonnen werden kann, muss zunächst das dynamische Verhalten der zufälligen Netzwerke, die den Ausgangspunkt einer Simulation bilden, bekannt sein. Für die in dieser Arbeit verwendeten Netzwerke mit Schwellenwertfunktionen, wurden diese Untersuchungen bisher nicht durchgeführt und so wurde im vierten Kapitel dieser Arbeit das Phasendiagramm für diese Art von Netzwerken mit Hilfe eines Mean-Field-Ansatzes bestimmt. Es fanden sich einige interessante Besonderheiten dieser Funktionenwahl, die Auswirkungen auf die Interpretation des mit ihr modellierten Zellzyklusregulationsnetzwerkes der Bäckerhefe haben.

Im evolutionären Kontext stellen die Booleschen Netzwerke ein gutes Modell einer Genotyp-Phänotyp-Abbildung dar. Hierbei stellt das Netzwerk selbst mit einem bestimmten Verknüpfungsmuster und einer bestimmten Wahl der Aktualisierungsfunktionen den Genotyp dar. Der Phänotyp als Ausprägung des Genotyps ergibt sich aus den Attraktoren, die ein Netzwerk erreicht. Mutationen, die in der Veränderung von Verknüpfungen oder Funktionen bestehen, wirken sich also auf den Genotyp aus, während Selektion auf den Phänotyp wirkt. Die in dieser Arbeit untersuchten Selektionskriterien verlangen die Robustheit der Attraktoren gegenüber kleinen Störungen sowie eine robuste Antwort der Netzwerke auf externe Stimuli. Unabhän-

gig von der Wahl des Selektionskriteriums oder der Netzwerkart, erreichen alle Netzwerke die maximal mögliche Fitness. Dabei ist das erreichte Maximum in der Fitnesslandschaft kein isolierter Gipfel, sondern ein ausgedehntes Plateau, auf dem sich die Netzwerke weit durch den Raum der möglichen Netzwerke bewegen können. Das heißt, es existieren neutrale Wege in den Fitnesslandschaften der Netzwerke, die sehr verschiedene Netzwerke verbinden, die alle die gleiche, durch das Selektionskriterium vorgegebene, Funktion erfüllen. Diese Erkenntnis knüpft an zahlreiche Studien von RNA- und Proteinmodellen an, in denen ebenfalls neutrale Pfade gefunden werden, die den gesamten Sequenzraum perkolieren, und an einige Studien an Netzwerkmodellen, deren Fitnesslandschaften ähnliche Eigenschaften zeigen. Dieses Ergebnis der Arbeit ist ein weiterer Hinweis darauf, dass die Existenz von weitreichenden neutralen Pfaden eine Eigenschaft vieler evolvierender Systeme zu sein scheint.

Die Untersuchung von komplexeren Netzwerkmodellen und komplexeren Selektionkriterien werden weitere Schritte auf dem Weg sein, biologische Evolution besser zu verstehen.

Da die Genduplikation eine große Rolle bei der Evolution von Genregulationsnetzwerken spielt [89], wäre es interessant, dies bei weitergehenden Untersuchungen zu berücksichtigen. Erwähnenswert hierbei ist, dass es trotz beobachteter Genduplikationen und obwohl viele Netzwerkkomponenten mit der selben Topologie existieren, keinen Hinweis auf die Duplikation von ganzen Netzwerkmotiven gibt. Somit deuten die gleichen Topologien in verschiedenen Netzwerken, auch in verschiedenen Spezies, auf konvergente Evolution hin [52, 22]. In [4] wurde untersucht, wie sich Duplikation und Divergenz eines Gens im Rahmen eines Booleschen Netzwerkmodells auf die Attraktorlandschaft auswirken. In [23] erlaubte das Kopieren und Löschen von Knoten das schnelle Anpassen einer Netzwerkpopulation an wechselnde Umweltbedingungen.

Literaturverzeichnis

- [1] ALBERT, RÉKA: *Scale-free networks in cell biology*. *Journal of Cell Science*, 118:4947–4957, 2005.
- [2] ALBERT, RÉKA und ALBERT-LÁSZLÓ BARABÁSI: *Statistical mechanics of complex networks*. *Review of Modern Physics*, 74:47–97, 2002.
- [3] ALBERT, RÉKA und HANS G. OTHMER: *The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster*. *Journal of Theoretical Biology*, 223:1–18, 2003.
- [4] ALDANA, MAXIMINO, ENRIQUE BALLEZA, STUART KAUFFMAN und OSBALDO RESENDIZ: *Robustness and evolvability in genetic regulatory networks*. *Journal of Theoretical Biology*, 245:433–449, 2007.
- [5] ALDANA, MAXIMINO, SUSAN COPPERSMITH und LEO P. KADANOFF: „*Boolean Dynamics with Random Couplings*“ in *Perspectives and Problems in Nonlinear Science*. Springer, 2003.
- [6] BABAJIDE, ADERONKE, IVO L. HOFACKER, MANFRED J. SIPPL und PETER F. STADLER: *Neutral networks in protein space: a computational study based on knowledge-based potentials of mean force*. *Folding & Design*, 2:261–269, 1997.
- [7] BARABÁSI, ALBERT-LÁSZLÓ und RÉKA ALBERT: *Emergence of Scaling in Random Networks*. *Science*, 286:509–512, 1999.
- [8] BASSLER, KEVIN E., CHOONGSEOP LEE und YONG LEE: *Evolution of Developmental Canalization in Networks of Competing Boolean Nodes*. *Phys. Rev. Lett.*, 93:038101, 2004.
- [9] BASTOLLA, U. und G. PARISI: *Relevant elements, magnetization and dynamical properties in Kauffman networks: A numerical study*. *Physica D*, 115:203, 1998.
- [10] BASTOLLA, UGO, MARKUS PORTO, H. EDUARDO ROMAN und MICHELE VENDRUSCOLO: *Connectivity of neutral networks, overdispersion and structural conservation in protein evolution*. *Journal of Molecular Evolution*, 56:243–254, 2003.
- [11] BASTOLLA, UGO, H. EDUARDO ROMAN und MICHELE VENDRUSCOLO: *Neutral Evolution of Model Proteins: Diffusion in Sequence Space and Overdispersion*. *Journal of Theoretical Biology*, 200:49–64, 1999.
- [12] BOLDHAUS, GUNNAR und KONSTANTIN KLEMM: *Regulatory networks and connected components of the neutral space - A look at functional islands*. Arxiv preprint arXiv:0904.4843.
- [13] BORNHOLDT, STEFAN und THIMO ROHLF: *Topological Evolution of Dynamical Networks: Global Criticality from Local Dynamics*. *Phys. Rev. Lett.*, 84:6114–6117, 2000.
- [14] BORNHOLDT, STEFAN und KIM SNEPPEN: *Neutral Mutations and Punctuated Equilibrium in Evolving Genetic Networks*. *Phys. Rev. Lett.*, 81:236–239, 1998.

-
- [15] BORNHOLDT, STEFAN und KIM SNEPPEN: *Robustness as an evolutionary principle*. Proc. R. Soc. Lond. B, 267:2281–2286, 2000.
- [16] BRAUNEWELL, STEFAN und STEFAN BORNHOLDT: *Superstability of the yeast cell-cycle dynamics: Ensuring causality in the presence of biochemical stochasticity*. Journal of Theoretical Biology, 245:638–643, 2007.
- [17] BRAUNEWELL, STEFAN und STEFAN BORNHOLDT: *Reliability of genetic networks is evolvable*. Phys. Rev. E, 77:060902(R), 2008.
- [18] CALCOTT, BRETT: *Assessing the fitness landscape revolution*. Biology and Philosophy, 23:639–657, 2008.
- [19] CIANDRINI, L., C. MAFFI, A. MOTTA, B. BASSETTI und M. COSENTINO LAGOMARSINO: *Feedback topology and XOR-dynamics in Boolean networks with varying input structure*. Phys. Rev. E, 80:026122, 2009.
- [20] CILIBERTI, S., O.C. MARTIN und A. WAGNER: *Innovation and robustness in complex regulatory gene networks*. PNAS, 104:13591–13596, 2007.
- [21] CILIBERTI, S., O.C. MARTIN und A. WAGNER: *Robustness can evolve gradually in complex regulatory gene networks with varying topology*. PLoS Comput. Biol., 3:164–173, 2007.
- [22] CONANT, GAVIN C. und ANDREAS WAGNER: *Convergent evolution of gene circuits*. Nature Genetics, 34:264–266, 2003.
- [23] CROMBACH, ANTON und PAULIEN HOGWEG: *Evolution of Evolvability in Gene Regulatory Networks*. PLoS Comput. Biol., 4:e10000112, 2008.
- [24] DERRIDA, B und Y POMEAU: *Random Networks of Automata: A Simple Annealed Approximation*. Europhys. Lett., 1:45–49, 1986.
- [25] DROSSEL, BARBARA: *Biological evolution and statistical physics*. Advances in Physics, 50:209–295, 2001.
- [26] DROSSEL, BARBARA: „Random Boolean Networks“ in: *Reviews of Nonlinear Dynamics and Complexity*, Band 1. Wiley, 2008.
- [27] ERDŐS, PAUL und ALFRÉD RÉNYI: *On Random Graphs*. Publicationes Mathematicae, 6:290–297, 1959.
- [28] ERDŐS, PAUL und ALFRÉD RÉNYI: *On the evolution of random graphs*. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5:17–61, 1960.
- [29] FERNÁNDEZ, PAU und RICARD V. SOLÉ: *Neutral fitness landscapes in signalling networks*. J. R. Soc. Interface, 4:41–47, 2007.
- [30] FESCHOTTE, CÉDRIC: *Transposable elements and the evolution of regulatory networks*. Nature Reviews Genetics, 9:397, 2008.
- [31] FLYVBJERG, HENRIK und N. J. KJÆR: *Exact solution of Kauffman’s model with connectivity one*. J. Phys. A, 21:1695, 1988.

-
- [32] FONTANA, W.: *Modelling 'evo-devo' with RNA*. *BioEssays*, 24:1164–1177, 2002.
- [33] FRETTER, CHRISTOPH und BARBARA DROSSEL: *Response of boolean networks to perturbations*. *Eur. Phys. J. B*, 62:365–371, 2008.
- [34] FRETTER, CHRISTOPH, AGNES SZEJKA und BARBARA DROSSEL: *Perturbation propagation in random and evolved Boolean networks*. *New J. Phys.*, 11:033005, 2009.
- [35] GAVRILETS, SERGEY: „*Evolution and Speciation in a Hyperspace: The Roles of Neutrality, Selection, Mutation, and Random Drift*“ in: *Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function*. Oxford University Press, New York, 2003.
- [36] GAVRILETS, SERGEY: „*Fitness landscapes*“ in: *Encyclopedia of Ecology*. Elsevier, 2008.
- [37] GREIL, FLORIAN und BARBARA DROSSEL: *Kauffman networks with threshold functions*. *Eur. Phys. J. B*, 57:109–113, 2007.
- [38] HAMER, CHRISTOPH: *Boolesche Genregulationsnetzwerke und Evolution*. Diplomarbeit, Technische Universität Darmstadt, AG Drossel, 2008.
- [39] HARRIS, STEPHEN E., BRUCE K. SAWHILL, ANDREW WUENSCH und STUART KAUFFMAN: *A Model of Transcriptional Regulatory Networks Based on Biases in the Observed Regulation Rules*. *Complexity*, 7:23–40, 2002.
- [40] JÖRG, T., O. C. MARTIN und A. WAGNER: *Neutral network sizes of biological RNA molecules can be computed and are not atypically small*. *BMC Bioinformatics*, 9:464, 2008.
- [41] KANEKO, KUNIHICO: *Evolution of Robustness to Noise and Mutation in Gene Expression Dynamics*. *PLoS ONE*, 2:e434, 2007.
- [42] KAPLAN, JONATHAN: *The end of the adaptive landscape metaphor?* *Biology and Philosophy*, 23:625–638, 2008.
- [43] KAUFFMAN, STUART, CARSTEN PETERSON, BJÖRN SAMUELSSON und CARL TROEIN: *Genetic networks with canalizing Boolean rules are always stable*. *PNAS*, 101:17102–17107, 2004.
- [44] KAUFFMAN, STUART A.: *Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets*. *Journal of Theoretical Biology*, 22:437–467, 1969.
- [45] KAUFFMAN, STUART A.: *The Origin of Order*. Oxford University Press, 1993.
- [46] KAUFFMAN, STUART A. und SIMON LEVIN: *Towards a General Theory of Adaptive Walks on Rugged Landscapes*. *Journal of Theoretical Biology*, 128:11–45, 1987.
- [47] KAUFFMAN, STUART A. und ROBERT G. SMITH: *Adaptive automata based on Darwinian selection*. *Physica D*, 22:68–82, 1986.
- [48] KAUFMAN, VIKTOR, TAMARA MIHALJEV und BARBARA DROSSEL: *Scaling in critical random Boolean networks*. *Phys. Rev. E*, 72:046124, 2005.
- [49] KESSELI, JUHA, PAULI RÄMÖ und OLLI YLI-HARJA: *Iterated maps for annealed Boolean networks*. *Phys. Rev. E*, 74:046104, 2006.
- [50] KIMURA, MOTOO: *Evolutionary Rate at the Molecular Level*. *Nature*, 217:624–626, 1968.

-
- [51] KIMURA, MOTOO: *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983.
- [52] KITANO, HIROAKI: *Biological Robustness*. Nature Reviews Genetics, 5:826–837, 2004.
- [53] KÜRTEEN, KARL E.: *Critical phenomena in model neural networks*. Phys. Lett. A, 129:157–160, 1988.
- [54] KUTSCHERA, ULRICH und KARL J. NIKLAS: *The modern theory of biological evolution: an expanded synthesis*. Naturwissenschaften, 91:255–276, 2004.
- [55] LAU, KAI-YEUNG, SURYA GANGULI und CHAO TANG: *Function constrains network architecture and dynamics: A case study on the yeast cell cycle Boolean network*. Phys. Rev. E, 75:051907, 2007.
- [56] LEWIS, TED G.: *Network Science: Theory and Applications*. Wiley, 2009.
- [57] LI, FANGTING, TAO LONG, YING LU, QI OUYANG und CHAO TANG: *The yeast cell-cycle network is robustly designed*. Proc. Natl. Acad. Sci. USA, 101:4781–4786, 2004.
- [58] LIU, MIN und KEVIN E. BASSLER: *Emergent criticality from coevolution in random Boolean networks*. Phys. Rev. E, 74:041910, 2006.
- [59] LODISH, HARVEY F., ARNOLD BERK, PAUL MATSUDAIRA, CHRIS A. KAISER, MONTY KRIEGER, MATTHEW P. SCOTT, S. LAWRENCE ZIPURSKY und JAMES DARNELL: *Molecular Cell Biology*. W. H. Freeman, 2004.
- [60] LUQUE, BARTOLO, FERNANDO J. BALLESTEROS und ENRIQUE M. MURO: *Self-organized critical random Boolean networks*. Phys. Rev. E, 63:051913, 2001.
- [61] LUQUE, BARTOLO und RICARD V. SOLE: *Phase transitions in random networks: Simple analytic determination of critical points*. Phys. Rev. E, 55:257, 1997.
- [62] MIHALJEV, TAMARA und BARBARA DROSSEL: *Scaling in a general class of critical random Boolean networks*. Phys. Rev. E, 74:046101, 2006.
- [63] MOREIRA, ANDRÉ AUTO und LUÍS A. NUNES AMARAL: *Canalizing Kauffman Networks: Nonergodicity and Its Effect on Their Critical Behavior*. Phys. Rev. Lett, 94:218702, 2005.
- [64] MORRIS, SIMON CONWAY: *Life's Solution: Inevitable humans in a Lonely Universe*. Cambridge University Press, Cambridge, 2003.
- [65] NYKTER, MATTI, NATHAN D. PRICE, MAXIMINO ALDANA, STEPHEN A. RAMSEY, STUART A. KAUFFMAN, LEROY E HOOD, OLLI YLI-HARJA und ILJA SHMULEVICH: *Gene expression dynamics in the macrophage exhibit criticality*. PNAS, 105:1897–1900, 2008.
- [66] PACZUSKI, MAYA, KEVIN E. BASSLER und ÁVARO CORRAL: *Self-Organized Networks of Competing Boolean Agents*. Phys. Rev. Lett, 84:3185–3188, 2000.
- [67] PIGLIUCCI, MASSIMO: *Sewall Wright's adaptive landscapes: 1932 vs. 1988*. Biology and Philosophy, 23:591–630, 2008.
- [68] PIGLIUCCI, MASSIMO und JONATHAN MICHAEL KAPLAN: *Making Sense of Evolution: The Conceptual Foundations of Evolutionary Biology*. University of Chicago Press, Chicago, 2006.

-
- [69] PLUTYNSKI, ANYA: *The rise and fall of the adaptive landscape?* *Biology and Philosophy*, 23:605–623, 2008.
- [70] RÄMÖ, P., J. KESSELI und O. YLI-HARJA: *Perturbation avalanches and criticality in gene regulatory networks*. *Journal of Theoretical Biology*, 242:164–170, 2006.
- [71] REIL, TORSTEN: „*Dynamics of Gene Expression in an Artificial Genome - Implications for Biological and Artificial Ontogeny*“ in: *Advances in Artificial Life: Proceedings of the 5th European Conference (ECAL '99)*. Springer, 1999.
- [72] RIDLEY, MARK: *Evolution*. Blackwell Science, 1996.
- [73] ROHLE, THIMO: *Critical line in random-threshold networks with inhomogeneous thresholds*. *Phys. Rev. E*, 78:066118, 2008.
- [74] ROHLE, THIMO und STEFAN BORNHOLDT: *Criticality in random threshold networks: annealed approximation and beyond*. *Physica A*, 310:245–259, 2002.
- [75] SAMUELSSON, BJÖRN und CARL TROIN: *Superpolynomial Growth in the Number of Attractors in Kauffman Networks*. *Phys. Rev. Lett.*, 90:098701, 2003.
- [76] SCHULTES, ERIK A. und DAVID P. BARTEL: *One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds*. *Science*, 289:448–452, 2000.
- [77] SCHUSTER, PETER, WALTER FONTANA, PETER F. STADLER und IVO L. HOFACKER: *From Sequences to Shapes and Back: A Case Study in RNA Secondary Structures*. *Proc. R. Soc. Lond. B*, 255:279–284, 1994.
- [78] SERRA, R., M. VILLANI, A. GRAUDENZI und S.A. KAUFFMAN: *Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data*. *Journal of Theoretical Biology*, 246:449–460, 2007.
- [79] SEVIM, VOLKAN und PER ARNE RIKVOLD: *Chaotic gene regulatory networks can be robust against mutations and noise*. *J. Theor. Biol.*, 253:323–332, 2008.
- [80] SHMULEVICH, ILJA, STUART A. KAUFFMAN und MAXIMINO ALDANA: *Eukaryotic cells are dynamically ordered or critical but not chaotic*. *PNAS*, 102:13439–13444, 2005.
- [81] SHMULEVICH, ILYA und STUART A. KAUFFMAN: *Activities and Sensitivities in Boolean Network Models*. *Phys. Rev. Lett.*, 93:048701, 2004.
- [82] STELLING, JÖRG, UWE SAUER, ZOLTAN SZALLASI, FRANCIS J. DOYLE III und JOHN DOYLE: *Robustness of Cellular Functions*. *Cell*, 118:675–685, 2004.
- [83] STROGATZ, STEVEN H.: *SYNC: The Emerging Science of Spontaneous Order*. Hyperion, 2003.
- [84] SZEJKA, AGNES: *Evolution kanalisierender Boolescher Netzwerke*. Diplomarbeit, Technische Universität Darmstadt, AG Drossel, 2006.
- [85] SZEJKA, AGNES und BARBARA DROSSEL: *Evolution of canalizing Boolean networks*. *Eur. Phys. J. B*, 56:373–380, 2007.

-
- [86] SZEJKA, AGNES und BARBARA DROSSEL: *Evolution of Boolean networks under selection for a robust response to external inputs yields an extensive neutral space.* Phys. Rev. E, 81:021908, 2010.
- [87] SZEJKA, AGNES, TAMARA MIHALJEV und BARBARA DROSSEL: *The phase diagram of random threshold networks.* New J. Phys., 10:063009, 2008.
- [88] SZÖLLŐSI, G. J. und I. DRÉNYI: *Congruent evolution of genetic and environmental robustness in micro-RNA.* Mol. Biol. Evol., 26:867–874, 2009.
- [89] TEICHMANN, SARAH A. und M. MADAN BABU: *Gene regulatory network growth by duplication.* Nature Genetics, 36:492–496, 2004.
- [90] WAGNER, ANDREAS: *Robustness, evolvability, and neutrality.* FEBS Letters, 579:1772–1778, 2005.
- [91] WAGNER, ANDREAS: *Robustness and evolvability: a paradox resolved.* Proc. R. Soc. B, 275:91–100, 2008.
- [92] WATSON, JAMES, NICHOLAS GEARD und JANET WILES: *Towards more biological mutation operators in gene regulation studies.* BioSystems, 76:239–248, 2004.
- [93] WATTS, DUNCAN J.: *Small Worlds: The Dynamics of Networks Between Order and Randomness.* Princeton University Press, 1999.
- [94] WILLADSEN, KAI und JANET WILES: *Robustness and state-space structure of Boolean gene regulatory models.* Journal of Theoretical Biology, 249:749–765, 2007.
- [95] WRIGHT, SEWALL: „*The roles of mutation, inbreeding, crossbreeding, and selection in evolution*“ in: *Proceedings of the Sixth International Congress on Genetics*, Band 1. Genetics Society of America, 1932.
- [96] ZHANG, YUPING, MINPING QIAN, QI OUYANG, MINGHUA DENG, FANGTING LI und CAO TANG: *Stochastic model of yeast cell-cycle network.* Physica D, 219:35–39, 2006.

Danksagung

An erster Stelle möchte ich Barbara Drossel danken, ganz besonders für die stete Unterstützung und Ermutigung während meiner Doktorarbeitsphase, für den Aufbau und Erhalt einer tollen Arbeitsgruppe und die Möglichkeit an zahlreichen Konferenzen und Sommerschulen teilzunehmen.

Der gesamten AG Drossel danke ich für die gemeinsamen Aktivitäten und die „Klassenausflüge“,

Tamara Mihaljev und Marianne Heckmann insbesondere für's Shoppen und Tratschen,

Tamara für ~~alte~~ Doktoranden jungdoktoranden und großschwesterliche Ratschläge,

Mari für Piratenzeiten und Ponys,

Florian Greil für's Quatschmachen und Ernstreden und für all die neuen Wörter,

Tiago de Paula Peixoto, Daniel Löb, Eva Gehrmann, Christian Guill, Manuel Bach und Yixian Song für alles was sciencefiction und pony war (und das war Einiges),

Christopher Priester für die Unterstützung bei den Simulationen am Ende.

Für das Korrekturlesen meiner Doktorarbeit möchte ich mich bei Florian Greil, Ulrich Seyfarth, Dimitar Ho, Christopher Priester, Eva Gehrmann, Daniel Lentz und Christiane Schabarum bedanken.

Für die seelische und moralische Unterstützung in der Endphase gilt ein ganz besonderer Dank meiner kleinen aber erlesenen „Sonntagsarbeitsgruppe“, mit Dimitar Ho und Sarah Böhm.

Dimitar danke ich für den Teil an Gelassenheit, der mir gefehlt hat, für die gemeinsame Nachtarbeit und die unfreiwilligen Auszeiten.

Für seelische, moralische und sonstige Unterstützung zu jeder Zeit einen riesigen Dank an meine Eltern, Doris Unger, Daniel Lentz und Christiane Schabarum.



Lebenslauf

Persönliche Daten

Name Agnes Szejka
Geburtsdaten 30.10.1981, Złotów (Polen)

Ausbildung

1992–1998 Gesamtschule, gymnasialer Zweig, *Ernst-Reuter-Schule, Dietzenbach*
1998–2001 Gymnasiale Oberstufe, *Oswald-von-Nell-Breuning-Schule, Rödermark*
Abschluss: Abitur, Note: 1,5
Okt. 2001–Okt. 2006 Studium der Physik, *Technische Universität Darmstadt*
Diplomarbeit: Evolution kanalisierender Boolescher Netzwerke
Abschluss: Dipl.-Physikerin, Note: sehr gut
Nov. 2006–Feb. 2010 Promotion im Fachbereich Physik, *Technische Universität Darmstadt, AG Prof. Dr. Barbara Drossel*

Publikationen

1. AGNES SZEJKA UND BARBARA DROSSEL
Evolution of canalizing Boolean networks
Eur. Phys. J. B, **56**, 373–380 (2007)
2. AGNES SZEJKA, TAMARA MIHALJEV UND BARBARA DROSSEL
The phase diagram of random threshold networks
New J. Phys. **10**, 063009 (2008)
3. CHRISTOPH FRETTER, AGNES SZEJKA UND BARBARA DROSSEL
Perturbation propagation in random and evolved Boolean networks
New J. Phys. **11**, 033005 (2009)
4. AGNES SZEJKA UND BARBARA DROSSEL
Evolution of Boolean networks under selection for a robust response to external inputs yields an extensive neutral space
Phys. Rev. E **81**, 021908 (2010)

-
5. SEBASTIAN ZAHNREICH, PETRA BOUKAMP, DAMIR KRUNIĆ, LARISSA MELNIKOVA, AGNES SZEJKA, BARBARA DROSSEL, MARCO DURANTE, SYLVIA RITTER AND CLAUDIA FOURNIER
Rearrangements between shortened telomeres and amplified chromosome fragments in human fibroblasts entering senescence
(in Vorbereitung)
6. AGNES SZEJKA, CHRISTOPHER PRIESTER UND BARBARA DROSSEL
Comparing the evolution of canalizing and random threshold networks
(in Vorbereitung)

Erklärung zur Dissertation

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst, keine anderen als die angegebenen Hilfsmittel verwendet und bisher noch keinen Promotionsversuch unternommen habe.

(A. Szejka)

