

Mechanisms for Emergence and Self-Organisation in Complex Adaptive Systems

A Network-Theoretical Perspective

PhD Thesis by
Gregory Paperin

Copyright Notices

Notice 1

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

Notice 2

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

A central question in complexity theory is how large-scale phenomena, such as self-organisation, perpetual novelty, and sustained diversity, emerge. Complex systems can be understood as networks of interacting components. The focus of this research is the role that the properties of such networks play in self-organisation and emergence in complex systems.

Based on the previously known concept of Dual Phase Evolution (DPE), I propose a theoretical framework, within which recurrent phase transitions in network connectivity underlie emergent phenomena in many systems. This DPE framework extends and refines the original concept.

Networks can exist in two general connectivity phases: well connected and poorly connected. DPE relates each of the two connectivity phases and the transition events between them to typical system dynamics. I analyse empirical and experimental evidence from published studies in areas as diverse as physics, biology, socio-economics, mathematics and computer science. The analysis implies that DPE is widespread and operates in many kinds of complex systems, where it drives emergence and self-organisation. What is more, the analysis uncovers hitherto unstudied deep similarities and common underlying processes between different complex systems.

To further understand the theoretical concepts of the DPE framework, I apply DPE in studies of mechanisms behind particular emergent properties in several types of complex systems:

Seeking to better understand the emergence of novelty and diversity in ecosystems, I develop and study an individual-based simulation model of adaptive radiation (speciation) in landscapes. Simulation results imply that recurrent external disturbances facilitate perpetual novelty and diversity in landscape populations through two complementary mechanisms: One mechanism constitutes recurrent DPE phase changes in landscape connectivity on several levels. The other mechanism is alteration of the environment in disturbed areas leading to modified selection regimes.

As a result of the simulation studies of landscape evolution, I develop a new genetic model that combines the advantages of two existing genetic models. The new model allows individual-based simulation studies of genetics on holey fitness landscapes (HFLs). Such fitness landscapes result from biochemical constraints to genetic viability and have previously only been studied analytically. Simulation studies of reproductive isolation uncover that when HFLs are considered, common predictions about maintenance of reproductive isolation in migrating populations change. Results also show that HFL-genetics can facilitate the emergence of stable hybrid populations, and the evolution of social selection through reinforcement.

Continuing to study and apply DPE, I investigate how DPE processes can lead to the emergence of important network topologies. Using simulation models, I demonstrate two possible mechanisms behind emergent connectivity phase transitions without facilitation by external stimuli. A study of social network models reveals simple mechanisms that lead to structures typical of some real social networks and points towards general principles for emergence of important topologies such as modularity. A study of a network model of co-operations in markets reveals further mechanisms behind the emergence of complex and hierarchical modularity.

Generative models for scale-free networks, that are ubiquitous in many natural systems, are well known, however, such models apply to growing networks. I propose and examine a generative model for scale-free topologies that can account for some scale-free networks of constant size found in nature.

A wider context for DPE as a framework for reasoning about complexity is provided by examining the relationship between DPE and other established concepts such as Self-Organised Criticality and the Adaptive Cycle. In conclusion, DPE complements other established theories.

In general, network-theoretical approaches, such as DPE, are powerful paradigms in understanding complexity. This thesis shows that recurrent changes in connectivity of component interaction networks constitute a broad mechanism for emergence and self-organisation in complex systems, and demonstrates this mechanism in several specific biological and socio-economic systems.

Acknowledgements

A PhD degree is a long and, at times, lonely journey, but it is also rewarding. I would not have been able to undertake this journey without the help and support of my academic advisers.

I am grateful to Prof. David Geoffrey Green for his wisdom and his patience. He taught me seeing the “bigger picture” in research and helped me directing my work to tackle problems that are both, interesting and relevant. David has originally suggested several of the key concepts that I extend and refine in this thesis.

In cross-disciplinary research, such as this thesis, it is a constant challenge to ensure that the results are both, relevant and accessible to people from various research backgrounds. I am grateful to Dr. Alan Dorin for his constructive criticism that always ensured that my work remains pertinent to researchers from a variety of disciplines related to complexity science.

I am also very grateful to Dr. Suzanne Sadedin, who became a personal friend and was a great support in weathering the more challenging periods of the last few years. Suzanne introduced me to the fascinating field of evolutionary biology. She was a great research partner and adviser. I would be both, pleased and honoured to work with David, Alan and Suzanne in future.

I also thank all of my friends and family who stuck with me over the last few years, even when at times I was carried away with work and neglected to keep in touch as much as I should have. Without your support, my PhD journey would have been much more difficult.

I would like to acknowledge that during my degree I was funded by scholarships of Monash Research Graduate School, the Australian COSNet research network, the Australian Research Council, and various one-off conference travel bursaries.

PhD is a long and often difficult journey. But the rewards outweigh the challenges by far. In future, when asked whether I would recommend taking up a PhD degree, I would recommend it without hesitation to anyone who has the right interests and talents. Research is work; but research is interesting, and research is fun.

Declaration of originality

I, Gregory Paperin, hereby affirm:

To the best of my knowledge, this thesis contains no material that has been accepted for the award of any other degree or diploma in any institution other than Monash University.

To the best of my knowledge, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.



† Melbourne, April 2010

Mechanisms for Emergence and Self-Organisation in Complex Adaptive Systems

A Network-Theoretical Perspective

Gregory Paperin
MSci (Hons)

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy



Clayton School of Computer Science and Software Engineering
Faculty of Information Technology
Monash University

April 2010

Table of contents

ABSTRACT	1
ACKNOWLEDGEMENTS	3
DECLARATION OF ORIGINALITY	3
TABLE OF CONTENTS	7
PART I: INTRODUCTION	11
1. INTRODUCTION	13
1.1. RESEARCH GOALS	14
1.2. OVERVIEW	15
1.3. MAIN RESEARCH CONTRIBUTIONS	17
1.4. LIST OF PUBLICATIONS	20
2. METHODOLOGICAL CONTEXT	23
3. THE LIVEGRAPH FRAMEWORK FOR EXPLORATORY DATA ANALYSIS AND VISUALISATION	27
3.1. MODELLING TECHNOLOGIES	27
3.2. USE CASE	28
3.3. AVAILABLE ALTERNATIVES	28
3.4. SOLUTION OUTLINE	29
3.5. IMPACT	31
PART II: BACKGROUND	33
4. COMPLEX SYSTEMS	35
4.1. THE SCIENCE OF COMPLEX SYSTEMS	35
4.2. KEY OUTCOMES OF COMPLEXITY SCIENCE	37
4.2.1. Natural sciences	37
4.2.2. Cybernetics	38
4.2.3. Computer science and Artificial Life	38
4.3. EMERGENCE IN COMPLEX ADAPTIVE SYSTEMS	39
5. NETWORKS IN COMPLEX SYSTEMS	43
5.1. BACKGROUND	43
5.2. ROLES OF NETWORKS	45
5.2.1. Component interaction networks	45
5.2.2. Primary and secondary interaction networks	45
5.2.3. Information networks	46
5.2.4. State transition networks	46
5.3. PHASE CHANGES AND CONNECTIVITY	49
5.4. TOPOLOGICAL PROPERTIES OF NETWORKS	49
5.5. CLUSTERS AND MODULES	51
5.6. NETWORK MOTIFS	51
5.7. STATE SPACES AS NETWORKS	52
6. SPECIATION THEORY AND BIOLOGICAL FITNESS LANDSCAPES	53
6.1. SPECIATION	53
6.2. FITNESS LANDSCAPES	55
6.2.1. Overview	55
6.2.2. Formal model	56
6.2.3. Fitness landscapes in speciation models	57

6.2.4. Nearly neutral networks	59
PART III: PHASE TRANSITIONS AND NETWORK CONNECTIVITY IN COMPLEX SYSTEMS	63
7. DUAL PHASE EVOLUTION	65
7.1. INTRODUCTION.....	65
7.2. CONNECTIVITY PHASES IN INTERACTION NETWORKS	66
7.3. CONSIDERING MULTIPLE NETWORKS.....	67
7.4. MECHANISMS FOR PHASE TRANSITIONS	68
7.5. SUMMARY	69
8. CONNECTIVITY PHASE TRANSITIONS IN NATURAL AND ARTIFICIAL SYSTEMS.....	71
8.1. NON-LIVING SYSTEMS.....	71
8.2. ECOSYSTEMS.....	73
8.3. LIVING SYSTEMS AT EVOLUTIONARY SCALES	73
8.4. PERCEPTION AND COGNITION	74
8.5. SOCIO-ECONOMIC SYSTEMS	75
8.6. MODELS OF MODULARITY	77
8.7. OPTIMISATION PROBLEMS.....	78
8.8. HEURISTIC ALGORITHMS	78
8.9. SUMMARY	79
PART IV: PERPETUAL NOVELTY AND SUSTAINED DIVERSITY IN BIOLOGICAL SYSTEMS	81
9. BACKGROUND	83
9.1. INTRODUCTION.....	83
9.2. PERPETUAL NOVELTY AND DIVERSITY IN CAS.....	85
9.3. EVIDENCE: RECURRENT ENVIRONMENTAL PERTURBATIONS IN BIOLOGICAL EVOLUTION	87
10. DPE AND PERPETUAL NOVELTY AND SUSTAINED DIVERSITY IN LANDSCAPE ECOSYSTEMS	89
10.1. SIMULATION MODEL OF ADAPTIVE RADIATION IN LANDSCAPES	90
10.1.1. Overview	90
10.1.2. Scenarios	91
10.1.2.1. Basis scenario	91
10.1.2.2. Disturbances	91
10.1.2.3. Sensitivity analysis	93
10.1.3. Results.....	93
10.1.4. Discussion	95
10.2. DETAILED ANALYSIS OF THE MODEL FITNESS LANDSCAPE	97
10.2.1. Choice of the genetic model.....	97
10.2.2. Equivalence between the employed genetic model and a standard niche-based additive quantitative trait model.....	98
10.2.3. Analytical examination	99
10.2.4. Numerical examination.....	102
10.2.4.1. Fitness distribution	102
10.2.4.2. Genotype reach cluster size	105
10.2.4.3. Genotype distances along adaptive paths	105
10.2.4.4. Number of advantageous mutations	106
10.2.4.5. Number of closest niches.....	107
10.2.4.6. Hybrid fitness	107
10.2.5. Conclusion.....	109
11. EVOLUTION ON HOLEY FITNESS LANDSCAPES	111
11.1. A GENOTYPE MODEL WITH ENVIRONMENT ADAPTATION TRAITS AND HOLEY FITNESS LANDSCAPE	113

11.2. HOLEY FITNESS LANDSCAPES IN SIMULATIONS	115
11.2.1. Feasibility of truly random distributions in computer simulations	115
11.2.2. Difficulties in building HFL models suitable for individual-based simulations	115
11.2.3. Creating a HFL suitable for individual-based simulations	117
11.3. EVOLUTIONARY PROPERTIES OF THE GENOTYPE MODEL WITH ENVIRONMENT ADAPTATION TRAITS AND HOLEY FITNESS LANDSCAPE	120
11.3.1. Adaptation properties	120
11.3.2. Speciation properties	123
11.3.3. Conclusion	126
11.4. HOLEY FITNESS LANDSCAPES AND THE MAINTENANCE OF EVOLUTIONARY DIVERSITY	127
11.4.1. Background	127
11.4.2. Methods	128
11.4.3. Results	130
11.4.4. Conclusion	134
11.5. APPLICATIONS OUTSIDE EVOLUTIONARY BIOLOGY	136
11.5.1. Premature convergence in genetic algorithms	136
11.5.2. Employing HFL genetics in optimisation algorithms	137
11.6. SUMMARY	139
 PART V: SELF-ORGANISATION IN COMPLEX NETWORKS	 141
12. EMERGENT CONNECTIVITY PHASE TRANSITIONS IN COMPLEX NETWORKS	143
12.1. INTRODUCTION	143
12.2. CONNECTIVITY PHASE TRANSITIONS IN AN ABSTRACT RESOURCE FLOW NETWORK MODEL	144
12.2.1. Model details	144
12.2.2. Results	145
12.2.3. Conclusion	146
12.3. CONNECTIVITY PHASE TRANSITIONS IN A COPYCAT NETWORK MODEL	148
12.3.1. Model details	148
12.3.2. Results	149
12.3.3. Conclusion	152
13. GENERATIVE MODELS FOR IMPORTANT NETWORK TOPOLOGIES	153
13.1. INTRODUCTION	153
13.2. MODULAR NETWORKS	155
13.2.1. Social network structures resulting from explicit dual phase processes	155
13.2.1.1. Model details	155
13.2.1.2. Results	156
13.2.1.3. Parameter configurations leading to other network topologies	159
13.2.2. Modular networks resulting from implicit dual phase processes	160
13.2.2.1. Basis scenario	161
13.2.2.2. Sensitivity analysis	163
13.2.2.3. Complex and hierarchical modules	164
13.2.2.4. Summary	168
13.3. SCALE-FREE NETWORKS	170
13.3.1. Model overview	170
13.3.2. Initial investigation	171
13.3.3. Systematic investigation	173
13.3.3.1. Relative length of local and global interaction phases	173
13.3.3.2. Power-law goodness of fit and estimated exponent	173
13.3.4. Improved generative model	176
13.3.4.1. Emergent two-phase dynamics in the updated model	179
13.3.5. Dependence of attractor edge density on network size	180
13.3.6. Summary	181
 PART VI: THE BIG PICTURE	 183
14. DPE AND OTHER FRAMEWORKS FOR REASONING ABOUT COMPLEX SYSTEMS	185
14.1. SELF-ORGANISED CRITICALITY	185

14.2. THE ADAPTIVE CYCLE	188
15. CONCLUSIONS.....	189
15.1. DIRECTIONS FOR FUTURE RESEARCH	190
15.1.1. Biological modelling	190
15.1.2. Evolutionary computation.....	190
15.1.3. Network models	190
15.1.4. DPE theory	191
15.1.4.1. DPE-interactions with external system environments	191
15.1.4.2. Other directions.....	192
15.2. VERDICT	193
 PART VII: APPENDIX	 195
A. TECHNICAL DETAILS OF THE SIMULATION MODEL OF ADAPTIVE RADIATION IN LANDSCAPES	197
A.1. MODEL CONFIGURATION.....	197
A.1.1. Landscape	197
A.1.2. Individuals.....	197
A.1.3. Lifecycle	198
A.1.4. Initialisation	199
A.1.5. Reproduction.....	199
A.1.6. Selection.....	200
A.1.7. Dispersal	200
A.2. ANALYSIS OF THE BASIS RUN SCENARIO	201
B. LIST OF FIGURES.....	205
C. LIST OF TABLES	207
D. INDEX OF TERMS AND ABBREVIATIONS.....	209
E. REFERENCES	213



Part I

Introduction

In this thesis I study the role of networks and their properties in complex systems. I highlight some previously undescribed underlying mechanisms common to many systems and apply these findings to gain new insights into some specific processes in biological evolution and in self-organisation of socio-economic groups. **Part I** lays out the thesis structure, summarises the main research contributions and discusses the tools and methods used in this research.

Chapter 1: Introduction.

This thesis studies mechanisms for emergence and self-organisation in complex systems. It concentrates on the role of networks and argues that recurrent connectivity changes represent a previously undescribed underlying mechanism common to many systems. These findings are used to gain new insights into some specific processes in biological evolution and in self-organisation of socio-economic groups. The first chapter provides an outline of this thesis, sets out the main goals, and summarises the key research contributions and published works.

Chapter 2: Methodological context.

Work presented in this thesis is located at the crossroads of complexity science, evolutionary biology, artificial life and network theory. Two main methods of study are employed: analysis and re-interpretation of existing literature aimed at identifying and highlighting shared mechanisms and common properties in known phenomena; and simulation models aimed at studying the effects of particular network properties in specific complex systems. I sketch out a science-philosophical context to motivate the employed methods and to qualify the results obtained on their basis.

Chapter 3: The LiveGraph framework for exploratory data analysis and visualisation.

An important contribution of this thesis is a software framework of research tools that was originally developed to facilitate research presented in this thesis. This framework – LiveGraph – is a suite of software tools and libraries for real-time data visualisation and exploratory analysis that incorporates features not available in other packages. After its initial publication as an open-source project, LiveGraph has been used by thousands of people and in several academic peer-reviewed publications. This chapter provides an outline of the frameworks and its capabilities.

Chapter 1:

Introduction

Our world is dominated by complex, networked systems of many kinds. Studying how large-scale features and properties emerge in such systems is crucial to improve our understanding of nature and our ability to advance science and society. In this thesis I study the role played by networks and their properties in natural and artificial complex systems. I uncover some previously undescribed mechanisms that govern behaviour of many systems and apply my findings to better understand some specific processes in biological evolution and in self-organisation of socio-economic groups. This chapter provides an outline of this thesis, sets out the main goals, and summarises the key research contributions and published works.

From finance markets to the World Wide Web, from genetic regulatory networks to food webs, science and society are dominated by complex, networked systems of many kinds. Studying how these systems self-organise, acquire features and properties, such as perpetual novelty and sustained diversity, is crucial in trying to understand and manage such systems.

It is increasingly recognised that large-scale, complex systems display emergent features that result not from the properties of individual system components, but from the cumulative effects of their interactions (see literature review in [section 4.1](#)). A fruitful paradigm is to represent components and their interactions within complex systems of different kinds as networks and to apply network and graph theory to study such systems (see review in [chapter 5](#)). This inter-disciplinary approach allows explicit characterisation of common mechanisms and aids cross-fertilisation between disciplines. Network theory and related methods from statistical physics found useful applications in fields as diverse as social sciences, evolutionary biology and cognitive neuroscience ([chapter 5](#)).

Abstracting complex systems as networks of component interactions raises many questions. For instance, what structures promote important properties such as resilience, diversity, modularity and novelty? What processes give rise to common topologies? And how do network properties influence the behaviour of systems?

In this thesis I approach these and related questions ([section 1.1](#)), and study the role played by networks in natural and artificial systems. Networks are inherent in all complex systems ([chapter 5](#)), and patterns and processes that occur within the underlying networks play a major role in the emergence of system wide features ([chapter 7](#)). Perhaps the most widely felt network processes are changes in edge density, especially the connectivity avalanche ([section 5.3](#)). My research shows ([chapter 8](#)) that phase changes occur in a wide range of complex systems. What is more, those phase changes can typically be traced to connectivity avalanches within underlying networks.

Focusing on the role of network connectivity in systems reveals many deep similarities between superficially unrelated systems and processes, pointing to common causal mechanisms. Different connectivity phases display characteristic features, such as apparent balance and stabilising selection in high connectivity phases, and vivid variation and exploration in low connectivity phases. In this thesis I argue that recurrent phase changes in connectivity of system components constitute a general mechanism behind some emergent phenomena in both, biological and other complex adaptive systems, as well as in non-living complex systems. As a consequence, I extend the concept of Dual Phase Evolution (DPE) and formulate it as a universal framework useful for reasoning about complexity ([chapter 7](#)). Interpreting processes as DPE reveals new useful insights, for instance about patterns in species evolution ([chapter 10](#)). It also suggests new methodologies, such as new generative models for important network architectures ([chapter 13](#)).

1.1. Research goals

The central aim of this thesis is to provide new insights into mechanisms behind emergence and self-organisation in complex systems. Primarily, I concentrate on the role of networks and network connectivity in self-organisation and emergence of complex properties such as perpetual novelty and sustained diversity.

To address this objective I focus on the following specific goals:

- 1) Elicit mechanisms by which network connectivity influences dynamics in different complex systems and facilitates self-organisation and emergence of large-scale properties.
- 2) Formulate the insights won by addressing goal (1) as general principles that describe relationships between network connectivity and system dynamics. Describe the role of these mechanisms in emergence and self-organisation in complex systems.
- 3) Investigate how the general principles formulated in goal (2) facilitate the emergence of specific complex properties in particular complex systems.
- 4) Investigate how the general principles formulated in goal (2) facilitate self-organisation towards complex underlying structures in particular complex systems.
- 5) Compare the general principles formulated in goal (2) with other established approaches to describing and reasoning about self-organisation and emergence in complex systems. Provide a critical comparison between the new insights and the established approaches.

Methodology:

The following approach is taken to address the above goals:

Goal 1: I examine a wide range of previous studies from different disciplines and highlight shared underlying similarities and mechanisms within the studied systems ([chapter 8](#)).

Goal 2: I extend the previously known concept of Dual Phase Evolution (DPE) and formulate it as a general framework for reasoning about complex systems rooted in network theory. The DPE framework describes two different connectivity phases in networks that underlie complex systems, and how these phases and the transitions between them relate to system dynamics ([chapter 7](#)).

Goal 3: I investigate how recurrent phase changes in landscape connectivity facilitate the emergence of perpetual novelty and sustained diversity in landscape ecosystems. The investigation is based on a simulation model of adaptive radiation (speciation) in a landscape, where connectivity is affected by intermittent disturbances ([chapter 10](#)).

Goal 4: I investigate how recurrent changes in density (connectivity) of interactions between system components lead to important topological structures in social and socio-economic networks. The investigation is based on several simulation models, where recurrent phase transitions in density of interactions between agents result in important topologies, such as modular and scale-free structures ([chapter 13](#)).

Goal 5: I compare the DPE framework to two other well-known frameworks for reasoning about complex systems (Self-Organised Criticality and the Adaptive Cycle), and highlight similarities and differences. DPE appears to be complementary to the other frameworks ([chapter 14](#)).

In the following [section 1.2](#) I provide a brief overview of the thesis structure. Thereafter, in [section 1.3](#), I highlight the key results and relate them to the above objectives.

1.2. Overview

This thesis is organised in six major parts. In the current [part I](#), I set out the scene for the thesis, lay out the research objectives, highlight the key contributions and describe primarily employed research methods. I also discuss an additional contribution that results from work described in this thesis: the LiveGraph software framework for exploratory analysis and visualisation of scientific simulation data. LiveGraph is suite of research tools. Originally developed within the scope of the work presented here, it is now employed by thousands of people worldwide and thus presents a significant contribution in itself. Aiming at conciseness, I avoid reviewing relevant background literature in [part I](#); instead, I refer to appropriate sections in [part II](#) where important background literature is reviewed in detail.

In [part II](#), I review the background literature and highlight established concepts and theories that are at the centre of this research. I begin with a brief general overview of complexity theory and its key results ([chapter 4](#)), and specifically consider the philosophical issues arising from studies of emergence ([section 4.3](#)). I continue with an overview of network theory ([chapter 5](#)), specifically of the results most relevant to the research described here. Finally, in [chapter 6](#), I review some biological concepts that play a role in my work that applies network theory to studies of ecological systems and speciation.

My original research contributions are discussed beginning in [part III](#). In [chapter 7](#) I introduce the main theoretical concept of this thesis – the framework of Dual Phase Evolution (DPE) – and discuss some of the theoretical corollaries that arise from the DPE framework. In [chapter 8](#) I apply the DPE framework to a wide range of previously studied natural and artificial complex systems and elicit the new insights that DPE reveals about those systems.

In [part IV](#), I investigate mechanisms for the emergence of self-organisation, perpetual novelty and sustained diversity in biological complex adaptive systems. The diversity and multiplicity of species is one of the most prominent attributes that signify novelty and diversity in biology. Thus, in [chapter 10](#) I study the dynamics of adaptive radiation (speciation) in landscape ecosystems using simulation experiments. In this study I concentrate on the role of interaction networks within landscapes and their connectivity. Results reveal that the role of network connectivity in determining speciation dynamics depends on assumptions about genetic properties of individuals made in the simulation experiments. Most notably, the fitness landscape that results from these genetic assumptions has strong influence on system behaviour. Thus, in [chapter 11](#), I undertake a detour from the network-theoretical theme of this thesis, and study in more detail, how assumptions about genetic fitness landscapes can influence predictions about the emergence and sustainability of novelty and diversity in biological ecosystems. These studies employ simulation models based on recent results from evolutionary speciation theory.

In [part V](#) I study processes that facilitate the emergence of important network topologies and structures. Connectivity phase transitions govern many system properties and an important question is thus, - *what mechanisms facilitate such transitions?* Simulation studies in [chapter 12](#) demonstrate different mechanisms by which recurrent connectivity transitions can emerge in network systems without facilitation by external events. However, connectivity density is not the only important network property that drives system dynamics. Other topological properties, including modular organisation ([section 13.2](#)) and power-law degree distributions ([section 13.3](#)) are known to facilitate emergent phenomena such as resilience, fault-tolerance and criticality. Thus, in [chapter 13](#), I study generative models for modular and scale-free networks.

In [part VI](#) I draw the results of work discussed in this thesis together and put them in a wider context through a critical comparison with other established frameworks for reasoning about complexity ([chapter 14](#)). In particular, I consider the concepts of Self-Organised Criticality and the Adaptive Cycle. In [chapter 15](#), I summarise the findings, identify gaps within my studies and discuss future research directions.

Overall, the structure of the thesis can be sketched as a loop touching on several aspects of network theory in complex systems research. That loop follows the key realisation that system behaviour and the topology of underlying networks interact in a feedback process that influences both, local component interactions and large-scale, emergent system properties (Figure 1.1).

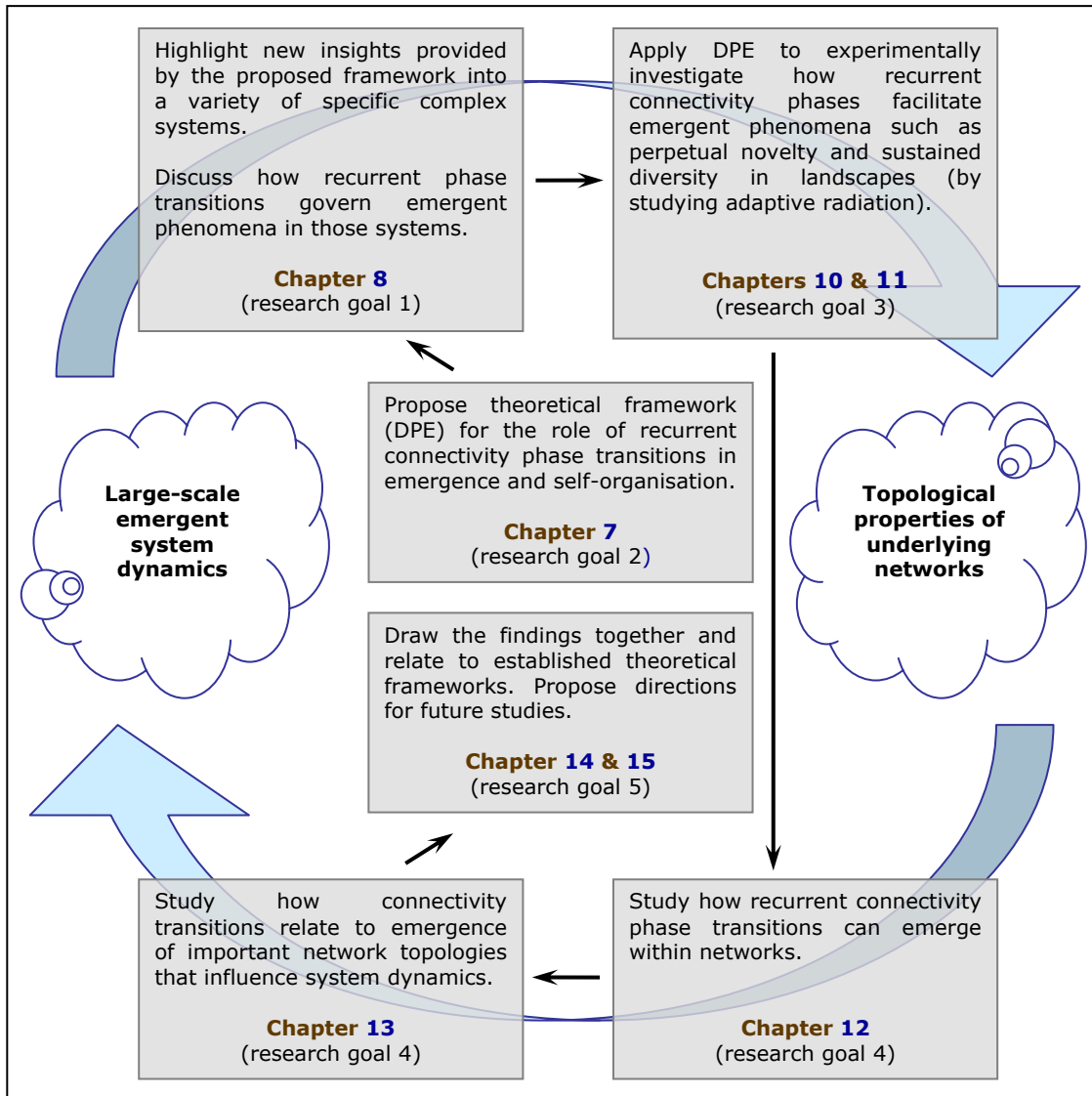


Figure 1.1. Thesis structure. The thesis is organised in a metaphorical loop that follows the feedback circuit between system dynamics and the topology of underlying networks.

1.3. Main research contributions

In this section I summarise the key contributions of this theses. As the thesis chapters group the different pieces of work together into common focus areas, this summary follows the chapter structure, whilst concentrating on chapters that provide key results. Where applicable, I relate each result to the *research goals* defined in [section 1.1](#).

Chapter 3. The LiveGraph framework for exploratory data analysis and visualisation.

This chapter describes a new data analysis tool that has become widely used in several areas:

- LiveGraph is a software framework for exploratory data analysis and real-time visualisation. Initially developed to facilitate research presented in this thesis, LiveGraph has been used by thousands of people and in several scientific publications (by unaffiliated authors). LiveGraph is a tool that allows performing tasks around simulation experiments more efficiently and is thus in itself an important contribution.

Chapter 7. Dual Phase Evolution.

This chapter establishes key relationships between connectivity phases in networks that underlie complex systems and system dynamics (*research goal 2*):

- Dual Phase Evolution (DPE) – the central theoretical concept of this thesis – originally described recurrent transitions between different kinds of dynamics in complex adaptive systems. In this chapter, DPE is extended and formulated as a general framework rooted in network theory. The framework relates connectivity of networks that underlie complex systems to typical system behaviour, and associates recurrent transitions between network connectivity phases to a number of emergent properties in complex systems.

Chapter 8. Connectivity Phase Transitions in Natural and Artificial Systems.

This chapter demonstrates how network connectivity phases influence behaviour of many complex systems and relates the results to concepts established by the DPE framework (*research goal 1*):

- An examination of a wide range of systems from physics, biology, sociology, economics, mathematics and computer science, based on previous studies, reveals that recurrent phase changes in network connectivity facilitate the emergence of many complex properties.
- Interpreting known and new phenomena within the DPE framework reveals previously undescribed underlying mechanisms and processes shared between many systems.

Chapter 10. DPE and Perpetual Novelty and Sustained Diversity in Landscape Ecosystems.

This chapter provides new insights into the relationship between disturbances, landscape connectivity, and the intermittent nature of species evolution in landscapes. It also sheds new light on mechanisms involved in perpetual novelty and sustained diversity observed in landscape ecosystems (*research goal 3*):

- A simulation model of adaptive radiation in landscapes shows that intermittent disturbances facilitate complex diversity and perpetual novelty in the ecosystem through two main mechanisms:

-
- First, disturbances affect the connectivity of two important interaction networks within the landscape. One is a network of reproductive interactions between individuals that facilitates gene flow and acts against adaptive radiation. Disasters separate populations and disrupt gene flow, which facilitates divergence through genetic drift. The second is a network of unoccupied cells. Disturbances free up additional cells thus providing connected invasion paths and aiding percolation by new species.
 - The second mechanism through which disturbances facilitate continuous novelty and high diversity in the model, is alteration of the environment in affected areas. Altered conditions result in directional selection towards new niches for extended periods of time, thus aiding local adaptations to establish stable populations.

Chapter 11. Evolution on Holey Fitness Landscapes:

This chapter takes a detour from the network theme of the thesis and considers the role of holey fitness landscapes in species evolution. I introduce a new genetic modelling technique that allows applying holey fitness landscapes, a hitherto purely analytical concept, in individual-based modelling, and use this technique to attain new insights into the maintenance of evolutionary diversity in parapatric populations (*these results address research goal 3, and also provide additional insights*):

- By combining two existing techniques, I introduce a genetic model that incorporates both, an explicit, biologically realistic treatment of genetic (in-)viability issues, and an inherent approach for relating genotype fitness to the ecological environment.
- Previously, holey fitness landscapes were not used in individual-based simulations due to inherent difficulties in crating a suitable, computationally efficient viability function. On the basis of the introduced genetic model, I demonstrate how these difficulties can be managed to create simulation models incorporating HFL-genetics.
- A numerical analysis of the genetic fitness landscape exhibited by the new model shows that the model exhibits many biologically realistic properties. This analysis can potentially inform and direct the design of future studies.
- Previous studies suggest that ecologically-based reproductive isolation will rapidly collapse under migration. However, based on the new techniques, I show that HFL can often maintain diversity and may facilitate exploration of new niches and speciation. Simulation results also point towards an important role of HFL-genetics in evolution of social selection through reinforcement.
- I discuss how the techniques developed in this chapter may prove useful in biologically-inspired engineering, for instance for combating premature convergence in evolutionary optimisation algorithms.

Chapter 12. Emergent connectivity phase transitions in complex networks:

This chapter demonstrates mechanisms for the emergence of connectivity phase changes without facilitation by external disturbances (*research goal 4*):

- In a model of resource flow in complex systems, connectivity transitions arise through a feedback between the properties of the resource distribution network and edge density.
- Simple behavioural rules, such as the human tendency to copy peer behaviour, can lead to emergent connectivity phases in a social network model. A causal mechanism for these phase transitions is identified in slow accumulation of specific small-scale structures and intermittent releases.

Chapter 13. Generative models for important network topologies:

This chapter introduces new generative models for key network architectures (e.g. modular and scale-free networks) that are found in social groups, as well as in various other systems (*research goal 4*):

- A social network model shows that both, modular structures, and circuit and chain structures emerge according to different parameter values. Several of the resulting structures bear close resemblance to some empirical social networks.
- A network model of an economy demonstrates the emergence of multi-level modules organised in complex hierarchies. The simple rules governing network evolution in this model imply that similar mechanisms may be present in other systems.
- A model of interactions in human and animal groups demonstrates a hitherto unknown, natural generative mechanism for scale-free networks that do not grow in size. Previous generative models for scale-free networks either operate on growing networks (e.g. preferential attachment) or are highly synthetic and unlikely to correspond to natural processes.

Chapter 14. DPE and other frameworks for reasoning about complex systems:

This chapter relates the theoretical concepts developed in this thesis to other established frameworks for reasoning about complex systems (*research goal 5*):

- DPE complements the notion of Self-Organised Criticality (SOC). SOC implies that CAS tend to self-organise towards a critical attractor. On contrary, DPE proposes that systems develop towards a simple attractor, and are repeatedly perturbed away by internal and external disturbances, thereby repeatedly crossing the critical threshold. In many instances these two cases may result in equivalent observations.
- The Adaptive Cycle (AC) theory describes recurrent transitions in dynamics of socio-economic systems between exploration of new development directions and exploitation of available resources. DPE generalises this concept in terms of network-theory and applies it to a wider range of systems.

The next section provides a brief overview of publications of the above key results.

1.4. List of publications

Most of the findings presented in this thesis have been published in peer-reviewed journals and conference proceedings. Some material presented in **part V** is currently in preparation for publication. An extended paper that summarises the key findings of this thesis, and reviews most of the material discussed here has been accepted for review by the Royal Society Journal Interface.

Throughout this thesis, I seek to avoid self-citations. Instead, at the beginning of every chapter or top-level section, I list publications that contain material presented in the respective chapter or section.

Here, I list all peer-reviewed publications containing material that results from my work undertaken during the time period when I was engaged in research presented in this thesis. Most of these publications (marked with "♦") contain material that has been incorporated into this volume.

Abstracts and additional details for the publications listed here, as well as the full text for most of these papers can be obtained from <http://www.paperin.org/publications>.

Legend:	
 - journal paper	 - edited volume
 - book chapter	 - conference paper
 - conference poster	♦ - publication is used in this thesis

-  **G. Paperin**, D. Green, S. Sadedin (2010): "Dual Phase Evolution in Complex Adaptive Systems". Journal of the Royal Society Interface. Accepted for review. ♦
-  **G. Paperin**, S. Sadedin (2010): "Dual Phase Evolution as a Framework for Understanding Complex Adaptive Systems". Springer. In press. ♦
-  **G. Paperin** (2010): "The Role of Networks in Computational Complex Systems". International Transactions on Systems Science and Applications. In press. ♦
-  **G. Paperin**, S. Sadedin (2009): "The Dual Phase Evolution Framework for Understanding Evolutionary Dynamics in Complex Adaptive Systems". In: "Proceedings of the 2009 International Conference on Evolutionary Computation (ICEC'09)". ♦
-  **G. Paperin**, S. Sadedin (2009): "Towards formalising the theory of Dual Phase Evolution". Presented at the 10th European Conference on Artificial Life (ECAL'09). ♦
-  S. Sadedin, **G. Paperin** (2009): "Implications of the social brain hypothesis for evolving human-like cognition in digital organisms". In "Proceedings of the 10th European Conference on Artificial Life (ECAL'09)". Springer.
-  S. Sadedin, **G. Paperin**, T. G. Leishman, eds. (2008): "Proceedings of the 12th Asia-Pacific Symposium on Intelligent and Evolutionary Systems (IES'08)". Monash University. ISBN 978-0-646-50671-5.
-  **G. Paperin**, D. G. Green, T. G. Leishman (2008): "Dual Phase Evolution and Self-Organisation in Networks". In "Proceedings of the 7th International Conference on Simulated Evolution And Learning (SEAL'08)". Springer. ♦
-  **G. Paperin**, S. Sadedin, D. Green, A. Dorin (2008): "Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity". In "Proceedings of the 11th International Conference on Artificial Life (ALife XI)", MIT Press, Cambridge, MA. ♦

-  **G. Paperin** (2008): "Using Holey Fitness Landscapes to Counteract Premature Convergence in Evolutionary Algorithms". In "Proceedings of the Graduate Student Workshop at the Genetic and Evolutionary Computation Conference 2008 (GECCO'08)". ACM Publishing. ♦
-  **G. Paperin** (2008): "Evolving sequence patterns for prediction of sub-cellular locations of eukaryotic proteins". In "Proceedings of the Genetic and Evolutionary Computation Conference 2008 (GECCO'08)". ACM Publishing.
-  **G. Paperin**, D. G. Green, A. Dorin (2007): "Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation". In: "Proceedings of 2007 International Symposium on Computational Models for Life Sciences (CMLS'07)", AIP Conf. Proc., Vol. 952, pp. 268-278. ♦
-  T. G. Leishman, D. G. Green, **G. Paperin** (2007): "Dual Phase Evolution – a mechanism for self-organisation and optimisation". In: "Asia-Pacific Symposium on Intelligent and Evolutionary Systems 2007 (APSIES'07)". ♦
-  **G. Paperin**, D. G. Green, S. Sadedin, T. Leishman (2007): "A Dual Phase Evolution model of adaptive radiation in landscapes". In: "Proceedings of the 3rd Australian Conference on Artificial Life (ACAL'07)", Springer LNCS, Vol. 4828/2007, pp. 131-143. ♦
-  **G. Paperin**, D. G. Green, S. Sadedin, T. G. Leishman (2007): "Complexity in Speciation: Effects of disasters on adaptive radiation in a Dual Phase Evolution model". In: "Proceedings of the 8th Asia-Pacific Complex Systems Conference (Complex'07)". ♦
-  **G. Paperin** (2007): "The LiveGraph framework for exploratory data analysis and visualisation". Software demonstration at the "8th Asia-Pacific Complex Systems Conference (Complex'07)". ♦
-  G. Briscoe, S. Sadedin, **G. Paperin** (2006): "Biology of Applied Digital Ecosystems". In: "Proceedings of Inaugural IEEE International Conference on Digital Ecosystems and Technologies 2007 (DEST'07)".
-  **G. Paperin** (2006): "Security of Communication and Quantum Technology". In: M. Quigley (ed.), "Encyclopaedia of Information Ethics and Security". Idea Group Publishing, 2007. ISBN 159140987X.

Chapter 2:

Methodological context

A central point of this thesis is a theoretical framework (termed DPE) for reasoning about complex systems that is centred around the role of connectivity in underlying interaction networks. In order to motivate and support the framework I conduct a study of literature from a wide range of disciplines aimed to identify and highlight shared mechanisms and common properties in different complex systems (chapter 8). The re-interpretation of existing literature provided in this study and the identification of hitherto unstudied common principles and mechanisms is an original contribution and constitutes a key result of this thesis. Further results presented in this thesis are obtained through simulation studies that apply insights gained from the theoretical DPE framework to models of particular complex systems (parts IV and V). This methodology is central to several sciences that study complexity. This chapter provides a brief science-philosophical context to motivate the employed methods and to qualify the results obtained on their basis.

Two research methodologies are central to this thesis. The first methodology is the review and re-interpretation of previous work aimed at revealing concealed mechanisms and previously unnoticed parallels in known phenomena. The second methodology is simulation modelling of complex systems aimed to study how the identified mechanisms drive specific behaviours in natural systems. In this chapter I motivate the applied methodology by providing a brief science-philosophical context.

Complexity theory aims to reveal deep similarities between superficially unrelated systems and processes by drawing together ideas from many different disciplines into a coherent and unified methodology [Nicolis & Prigogine, 1977a; Manson, 2001]. A key challenge therein is overcoming incompatible terminologies, competing theories of equivalent content and sometimes also entrenchment by traditionalists within different cradle sciences of complexity (chapter 4). Scientists tend to pay attention to problems defined by the concepts and methods readily available [Kuhn, 1977]. Thus, some facts, although known, are pushed to the periphery of scientific investigation, either because they are deemed irrelevant, or because their study would demand unavailable data or techniques. Bringing the relevance of these facts into focus presents a valuable and important contribution to science: it can catalyse re-conceptualisation and cause impacts that can range from an undramatic reorganisation of concepts to a scientific revolution [Di Paolo et al., 2000].

In chapter 7 I introduce a theoretical framework (termed DPE) for reasoning about complex systems. The framework is centred around the realisation that recurrent phase changes in connectivity of underlying networks play a key role in emergence and self-organisation within many kinds of systems. In chapter 8 I provide evidence for the theoretical ideas presented in chapter 7 through a study of a large body of literature from different fields that highlights common processes and re-interprets some of the original results. This literature study is a key contribution of this thesis. It does not only motivate and direct the simulation studies conducted in the later chapters. It also demonstrates that DPE is a diverse and useful interpretation framework for future research. What is more, it uncovers underlying mechanisms and deep similarities between systems, not always noticed by the original authors. Thus, the literature study is not a literature review. A literature review that provides a context and motivation for the research described in this thesis is provided in chapters 4, 5, and 6. Conversely, the literature study in chapter 8 takes the role of data analysis performed on existing ideas and results rather than on newly collected data.

In parts IV and V of this thesis I study the applications of the DPE framework to different complex systems. The main method of these studies is simulation. Simulation, particularly computer simulations, is a relatively new method of scientific enquiry that only became available with the advent of affordable computing power provided by modern IT systems. Thus, simulation models as a scientific method are actively discussed by science philosophers and practicing researchers [e.g. Maynard Smith, 1974; Kawata & Toquenaga, 1994; Bedau,

1998; 1999; Di Paolo et al., 2000; Grimm et al., 2005]. It is useful to sketch out some of the points made in such discussions.

The traditional approach to system modelling is to express system dynamics in terms of equations, usually differential equations. Commonly, such approach describes large-scale, average system behaviour [Castiglione, 2006]. Mathematical assumptions made to construct tractable equation-based models can often be relaxed in a simulation model [Taylor & Jefferson, 1995]. For instance, traditional models in population biology often assume random mating and unstructured populations. Simulations models, however, allow for structured, spatially explicit populations and other evolutionary phenomena that are difficult to treat analytically, such as frequency dependent selection [Di Paolo et al., 2000].

Simulation models can be classified in two broad categories: traditional numerical simulations and individual-based models (also termed agent-based models) [Grimm & Railsback, 2005; Castiglione, 2006]. The distinction is not always clear-cut and combined approaches are possible. *Numerical simulations* are mainly based on routines for solving the equations that describe the behaviour of a system. For instance, they are employed when an analytical solution to the equations is intractable or when the behaviour of a system over time is studied. The numerical approach is often limited to modelling the average behaviour of systems [Castiglione, 2006]. *Individual-based models* emphasise individual low-level system components and their interactions [Kawata & Toquenaga, 1994]. They have the potential to explore high-level aggregate behaviours and can be used to model phenomena that exhibit self-organisation and emergence [Taylor & Jefferson, 1995]. For instance, individual-based simulations have become widespread in ecosystems modelling after it became common to treat ecosystems as CAS [Di Paolo et al., 2000; Grimm et al., 2005].

The view of the role of simulation experiments varies among researchers. One extreme position is that simulation models need to act as maximally realistic replicas of systems [Maynard Smith, 1974; Kitano et al., 1997]. Such models must be validated against empirical or experimental data and can then be used for quantitative predictions. Models calibrated and validated in this way can be used for virtual experiments and for exploring what-if scenarios [Kitano et al., 1997; Di Paolo et al., 2000].

Conversely, Bedau [1998; 1999] sees simulation models as abstract, computational thought experiments that may be far away from actual natural systems. Such models can capture essential processes and thus provide instances of phenomena of interest, while abstracting away from micro-details gaining simplicity and universality [Bedau, 1999]. Many researchers (e.g. see Di Paolo et al. [2000] for a discussion) take an intermediate position between seeing simulations as tools for realistic numerical predictions and as computational thought experiments.

Simulation models that do not make exact quantitative predictions of system behaviour that can be calibrated against empirical data can be used to study qualitative patterns in systems dynamics. How can such models be validated? Grimm et al. [2005] suggest an approach termed “pattern-oriented modelling”: Simulation models should not only reproduce the qualitative behaviour it was designed to study, but also other emergent behaviours observed in the real system that were not at the focus during model design. This is not an obligatory requirement in all cases, particularly as model simplicity may need to be traded off against the complexity required for multiple emergent patterns, however, when qualitative models exhibit multiple large-scale behaviours that are also present in the real system, it can be taken as a strong sign of model validity.

Due to the nature of complex systems, some hypotheses about their internal mechanisms are inherently difficult to test empirically. Such testing requires accurate empirical data over extended time periods, which can stretch into years or even millennia. Explanatory variables are typically numerous and cannot be manipulated, experiments are frequently unfeasible and are replaced by observations and data often collected for other purposes. As a result, thought experiments and simulation models are essential tools in complexity research. However, the power of qualitative individual-based models goes beyond their role as computational thought experiments. They can be used to assess and verify intuition and reasoning about emergent phenomena. Di Paolo et al. [2000] exemplify this by considering a thought experiment discussed by Gould [1990]: If the Earth’s geological history was replayed, would today’s Earth look similar to what we observe, or would it have strongly diverged due to differences in small random events? In particular, would intelligent life have

evolved? Gould's assessment [1990] is that evolution of intelligent life forms in such case would be highly unlikely. However, Di Paolo et al. [2000] follow Bedau [1998] in that such a thought experiment is inconclusive until a computer simulation can verify either conclusion.

There are numerous examples of how qualitative simulation models solved key problems by demonstrating how particular theoretical concepts can work in dynamical systems [Di Paolo et al., 2000]. An example is the Daisyworld model (section 4.3) of a planet inhabited by only two species. The model demonstrates that global self-regulation can emerge from only local interactions without a global controller – a concept that was considered highly controversial before the model was published (section 4.3). Another example is the demonstration of the Baldwin effect by Hinton and Nowlan [1987]. The Baldwin effect, which suggests that evolving increased ability for life-time learning can speed-up evolution, received limited attention for over 9 decades since its original publication [Baldwin, 1896]. The scepticism was partly due to an incorrect association with Lamarckism, partly due to the absence of a convincing argument showing the Baldwin effect would work [Di Paolo et al., 2000]. Hinton and Nowlan [1987] presented a simple individual-based simulation model that demonstrated the effect. Although the model did not lead to any novel results, it gave a clear and convincing demonstration of the Baldwin effect and sparked a large wave of exploration in the area [Di Paolo et al., 2000].

Individual-based simulation models present one of the main research methods in many sciences studying complex systems including ecology [Di Paolo et al., 2000; Grimm et al., 2005] and Artificial Life [Kawata & Toquenaga, 1994]. As such, they are a well suited tool to study concepts central to this thesis.

Chapter 3:

The LiveGraph framework for exploratory data analysis and visualisation

LiveGraph is a software framework for real-time data visualisation and exploratory analysis that incorporates features not available in other packages. Initially, it has been developed to facilitate research presented in this thesis. However, since its publication, LiveGraph has been used by thousands of people and in several academic peer-reviewed publications. Thus, the framework and its software tools are in themselves an important contribution to the area of scientific modelling. This chapter provides an outline of the frameworks and its capabilities.

Any scientific progress depends on the availability of appropriate tools and methods required to achieve it. Thus, not only the discovery of new insights, but also the development of new tools that make such insights possible must be considered important contributions to the respective scientific areas. Using computer simulations, in particular individual-based computer simulations, for research in large-scale natural systems is a methodology that is still young (chapter 2), and the set of tools and methods has not yet reached the level of maturity and standardisation, as it is the case for some traditional sciences.

One important contribution of work described in this thesis is the development of the LiveGraph framework. LiveGraph is a software system directed towards efficient exploratory analysis of data produced by computer simulations. It has been developed in order to facilitate the research presented in this thesis, but has been used by thousands of people since for both, research and engineering purposes. In this chapter I give a brief overview of the system.

3.1. Modelling technologies

When building simulation models, researchers have the general choice between using a specialised individual-based modelling platform (see examples in Table 3.1), or developing their models from the first principles using any suitable programming language. The decision towards one or the other approach depends primarily on the capabilities of available choices, on suitability for the specific research task, availability and compatibility of additional tools and existing modules, and on the prior knowledge and expertise of the researcher. Typically, modelling platforms provide higher-level modelling and analysis tools of varying complexity, while general purpose programming languages allow more freedom and flexibility in model design. Most models presented in this thesis have been built using the Java™ programming language. This choice reflects several considerations:

- Java™ is a general-purpose programming language that allows maximum flexibility in model design and does not lock the user into a specialised framework.
- Java software can be executed on a common office PC and on a high-performance grid without modifications. When developed wisely, the runtime-performance of Java programs is very high, far above most high-level agent-modelling platforms.
- High proficiency and prior experience of the author in Java and related technologies.

Title	Web site	Title	Web site
AnyLogic	http://www.xjtek.com	SeSAm	http://www.simsesam.de
EcoLab	http://ecolab.sourceforge.net	Swarm	http://www.swarm.org
NetLogo	http://ccl.northwestern.edu/netlogo/	Repast	http://repast.sourceforge.net

Table 3.1. Agent-based modelling platforms. Listed is a small selection of popular software suites for individual-based simulations.

3.2. Use case

Due to the nature of complex systems, early stages of modelling work require developing intuition for a system at the focus of investigation. Before a research hypothesis can be formulated and rigorous experiments can be performed, it is necessary to gain an initial understanding of the model and to explore its behaviour in many scenarios and under different parameter values.

Such preliminary analysis typically involves eyeballing techniques which rely on an appropriate visual representation of simulation data. Often, the conclusion of such analysis is that the simulation must be repeated with a different set of parameters. Simulation runs can be lengthy, taking many hours, or even several days. However, in many cases it is possible to recognise that simulation parameters must be changed already soon after starting the simulation, provided it is possible to inspect the data instantly, while the simulation is being executed.

During early work on simulation models discussed in this thesis I realised that almost every time I created a new simulation model, a new data visualisation module needed to be created, in order to allow watching how the simulation variables change during the model runtime. Typically, such data visualisation modules are graph plotters that display data directly known to the simulation, or immediately derived variables (e.g. number of agents, relative proportions of populations, fitness and so on). More advanced data analysis functions and statistical measures typically become important during later analysis stages, when rigorous experiments are performed and completed simulation runs are analysed.

Repeated development of software modules for exploratory data visualisation is an inefficient approach that requires time and resources that could otherwise be invested into the development of the primary model. Thus, the need for an appropriate, re-usable tool was apparent. Such tool needed to be generic, useful for a wide range of different modelling techniques, and applicable by researchers with varying backgrounds and levels of programming expertise.

3.3. Available alternatives

A wealth of data plotting software tools of varying capabilities are available. Some of such systems provide advanced features for comprehensive data analysis in a post-simulation phase. For instance, *GNUPlot* (www.gnuplot.info) is a very powerful plotter system featuring its own scripting language. Computer algebra and numerical computation packages such as *Mathematica* (www.wolfram.com), *Maple* (www.maplesoft.com) or *Mathlab* (www.mathworks.com) also offer advanced plotting facilities along with powerful statistical and mathematical analysis tools. However, such tools (unless directly used to develop the entire simulation) require significant effort in data preparation and provide a powerful, but slow-to-use user interface. Fast, exploratory data visualisation in real-time is *not* the primary focus of such software packages.

Extensive research revealed a number of available tools for real-time data visualisation. Most notably:

- **RealView** (www.virtualworkbench.com/realview.html): This package combines some convenient data analysis features in an easy-to-use interface. However, the software is not available for free and it is primarily targeted towards visualisation of real-time data produced by hardware devices, making it not straight forward to use with computer simulations.
- **Universal Real-Time Software Oscilloscope Library** (www.oscilloscope-lib.com): This free package provides a library that can be used for integration into 3rd-party software, and also includes a stand-alone demo version. However, its user interface lacks a straight-forward possibility for concurrently selecting and comparing many data series in an efficient way.

- RtPlot (members.surfbest.net/dbai@surfbest.net/RtPlot/): This package was in an early stage of development at the time this investigation was undertaken. The current version (April 2010) includes powerful features for real-time monitoring of simulation data, however, the user interface is not targeted towards quickly switching and comparing of different data series, particularly in cases when there are dozens of series to explore, as is often the case in individual-based models.

In summary, no suitable software for efficient exploratory data visualisation analysis was available.

3.4. Solution outline

Based on the above research of available tools, I have scoped and developed a new software system for exploratory and real-time data analysis, specifically directed towards scientific simulations that produce dozens and hundreds of data series, with thousands and millions of data points per series. The package is titled "LiveGraph". It is available (along with source code and documentation) from <http://www.live-graph.org> [Livegraph Open Source Project, 2008].

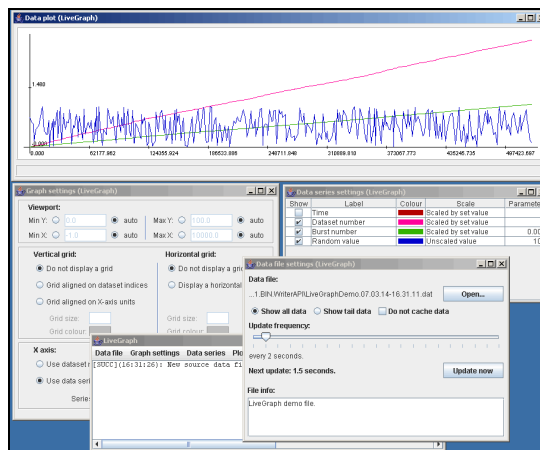


Figure 3.1. LiveGraph framework for exploratory data analysis. Depicted is a screenshot of the real-time plotter module from an early software version.

The LiveGraph framework includes the following distinctive features, not otherwise available in a single free or commercial software product:

- A real-time plotter that can automatically update graphs of data while it is still being computed by an application.
- Concise and simple point-and-click interface that allows quickly selecting and comparing data series even in applications that output over 1000 series simultaneously.
- Transformation of data series for visual comparison by the virtue of a single click. Transformations include: all-on-the-same-scale, linear, logarithm, unit interval, and time-plot transformations.
- The framework is Java-based and can run on any computer system. However, it is easily integrated with applications written in any programming language. Examples and tutorials are provided on the web site.

-
- LiveGraph reads files in a simple CSV-style format. For applications developed in Java, LiveGraph additionally provides an API that handles all data logging and persistency issues.

The LiveGraph framework consists of three major components: *a real-time plotter program, an open data file standard and several integration APIs.*

Real-time plotter:

The key component of LiveGraph the real-time plotter application ([Figure 3.1](#)). The plotter reads the LiveGraph data file (or any other CSV file) and plots the data series contained therein on the screen.

The plotter has two distinctive features which are not available elsewhere in this combination:

- LiveGraph can automatically re-examine the data file at short time intervals (anything between 100 Hz and once per hour) and update the graphs on the screen in real-time. By using specialised caching techniques specifically developed for this purpose, LiveGraph uses only very few computational resources for the updates, even if the data file grows very large and fast (e.g. several gigabytes).
- LiveGraph has an efficient graphical user interface optimised for 100s of data series. While many plotters trade off powerful functions for a complex, slow-to-use user interface, LiveGraph offers only a few data transformation functions and makes them available by a single click in the main legend window. LiveGraph also automatically detects data series and their headings in the data file and provides a very concise user interface for selecting the graphs to display by virtue of a single click. This allows keeping an overview over hundreds or even thousands of graphs.

Open data file definition:

The LiveGraph framework provides an open standard for data files containing transferable data logs. The file format is closely based on the comma separated values (CSV) text file format. LiveGraph defines a few small but important extensions, such as data column header definitions and file description annotations.

This open format definition allows creating LiveGraph compatible data files in any programming language extremely easily.

In addition, this format is readily understood by other frameworks that can read CSV data, including the standard office software.

Integration APIs:

LiveGraph provides an open and thoroughly documented API. This allows researchers and programmers to seamlessly integrate any of LiveGraph's functionality into their own software. LiveGraph can be integrated on several levels:

- Integrate LiveGraph into any (non-Java) application. LiveGraph can be customised via settings files and command line parameters. Any application can start a LiveGraph process and automatically set up all options that are otherwise available through the user interface. In addition, the simplicity of the data file format makes it straight forward for developers who use programming languages other than Java to create appropriate data files using their own routines.
- Writer & logger API for Java. This API provides application developers with a flexible, fast and easy way for writing data generated by an application into LiveGraph CSV files. The API offers advanced functions such as caching and log file management.
- Application interface. Java developers have the option to start LiveGraph as a separate process or to start it within the same process as the main application. The latter allows taking advantage of the application interface exposed by LiveGraph. The application

interface provides an API to control a LiveGraph instance while it runs. This includes full real-time control over all settings.

- LiveGraph plotter API - LiveGraph's user interface can be fully integrated into a Java application or into a web-applet. The API exposes all settings windows as well as the plot window allowing integrating a LiveGraph instance seamlessly into 3rd party user interfaces. Data can be passed through memory and without a detour via a hard drive file.
- LiveGraph is open source. It is well documented and commented. Researchers and developers can modify and re-distribute the source code according to their needs.

3.5. Impact

I made LiveGraph first publicly available in March 2007. Since then, I released several updates and fixes. The currently publicly available version was released in late 2008, however, future updates are planned for 2010. As of April 2010, LiveGraph has been downloaded over 27500 times in total, averaging at about 900 downloads over the last 12 months (Figure 3.2).

As a tool for analysis and visualisation of scientific simulation data, the LiveGraph framework has made a significant contribution. This is signified by the large number of downloads and active users. The framework was presented at an international conference for complex systems [Paperin, 2007], and became an essential tool that facilitated work explored in this thesis, as well as related and unrelated work of many of my colleagues. The framework has been applied in a number of engineering contexts (see web), and has been referenced by authors who are not affiliated with me in several peer-reviewed publications [e.g. Escobar, 2007; Sain et al., 2009a; 2009b; Tadele, 2009; Jung & Lim, 2010].

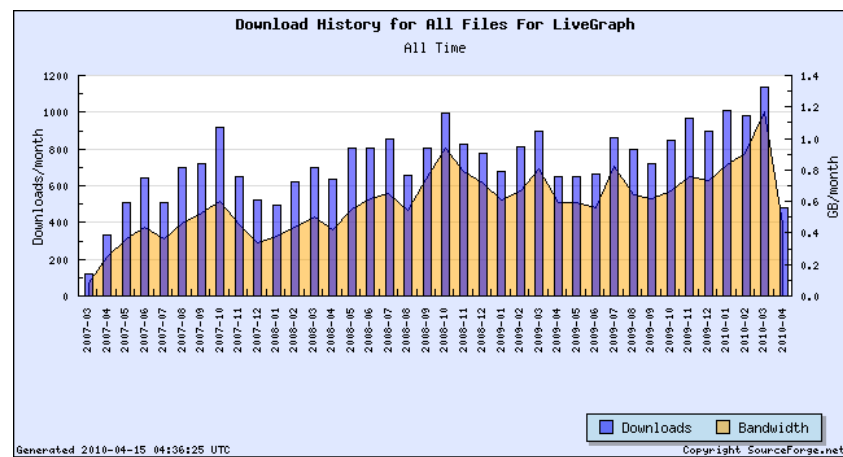


Figure 3.2. LiveGraph download statistics. The current statistics can be obtained from:
http://sourceforge.net/project/stats/detail.php?group_id=191061&uqn=live-graph&type=prdownload&mode=alltime&file_id=0



Part II

Background

This part provides the context for the research presented in this thesis. I provide a general overview of the research areas that deal with the key questions tackled in this thesis, review the background literature and describe the key concepts.

Chapter 4: Complex Systems.

Complexity science is an interdisciplinary research area that studies the general properties of natural and artificial systems that exhibit properties such as non-linear dynamics, adaptation, self-organisation, perpetual novelty and emergence. I review some of the key discoveries of complexity science and describe some of the key concepts, in particular the concept of emergence.

Chapter 5: Networks in Complex Systems.

Complex systems are composed of many components that interact in intricate patterns which can be represented as networks. Topological properties of these networks have profound effects on the dynamics of the systems they underlie. Such effects are one of the foci of this thesis. I thus provide a brief review of network theory from the perspective of complexity research and highlight the key relationships between widespread network structures and complex system dynamics.

Chapter 6: Speciation theory and biological fitness landscapes.

The multiplicity of species – reproductively isolated populations – is a primary example for biological variety and diversity. Mechanisms leading to speciation differ according to type of reproductive isolation and to geographical factors. A key concept in evolutionary speciation theory is the fitness landscape. Recent research reveals that many biological fitness landscapes are well represented by nearly neutral fitness networks in genotype space. Such fitness landscapes have been termed holey. This chapter briefly develops some key biological concepts used in this thesis, including common modes of reproductive isolation and speciation. The concept of fitness landscapes, in particular, holey fitness landscapes, is developed from the perspective of speciation theory.

Chapter 4:

Complex Systems

Complex systems are systems that display emergent properties that cannot be fully understood by a reductionist analysis of system components. Complexity science is a research agenda that uses inter-disciplinary approaches and new methodologies to fill this gap. This chapter sets a general context for the work presented in this thesis. It presents a brief overview of complexity theory and its hallmark achievements, and provides a critical discussion of the phenomenon of emergence.

4.1. The science of complex systems

Since the ancient Greeks western science has been dominated by strict, hierarchical separation of subject domains. The dominance of the reductionist approach to the world lead to a an increasing separation of research areas into a rigid classification with little cross-fertilisation [Anderson, 1972]. This has been a fruitful approach as long as the systems studied were relatively simple and the experts could benefit from a concentration on their domain and from non-interference caused by consideration of phenomena “outside the scope of interest”. As a result, different sciences have developed incompatible terminologies for similar phenomena, reinvented similar theories and introduced similar methodologies under different names, thus impeding inter-disciplinary exchange [Hjorland & Pedersen, 2005; Szostak, 2008].

However, as modern science embraces larger systems in more dynamic and complex environments, the drawbacks of traditional scientific reductionism and rigid disciplinary ontology are increasingly recognised. Systems such as landscapes and ecosystems, insect colonies, organisms, cells, human economies and organisations, telecommunication networks and many others including our planet and our civilisation cannot be understood as a whole by reducing them to their constituents [Anderson, 1972]. The complexity of these entities and their insusceptibility to traditional techniques call for new inter-disciplinary approaches and non-traditional methodologies. Such approaches, directed towards cross-fertilisation between natural, social, mathematical and computational sciences contributed to a new thinking paradigm and a new research agenda of *complexity science*. Complexity science aims to draw together ideas from many different disciplines into a coherent and unified methodology. Complexity research has revealed many deep similarities between superficially unrelated systems and processes. For example, epidemics, wildfires and nuclear fission differ enormously in terms of the objects involved, yet they can all be usefully treated as instances of percolation [Isichenko, 1992; Grimmett, 1999].

The term *complexity* has been used in various contexts. For instance, Manson [2001] breaks the term down into three major divisions: algorithmic complexity, deterministic complexity, and aggregate complexity.

Algorithmic complexity expresses the difficulty to describe a particular system. The mathematical complexity theory, for instance, expresses the effort required to solve a mathematical problem [Hartmanis & Stearns, 1965]. Other information-theoretical definitions treat complexity as the size of the simplest computer program that can reproduce the behaviour of a given system (e.g. [Solomonoff, 1964; Kolmogorov, 1965; Wallace & Boulton, 1968; Chaitin, 1969]).

Deterministic complexity is a notion that incorporates a number of concepts including feedback controls, sensitivity, bifurcation, deterministic chaos and fractals [Manson, 2001]. A well-known notion in deterministic complexity is the “Butterfly Effect”,- the realisation that a flap of a butterfly’s wings can affect a distant major weather system through a series of chained consequences, or, in general, that small changes in initial conditions can lead to major changes in system behaviour. The technical repertoire of deterministic complexity is dominated by non-linear and fractal mathematics and statistics.

Aggregate complexity concentrates on how individual elements work together to create larger systems with complex behaviour. It emphasises the importance of component interactions, dynamic structure and the surrounding environment for the behaviour and the emergent properties of complex systems. Aggregate complexity is primarily concerned with mathematically untreatable aspects of natural and artificial systems. Researchers employ traditional along with non-reductionist approaches and rely on computational methods and simulations as analysis tools [Gligor & Ignat, 2001; Ehrhardt et al., 2005]. This view of complexity is increasingly at the focus of complexity research and shall also be the main focus of this thesis.

In this context, the etymological roots of the word *complexity* are curious. The word originated from the Latin *complexus*, which comes from the Greek *pleko* meaning to twine or to plait. Philologically, a complex system is thus something consisting of interwoven parts [Lenton & Van Oijen, 2002].

The above classification lays out various approaches to defining a complex system. Here, I concentrate on aggregate and emergent aspects of complexity. Various similar definitions of complexity concentrating on these aspects have been given (e.g. [Coveney & Highfield, 1996; Dent, 1999; Goldstein, 1999; Casti, 2009]). Here, I follow the definition provided in Paperin et al. [2010]:

Complex systems are systems that are composed of many interacting components and display behaviour that is difficult to explain through an analysis of the systems' constituent parts. Such behaviour is called "emergent". Complex adaptive systems (CAS) are complex systems that have the capacity to adapt to their environment through an evolution-like process.

4.2. Key outcomes of complexity science

In this section I provide a brief overview of the most important achievements and seminal works of complexity science, roughly grouped by the cradle sciences of complexity. A comprehensive account of contributions to complexity science originating from all the relevant disciplines is far beyond the scope of this work. The current overview is merely intended to sketch out the broad scope of complexity research. The choice of literature cited here largely follows a citation frequency analysis by Heylighen [1997] with some focus on work that is particularly relevant to the issues discussed in later chapters. More comprehensive reviews are provided by [Manson, 2001], [Heylighen, 1997], [Dent, 1999], as well as [Coveney & Highfield, 1996].

4.2.1. Natural sciences

On the search for biochemical origins of life, Eigen introduced the concept of autocatalytic *hyper-cycles* of chemical reactions that contain other autocatalytic cycles, and the concept of *quasispecies* – groups of self-reproducing genotypes or chemical molecules that are subject to very high mutation rates [Eigen & Schuster, 1979].

The modern understanding of our living environment is largely due to Darwin's *theory of evolution by natural selection* [Darwin, 1859]. It took seventy years until the evolutionary theory received an in-depth mathematical underpinning by Fisher [1930]. The metaphor of a biological *fitness landscape*, a visualisation of individuals' and populations' adaptedness to their environment, was coined by Wright [1932]. Eldredge and Gould [1972] have showed by an analysis of the fossil record that species evolution on earth underwent a pattern of *punctuated equilibria*, i.e. long periods of stasis punctuated by bursts of change. An explanation for this pattern remains a debated research topic and work presented in this thesis may shed additional light on this question.

Many principles and laws that are apparent in biological systems appear to be generic and are known to govern other complex systems:

Prigogine along with colleagues demonstrated that systems can exist in a non-equilibrium state as they *self-organise* at the expense of increasing entropy on the outside of the system [Nicolis & Prigogine, 1977b; Prigogine, 1980]. Such systems are called *dissipative*. This insight allows drawing important parallels between biological CAS, which are known to develop internal order, and thermo-dynamical systems.

Research into non-linear physical systems lead to the concept of *deterministic chaos*, a deterministic process that is extremely sensitive to initial conditions [Poincaré, 1890; Crutchfield et al., 1986]. Closely related is *catastrophe theory* developed by Thom [1975] and others. Catastrophe theory provides a mathematical treatment of dynamical systems that can exhibit large-scale changes in response to small-scale triggers. Among others, chaotic systems were studied by Feigenbaum [1983] who coined the term *bifurcation*, a mathematical concept describing a drastic change in qualitative behaviour in response to a small change of control variables.

A related area of mathematics that bears key importance for complexity is the field of *fractal* mathematics, founded by Mandelbrot, that studies *self-similar* entities of "broken" dimensionality. Mandelbrot [1983] has noticed that many natural objects exhibit self-similarity, e.g. the structure of a tree trunk and the main branches is reflected on scales of small branches through to leaf veins. Fractals are mathematical objects in which similar patterns occur on different scales, often, according to *power laws*. Power laws are also widely found in natural systems. Bak introduced the term *self-organised criticality* that describes systems that have a critical attractor and produce power-law distributed data [Bak et al., 1988; Bak, 1999].

4.2.2. Cybernetics

Insights gained from studying specific natural complex systems have been abstracted and generalised by *cybernetics* and *general systems theory*. Traditionally, systems theory focuses on equilibrium systems. Although complexity science's prime focus is on non-equilibrium dynamics, it has been strongly influenced by cybernetics. For instance, Ashby [1964] introduced the law of *requisite variety*, a notion that dynamical systems always self-organise and that any system regulator seeking to stabilise a system, is also required to be (or incorporate) a model of that system. Von Foerster's *second-order cybernetics* principle [Von Foerster, 1981] states that many self-regulating complex systems are bound to change themselves by the very process of self-modelling and self-regulation. His crucial work on self-organisation and self-reproduction [Von Foerster, 1960] has been extended by other's work on autonomous reproducing systems (e.g. [Varela, 1979]). Bateson [1972] established several crucial parallels between natural evolution and human mind.

4.2.3. Computer science and Artificial Life

The necessity of simulation tools for the study of complex systems along with increasing availability of computing power lead to a strong interest in the field by computer scientists and information theorists. Classical *information theory*, coined by Shannon [1993], is a basis for various formal treatments of complexity. Ideas expressed by the economist and philosopher Friedrich Hayek, the psychologist Donald Hebb and other researchers have eventually lead to the construction of *artificial neural networks* [Bishop, 1995]. Insights from biology lead to the field of *evolutionary computation*, of which *genetic algorithms* [Holland, 1992; 1995] are perhaps the most widely known example. Kauffman's theoretical and computational work on Boolean abstractions of gene regulatory networks and on fitness landscapes [Kauffman, 1993] resulted in a further argument that an ability to self-organise is a required factor in open-ended evolutionary systems.

The successful application of simulation experiments to complexity lead to the emergence of a new discipline of *Artificial Life* [Langton, 1986; Bedau et al., 2000; Bedau, 2002]. Artificial Life is concerned with the development of artificial systems (either in software or in hardware or in "wetware") that exhibit life-like properties. This new science lead to important results and new understanding of evolution (e.g. [Ray, 1991; Adami & Brown, 1994]), cognition (e.g. [Steels & Brooks, 1995; Sadedin & Paperin, 2009]) and other processes in biology and in human societies (e.g. [Axelrod & Hamilton, 1981; Palmer et al., 1994]).

A popular modelling methodology for chaotic systems used by Artificial Life researchers are *cellular automata* [Von Neumann, 1966], a spatially explicit model of a discrete process, commonly used in conjunction with computer simulations. Wolfram [1984; 2002] developed a classification of cellular automata according to their dynamical behaviour and potential for universal information processing capability. Langton [1990; 1991] used cellular automata to study self-organised criticality and coined the term *edge of chaos* to describe their critical attractor. Critical phase transitions have previously been described by Erdős & Rényi [1960] in their analytical *random network model*.

At the end of the first decade of the new millennium, complexity researches increasingly recognise the importance of network theory in explaining many complex phenomena [Albert & Barabasi, 2002; Barabasi, 2003; Newman, 2003]. The role of networks, specifically network connectivity, in system dynamics is a central topic of this thesis. Chapter 5 provides a further context for this work through a brief review of relevant network-theoretical research.

4.3. Emergence in complex adaptive systems

The definition of a complex system given in [section 4.1 above](#) is used in that or a similar form by many authors (e.g. [Coveney & Highfield, 1996; Levin, 1998; Dent, 1999; Goldstein, 1999; Lenton & Van Oijen, 2002; Casti, 2009]). However, a critical consideration of this definition and of the phenomenon of *emergence*, which is central to the definition, is useful for a better understanding of the term's scope.

Dynamical systems can be classified in terms of their behaviour into simple, chaotic, ordered and critical systems [Lenton & Van Oijen, 2002]. *Simple systems* (i.e. not complex systems) are systems whose properties can be fully explained in terms of their components (e.g. a pendulum). *Chaotic systems* exhibit unpredictable behaviour described by non-linear equations that model their dynamics, giving high sensitivity to initial conditions. *Ordered systems* exhibit a tendency to reside in one state of a limit cycle of a small number of states. *Critical systems* exhibit behaviour "between" the frozen balance of ordered systems and the unpredictability of chaotic systems. These are the systems that commonly exhibit emergence and self-organisation [Langton 1991; Kauffman, 1993; Bak, 1999]. Complexity theory and Artificial Life can be seen as extensions of control theory and cybernetics, which focus on predictable feedback and feed-forward systems. Conversely, chaos theory focuses on systems with unpredictable behaviour [Lenton & Van Oijen, 2002].

The concept of emergence has been defined differently by various researchers from different cradle sciences of complexity and thus there is some philosophical controversy about the precise meaning of the concept [Kawata & Toquenaga, 1994]. For instance, in ecology the term is related to the debate of holism versus reductionism [Wilson, 1988]. A possible definition is that emergent properties are properties that are inherently not derivable from the parts of a system [Allen & Starr, 1982]. Generally, the term of emergence implies the appearance of higher-level phenomena that are "unexpected" from the organisation, the behaviour or the rules at lower levels [Lewin, 1992].

The use of terms such as "unexpectedness" or "unpredictability" to define emergence has been criticised as they may arguably represent not an a priori property of a phenomenon, but merely our inability to explain it, for instance due to insufficient knowledge or lack of data [Edson et al., 1981; Allen & Starr, 1982]. Gödel's incompleteness theorems [Gödel, 1931] imply that in formal systems, for any set of rules (axioms) there are true statements that cannot be proven on the basis of those rules, and that for some true statements it cannot be determined in finite time whether they are provable under any given set of axioms. The philosophical implication for natural systems is that for some phenomena it may be impossible to determine whether they are emergent.

What are, specifically, these properties and behaviours that we call "emergent"? A comprehensive list of such properties is inevitably arguable, but there is a maturing consensus in the literature about some of these properties [Kauffman, 1993; Holland, 1995; Levin, 1998; Manson, 2001; Lenton & Van Oijen, 2002]:

- Sustained diversity of local components
- Non-equilibrium dynamics in the absence of a global controller
- Perpetual novelty
- Permanent adaptation
- Modular and hierarchical organisation across many orders of magnitude
- Resilience and robustness towards disturbances

A common attribute to all of the above properties is that they are observed in systems that are capable of adapting to their environment. In fact, Levin [1998] defines *complex adaptive systems* (CAS) as systems that exhibit the above properties. Some researchers (e.g. [Langton 1991; Kawata & Toquenaga, 1994; Taylor, 2003]) go further to suggest that a necessary condition for emergence is adaptation and self-regulation through feedbacks across different levels of magnitude. Thus, a global phenomenon is emergent from the local dynamics of a set of interacting components when such phenomenon adjusts its own environment by manipulating the local dynamics from which it emerges.

A hallmark concept in emergent self-regulation is the Gaia theory [Lovelock, 1972; Lenton & Van Oijen, 2002]. Gaia is postulated as a “living” complex entity that includes the planet Earth along with its atmosphere, oceans, soil and the biosphere – a complete control system that seeks an optimal environment for life and maintains it in a homeostasis through feedback and feed-forward interactions. Gaia’s biota maintains a complex structure and constantly modifies its physical and chemical environment to improve and sustain conditions suitable for life. This happens in spite of significant disturbances such as meteor impacts, increase in the sun’s luminosity or changes in earth mantle.

Gaia is an open system. The second law of thermodynamics implies that true self-organisation cannot occur since growing order is possible only in thermodynamically open systems. Thus, Gaia as well as any self-organising system must be dissipative, taking in energy and putting out entropy [Lenton & Van Oijen, 2002].

The Gaia theory has been challenged to explain how global-scale order and self-regulation can arise without planet-level selection or teleology (global controller) [Doolittle, 1981]. Such queries led to the formulation of the Daisyworld model [Lovelock, 1972; Watson & Lovelock, 1983] – a simple model that shows how self-regulation can arise on an imaginary planet that contains populations of only two species: black daisies and white daisies. Black daisies absorb sunlight and thus heat their local environment; white daisies reflect sunlight, which has a cooling effect. All daisies can only exist when the environment temperature is within a certain interval suitable for life. The Daisyworld planet orbits a sun that steadily increases its luminosity (as all stars do [Lovelock, 1972]) and heats the planet. Thus, the temperature of a bare planet is suitable for life only during a short interval of the planet’s entire life span. However, the Daisyworld planet not only significantly widens the time interval where life-suitable temperature is sustained, but also maintains the temperature at a near constant optimal level: Initially, Daisyworld is dominated by black daisies that retain extra heat from the young star and improve their living conditions. As the star brightens, Daisyworld becomes dominated by white daisies that cool the planet to optimal living conditions. No global controller is involved – the daisies are subject to selection according to their local environment only. A large number of variations of the original Daisyworld model have been created to explore more complex scenarios and self-regulation has repeatedly been found in many cases (see [Wood et al., 2008] for a review).

Daisyworld is a very simple model of Gaia. Should the self-organisation that arises in the model be called emergent? The model exhibits cross-level feedback between daisies which modify their local environment and the global planet climate, as well as adaptation to the external environment – the heating star. But does the temperature regulation process qualify as “unexpected”, let alone “unpredictable”? The model is not sufficiently complex to include a number of other typical CAS properties, for instance perpetual novelty or hierarchical organisation. Whether self-regulation in Daisyworld is truly emergent may remain arguable [Lenton & Van Oijen, 2002]. Here, I qualify specific differences between the dynamics of Daisyworld and Gaia (i.e. the planet Earth).

Complex systems can be described in terms of their order dynamics and control [Jørgensen & Straskraba, 2000; Lenton & Van Oijen, 2002]:

- 1) First-order dynamical systems have fixed parameters and fixed structure. The properties of system components and their interactions do not change. Thus, such systems can only achieve feedback control. An example is the original Daisyworld model described above.
- 2) Second-order dynamical systems have variable parameters and variable states, but fixed structure. They can adjust to external conditions through self-adaptation of parameters of their internal functions. For instance, second-order control can be enabled in the Daisyworld model by providing a mechanism for black and white daisies to adjust the amount of temperature absorbed or retained (this is fixed for each type of daisies in the original model).
- 3) Third-order dynamical systems have variable structure as well as variable parameters and states. They can self-organise by varying their structure over time as a response to changing state or external conditions. An example is a social group where individuals can leave and join and where different collaborative interactions between individuals can appear and disappear over time.

- 4) Fourth-order dynamical systems are systems where changes in parameters, states and structures lead to changes in the goal functions that affect selective pressures and determine the long-term trajectory of the system. Such systems can thus exhibit evolutionary dynamics. Examples include real large-scale ecosystems, human societies and economies, as well as Gaia itself.

Third-order, and particularly fourth-order systems tend to exhibit complex feedback interactions between control mechanisms of various orders. Changes in internal structure can affect evolutionary constraints. New evolutionary constraints can in-turn affect lower-level control mechanisms and facilitate changes in properties of individual system components and in their interaction structures [Lenton & Van Oijen, 2002].

One of the key questions in Artificial Life is to understand the requirements of open-ended evolution and to create a system capable of open-ended evolution in a computer [Bedau et al., 2000; Green & Newth, 2001]. Although limited progress has been made towards this goal (e.g. [Ray, 1991; Adami & Brown, 1994]), no artificial evolutionary system so far has been able to demonstrate all of the key emergent properties listed earlier. I believe, that an issue inherent in most models intended to reproduce life-like evolution is that they concentrate on providing mechanisms for some of the above emergent properties, while neglecting others. Many such models fail to allow for fourth-order control mechanisms [Kawata & Toquenaga, 1994].

First and second-order dynamical systems can usually be fairly well understood. Even when a formal analysis appears intractable, such systems can usually be modelled using different simulation techniques [Jørgensen & Straskraba, 2000]. Multi-level self-regulation can arise already in first-order control systems (e.g. Daisyworld). Whether or not such phenomena are unexpected and thus emergent remains to be settled by science philosophers. Here, I use the term emergence loosely for any complex process in which not all causal mechanisms are understood, either due to an inherent inability to understand a process with available methods, due to lack of data, or merely due to insufficient concentration of research efforts to date. In a sense, the central focus of this thesis is to contribute towards limiting the abundance of emergence in our world: by identifying and describing some of the causal mechanisms behind emergent phenomena.

Chapter 5:

Networks in Complex Systems

Networks are inherent in all complex systems, and patterns and processes that occur within the underlying networks play a fundamental role in the emergence of system wide properties. Networks play different roles in complex systems models and can be used to describe system structure as well as system dynamics. Particular topological properties of underlying networks are associated with specific large-scale behaviours. This chapter presents a brief review of network theory from the perspective of complexity science and provides a specific context for key questions approached in this thesis.

Material presented in all sections of this chapter is based on the following publications:

 G. Paperin (2010):
The Role of Networks in Computational Complex Systems.
International Transactions on Systems Science and Applications. In Press. Ref: [Paperin, 2010].

A *network* (also *graph*) is a collection of entities with connections between them. The entities are termed *vertices* or *nodes*. The connections are termed *edges* or *links*. Networks are abundant in the world. Examples include networks of social acquaintances, communication networks (e.g. the internet), transportation networks (e.g. a road network) organisational networks, ecological food webs, gene interaction networks, neural networks and many others [Newman, 2003]. A central point of this thesis is to show that networks influence and govern a wide range of emergent properties in complex systems.

In this chapter I briefly review the current state of research in network theory. I concentrate on the role of networks in complex systems. A complete review of network theory is outside the scope of this work. Comprehensive accounts of general network theory including basic terminology and mathematical models can be found in review papers such as [Newman, 2003] or [Albert & Barabasi, 2002], and in textbooks on graph theory (e.g. [Biggs et al., 1986; Gross & Yellen, 2003]).

5.1. Background

Network theory has historically emerged as a branch of graph theory [Newman, 2003]. Science historians (e.g. [Biggs et al., 1986]) traditionally attribute the roots of the graph theory to Euler's seminal 1735 presentation on the *Seven Bridges of Königsberg Problem* (published in [Euler, 1741]). However, the term "graph" has only been introduced by Sylvester [1878] in a paper that brought together related ideas from algebra and chemistry. Since then it became increasingly apparent that networks underlie a wide range of systems and processes. Green [1993] has shown that systems studied by complexity science are isomorphic to directed graphs. This realisation led to an increasingly fruitful paradigm of abstracting complex systems as networks of simple elements. It is now recognised that studying the topological properties of networks provides important insights into the functioning of complex systems, and that such studies are essential steps towards general theories of complexity [Bedau et al., 2000; Alon, 2003].

The mathematical graph theory is a fundamental notion in discrete mathematics [Newman, 2003]. Beginning in the second half of the 20th century, a new paradigm in network research has emerged: Previously, researchers focused on small graphs and the properties of individual nodes and edges. The availability of modern-day computing power, storage and communication allows gathering and process previously unseen amounts of data. This allowed studying large-scale topological and statistical properties of networks containing thousands and millions of nodes. The new approach led to a change in the kind of questions asked by network scientists: For instance, efficiently finding the shortest path length

between two given nodes is a key problem in a small graph; but in a graph consisting of millions of nodes, efficiently estimating the mean shortest path length between any two vertices can be more relevant. Similarly, in a small graph an analyst may want to determine the particular node that is most critical for the graph connectivity if removed; in a giant network it may be more relevant to determine the proportion of nodes that can be removed before the connectivity structure of the network is significantly affected in some specific way [Newman, 2003].

Modern network theory has three important foci [Barabasi, 2003; Newman, 2003]:

- Identifying important statistical properties (e.g. degree distributions, connectivity densities) and describing appropriate measures.
- Creating models to investigate how important properties arise and how they relate to each other. For instance, random networks with high connection density tend to exhibit low average path length.
- Understanding how the behaviour of networked systems is affected by specific network properties and why specific properties are dominant in certain kinds of natural networks. For instance: How does the average path length affect the stability and efficiency of communication networks? Or why do degree distributions in many social networks follow a power law?

Although significant research potential remains in all three of the above themes, network theory has made significant progress in the first two areas, and studies of the effects of network structure on system behaviour bear the largest potential for new discoveries [Newman, 2003]. One of the central topics of this thesis is to investigate how changes in network connectivity affect the dynamics of complex systems.

5.2. Roles of networks

In the context of ubiquity of networks it is important to distinguish the different *roles* that networks can play within systems. As the term implies, a particular network can assume different roles in different situations. Here, I distinguish 4 general roles: component interaction networks (primary and secondary), information networks and state transition networks.

5.2.1. Component interaction networks

Complex systems are systems that are composed of many interacting components ([chapter 4](#)). Thus, *interaction networks* underlie all complex systems [Green, 1993]. *Nodes* (vertices) in such networks represent system components, and *links* (edges) represent pair-wise interactions between the components. Broadcast-style interactions can be represented as several pair-wise interactions.

I refer to *interactions* as “local” or “global”. *Local interactions* are interactions between nodes within a defined neighbourhood. There are various possible ways of defining a neighbourhood relation for nodes, including the nodes’ spatial location in the real world, node similarity, or affiliation to a common subsystem. Many systems contain several distinct networks and the scope of some interactions can be described in terms of another network. For example, in an online social network, where nodes represent individuals and edges represent friendship ties, an email communication between two friends can be seen as a local interaction and an email between unconnected participants without common friends can be seen as distant or global interaction. In this sense, I use the term *global interactions* for pair-wise interactions between any nodes in the network, that do not belong to the same neighbourhood.

5.2.2. Primary and secondary interaction networks

When studying the role of networks in complex systems, it is crucial to keep in mind that in most natural systems networks do not exist *a priori*. Much rather, they are analytical constructs that our mind creates in order to make sense of structures present in complex systems. For instance, it is common to speak of one’s friends as of a social network. However, a person’s group of friends is not connected to each other in any physical way. Rather, the social friendship network is a mental construct formed on the basis of opinions and attitudes people may have towards their friends and acquaintances. Some systems, however, possess an obvious physical network structure. For instance, a circuit board is a physical network of electronic components connected by conducting links. Notably, however, a circuit board can contain several other “less obvious” physical or non-physical networks, for instance a network of sub-circuits that perform common tasks, a network carrying clocking signals, a network carrying data signals (for digital boards), or a network of cooling air flow across the board.

For many studies it is convenient to differentiate between *primary* and *secondary* interaction networks. A primary network is a network that for the purpose of a particular study describes the fundamental structure of a system. Other interaction networks in a system are secondary. Consider again the example of an online social network, where nodes represent individuals and edges represent friendship ties. A study may consider this a primary network and study several secondary interaction networks that arise within the system (e.g. who emails whom – communicative interactions, who personally knows whom – off-line interactions and so on). However, a different study may represent the same system on the basis of a different primary network. For instance, communication may represent the primary interaction (nodes still represent individuals, links connect people who exchange email). A study may then ask how communication affects friendship ties, i.e. the friendship network is secondary in this context.

In summary, the distinction between primary and secondary interaction networks is not an inherent property of a particular system, but of the model used to study it. As complex

systems can contain many different interaction networks, a clear conceptual distinction between primary and secondary networks can be crucial for formulating concise and useful models (e.g. see simulation models in [part V](#)).

5.2.3. Information networks

A state change in a system component can cause an interaction with a neighbouring component that can in turn lead to a chain of mutual interactions with other components. This can eventually affect components that never directly interact with the initially modified component. Although the nodes at the ends of such chain of interactions may not be *directly* connected in an interaction network, they can still exchange information. Thus, every interaction network has a corresponding information network with the same nodes, where a pair of nodes have a common edge exactly when there is a path between those nodes in the interaction network. In other words, while interaction networks imply direct interaction, information networks describe direct or indirect transmission of stimuli as a result of one or more interactions ([Figure 5.1](#)).

The distinction between information and interaction networks can relate to the order of magnitude at which a system is studied. For instance, in a study of the internet as a physical network, computers in different cities do not interact directly, even when exchanging information. However, in a study of the internet as a means of personal communication, any two computers exchanging digital messages may be considered to interact directly.

5.2.4. State transition networks

These networks describe the dynamics of a system rather than its internal organisation. Nodes correspond to system states and edges to possible transitions between the states. State transition networks are discussed in [section 5.7](#).

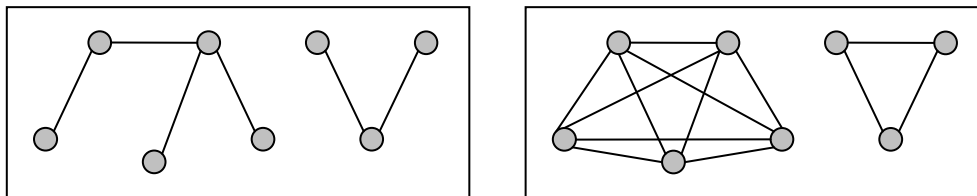


Figure 5.1. Interaction and information networks.

Left: An interaction network consisting of two clusters.

Right: The corresponding information network (same nodes). Every pair of nodes that is connected by a path in the interaction network, is connected directly in the information network.

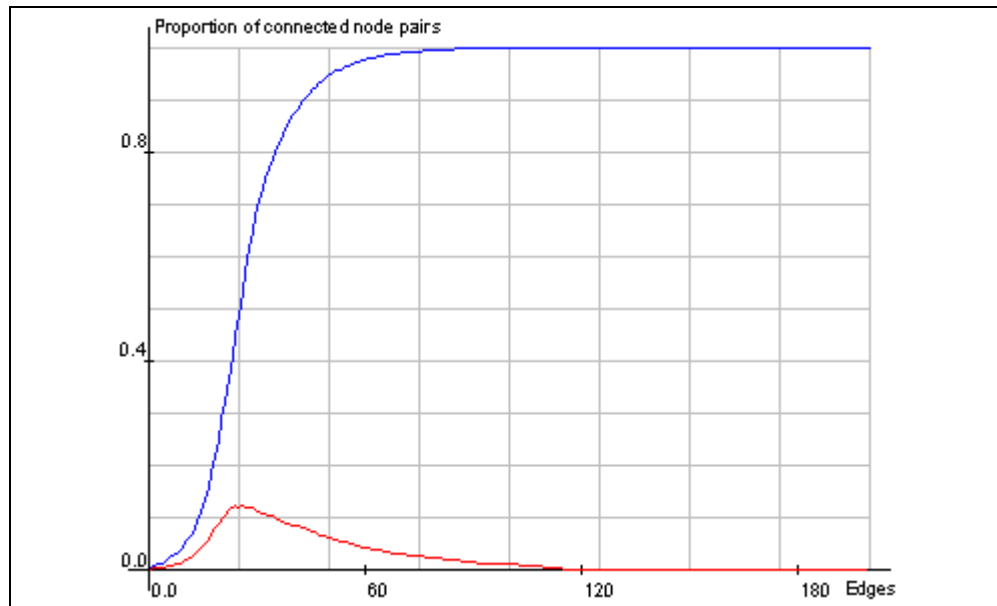


Figure 5.2. Critical connectivity phase transition in networks. Random additions of edges to a network lead to a critical transition from many separate components to a single super-component.

Top blue line: Proportion of node pairs that are connected to the same component as a function of the number of edges, averaged over 1000 random networks.

Bottom red line: Standard deviation of the proportion of connected node pairs in the same 1000 networks.

Plotted using the LiveGraph framework [Livegraph Open Source Project, 2008].

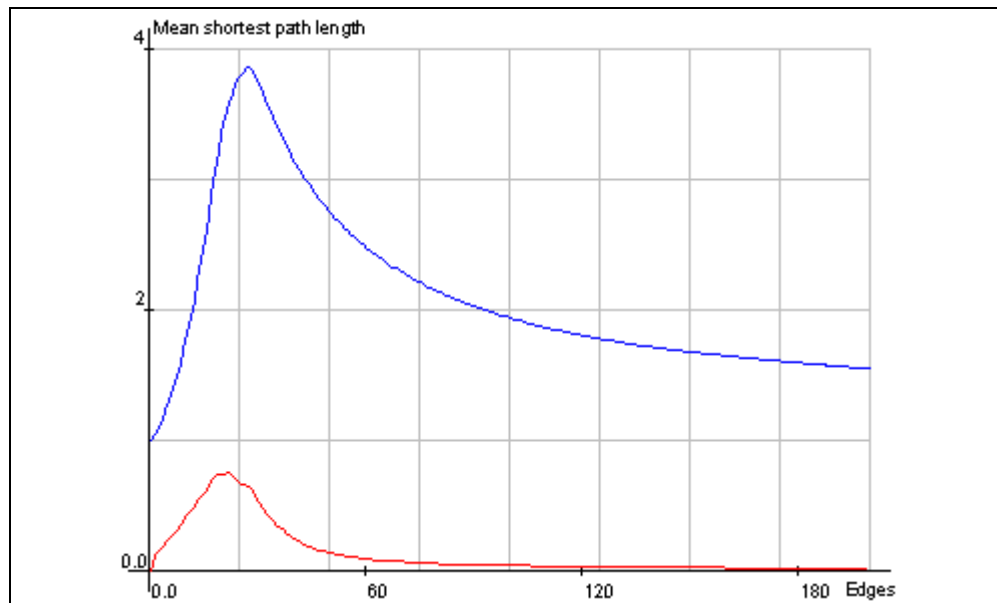


Figure 5.3. Unpredictable dynamics near critical threshold. In very sparse networks the paths length are short because only very small connected components exist. As edge density approaches the critical threshold, the network becomes largely connected, and path lengths become very long and highly stochastic. As further edges are added, path lengths become shorter and more predictable.

Top blue line: Mean shortest path length between all node pairs in a network with 30 nodes as a function of the number of edges, averaged over 1000 random networks.

Bottom red line: Standard deviation of the mean shortest path length in the same 1000 networks.

Plotted using the LiveGraph framework [Livegraph Open Source Project, 2008].

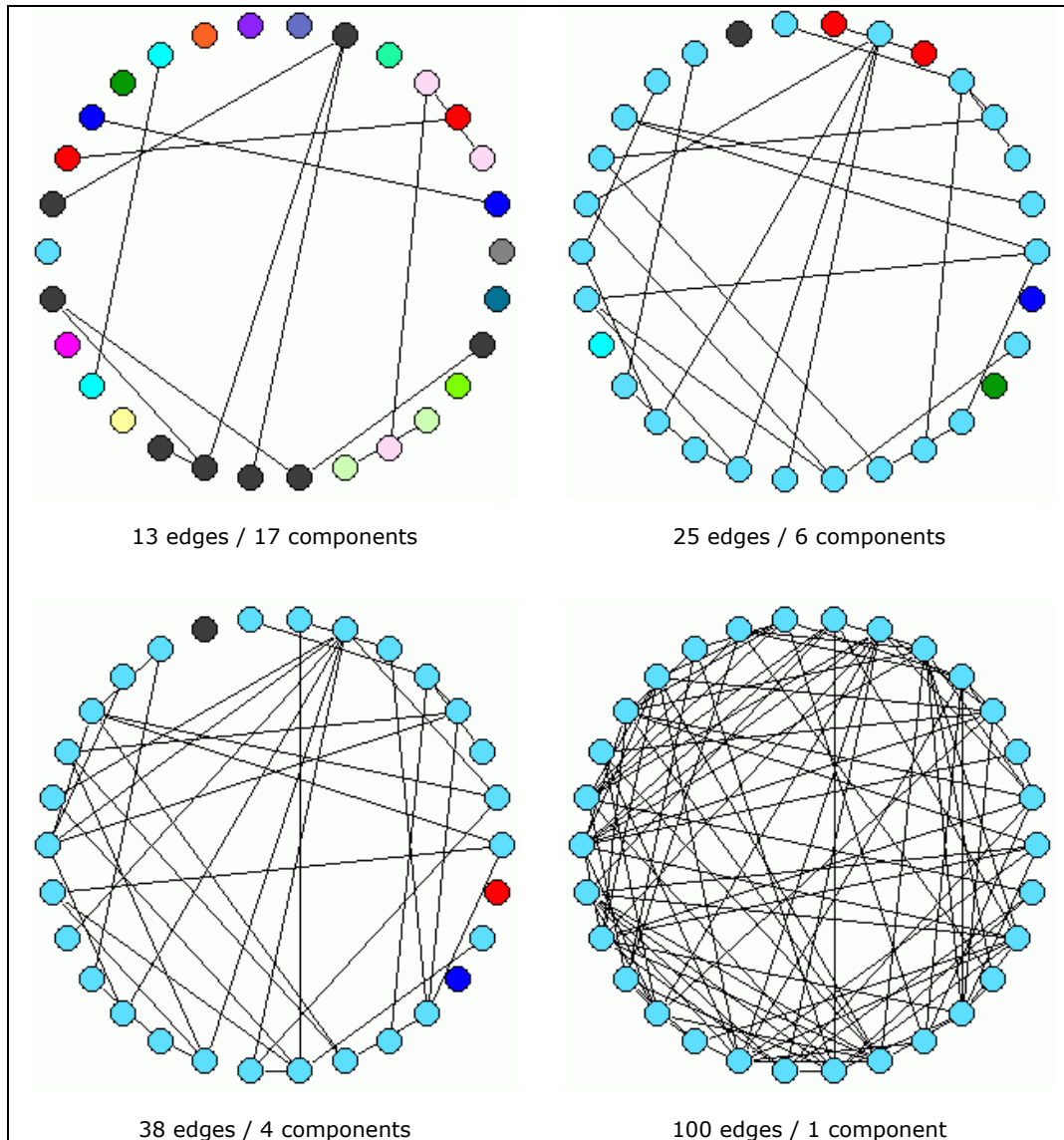


Figure 5.4. Connectivity avalanche in networks. Random additions of edges to a network lead to a critical transition from a poorly connected to a well connected state. Depicted is a typical network with 30 nodes with random edges progressively added to it. Nodes that belong to the same connected component are coloured in the same colour. As more edges are added, the network components merge.
Plotted using the Pajek software [Batagelj & Mrvar, 1996].

5.3. Phase changes and connectivity

Perhaps the most important consequence of the universality of networks is the *connectivity avalanche*. Erdős and Rényi [1960] showed that random additions of edges to a graph lead to a critical transition from disconnected to largely connected dynamics (Figure 5.4). This mechanism underlies phase changes in many systems, such as freezing of water, crystallization, and nuclear chain reactions [Schroeder, 1991]. At the *critical threshold point*, the probability of a percolation process spreading throughout the entire system (e.g. of a disease spreading throughout a population) undergoes an abrupt change from near 0 to near 1 (for sufficiently large systems). Below the critical connectivity threshold, networks typically consist of many small separate components. Above the critical density, most such components become connected into a single giant component that includes most of the network nodes (Figure 5.2). Therefore, connectivity patterns of interaction networks can determine whether stimuli propagate locally or globally in the system. When the connectivity of networks falls near the critical percolation threshold, propagation of stimuli is highly stochastic [Grimmett, 1999] (Figure 5.3).

In networks where the connectivity density is near the critical threshold, the sizes of subsystems affected by stimuli propagations follow a *power-law distribution*. That is – small effects are more common, but effects of an n -fold magnitude occur with a frequency in the order of n^{-1} . This *scale-free* effect can be observed in a variety of systems [Bak, 1999]. Bak [1988] has thus argued that many systems organise the density of their internal interactions towards the critical threshold. This phenomenon has been termed *self-organised criticality* (SOC). Since power-law distributions in empirical data can arise for a number of different reasons [Newman, 1997; Eliazar & Klafter, 2009], the SOC theory remains actively debated (see section 14.1).

5.4. Topological properties of networks

Topological network structures (patterns of connections among nodes) play important roles in systems and can influence their behaviour. Such patterns include circuits, trees, modules and clusters, scale-free networks, and small-world structures. For instance, *circuits* produce feedback (both positive and negative); and *trees* (networks containing no circuits) are associated with hierarchies, encapsulation and modularity [Paperin et al., 2010].

One important topological network property is the distribution of node degrees. (The degree of a node is the number of edges connected to that node.) Erdős and Rényi [1960] studied *Poisson random graphs* – networks in which links are created between randomly and independently selected nodes. Such networks show a binomial node degree distribution (or a Poisson degree distribution in the limit of a large graph size). Much of the common intuition about network topology as well as many analytic results have been obtained on the basis of these simple networks [Newman, 2003]. However, the informative value of Poisson random graph models is limited, as such networks are rarely found in natural systems, and they do not exhibit many of the statistical properties found in natural networks. For instance, Poisson random graphs exhibit a low level of clustering, there is no degree correlation and they are not suitable for local navigation using heuristics [Newman, 2003].

Most real-world networks exhibit strongly right-skewed degree distributions: They have a large number of low degree nodes and a smaller number of high degree nodes that act as hubs. Some such networks exhibit an exponential node degree distribution, i.e. the proportion of nodes with degree k falls off as $\lambda e^{-\lambda k}$, for some constant λ . An example is the power distribution grid in the western United States [Watts & Strogatz, 1998].

A particularly well studied class of networks exhibit a node degree distribution that follows a power-law [De Solla Price, 1965; Barabasi & Albert, 1999; Albert & Barabási, 2000], i.e. the number of k -degree nodes falls as $k^{-\alpha}$ for some constant α (Figure 5.5, a). Such networks are termed *scale-free networks* [Barabasi & Albert, 1999]. Scale-free networks are frequently related to systems with SOC-style behaviour [Albert & Barabasi, 2002]. They are found in a variety of complex systems including scientific collaboration networks [De Solla Price, 1965],

protein interaction networks [Wuchty, 2001], telecommunication networks including the WWW [Barabasi et al., 2000], and many others [Barabasi, 2003]. Scale-free networks play an important role in the emergence of some widespread complex properties. For instance, they influence epidemic spreads [Pastor-Satorras & Vespignani, 2001], exhibit a high random fault tolerance [Jeong et al., 2000], and show the small world property [Watts & Strogatz, 1998].

Small-world networks are networks with predominantly local connections and a few non-local cross-connections (Figure 5.5, b). This topology occurs in scale-free networks, as well as in networks with other degree distributions. Formally, the small-world property is present in networks where the average shortest path length between nodes grows logarithmically (or slower) as a function of network size (for a given mean node degree) [Newman, 2003]. Natural small-world networks tend to be highly clustered and have a small average shortest path length between the nodes [Watts & Strogatz, 1998]; they are robust against random edge deletions [Wang & Chen, 2003]. Small-world networks play an important role in many complex phenomena. They have been discovered in a variety of systems including many social networks [Davidsen et al., 2002], biological metabolic networks [Wagner & Fell, 2001], disease spread, telecommunication and distribution networks [Watts & Strogatz, 1998; Newman, 2003; Wang & Chen, 2003].

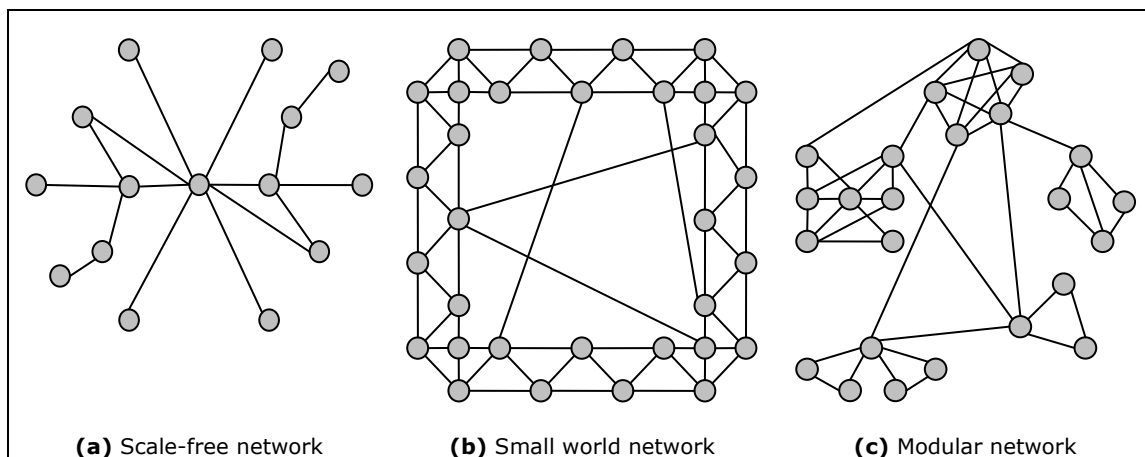


Figure 5.5. Some common structures in networks.

(a) A scale free network has a negative exponential distribution of edges per node.

(b) A small world has a regular structure of local connections with some randomly placed long-range connections.

(c) A modular network consists of sub-components (modules) that are well connected internally, but poorly connected to the rest of the network.

5.5. Clusters and modules

Community structure is another key attribute of network topology. Two particular measures are often applied in this context – the level of node clustering and modularity.

Clustering is a measure of the abundance of transitive structures in a network, i.e. of how often it is the case that two nodes that share a neighbour are themselves connected. Girvan and Newman [2002] have determined the clustering in many natural networks. Formally, clustering is measured by the clustering coefficient C [Newman, 2003]:

$$C = 3 \cdot (\text{number of triangles in the network}) / (\text{number of connected node triples})$$

Modules are network substructures that exhibit a significantly higher connection density within themselves compared to the rest of the network (Figure 5.5, c). Modules can be hierarchical, i.e. they can consist of sub-modules. In opposite to clustering, modularity appears harder to define, detect and quantify in a generic context. Some methods for discovering modular structures in networks have been developed (e.g. [Van Dongen, 2000b; Newman & Girvan, 2004]), but many open questions remain, particularly in regard to discovering hierarchical modules.

Modular structures facilitate reusability of parts and constrain feedbacks. Such structures are thought to enhance various emergent system properties including stability [Variano et al., 2004], robustness [Kitano, 2004; Ma et al., 2006], as well as evolvability and capacity to adapt [Wagner, 1996; Yang, 2001]. Modularity is also ubiquitous in human problem solving. Various engineering techniques seek to ensure modularity in processes (e.g. divide-and-conquer algorithms [Knuth, 1998]) and in structures (e.g. object-oriented software [Booch et al., 2007]). Some mechanisms that may produce modularity in biological systems have been proposed [Wagner, 1996; Wagner & Altenberg, 1996], however, the generic mechanisms involved are still poorly understood (however, see section 13.2).

5.6. Network motifs

Network properties discussed in the above sections are global, they relate to the large-scale topology of a network. Recent research suggests that widely occurring local network structures may also play a role in system dynamics. For instance, Milo et al. [2002] determined that in many natural networks certain small sub-graphs occur significantly more often than other sub-graphs. These frequent sub-graphs – termed *motifs* – also occur significantly more often than in random networks that have otherwise similar properties.

Specific motifs are particularly abundant in diverse systems that share specific functional similarities. For instance, Milo et al. [2004] found that networks that share similar requirements for information processing also share motifs. An example are gene regulatory networks in bacteria and stimuli response networks in yeast, which both perform sensory transcription tasks. Some topical link networks in WWW share motifs with particular social networks. Co-location networks from different unrelated languages exhibit clear similarity of local structures [Milo et al., 2004].

It has been conjectured [Milo et al., 2002; Kashtan et al., 2004] that network motifs perform specific, small-scale, functional roles. Based on this idea, Itzkovitz et al. [2005] used motifs to detect coarse-grained structures and super-motifs that allow analysis of network patterns that are otherwise difficult to detect.

5.7. State spaces as networks

The state spaces of dynamic systems form networks in which the states are nodes and the transitions define edges [Green, 1993]. Thus, system dynamics can be modelled in terms of *state transition networks*, allowing the application of graph-theoretical analysis techniques (see also [section 5.2 above](#), in particular [sub-section 5.2.4](#)).

The way in which walker can navigate along a state transition network depends on the network topology. Thus, understanding topological properties of state transition networks in particular systems can help understand the system behaviour. Sparse connectivity of state transition networks often implies simple behaviour, while richly connected state transition networks are associated with chaotic behaviour. A connectivity phase change between disconnected and connected state transition networks often marks a transition between simple and chaotic behaviour. Langton [1990; 1991] coined the term *edge of chaos* when investigating such phase transitions in systems using cellular automation models. Langton as well as some other researchers (e.g. [Langton 1990; 1991] [Packard, 1988]) have speculated that only systems whose state space connectivity is close to the critical threshold are capable of universal computation. However, later studies (e.g. [Mitchell et al., 1993]) have shown that the evidence is not conclusive.

A widely used state transition network model is a *Markov network* [Markov, 1906]. Markov networks are used primarily for modelling state transitions in discrete stochastic systems (e.g. see [Meyn & Tweedie, 2009] for a review). In a Markov state transition network, every node represents a complete state of the entire system and the edges, annotated with transition probabilities, represent potential changes of state at any given time. Similar state transition graphs have also been applied to non-discrete systems, for instance under the term *configuration state network* [Gfeller et al., 2007].

The concept of state transition networks is related to two other widely used concepts: *stability landscapes* and *fitness landscapes*.

Fitness is a notion of relative adaptedness to an environment or a task, or of relative success under competitive selection. The fitness landscape is a mapping of possible states of an entity (e.g. individual genes, parameter values) to the corresponding fitness values. Stability landscapes can be seen as “inverted” fitness landscapes ([section 6.2](#)). Entities evolving under selection on a fitness landscape tend to “climb” fitness peaks and to “avoid” fitness valleys. Thus, fitness landscapes can be used to describe and predict system dynamics ([section 6.2](#)).

State transition networks represent system dynamics directly rather than modelling fitness/stability topology. However, these concepts are closely related because the dynamics of systems are directly influenced by their fitness/stability landscapes. Systems located close to an attractor (fitness peak) reside in areas with likely transitions to states closer to the attractor and fewer or less probable transitions away from the attractor. Crossing stability ridges (fitness valleys) amounts to traversing from one region of a state network into another region where transition probabilities are biased towards a different attractor (fitness peak).

Chapter 6:

Speciation theory and biological fitness landscapes

The multiplicity of species is a primary example for biological variety and diversity. Species can be understood as reproductively isolated (sub-)populations and the mechanisms leading to speciation differ depending on the type of reproductive isolation and geographical relation of habitats. A key concept in evolutionary speciation theory is the fitness landscape (FL). Recent research reveals that the traditional view of rugged FLs is flawed, and that biological FLs can be modelled as (nearly) neutral fitness networks in genotype space, termed holey fitness landscapes (HFL). This chapter provides a brief overview of key biological concepts that are central to the work presented in this thesis.

The material presented in this section is based on the following publications:

-  G. Paperin, S. Sadedin, D. Green, A. Dorin (2008):
Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity.
11th International Conference on Artificial Life (ALife XI). Ref: [Paperin, Sadedin, et al., 2008].
-  G. Paperin, D. G. Green, A. Dorin (2007):
Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].

In this chapter I briefly review the key biological concepts that are central to the work presented in this thesis. Biological diversity is primarily expressed in the striking multiplicity of different varieties and adaptations. In [section 6.1](#) I discuss the different modes of reproductive isolation and speciation known in biology. A key concept in speciation theory is the fitness landscape (FL). In [section 6.2](#) I discuss FLs with a specific concentration on the biological perspective. A brief overview of the development of the FL concept in evolutionary theory arrives at the modern understanding that high-dimensional biological FLs form *holey* rather than *rugged* topographies. Such adaptive landscapes can be understood as (nearly) neutral networks and are known under the term *holey fitness landscapes* (HFLs).

6.1. Speciation

The modern evolutionary theory takes origin from Darwin's seminal book on the origin of species [Darwin, 1859]. An accessible introduction into theoretical evolutionary biology from today's perspective is provided by Schilthuizen [2001]. An excellent technical overview of the field from the perspective of mathematical modelling and numerical simulations is provided by Gavrillets [2004].

Speciation is the event or the process leading to the formation and establishment of new species [Schilthuizen, 2001]. So, what is a *species*? Intuitively, a species is a group of individuals that have biological characteristics that are different from the characteristics of all other individuals. A precise species definition remains a topic of current debate. Various definitions have been suggested (for instance [Mallet, 1995; De Queiroz, 1998; Harrisson, 1998; Shaw, 1998; Templeton, 1998]), and each has their pros and contras in specific research contexts. Here I follow the mainstream literature [Dobzhansky, 1937; Mayr, 1942] in the understanding that: two populations are considered different species if they are reproductively isolated. I further agree with Gavrillets [2004] that the reproductive isolation does not need to be absolute and some gene flow may be permitted between species.

Reproductive isolation is a reduction or prevention of gene flow between (sub-)populations. [Futuyma, 1987; Gavrilets, 2004]. There are different types of reproductive barriers that can prevent gene flow:

Prezygotic reproductive barriers (barriers that prevent fertilisation) include geographical or seasonal separation that prevents encounters, as well as behavioural or mechanical differences that prevent copulation, and incompatibility of gametes that prevents fertilisation. *Postzygotic* barriers (barriers acting after fertilisation) include inviability of the embryo or the developed hybrid, sterility, and inviability of hybrid offspring. Such barriers may be caused by genetic incompatibilities or by hybrid maladaptedness to the ecological environment [Gavrilets, 2004].

The dynamics of biological evolution and speciation can be influenced by factors such as natural and social (e.g. sexual) selection, mutation, recombination and random genetic drift. Deterministic factors, such as natural and social selection, exhibit stronger effects on large populations. Stochastic factors, such as random genetic drift, particularly affect smaller populations [Maynard Smith, 1958; Gavrilets, 2004].

Three geographical modes of speciation are usually differentiated.

Allopatric speciation occurs when two populations of the same species become geographically isolated by some means, and then diverge to become different species. This form is the most common type of speciation [Mayr, 1942; Schilthuizen, 2001].

Sympatric speciation is speciation without geographic isolation or speciation within a single population. Formally, it can be defined as speciation within a population where mating is not dependent on the birth location of mating partners [Gavrilets, 2004]. This type of speciation is considered rare, although various examples have been observed [Schilthuizen, 2001].

Parapatric speciation can be considered as the intermediate case, where the diverging (sub-)populations are partially geographically isolated. It can also be viewed as the general case. If the migration rate between two diverging populations is near zero, it is a case of allopathy, if the migration rate is very high, it is a case of sympatry, otherwise it is a case of parapatry [Endler, 1977; Schilthuizen, 2001].

Notably, the terms *allopatric*, *sympatric* and *parapatric* are not used consistently in the literature. Sometimes, the term *sympatric* is used as the opposite of *allopatric*, and thus includes both, the *sympatric* and the *parapatric* modes described above [Schilthuizen, 2001]. At some occasions other meanings are associated with these terms. Here, I use these terms in-line with the above descriptions.

A more comprehensive overview of speciation modes and the conditions under which they may or may not lead to the formation of a new species can be found in [Gavrilets, 2004]. A less technical discussion is available in [Schilthuizen, 2001].

Adaptive radiation is the diversification of one or a few ancestral species into a larger number of species that evolve to use different ecological niches [Gavrilets, 2004]. In a sense, adaptive radiation can be viewed as the "bigger picture" of speciation. The term *speciation* normally refers to a single speciation event (that may actually be taking place over many thousands of years); the term *adaptive radiation* is used to refer to the dynamics affecting the development of a variety of species within a larger biological environment. Some writers, however, use these terms fairly loosely. For instance, *speciation* can sometimes refer to a particular speciation event, to the process of appearance of a new species, to the adaptive radiation process in some environment, or all three of these things at the same time [Schilthuizen, 2001].

6.2. Fitness landscapes

The material presented in this section is based on the following publications:

-  G. Paperin, S. Sadedin, D. Green, A. Dorin (2008):
 Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity.
 11th International Conference on Artificial Life (ALife XI). Ref: [Paperin, Sadedin, et al., 2008].
-  G. Paperin, D. G. Green, A. Dorin (2007):
 Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
 2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].

Fitness landscape (FL) is a concept that originated in evolutionary biology and has found wide applicability in other areas, such as evolutionary computation. In evolutionary biology the *fitness* of an individual is a measure of the survival and reproductive success of that individual's offspring relative to others [Gavrilets, 2004]. More generally, fitness is a notion of relative adaptedness to an environment or a task, or of relative success under competitive selection. The fitness landscape is a mapping of possible system states (e.g. phenotypic traits, parameter values, strategy choices) to the corresponding fitness values. Entities evolving under selection tend to "climb" fitness peaks and to "avoid" fitness valleys [Wright, 1932; Gavrilets, 2004]. Fitness landscapes play a key role in [chapters 10 and 11](#) that investigate mechanisms behind the emergence of sustained diversity and perpetual novelty in biological CAS.

Concepts that are very similar to FLs are known in many sciences. In physics, *spin-glasses* can be described in similar form [Mezard et al., 1987; Stadler, 2002]. Chemical *potential energy surfaces* are a further related concept [Mezey, 1987], as well the *energy landscapes* used in models of protein and nucleotide sequence folding [Dill et al., 1995; Flamm et al., 1999]. Kollman et al. [1992] demonstrate how the metaphor is used to develop a notion of an electoral landscape in a social voting model. In ecology, *stability landscapes* map system states to relative stability, with attractors represented as sinks and ridges representing thresholds between basins of attraction. The basin of an attractor is the area of the stability landscape (i.e. set of system states) from which from which the system tends to develop to that attractor. Systems are generally resilient against forcing that drives them away from the attractor of the basin in which they are currently located. [Scheffer & Carpenter, 2003]. Thus, systems are said to "fall" towards the nearest attractor. Metaphorically, stability landscapes can be seen as "inverted" FLs. Peaks on a fitness landscape correspond to attractors and fitness valleys correspond to the ridges between basins of attraction.

In this section I concentrate on biological FLs and briefly review FL concept from the perspective of evolutionary biology.

6.2.1. Overview

In biology, the term *fitness landscape* (FL) was coined by Wright [1932] to represent the fitness of all conceivable individuals relative to their traits. He envisaged a rugged landscape, where peaks represented combinations of traits with high fitness separated by valleys of low-fitness trait combinations ([Figure 6.1](#)). On this landscape, selection drives populations uphill.

There are several related, but different definitions for FLs in the literature. A general overview over the area is given by Stadler [2002]; Gavrilets [2004] provides a detailed discussion that focuses on evolutionary speciation theory. There, some definitions assign a fitness value to every individual, while others assign a single average fitness value to a whole (sub-)population. The work discussed here is based around individual-based simulation experiments, thus I define a FL over the space of all possible individual genotypes.

Generally, a FL can be visualised as a plot of a "goodness" function that assigns a fitness value to every possible individual genotype. The genotypes are ordered in an abstract space

in which the distance between the genotypes describes the probability or easiness of the transformation of one genotype into another. In many cases, FLs are used as a metaphorical concept that aids an intuitive understanding of evolutionary dynamics. However, in evolutionary biology, a formally defined construct is often used for mathematical models (see overview by Gavrillets [2004]).

In evolutionary biology the notion of FLs has been criticised (e.g. [Provine, 1986; Arnold et al., 2001]). FLs are usually considered static. In reality, fitness is the ability to survive and reproduce in a dynamic environment that is constantly changing through co-evolutionary dynamics and external disturbances. The effect of phenotypic traits in such environments, and therefore of the gene complexes affecting the traits will change over time. Therefore the assumption that the corresponding FLs are static is a simplification. Some models account for this by using a FL that changes with time to reflect changes in the environment (e.g. [Kauffman, 1993]). However, the genes underlying species differences and reproductive isolation are, in general, not strongly affected by the environment and FLs are, therefore, widely accepted as a useful abstraction in theoretical biology [Gavrilets, 2004].

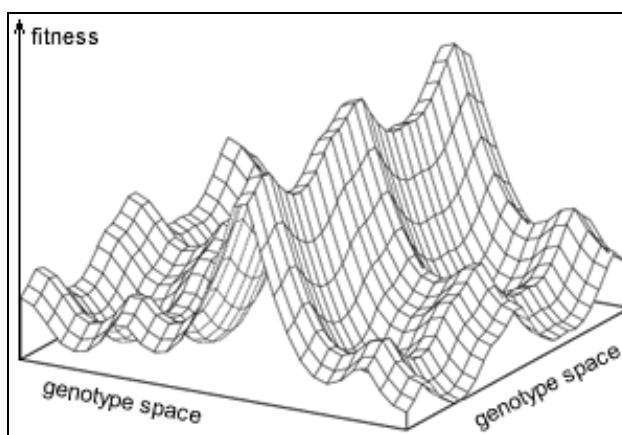


Figure 6.1. Adaptive fitness landscape metaphor. Wright [1932] envisaged a rugged landscape representing the fitness of all conceivable individual genotypes, where peaks corresponding to high-fitness genotypes are separated by valleys of low-fitness genotypes. (Original figure from <http://www.compapp.dcu.ie/~humphrys/Notes/GA/evolution.html>. Modified and reproduced with permission by Mark Humphrys, Dublin City University, Ireland.)

6.2.2. Formal model

A *genotype* is the complete set of all genes carried by an individual. In the models discussed here, individuals are represented by their genotypes, thus a genotype is equivalent to an individual. A genotype consists of an ordered set of loci. A *locus* is stretch of an individual's DNA (i.e. an individual's genotype) at a specific location. The actual DNA sequence at a given locus is an *allele*. In the models discussed here, a genotype is represented as a fixed-length string. Every position of the string corresponds to a locus and the set of characters or values permitted at that position corresponds to the set of possible alleles at that locus.

Formally, let $\mathbb{G}$ be the set of all possible genotypes. A genotype $G \in \mathbb{G}$ of an individual λ is represented as $G_\lambda = (g_1, \dots, g_k, \dots, g_L)$, where L is the number of loci and g_k are the alleles at loci k . Assume that each locus can accommodate one of A different alleles: $g_k \in \{a_1, a_2, \dots, a_A\}$. Note, however, that the model assumption that each of the L loci can accommodate the same number A of distinct alleles does not generally apply in real organisms.

This model only describes the gene configuration of an individual and says nothing about the reproduction mechanism or the genes relationship to each other. Hence, within this model a diploid organism with L alleles is equivalent to a haploid organism with $2L$ alleles. For simplicity I will thus only consider haploid individuals.

For any genotype $G \in \mathbb{G}$ let $neighbours(G) \subseteq \mathbb{G}$ denote the set of all genotypes that can be obtained from G by a single atomic mutation.

Here, I define an *atomic mutation* as a single one-point mutation: $G' \in \mathbb{G}$ is a neighbour of G iff G' and G are same except at some locus g and the allele of G at locus g can directly mutate into allele of G' at locus g .

Let $gdist(G_1, G_2)$ be the minimum number of atomic mutations required to obtain $G_2 \in \mathbb{G}$ from $G_1 \in \mathbb{G}$. Then:

$$G_2 \in neighbours(G_1) \Leftrightarrow gdist(G_1, G_2) = 1$$

Here, I assume that mutations are reversible, i.e. the probability of a mutation from an allele a to allele a' at some locus g is the same as the probability of mutation from a' to a at g . In general, this is not always be the case in biological organisms, however, this does not affect the following discussion. Thus, $dist$ is symmetric:

$$G_2 \in neighbours(G_1) \Leftrightarrow G_1 \in neighbours(G_2) \Leftrightarrow gdist(G_1, G_2) = 1 \Leftrightarrow gdist(G_2, G_1) = 1$$

The *genotype space* is the set $\mathbb{G}$ of all genotypes with the distance function $gdist: \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{N}$.

The *fitness* of a genotype is a notion of relative adaptedness of the genotypic traits to an environment or a task, or of relative success under competitive selection. Biologists use fitness as a measure of the survival and reproductive success of individuals carrying a certain genotype relative to others. A *fitness function* is a function that assigns some fitness value to every genotype. Formally: $fitness: \mathbb{G} \rightarrow \mathbb{R}$. The *fitness landscape* is the hyperplane that contains all points $fitness(G)$ for all $G \in \mathbb{G}$ ordered under $gdist$.

6.2.3. Fitness landscapes in speciation models

In this subsection I give a brief summary of the historical development of the FL concept in evolutionary biology and highlight the issues arising from viewing adaptive radiation as a process of adaptation on a FL. In [section 6.2.4 below](#) I discuss how some the issues raised here can be resolved.

If every individual genotype can accommodate A alleles at each of L loci, then there are A^L possible distinct genotypes. The number of neighbours of a genotype $G \in \mathbb{G}$ is $|neighbours(G)| = L \times (A - 1)$ (because any of the L loci can change to any of the possible A alleles except the one that is currently at that locus).

The dimensionality of a space is (roughly speaking) the number of (linearly independent) coordinates required to describe a point in that space. Gavrillets [2004] provides an in-depth discussion of FL dimensionality. Here, I concentrate on issues relevant for the discussion in the following chapters. The dimensionality of the above FL construct is by one larger than the dimensionality of the genotype space (the additional fitness dimension):

$$dim(fitness: \mathbb{G} \rightarrow \mathbb{R}) = L \times (A - 1) + 1$$

The implication of this is that the dimensionality of biological FLs is typically extremely high. Consider, for instance, a population of individuals with $L = 100$ loci and $A = 3$ alleles (these are very low numbers, even the genotypes of primitive bacteria are much larger). The number of possible genotypes is then $3^{100} \approx 5 \times 10^{47}$. This number is by far larger than the size of any natural population, implying that only a small proportion of possible genotypes can actually occur in individuals. The dimensionality of the corresponding FL is $100 \times (3 - 1) + 1 = 201$.

Wright's original FL metaphor implied that a FL is a "rugged" surface, with many fitness peaks separated by fitness valleys ([Figure 6.1](#)). Peaks represent genotypes posing "good" solutions to the problem of survival, and valleys represent low-fitness genotypes, that may arise either due to deleterious mutations or because they include incompatible genes. Selection drives species populations up the peaks and away from the valleys. In such model,

distant peaks corresponding to distant genotypes can be thought of as different species – the further, the less related.

Due to selection pressures populations cannot “escape” fitness peaks without additional forces, and thus, a population that reached such a peak will stop evolving. This poses an important question for evolutionary theory: In order for a new species to arise, part of the population must escape a local fitness peak and develop towards a different region in the gene space, so what are the factors that facilitate this process? An obvious approach is to critically consider the simplifying assumption that FLs are static. As this is not always the case in reality, peak regions may become less fit as the environment changes over time, and populations located around such areas can evolve towards better adaptations. However, genes underlying reproductive isolation and thus species differences are, in general, not strongly affected by the environment [Gavrilets, 2004]. Hence, the mechanisms causing different parts of a population to evolve towards different adaptations and to develop reproductive isolation must be clarified.

Stochastic factors such as *genetic drift* may act against natural selection and help overcoming fitness valleys, particularly for small populations, however such factors can only account for selected types of speciation [Gavrilets, 2004].

A key consideration relates to the topological structure of FLs. Much of intuition about evolutionary dynamics is obtained by considering the metaphor or a 3-dimensional FL. However, the mathematical properties of high-dimensional spaces are often different from the properties of the 3-dimensional world [Gavrilets, 2004], and some of the intuition won in 3 dimensions may not apply.

Indeed, peaks in a low-dimensional spaces become saddle points in higher-dimensional spaces (Figure 6.2). As a result, it has been suggested that the FLs may actually possess a single global maximum that can be reached by hill climbing from (almost) any point [Provine, 1986]. However, the local-maxima-to-saddle-point transformations are outnumbered by the appearance of new peaks in higher dimensions. The single-peak model, although does not apply in general, is nevertheless useful as an approximation when investigating the local behaviour of a population close to a fitness peak [Gavrilets, 2004].

Kimura [1983] suggested that most genetic changes on the biochemical level may actually be fitness-neutral and that FLs are thus (almost) flat. This neutral theory implies that the main forces behind speciation are stochastic genetic divergence, in particular due to genetic drift. However, in nature, a vast number of gene combinations are inviable because they contain deleterious genes or groups of incompatible genes. Neutral fitness landscapes fail to account to this fact.

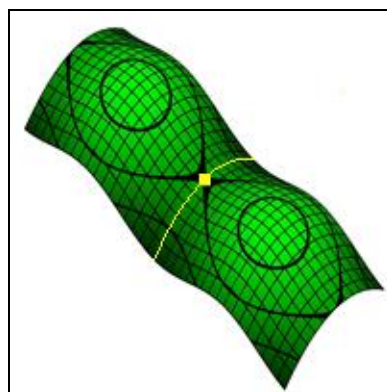


Figure 6.2. Local maxima can become saddle points in higher dimensions. The yellow point is a maximum of the 2-dimensional curve (yellow), however, it becomes a saddle point between the two hills in 3 dimensions. Two new maxima (the hill peaks) appear in 3 dimensions.
(Original image, obtained from http://en.wikipedia.org/wiki/File:Saddle_pt.jpg, is in the public domain.)

A well known FL model that can mediate between smooth and rugged surfaces is the N - k *landscape* introduced by Kaufman [1993]. There, each of the L diallelic (binary) loci belongs to an epistasis-group with k other loci (*epistatic* loci are loci whose effects influence each other). (In the original paper the number of loci was denoted by N , not L , thus the term N - k *landscape*.) Each locus g_i ($i = 1 \dots L$) has a fitness contribution

$$f_i = \text{fitness}(g_i, (g_{i,1}, g_{i,2}, \dots, g_{i,k}))$$

where $g_{i,1}, g_{i,2}, \dots, g_{i,k}$ are the loci with which g_i interacts epistatically. The fitness values f_i are independently drawn from a uniform random distribution for all possible gene combinations. The total fitness of a genotype is:

$$\text{fitness}(G) = \sum_{i=1}^L \frac{f_i}{L}$$

The parameter k that defines the size of epistatic groups has a direct effect on the ruggedness of the FL. Small k yield smooth landscapes and large k yield very rugged landscapes. N - k landscapes find a wide use in biological modelling because the k parameter may be conveniently used to tune the ruggedness of the landscape to the requirements of the respective problem at hand. However, N - k landscapes suffer from the same issues as the previous models: Small values of k do not account for the large proportion of inviable genotypes found in nature, and large values of k produce rugged FLs in which stochastic crossing of fitness valleys is unlikely.

6.2.4. Nearly neutral networks

In the previous sections I introduced the notion of FLs and discussed several FL models. Gavrillets [2004] reviews various speciation models on rugged FLs and concludes that on such landscapes the transition between fitness peaks required for speciation, is, in general, unlikely. Two alternative FL configurations (single peak and flat landscapes) attempt to address some of the drawbacks of the original rugged FL concept. However, such models are appropriate only in specific speciation scenarios. In this section, I describe the concept of evolution on *nearly neutral networks* that combines key ideas of rugged and neutral FLs. The notion of evolution on nearly neutral networks lead to the formulation of the *holey fitness landscape* (HFL) concept [Gavrillets & Gravner, 1997; Gavrillets, 2003; 2004]. Generally, a HFL is “an adaptive landscape where relatively infrequent high-fitness genotypes form a contiguous set that expands throughout the genotype space” [Gavrillets, 2003].

To build some intuition for the HFL model, recall a few results from percolation theory which plays an important role in the analytical treatment of HFLs. Consider a 2-dimensional lattice of cells which can assume one of two states: “black” or “white” (Figure 6.3). Let every cell be black with some probability p independently of all other cells, or white with probability $1 - p$. If p is small, the lattice will contain a few black cells, which may be grouped in a number of small, isolated components. As p increases, these components grow and merge. Once p crosses a certain threshold p_c , most of the black cells merge together into a single giant component that percolates the whole lattice (Figure 6.3). This process of lattice percolation is closely related to the process of network percolation described in section 5.3. For a 2-dimensional square lattice the *percolation threshold* is known to be $p_c \approx 0.5927$ [Newman & Ziff, 2000]. However, for higher dimensional lattices the percolation threshold lies around the reciprocal of the lattice dimension [Gavrillets & Gravner, 1997; Grimmett, 1999], meaning that for a high dimension lattice a small proportion of black cells is sufficient for the emergence of a giant percolating component of connected black cells.

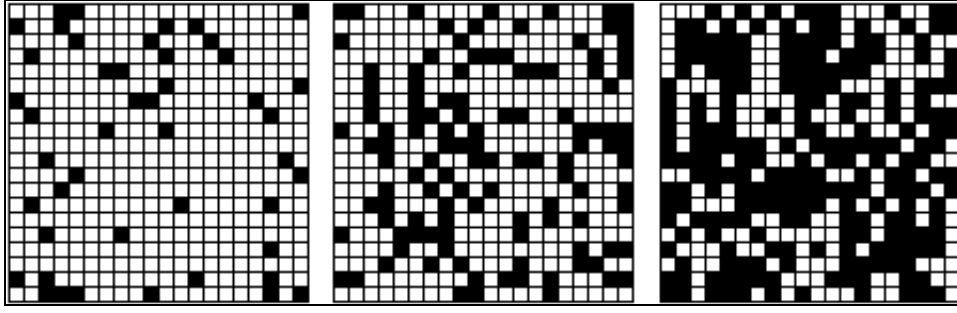


Figure 6.3. Percolation on a square lattice. The cells are black with probability $p = 0.1$ (left), $p = 0.3$ (middle) and $p = 0.6$ (right).

For the HFL model, assume that a genotype is *viable* with probability p independent of all other genotypes, and *inviable* with probability $1 - p$. Let $\mathbb{V} \subseteq \mathbb{G}$ be the set of all viable genotypes (in reality, because p is small, $\mathbb{V} \subset \mathbb{G}$). Two viable genotypes $G_1, G_2 \in \mathbb{V}$ are *connected*, iff G_2 can be obtained from G_1 by successive atomic mutations such that all genotypes along the mutational path are also viable (note, that *connectivity* is a symmetric relation under the current assumptions). Similarly to the above lattice model, when p is small, the viable genotypes will generally not be connected. As the number of viable genotypes grows (p increases), viable genotypes merge into connected components. There is a threshold p_c , such that if $p > p_c$, a giant connected component of viable genotypes emerges. Recall that the dimensionality of the genotype space is $\dim(\mathbb{G}) = L \times (A - 1)$, where L is the number of loci and A is the number of alleles at each locus. It can be shown [Gavrilets, 1997; 2004] that a positive proportion of all viable genotypes $G \in \mathbb{G}$ belong to a single connected component if:

$$p \geq p_c \approx \frac{1}{\dim(\mathbb{G})} = \frac{1}{L \cdot (A-1)}$$

In that case $\mathbb{V}$ is said to percolate $\mathbb{G}$. It can be shown [Reidys, 1997], that *all* viable genotypes will belong to a single component if:

$$p > 1 - \frac{1}{A-1\sqrt{A}}$$

Recall the previous example with $L = 100$ loci and $A = 3$ alleles. In that case $\mathbb{V}$ percolates $\mathbb{G}$ already when $p \geq p_c \approx 1 / 200$. For a biologically more realistic, but still small genotype size with $L = 10^5$ and $A = 5$, the giant connected component emerges at $p_c \approx 20 \times 10^{-7}$. The implication for high-dimensional FLs is that even very small values of p will result in an extensive network of viable ridges extending through the genotype space. The traditional picture of rugged highly-dimensional FLs is therefore misleading, as high-dimensional landscapes are characterised by the existence of percolating networks. In the next few paragraphs I summarise the arguments presented by Gavrilets [2004, chap. 4] to show that if the fitness of the genotypes is not restricted to *viable* or *inviable*, a large number of such networks emerges, each containing genotypes from a narrow fitness band. Such networks are thus termed *nearly neutral* (in respect to fitness changes). Among these networks, those with high fitness are particularly important as adaptive walks along such networks can proceed very far without any substantial loss to fitness.

Assume that the fitness of the genotypes is uniformly randomly distributed in $[0, 1]$. Let $f_1 \in [0, 1]$ and $f_2 = f_1 + \varepsilon$ for some small positive ε , such that $f_2 \in [0, 1]$. A genotype $G \in \mathbb{G}$ is in the fitness interval $[f_1, f_2]$ iff $f_1 < \text{fitness}(G) \leq f_2$. Note that belonging to a certain fitness interval is analogous to being viable in the previous argumentation, so if $\dim(\mathbb{G})$ is large and $\varepsilon > 1 / \dim(\mathbb{G})$, there exists a percolating network of genotypes that belong to the $[f_1, f_2]$ -fitness-interval [Gavrilets & Gravner, 1997; Gavrilets, 2004]. The smaller ε , the more similar will be the fitness values of the genotypes that belong to such a network, forming therefore, a (nearly) neutral network. Note that there will be several non-overlapping such networks. In fact, there will be approximately $1 / \varepsilon \approx \dim(\mathbb{G})$ of them.

More generally, let the fitness values of the genotypes be independently randomly distributed according to some distribution with the probability density function $\mathcal{D}(f)$, where $f = \text{fitness}(G)$ and $G \in \mathbb{G}$. The genotypes within the fitness interval $[f_1, f_2]$ form a percolating network iff :

$$\int_{f_1}^{f_2} \mathcal{D}(f) df > \frac{1}{\dim(\mathbb{G})}$$

That is, iff the proportion of genotypes in $[f_1, f_2]$ is larger than $1 / \dim(\mathbb{G})$.

Gavrilets [2004] provides a variety of evidence for the existence of nearly neural networks in his book. He coins the term *holey fitness landscape*, a landscape that encompasses a nearly neural network of genotypes in a relatively high fitness interval. The *holes* in the landscape represent genotypes that do not belong to the interval, including both, lower fitness valleys, as well as largely inaccessible fitness peaks. Evolutionary models on HFLs combine the desirable properties of rugged and flat landscapes.

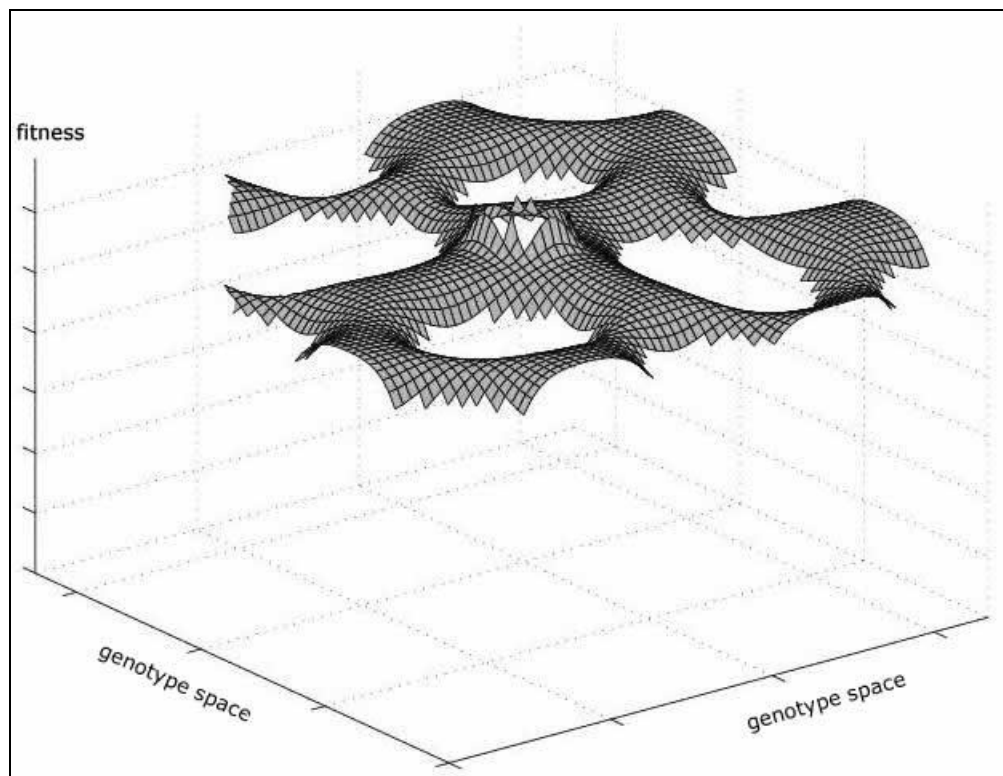
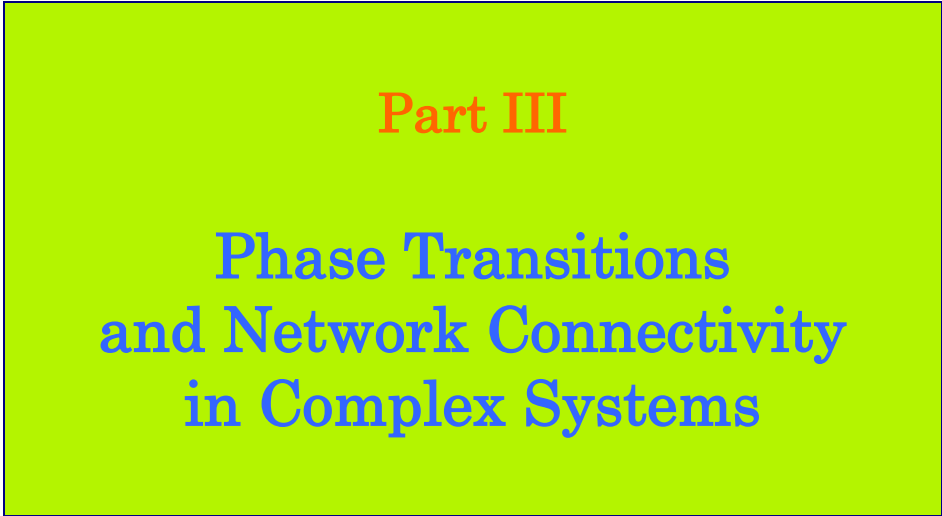


Figure 6.4. Holey fitness landscape (HFL) concept. HFL is an adaptive landscape where relatively infrequent high-fitness genotypes form a contiguous network that expands throughout the genotype space. The holes represent both, lower fitness valleys as well as largely inaccessible fitness peaks.
(Original figure obtained with permission to modify and reproduce from Prof. S. Gavrilets, personal communication.)



Part III

Phase Transitions and Network Connectivity in Complex Systems

This part establishes the core terminology that is introduced in this thesis and explains the main theoretical concept proposed in this work – the Dual Phase Evolution framework (DPE). A wide range of evidence for DPE processes in complex systems is discussed and an understanding of how DPE processes can be identified and described in particular systems is developed.

Chapter 7: Dual Phase Evolution.

The topology of networks that underlie complex systems is a key factor that influences system dynamics. Networks can exist in two general connectivity phases: well connected and poorly connected. The DPE framework relates network connectivity to typical system behaviour and associates recurrent transitions between network connectivity phases to a number of emergent properties in complex systems. This chapter introduces DPE, and explores important related issues, such as the mechanisms behind connectivity phase transitions.

Chapter 8: Connectivity Phase Transitions in Natural and Artificial Systems.

This chapter examines a wide range of empirical and experimental evidence from physical, biological, social, economic, mathematical and computational systems drawn from previous studies. The evidence supports the central ideas of the DPE framework by demonstrating that DPE phase changes can drive the emergence of different complex system properties. Moreover, the evidence uncovers that many of the examined systems share underlying DPE mechanisms and processes that were not previously described.

Chapter 7:

Dual Phase Evolution

Complex systems can be understood as networks of interacting components, and the topology of component interaction networks is a key factor that drives system dynamics. Networks can exist in two general connectivity phases: well connected and poorly connected. The Dual Phase Evolution (DPE) framework relates each of the two connectivity phases in interaction networks to typical system dynamics and postulates that recurrent transitions between the connectivity phases may account for a number of emergent properties in complex systems. After introducing DPE, this chapter explores some related issues, such as the general mechanisms that can cause connectivity phase transitions.

Material presented in all sections of this chapter is based on the following publications:

- 📄 G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
- 📄 G. Paperin, S. Sadedin (2009):
Towards formalising the theory of Dual Phase Evolution.
10th European Conference on Artificial Life (ECAL'09). Ref: [Paperin & Sadedin, 2009b].
- 📄 G. Paperin, S. Sadedin (2009):
The Dual Phase Evolution Framework for Understanding Evolutionary Dynamics in Complex Adaptive Systems.
2009 International Conference on Evolutionary Computation (ICEC'09). Ref: [Paperin & Sadedin, 2009a].
- 📄 G. Paperin, D. G. Green, T. G. Leishman (2008):
Dual Phase Evolution and Self-Organisation in Networks.
7th International Conference on Simulated Evolution And Learning (SEAL'08). Ref: [Paperin, Green, et al., 2008].
- 📄 G. Paperin, D. G. Green, S. Sadedin, T. G. Leishman (2007):
A Dual Phase Evolution Model of Adaptive Radiation in Landscapes.
3rd Australian Conference on Artificial Life (ACAL'07). Ref: [Paperin, Green, et al., 2007b].
- 📄 G. Paperin, D. G. Green, S. Sadedin, T. G. Leishman (2007):
Complexity in Speciation: Effects of Disasters on Adaptive Radiation in a Dual Phase Evolution Model.
8th Asia-Pacific Complex Systems Conference (Complex'07). Ref: [Paperin, Green, et al., 2007a].

7.1. Introduction

In this chapter I present a theoretical framework for reasoning about complex systems. The framework implies that recurrent phase changes in interaction density of system components (i.e. in connectivity of underlying interaction networks) constitute a general mechanism for self-organisation and underlie many emergent phenomena. The framework identifies properties that are typical of systems in different connectivity phases as well as characteristics commonly associated with the phase transitions.

Inspired by the two general connectivity phases in which networks that underlie developing systems can exist (well connected and poorly connected), the framework is termed Dual Phase Evolution (DPE). The term *Dual Phase Evolution* has been previously coined by Green [e.g. Green, 2000; Green et al., 2000; Green & Sadedin, 2005; Green et al., 2006] to generally refer to recurrent changes between connectedness and separation in the development of systems. Here, I formulate DPE as a well-defined framework rooted within network theory.

In subsequent chapters I will show that DPE is a universal framework that is widely useful for reasoning about complexity.

7.2. Connectivity phases in interaction networks

Networks are inherent in all complex systems, and patterns and processes that occur within the underlying networks play a fundamental role in the emergence of system wide features ([chapter 5](#)). Perhaps the most widely felt network processes are changes in edge density, especially the connectivity avalanche ([section 5.3](#)). Evidence presented in [chapter 8](#) demonstrates that phase changes occur in dynamics of complex systems of virtually every kind. What is more, such phase changes can be traced to phase transitions between *well connected* and *poorly connected* phases in underlying interaction networks.

In the *well connected phase*, typified by high edge density and short characteristic path lengths, direct interactions can occur between most system components. Thus, the well connected phase is typified by *global interactions* in the sense of [section 5.2.1](#). Since processes can spread rapidly in well connected networks (see [chapter 5](#), in particular network theory reviews cited there), such systems often exhibit little local variation, but may be highly variable¹. Dense interactions among control variables tend to stabilise systems, but also lead to increased coupling thereby reducing robustness towards external perturbations (see also [section 14.2](#)). This makes highly connected systems hard to predict in their responses to external stimuli. Evolutionary selection in well connected systems acts globally and has stabilising effects: Local perturbations, if advantageous, spread rapidly, generating uniformity. Consequently, state space exploration is reduced and systems tend to converge towards the nearest basin of attraction.

In the *poorly connected (or disconnected) phase* edge density is low, characteristic path lengths are long, and the network typically consists of several sub-networks. Interactions in the poorly connected phase mainly occur locally or only within sub-networks (see *local interactions* defined in [section 5.2.1](#)). Propagation of stimuli is locally constrained in poorly connected systems ([section 5.3](#)). Systems of this kind typically exhibit strong local variation, but little large-scale variability. Such decoupled dynamics imply that large-scale responses to external stimuli can be described in terms of sub-systems and are thus easier to predict. For CAS, evolutionary competition in poorly connected systems is locally constrained permitting exploration of alternative adaptation strategies and directional selection towards novel niches.

Definition:

Dual Phase Evolution (DPE) occurs when networks that dominate the dynamics of an evolving system repeatedly switch between well connected and poorly connected phases ([Figure 7.1](#)).

DPE phase transitions are common to adaptive as well as to other complex systems that exhibit self-organisation (e.g. physical systems discussed in [section 8.1](#)). The term *Dual Phase Evolution* thus implies a general process of complex development, rather than evolution in a strict biological sense. In the following chapters I argue that DPE-driven self-organisation can account for properties widely observed in complex systems, including sustained diversity, perpetual novelty, modularity and scale-free topologies (see [parts IV and V throughout](#)).

While topological properties of underlying *interaction networks* can drive system dynamics, the topologies of *state transition networks* are an emergent property of the system ([section 5.7](#)). DPE phases of local interactions and enhanced evolutionary exploration often correspond to phases of dense connectivity in state networks with increased transition possibilities between attractor basins.

¹ Wagner & Altenberg [1996] attribute precise and different meanings to the terms variation and variability: *Variation* refers to the actually present differences among components in a system. In contrast, *variability* describes the potential to vary. Thus, variation is a property of a set of items that can be observed directly. Variability is a dispositional concept that describes the way in which a system behaves as a whole, for instance in response to external stimuli.

The transition between poorly connected and well connected dynamics does not always require that the critical percolation density threshold is crossed. For instance, some networks above the critical threshold may still be sparse and exhibit long path lengths between nodes. Thus, stimuli may lose their significance by the time they are transmitted via a series of interactions, leading to poorly connected dynamics. An increase in connectivity can reduce path lengths and cause a transition to global (well connected) dynamics. Examples can be found in collaborative social groups: Although information can always be transferred between members via common collaborators, this is often perceived as inefficient, and projects are usually organised as a succession of work in small groups (sparse interactions) and larger meetings (dense interactions).

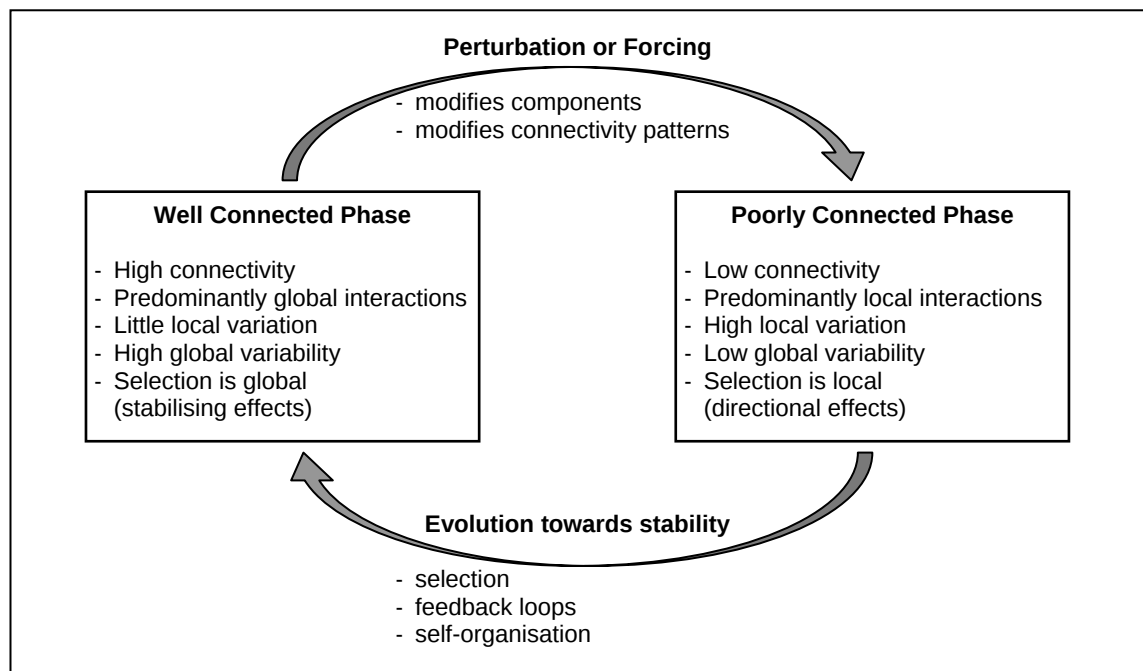


Figure 7.1. The mechanism of Dual Phase Evolution. Systems flip between poorly connected and well connected phases. Perturbations or slow forcing – arising externally or internally – disrupt systems causing connectivity phase transitions in underlying networks. Internal processes restore old and create new interactions.

7.3. Considering multiple networks

An important consideration when linking network phases to system dynamics is that at any one time different networks within a system may be in different connectivity phases. For example, a cell can be viewed as an ensemble of various networks, e.g. protein-protein, protein-DNA, and protein-metabolite networks [Alon, 2003]. Specific system properties may be generated by single networks or by feedback between multiple networks.

For instance, at least two major networks can be identified for plant evolution in landscapes [Paperin, Green, et al., 2007b]. One is a network of reproductive interactions between patches of interbreeding plants. It facilitates gene flow and acts against adaptive radiation. The second is a network of unoccupied patches. An increased presence of such patches disrupts gene flow and aids percolation by newcomer species through increasing the connectivity of their invasion paths. A well connected phase in the first network (reproductive interactions) is signified by predominant stability in landscape composition, while the poorly connected phase is signified by variation and evolutionary exploration. However, the impact of connectivity in the second network (invasion paths) is essentially opposite. Crucially, from either perspective, connectivity phase transitions are linked to significant changes in system dynamics (DPE processes in landscape evolution are considered in more detail in [chapters 9 and 10](#)).

7.4. Mechanisms for phase transitions

There are at least three kinds of processes that can cause connectivity phase changes in systems: external disturbance events, slow external forcing and internal feedback processes. A wealth of specific examples of all three of these mechanisms are discussed in [chapter 8](#). Here, I discuss some of the general mechanisms through which those processes may affect system dynamics.

External disturbances are a very common mechanism behind critical transitions in system dynamics. Disturbances can influence system behaviour by affecting system components or their interactions (give examples showing what you mean by “external”? I assume things like a volcano or the sun going out... i.e. phenomena that are not actually generated by the elements of the system itself.), as well as by affecting the external system environment (see [section 15.1.4.1](#) for a further discussion of environmental influences). For example:

- **Components:** Keystone species are important components in ecological systems. A disturbance of a keystone species can affect a complete food web have far reaching consequences for an ecosystem [Paine, 1969; 1995].
- **Interactions:** Landscape composition is known to persist against strong external forcing, but changes dramatically when perturbations affect landscape connectivity thus changing the topology of internal interactions. This issue is explored in detail in [chapter 10](#).
- **Environment:** Environment changes can affect system constraints and selective pressures (e.g. warmer climate, less sunlight, etc.). Thus, disturbances can influence the evolutionary development of a system even if they do not directly affect it. (See also [section 15.1.4.1](#) and [section 8.6](#).)

Slow forcing caused by external factors can lead to critical phase transitions in system dynamics [Gunderson & Holling, 2002]. Typically, this occurs when a system is pushed across a bifurcation threshold between alternate attractor basins. For example, in drylands, when rainfall declines or grazer populations increase beyond a threshold, erosion removes topsoil leading to a shift from savannah to desert as ecological attractor states [Phillips, 1993].

Internal feedback processes can result in recurrent phase transitions in connectivity of intrinsic interaction network without any external stimuli. Such feedbacks are possible since the relationship between system behaviour and the topology of internal interactions is bidirectional: component interactions drive system behaviour, but are also shaped by emergent system dynamics [Newman, 2003; Garlaschelli et al., 2007]. Intrinsically emerging connectivity transitions can be observed in systems as diverse as stellar systems ([section 8.1](#)), social networks ([section 8.5](#)) and the human brain ([section 8.4](#)). In [chapter 12](#) I discuss two simulation models that demonstrate such processes.

7.5. Summary

The Dual Phase Evolution framework presented in this chapter is the central topic of this thesis. It is thus useful to summarise the key points raised in this chapter.

- Complex systems can be understood as networks of interacting components, and the topology of component interaction networks is a key factor that drives system dynamics.
- Networks can exist in two general connectivity phases: well connected and poorly connected.
- Each of the two connectivity phases in underlying interaction networks relates to specific typical system dynamics.
- DPE occurs when networks that drive system dynamics recurrently switch between the two connectivity phases. Such recurrent connectivity phase transitions may account for a number of emergent properties of complex systems.
- Systems can be understood by reference to multiple interaction networks that may be in different connectivity phases at any one time.
- Mechanisms causing connectivity phase transitions may be system internal or external and include disturbances, slow forcing and feedback processes.
- Connectivity phases in interaction networks can often be related to the density of the current region of system's state transition network. Phases of sparse component interactions are associated with strong variation within systems; phases of dense component interactions are associated with uniformity and balance.

In the next chapter I will present a large body of empirical and experimental evidence, drawn from previous studies, to further support the DPE framework and to demonstrate how its application can be useful for uncovering similarities and common causal mechanisms between seemingly unrelated systems. The subsequent chapters concentrate on how DPE processes may account for specific emergent properties in complex systems.

Chapter 8:

Connectivity Phase Transitions in Natural and Artificial Systems

The DPE framework associates different connectivity phases in networks that underlie complex systems and transitions between these phases with specific system behaviours. In this chapter I examine empirical and experimental evidence from physical, biological, social, economic, mathematical and computational systems drawn from previous studies. The evidence demonstrates how DPE phase changes can drive the emergence of different complex properties in systems. Moreover, the evidence uncovers that many of the examined systems share underlying DPE mechanisms and processes that were not previously described.

Material presented in all sections of this chapter is based on the following publications:

-  G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
-  G. Paperin, S. Sadedin (2009):
The Dual Phase Evolution Framework for Understanding Evolutionary Dynamics in Complex Adaptive Systems.
2009 International Conference on Evolutionary Computation (ICEC'09). Ref: [Paperin & Sadedin, 2009a].

In the previous chapter I introduced the DPE framework for reasoning about complex systems. The central idea of DPE is that recurrent phase changes in connectivity of interaction networks facilitate a general mechanism for self-organisation and emergence in many complex systems. DPE identifies properties that are typical of systems in different connectivity phases and highlights characteristics commonly associated with phase transitions.

In this chapter I present a wide range of empirical and experimental evidence for DPE processes in complex systems. In many cases, this evidence allows new insights into the behaviour of such systems by accentuating hitherto overlooked mechanisms and highlighting previously unknown parallels and common mechanisms.

Beginning with physical complex systems, I examine biological and socio-economic systems, as well as mathematical and computational systems. The evidence presented here is drawn primarily from previous studies. In later chapters I discuss detailed simulation experiments directed to investigate how DPE can account for specific emergent properties of complex systems.

8.1. Non-living systems

DPE phase changes play a key factor for self-organisation in physical systems across many scales. Here, I discuss examples spanning geomorphologic processes in landscapes through to star formation in galaxies back to DPE processes at the molecular level employed in manufacturing and biotechnology.

Recurrent changes in connectivity of underlying networks play an important role in landscape geomorphology. For instance, repeated flips in connectivity between water bodies drive the formation of many typical geographical features within river basins. For instance, backswamps are moist, swampy areas of a floodplain located away from the main stream [Ritter, 2006]. They are formed and sustained through recurrent connections to the main channel occurring during floods. Heavier sediment particles are dropped by the flood water

close to the main channel and form natural levees [Hudson, 2005], while finer alluvium is carried a greater distance. The fine material can hold significant amounts of water and drains slowly creating wetland areas. After the flood retreats, the alluvium becomes denser and swamps can develop. Backswamps can increase in size and volume through repeated restoring and breaking of water connectivity to the main channel by subsequent floods [Summerfield, 1991; Leopold et al., 1995].

Recurrent water connectivity also affects the morphology of braided river channels [Reinfelds & Nanson, 1993; Warburton, 1994]. These are initially formed when a stream is divided into several smaller channels by the accumulation of in-channel deposits. When water levels rise, for instance through seasonal discharge increase, the braids can become reconnected; when water levels falls subsequently, the stream often settles on a new, more efficient, layout of channel braids. Such repeated transitions of between-channel connectivity contribute to continual morphological diversity within a floodplain [Summerfield, 1991; Reinfelds & Nanson, 1993; Warburton, 1994; Leopold et al., 1995; Ritter, 2006].

On larger scales, geomorphologic landscape connectivity describes the extent to which sediments are transported between distant sites by erosive processes, such as streams and flows [Brierley et al., 2006]. Such landscape connectivity has crucial impacts on landscape morphology [Fryirs et al., 2007; Hazel, 2008] and biophysical fluxes [Brierley et al., 2006]; it affects the transmission of perturbation responses through catchments and is thus a key factor for landscape sensitivity towards disturbances [Thomas, 2001; Brierley et al., 2006; Fryirs et al., 2007].

A number of factors can impede or facilitate sediment transportation, resulting in constant fluctuations in regional connectivity patterns [Brierley et al., 2006; Fryirs et al., 2007]. For instance, slope-channel connectivity transitions relate to the frequency with which channel processes rework materials derived from hillslopes, and channel-floodplain connectivity is driven by frequency and magnitude of flooding events [Fryirs et al., 2007]. Dynamically occurring landscape features affect sediment transportation to and through channels, effectively adding and removing transportation links. For instance, gully systems, which facilitate sediment transportation towards larger channels, often demonstrate disconnecting and reconnecting behaviour [Heede et al., 1988]. Conversely, buffers or barriers can impede sediment flow with different frequencies [Fryirs et al., 2007].

Intrinsic changes in spatio-temporal landscape connectivity patterns cause changes to operation of geomorphologic processes and are thus a significant source of morphological variety [Hazel, 2008]. Such connectivity phases influence size and location of catchment areas [Fryirs et al., 2007], as well as composition patterns of sediment deposits [Hazel, 2008]. Landscape connectivity phases thus bears profound significance for inferences about past climates and the nature and mechanisms of landscape change [Thomas, 2001; Fryirs et al., 2007].

DPE processes can also be observed at a cosmological scale, where network phase transitions appear to play an important role: Smolin [1995] proposes that the pattern of star formations in spiral galaxies can be modelled as a propagation process along a network of dense gas clouds. For stars to form, a cloud of molecular gas must cool to a temperature that permits gravitational collapse. Cold cloud formations tend to be stable, but external perturbations can initiate a collapse [Smolin, 1995]. A primary source of such perturbations are other recently formed stars and supernovae [Elmegreen, 1991]. Consequently, star formation may be self-propagating [Seiden & Schulman, 1986; 1990]. Smolin [1995] suggests that the pattern of observed star formations in many spiral galaxies can be explained if the connectivity of the propagation network repeatedly fluctuates across the critical density. Temperature oscillations required for these fluctuations may arise by feedback between two processes [Parravano, 1988; Parravano et al., 1990]: As the temperature falls, giant molecular clouds condense. External shocks then trigger the formation of new stars. Radiation from these stars heats the surrounding molecular gases, inhibiting further star formation. When the fuel of the heating stars is exhausted, radiation declines and the surrounding gases cool again to a level permitting star formation. Given enough molecular clouds, such a self-propagating network of star formation can maintain itself throughout large parts of a galaxy [Smolin, 1995].

Alternating phases are common in advanced manufacturing, especially in metallurgical processes involving manipulation of atomic lattices. The process of repeated annealing is a

standard technique for increasing material stability in products manufactured from many solid materials [Henley, 1916; Van Vlack, 1980]. It works by rapid heating and slow cooling of the material thus quickly breaking the connectivity of molecular bonding networks and letting them settle to a more stable configuration. The process must be repeated several times to achieve the stabilising effect.

Polymerase Chain Reaction (PCR) is a widely used biochemical technique that produces large numbers of copies of a DNA sequence. Cycles of heating and cooling break the molecular bonding of DNA double-strands, and then rebuild the strands using single-stranded DNA fragments, free-floating DNA molecules and specific helper-chemicals. Every cycle doubles the amount of DNA and many heating-cooling transitions are necessary to obtain a useful amount of DNA [Mullis, 1990].

8.2. Ecosystems

Interactions among organisms form spatial networks that are structured by many processes, ranging from climatic to geological to ecological to socio-economic. For example, plants are confined to networks of habitat sites that are connected by dispersal. Processes that structure landscapes, such as fire, often induce phase changes in habitat connectivity for particular species [Saunders et al., 1991]. Such phase changes may have profound ecological and evolutionary consequences, which we discuss below.

Ecological and evolutionary dynamics differ strongly between fragmented and connected habitat networks. Highly-connected habitat networks allow processes such as population explosions, adaptive genes, diseases and predators to spread throughout a landscape, homogenizing selection pressures and potentially inhibiting local adaptation or even triggering extinction. Conversely, fragmented habitat networks may increase species extinctions through the gradual loss of isolated demes, increase genetic drift, reduce genetic diversity, and reduce adaptive response to selective pressure [Young et al., 1996]. For species that are adapted to disturbed or marginal habitat (e.g. weeds), fragmentation of undisturbed habitat often provide a connected network of disturbed habitat, hastening their spread [Kruess & Tscharntke, 1994; With, 2002].

The ecological impact of repeated cycles of phase transitions in habitat connectivity is illustrated by postglacial pollen records of North American forests [Davis, 1979]. Species composition in forests remained largely static ("pollen zones") despite gradual climate change during this time, but underwent dramatic transitions following major fires [Green, 1982]. Migrating species were constrained spatially due to the overwhelming abundance of seed sources from dominant species, even though a changing climate might favour better-adapted invaders: it was only when dominant species were cleared away by fire that invading species could succeed. Fire mediated zone boundaries have since been found at sites in many regions [Singh & Geissler, 1985; Tinner et al., 1999; Kershaw et al., 2002].

8.3. Living systems at evolutionary scales

For long-term evolutionary dynamics, the effects of repeated phase transitions in habitat networks are less transparent but no less crucial. Glacial cycles present a well-studied example. Cooling climates can result in falling sea levels, which create land bridges among islands, forming a connected habitat for many coastal species. Simultaneously, many widespread continental species have undergone repeated, drastic reductions in range during glacial periods. Due to topography and other geographic factors, their habitats fragmented and these species became confined to a small number of isolated refugia [Willis et al., 2004]. While isolated in refugia, these species diverged genetically such that their modern-day descendants, although now well connected throughout their geographic range, still exhibit marked patterns of population differentiation reflecting these periods of isolation [Hewitt, 1996].

Parallel patterns of post-glacial migration by multiple species out of the same refugia can explain genetic suture zones, which are geographic regions where numerous hybrid zones in

different species co-occur along the same geographic lines [Hewitt, 2004; Swenson & Howard, 2005]. These lines often also correspond to habitat barriers and range boundaries for adjacent-living sister species, suggesting that repeated glacial cycles act as an engine of speciation [Hewitt, 1996]. For example, meta-analysis of mitochondrial DNA studies in 63 bird species suggested that their adaptive radiations resulted from repeated population fragmentation events triggered by glacial cycles over the last 3 million years [Avice & Walker, 1998]. Similarly, the evolution of new grasshoppers, macaques, pond-skaters, cicadas, bees, butterflies and beetles on Sulawesi has been attributed to habitat fragmentation caused by climate change [Butlin et al., 1998].

Increasing habitat connectivity allows species to spread and diversify, implying that phase changes in habitat connectivity may rapidly transform biota across very large areas. Rapid evolution of new ecotypes and species is often traceable to invasion of new habitats, especially when selection pressures are spatially distinct and mating is local [Lande, 1981; Gavrillets & Vose, 2005]. For example, the shore-dwelling snail *Littorina saxatilis* underwent repeated parallel ecotype evolution within islands that arose in the last ten thousand years [Johannesson, 2001]. Similarly, replicate adaptive radiations in African cichlids followed their invasion of unoccupied lakes [Sturmbauer, 1998]; distinct stickleback ecotypes have arisen repeatedly following migration into streams [Rundle et al., 2000]; and the apple maggot fly, an invader that arrived in the United States during the 1800s, has since diverged into several specialised forms that mate and feed on different native plants [Filchak et al., 2000]. In each case, increased habitat connectivity allowed invasion of novel environments, triggering diversification.

Large-scale habitat fragmentation thus influences evolutionary novelty and diversity in several ways. Long-term isolated habitat patches are vulnerable to extinction [Andren, 1994], but facilitate genetic divergence through drift, local selection, and sexual selection [Endler, 1977]. Newly connected habitats allow invasions by species which may then diversify through ecological speciation (but also drive existing species to extinction) [Schluter, 2000]. Habitats that are connected over long periods permit the global spread of adaptations and well-adapted species, but eliminate less competitive forms. Thus each connectivity phase, and the transition between the phases, provides fundamentally different evolutionary conditions.

A phase transition may also occur in evolutionary state transition networks. Eldredge and Gould [1972] documented the pattern of punctuated equilibrium: at large timescales, the fossil record is dominated by long periods of apparent stasis punctuated by mass extinctions, which are followed by brief periods of rapid change and diversification. Remarkably, the resultant qualitative pattern resembles that of the pollen record data described above, but on much larger timescales [Gunderson & Holling, 2002]. Gould [2002] argued that evolving species rapidly reach local optima due to the constraints of existing adaptations. Consequently, evolution of novel adaptations is confined to periods of reduced competition occurring mainly after mass extinction. In other words, the state transition network for evolving species becomes sparser and ultimately fragments in the absence of disturbance; by deleting competitors, mass extinction provides a richly connected state transition network for the survivors.

8.4. Perception and cognition

A number of seminal studies suggest that mechanisms involving connectivity phase changes may play an important role in human perception and cognitive development.

One of the most influential theories in twentieth century psychology was Piaget's theory of cognitive development [Piaget et al., 1985; Piaget & Inhelder, 2000], which is built around *schemas*: mental constructs used in perception, memory, and problem solving. Schemas are formed by two processes: *assimilation* and *accommodation*. Assimilation is the incorporation of new experiences into existing schemas, thereby refining and reinforcing them. Accommodation is the process of coping with novel situations by creating, combining and forming new schemas. Piaget proposed a development mechanism (*equilibration*) that mediates between phases of assimilation and accommodation. In a network model, where nodes represent schemas and edges their relationships, assimilation involves mainly local

interactions, allowing new concepts to develop in relative isolation. In contrast, accommodation is a “global” phase in which links can form between any relevant schemas.

Experiments conducted by Freeman [1975] show that cognition involves sudden transitions in cortical EEG oscillations. Sensory stimuli trigger rapid, chaotic firing of neurons. Subsequently, feedback loops gradually re-synchronise neural firing via a process of pattern selection. These changes suggest two apparent phases: the sensory stimulus flips the neural network into a phase of high connectivity and chaotic activity, followed by gradual decline in overall activation and increasingly constrained local activity. Freeman conjectured that the sensitivity to initial conditions displayed by the cortex during these phase changes enhances the brain’s ability to discriminate sensory inputs [Freeman, 1991; 1992].

Sergent and Dehaene [2004] postulated that consciousness acts as a “global workspace” for brain activity. In their model, sensory stimuli are normally processed subconsciously within local neural circuits. However, specific stimuli may become amplified by the local network, activating long-range connections, particularly via the fronto-parietal cortex which connects many different brain regions. This long-range activation can trigger a sudden, global activation at the whole-brain scale typical of conscious processing.

Extending the above ideas, the daily sleep-wake cycle suggests the possibility of other alternating phase transitions in the brain. In particular, dreams appear to play a crucial role. The nature of dreams is still far from clear, but amongst the many theories a common theme is to treat memory implicitly as a network. The nodes of this network are variously interpreted as mental schemas [e.g. Coutts, 2008], semantic memories [e.g. Stickgold et al., 2001] or other instances of perceptions and experiences. Waking hours can therefore be regarded as a growth phase in which new nodes and edges are added to the network and dreaming as a sorting phase, in which nodes and edges are reinforced or removed [e.g. Blechner, 2001]. Several authors have likened dreams to maintenance operations carried out when an information system is off-line [e.g. Crick & Mitchison, 1983]. Bar-Yam [1997] proposed that the different alternating sleep modes observed in humans, signified by typical EEG patterns, directly correspond to different levels of connectivity between subdivisions within the brain.

8.5. Socio-economic systems

Network thinking is becoming increasingly important in social sciences [Bianconi et al., 2005] and researchers find methods from statistical physics useful for analysing socio-economic network structures [Hors & Lordon, 1997; Roehner, 2005].

In economics, some recent research has focused on emergent phenomena [Marengo & Willinger, 1997] that cannot always be understood with established analytical methods and traditional theories such as the general equilibrium theory or game theory [Kirman, 1997; Morel & Ramanujam, 1999]. These theories concentrate on specific kinds of interactions: In competitive general equilibrium theory individuals interact with the market – global interactions; in game theory individuals interact with each other – local interactions [Davis, 2007]. Many authors [e.g. Kirman, 1997; Davis, 2007] stress that consolidating theories of economic interactions on different scales is central to understanding economy as a complex adaptive system.

Local and global interaction patterns have been identified in various empirical studies. Cobb [1999], for instance, analyses the role of local (personal) and global (corporate) networks in creation of reputation among competing offshore financial centres. Iori et al. [2008] analyse the network topology of Italian overnight bank transactions and find recurrent increases in transaction density close to common month-end closing dates implying increased systemic risk due to higher dependence and exposure. On a wider scale, Mann [1997] analyses threats posed to nation-states by modern socio-economical and geo-political developments. He classifies political and socio-economic interactions into distinct networks according to their “distance”: local, national, inter-national, trans-national and global. For instance, capitalist globalisation increases the significance of inter-national and trans-national networks, as a result weakening the northern states (e.g. the EU), yet strengthening southern nation-states by facilitating their economic development.

Some researchers see drastic social phenomena as phase transitions [Gligor & Ignat, 2001; Levy, 2005] and empirical studies show that social systems can exhibit basins of attraction. For example, Crane [1991] shows that ghetto-like conditions spread through neighbourhoods in a way suggesting the presence of a critical threshold between constrained and global propagation dynamics. It has been suggested [Kuran, 1989] that political revolutions are motivated by “preference falsification” – a critical transition in mass attitude resulting from a mismatch between individuals’ preferences drawn from their personal interactions (private preference) and from their global interactions (public preference). This situation can occur when people support a regime (e.g. for fear of prosecution) despite privately hating it.

Social networks lend themselves particularly well to approaching social phase transitions with graph-analytical techniques due to their explicit network structure. Since the seminal study [De Solla Price, 1965] that led to the discovery of scale-free topologies (see also [section 5.4](#) and [section 13.3](#)) scientific co-authorship networks have attracted much interest. Numerous snapshot [e.g. Newman, 2001b] and longitudinal studies [e.g. Newman, 2001a] have been undertaken, including efforts to relate collaboration connectivity to geographical distance [Katz, 1994]. Connectivity transitions are observed in some, but not all social networks. For instance, since co-authorship links do not disappear, connectivity of such networks tends to increase monotonically.

Links in other types of social networks, however, can decay over time, and such networks tend to exhibit recurrent connectivity changes: A study of professional affiliations in biotech industry [Casper, 2007] has shown that over time network density alternates around very low values while average path length increases, implying the dominance of local connections, probably influenced by company associations. A study of an email communication network in a university [Kossinets & Watts, 2006] has shown significant seasonal changes in connectivity density leading to variations in size of the giant component. A longitudinal study of the Enron corporation email corpus [Diesner & Carley, 2005] concluded that the communication networks become significantly denser and less hierarchical at times of crisis.

Detecting dynamic effects such as phase transitions in empirical data is inherently more difficult than detecting static properties such as scale-free or small world structures. It requires accurate empirical data over extended time periods, which is rarely available in socio-economic sciences [Ehrhardt et al., 2005]. Explanatory variables are typically numerous and cannot be manipulated, experiments are frequently unfeasible and are replaced by surveys and data often collected for other purposes [Gligor & Ignat, 2001]. As a result, simulation models of social networks have received wide attention.

For instance, Ehrhardt et al. [2005] and Marsili et al. [2004] have found that many social networks can exhibit discontinuous phase transitions in their structure. Ehrhardt et al. [2005] demonstrate that in volatile environments social networks can exhibit *emergent* recurrent phase transitions in connectivity density that are caused through internal feedbacks and without any external forcing or disturbances. Solomon et al. [2000] studied a market percolation model of products through social influence and conclude that under simplifying assumptions the connectivity density converges towards the critical threshold, and population sizes affected by products follow a power-law distribution. In many markets the distribution of such product impacts is bimodal [Solomon et al., 2000]. It can be expected that under relaxed assumptions the connectivity of the social influence network will fluctuate across the critical density leading to such distribution.

Castellano et al. [2000] studied a model previously introduced by Axtell, Axelrod et al. [1996] that investigates social consensus and demonstrates emergent dual phase dynamics. In the model, agents exist on a lattice and communicate with their neighbours. Communication links are only active if agents agree and differ on at least one of their “cultural beliefs” each, and agents become more similar in their beliefs during communication. Depending on parameter values, the model either converges on a dominant belief set or on a highly fragmented belief landscape with no active communication links remaining. In cases where a single belief set dominates, the active link density displays emergent connectivity phase transitions: Initially it declines until it reaches a very low value, then it rises well over the percolation threshold and subsequently declines again until it reaches zero. In the real-world social groups do not exist in isolation. When groups that reached consensus come into contact, new active communications links appear. Thus, repeated cycles of creating active links and then using them to achieve new consensus can be expected to occur in real societies.

8.6. Models of modularity

Changing environmental conditions affect the fitness landscape of evolving entities and thus the structure of the state transition networks (section 5.7). Moreover, as evolution can be seen as a search problem for better adapted individuals, changing environments can affect the connectivity structure of the corresponding search space (see also section 8.8).

Arguments from theoretical biology imply that environmental changes that affect selection pressures acting on different traits may be the mechanism behind genetic modularity [Wagner, 1996]: Most individual traits are subject to stabilising selection. However, depending on current environmental conditions, specific groups of traits can be subject to directional selection. Such trait groups are likely to evolve pleiotropic genes (genes that influence multiple traits). Conversely, genes that simultaneously influence traits that are subject to directional and stabilising selection are likely to be selected against. Thus, genes involved in traits that are often selected for at the same time, may organise themselves into modules that are relatively independent from other trait modules.

The difficulty in formal mathematical and experimental treatment of biological CAS has promoted simulation models as a popular tool for their analysis. Several simulation models support the argument that environmental changes that induce connectivity transitions underlie the evolution of modularity. For instance, Kashtan and Alon [2005] investigated the effect of changing environments on network evolution. They evolved digital circuits and neural networks with objective functions that repeatedly switched between different goals thus recurrently altering the connectivity structure of the search space. Each goal was composed of a different combination of sub-goals. The evolved networks contained densely connected functional modules that were loosely connected to the rest of the network. Network evolution in varying environments also led to the formation of network motifs [Milo et al., 2002] found in real systems such as signal transduction and synaptic neuronal networks [Kashtan & Alon, 2005]. Subsequent work [Kashtan et al., 2007] showed that variability in environmental constraints increased the speed of evolution (decreased the number of generations required to match the objective function). Further experiments [Kashtan et al., 2009] showed that phase changes in habitat connectivity caused by induced extinctions lead to less specialisation and increased modularity in internal structure of simulated organisms.

Similar results have also been obtained in numerical simulations of mathematical systems. For instance, Lipson et al. [2002] represented individuals as matrices. With increasing rate of change of the target requirements, evolving matrices became less coupled and more modular. Variano et al. [2004] investigated network properties of linear systems evolved using a genetic algorithm. They evolved directed graphs, each represented as an edge weight matrix M , and selected for linear stability of the resulting dynamical system $x' = Mx$. The resultant networks were more modular and hierarchical than random networks with similar structural properties, and were robust to random rewiring. These simulations make few assumptions about the individuals or their environment, implying that modularity in network structures may be fundamentally associated with recurrent phase transitions in search spaces caused by varying environments [Variano et al., 2004].

8.7. Optimisation problems

Phase transitions occur in the complexity of optimisation problems. The k -SAT problem [Cook, 1971] requires finding an assignment for a set of Boolean variables that satisfies a number of constraints (such assignments are called “ground-states”). Such problems can be seen as networks where nodes represent variables and links connect variables that participate in the same constraint. In general, k -SAT can only be solved in a timeframe that is exponential in the order of participating Boolean variables (assuming the NP-completeness conjecture [Cook, 1971] holds). However, if the number of constraints is small (the density of the constraint network is low), k -SAT instances can often be solved efficiently. Numerical experiments performed on k -SAT [Achlioptas et al., 2005; Hartmann & Weigt, 2006], graph colouring [Achlioptas et al., 2005] and other problems show that as the number of constraints (i.e. the connectivity of the constraint network) increases, the probability that all constraints can be satisfied concurrently undergoes a sharp transition from near 1 to near 0. Simultaneously, the complexity of determining the ground-states undergoes a sharp transition from “easy” to “complex”. Because these problems are NP-complete, a large body of practically occurring constraint optimisation problems can be reduced to these. The complexity phase transition is thus thought to be a widely occurring phenomenon. The review by Martin et al. [2001] treats complexity phase transitions in optimisation in detail.

8.8. Heuristic algorithms

Phase transitions within interaction networks are implicit in many optimisation algorithms. They take the form of mediation between local and global search. For instance, in simulated annealing [Kirkpatrick et al., 1983; Cerný, 1985] the temperature schedule is used to arbitrate between local and global search steps. This corresponds to mediating between neighbourhood interactions (short search steps) and global interactions (distant search steps) between candidate solutions. Another example is the great deluge algorithm [Dueck, 1993] that uses the analogy of a walker who walks around the solution space avoiding water floods of different levels. The water level thus controls the connectivity of the transportation network in the solution space acting similar to an annealing schedule and thus mediating between local (high water level, poorly connected transportation network) and global (low water levels, well connected transportation network) search. Similar ideas have been employed to improve the performance of a variety of optimisation techniques that are prone to being caught in undesirable local optima when applied to non-smooth search spaces. This includes for instance the backpropagation learning algorithm for artificial neural networks [e.g. Ramamoorthy & Shekhar, 1989], the particle swarm optimisation algorithm [e.g. Wang & Li, 2004; e.g. Liua et al., 2005], genetic programming [e.g. Cordon et al., 2002] and support vector machines [e.g. Sun & Sun, 2005; Lin et al., 2008]. In all these cases the connectivity of the search space is changed from well connected (global search, exploration) to poorly connected or disconnected (local search, exploitation). The phase transition occurs in that direction and only once or a few times. However, natural processes are typified by repeated connectivity phase transitions in both directions.

Arguably, optimisation algorithms supplemented with simulated annealing style techniques may be improved by incorporating repeated connectivity phase transitions in both directions. For instance, this has been demonstrated by Kirley et al. in a modified version of the cellular genetic algorithm [Whitley, 1993; Alba & Dorronsoro, 2008] based on insights from population dynamics and landscape ecology [Kirley et al., 1998; Kirley, 2002]. A population of individuals (candidate solutions) evolves on a grid with local competition and mating. The grid is subjected to intermittent “disasters” that wipe out areas of the grid and make them temporarily uninhabitable. This cellular genetic algorithm modified in this way outperforms the standard algorithm on a number of hard test problems [Kirley et al., 1998; Kirley, 2002]. Two interaction networks are important here: The network of populated grid cells determines the amount of gene (information) flow. When it is fragmented, the population evolves not as a whole, but as divergent sub-populations, which can slow premature convergence. The second is the network of the free grid cells. When this network is well connected, free cells can be repopulated without substantial fitness competition. Despite an initial fitness disadvantage such candidate solutions can potentially evolve towards a different, possibly

better local optimum. Similar ideas have been employed in a number of different parallel genetic algorithms; a comprehensive review is provided by Cantú-Paz [1998].

Another way to take advantage of repeated changes in connectivity of the search space is based on combining genetic algorithms with Lamarckian Evolution or the Baldwin Effect. Lamarckian Theory [Lamarck, 1815] held that traits acquired (learned) during a life time can be inherited into the next generation; the Baldwin Effect [Baldwin, 1896] emphasises the importance of selection for learning ability rather than for specific genetically “hardcoded” abilities. In biology, Lamarckism has been displaced by Darwin’s evolutionary theory and the Baldwin Effect remains debated, but both ideas have been applied successfully in computing [e.g. Morris et al., 1998]. Artificial Lamarckian Evolution uses local search at every generation to improve candidate solutions which are then recombined through genetic operations that permit more distant (global) jumps in the search space [Gen & Cheng, 1997]. Conversely, when Baldwin Effect is applied to artificial evolution, results of local search performed at every generation are used to evaluate the fitness of candidate solutions, but are not fed back into the evolving genotypes. In non-smooth search spaces, the resulting selective pressure towards good starting points for local optimisation helps directing the global evolutionary search [Hinton & Nowlan, 1987]. For some optimisation problems, hybrid methods between Lamarckism and the Baldwin Effect have proven successful [e.g. Ishibuchi et al., 2005].

The Markov clustering algorithm (MCL) [Van Dongen, 2000b; 2000a; 2000d; 2000c] is a further example for a heuristic algorithm based on repeated phase changes. The MCL algorithm finds cluster structure in graphs by simulating random walks without the need for additional information such as number of clusters or distances between cluster items. The algorithm is based on two alternating operations: The first operation (“expansion”) computes a power of the Markov matrix that defines the graph. This computes the probability of multi-step transitions along the graph and reinforces intra-cluster links. The second operation (“inflation”) computes a power > 1 of each matrix element. This re-enforces stronger links in relation to weak links. This process of repeated flips between intra-cluster interactions and direct neighbour interactions eventually leads to the separation of the graph into different clusters. The MCL algorithm has proven successful in identifying cluster structure in networks obtained from empirical data [e.g. Enright et al., 2002] and from simulation studies (e.g. [section 11.4.2](#)).

8.9. Summary

In this chapter I highlighted the role played by networks in natural and artificial systems. As this analysis shows, phase changes occur in complex systems of virtually every kind. What is more, in almost every case, those phase changes can be traced to connectivity avalanches within underlying networks. This uncovers a common pattern across many kinds of systems: Recurrent phase changes in network connectivity facilitate emergence and self-organisation.

In the following chapters I will use this insight to study specific systems. Using simulation models I first investigate the role of connectivity changes on evolution of landscape ecosystems ([part IV](#)). In [part VI](#) I research the role of changes in connectivity in the formation of important network structures, such as modular and scale-free topologies.

Part IV

Perpetual Novelty and Sustained Diversity in Biological Systems

Here I seek to improve our understanding of mechanisms behind the emergence of two crucial properties of CAS – perpetual novelty and sustained diversity. I concentrate on two key areas: First – recurrent external disturbances that alter local environments, unbalance systems away from attractor states, and cause DPE connectivity transitions in interaction networks that underlie bio-ecosystems.

Second – complex, biologically realistic fitness landscapes. Recent research provides new insights into topology of biological fitness landscapes, and models based on these insights offer new conclusions about genetic diversity and discovery of novel ecological habitat niches.

Chapter 9: Background.

Theoretical implications about novelty and diversity in biological CAS provided by the DPE framework stress the role of external disturbances in causing connectivity phase changes and pushing systems away from attractors. Empirical evidence from previous studies supports these implications.

Chapter 10: DPE and Perpetual Novelty and Sustained Diversity in Landscape Ecosystems.

A simulation model of adaptive radiation in landscapes is presented. Model organisms normally exist within a connected landscape, intermittent disturbances (fires, impacts, etc.) flip the system into a poorly connected phase and affect the environment in disturbed areas. Results show that this DPE mechanism facilitates complex diversity and perpetual novelty in the ecosystem. An analysis of the fitness landscape employed in the simulation provides a richer context for the results, and shows that the employed genetic model exhibits biologically unrealistic properties, which, however, do not affect validity of the results.

Chapter 11: Evolution on Holey Fitness Landscapes.

Based on the discussion in the previous chapter and recent results in speciation theory, a new genetic model is proposed that combines the advantages of existing techniques, while avoiding some of their drawbacks. A numerical analysis of the genetic fitness landscape exhibited by the new model informs future studies and shows that the model exhibits many biologically realistic properties, including properties implied by the holey fitness landscape concept. A simulation of parapatric populations with holey fitness genetics provides new insights into maintenance of reproductive isolation and diversity under migration. Techniques used in the simulation experiments may also be employed for to improve some evolutionary optimisation algorithms.

Chapter 9:

Background

CAS exhibit sustained diversity and perpetual novelty, however, the mechanisms behind the emergence of these properties are not well understood. DPE provides a key mechanism for novelty and diversity in biosystems – recurrent connectivity phase changes caused by intermittent external disturbances. Previous studies reveal a wealth of empirical evidence for the role that connectivity changes caused by disturbances play in species variety and other kinds of ecosystem diversity. Subsequent chapters use simulation experiments to investigate the theoretical ideas developed in this chapter in more detail.

The material presented in this chapter is based on the following publications:

-  G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
-  G. Paperin, D. G. Green, S. Sadedin, T. Leishman (2007):
A Dual Phase Evolution Model of Adaptive Radiation in Landscapes.
3rd Australian Conference on Artificial Life (ACAL'07). Ref: [Paperin, Green, et al., 2007b].

9.1. Introduction

CAS exhibit a sustained diversity of their locally interacting components. In the absence of a global controller CAS exhibit far-from-equilibrium dynamics, and permanent novelty and adaptation [Holland, 1995; Levin, 1998]. However, our understanding of how these properties emerge remains limited. In this part of the thesis I investigate some of the mechanisms that facilitate and maintain sustained diversity and perpetual novelty in biological CAS. In particular, I concentrate on two of such mechanisms: DPE phase changes in habitat connectivity ([chapter 10](#)) and complex biological fitness landscapes ([chapters 10 and 11](#)).

I begin by discussing the insights that the DPE framework provides into diversity and novelty in CAS ([section 9.2](#)). External disturbances play a key role in the emergence of sustained diversity and novel adaptations in at least two respects: First, recurrent disturbances alter the organisation of component interactions in systems resulting in ever new self-reorganisation of interaction networks. Second, disturbances unbalance systems away from local optima states thereby pushing them into phases of evolutionary exploration and reorganisation. [Section 9.3](#) discusses a number of previously published empirical studies that provide evidence for the role of disturbances in facilitation of biological novelty and diversity.

To support my findings, I introduce an individual-based simulation model of species evolution in landscapes ([chapter 10](#)). In the simulation, recurrent phase transitions in landscape connectivity caused by intermittent disturbances lead to continuous usage of new habitat niches and to increased diversity. Analysis of simulation results reveals that adaptive radiation is facilitated not only through changes in landscape connectivity, but also through environmental alterations in disturbed areas that affect the evolutionary FL. Thus, I conclude [chapter 10](#) with a detailed examination of the model FL and its evolutionary properties. The examination shows that the model FL exhibits a number of properties not found in biological FLs, however, these differences do not affect the validity of simulation results.

Based on examination of the model FL and on recent results in speciation theory, I propose and apply an improved genetic model ([chapter 11](#)). The model combines existing techniques and incorporates both, an explicit, biologically realistic treatment of biochemical genotype viability, as well as an approach for relating genotype fitness to the ecological environment

([section 11.1](#)). A numerical analysis of the FL structure exhibited by the proposed genetic model confirms that the model results in several biologically realistic properties ([section 11.3](#)).

Application of the analytical HFL concept within individual-based computer simulations is problematic due to differences between analytical and numerical approaches ([section 11.2.2](#)) and due to the exponential-time algorithms involved ([section 11.2.3](#)), however I develop techniques that allow creating and using explicit HFL models with acceptable run-time performance ([section 11.2.3](#)). Based on the new techniques I examine how HFL-genetics change our understanding of the maintenance of genetic diversity in parapatric (partially separated) populations in the face of migration ([section 11.4](#)). Earlier studies implied that reproductive isolation caused by ecological factors will collapse under migration. However, simulation results show that HFL-genetics can often maintain diversity and facilitate exploration of new niches and speciation.

The new simulation techniques may also prove useful in biologically-inspired engineering. For instance, I speculate that the technique used for creating HFLs suitable for individual-based simulations of biological systems can also be useful for combating premature convergence in evolutionary optimisation algorithms ([section 11.5](#)). More generally, work presented in [chapter 11](#) allows to conclude that HFL-genetics provide a fruitful concept in future studies of adaptation and diversity in natural and artificial systems.

9.2. Perpetual novelty and diversity in CAS

The material presented in this section is based on the following publications:

-  G. Paperin, D. Green, S. Sadedin (2010):
 Dual Phase Evolution in Complex Adaptive Systems.
 Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
-  G. Paperin, D. G. Green, S. Sadedin, T. Leishman (2007):
 A Dual Phase Evolution Model of Adaptive Radiation in Landscapes.
 3rd Australian Conference on Artificial Life (ACAL'07). Ref: [Paperin, Green, et al., 2007b].

There is a large amount of evidence (section 9.3) that CAS generally tend to self-organise towards a stable, balanced state. A number of adaptive mechanisms present in CAS cause these systems to exhibit little large-scale variation over long periods of time: For instance, analytical [Watson & Lovelock, 1983; Weber, 2001] and computational [Lenton & Van Oijen, 2002] models show that lower-order local dynamics (e.g. feedback processes) can stabilise systems over a large range of external forcing, and that higher order local dynamics (e.g. evolutionary processes) can greatly increase the stabilising effect (see section 4.3 for a discussion of different orders of system dynamics). Established ecosystems, for example, tend to be resilient against forcing that could drive them away from current attractors [Gunderson & Holling, 2002; Scheffer & Carpenter, 2003; Walker et al., 2004].

However, the adaptive forces that underlie global stability of CAS also inhibit novelty. Stabilising selection acting on system components at various scales, as well as on topology and interactions, may drive a system as a whole to a local optimum state, halting innovation [Holland, 1995]. At least two mechanisms that can act against such long-term stasis can be identified:

One such mechanism is *co-evolution*. Local adaptation of system components driven by selection may affect the selection criteria (the fitness landscape) for other components, which will also adapt as a result. The adaptation of the components affected in such a way may in turn cause changes in the fitness landscapes of other components, including the components which initiated the changes in the first place. The feedback loops which can arise in this way may function as sources of perpetual novelty because the selection acts on random variation and the results of such feedbacks may be different for each loop. However, it is not clear that co-evolution is capable of providing the degree of innovation observed in many natural CAS. For instance, current models suggest that selection, not variation, drives speciation and that co-evolutionary feedback loops are likely to rapidly (on evolutionary timescales) lead to stable local optimum states [Gavrilets, 2004].

The second mechanism that may function as a source of continual novelty in CAS is *recurrent external disturbance*. Disturbances can cause systems to flip from high connectivity phases dominated by stabilising selection to low connectivity phases of evolutionary exploration leading to ever new and diverse adaptations.

It has been shown, that evolutionary innovations in various natural CAS often coincide with external perturbations (e.g. [Green, 1982], and section 9.3 below). External disturbances can affect both system components and interaction networks, thus moving systems away from local optima (and so increasing the connectivity of corresponding state transition networks). While a disturbed system remains far from a fitness optimum, any chance variation of local components or substructures may provide a better adaptation to the local constraints and selection will facilitate the proliferation of such change. Over time, components and their interactions at multiple scales are driven towards new local optima and selection again inhibits variation and stabilises the system (state transition network connectivity decreases). Future perturbations continuously affect the system and flip it between different connectivity phases leading to ever new and diverse adaptations.

The above mechanism results in a fruitful interpretation of the DPE phase change mechanism: CAS develop towards a *balanced phase*, where they exhibit little large-scale variation (on evolutionary timescales). In this state they are stabilised through self-

organisational processes of various orders (section 4.3) including feedback and selection. The balance phase is recurrently disturbed by external perturbations, which unbalance systems and flip them into a phase in which they exhibit variation on all scales. Over time, stabilising processes drive systems into a new balance-state (Figure 9.1).

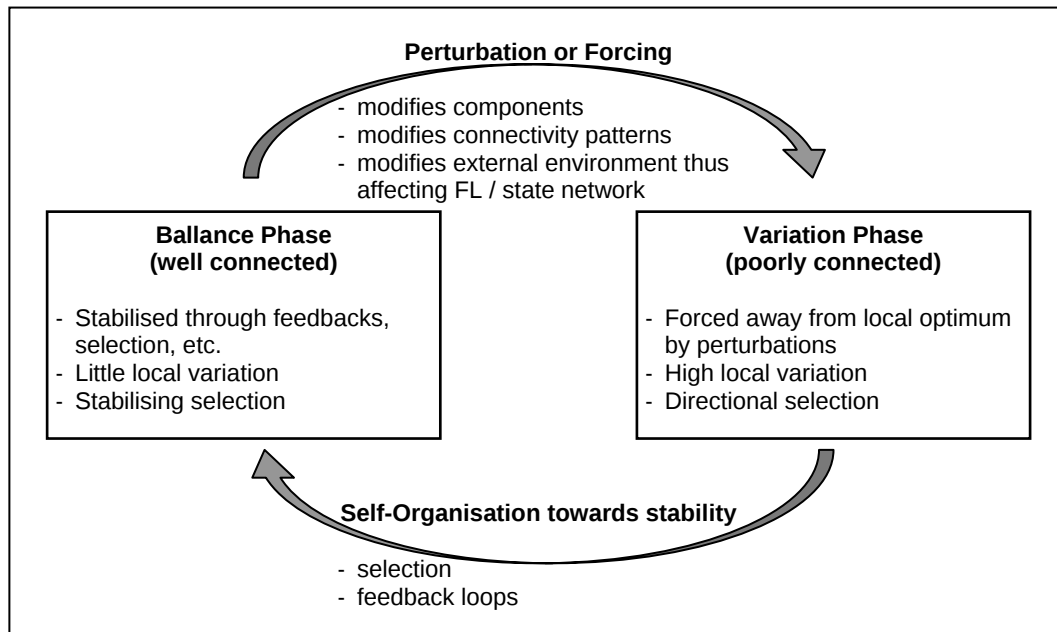


Figure 9.1. DPE phase transition cycles from the perspective of CAS dynamics. Systems flip between balance and variation phases. External disturbances unbalance stable systems, variation facilitates evolutionary exploration, internal pressures drive the system into a new stable state. Compare this figure to Figure 7.1 on page 67.

9.3. Evidence: Recurrent environmental perturbations in biological evolution

The material presented in this section is based on the following publications:

 G. Paperin, D. G. Green, S. Sadedin, T. Leishman (2007):
A Dual Phase Evolution Model of Adaptive Radiation in Landscapes.
3rd Australian Conference on Artificial Life (ACAL'07). Ref: [Paperin, Green, et al., 2007b].

In [chapter 8](#) I discussed a large amount of empirical evidence for the role of DPE phase transitions in the emergence of different complex system properties. Here, I add to that evidence by reviewing a range of previous studies, while concentrating in particular on the role that connectivity changes caused by external environmental perturbations play in speciation and ecosystem diversity.

Evolution occurs in fits and starts. This pattern of change is clear in the geologic record. The system of geologic classification reflects a history in which similar assemblages of fauna and flora predominated for long time periods of time, often tens of millions of years. These periods are punctuated by abrupt changes in species composition. Research has revealed that the changes between geologic eras are associated with mass extinction events. Recognising this pattern led Eldredge and Gould [1972] to put forward the idea of *punctuated equilibrium*, in which the general pattern of evolution is constant composition punctuated by mass extinctions, followed by brief periods of rapid adaptive radiation.

Alvarez et al. [1980] found evidence that the Cretaceous-Tertiary boundary was associated with an impact of a large comet. Research since then has produced evidence that asteroid impacts, volcanic activity and climate change are associated with many other geological boundaries as well.

Green et al. [2000] proposed a mechanism to explain the above observations: For most of the time, competition for space or niches within a landscape impose selective pressure. Established populations restrict the spread of invaders. Widespread populations are genetically connected and genetic variation is suppressed. When a major disturbance occurs, the above patterns are reversed: vast areas of free space are opened up; suppressed species are free to expand into the new territory; selective pressure becomes negligible; and established populations become fragmented.

There are striking similarities between species evolution, on a scale of millions of years, and forest change, which occurs on a scale of thousands of years. Forest history, as recorded by preserved pollen, shows that during postglacial (the last 12,000 years), forest composition changed in fits and starts. This pattern is reflected by the systematics used by palaeontologists, who divide the postglacial history into pollen zones. The zones have more or less constant composition, with rapid changes from one zone to another. Studies have shown forest composition remained resilient against climatic changes, and that major forest fires triggered rapid transformations, with the new species composition being determined by climate at the time [Green, 1982].

In certain regions, recurrent fluctuations in landscape connectivity have been linked to the evolutionary radiation of whole groups of animals. In lakes of east Africa, for instance, the explosive speciation in cichlid fishes has been linked to intermittent changes in water level [Sturmbauer & Meyer, 1992; Kornfield & Smith, 2000]. During periods of high water level, environments are connected, but become fragmented when water levels are low. Similarly, Hewitt [2004] argues that repeated glaciations throughout the Quaternary (period from ca. 2.6×10^6 years ago to present) caused species ranges in North America and Europe to fragment, leaving surviving populations in isolated refugia. These isolated populations diverged genetically, but later reunited, creating a complex genetic patchwork within some species, and sometimes leading to speciation. Affected species include the European hedgehog, *Chorthippus* grasshoppers and bears. Numerous similar parapatric species occur in the mountains of Sulawesi, Indonesia, that are thought to have diverged during periods of

habitat fragmentation [Butlin et al., 1998]. Taxa include *Chitaura* grasshoppers, macaques, pond-skaters, cicadas, carpenter bees, butterflies, limacodid moths and tiger beetles. Likewise, Amazonian insects are thought to have diversified in response to fluctuating connectivity in forest canopy density [Cowling et al., 2001].

Habitat fragmentation at fine temporal scales does not always lead to speciation: instead, the outcome may be formation of genetic suture zones, where populations that have diverged while separated meet and interbreed. Some authors suggest that fragmentation may even contribute to evolutionary stasis. For example, Bennet [2004] argues that Milankovitch climate oscillations, which occur on the order of 10^4 – 10^5 years, cause continual changes in the direction of selection, preventing (sub-)populations from adapting locally and therefore speciating. Similarly, Coope [2004] notes the prevalence of stasis among temperate Quaternary insect species despite the appearance of incipient ecological species in modern fragmented fauna. Clearly, there is still much to be discovered about the impact of habitat fragmentation on evolution. The simulation model investigated in the next chapter is one step in this direction.

Chapter 10:

DPE and Perpetual Novelty and Sustained Diversity in Landscape Ecosystems

A simulation model of adaptive radiation in landscapes is presented. In the model, organisms normally exist within a connected landscape in which selection maintains them in a stable state. Intermittent disturbances (such as fires, commentary impacts) flip the system into a poorly connected DPE phase, in which populations become fragmented, freeing up areas of empty space in which selection pressure changes and genetic variation predominates. The simulation results show that this DPE mechanism facilitates the appearance of complex diversity and perpetual novelty in a landscape ecosystem. A detailed analysis of the FL that emerges from the genotype model employed in the simulation provides a better interpretation context for the results, assesses the role of directional selection caused by environmental alterations in disturbed areas. The analysis shows that the genetic FL employed in the simulation model exhibits several biologically unrealistic properties, however, these issues do not affect the validity of the results. The subsequent chapter discusses improvements to the genetic FL utilised here.

The material presented in this chapter is based on the following publications:

-  G. Paperin, D. G. Green, S. Sadedin, T. Leishman (2007):
A Dual Phase Evolution Model of Adaptive Radiation in Landscapes.
3rd Australian Conference on Artificial Life (ACAL'07). Ref: [Paperin, Green, et al., 2007b].
-  G. Paperin, D. G. Green, A. Dorin (2007):
Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].

One mechanism that can facilitate perpetual evolutionary novelty in ecosystems is environmental disturbance. It acts by affecting the connectivity of interactions within ecosystems and by unsettling systems away from local optima states (sections 9.2 and 9.3). To support this proposition I use a series of simulation-based experiments with a landscape ecosystem model. The experiments concentrate of the role that DPE phase changes caused by disturbances play in evolutionary dynamics of landscapes.

The experiments are performed on an individual-based simulation model of species evolution in landscapes. In the simulation, recurrent phase transitions in landscape connectivity caused by intermittent disturbances lead to continuous usage of new habitat niches and to increased diversity. Analysis of simulation results reveals that adaptive radiation is facilitated not only through changes in landscape connectivity, but also through environmental alterations in disturbed areas that affect the evolutionary FL. Thus, I conclude this chapter with a detailed examination of the model FL and its evolutionary properties. The examination shows that the model FL exhibits a number of properties not found in biological FLs, however, these differences do not affect the validity of simulation results. Based on examination of the model FL and on recent results in speciation theory, I propose and apply an improved FL model in the next chapter.

10.1. Simulation model of adaptive radiation in landscapes

The material presented in this section is based on the following publications:

📖 G. Paperin, D. G. Green, S. Sadedin, T. Leishman (2007):
A Dual Phase Evolution Model of Adaptive Radiation in Landscapes.
3rd Australian Conference on Artificial Life (ACAL'07). Ref: [Paperin, Green, et al., 2007b].

In order to further investigate the effect of the DPE-process on evolutionary ecosystems, I use an individual-based simulation model. The model investigates the consequences of recurrent DPE phase changes caused by disturbances on adaptive radiation exhibited by a population of organisms in a landscape. The current model is based on a previous model of adaptive radiation [Gavrilets & Vose, 2005] that did not incorporate DPE.

10.1.1. Overview

Here I describe the model and give parameter values used in the basis scenario. In [appendix A](#) I provide additional details. Sensitivity analysis that systematically varied parameter values is discussed in [appendix A](#) and [section 10.1.2.3](#), as applicable.

The model consists of a population of haploid individuals situated on a two-dimensional landscape. The landscape consists of a 100×100 grid of cells. Each cell has a maximum carrying capacity of up to 4 individuals

The environment allows 60 possible habitat niches, where each niche is represented by a string of 20 bits; a bit represents the requirement, that a particular trait must be present (1) or absent (0) in an individual in order to be well adapted to that niche. The niches are not location specific ([section A.1.1](#)). Individuals' genotypes are also represented by bit-strings of length 20; the bits represent the presence (1) or absence (0) of the above traits (i.e. genotype = phenotype). The individuals evolve to adapt to one of the niches. At each time-step an individual is assumed to occupy the niche which best matches the individual's traits ([section A.1.2](#)).

The *fitness* of an individual is determined in proportion to the hamming distance between the individual's genotype and its niche and is scaled by the niche condition ([section A.1.2](#)). The *niche condition* is a number between 0 and 1 that describes how appropriate a particular niche is within the current environment (e.g., in a desert environment, the niche "hot & dry sand" may have *condition* = 1 and individuals well adapted to that niche will have a high fitness; individuals which are equally well adapted to the niche "cold and wet soil" will have a low fitness as such niche will have a much lower condition in this climate). 30 out of the 60 model niches are called *normal*, they have a *condition* = 1. The other 30 niches describe the environmental conditions short after a disaster. In the basis setup, in which there are no disturbances, the condition of these *disaster-niches* is set to 0.2.

The life-cycle of the model organisms is reproduction — selection — dispersal ([section A.1.3](#)), the generations are non-overlapping. Individuals mate within their occupied cell only. Each individual in a cell is selected once as *mother*. A *father* is then randomly selected from the remaining individuals of the cell (regardless of their niche). If there is only one individual in a cell, it will engage in hermaphroditism. The number of offspring each couple produces is Poisson-distributed with $\lambda = 5$. Reproduction is through free recombination with each trait being inherited from either parent with equal probability. Each offspring is subject to a mutation rate of 10^{-5} per gene, which corresponds to background mutation rate in nature [Russell, 1994]. Individuals mate regardless of their habitat niche. As only 60 of all possible 20-bit strings represent actual niches, the density of well-adapted genotypes is low. Hence, offspring of individuals existing in different niches are likely poorly adapted to either niche and have low fitness. Sub-populations adapted to different niches are thus postzygotically reproductively isolated and represent different species ([section A.1.5](#)).

Once all individuals have mated, the old population is replaced by the offspring. It is then determined which individuals of the new population will survive to the age of reproduction, by reducing the number of individuals in each cell to its carrying capacity. The probability of survival is given directly by an organism's fitness (section A.1.6). Finally, the surviving individuals may disperse across the landscape. With probability of 0.1, each individual will migrate to one of the neighbouring cells (section A.1.7). After all individuals have migrated, the current generation will engage in reproduction and complete the live-cycle.

The model is initialised with a small population of clones of 2 randomly chosen individuals and thereafter simulated for 40 000 generations (section A.1.4).

There are three basic measurements that describe the general behaviour of the model:

- The *number of utilised niches* (or the number of species) at any given generation t is the number of distinct niches at time t occupied by at least one individual.
- The *niche proportion* (species proportion) of niche M at generation t is the proportion of individuals that occupy niche M out of the total modal population at time t .
- The *average niche fitness* of niche M at time t is the average fitness of all individuals that occupy M at time t .

10.1.2. Scenarios

10.1.2.1. Basis scenario

Initially, a series of model runs was performed without any disturbances. The results are generally in line with intuitive expectations and with the dynamics observed by Gavrillets and Vose [2005]. Here I outline the basis dynamics of the model in sufficient detail to interpret further experimental results. A more detailed description and an analysis is given in appendix section A.2.

At the beginning of the simulation there is a burst of adaptive radiation which leads to a large number of niches being utilised. After a while (typically 5 to 10 thousand generations), the number of utilised niches begins to decline slowly. Most niche proportions fall below 0.1 and then either engage in uncorrelated fluctuations or decline to 0. The proportions of 1 to 3 niches typically remain above the threshold of 0.1. Out of these "dominating" niches, one typically grows in proportion slowly, while the others decline accordingly. Sometimes, two rather than one niches gain a stable proportion while the others decline (Figure 10.1, Top).

In addition, some patterns occurred that were not initially expected from the model, but are known to occur in nature. These patterns include the dominance of lower-fitness populations over closed spatial patches [Green, 1989] and the occurrence of stable hybrid zones [Barton & Hewitt, 1985] maintained by a balance between dispersal and selection against hybrids. These observations provide an indication that the model correctly captures the relevant landscape dynamics and support the validity of experimental results (chapter 2 and [Grimm et al., 2005]).

10.1.2.2. Disturbances

The basis scenario was modified such that at each generation a disaster occurred with probability 5×10^{-5} . During each disaster, all individuals in most landscape cells are wiped out. The affected cells are selected by setting a random point in the landscape as disaster centre and wiping out all cells within a certain radius around the centre. The radius is drawn from a normal distribution with a mean of 30 cells. Analysis (section 10.1.4) reveals that if the average radius exceeds a certain threshold, the model behaviour is not sensitive to the impact radius. Then, a new impact centre is randomly selected. This process is repeated until 95% of cells are wiped out. Whatever the nature of a disaster (bush fire, volcanic eruption, disease, etc) it does not only wipe out the population in affected areas, but also alters the local environment. To model this, the conditions of normal niches in areas affected by the

disaster are reduced to 0.2, and the conditions of the disaster-niches in these areas are raised to 1. Sensitivity analysis (section 10.1.4) reveals that reducing the normal niche condition to any value below 0.5 does not significantly affect the qualitative system behaviour. This altered environment is maintained at the impact sites for several thousand generations and then stepwise returned to normal using a linear interpolation between the disaster environment and the normal environment. This strategy is similar to a number of cellular landscape models that incorporate disturbances (e.g. [Green, 1989]).

The general model behaviour in the disaster scenario can be described as follows: A disaster kills a large part of the population, wiping out entire niche-populations. The number of utilised niches therefore drops. However, the remaining population can now expand into the freed-up areas and colonise the disaster-niches available in the disturbed landscape, which leads to a quick growth of utilised niches. As the disturbed areas recover from the disaster, some of the normal niches are re-populated. In addition, some of the normal niches not used before the disturbance are also utilised. By the time the landscape has fully recovered, the number of utilised niches is typically large. As the landscape is fully populated at this point, selection now slowly reduces the number of utilised niches until the next disturbance (Figure 10.1).

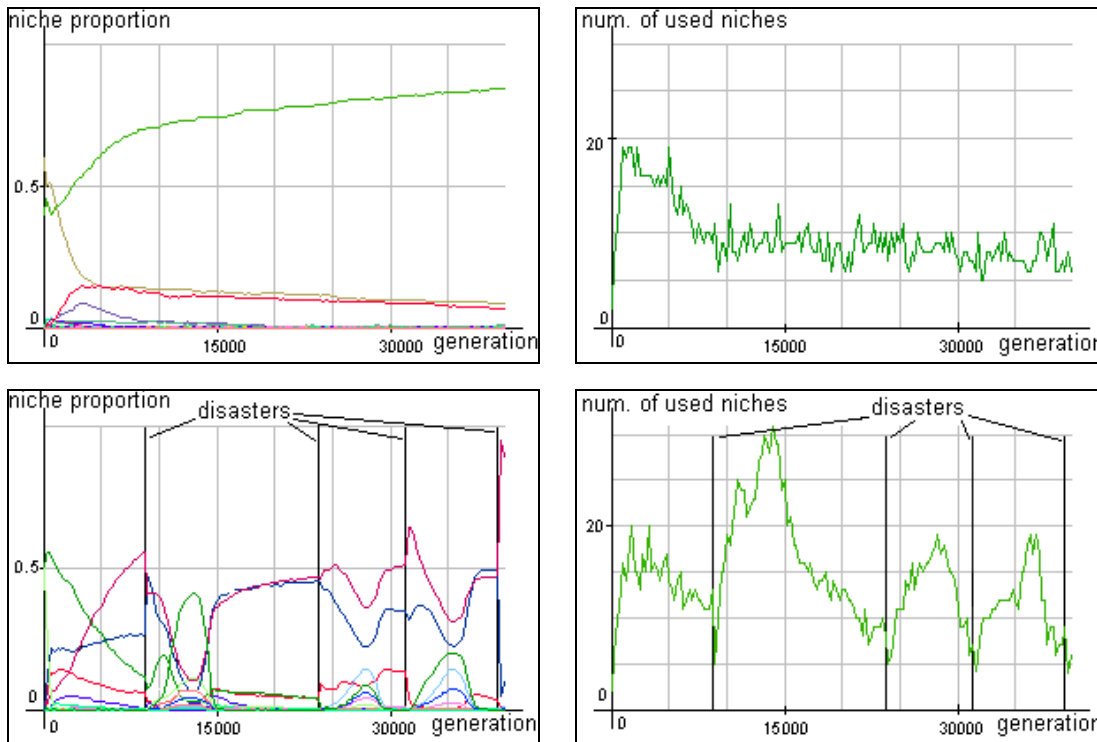


Figure 10.1. Typical model runs.

Basis case (top). **Left:** All niche proportions are displayed. An initial burst of adaptive radiation is followed by populations of 3 niches taking over the entire population. After 40,000 generations it can be expected that a single niche will soon take over the entire landscape. **Right:** Number of utilised niches at each generation from the same simulation run. The graphs suggests that at any time, there are several niches that are occupied but have very low proportions.

Disaster case (bottom). **Left:** All niche proportions are displayed. While in long phases between disturbances the dynamics are similar to the basis scenario, disasters are followed by phases of intense variation. After the effects of a disaster have faded out, new niches appear to be colonised. **Right:** the number of utilised niches drops immediately after a disaster as some niche populations are wiped out. Reduced competition in the freed-up areas allows new genetic variations to colonise new niches.

10.1.2.3. Sensitivity analysis

Sensitivity analysis reveals that the model is sensitive to extremely small parameter values. For instance, the disaster dynamics described here could not always be observed on grids of size 30×30 and smaller. This may be because the spatial distribution of genotype differences is too small on grids of this size. Runs involving 8 and less niches also did not always produce the disaster scenario dynamics. This was mostly due to the fact that all niches were utilised and no further diversification was possible. In addition, selection will always drive the genotypes towards the available niches and a very small number of niches will work strongly against diversification. A very small number of traits (< 10) also tended to not produce interesting dynamics. This is probably due to the small number of possible genotypes in such a setup and a small mutation probability per genotype. Once the parameters exceed some threshold, the qualitative dynamics of the model are not sensitive to specific values. In general, the parameters used for both, the basis and the disaster scenario, lie well above these thresholds. Of course, particularly large parameter values lead to unreasonably long simulation run times. The strategy for choosing parameter values was therefore to choose values as large as possible, while still achieving a reasonable computation time (one model run on a workstation – Pentium 4 Prescott, 3.2 GHz, 1GB RAM – takes approx. 9 hours).

10.1.3. Results

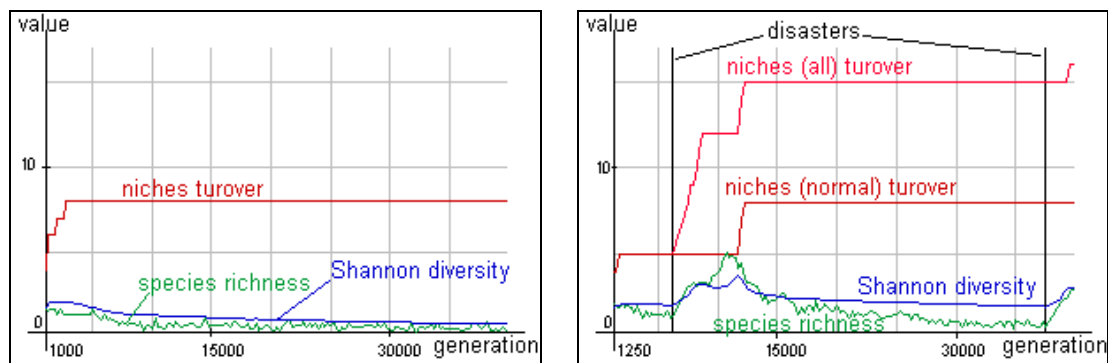


Figure 10.2. Measurement results of typical runs. Measurements were taken every 250 generations. Measurements taken during variation phases (when *crowdedness* < 0.9) are removed.

Basis scenario (left): Initially, a burst in *niches turnover*, *Shannon diversity* as well as *species richness* can be observed. After a while, selection inhibits diversity and further turnover of niches.

Disaster scenario (right): Variation phases induced by disasters are followed by a jump in diversity as well as the discovery of new habitat niches.

The main research question posed to the current model is to investigate the connection between disturbances and evolution in landscapes. The hypothesis is that by affecting landscape connectivity disturbances facilitate novelty and diversity. Two quantities measured in the model offer insights into this connection: niche turnover and diversity.

Niche turnover is the number of niches that have been utilised throughout the history of the landscape up to a certain point of time. Note that as niche populations are expected to be reproductively isolated (section A.1.5), this corresponds to *species turnover*.

Diversity is a necessary basis on which selection acts. While continual evolutionary innovation does not necessarily lead to increased diversity, a certain degree of diversity is required for such innovation. There are various ways to measure diversity [Purvis & Hector, 2000; Rojas, 2004]. As phylogeny (taxonomy) based measures are not meaningful within the model, I apply several information theoretic measures: *species richness* (taken here as *niches richness*) as well as the *Simpson*, *Shannon* and *Pielou* measures of diversity [Rojas, 2004].

In order to ensure the comparability of results between the scenarios, I calculate the turnover of *all* niches as well as the turnover of *normal niches only*. This is because disaster-niches are expected to be rarely utilised in the basis scenario.

	Simpson diversity				Shannon diversity				Pielou diversity			
	Min val.	Max val.	Mean	Std. div.	Min val.	Max val.	Mean	Std. div.	Min val.	Max val.	Mean	Std. div.
Basis scenario	0.01	0.75	0.42	0.19	0.04	2.20	1.10	0.50	0.03	0.58	0.35	0.13
Disaster scenario	0.18	0.90	0.60	0.17	0.47	3.81	1.87	0.76	0.26	0.69	0.46	0.11

	Species richness				Niches turnover (all)				Niches turnover (normal)			
	Min val.	Max val.	Mean	Std. div.	Min val.	Max val.	Mean	Std. div.	Min val.	Max val.	Mean	Std. div.
Basis scenario	0.09	1.89	0.82	0.52	2	9	5.38	1.88	2	9	5.29	1.79
Disaster scenario	0.09	4.15	1.86	1.06	6	22	14.8	3.7	3	11	8.05	1.88

Table 10.1. Niche turnover and diversity measured at the end of each run. 20 control runs were performed for each scenario. Given are the minimum and the maximum values, the average, and the standard deviation. Measurements taken during variation phases (when *crowdedness* < 0.9) are discounted. Two-sample *t*-tests confirm with confidence 99% ($\alpha = 0.01$) that all means are higher in the disaster scenario.

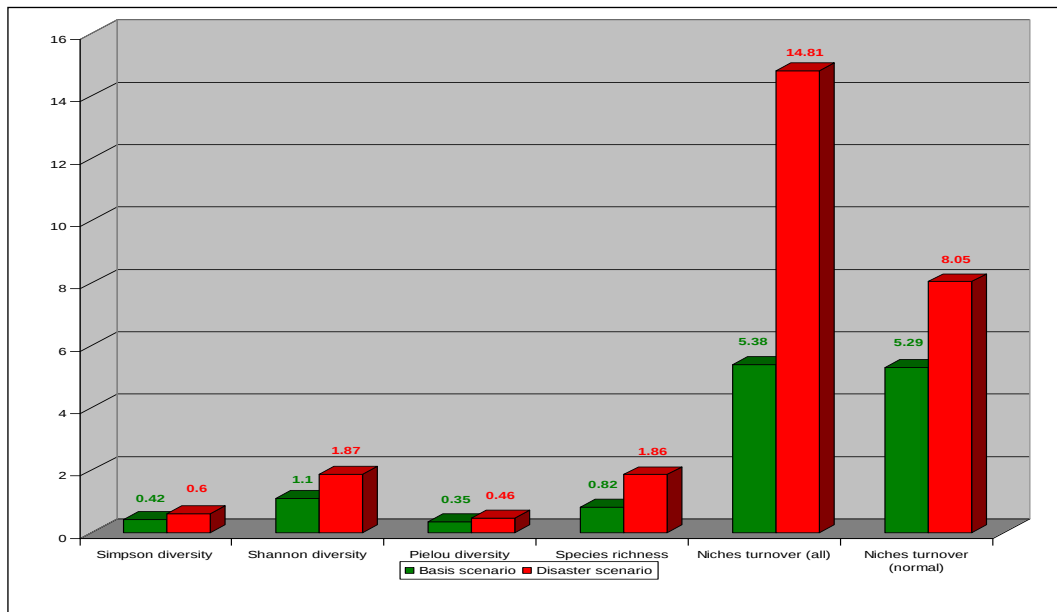


Figure 10.3. Niche turnover and diversity measured at the end of each run. 20 control runs were performed for each scenario and measurements were taken at the end of each run. Plotted are the mean values for the different diversity and turnover measures over the 20 runs (data from [Table 10.1](#)). Two-sample *t*-tests confirm with confidence 99% ($\alpha = 0.01$) that all means are higher in the disaster scenario.

During variation phases following disasters, the model population exhibits increased diversity. This is expected and the primarily question in this context is whether this diversity persists into the next balance phase. To verify that, diversity measurements taken during variation phases need to be discounted. As selection in the model is due to restricted carrying capacity of the landscape cells, selection is inhibited while the landscape is under-

populated. Let *crowdness* be the degree to which the landscape is populated relative to its maximum possible population. Then, *variation phases* in the model are defined as phases during which *crowdness* < 0.9.

$$\text{crowdness} = \frac{\text{current population}}{\text{maximum population}} = \frac{\text{current population}}{\text{number of grid cells} \times \text{cell carrying capacity}}$$

For both, the basis and the disaster scenarios, 20 control runs are performed. At intervals of 250 generations, the turnover of utilised niches and the four measures of diversity described above are measured. Measurements taking during variation phases (when *crowdness* < 0.9) are discounted. The results are shown in Table 10.1 and Figure 10.3. Characteristic simulation dynamics are shown in a plot of a typical experimental run (Figure 10.2).

10.1.4. Discussion

The results of the simulation experiments show that recurrent disturbances facilitate perpetual novelty and diversity in the model. Indeed, all four measures of diversity as well as both measures of niche turnover are consistently higher in the disaster scenario compared to the scenario without disasters. Two-sample *t*-tests confirm with confidence 99% ($\alpha = 0.01$) that in the disaster scenario, the means of all of the above statistics are significantly higher (Table 10.1 and Figure 10.3).

Theoretical arguments (section 9.2) and the emergence of several of nature-like dynamics within the model (section A.2) imply that these results are likely applicable to real landscape ecosystems [Grimm et al., 2005].

As discussed in section 9.2, co-evolution is a possible mechanism behind natural novelty. However, the present model does not include any species interactions, supporting the earlier argument that sustained novelty does not depend on co-evolution. Although such investigation is beyond the scope of this study, it would nevertheless be interesting to see this work extended to incorporate co-evolution in some form. For instance, this may be done by varying the *normal* niches over time through introducing new niches, removing existing niches or varying the niche conditions.

It is important to consider the specific means in which disturbances facilitate novelty and diversity in the model. Are these mechanisms the DPE phase changes suggested in section 9.2, or are there other forces at work? There are two ways in which disturbances facilitate continuous novelty in the model.

First, disturbances affect the connectivity of two important interaction networks within the landscape. One is a network of reproductive interactions between individuals in different cells that occur through migration. It facilitates gene flow and acts against adaptive radiation. By wiping out large areas, disasters separate the remaining populations in disconnected islands, preventing gene flow between sub-populations. This can lead to divergence of two population islands through genetic drift. The second is a network of unoccupied cells. An increased presence of such cells, caused by disturbances, disrupts gene flow and aids percolation by new species through providing invasion paths and areas with altered selection regimes. In a well-adapted population, random mutations are likely to be selected out. As selection is temporarily inhibited after a disaster (because the freed-up cells are not filled to their carrying capacity), mutations can lead to adaptations to new niches, even if the fitness within those niches it initially low.

The second mechanism through which disturbances facilitate continuous novelty in the model, is the altering of the environment that leads to increased *condition* of disaster-niches and poorer *condition* of normal niches in disturbed areas. The altered conditions result in directional selection towards disaster niches even after the disturbed areas are repopulated. By themselves, altered niche conditions are not expected to lead to increased adaptive radiation: even lower-fitness populations can persist once they occupy a significant habitat area for a long time (section A.2). However, when altered niche conditions occur in areas that have previously been freed-up by disasters, they can facilitate the establishment of new species adapted to the disaster environment. Such adaptations will be amplified by the high

condition of disaster-niches in disturbed areas, thus allowing new species to establish stable habitats against competition from readily established, more numerous populations.

In order to further investigate the relative importance of the two above mechanisms for facilitation of adaptive radiation through disturbances, I performed a short series of additional experiments.

In the first set of experiments I varied the radii of disaster impacts. In large, the results show that when the radii of the disasters are small, the dynamics of the disaster scenarios cannot be observed. Small disaster sizes are essentially equivalent to selecting the cells affected by disasters in an independently randomly distributed manner (Figure 10.4). Rather than freeing up large areas, such method leads to freeing up many small patches which will be immediately repopulated by individuals from remaining population sites, which will, on average, be close by. The resulting variation phase of inhibited selection will be very short and not sufficient for sub-populations to diverge. In nature, sites affected by disasters are not independently distributed. If a landscape patch was affected by a disaster, the nearby patches are more likely to be affected than those far away. Although further research of the relationship between particular disaster types (impact distribution, size, etc.) and their effects (diversity, niche utilisation, etc.) is beyond the scope of this work, such research may provide interesting insights.

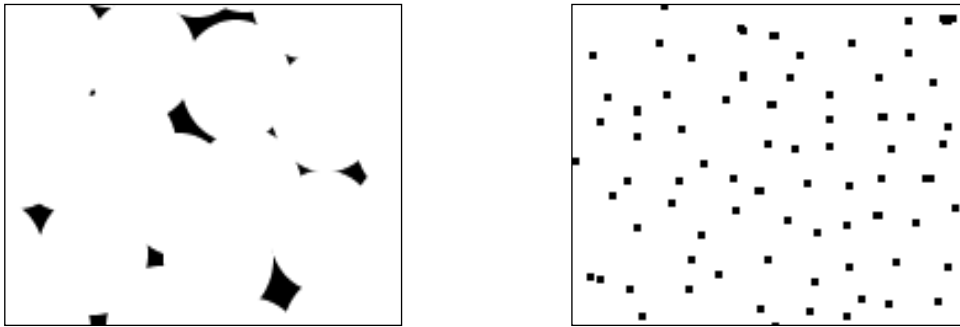


Figure 10.4. Schematic representation of different disaster sizes. Areas cleared by a disaster are white, remaining vegetation – black.

Left: Large radii of disaster impacts free up to vast continuous areas. Due to limited migration such areas take long time to be re-colonised and the altered selection regime lasts longer.

Right: Small radii of disaster impacts result in an even distribution of remaining unaffected patches across the landscape. Starting from these patches, extinction areas can be re-colonised quickly and the altered selection regime lasts for a shorter time.

In a further set of experiments I modified the disasters to affect the niche conditions without killing off existing individuals. As expected, the dynamics of this scenario do not differ significantly from the basis scenario that does not incorporate any disturbances at all (data not shown).

Finally, I performed a set of experiments where disasters free up the affected areas, but do not alter the environment (i.e. niche conditions remain constant). In large, new species occur after disturbances in the same way as in the original disaster scenario. However, they typically do not establish stable populations and are quickly displaced by previous species. After measurements performed when *crowdness* < 0.9 are discounted, the increase in niche turnover and diversity is not significant relative to the basis scenario.

In summary, the results show that both, recurrent DPE phase changes in landscape connectivity as well as alteration of the environment in disturbed areas are necessary for sustained perpetual novelty and diversity to emerge in the model. This conclusion corresponds to previous research on evolution of landscape composition. For instance, Green [1982; 1989] has arrived at similar conclusions when analysing the effects of forest fires on the basis of empirical data.

In the next section I examine the FL used in the model. The analysis is aimed to elicit a better context for the simulation results by achieving a better understanding of the FL.

10.2. Detailed analysis of the model fitness landscape

The material presented in this section is based on the following publications:

-  G. Paperin, D. G. Green, S. Sadedin, T. Leishman (2007):
 A Dual Phase Evolution Model of Adaptive Radiation in Landscapes.
 3rd Australian Conference on Artificial Life (ACAL'07). Ref: [Paperin, Green, et al., 2007b].
-  G. Paperin, D. G. Green, A. Dorin (2007):
 Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
 2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].
-  G. Paperin (2007):
 PhD candidature confirmation report.
 Report reviewed and accepted by an independent examination panel of the Faculty of IT, Monash University.

The simulation model presented in the previous section showed that recurrent disturbances facilitate perpetual novelty and diversity in landscape ecosystems through two main mechanisms: recurrent DPE phase transitions in landscape connectivity, and alteration of the environmental conditions in disturbed areas. Such alterations affect the FL and change local evolutionary dynamics. This result prompts for a further analysis of the FL used in the simulation. In this section I discuss the FL that results from the genotype model employed in the simulation and perform analytical and numerical analyses of evolutionary properties of the FL.

10.2.1. Choice of the genetic model

The genetic model employed here is based on a previous study of adaptive radiation in landscapes that uses a standard genetic model (section 10.2.2), however, there are other standard FL models frequently used in simulation studies of evolutionary dynamics. For instance, the N - k FL model is frequently employed due to its simplicity and theoretical relevance, as well as due to a convenient possibility to tune the ruggedness of the FL via the k parameter (section 6.2.3). However, the present simulation concentrates on creation of novelty by means of adaptive radiation. Hence, it requires a notion of species and thus of reproductive isolation.

Genetic models such as the N - k FL do not incorporate a notion of reproductive isolation necessary to investigate adaptive radiation. A common approach to defining reproductive isolation in such models is to assume that two genotypes that have sufficiently diverged from one another are incompatible (see [Gavrilets, 2004] for examples); i.e. reproductive isolation is a (possibly smoothed) threshold function of genetic divergence between individuals. This threshold can have great impact on the model's dynamics, but any threshold is necessarily arbitrary – a highly undesirable situation.

An alternative approach to determining the genetic compatibility of two model individuals is to simulate a number of hypothetical mating events and to determine the viability of offspring. However, as the number of individuals in simulations may be large, and the number of possible mating pairs even larger, this approach is often impractical. Moreover, many simulations, such as the model presented here, are not only interested in reproductive isolation between pairs of individuals, but also in separating the individuals in reproductively isolated groups (species). Determining such groups, their number and their spatial distribution may require further arbitrary parameters and can be computationally infeasible, even if pair-wise genetic compatibility is known. (Note – determining a degree of reproductive isolation in simulations incorporating a relatively small number of individuals may be feasible – see model in section 11.4.)

The above arguments imply that a genetic FL model suitable for simulations such as the one discussed in this chapter must at least fulfil the following requirements:

- It must be computationally feasible to determine whether any two individuals are reproductively isolated.
- It must be computationally feasible to cluster the entire population in reproductively isolated groups.
- Unsupported or arbitrary parameters (e.g. divergence thresholds) must be avoided where possible.

The study by Gavrillets and Vose [2005] has a focus that is related to the current study (adaptive radiation dynamics in landscapes) and utilises a standard genetic model that fulfils the above requirements. Thus, a modified, but equivalent (section 10.2.2) genetic configuration was chosen for the current model.

10.2.2. Equivalence between the employed genetic model and a standard niche-based additive quantitative trait model

The significance of the analysis of the FL employed in the current simulation goes beyond the study presented in this chapter. The analysis also applies to a standard genetic model employed in other studies. This correspondence is shown in this sub-section.

The genotype model used in the present simulation study is inspired though a previous study by Gavrillets and Vose [2005]. The FL in that study is based on a standard genetic model – the *additive quantitative trait model* [Gavrillets, 2004, chap. 2 and 8]. Unlike the approach taken here, the additive quantitative trait model has an explicit genotype-phenotype distinction. That genetic model has been used in a number of previous studies (e.g. see several listed by Gavrillets [2004]), in particular it can be applied to environmental requirements towards individual traits, such as niche requirements, as it is done here.

The genetic model employed in the simulation described in this chapter is equivalent, and computationally more efficient, to the additive quantitative trait model. The equivalence applies when number of traits in the current simulation model equals to *number of traits* $\times$ *number of loci per trait* in the standard model.

In the additive quantitative trait model *with niches*, the phenotype Ph_λ of an individual λ is represented by T real values between 0 and 1. Each value describes the degree to which a certain trait is expressed in the phenotype. The genotype G_λ of λ is represented by a sequence of L diallelic loci (say, 1 or 0). The genotype is split into non-overlapping groups of r loci per group ($L = r \times T$), the loci in each trait group control the expression value of one trait, which is determined by the proportion of 1s in that group.

Populations adapt to live in ecological niches. A niche M is represented by a binary string: $M = (m_1, m_2, \dots, m_T)$, where $m_i \in \{0, 1\}$, $i = 1 \dots T$. The binary values m_i are the trait requirements of the niche M : if $m_i = 1$, then trait i must be present in an individual in order to be fully adapted to niche M ; if $m_i = 0$ – absent. The niche-distance of an individual λ with respect to a niche M is the average distance between the trait values of λ and the trait requirements of M .

For example, let $r = 3$, $G_\lambda = (000\ 001\ 011\ 010\ 110\ 111\ 101\ 100)$, $M = (10011101)$:

Genotype G_λ :	0 0 0	0 0 1	0 1 1	0 1 0	1 1 0	1 1 1	1 0 1	1 0 0
Phenotype Ph_λ :	0.000	0.333	0.667	0.333	0.667	1.000	0.667	0.333
Niche M :	1	0	0	1	1	1	0	1
Niche-distance:	1	1/3	2/3	2/3	1/3	0	2/3	2/3

$$dist(Ph_\lambda, M) = \frac{1}{8} \times \left(1 + \frac{1}{3} + \frac{2}{3} + \frac{2}{3} + \frac{1}{3} + 0 + \frac{2}{3} + \frac{2}{3} \right) = \frac{1}{8} \times \frac{13}{3} = \frac{13}{24} \approx 0.54167$$

The fitness of λ is then a continuous, non-increasing function of $dist(Ph_\lambda, M)$.

Observe that the proportion of 1s in a trait-group of size r is equivalent to the hamming distance of such trait group to a sequence of r 0s normalised by r . For a niche M , generate a niche M' by replacing each 0 with a sequence of r 0s and each 1 with a sequence of r 1s. Now, instead of calculating the distance between Ph_λ and M , calculate the distance between G_λ and M' and normalise by the number of components.

In the above example, $M' = (111\ 000\ 000\ 111\ 111\ 111\ 000\ 111)$ and:

Genotype G_λ :	0	0	0	0	0	1	0	1	1	0	1	1	0	1	1	1	1	0	1	1	0	0
Niche M' :	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	1
Niche-distance:	1	1	1	1	0	0	1	0	1	1	1	0	1	0	0	1	0	0	0	1	0	1

$$dist(Ph_{\lambda}, M') = \frac{1+1+1+0+0+1+0+1+1+1+0+1+0+0+1+0+0+0+1+0+1+0+1+1}{8 \times 3}$$

and

$$dist(Ph_{\lambda}, M') = \frac{13}{8 \times 3} = \frac{13}{24}$$

as before. Observe that the ordering of the traits does not play a role in the model. Flipping any one bit in the genotype will either increase or decrease the niche distance by $1/24$ (generally, by $1/L$). Clearly, this is also the case under any permutation of the trait order. Therefore, it is not necessary to group the trait requirements of M' into consecutive sequences of r 1s or 0s. The fitness of λ in M' can be calculated directly as a function of its genotype G_λ : $fitness_{SM}(\lambda)$ is a continuous, non-increasing function of $dist(G_\lambda, M')$. Gavrillets [2004, chap. 2 and 8] uses a Gaussian function, but other choices are possible. Note that when L is constant, the niche-distance function may not require normalisation, if a suitable fitness function is used.

As a result, the additive quantitative trait model with niches described above is equivalent to a modified model that does not explicitly incorporate a phenotype. Instead, a larger number of discrete-valued traits, which are directly determined from the genotype, can be used. In the above notation this means:

$$T_{NewModel} = L_{NewModel} = L_{OldModel} = r_{OldModel} \times T_{OldModel}$$

The above modification simplifies the original genetic model and makes both, an analytical treatment and a computer implementation much easier. In addition, a computer implementation of the modified model can largely be realised using bit-arithmetic, which executes greatly faster than the floating point calculations required for the original approach.

10.2.3. Analytical examination

Let $\mathbb{G}$ be the set of all bit-sequences of length L and assume that there are S distinct *goal sequences* $M_1, M_2, \dots, M_S$ ($M_i \in \mathbb{G}$, $i = 1 \dots S$). Denote the set of all goal sequences as $\mathbb{M}$, i.e. $\mathbb{M} = \{ M_1, M_2, \dots, M_S \} \subseteq \mathbb{G}$. Let $dist(G, M_i)$ be the hamming distance between a sequence $G \in \mathbb{G}$ and a goal sequence M_i .

A closest goal M_G^* of a bit-sequence G is any goal M_i such that:

$$M_G^* = \arg \min_{M_i} dist(G, M_i)$$

Note that a sequence G may have multiple closest goals.

The *closest-goal distance* $d_{\min}(G)$ of a bit-sequence G is the hamming distance between G and any closest-goal M_G^* of G :

$$d_{\min}(G) = \text{dist}(G, M_G^*) = \min_{i=1 \dots S} (\text{dist}(G, M_i))$$

The process of evolutionary adaptation towards habitat niches in the current simulation model can be formulated as an optimisation problem as follows:

Starting with two initial bit-sequences randomly and independently drawn from $\mathbb{G}$, use an evolutionary algorithm (i.e. the model lifecycle) to evolve a population of bit-sequences $G \in \mathbb{G}$ while seeking to minimise the closest-goal distances towards a fixed set $\mathbb{M}$ of 60 goal sequences randomly and independently drawn from $\mathbb{G}$.

In this formulation, a bit-sequence corresponds to a genotype and represents a point in the genotype space. Each of the S goal sequences corresponds to a niche and represents one global optimum on the fitness landscape.

For the purpose of the following discussion, assume that that L is not small. Say, $L > 10$.

Lemma 10.1:

Let $S = 1$. Then, the number c_d of possible bit-sequences G with the goal distance $d = \text{dist}(G, M_1)$ is:

$$c_d = \binom{L}{d} = \frac{L!}{d! \cdot (L-d)!}$$

Indeed:

Consider $d = 0$. Clearly, there is exactly one sequence such that $d = 0$: M_1 itself.

Consider $d = 1$. There are L distinct sequences such that $d = 1$. Each can be obtained by flipping any one bit of the goal sequence.

Consider $1 < d \leq L$. For any such d , a corresponding bit-sequence can be obtained from the goal sequence by flipping any d bits, but no more than L bits can be flipped. There are $\binom{L}{d}$ distinct ways to choose the bits to flip, q.e.d.

Lemma 10.2:

Let $S = 2$. Then, the largest possible closest-goal distance of any bit-sequence is $\max_2(d_{\min}) = L - 1$.

Indeed:

There is exactly one bit-sequence G_1 that has a hamming distance $\text{dist}(G_1, M_1) = L$ towards M_1 , and only one bit-sequence G_2 that has a hamming distance $\text{dist}(G_2, M_2) = L$ towards M_2 (Lemma 10.1). Since M_1 and M_2 are distinct, G_1 and G_2 are also distinct. But as there is only one bit-sequence at distance L to M_1 (namely G_1), G_2 must be closer to M_1 than G_1 : $\text{dist}(G_2, M_1) \leq L - 1$. By an analogous argument, $\text{dist}(G_1, M_2) \leq L - 1$.

Indeed, there may be a bit-sequence with the hamming distance $d_{\min} = L - 1$: assume M_1 and M_2 differ only at one bit. Consider a bit-sequence G' obtained from M_1 by reversing all bits. For such sequence, $\text{dist}(G', M_1) = L$ and $\text{dist}(G', M_2) = L - 1$, so the distance of G' to its closest goal is $d_{\min}(G') = L - 1$, q.e.d.

Lemma 10.3:

The previous results are now generalised. Let $S \in \mathbb{N}$ be the number of distinct goal sequences such that $1 < S \leq 2^L$. Let $\max_s(d_{\min})$ be the largest possible closest-goal distance. Then, $\max_s(d_{\min}) = L - k$ for a k such that

$$\binom{L}{k-1} + 1 < S \leq \binom{L}{k} + 1$$

Indeed, this can be seen as a generalisation of [Lemma 10.2](#):

For $S = 3$, $\max_3(d_{\min}) = L - 1$ by the same argument as in [Lemma 10.2](#). In fact, that argument applies for all $1 < S \leq L + 1$, because there are up to $L - 1$ goal sequences that have a hamming distance 1 to some arbitrary goal sequence M_1 .

For $S = L + 2$, the largest possible closest goal distance is $L - 2$. This follows from the same argument as above, only in this case there are goal sequences that have a hamming distance 2 towards each other. This applies for all cases where the distance between any two goal sequences is not larger than 2. By [Lemma 10.1](#), this is the case when:

$$L + 1 < S \leq \frac{L \cdot (L-1)}{2} + 1$$

or

$$\binom{L}{1} + 1 < S \leq \binom{L}{2} + 1$$

Exactly the same argument can be extended to any number S of goal sequences. When there are at least two goal sequences that have a distance k to each other, the largest possible closest goal distance is $L - k$. By setting $d = k$ in [Lemma 10.1](#) it is clear that it applies exactly when

$$\binom{L}{k-1} + 1 < S \leq \binom{L}{k} + 1$$

, q.e.d.

The above analysis determines the *maximum possible* closest-goal distances. The closest-goal distances are at maximum, when the goal sequences differ by the least possible amount of bits. The maximum possible closest-goal distance corresponds to the smallest possible genotype fitness and hence determines a limit for the valley depth and the slope steepness of the FL. The primary point of interest is, however, the *distribution* of genotype fitness when the goal sequences are randomly and independently selected from $\mathbb{G}$. This corresponds to distribution of closest-goal distances $d_{\min}(G)$ when the sequence space $\mathbb{G}$ contains S randomly and independently distributed goal sequences.

Lemma 10.4:

Let $S = 1$ and recall that c_d is the number of bit-sequences $G \in \mathbb{G}$ such that $\text{dist}(G, M_1) = d$. Then, c_d is binomially distributed with the probability parameter $p = 0.5$, i.e. $c_d \sim \text{Bin}(L, 0.5)$.

Indeed. There are 2^L bit-sequences of length L , and by [Lemma 10.1](#), there are $c_d = \binom{L}{d}$

sequences at distance d . Hence, the frequency f of bit-sequences at distance d to M_1 is $\frac{c_d}{2^L}$:

$$f = \frac{c_d}{2^L} = \frac{\binom{L}{d}}{2^L} = \binom{L}{d} \cdot \left(\frac{1}{2}\right)^L = \binom{L}{d} \cdot \left(\frac{1}{2}\right)^d \cdot \left(\frac{1}{2}\right)^{L-d} = \binom{L}{d} \cdot \left(\frac{1}{2}\right)^d \cdot \left(1 - \frac{1}{2}\right)^{L-d} \sim \text{Bin}\left(L, \frac{1}{2}\right)$$

, q.e.d.

The mean of this distribution is $\frac{L}{2}$ and the variance is $\frac{L}{4}$. When both, L and the mean are large, this distribution can be approximated by a Gaussian distribution: $f \sim \text{Norm}(\frac{L}{2}, \frac{L}{4})$. Then, ca. 68% of goal distances lie within one standard deviation of the mean and ca. 95% of goal distances lie within 2 standard deviations of the mean.

In other words, the distance to the goal of about 68% of sequences lies in the interval $\left[\frac{1}{2}(L-\sqrt{L}), \frac{1}{2}(L+\sqrt{L})\right]$, and of about 95% – in the interval $\left[\frac{1}{2}(L-2\sqrt{L}), \frac{1}{2}(L+2\sqrt{L})\right]$.

For $L = 15$ it means respectively [6, 9] and [4, 11]; for $L = 20$, [8, 12] and [6, 14]; for $L = 100$, [45, 55] and [40, 60].

This implies that for a single habitat niche, the fitness of most genotypes is very similar and average. There are a few low fitness genotypes and a few high fitness genotypes with the global optimum being reachable from any point via a sequence of advantageous atomic mutations. As fitness in the model is a linear function of the closest-goal distance, the FL landscape around the majority of genotypes is nearly flat and selection pressures are thus weak.

For $S > 1$ there are fewer genotypes of very low fitness (maximum closest-goal distance) ([Lemma 10.3](#)). For an independent distribution of goal sequences, the FL structure is expected to be even flatter. An analytical treatment of this general case is complex due to a large number of conditional probabilities involved. Instead, I perform a numerical analysis through a series of Monte Carlo simulations in the next section.

10.2.4. Numerical examination

In the previous section I provided an analytical examination of the FL used in the simulation study. The analysis also applies to a more general additive quantitative trait model with niches ([section 10.2.2](#)). Here I continue the analysis with numerical techniques. A range of statistical analysis techniques that can be used for FLs is discussed by Stadler [2002]. Here, use some of such techniques, as well as some methods specifically designed for the current investigation. The main objective is to obtain qualitative results about the FL structure that affect the interpretation of the experimental results.

10.2.4.1. Fitness distribution

The fitness distribution corresponds to the *distribution of closest-goal distances*. To determine this distribution, a complete genotype space $\mathbb{G}$ is created for $L = 10, 15$ and 20 loci. For each genotype space $S = 2, 10, 50, 100$ and 500 random goal sequences (niches) are randomly and independently selected from $\mathbb{G}$ and the resulting frequency distribution $\mathcal{D}(d_{\min})$ of the closest-goal distance $d_{\min}(G)$ for all $G \in \mathbb{G}$ is generated. 20 runs are undertaken for each pair (L, S) with new randomly created goal sequences for each run, the results are averaged. Over the 20 runs, the mean $\mu(d_{\min})$, the variance $\sigma^2(d_{\min})$ and the standard deviation $\sigma(d_{\min})$ are computed. The results are shown in [Table 10.2](#), [Table 10.3](#), [Table 10.4](#) and in [Figure 10.5](#).

The measurements show that the distribution $\mathcal{D}(d_{\min})$ is bell-shaped and slightly skewed towards smaller values. As the number of goal sequences increases, largest closest-goal distance shrinks slowly, however, still faster than the largest *possible* closest-goal distance determined in [section 10.2.3](#).

The fitness is a linear function of the closest-goal distance, thus $\mathcal{D}(d_{\min})$ allows to draw conclusions about the fitness distribution on the simulation FL and the resulting evolutionary dynamics.

10. DPE and Perpetual Novelty and Sustained Diversity in Landscape Ecosystems

In cases where the number of niches is not very small (say, say $S \geq 10$), the poorest fitness is fairly high, which means that fitness hills are gently sloped, and hence, selection is weak. This conclusion confirms the assumption made at the end of [section 10.2.3](#). The majority of genotypes have similar fitness. This is an indication for the existence of neutral networks within the FL, although with very shallow canyons around the network ridges. Also, the small variance of the fitness distribution means that a recombination of two randomly selected genotypes is likely to have a fitness value that is close to its parents fitness.

$d_{\min}$	Average number of genotypes				
	$S = 2$	$S = 10$	$S = 50$	$S=100$	$S=500$
0	2	10	50	100	500
1	20	95	384	597	523
2	88	338	532	324	1
3	214	430	58	3	0
4	305	143	0	0	
5	252	8			
6	115	0			
7	26				
8	2				
9	0				
$\mu(d_{\min})$	4.11	2.61	1.58	1.22	0.51
$\sigma^2(d_{\min})$	1.69	0.81	0.45	0.37	0.25
$\sigma(d_{\min})$	1.30	0.90	0.67	0.61	0.50

Table 10.2. Frequency distribution of the closest-goal distance $d_{\min}$ on the model fitness landscape for $L = 10$ loci, averaged over 20 runs.

$d_{\min}$	Average number of genotypes				
	$S = 2$	$S = 10$	$S = 50$	$S=100$	$S=500$
0	2	10	50	100	500
1	30	150	741	1464	6650
2	209	1029	4744	8549	20546
3	899	4126	13719	17070	5068
4	2623	9632	11946	5482	4
5	5379	11409	1557	74	0
6	7740	5519	10	0	
7	7699	859	0		
8	5171	33			
9	2270	0			
10	627				
11	107				
12	11				
13	1				
14	0				
$\mu(d_{\min})$	6.42	4.56	3.27	2.81	1.92
$\sigma^2(d_{\min})$	2.55	1.25	0.74	0.61	0.41
$\sigma(d_{\min})$	1.60	1.12	0.86	0.78	0.64

Table 10.3. Frequency distribution of the closest-goal distance $d_{\min}$ on the model fitness landscape for $L = 15$ loci, averaged over 20 runs.

$d_{\min}$	Average number of genotypes				
	$S = 2$	$S = 10$	$S = 50$	$S=100$	$S=500$
0	2	10	50	100	500
2	40	200	999	1998	9949
3	380	1897	9447	18800	89929
4	2278	11317	54940	106037	397708
5	9658	46870	203227	342670	469104
6	30639	137339	410223	450539	54354
7	74739	271224	314757	125744	32
8	141081	322906	53936	2686	0
9	204699	198448	986	1	
10	225234	52997	1	0	
11	184756	5203	0		
12	110686	164			
13	47241	2			
14	13959	0			
15	2781				
16	369				
17	32				
18	2				
19	0				
$\mu(d_{\min})$	8.73	6.60	5.08	4.53	3.47
$\sigma^2(d_{\min})$	3.38	1.68	1.01	0.84	0.59
$\sigma(d_{\min})$	1.84	1.30	1.00	0.92	0.77

Table 10.4. Frequency distribution of the closest-goal distance $d_{\min}$ on the model fitness landscape for $L = 20$ loci, averaged over 20 runs.

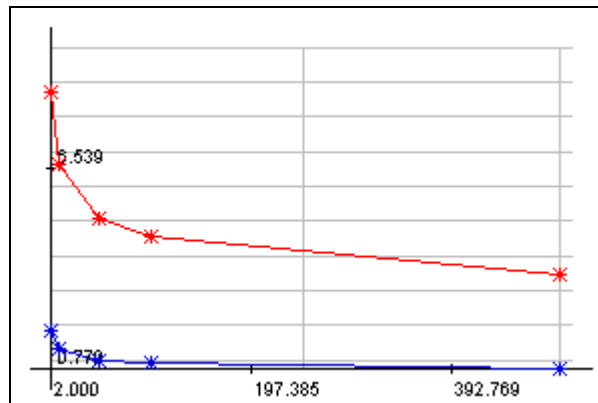


Figure 10.5. Frequency distribution of the closest-goal distance. The mean $\mu(d_{\min})$ (top red line) and the standard deviation $\sigma(d_{\min})$ (bottom blue line) of the closest goal distances for $L = 20$ loci and $S = 2, 10, 50, 100$ and 500 goal sequences. Other values of L yield a qualitatively similar picture. X-axis: number of goal sequences. Y-axis: mean and standard deviation. (The rounding artefacts on axes labels are due to an issue in an early version of LiveGraph that was used to render this figure (chapter 3).)

10.2.4.2. Genotype reach cluster size

In this and the following sub-sections I continue the analysis of the FL structure. All of the described experiments were initially performed for the same pairs (L, S) as in the previous section. The qualitative results are consistent across all pairs. The particular values of L and S are model specific and the main objective is to obtain a qualitative description of the FL. Thus, I illustrate the following discussion with data for a configuration of small values ($S = 10, L = 15$) only. For all measurements, the variation of values over different runs is small, implying consistent results.

Define the reach cluster $reach(G) \subseteq \mathbb{G}$ for a genotype $G \in \mathbb{G}$ as the set of all genotypes that can be reached from G by a number of successive atomic mutations without lowering the fitness. I independently choose 5000 random distinct genotypes and determine the size $|reach(G)|$ of the corresponding reach cluster. The upper bound on the reach cluster size of any $G \in \mathbb{G}$ is $|\mathbb{G}| - 1 = 2^L - 1 = 2^{15} - 1 = 32767$. Results are shown in Table 10.5.

The spike in genotype frequency for large cluster sizes indicates that there is a large number of genotypes for which the reach cluster includes almost all genotypes. As $reach(G)$ contains only genotypes that are no less fit than G , these must be low-fitness genotypes. This implies that most low-fitness genotypes can evolve into almost any genotype by a series of advantageous mutations.

The spiky nature of the distribution in the face of the fact that most genotypes are of average fitness (section 10.2.4.1), implies that once a low-fitness genotype improved fitness along some evolutionary path, many alternative evolutionary paths become unavailable.

Cluster size		Number of genotypes	
from	to		
0	3	24	(0.48%)
4	10	642	(12.84%)
11	50	856	(17.12%)
51	200	182	(3.64%)
201	2000	140	(2.80%)
2001	5000	0	(0.00%)
5001	10000	0	(0.00%)
10001	15000	875	(17.50%)
15001	20000	0	(0.00%)
20001	25000	0	(0.00%)
25001	30000	1573	(31.46%)
30001	32768	708	(14.16%)

Table 10.5. Frequency distribution of 5000 random genotypes according to the corresponding reach cluster sizes.

10.2.4.3. Genotype distances along adaptive paths

The next experiment elicits information about fitness canyons and the ruggedness of the FL by measuring the shortest adaptive paths that do not include deleterious mutations between pairs of genotypes.

The average number of mutations separating two random genotypes is $L / 2 = 15 / 2 = 7.5$. I compare this with the number of *non-deleterious* mutations separating two random genotypes. A mean of this number that is significantly larger than $L / 2$, would indicate that detours must be taken to avoid valleys and thus a high degree of FL ruggedness.

Table 10.6 shows the lengths of the shortest adaptive walk with non-deleterious mutations for 5000 randomly selected genotype pairs. The mean of the path lengths is larger than $L / 2$, but variance is also large, implying that detours are being taken consistently, but they are not long. Over 50% of the random genotype pairs do not have a non-deleterious mutational connection at all. In combination with the previous result, this means that

although a low-fitness ancestor may evolve to use almost any niche, two diverged sub-populations will continue to diverge. This is an indication for decreased gene flow between diverged high fitness populations.

Path length	Number of pairs	Path length	Number of pairs
0	0 (0.00%)	9	397 (7.89%)
1	1 (0.02%)	10	320 (6.36%)
2	4 (0.08%)	11	152 (3.02%)
3	49 (0.97%)	12	68 (1.35%)
4	64 (1.27%)	13	32 (0.64%)
5	153 (3.04%)	14	16 (0.32%)
6	284 (5.65%)	15	4 (0.08%)
7	412 (8.19%)	17	1 (0.02%)
8	499 (9.92%)	18	0 (0.00%)
Mean length:	8.01	No path:	2574 (51.17%)
Std. Dev.:	2.07		

Table 10.6. Frequency distribution of 5000 random genotype pairs according to the length of the shortest adaptive walk between the genotypes that contains non-deleterious mutations only.

10.2.4.4. Number of advantageous mutations

Next, consider the number of advantageous atomic mutations for each genotype. For each $G \in \mathbb{G}$, generate all $L = 15$ possible atomic mutations and determine $N^+(G)$ – the number of such mutations that have a fitness strictly higher than G . A typical result is presented in Table 10.7. The distribution of genotypes according to N^+ is strongly skewed towards smaller values and the small number of better-fitness neighbours typically does not exceed $L / 3$.

The genotypes G with $N^+(G) = 0$ correspond to the $S = 10$ goal sequences representing global fitness maxima. It is easy to see that the FL has no local maxima that are not also global:

Assume $M_G^* = (m_1, \dots, m_i, \dots, m_L)$ is a closest goal for a genotype $G = (g_1, \dots, g_i, \dots, g_L)$ such that $d_{\min}(G) = \text{dist}(G, M_G^*) > 0$. Then, there is at least one bit g_q in G that is different from the corresponding bit m_q in M_G^* . By flipping g_q in G , a new genotype G' is obtained such that $\text{dist}(G', M_G^*) < \text{dist}(G, M_G^*)$. Thus G is not a local fitness optimum, q.e.d.

The presence of non-global maxima is typical and important feature of biological FL (section 6.2) and the absence of such maxima in the simulation FL indicates that some of evolutionary dynamics of the model may not fully reflect circumstance found in nature.

$N^+(G)$	Number of genotypes	$N^+(G)$	Number of genotypes
0	10 (0.03%)	8	1762 (5.38%)
1	150 (0.46%)	9	2209 (6.74%)
2	1008 (3.08%)	10	1783 (5.44%)
3	3758 (11.47%)	11	942 (2.88%)
4	7702 (23.51%)	12	805 (2.46%)
5	7282 (22.22%)	13	635 (1.94%)
6	2764 (8.44%)	14	422 (1.29%)
7	1156 (3.53%)	15	380 (1.16%)

Table 10.7. Frequency distribution of genotypes according to the number of better fitness neighbours.

10.2.4.5. Number of closest niches

It is possible that a bit-sequence $G \in \mathbb{G}$ has multiple closest-goal sequences. Formally, if $M_G^* \in \mathbb{M}$ is a closest goal of G , then there may be another goal sequence $M_G^\# \in \mathbb{M}$ such that $M_G^* \neq M_G^\#$ and $\text{dist}(G, M_G^*) = \text{dist}(G, M_G^\#)$.

For all genotypes G , I measure the number m_G of equidistant closest-goal sequences. Results (Table 10.8) show that ca. two thirds of genotypes have a unique closest-goal sequence, however, the remaining genotypes have two or more equidistant goals.

The implication for the simulation model is that care must be taken to always assign a genotype that has multiple closest goals to the same niche. For the calculation of fitness, the choice of any of the closest-goal sequences is irrelevant, however, this choice is important for an determining of the number of species (utilised niches), species diversity and turnover, and habitats occupied by niche populations. The approach taken uses an arbitrary ordering of goal sequences that is fixed throughout the simulation. For genotypes with multiple closest goals, the goal with the smallest index in that ordering is always chosen. The resulting preference may affect the model dynamics in the sense that populations of niches with smaller indices have a selective advantage via niche choice. This, however, does not affect any of the research results.

m_G	Number of genotypes
1	20670 (63.08%)
2	7457 (22.76%)
3	3018 (9.21%)
4	1112 (3.39%)
5	393 (1.20%)
6	113 (0.34%)
7	5 (0.02%)

Table 10.8. Frequency distribution of genotypes according to m_G – the number of equidistant closest-goal sequences.

10.2.4.6. Hybrid fitness

In a final series of tests I investigate the offspring of random pairs of individuals from different groups. In particular, I verify whether the hybrid fitness (the fitness of progeny of individuals from different habitat niches) is lower than the parents fitness. These tests will determine whether the species concept in the model is justified and whether populations adapted to different niches can be considered different species.

Every measurement is averaged over 100000 independently selected random pairs of parents, for which offspring are created using free recombination (as in the simulation model) and no mutation.

When completely random parents are considered, the mean offspring fitness is similar to the mean fitness of the population (as any child can be expressed as a mating outcome). 11% of random individual pairs belong to the same niche, 89% of pairs belong to different niches. I differentiate these cases and perform measurements for 100000 random pairs from the same niche, and for 100000 random pairs from different habitat niches.

Offspring from parents existing in the same habitat niche:

65% of offspring from parents that exist in the same habitat niche have the same fitness as their parents (or a fitness that is between the fitness values of parents).

72% of offspring exist in the same niche as their parents, 28% of offspring exist in a different niche. Out of those who remain in the parents' niche, 27% have a higher fitness, and 7% have a lower fitness, which means that fitness can be increased through

recombination. From offspring that exist in a different closest niche from their parents, 21% have higher fitness (in the new niche), and 14% have lower fitness.

The informative value of these measurements is limited because parent pairs of any fitness are selected, and thus no specific niche adaptation is expected. Thus, I repeat the experiment while selecting only parents with *closest-niche distance* ≤ 4 (i.e. a fitness higher than population average, [section 10.2.4.1](#)).

For such parents, a larger proportion (82%) of offspring belong to the same niche as their parents. This was expected for mating between better adapted individuals. For the offspring that exist in a different niche from their parents, the proportion of progeny with fitness lower than parents grows (34%), and the proportion of higher-fitness progeny (15%) falls.

Offspring from parents existing in different habitat niches (hybrids):

In total, 55% of hybrid progeny from any-fitness different-niches parents exist in the same niche as one of the parents, 45% exist in a niche that is different to the habitat niche of either parent.

66% of hybrids have a fitness value that is either the same or between the fitness values of their parents. 16% have a higher fitness compared to their parents, 18% have a lower fitness. This shows that hybrids are *not* more likely to have a lower fitness than a random individual. In fact hybrid fitness distribution is similar to the fitness distribution of the entire population ([section 10.2.4.1](#)). However, since measurements are performed on parents of any fitness, offspring properties were expected to be similar to the general population.

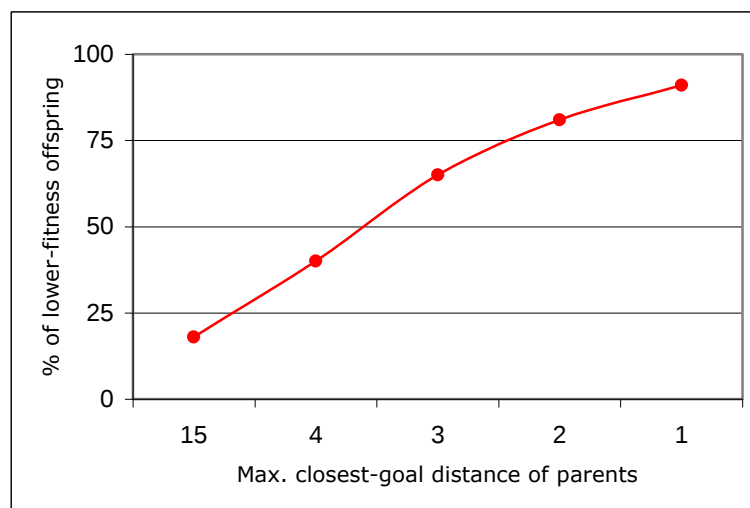


Figure 10.6. Proportion of hybrids with lower fitness than both parents. For 100000 random pairs of parents from different habitat niches, the proportion of offspring that have a larger closest-goal distance than either parent is calculated ($L = 15$, $S = 10$). Parents are chosen randomly and independently from different niches such that the largest closest-goal distance of either parent is not larger than 15, 4, 3, 2 or 1 bits.

As previously, the measurements are repeated while selecting only parents with *closest-niche distance* ≤ 4 . For such pairs, only 47% of offspring have the same fitness as the parents, 13% have a higher fitness, and 40% have a lower fitness. This is higher than the proportion of lower-fitness offspring from same-niche parents, indicating that gene flow between adapted niche populations is reduced, but is still significant.

Compared to any-fitness-parents, a higher proportion (62%) of hybrid offspring belong to one of the parents' niches. From these hybrids, 37% have a lower fitness than the parents, and 14% have higher fitness.

For hybrids that do not belong to either of the parents' niches, 46% have lower fitness, and 10% have higher fitness than the parents. This shows that hybridisation leading to a new habitat niche is associated with a drop in fitness.

The above measurements reveal that even for above-average pairs of parents from different niches, only 40% of offspring are less fit than the parents. I repeat the measurements for hybrids of parents with even higher fitness. For pairs with *closest-niche distance* ≤ 3 , 65% of hybrids have lower fitness than both parents. For *closest-niche distance* ≤ 2 , 81% of hybrids are less fit; and for *closest-niche distance* ≤ 1 – 91%. The results are shown in [Figure 10.6](#).

The measurements indicate that for highly adapted populations one *can* speak of reduced gene flow between niches. However, only for very-high-fitness populations the gene flow is reduced to a minimum.

10.2.5. Conclusion

The above discussion reveals several evolutionary properties of the adaptive FL used in the simulation. The particular genotype model was chosen for the experiments because it was used in previous simulation studies of adaptive radiation in landscapes ([section 10.2.1](#)), and because it is equivalent to a standard additive trait model ([section 10.2.2](#)).

When fitness is a linear function of the closest-niche distance of a genotype and there is only one habitat niche available, the distribution of genotype fitness is binomial ([section 10.2.3](#)). As the number of niches increases, the distribution remains approximately bell-shaped, but becomes increasingly skewed towards higher fitness values ([section 10.2.4.1](#)).

Low-fitness genotypes can potentially evolve towards any niche, but once adaptation steps towards some niche have been made, the number of potential adaptive trajectories fall quickly in a non-smooth fashion ([section 10.2.4.2](#)). The majority of genotype pairs do not have a connecting adaptive path ([section 10.2.4.3](#)). These properties reflect the general characteristics found in biological FLs ([section 6.2](#)).

However, contrary to biological FLs, the FL employed in the simulation has no local optima that are not also global. The number of advantageous mutations resulting in viable genotypes is a significant proportion of the number of loci ([section 10.2.4.4](#)). When adaptive paths between genotypes exist, their shortest lengths tend to be similar to direct mutational paths ([section 10.2.4.3](#)). Thus, the FL has a low degree of ruggedness (i.e. there are few sudden fitness jumps), and fitness canyons are shallow if present at all ([sections 10.2.4.1 and 10.2.4.3](#)). All possible genotypes are viable.

A significant proportion of genotypes have multiple close niches. Such genotypes can be interpreted as generalist or undifferentiated. Hybrids between genotypes adapted to different niches have a high probability to be not less fit than their parents, unless the parents are very high fitness individuals. Such hybrids can facilitate significant gene flow between different niche populations. This indicates that the assumption that niche populations within the simulation exhibit postzygotic isolation and are thus equivalent to species ([section A.1.5](#)) does not hold. However, this realisation does not affect the conclusions based on the simulation results about the effects of disturbances and changes in landscape connectivity on diversity and continuous discovery of habitat niches. Simply, the term *species* is not appropriate. The above FL analysis reveals that simulation results allow drawing conclusions about continuous discovery and turnover of new *habitat niches*, not new *species*; as well as about lasting diversity of *adaptations to different niches*, not *different species*.

The dynamics observed in the simulation did not depend on reproductive isolation of niche populations. It is thus likely that if the utilised genotype model was modified to increase reproductive isolation of individuals from different niches and to minimise gene flow between niche populations, same general results were observed.

In the next chapter I propose a new genotype model based on the additive trait model utilised here and the HFL model ([section 6.2.4](#)). I analyse the proposed technique and show that it does not exhibit many of the drawbacks of the genotype model used in this chapter.

Chapter 11:

Evolution on Holey Fitness Landscapes

Simulation models of evolution often incorporate disruptive (diversifying) selection while ignoring the viability and genomic compatibility issues. Other methods, such as the HFL, provide a biologically realistic model of genetic (in-)compatibilities, but do not incorporate ecological selection and are not explicitly spatial. Here, I combine existing techniques and introduce a genetic model that incorporates both, an explicit, biologically realistic treatment of genetic (in-)viability issues, and an inherent approach for relating genotype fitness to the ecological environment. The proposed model is can be used for within simulations studies incorporating HFL-genetics. A numerical analysis of the genetic FL exhibited by the model informs future studies and shows that the FL exhibits many biologically realistic properties.

Based on the new techniques I examine how the HFL changes predictions for the maintenance of genetic diversity in the face of migration. Previous models suggest that ecologically-based reproductive isolation will rapidly collapse under migration. However, my simulation results indicate that an underlying HFL can often maintain diversity and may facilitate exploration of new niches and speciation. The techniques employed may also prove useful in biologically-inspired engineering, for instance for combating premature convergence in evolutionary optimisation algorithms. More generally, the results imply that HFL genetics should not be neglected in studies of adaptation and diversity.

The material presented in this chapter is based on the following publications:

-  G. Paperin, S. Sadedin, D. Green, A. Dorin (2008):
Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity.
11th International Conference on Artificial Life (ALife XI). Ref: [Paperin, Sadedin, et al., 2008].
-  G. Paperin (2008):
Using Holey Fitness Landscapes to Counteract Premature Convergence in Evolutionary Algorithms.
Satellite Workshop at the Genetic and Evolutionary Computation Conference 2008 (GECCO'08). Ref: [Paperin, 2008b].
-  G. Paperin, D. G. Green, A. Dorin (2007):
Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].

In the previous section I used a simulation model to investigate the role of DPE connectivity changes caused by environmental disturbances in facilitation of evolutionary diversity in landscapes. The simulation employed a genotype model known as *additive qualitative trait model*. The model has been used in a range of previous studies (sections 10.2.1 and 10.2.2) due to a number of advantages. In particular, it offers a convenient way of relating genotype fitness to the environment, for instance by providing a natural approach to defining adaptedness to habitat niches. This can be applied in a variety of individual-based simulation experiments. Environmental conditions such as niches can be varied geographically and over time allowing to investigate a wide range of evolutionary scenarios.

However, the fitness landscape (FL) that emerges from the genotype model used in the last chapter also suffers from a number of drawbacks (section 10.2): It has a low degree of ruggedness and there are no local optima. Fitness valleys are shallow and all possible genotypes are viable. Recombinations of genotypes adapted to different habitat niches are likely to have an average fitness. These features are not found in biological FLs (section 6.2). Models based on FL of this kind tend to ignore viability and genomic compatibility issues or make strong simplifying assumptions about such incompatibilities (e.g. [Orr, 1995]) [Paperin, Sadedin, et al., 2008].

In section 6.2.4 I described the concept of *holey fitness landscapes* (HFLs). HFLs are thought to underlie the evolution of lasting, effective barriers to gene flow that appear during adaptive radiation and they do not exhibit the above drawbacks. However they do not

provide some of the advantages of the additive trait model. There are a number of analytic models of adaptive radiation based on HFLs (e.g. see [Gavrilets, 2004, part 1]), however they do not incorporate ecological selection and are not explicitly spatial. HFL is an analytical concept that is difficult to apply in individual based simulations (section 11.2.2). There is no implicit notion of fitness (except a binary viable/inviable distinction), and no natural approach for relating genotypes to phenotypes or to adaptedness to particular environmental conditions.

In this chapter introduce a new genotype model that combines the advantages of HFL and the additive trait model, and avoids many of their drawbacks. After formally defining the new model (section 11.1), I identify the difficulties that arise when using the model in individual-based simulations and discuss how these difficulties can be overcome (section 11.2).

Using an implementation based on that discussion I perform a detailed analysis of the FL that results from the proposed genotype model (section 11.3). The analysis considers both, adaptation and speciation properties of the FL. While concentrating on the combined model introduced in this chapter, the analysis also allows to draw conclusions about some topological properties of HFLs. In large, the analysis shows that the FL of the proposed genotype model successfully combines several desired properties of the qualitative trait and the HFL models while avoiding many drawback. Future studies based on the proposed genetic model are informed by this analysis.

In order to further investigate the role of HFLs in evolutionary dynamics, I introduce an individual-based, spatially explicit simulation model that investigates the ability of HFL-genetics to facilitate and maintain reproductive isolation between populations (section 11.4). In the model, populations of individuals with HFL-genotypes evolve in isolation for a large period of time and then come into contact. The results show that despite a significant amount of migration between the reconnected populations and random mating, the populations can remain reproductively isolated due to genetic incompatibilities resulting from the HFL. This has important implications for mechanisms for evolution of sexual selection and speciation.

The techniques used in the HFL-simulation may be applied in other, non-biological computation studies. For instance, I propose a mechanism for using HFL structures to tackle the problem of premature convergence in genetic and other evolutionary optimisation algorithms (section 11.5).

Overall, this chapter demonstrates that representing FLs within individual-based simulation models in a biologically plausible way can improve predictions about emergent diversity and perpetual novelty in nature. The approach described here may also be usefully adapted into engineering to tackle diversity-related problems in evolvability and evolutionary search.

11.1. A genotype model with environment adaptation traits and holey fitness landscape

The material presented in this section is based on the following publications:

 G. Paperin, D. G. Green, A. Dorin (2007):
Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].

In this section I introduce a new genotype model that combines several advantages of the additive trait model (section 10.2.2) and the HFL model (section 6.2.4), while avoiding several of their drawbacks. The advantages provided by the new model include for instance the requirements outlined in section 10.2.1, such as the possibility to efficiently assess the degree of reproductive isolation between individuals and populations, and the absence of arbitrary tuning parameters.

The genotype is represented in the same way as in the previous simulation model (section A.1.1). To recap: A genotype is modelled as a collection of diallelic loci, although the model can be easily extended to any number of alleles. Each locus determines the value of a phenotypic trait. Here, there is a 1:1 correspondence between the diallelic loci and the presence or absence of a corresponding trait, however, in section 10.2.2 I showed that this is equivalent to using several loci per real-valued trait with fewer traits. Each trait contributes to the fitness of the individual within the current environment and the overall fitness of a genotype is based on the average of the fitness contributions of individual traits.

Formally, the genotype $G_\lambda \in \mathbb{G}$ of an individual λ is represented by a binary string of length L , where $\mathbb{G}$ is the set of all possible genotypes, and L is the number of loci: $G_\lambda = (g_1, g_2, \dots, g_L)$, $g_k \in \{0, 1\}$, $k = 1 \dots L$. Each locus g_k determines whether the trait k is expressed in the phenotype of λ .

Individuals adapt to live in ecological niches. A niche M is represented by a binary string: $M = (m_1, m_2, \dots, m_L)$, where $m_i \in \{0, 1\}$. The binary values m_i are the *trait requirements* of the niche M : if $m_i = 1$, then trait i must be present in an individual in order to be fully adapted to niche M ; if $m_i = 0$, then trait i must be absent.

The degree of adaptation of an individual λ to a niche M is described by the *niche-distance* of λ with respect to M . The niche distance is defined as the as the hamming distance between the trait values of G_λ and the trait requirements of M :

$$\text{dist}(G_\lambda, M) = \sum_{i=1}^L |m_i - g_i|$$

For example, assume $L = 10$, $G_\lambda = (1010100111)$ and $M = (1001010100)$. We have:

Genotype G_λ :	1	0	1	0	1	0	0	1	1	1
Niche M :	1	0	0	1	0	1	0	1	0	0
Niche-distance at each locus:	0	0	1	1	1	1	0	0	1	1

$$\text{dist}(G_\lambda, M) = 0 + 0 + 1 + 1 + 1 + 1 + 0 + 0 + 1 + 1 = 6$$

The fitness $\text{fitness}_M(\lambda)$ of an individual λ within a niche M can be defined as any continuous, non-increasing function of $\text{dist}(G_\lambda, M)$. Gavritels [2004, chap. 2 and 8] uses a Gaussian function, but other choices are possible. For the rest of this chapter I assume that fitness is a real number within the unit interval with larger values describing fitter individuals ($0 \leq \text{fitness}(\lambda) \leq 1$, $\text{fitness}(\lambda) \in \mathbb{R}$).

Assume that there are S distinct niches $M_1, \dots, M_S$. A *closest niche* M_λ^* for an individual λ is a niche with the smallest distance:

$$M_\lambda^* = \arg \min_{M_j} \text{dist}(G_\lambda, M_j), \quad j = 1 \dots S$$

Note that an individual may have more than one closest niche. Assume that individuals always exist within a closest habitat niche. Thus, fitness is calculated with respect to such niche: $\text{fitness}(\lambda) = \text{fitness}_{M_\lambda^*}(\lambda)$. Observe that if an individual has multiple equidistant closest niches it does not affect its fitness. Individuals that do not have a unique closest niche are called *undifferentiated*.

Recall the notion of HFL (section 6.2.4). That notion can be applied to the above genotype model as follows:

Let $\mathbb{V} \subseteq \mathbb{G}$ be the set of all *viable* genotypes and assume that a genotype $G \in \mathbb{G}$ is viable ($G \in \mathbb{V}$) with probability p and *inviable* ($G \notin \mathbb{V}$) with probability $(1 - p)$.

For each $G \notin \mathbb{V}$, set some small fitness value $\text{fitness}(G) = \text{fit}_{\text{Inv}} \geq 0$. For the viable genotypes, define a minimum fitness fit_v ($\text{fit}_{\text{Inv}} \ll \text{fit}_v < 1$). For each $G \in \mathbb{V}$, let $\text{fitness}(G)$ be a continuous, non-increasing function of $\text{dist}(G, M_\lambda^*)$, such that $\text{fit}_{\text{Inv}} \leq \text{fitness}(G) \leq 1$.

In other words, the fitness of inviable genotypes is a small, constant, non-negative value fit_{Inv} ; the fitness of viable genotypes is a value ranging between fit_v and 1, where fit_v is significantly larger than fit_{Inv} (it can be shown [Gavrilets, 2004, chap. 4] that holey FL models are not particularly sensitive to the difference between fit_{Inv} and fit_v). Notably, for the special case $\text{fit}_v = 1$, this model corresponds to the standard HFL model.

11.2. Holey fitness landscapes in simulations

The material presented in this section is based on the following publications:

-  G. Paperin, S. Sadedin, D. Green, A. Dorin (2008):
 Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity.
 11th International Conference on Artificial Life (ALife XI). Ref: [Paperin, Sadedin, et al., 2008].
-  G. Paperin, D. G. Green, A. Dorin (2007):
 Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
 2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].

There is a number of models of adaptive radiation based on HFLs, however they are analytic and do not consider ecological selection, spatial distribution and other scenarios that are better explored with simulation models. The absence of *in silico* models of HFLs may be due to a number of difficulties that arise then realising a HLF in a computer model.

11.2.1. Feasibility of truly random distributions in computer simulations

Recall that in a HFL, every genotype $G \in \mathbb{G}$ is viable with a small probability p *independently* of all other genotypes. Before discussing approaches to creating any set of entities that are randomly and independently selected using a computer, it is important to highlight the general limitations of modern computers in this respect. The usual source of random numbers in computers are *pseudo-random* number generators. These are functions, that generate values that approximate the properties of truly random numbers. Usually, these are realised by iterative algorithms, which, when initialised with a unique seed, produce a sequence of numbers with an approximately uniform distribution and a long cycle period within some target interval. Good pseudo-random number generators produce sequences that are indistinguishable from truly random numbers for most purposes. Still, specialised tests can detect statistical artefacts in sequences produced by many common pseudo-random number generators. Also, when initialised with the same seed, such generators produce exactly the same sequence every time, and the sequence repeats once the period length, however long, has been reached. Truly random number generators can be implemented in hardware, however these tend to be slow and expensive and are thus not generally available on systems commonly used for scientific simulations. Modern mainstream programming languages include algorithms for pseudo-random number generation in their standard libraries. For most purposes, these algorithms provide sufficiently randomly and independently distributed values, when initialised with a unique seed drawn from the computer clock. In all experiments described here I use the generator provided by the standard Java 6 API (`java.util.Random`) that uses the algorithm described by Knuth [1998, vol. 3, section 3.2.1]. Sequences produced by this pseudo-random number generator are sufficiently uncorrelated and have a sufficiently long cycle for all present purposes. Thus, throughout the following discussion I use the term *random* to refer to both, truly random and pseudo-random values.

11.2.2. Difficulties in building HFL models suitable for individual-based simulations

Let $viable(G): \mathbb{G} \rightarrow \{0, 1\}$ be a binary function that determines whether G is viable ($viable(G) = 1$), or inviable ($viable(G) = 0$). In a simulation, a particular G may be encountered at many occasions and the value of $viable(G)$ for a given G must always be the same. Generally, there are two ways to achieve that. The first approach is to design the function $viable(G)$ so that it can be computed in a computationally efficient manner. Then, the function value can be recomputed at every occasion G is encountered without compromising an acceptable run-time performance. The second approach is to compute the

value of $\text{viable}(G)$ only once, either when G is first encountered, or at the prior to the simulation, and then to look up and reuse the computed value every time G is encountered again.

When using the first approach, it must be ensured that $\text{viable}(G)$ always results in the same value for the same G and yet produces independent random values for different G . This can be achieved by assigning a constant random real number dice uniformly distributed in $[0, 1[$ to every G independent of all other genotypes. When $\text{dice} < p$, then $\text{viable}(G) = 1$; otherwise $\text{viable}(G) = 0$. A technique that ensures that the same dice value with the above properties is always assigned to a given G can be borrowed from pseudo-random number generators available on all modern computers (section 11.2.1). Such generators produce the same sequence of (pseudo-)random numbers for any given seed, but a different sequence for different seeds. Thus, when a pseudo-random generator is seeded with G , it will produce a dice value with the above desired properties. As only one random number per genotype is required, this approach can be implemented very efficiently.

Recall that $\mathbb{V}$ denotes the set of viable genotypes. By definition, $\mathbb{V} \subseteq \mathbb{G}$, however, since p is small, in practice $\mathbb{V} \subset \mathbb{G}$. It can be shown [Gavrilets & Gravner, 1997] that in a HFL the majority of viable genotypes $G \in \mathbb{V}$ belong to a single largest connected component $\mathbb{V}' \subseteq \mathbb{V}$. The size of $\mathbb{V}'$ is of the order of $2^L \cdot p$, and $\mathbb{V}'$ percolates $\mathbb{G}$. The proof uses the idea of a surviving branching process to estimate the size of $\mathbb{V}'$.

Assume that $p = p_c = 1 / (L - 1)$. The probability that the branching process dies at any specific branching point is given by

$$(1-p)^{L-1} = \left(1 - \frac{1}{L-1}\right)^{L-1} \approx \left(1 - \frac{1}{L}\right)^L$$

This means that the above statement holds with probability 1 when $L \rightarrow \infty$. For finite but large L this probability is close to 1; however, for smaller L , the probability of the emergence of the giant connected component is smaller.

In natural populations, L is very large, but in an individual-based simulation, the genotype of each individual must be modelled explicitly, held in computer memory and processed by various operations. In practice, this limits the number of loci L to relatively low values. If L is small, $\mathbb{V}'$ can be expected to be small relative to $\mathbb{V}$, i.e. $\mathbb{V}$ can be expected to consist of a large number of small clusters that are not connected to each other. In such case, an adaptive walks starting at some $G \in \mathbb{V}$ cannot proceed far and many key evolutionary properties of HFLs do not hold.

Thus, although a suitable algorithm for re-computing $\text{viable}(G)$ each time G is encountered can be developed as shown above, such approach is not feasible for practical values of L , because the resulting neutral fitness network is not likely to be connected.

Note, however, that even when L is small, the probability that $\mathbb{V}'$ contains most of $\mathbb{V}$ is not large, but positive. For any given L and $p \geq p_c$, consider all possibilities for selecting $\mathbb{V}$ from $\mathbb{G}$. For most of such possibilities, $\mathbb{V}'$ is small, but there are some choices for which $\mathbb{V}'$ is large (Figure 11.1).

Any selection of $\mathbb{V}$ from $\mathbb{G}$, for which the giant connected component $\mathbb{V}'$ emerges, is an approximation of a HFL for large L . For any such selection, all crucial features of $\mathbb{V}'$, $\mathbb{V}$ and $\mathbb{G}$ hold and no assumptions are violated. The key evolutionary properties of HFLs [Gavrilets & Gravner, 1997; Gavrilets, 2004] hold in these cases. In the next sub-section I describe an algorithm for selecting $\mathbb{V}$ from $\mathbb{G}$ such that the giant connected cluster $\mathbb{V}'$ is guaranteed to emerge. The resulting set of genotypes can be used as a basis for individual-based simulations exploring HFL genetics, in particular for the genotype model proposed in section 11.1 above. The algorithm is based on the second type of approach to creating the function $\text{viable}(G)$ discussed earlier: all viable genotypes are determined before any individual-based simulation commences. During the simulation, the value $\text{viable}(G)$ is determined through a look-up.

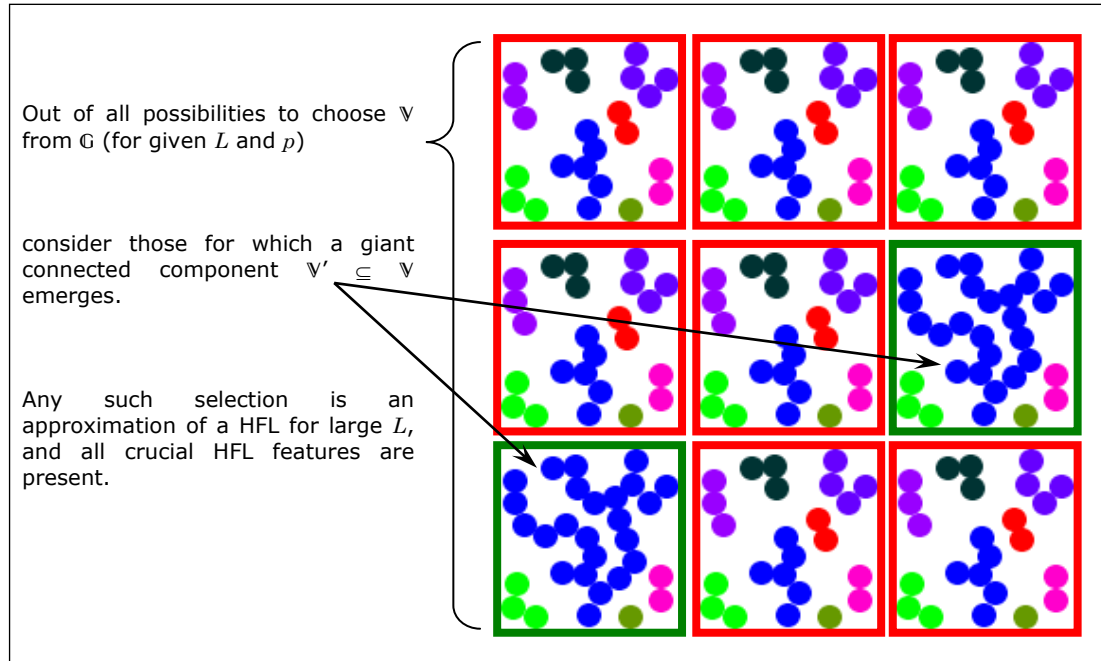


Figure 11.1. Approach to independent random selection of viable genotypes for a HFL suitable for computer modelling.

11.2.3. Creating a HFL suitable for individual-based simulations

Here I describe an approach to implementing the FL model described in [section 11.1](#). As discussed in the previous section, the main challenge is to select a set of viable genotypes $\mathbb{V} \subset \mathbb{G}$ such that most $G \in \mathbb{V}$ belong to a single connected component $\mathbb{V}' \subseteq \mathbb{V}$. Once this task is accomplished, it is relatively straight forward to determine the closest niche(s) for each genotype and thus the fitness. Finally, an efficient look-up table can be built to contain all $G \in \mathbb{V}$. Such table can be used to quickly determine whether G is viable at later occasions.

Formally, the problem is to creating a component $\mathbb{V}$ in a 2^L -dimensional discrete space $\mathbb{G}$, which has the following properties: (1) $\mathbb{V}$ is connected; (2) $\mathbb{V}$ percolates $\mathbb{G}$, i.e. any point $I \in (\mathbb{G} \setminus \mathbb{V})$ comes near a point $V \in \mathbb{V}$; (3) the points $V \in \mathbb{V}$ are uniformly distributed in $\mathbb{G}$.

First determine the “target” size v of $\mathbb{V}$ ($v = |\mathbb{V}|$). Note that v is binomially distributed: $v \sim \text{Bin}(|\mathbb{G}|, p)$. This can be approximated by a Gaussian, and v can be drawn from a normal distribution provided by a standard random number generator.

Phase one:

Begin with an empty $\mathbb{V}$. Randomly pick points in $\mathbb{G}$ and assign them to $\mathbb{V}$. After a point is picked, group all points in $\mathbb{V}$ are into connected components and determine the largest component $\mathbb{V}'$. If the size of the largest component is smaller than v , repeat the procedure ([Figure 11.2](#)). At some time, the largest connected component $\mathbb{V}'$ will contain v or more points. At this time, discard all other components, now $\mathbb{V} = \mathbb{V}'$ is a single connected component. As all points are picked independently and randomly, $\mathbb{V}$ is also uniformly distributed in $\mathbb{G}$, and as shown in [Gavrilets & Gravner, 1997], $\mathbb{V}$ percolates $\mathbb{G}$ even for relatively small p .

Phase two:

While $\mathbb{V}$ consisted of several disconnected components, adding a single point to $\mathbb{V}$ increased the size of the largest component by one or more points, since the inserted point could connect the largest component with another component. Therefore, after the first phase is complete, it is likely that $|\mathbb{V}| > v$. In such case, a number of points must be removed from $\mathbb{V}$.

In order to do that, randomly select a point $V \in \mathbb{V}$ and check whether removing V would split $\mathbb{V}$ in two disconnected components. If $\mathbb{V} \setminus V$ is still connected, then V is removed, otherwise a new point is picked (Figure 11.3). This procedure is repeated until $|\mathbb{V}| = v$.

The described procedure is computationally intensive, particularly as v grows exponentially with L . However, as in practice L is often not large, the algorithm can run in acceptable time. Once $\mathbb{V}$ is created, all subsequent look-ups are fast. Moreover, various optimisations can be used for the algorithm. For instance, the clustering operations during the first phase do not need to occur after each point is added to $\mathbb{V}$, but can be carried out at larger intervals. In specific cases during the second phase, it is possible to remove several points from $\mathbb{V}$ at once. Additionally, v does not need to be fixed up-front. Instead, during the second phase it can be estimated whether the current size $|\mathbb{V}|$ is likely to be drawn from $\text{Bin}(|\mathbb{G}|, p)$ and thus to accept a larger number of possible v .

A reference implementation of the above algorithm in Java 6 creates a $\mathbb{V}$ for $L = 22$ in approximately 2.5 minutes when executed on a 3.2 GHz Pentium Prescott PC with 1GB RAM running Win XP. The initial study was conducted in 2007, and advances in hardware performance and memory availability that occurred since then allow creating $\mathbb{V}$ for $L = 26$ in approximately the same time. Since the exponential-time algorithm to create $\mathbb{V}$ only needs to be run once (for every simulation run), and subsequently look-up is fast, simulation models with genotype length $L = 32$ on the basis of the above technique are feasible today.

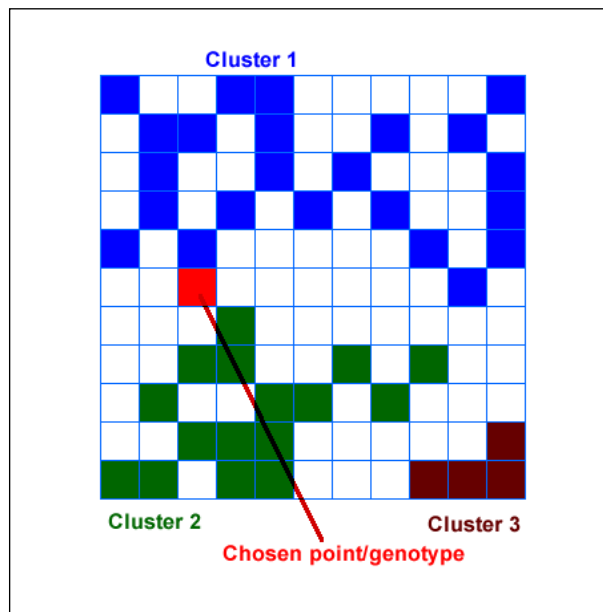


Figure 11.2. Phase 1 of algorithm for selection of viable genotypes for a HFL computer model. Randomly pick points in G and assign them to $\mathbb{V}$. After a point is picked, all points in $\mathbb{V}$ are grouped into connected components and the largest component $\mathbb{V}'$ is determined. Repeat until $\mathbb{V}'$ contains a sufficient number of points, and discard all other components except $\mathbb{V}'$. Now $\mathbb{V} = \mathbb{V}'$ is a single connected component.

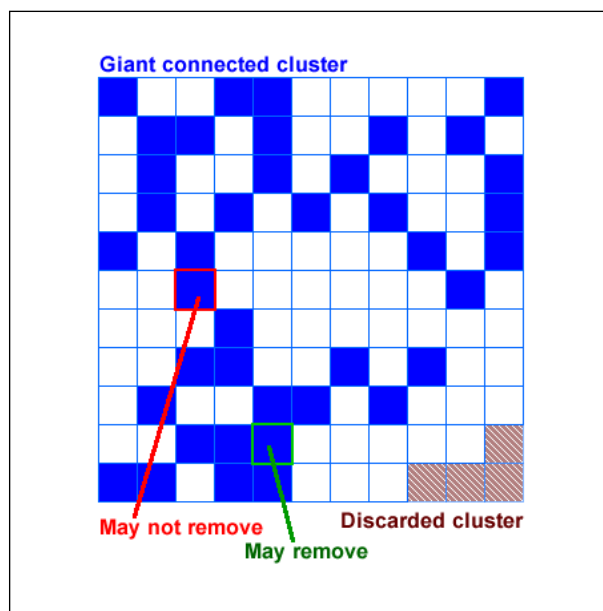


Figure 11.3. Phase 2 of algorithm for selection of viable genotypes for a HFL computer model. While $\mathbb{V}$ consisted of several components, adding a point to $\mathbb{V}$ could increase the size of $\mathbb{V}'$ by more than 1 point (may connect two clusters). Thus, after phase 1 $|\mathbb{V}|$ may be too large. Need to remove some points: Randomly select a $V \in \mathbb{V}$ and check whether removing V would split $\mathbb{V}$ in two disconnected components. If not - remove V , otherwise pick a different point. Repeat until $\mathbb{V}$ has the right size.

11.3. Evolutionary properties of the genotype model with environment adaptation traits and holey fitness landscape

The material presented in this section is based on the following publications:

📖 G. Paperin, D. G. Green, A. Dorin (2007):
Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation.
2007 International Symposium on Computational Models for Life Sciences (CMLS'07). Ref: [Paperin, Green & Dorin, 2007].

Here, I perform an extensive numerical analysis of the evolutionary properties of the FL that results from the genotype model proposed in [section 11.1](#). The aim is to provide detailed information about the FL in order to inform the design and the analysis of evolutionary models based on this method. A range of different techniques for FL analysis is discussed by Kaufman [1993] and Stadler [2002]. Here I use a mixture of such techniques as well as some approaches developed specifically for this case. Note that while the analysis in [section 10.2 above](#) was specifically targeted to provide a better context for the simulation results presented in that chapter, here I aim for a more generic discussion that shall inform future studies. Thus, I explicitly list data for more different combinations of (L, S) values and provide statistically significant averages where possible. Note that the following analysis is also significant for the standard HFL model, as the standard HFL model is a special case of the current FL when the fitness function has a constant high-fitness value for all viable genotypes.

Recall that L is the *number of traits* in the model. Due the employed representation of genotypes and niches, L is also the *number of trait requirements of a niche*, the *number of loci in a genotype* and the *length of bit-strings that represent genotypes and niches*. Recall that S is the *number of distinct habitat niches*.

The fitness of all $V \in \mathbb{V}$ is significantly higher than the fitness of all $I \in (\mathbb{G} \setminus \mathbb{V})$. In fact, all $I \in (\mathbb{G} \setminus \mathbb{V})$ are inviable. Thus all evolutionary dynamics involve $V \in \mathbb{V}$ only. In all following measurements I therefore only consider the viable genotypes $V \in \mathbb{V}$. For each set of parameters all quantities discussed below are measured on 20 random FLs. This means that for every type of measurement, a new set $\mathbb{V} \subset \mathbb{G}$ is selected from scratch according to the algorithm described in [section 11.2.3](#) 20 times, and a new random set $\mathbb{M}$ of niches is selected in $\mathbb{G}$ at each occasion. The results are averaged. The possible parameters are L , S , and p - the probability that a genotype is viable. As p merely affects the size of $\mathbb{V}$, not its structure, the value $p = 1.05 / (L - 1)$ is used throughout. L takes the values 12, 14, 16, 18, 20 and 22. For each L , S takes the values 10, 50, 100 and 250. Various choices of the fitness function and the values for fit_{inv} and fit_V affect the smoothness of the FL, but not its structure. In order to avoid artefacts introduced by those values, the measurements relate to the *closest-niche distance* of genotypes, rather than their fitness.

11.3.1. Adaptation properties

The expected number of *viable genotypes* is $1.05 \times 2^L / (L - 1) \approx 2^L / L$, which I confirmed experimentally. As shown in [Table 11.1](#), the *proportion of local optima* in $\mathbb{V}$ drops when S becomes larger. This suggests that the FL becomes less rugged as the number of niches increases.

The *number of viable neighbours* of a $V \in \mathbb{V}$ does not depend on the number of niches ([Table 11.2, left](#)). The number of neighbours decreases as L increases, which is expected as p is a reciprocal of L . [Table 11.2 \(left\)](#) shows that the proportion of viable neighbours is approximately $2p$. The *number of neighbours with smaller closest-niche-distance* reflects the number of advantageous mutations for a genotype. Equally, the *number of neighbours with smaller or equal closest-niche-distance* reflects the number of non-deleterious mutations. Similarly to the total number of possible mutations, the numbers of advantageous and non-

deleterious mutations weakly depends on L (Table 11.2, centre & right). For a fixed L , the number of non-deleterious mutations grows when S increases (Table 11.2, right), which can be expected since the FL becomes smoother. The number of advantageous mutations decreases when S increases (Table 11.2, centre) which suggest that when the number of niches is large, a larger number of mutations are neutral.

The *length of a random adaptive walk* measures the number of non-deleterious mutations that occur before a local optimum is found. 10 random walks are simulated using each of the viable genotypes as a starting point. At each step, one of the non-deleterious neighbour genotypes of the current genotype is chosen with equal probability. To avoid infinite loops, a walk is prevented from visiting a genotype more than once. Table 11.3 (left) shows that for a fixed L , the average walk length increases as S grows, which is what can be expected for increasingly smooth FLs. A related measure is the *reach cluster size* – the total number of genotypes that can be potentially reached by an adaptive walk starting from each of the viable genotypes. Table 11.3 (centre) shows that the reach cluster sizes behave similarly to the lengths of adaptive walks. This means, that when the number of niches becomes larger and the FL becomes smoother, a random viable genotype can evolve into a larger number of different genotypes. A similar observation applies to the *basin sizes of local optima*: for larger S , more genotypes may evolve towards a given local optimum (Table 11.3, right).

L	S	% local optima
12 $p \approx 0.096$	10	17%
	50	14%
	100	15%
	250	9%
14 $p \approx 0.081$	10	20%
	50	15%
	100	13%
	250	13%
16 $p = 0.070$	10	21%
	50	18%
	100	16%
	250	13%
18 $p \approx 0.062$	10	24%
	50	21%
	100	20%
	250	18%
20 $p \approx 0.055$	10	24%
	50	22%
	100	21%
	250	20%
22 $p = 0.050$	10	24%
	50	23%
	100	22%
	250	21%

Table 11.1. Percentage of local optima genotypes in respect to all viable genotypes. Results averaged over 20 random FLs for each combination of L and S .

L	S	Viable neighbours	Better neighbours	Better or equal neighbours
12 $p \approx 0.096$	10	2.5 (0.204)	1.0 (0.081)	1.5 (0.124)
	50	2.5 (0.206)	0.9 (0.076)	1.6 (0.130)
	100	2.4 (0.204)	0.9 (0.073)	1.6 (0.150)
	250	2.4 (0.202)	0.6 (0.052)	1.8 (0.106)
14 $p \approx 0.081$	10	2.4 (0.175)	1.0 (0.069)	1.5 (0.111)
	50	2.5 (0.176)	0.9 (0.064)	1.6 (0.113)
	100	2.5 (0.176)	0.9 (0.063)	1.6 (0.113)
	250	2.5 (0.176)	0.9 (0.063)	1.6 (0.092)
16 $p = 0.070$	10	2.5 (0.154)	1.0 (0.062)	1.5 (0.098)
	50	2.5 (0.154)	0.9 (0.056)	1.6 (0.100)
	100	2.5 (0.154)	0.9 (0.054)	1.6 (0.101)
	250	2.5 (0.154)	0.8 (0.053)	1.6 (0.073)
18 $p \approx 0.062$	10	2.2 (0.123)	0.9 (0.050)	1.3 (0.079)
	50	2.2 (0.123)	0.8 (0.044)	1.4 (0.080)
	100	2.2 (0.123)	0.8 (0.043)	1.4 (0.082)
	250	2.2 (0.123)	0.7 (0.041)	1.5 (0.065)
20 $p \approx 0.055$	10	2.2 (0.111)	0.9 (0.046)	1.3 (0.070)
	50	2.2 (0.111)	0.8 (0.041)	1.4 (0.072)
	100	2.2 (0.111)	0.8 (0.039)	1.4 (0.073)
	250	2.2 (0.111)	0.8 (0.038)	1.5 (0.059)
22 $p = 0.050$	10	2.2 (0.100)	0.9 (0.042)	1.3 (0.063)
	50	2.2 (0.100)	0.8 (0.038)	1.4 (0.064)
	100	2.2 (0.100)	0.8 (0.036)	1.4 (0.066)
	250	2.2 (0.100)	0.8 (0.034)	1.4 (0.130)

Table 11.2. Number of possible adaptive mutations per genotype. Shown are: Number of viable neighbours per genotype; Number of neighbours with smaller closest-niche-distance per genotype; Number of neighbours with smaller or equal closest-niche-distance per genotype. **In brackets:** The respective number of neighbours normalised by L . Results averaged over 20 random FLs for each combination of L and S .

L	S	Random adaptive walks length	Reach clusters size	Basin size of local optima
12 $p \approx 0.096$	10	84	9	14
	50	230	11	19
	100	423	10	13
	250	825	35	5
14 $p \approx 0.081$	10	60	10	15
	50	97	13	21
	100	173	13	24
	250	337	12	21
16 $p = 0.070$	10	54	12	17
	50	91	15	21
	100	106	19	28
	250	138	22	35
18 $p \approx 0.062$	10	70	7	9
	50	112	8	11
	100	129	9	12
	250	148	10	13
20 $p \approx 0.055$	10	56	7	9
	50	105	8	11
	100	124	9	12
	250	142	10	13
22 $p = 0.050$	10	58	7	9
	50	96	8	11
	100	114	9	11
	250	134	10	13

Table 11.3. Measures of possible adaptation trajectories. Shown are: Mean length of random adaptive walks; Mean size of reach clusters for each viable genotype; Mean size of (possibly overlapping) attraction basins of local optima. Results averaged over 20 random FLs for each combination of L and S .

11.3.2. Speciation properties

It is possible for a genotype to have the same niche-distance from many niches. Genotypes that have more than one unique closest niche must be considered *undifferentiated* (note that such genotypes may still have high fitness if their niches are sufficiently similar). Important questions are: What proportion of viable genotypes are undifferentiated? and; How do their niche distances differ from the niche distances of genotypes which have a unique niche? When the number of niches is large, a larger part of the genotype space is expected to become “border lines” between the niche basins. Table 11.4 confirms this intuition: when L is fixed, the *number of genotypes with unique niches* falls when S increases. An exception occurs when L is low. This may be due to the very low genotypes/niches ratio. The table also shows that the *proportion of undifferentiated genotypes* does not tend to vary with L .

Table 11.5 shows that the genotypes differentiated towards a unique niche are on average fitter than undifferentiated genotypes. Other measurements (not shown) demonstrate that the minima and maxima of closest-niche distances also tend to be lower for differentiated genotypes. It can therefore be expected, that populations evolving on this FL will tend to adapt towards unique niches.

Now, I verify the assumption that genotypes from different niches are likely to produce inviable offspring causing reproductive isolation between niche populations. For each viable genotype differentiated towards a unique niche, 5 random mating partners are selected from the same niche and 5 from a different niche (a pair may be chosen more than once if a niche contains less than 6 genotypes). For each of the resulting pairs, an offspring is created using free recombination (i.e. each locus is inherited from either parent with equal probability). Offspring viability is determined. As before, all measurements are repeated on 20 random FLs and averaged. In order to investigate how the degree of adaptation affects the result, I repeat the measurements for parents with a closest-niche distance smaller or equal to different values $d_{\min}$.

The measurements show that the proportion of viable offspring of parents from *different niches* does not depend on the number of niches S (Table 11.7). Conversely, the proportion of viable offspring of parents from the *same niche* increases with S (Table 11.6). This is because a larger number of niches leads to smaller, more specialised and constrained niche-populations with more similar organisms within each population.

For parents from the *same niche*, the proportion of viable offspring varies between ca. 40% ($L = 12$) and ca. 8% ($L = 22$). For parents from *different niches*, the proportion is significantly lower (from ca. 20% for $L = 12$ it falls to ca. 6% for $L = 22$).

For fitter parents from the *same niche*, the proportion of viable offspring rises sharply, however, for parents from *different niche-populations* the corresponding increase is negligible. For $L = 22$ and $d_{\min} = 3$, ca. 20% of offspring are viable for *same niche* parents, and only ca. 5.5% for *different niche* parents; for $d_{\min} = 2$: ca. 40% for *same niche* parents, and ca. 5.6% for *different niches*. In summary, recombination between viable genotypes with above-average fitness from *different niches* is only marginally more likely to lead to viable offspring than a random genotype (recall, $p = 0.05 = 5\%$). On contrary, recombination within the *same niche* is significantly more probable to result in viable genotypes. This implies that the assumption that niche populations on this FL are reproductively isolated is indeed correct.

L	S	% differentiated	% undifferentiated
12 $p \approx$ 0.096	10	66%	34%
	50	60%	40%
	100	63%	37%
	250	78%	22%
14 $p \approx$ 0.081	10	68%	32%
	50	58%	42%
	100	57%	43%
	250	60%	40%
16 $p =$ 0.070	10	68%	32%
	50	58%	42%
	100	56%	44%
	250	55%	45%
18 $p \approx$ 0.062	10	69%	31%
	50	59%	41%
	100	56%	44%
	250	53%	47%
20 $p \approx$ 0.055	10	69%	31%
	50	61%	39%
	100	57%	43%
	250	54%	46%
22 $p =$ 0.050	10	72%	28%
	50	62%	38%
	100	58%	42%
	250	55%	45%

Table 11.4. Proportions of differentiated and undifferentiated genotypes. Shown are: Percentage of viable genotypes adapted to a unique niche in respect to all viable genotypes; Percentage of viable genotypes that have a multiple equidistant closest-niches. Results averaged over 20 random FLs for each combination of L and S .

L	S	$d_{\min}$ of differentiated G	$d_{\min}$ of undifferentiated G
12 $p \approx$ 0.096	10	2.73	3.84
	50	1.30	2.44
	100	0.76	1.89
	250	0.19	1.25
14 $p \approx$ 0.081	10	3.63	4.75
	50	2.18	3.26
	100	1.63	2.75
	250	0.95	2.07
16 $p =$ 0.070	10	4.52	5.62
	50	3.05	4.08
	100	2.49	3.54
	250	1.79	2.87
18 $p \approx$ 0.062	10	5.41	6.56
	50	3.89	4.90
	100	3.33	4.33
	250	2.65	3.67
20 $p \approx$ 0.055	10	6.21	7.39
	50	4.65	5.68
	100	4.09	5.08
	250	3.41	4.38
22 $p =$ 0.050	10	7.06	8.31
	50	5.41	6.47
	100	4.83	5.84
	250	4.14	5.09

Table 11.5. Average closest-niche distances of genotypes. Shown are: Mean closest-niche distances of genotypes adapted to a unique niche; Mean closest-niche distances of genotypes that have a multiple equidistant closest-niches. Results averaged over 20 random FLs for each combination of L and S .

L	S	$d_{\min} = \infty$	$d_{\min} = 5$	$d_{\min} = 4$	$d_{\min} = 3$	$d_{\min} = 2$	$d_{\min} = 1$
12 $p \approx 0.096$	10	37.4	37.8	38.0	43.4	59.2	92.1
	50	68.1	68.0	67.9	68.0	71.5	93.6
	100	84.1	84.2	84.2	84.3	84.4	95.1
	250	98.1	98.2	98.4	98.2	98.2	98.6
14 $p \approx 0.08$	10	23.6	23.8	26.1	33.9	53.9	91.1
	50	42.2	42.2	42.4	43.4	57.7	92.1
	100	55.7	55.7	55.7	55.9	62.6	92.4
	250	78.2	78.1	77.7	78.0	78.5	94.3
16 $p = 0.070$	10	15.8	16.9	20.1	29.0	50.9	90.9
	50	25.2	25.4	25.7	31.3	51.7	91.0
	100	33.0	32.9	33.1	35.0	52.6	91.4
	250	48.4	48.6	48.8	48.3	57.0	91.9
18 $p \approx 0.062$	10	10.2	11.7	15.2	23.1	46.1	92.0
	50	14.7	14.6	16.3	23.8	46.2	91.4
	100	18.1	18.1	18.5	24.8	46.3	91.3
	250	25.8	25.9	25.9	28.0	47.2	91.8
20 $p \approx 0.055$	10	7.7	9.8	13.1	21.1	43.8	91.5
	50	10.1	10.6	13.5	21.2	43.4	91.7
	100	11.9	12.0	14.1	21.4	43.4	91.3
	250	15.5	15.7	15.9	22.0	44.3	91.5
22 $p = 0.050$	10	6.4	8.4	11.4	18.9	41.7	91.0
	50	7.6	8.7	11.5	19.0	41.4	91.6
	100	8.4	9.1	11.7	19.1	41.8	91.1
	250	10.3	10.4	12.2	19.4	41.5	91.2

Table 11.6. Proportion of viable offspring of parent genotypes from the name niche.

Shown are the percentages of viable offspring from random parents originating from the same habitat niche. $d_{\min}$ is the maximum closest-niche distance permitted for each parent. Results averaged over 20 random FLs for each combination of L and S .

L	S	$d_{\min} = \infty$	$d_{\min} = 5$	$d_{\min} = 4$	$d_{\min} = 3$	$d_{\min} = 2$	$d_{\min} = 1$
12 $p \approx 0.096$	10	19.4	19.4	19.6	19.9	21.3	22.1
	50	19.9	20.0	20.1	20.0	20.3	21.3
	100	20.1	19.8	20.1	19.9	20.1	20.6
	250	20.3	20.2	20.1	20.3	20.3	20.3
14 $p \approx 0.08$	10	13.5	13.6	13.8	14.5	15.0	14.6
	50	14.3	14.3	14.2	14.3	14.9	15.3
	100	14.6	14.8	14.4	14.4	15.0	15.4
	250	14.4	14.2	14.3	14.3	14.3	15.0
16 $p = 0.070$	10	10.5	10.4	10.7	11.3	11.6	11.7
	50	10.7	10.8	10.7	11.1	11.6	11.3
	100	10.7	10.6	10.8	11.1	11.3	11.7
	250	10.7	10.7	10.9	10.7	11.2	11.4
18 $p \approx 0.062$	10	7.6	7.6	7.6	7.8	8.1	8.0
	50	7.7	7.8	7.7	8.0	8.2	8.0
	100	7.9	7.7	7.7	7.8	8.2	8.1
	250	7.8	7.9	7.9	7.8	8.1	8.0
20 $p \approx 0.055$	10	6.4	6.3	6.5	6.4	6.6	6.6
	50	6.3	6.3	6.5	6.7	6.5	6.6
	100	6.4	6.6	6.5	6.6	6.6	6.5
	250	6.4	6.4	6.5	6.5	6.6	6.7
22 $p = 0.050$	10	5.4	5.5	5.5	5.5	5.8	5.6
	50	5.5	5.5	5.5	5.7	5.7	5.5
	100	5.5	5.4	5.6	5.4	5.7	5.7
	250	5.6	5.5	5.4	5.6	5.6	5.6

Table 11.7. Proportion of viable offspring of parent genotypes from different niches.

Shown are the percentages of viable offspring from random parents originating from different habitat niches. $d_{\min}$ is the maximum closest-niche distance permitted for each parent. Results averaged over 20 random FLs for each combination of L and S .

11.3.3. Conclusion

The above results show that the FL of the genotype model introduced in this chapter exhibits a number of desirable nature-like properties [Kauffman, 1993; Gavrillets, 2004]. In particular, it improves on the structure of the FL employed in the simulation from [chapter 10](#) in several respects: The majority of the genotypes are inviable. For viable genotypes, the number of atomic mutations leading to viable adaptations is small; the proportion of non-deleterious mutations is even smaller. The number of local optima is large, and fitness canyons are numerous and deep. The overall surface of the FL is rugged. Moreover, the proportion of viable recombinations of genotypes that have an average or above-average fitness within the same habitat niche is by an order of magnitude higher than the proportion of viable recombinations of genotypes adapted to different niches. The latter proportion is only marginally higher than the total proportion of viable genotypes. This implies that different niche populations exhibit postzygotic reproductive isolation and that individuals can be grouped into species according to their habitat niche without the use of arbitrary parameters. In [section 11.4](#) I discuss a series of individual-based simulation experiments that show that postzygotic reproductive isolation between different populations based on a HFL genetic model is indeed strong and persistent.

The present model does not attribute any meaning to specific traits, however such meaning may be introduced. This can allow an investigation of a number of speciation scenarios, that combine postzygotic reproductive isolation inherent in this model with other kinds of reproductive isolation, for instance according to trait preference. This model is thus particularly exciting for potential investigations of the relationship between reproductive isolation caused by social selection (e.g. sexual selection) and genetically based reproductive isolation.

Moreover, the present model can be used in scenarios that incorporate dynamical environments or co-evolution. The survival viability of specific genetic sequences in nature is determined by biochemical constraints and does not change over time, however the fitness given by the degree of adaptation to different environmental niches may change quickly (on evolutionary timescales). This is conveniently reflected in the structure of the model. While the set $\mathbb{V}$ of viable genotypes remains constant over time in most evolutionary scenarios, the number and availability of habitat niches can change dynamically. Additionally, niche-specific fitness functions may be used in a way similar to the simulation model from [chapter 10](#): For instance, in a hot and dry environment with only a few water patches, the ecological niche of living in a wet environment will provide low fitness, even to individuals that are well adapted to such a niche. This can be modelled by introducing niche-specific constraints to the fitness function that determines the fitness of an individual based on its niche distance. This approach allows realising spatially-explicit models in which the environment varies in space as well as in time.

The current FL reduces to the standard HFL model in the case when the fitness function does not depend on the closest-niche distance. Thus, the above analysis of the FL properties bears wider significance as it provides insights into the topology of HFLs. The above numerical analysis is generally in-line with analytical insights into HFLs [Gavrillets & Gravner, 1997]. It also provides new numerical results that could not be obtained analytically. For instance, the measurement of the number of viable mutants ([Table 11.2](#)) and the proportion of viable recombinations ([Table 11.7](#)) have a meaningful interpretation without a niche context.

In summary, the family of FL models presented in this paper provides a useful tool for the investigation of a wide range of biologically plausible evolutionary scenarios. I am looking forward to seeing future research that combines individual-based simulations with the genetic model presented here. In the next section I begin laying the ground and introduce an individual-based simulation model aiming to investigate the degree and the robustness of reproductive isolation provided by HFL-genetics.

11.4. Holey fitness landscapes and the maintenance of evolutionary diversity

The material presented in this section is based on the following publications:

 G. Paperin, S. Sadedin, D. Green, A. Dorin (2008):
Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity.
11th International Conference on Artificial Life (ALife XI). Ref: [Paperin, Sadedin, et al., 2008].

Recent results in theoretical evolutionary biology show that high-dimensional FL form “holey” rather than “rugged” topographies. This lead to the formulation of the HFL model (section 6.2.4). However, the implications of this finding for biological and artificial life systems remain largely unexplored. One of the reasons for this gap can be attributed to serious difficulties in the implementation of individual-based HFL models (section 11.2.2). Previously, I introduced a method for integrating HFLs in individual-based models that overcomes these difficulties (section 11.2.3) and performed a detailed numerical analysis of the properties of HFLs created using that method (section 11.3). Here, I use an individual-based, spatially explicit simulation model to examine how HFL changes predictions for the maintenance of genetic diversity in the face of migration. Previous models suggest that ecologically-based reproductive isolation will rapidly collapse under migration. However, my results indicate that an underlying HFL can often maintain diversity in this situation. Additionally, hybrid species emerge frequently when HFL genetics are simulated, but are usually doomed to extinction because of small population sizes. However, hybridisation can also lead to novel adaptations and potentially the exploitation of new ecological niches. More generally, the results imply that HFL genetics should not be neglected in studies of adaptation and diversity.

11.4.1. Background

The processes underlying the emergence and persistence of diversity form a key topic in evolutionary theory. Analytical models have provided considerable insight into these issues, but integrating the findings from different theoretical approaches remains a formidable challenge. In particular, the relationship between genetic diversity and reproductive isolation – widely considered the defining feature of biological species [Dobzhansky, 1937; Mayr, 1942] – remains controversial [Mallet, 1995; De Queiroz, 1998; Harrison, 1998; Shaw, 1998; Templeton, 1998]. Here, I explore the dynamics of reproductive isolation on HFLs, a genetically realistic FL, within an individual-based, spatially explicit model.

Reproductive isolation is often seen as a requirement for biological diversification because it permits the coexistence of different lineages with co-adapted genomes. However, the origin and persistence of reproductive isolation requires special circumstances. A mutant individual that is reproductively isolated from the surrounding population will rarely be successful. For this reason, speciation is usually thought to occur between spatially separated populations that acquire incompatible alleles through drift or selection [Orr, 1995; Gavrillets, 2003; 2004]. However, even in this scenario, the maintenance of reproductive isolation presents a theoretical challenge: even moderate migration between two populations leads to selection against incompatible alleles, and the extinction or merging of incipient species is likely. Likewise, when reproductive isolation is based on ecological divergence or mating barriers, it is often transient, collapsing when selection pressures change.

Recent theoretical advances suggest that assumptions about the relative fitness of different combinations of traits have profound implications for our understanding of these problems [Gavrillets & Gravner, 1997; Gavrillets, 2004]. Several implications of this insight and the resulting HFL model for speciation theory are explored analytically by Gavrillets [2004].

It is known that diversification occurs easily in large spatial environments with disruptive ecological selection, and there will often be restricted gene flow between the resultant ecotypes, but how enduring reproductive isolation occurs remains unclear. Gene flow barriers induced by mating barriers – even with strong ecological selection – appear to be transient. Models of adaptive radiation and ecological speciation in general deal with this simply by

setting a threshold level of gene flow that they regard as acceptable, but this is unsatisfactory in that such species can merge back together as soon as selective pressures change. HFLs are thought to underlie the evolution of lasting, effective barriers to gene flow that appear during adaptive radiation, however this has not been further explored *in silico* (section 11.2.2).

11.4.2. Methods

My objective is to investigate the extent to which HFL can sustain existing reproductive isolation between spatially isolated populations under different levels of migration. For this I created an individual-based simulation model in which the individuals are located on a homogeneous landscape consisting of cells. Genotype fitness is defined by the HFL, and no habitat niches are unitised. Thus, an individual is either viable (fitness = 1), or inviable (fitness = 0). Individuals mate with other individuals within the same cell and then migrate to a neighbouring cell with a certain probability. I use a number of neutral loci to measure the level of gene flow between the populations in different cells for different migration rates. This technique is widely employed in simulation models of this kind (e.g. [Gavrilets & Vose, 2005]).

Individuals:

In this model the individuals are represented by their genotype, which consists of two sections: a *coding* section and a *neutral* section. The coding section consists of a number of diallelic loci that are assumed to code for vital traits. The *coding section* of a genotype is used as a parameter to the *viable* function of the HFL in order to determine whether an individual is viable. I experimented with 20 to 28 coding loci (not shown here) and found that the particular number does not affect the results significantly. In the experiments reported here I use $L=26$, which represents a trade-off between richer genotypes and the computational resources required to complete a large number of simulation runs. The *neutral genotype section* consists of 5 loci with 128 different alleles possible at each locus. The neutral loci do not affect the fitness (viability) of an individual and are used to measure the genetic divergence between individuals (Figure 11.4).

The lifecycle of a model individual is reproduction – selection – migration. Generations are non-overlapping.

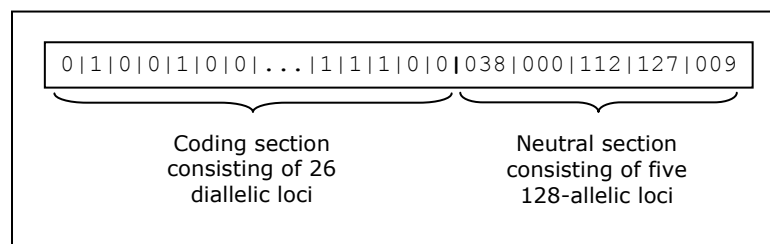


Figure 11.4. An example of a model genotype.

Reproduction:

Individuals mate only with other individuals within the same cell of the spatial landscape. Each individual in a cell is selected once as a mother. For each mother, a partner is randomly selected from the same cell (selfing is permitted). The number of offspring for each pair is drawn from a Poisson distribution with a parameter $\lambda = 4$ (values in range $\lambda = 2..10$ did not affect the results significantly). The genotype of each offspring is determined through free recombination of the parents' genotypes (i.e. the allele at each locus is inherited from each parent with equal probability independent of other loci). Each locus of the offspring is mutated with a probability 10^{-4} (values in the range 10^{-3} to 10^{-5} are commonly used in biological models of this kind, e.g. [Gavrilets & Vose, 2005; Paperin, Green, et al., 2007b]). If a coding locus is mutated, its binary value is flipped. The neutral loci are subject to a

circular stepwise mutation model [Ohta & Kimura, 1973]. If the coding section of an offspring's genotype is determined to be viable by the HFL model, the offspring is added to the new generation, otherwise it is discarded immediately. After all offspring for all pairs of parents have been determined, the old generation is discarded and replaced with the new population.

Selection:

All individuals within a single cell of the spatial landscape compete to survive to the age of reproduction (note that this approach of selecting individuals *for survival* is different from the approach commonly used in genetic algorithms, where all individuals survive and then compete to be selected *for reproduction*). All surviving individuals reproduce and their progeny compete to reach a mature age, which normally requires acquiring environmental resources. Each landscape cell is assumed to have a certain maximum carrying capacity *CapMax*, i.e. to provide enough resources for the survival of *CapMax* mature individuals. If a cell is inhabited by no more than *CapMax* individuals, all survive. Otherwise *CapMax* individuals are selected with equal probability and the rest are discarded (as in this HFL model a particular individual is either fit or inviable).

Dispersal:

Individuals that reach maturity have a certain probability of migrating to one of the neighbouring spatial landscape cells. To avoid edge artefacts the landscape is represented as a torus. The effect of different migration rates is discussed in [section 11.4.3](#), Results.

Experimental scenarios:

In order to investigate how spatial distance affects the results I consider different grid layouts. The simplest case is a 1×2 grid. Other investigated grid sizes are 2×2 cells and 3×3 cells. The results ([section 11.4.3 below](#)) allow to draw conclusions on how the dynamics of reproductive isolation will behave on larger landscapes. Each cell is initialised with a random viable individual with alleles at neutral loci all set to 0. Initially, I disable any migration between the cells and iterate the model for 100 thousand generations in order to allow the allele distribution to reach equilibrium. I then turn on migration at a specific rate (see [section 11.4.3](#)) and iterate the model for 300 thousand further generations. Measurements are taken every 1000 generations.

Measured quantities:

A quantity of prime interest in this model is the number of reproductively isolated groups (RI groups) present in the model at any one time as well as various attributes of such groups. I am interested in groups of genotypes that could inter-mate successfully producing likely viable offspring, not in groups of individuals who actually do so. Finding such groups is difficult as the groups may be partially overlapping (e.g., a genotype can successfully mate with two genotypes that cannot mate with each other) and the genetic distance between groups is initially unknown and may vary. In order to cluster the genotypes of a population into RI groups I employ the Markov Clustering algorithm (MCL) [Van Dongen, 2000b], as it does not require a distance threshold parameter and because it has been successfully applied to a similar task – clustering protein sequences into families [Enright et al., 2002] (here, I essentially cluster gene sequences into families). For that I first calculate a reproductive success probability matrix for all genotypes in the population. The probability of reproductive success of two genotypes is estimated by simulating a large number of crossovers between the genotypes and considering the proportion of crossovers that result in viable offspring. The matrix is then used as input to the clustering algorithm. To further verify the applicability of MCL to the model, I apply this algorithm the model discussed in [chapter 10](#). Recall that that model uses a similar genotype representation (a string of diallelic loci). There, RI groups were determined simply by asserting to which habitat niche genotypes were best adapted. Tests show that the RI groups determined by the clustering correspond to the groups determined by assigning the genotypes to niches ([Figure 11.5](#)).

On the basis of the RI groups I measure the average genetic divergence in neutral loci between the groups using the fixation index F_{st} . A number of slightly different approaches to

calculating F_{st} have been proposed in literature. Here I follow the approach taken by Hudson et al. [1992], who defines F_{st} as the ratio of genetic divergence within a group and the divergence between the groups. More specifically, for every pair of genotypes within a group C , measure the stepwise genetic distance at neutral loci, i.e. the minimum number of stepwise mutations necessary to obtain one genotype (the neutral section thereof) from the other. This allows calculating the average genetic distance $d_w(C)$ *within* a group C . To calculate the average genetic distance $d_B(C)$ *between* C and all other groups, measure the pair-wise distances between all genotypes that belong to C and all genotypes that do not belong to C .

The group fixation index is then $F_{st}(C) = 1 - d_w(C) / d_B(C)$ and the overall fixation index F_{st} is the average of $F_{st}(C_i)$ for all groups C_i . Note that groups of different sizes are treated equally in this approach. $F_{st} \approx 1$ implies high divergence, and $F_{st} \approx \pm 0$ implies low divergence.

Another important quantity is the number of *distinct coding sections*. This essentially corresponds to the number of *distinct phenotypes* in the population (recall, the neutral genotype sections can be thought of as introns and are used only to measure divergence). The number of distinct coding sections is expected to be higher than the number of RI groups (because some coding sections are reproductively compatible), but significantly lower than the number of distinct neutral sections (because phenotypes are constrained by selection for viability and reproductive compatibility with others).

The *turnover of viable coding genotype sections* (i.e. *turnover of viable phenotypes*) in the population is the number of distinct *viable* coding sections that have been present in the model population from the start until a given time. This value can be used to describe the rate at which novel adaptive phenotypes are evolved. On this basis, I relate the degree of reproductive isolation and the migration rate to evolutionary novelty within the system.

For each of the scenarios discussed below I performed 10 independent model runs and averaged the results.

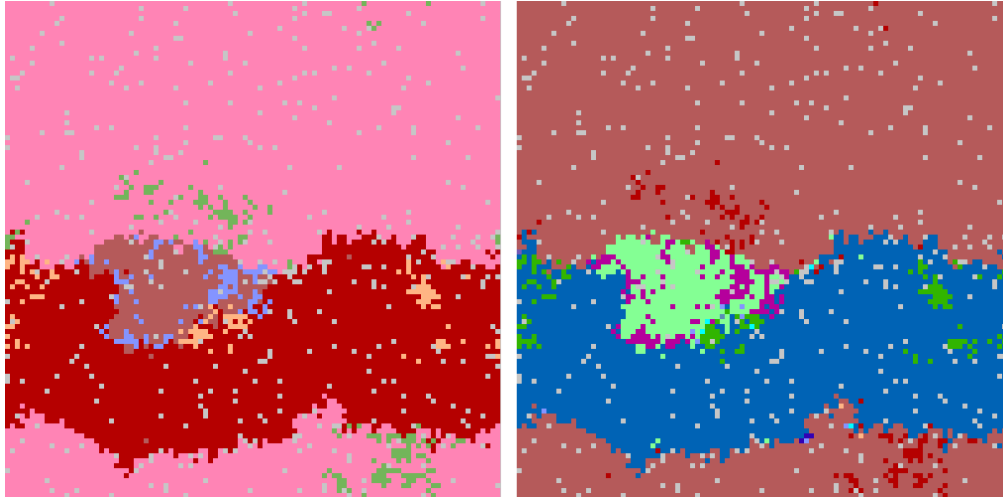


Figure 11.5. Using Markov Clustering (MCL) for determining reproductively isolated groups. Depicted is a snapshot of a spatial landscape (100×100 grid) from the model in chapter 10. Each cell is coloured according to the cluster to which the majority of the genotypes inhabiting the cell belong. **Left:** the genotypes were assigned to RI groups using the MCL algorithm. **Right:** the genotypes were assigned to RI groups according to the ecological niche to which they are best adapted. Although represented by different colours, both groupings are largely the same.

11.4.3. Results

Consider first the 2×2 layout. As a basis for comparison I performed a set of runs with a migration rate of 0%. As expected, the number of reproductively isolated groups (RI groups)

corresponds to the number of cells (i.e., 4), the divergence at neutral loci grows (F_{st} approaches 1) and the number of distinct coding genotype sections in the population fluctuates around a value slightly higher than the number of RI groups – due to viable mutants and drift (see [Figure 11.7](#)). In one of the runs, two of the cells appeared not to be reproductively isolated as the random founder individuals were genetically similar by chance.

In the next scenario I increased the migration rate to 1% after the first 100,000 generations. This led to a slight increase in the number of distinct coding sections in the population which is due to viable hybrids resulting from breeding with immigrants. Some of these hybrids spontaneously form RI groups, however such groups cannot persist due to low population numbers in comparison to native populations. These viable hybrids facilitate a limited gene flow between the populations: after 300,000 generations F_{st} has decreased to ca. 0.8 ([Figure 11.7, D](#)).

In the first 100,000 generations, when migration rate is 0%, the turnover of viable coding genotype sections increases at a small rate due to genetic drift. Once migration is enabled, the turnover grows at a higher rate which suggests that more new viable genotypes are discovered through hybridisation than through generic drift. Indeed, this behaviour is even more pronounced for higher migration rates ([Figure 11.7, C](#)).

In the next scenario the migration rate was set to 5% after the first 100,000 generations. Qualitatively, the results are similar to the 1% scenario. Quantitatively, the gene flow between the populations is higher (F_{st} falls to ca. 0.7). The higher migration rate leads to an increased probability for formation of reproductively isolated hybrid groups ([Figure 11.7, A](#)). Genetic drift within a larger number of RI groups as well as hybridisation between more diverse individuals leads to a larger number of coding genotype sections in the population ([Figure 11.7, B](#)) and to a higher rate of discovering new viable adaptations ([Figure 11.7, C](#)). Further rises in the migration rate to 10% (not shown), 15% and 20% ([Figure 11.7](#)) increase the strength of the above effects.

When the migration rate is set to 25% or more, reproductive isolation between groups can no longer be sustained. A large number of reproduction events that lead to inviable offspring have a destabilising effect on the population size. Under such conditions, there is a high chance of extinction for any native cell population. Once an immigrant population has become established in a cell, a positive feedback loop is created: For individuals of the native population, the chance of having viable offspring is decreased by the presence of the invaders who may be selected as mating partners. At the same time, the chance of new invaders to successfully reproduce increases in a growing immigrant population. As seen in [Figure 11.7, A](#), the number of RI groups collapses to 1 under 25% migration. Sporadically, small RI groups arise due to drift, but do not persist long enough to achieve a significant divergence in neutral loci ([Figure 11.7, D](#)). The main population evolves as a single RI group. As a consequence, the number of distinct coding sections in the population is very small ([Figure 11.7, B](#) and [Figure 11.7, C](#)).

In order to investigate how spatial distance affects the above results I repeated the experiments on a 1×2 grid. In large, the model behaviour is similar, however the migration rate has a larger impact on the smaller landscape:

Readily a migration rate of 1% causes F_{st} to decrease to ca. 0.5 after 300,000 generations of migration ([Figure 11.6, C](#)). A migration rate of 10% causes the generic divergence of the two present RI groups to decrease to insignificant levels within 50,000 generations of migration. However, reproductive isolation can be sustained at 10% and 15% migration – the number of RI groups stays around 2 which shows that the significant gene flow is not sufficient to break reproductive isolation and must occur through viable hybrids, who, however, cannot establish a separate reproductively isolated population. This is apparent in the fact that the number of distinct coding sections in the population remains small (not shown) suggesting that hybrids occur between the same genotypes. This conclusion is further supported by the turnover rate of the coding sections ([Figure 11.6, B](#)): After an initial increase similar to the 2×2 scenarios, the turnover rate slows down to a level close to the rate before migration was turned on, showing that the two populations have reached an equilibrium and that further genetic innovation is due to drift. At 20% migration reproductive isolation collapses rapidly and the entire model population evolves as a single reproductive group ([Figure 11.6](#)).

Next, I repeated the experiments on a 3×3 grid. As expected, a larger grid makes it possible to sustain reproductive isolation at higher migration rates. At 30% migration, reproductive isolation is sustained and the number of RI groups lies above 40. At 35% migration, reproductive isolation collapses in a way similar to the previous scenarios (data not shown).

In order to assert that the above dynamics are caused by the HFL genetics, I simulated all of the above scenarios without the HFL. In these control runs all individuals are viable and selection is thus random. In this context, reproductive isolation cannot be defined and the number of RI groups and F_{st} cannot be measured. However, a related measure is the average genetic divergence D_B at neutral loci between all individuals of the entire model population. In the presence of groups without gene flow between them, D_B is expected to grow as the neutral loci in such groups will diverge. I measured D_B for all grid sizes discussed earlier. As expected, for a migration rate of 0%, D_B steadily grows. However, for all grid sizes, a migration rate of 1% is sufficient to cause D_B to drop sharply and to remain low for the rest of the simulation (Figure 11.8). This indicates that without HFL-genetics (or other reproductive barriers occurring in nature) reproductive isolation cannot be sustained even for small migration rates.

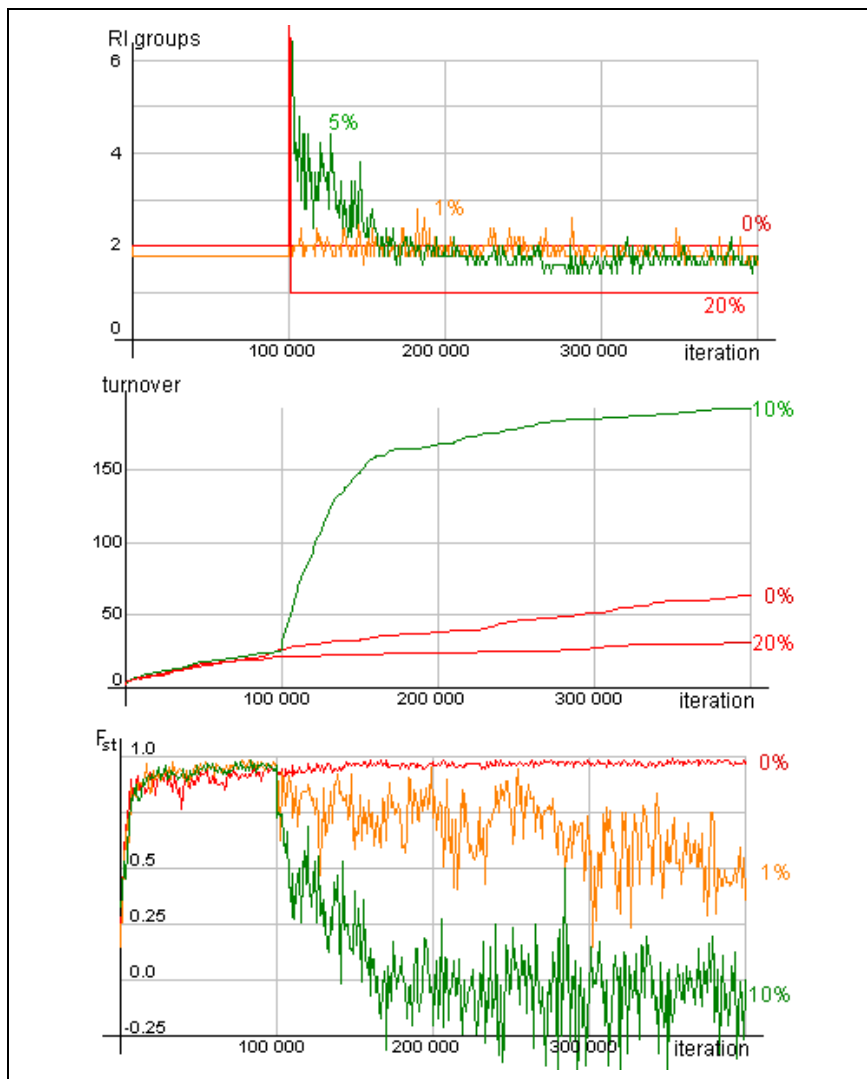


Figure 11.6. Evolution on a 1×2 grid for different migration rates. Shown migration rates: 0% (red), 1% (orange), 5% (green), 10% (green) and 20% (red). Data averaged over 10 runs. Some values omitted for clarity.

A (top): Number of RI groups.

B (middle): Turnover of viable coding genotype sections.

C (bottom): Genetic divergence measured using the F_{st} .

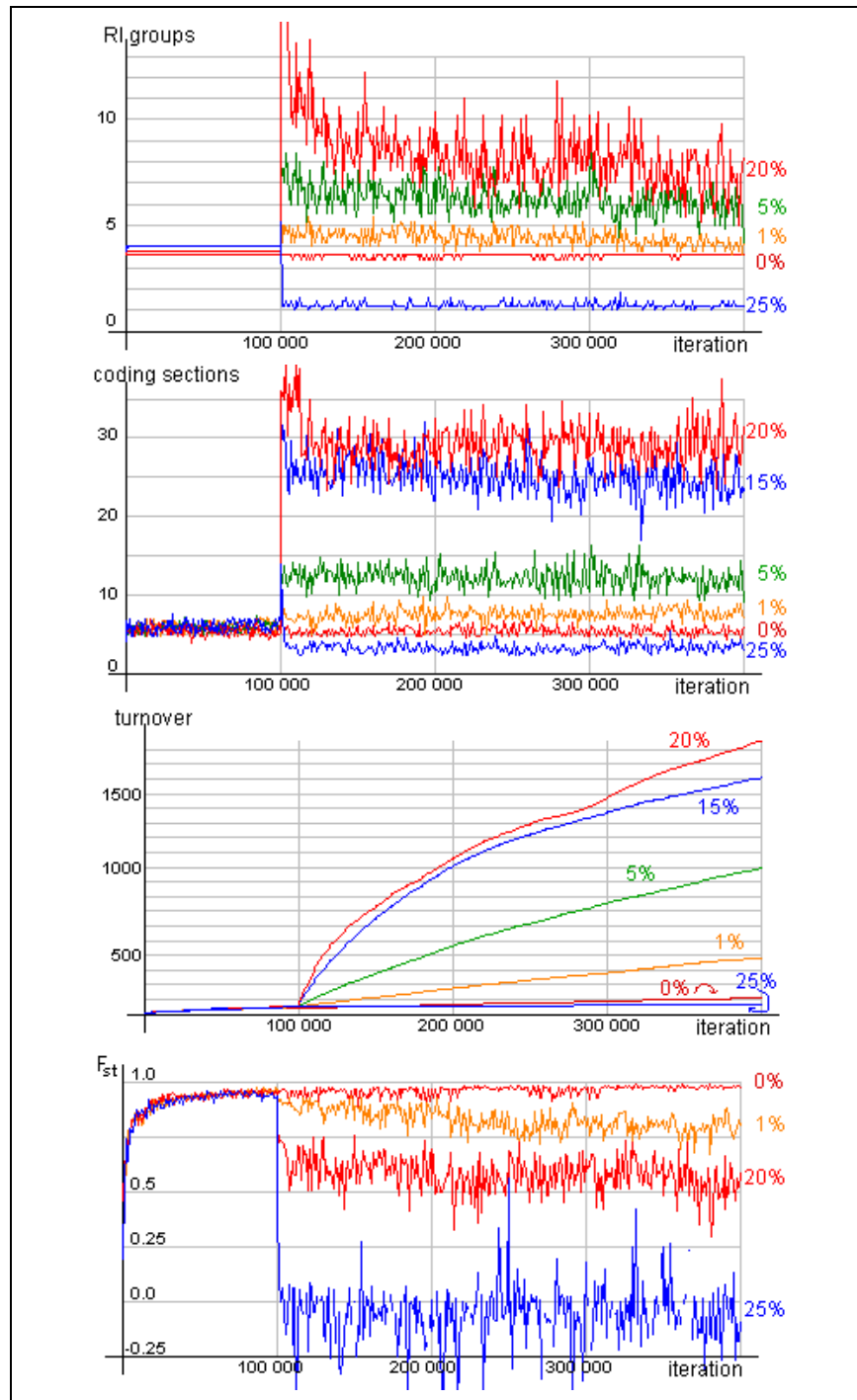


Figure 11.7. Evolution on a 2x2 grid for different migration rates. Shown migration rates: 0% (red), 1% (orange), 5% (green), 15% (blue), 20% (red) and 25% (blue). Data averaged over 10 runs. Some values omitted for clarity.

A (top): The number of RI groups increases when the migration rate is higher. For very high migration rates the whole model population collapses into a single reproductive group.

B (2nd from top): The number of distinct coding genotype sections in the population increases when the migration rate is high. As the population collapses to a single reproductive group at very high migration rates, the number of distinct coding sequences falls.

C (3rd from top): The rate of evolving new viable coding genotype sections increases when migration rate is higher due to drift in a larger number of IR groups and due to hybridisation between more RI groups. As the population collapses into a single reproductive group at very high migration rates, the number of coding sequences falls.

D (bottom): Genetic divergence between RI groups measured using the fixation index F_{st} . Higher migration rates lead to increased gene flow and thus lower genetic divergence.

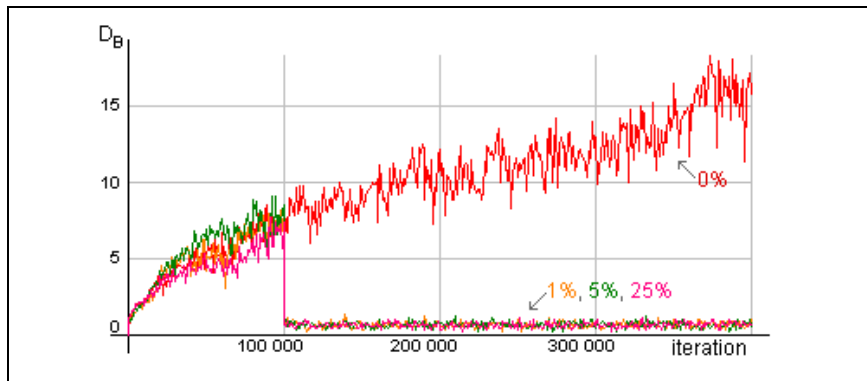


Figure 11.8. Average genetic divergence D_B at neutral loci between the individuals of the entire population in neutral evolution (without HFL). The average genetic distance grows when migration rate is 0%. The average genetic distance quickly collapses to a small value above 0 (due to drift) for all other migration rates (1%, 5%, 25% are shown). This behaviour is largely the same for all grid sizes considered (1×2 grid shown here).

11.4.4. Conclusion

The role of spatial separation in facilitating reproductive isolation is well known [Wright, 1932]. The difference between the three spatial scenarios demonstrates this effect. In order for an allele to pass from one cell to another non-adjacent cell it must first become established in the intermediate locations. Strong reproductive isolation induced by the HFL enhances this effect. Thus, hybrid zones and divergent satellite populations may provide a stronger barrier to gene flow than often assumed.

The effect of higher mutation rates on the number of distinct viable coding sections in the population is stronger than on the number of RI groups. This suggests that despite HFL, a small proportion of hybrids is viable and does not exhibit reproductive isolation from the main population. It is these viable hybrids that facilitate the gene flow between reproductively isolated populations. However, the small effect of an increasing migration rate on the number of RI groups implies that some hybrid populations exhibit real reproductive isolation and are not simply fuelled by repeated hybridisation with immigrants.

The common assumption is that hybrid zones are maintained by an interaction between continuous hybridisation and selection against hybrids. Reproductive isolation between the hybrids and the main population is often attributed to ecological preferences to a specific environment within the hybrid zone and not to genetic incompatibility. If such a specific ecological environment is altered, the hybrids become disadvantaged. As a result they become extinct either through selection against them or by adapting to the main environment thus removing reproductive isolation between the hybrids and the main population. However, hybrid populations that have strong genetic incompatibilities with the main population caused by HFL-genetics are more likely to persist. In the present simulations, such populations are short-lived because their small initial population size and the absence of prezygotic isolation (i.e., reproductive isolation caused by not mating with members of other groups rather than by offspring inviability) make it unlikely that they successfully reproduce for a large number of consecutive generations. However, in the presence of a free ecological environment niche within the hybrid zone, hybrid groups can multiply in numbers and persist. These populations, once numerous, are less likely to be affected by a disturbance of their specific ecological niche due to the strong genetic reproductive isolation between the hybrids and the main population. This can allow the hybrid population to further diverge eventually forming prezygotic reproductive isolation and thus to speciate. Although further data are required, this observation provides potential support for the analogy of novel species to point mutations implicit in some recent ecological [Hubbell, 2001] and macro-evolutionary [Gould, 2002] theories.

As discussed in [section 11.4.3 above](#), for relatively high migration rates, an immigrant (not hybrid) population that became established in a new environment is likely to induce a positive feedback loop leading to the extinction of the native population: A large number of immigrants who act as potential mating partners in the absence of prezygotic reproductive isolation decreases the chance of native inhabitants to have viable offspring and increases the chance of further invaders to successfully reproduce. This may lead to reinforcement – the evolution of sexual selection and thus prezygotic mating barriers in response to selection against hybrids ([Figure 11.9](#)). Reinforcement is a controversial topic in speciation theory [[Spencer et al., 1986](#); [Butlin, 1987](#)]. However, as argued in the previous paragraph and supported by current results, reproductive isolation generated by the HFL is often resistant to mutations reducing hybrid disadvantage. Thus, reinforcement may be more likely in the context of HFL-genetics than previous models indicate [[Spencer et al., 1986](#); [Butlin, 1987](#)].

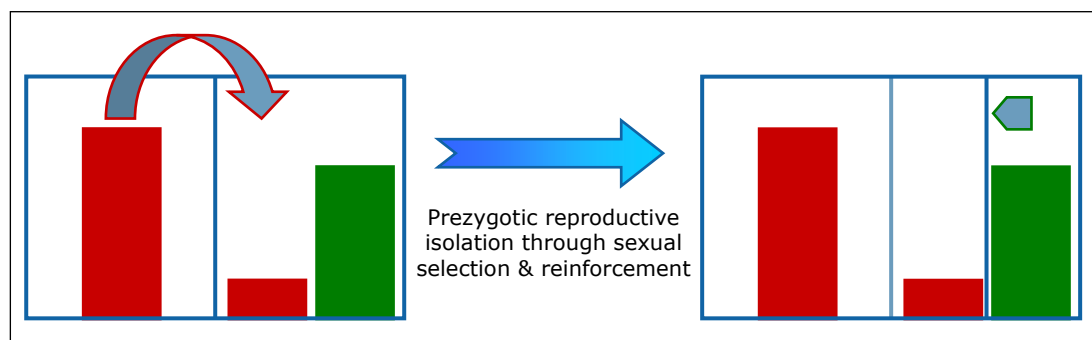


Figure 11.9. Reinforcement and evolution of sexual selection as a result of selective pressure to avoid mating with immigrants. Immigrant population becomes established in a cell. As mating is random, the chance of native individuals having viable offspring is decreased since hybrid offspring are likely inviable. Conversely, as the immigrant population grows, the invaders' chance to successfully reproduce increases. The resulting positive feedback loop is likely to drive the native population to extinction unless they develop a form of sexual selection to mate only with other native individuals. The resulting prezygotic reproductive barrier leads to speciation.

11.5. Applications outside evolutionary biology

The material presented in this section is based on the following publications:

 G. Paperin (2008):
Using Holey Fitness Landscapes to Counteract Premature Convergence in Evolutionary Algorithms.
Satellite Workshop at the Genetic and Evolutionary Computation Conference 2008 (GECCO'08). Ref: [Paperin, 2008b].

In the previous section I showed that a genotype model that results in a HFL can facilitate and maintain strong reproductive isolation and thus diversity. In biological systems, on-going diversity is a widely emerging phenomenon that we seek to understand. In some engineering problems, creating and maintaining diversity is a fundamental challenge. Here, I seek to apply insights about natural diversity gained earlier to computational optimisation problems.

A key problem in evolutionary optimisation, for instance – in genetic algorithms, is *premature convergence*. While a number of methods exist to approach this issue, they usually require problem specific calibration or only partially resolve the issue, at best by delaying premature convergence of an evolving population. In this section, I speculate that the genetic model employed in the simulation experiments in [section 11.4 above](#) can be used to combat premature convergence in evolutionary optimisation algorithms.

11.5.1. Premature convergence in genetic algorithms

Premature convergence is a common problem in genetic and other evolutionary algorithms. A number of approaches have been introduced to counter against it, however, solutions are subject of on-going research. In nature, populations often maintain a resilient genetic diversity under strong selection pressures (e.g., see throughout [part IV](#) of this thesis). Thus, various generic approaches to achieving diversity in evolutionary optimisation have been inspired by mechanisms thought to facilitate and maintenance diversity in nature. Most of these methods are variants of the so-called *niching* approach [Mahfoud, 1995] (e.g. crowding [De Jong, 1975], sharing [Goldberg & Richardson, 1987], island model [Whitley et al., 1999]). Other approaches attempt to vary environmental factors such as maximum population size [Liu et al., 2000], or the evolutionary goal [Kashtan et al., 2007], however the latter processes are not always present in diverse natural populations.

In general, niching aims to introduce a degree of reproductive isolation between groups of candidate solutions in order to concentrate the evolutionary search on different regions of the solution space. A recurring difficulty in applying niching-based algorithms is that the optimal degree of reproductive isolation and the number of reproductively isolated groups (RI groups) are usually problem-specific and must be artificially tuned or set arbitrary. In addition, while often successfully delaying or slowing premature convergence, niching algorithms are rarely successful at preventing it completely [Mahfoud, 1995]. This difficulty corresponds to a number of results from theoretical biology: In biological terms, niching introduces *prezygotic* reproductive isolation. However, the maintenance of sustained prezygotic reproductive isolation presents a theoretical challenge for biologists. To recap from [section 11.4.1](#): Prezygotic reproductive isolation based on ecological divergence or physical barriers is often transient, collapsing when selection pressures change; even moderate migration between populations leads to high gene flow making the extinction or merging of RI groups likely. On the contrary, reproductive isolation is likely to be sustained once *postzygotic* reproductive barriers have evolved and genetic incompatibilities make hybrid viability unlikely ([section 11.4](#)). Previously in this chapter I have shown that such postzygotic reproductive barriers emerge for instance when a fitness function is employed that results in a HFL.

11.5.2. Employing HFL genetics in optimisation algorithms

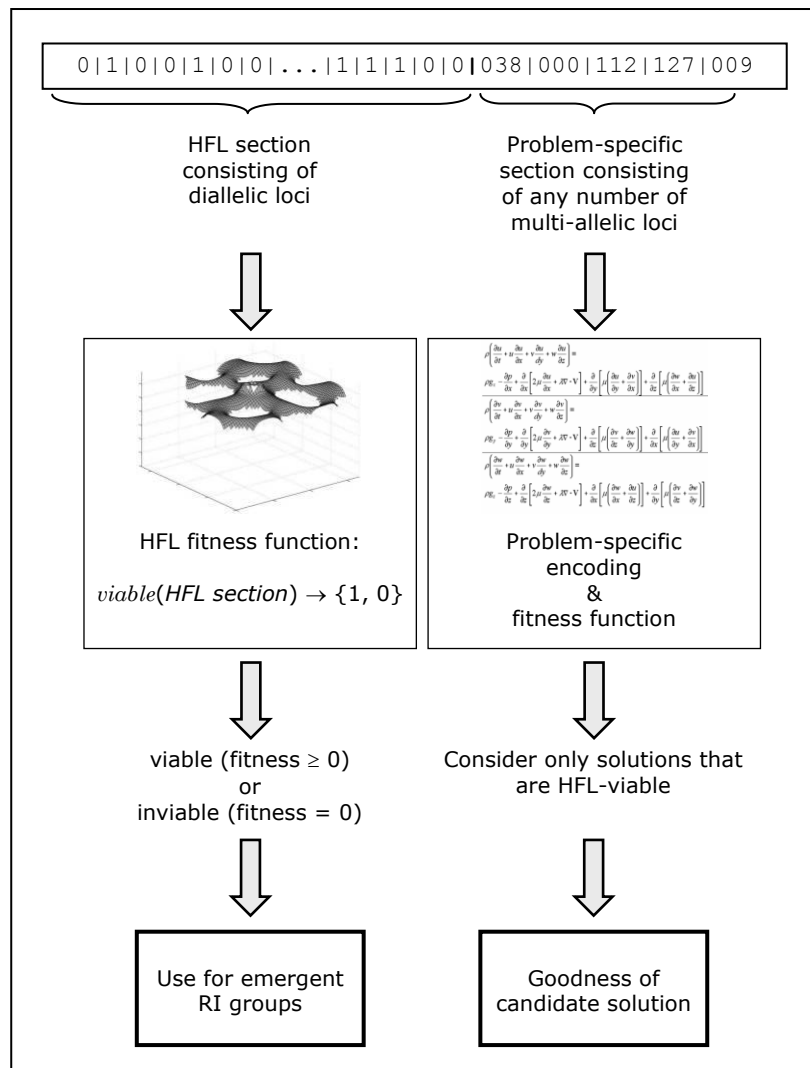
In nature the “holey” structure of HFLs arises implicitly through gene incompatibilities. In the computer models described in this chapter, HFL was modelled explicitly. This explicit model may be used to support and, in some circumstances, to replace traditional niching algorithms. In particular, the island model [Whitley et al., 1999] bears resemblance to the simulation in [section 11.4 above](#). There, I used the neutral loci only to measure genetic divergence. However, in a genetic algorithm, these loci may be used to encode candidate solutions instead ([Figure 11.10](#)). The fitness of such composite solutions may be measured as follows: If the HFL-section is inviable, the overall fitness of the composite candidate solution is zero. If the HFL-section is viable, then the overall fitness is determined by evaluating the loci that encode the candidate solution in a problem-specific way. In that case, terming these loci *neutral* would be inappropriate as they are crucial to fitness of HFL-viable candidate solutions. Reproduction of such composite individuals can be realised according to the two different genotype sections: The HFL-sections can be reproduced similar to the earlier model – through free recombination and bit-flip mutations. The problem-specific sections can be reproduced in a problem-specific way.

The small effect of an increasing migration rate on the number of RI groups observed in the earlier experiments implies that some hybrid populations exhibit real reproductive isolation and are not simply fuelled by repeated hybridisation with immigrants. In population biology, hybrid populations that have strong genetic incompatibilities with the main population (such as those caused by HFL-genetics) are thought to be most stable. In the current simulations such populations are short-lived because their small initial population size and the absence of prezygotic isolation make it unlikely that they successfully reproduce for a large number of consecutive generations. However, in the presence of a free niche, these hybrid groups can reproduce and persist. Such a niche may be given by under-explored areas of the solution space of an optimisation problem. Thus, the current model may be used as a mechanism for dynamic discovery and maintenance of multiple search directions in genetic and other evolutionary algorithms.

Because a large proportion of genotypes is HFL-inviable, advantageous adaptations in problem specific sections may be prematurely discarded when paired with inviable HFL-sections. Thus, the speed of discovery of good solutions with the proposed approach is expected to be slow. However, the candidate population is expected to exhibit strong and robust diversity and maintain multiple foci in the search space. This approach is thus best suited to problems where the speed of search is not of crucial significance and where premature convergence and local optima present important challenges. Evolution on HFLs in nature leads to diverse populations that perform robustly under unexpected disturbances and are able to adapt to unforeseen circumstances. It is this type of tasks where engineering methods often fail and where nature-inspired HFL-based methods may yield a substantial benefit.

Recall that the HFL model discussed here is explicit, while in nature, HFL arises implicitly because the majority of biochemically conceivable genotypes gives rise to inviable phenotypes. It is important to note that this is paralleled by many optimisation problems in which a valid solution encoding can result in an illegal or irrelevant solution. For instance, candidate solutions to the travelling salesman problem that contain incomplete loops or duplicate stops are inviable. Candidate solutions to multiobjective optimisation problems that conflict with one or more constraints can also be assumed inviable. In such problems, evolutionary search is carried out in an HFL-like search space, if viable genotypes (representations) build a connected component. Thus, a better understanding of structure and dynamics of HFLs may provide new insights for the solution of such problems. A step in this direction is undertaken in [section 11.3](#), where I discussed a numerical analysis of a biological niche model with HFL genetics.

Further experiments on particular optimisation problems are necessary to access the benefits of the proposed approach more specifically. Although that research is beyond the scope of this particular work, I would be remarkably excited to see further results in this area. A presentation of these proposals at a satellite-workshop of a major evolutionary computation conference [Paperin, 2008b] received significant interest from delegates.



11.6. Summary

In this section chapter I introduced a genotype model based on a combination of the quantitative trait model with ecological niches and the HFL model ([section 11.1](#)). The application of the proposed model in a computer simulation is not straight forward due to a number of constraints and differences between the analytical HFL concept and the requirements posed by feasible simulation experiments ([section 11.2.2](#)). I show an approach to overcome these difficulties and build a computer model that can be employed in individual-based simulations ([section 11.2.3](#)). On this basis I perform a structural analysis of the proposed genetic model that will inform future studies based on this work ([section 11.3](#)). The structural analysis concludes that the new model exhibits many desired nature-like properties while avoiding several drawbacks of the quantitative trait and the HFL models.

Based on the new technique for incorporating HFL-genetics into individual-based simulations, I describe a computer model that investigates the degree and the robustness of reproductive isolation provided by HFL-genetics ([section 11.4](#)). Simulation results clearly show that genetic incompatibilities caused by the HFL lead to maintenance of reproductive isolation between parapatric sub-populations despite significant migration and the absence of prezygotic mating barriers. These results may have important implications for the understanding of the role of reinforcement in evolution of sexual selection and speciation.

The techniques employed in the simulation model may be adapted for engineering problems, for instance for maintenance of diversity within populations of candidate solutions in various evolutionary algorithms ([section 11.5](#)).

The notion of HFLs, while largely unchallenged, has arguably received insufficient attention from theorists. The work presented here shows that simulating plausible FLs can considerably change predictions about the maintenance of diversity and the emergence of new adaptations in natural and artificial systems. The current approach may be useful in further exploring these issues and related problems of adaptive radiation, evolvability and evolutionary search. From the perspective of evolutionary biology, representing FLs in a more plausible way may facilitate the acceptance of individual-based techniques along with traditional, differential-equations based simulations. These techniques can help resolving many questions that are difficult to approach with conventional approaches, such as questions related to spatially-explicit scenarios, and scenarios where analytical models are not available. From the perspective of Artificial Life research, representing fitness landscapes in a biologically plausible way may facilitate ongoing adaptive exploration and the continuous generation of novelty in evolving populations. From the perspective of complexity science, representing fitness landscapes in a biologically plausible way may help understanding general mechanisms behind diversity, perpetual novelty and robustness in CAS.

Part V

Self-Organisation in Complex Networks

Topological properties of networks that underlie complex systems influence system behaviour in a fundamental way. Thus, understanding mechanisms behind dynamic and static topologies is essential for understanding complex system behaviour. Here, I discuss several network models that facilitate in emergence of key topological constructs from simple local rules. Studied dynamic and static network constructs include large-scale recurrent connectivity transitions, modularity, and scale-free networks.

Chapter 12: Emergent connectivity phase transitions in complex networks.

Connectivity phase transitions can emerge through mechanisms such as external disturbances, slow forcing and internal feedback. Simulation experiments on a model of resource flow in complex systems indicate connectivity changes facilitated by feedback between flow network properties and edge density. A social network model demonstrates recurrent non-linear transitions in connectivity and dynamical behaviour caused by slow accumulation of specific small-scale structures and intermittent releases.

Chapter 13: Generative models for important network topologies.

Modular structures, circuit, and chain structures found in real social networks, emerge according to different parameter values in a social network model. In a network model of an economy, the emergence of multi-level modules organised in complex hierarchies is facilitated by growth of certain types of local interactions. An affiliation network model of human and animal groups demonstrates a previously unstudied natural process resulting in scale-free networks of constant size.

Chapter 12:

Emergent connectivity phase transitions in complex networks

Recurrent phase transitions in connectivity density of interaction networks that underlie complex systems can emerge through different mechanisms including external disturbances, slow forcing and internal feedback. Preliminary experiments on a simulation model of resource flow in different complex systems demonstrates how connectivity phase transitions can emerge through a feedback between efficiency-properties of the resource distribution network and edge density. Another abstract social network model demonstrates how recurrent non-linear connectivity phase transitions can emerge through simple local rules that govern individual actions and relationships. In the model, slow accumulation of specific small-scale structures leads to stochastic large-scale transitions between phases of different model dynamics and connectivity.

Some of the material presented in this chapter is based on the following publications:

 G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].

 G. Paperin, S. Sadedin (2009):
The Dual Phase Evolution Framework for Understanding Evolutionary Dynamics in Complex Adaptive Systems.
2009 International Conference on Evolutionary Computation (ICEC'09). Ref: [Paperin & Sadedin, 2009a].

12.1. Introduction

Throughout this thesis I have argued that topological properties of networks that underlie complex systems play a fundamental role for the behaviour of such systems. Specifically, connectivity (density) of interaction networks is a key factor in facilitating many emergent phenomena. In this and the next chapter I am closing the loop by considering mechanisms that lead to emergence of particular important network topologies.

Previously have argued that recurrent phase transitions in connectivity of networks that underlie systems facilitate the emergence of many complex properties. In [section 7.4](#) I identified three mechanisms that can cause recurrent connectivity phase transitions: external disturbances, slow forcing, and internal feedbacks. In [part IV](#) I investigated some of the effect of the first of these mechanisms – external disturbances – in detail. In this chapter I use simulation models to provide examples of how recurrent phase changes in system connectivity can emerge without facilitation by external disturbances.

The discussion in this chapter is brief. The main objective is to demonstrate that recurrent connectivity changes *can* emerge through the above processes. In the next chapter I discuss in more detail, how such emergent connectivity transitions may be responsible for the emergence of important network topologies, such as modular and scale-free structures.

12.2. Connectivity phase transitions in an abstract resource flow network model

The material presented in this section is based on the following publication:

 G. Paperin, S. Sadedin (2009):
The Dual Phase Evolution Framework for Understanding Evolutionary Dynamics in Complex Adaptive Systems.
2009 International Conference on Evolutionary Computation (ICEC'09). Ref: [Paperin & Sadedin, 2009a].

In this section I present a model that captures some of the major features of resource flow in real-world CAS. For instance, the energy flow through food webs in ecosystems follows patterns similar to those described here. Resource flow between primary and intermediate producers, and end-consumers in economies also follows a similar pattern. Thus, the results obtained from this abstract model allow conclusions about a different real-world CAS.

In large: A system is represented as a network of *component nodes*. Such nodes require resources (here – energy) to exist. Energy enters the system through a *source node* and is distributed through directed edges that connect component nodes to the source and to each other. When several nodes are connected downstream to a node, they compete for the energy that flows out from that node. The competition incurs an *energy cost* and the remaining energy is distributed between the downstream nodes. The transfer of energy from one node to another corresponds to *energy conversion* in nature (e.g. trophic interactions in an ecosystem, asset transfer in an economy) and incurs an energy cost. Thus, every time energy flows across an edge from one node to another, some of that energy is used up. Over time, new component interactions are established, components enter and leave the system, and components that collect large amounts of energy can reproduce.

The preliminary results presented here indicate that the network structure that emerges in this model shares many parallels predicted by the DPE framework and by the framework of Adaptive Cycles (section 14.2): Over time, the network develops towards a robust configuration where a relative small number of edges can distribute the energy between the network components efficiently. Once such robust and efficient configuration is discovered, it persists for a long time, while additional relationships and interactions (edges) within the system develop. Such ever denser interactions around the stable core-configuration incur additional maintenance costs and make the resource distribution increasingly inefficient. Eventually the network reaches a state where the dense interactions can no longer be sustained, and a catastrophic collapse occurs, during which many components and resource distribution paths fail and disappear. Following such collapse, new stable system configurations are discovered and the cycle re-iterates.

12.2.1. Model details

The model is a network that consists of a single *source node* and several *component nodes* connected via directed edges that transport energy. Component nodes require energy to sustain themselves. The source node does not require energy, instead it produces a constant amount of energy at each iteration. Each model iteration consists of three stages: energy propagation, node maintenance and structural modification.

Energy propagation:

At the start of each iteration each component node c passes a proportion of its stored energy f_c along its downstream connections (i.e. outgoing edges). Total energy propagated downstream by c is $d_c = f_c \cdot (1 - r_c)$, where the retention factor $0 \leq r_c < 1$ is a random number drawn when c is created. The remaining energy ($f_c - d_c$) is retained within the node. If c has no outgoing links, all of f_c is retained. Nodes at the end of downstream edges of c compete for the energy propagated by c . Competition for resources in real systems requires energy. This is modelled by a competition cost factor $k_c = 1 / (1 + e^{2 \cdot (l_c - i_c)})$, where l_c is the number of

downstream edges from c , and $i_c > 0$ is a random number drawn when c is created, it is the maximum value of l_c such that most energy is not wasted by competition expenses. Each of the l_c downstream edges receives an equal amount of $(d_c \cdot k_c / l_c)$ units of energy from c . Any energy conversion in nature comes with a loss. To model this, every edge g has a flow efficiency value w_g associated with it, such that the amount of energy actually arriving at node c_q from node c_p is $u_{q,p} = (d_c \cdot k_c / l_c) - w_{g(p,q)}$, where $g(p, q)$ is the edge from c_p to c_q and $w_{g(p,q)}$ is a random number drawn when g is created.

Node maintenance and reproduction:

After all nodes have propagated energy downstream, the total available energy f_c at each component node c is equal to the amount of energy retained by c during the propagation stage plus the sum of the incoming energy from all upstream edges. Every c has an associated maintenance cost $m_c > 0$ selected randomly when c is created. To maintain its existence, every c expends m_c energy units per iteration. If $m_c > f_c$, then c dies and is removed from the system along with all connected up- and down-stream edges. The source node never dies. If c accumulates a large amount of energy, it reproduces. This happens by creating a duplicate copy h of c . The offspring h receives the same number of edges as c . Each of these edges may be connected either to the respective partner of c , or to any other random node with equal probability, thus modelling random mutation. The reproduction process consumes an amount of energy significantly larger than m_c and remaining energy is divided evenly between c and h .

Structural modification:

At every iteration, a new component or a new edge is introduced into the network with a small probability. When a new component c_n is introduced, for every existing node p , an edge $g(p, c_n)$ is added with a small probability. New edges connect two randomly selected existing nodes. Similarly, nodes and edges are removed from the network with a small probability at each iteration simulating external disturbances.

12.2.2. Results

Model dynamics, explored under a range of parameter values, demonstrate a DPE-behaviour with phases of high and low network connectivity. A detailed treatment of the presented model, including sensitivity analysis, and an investigation of which particular model features and parameter values facilitate the behaviour discussed here, is beyond the scope of this discussion. Preliminary results indicate that some features of the model may be simplified without compromising the DPE-behaviour discussed here. Such simplifications can make the model both, easier to analyse, and more applicable as a model of several natural CAS. That work is likely to lead to interesting insights and may result in a future publication of the findings. Here, I concentrate on three patterns in the dynamics of this model that can be observed under a range of parameter values. The key observation in the context of this thesis is the occurrence of recurrent connectivity phase transitions. In this model, such transitions are caused by internal feedback between the efficiency of the energy transportation network, and the edge density in the model (pattern C below).

Some of the indicators of network dynamics in this model are the number C of component nodes, the total amount E of energy stored by all component nodes in the system, and the network edge density D . The maximum node age B is an indicator on internal stability of the system as it shows how long nodes can persist.

A. External stress facilitates internal stability:

In the absence of structural modifications (external disturbances), when probability of random node and edge removal is zero, the number C component nodes and amount E of energy in the system are lower on average compared to cases with structural modifications. This initially unexpected result can be explained within the scope of the DPE framework: In the absence of disturbances, unfavourable configurations can only be removed through node starvation. In the presence of structural modifications that propagate through the system by cutting off nodes and reducing connectivity, the remaining network sub-structures exhibit

more efficient and robust connectivity patterns. Additionally, newly created nodes can better compete with established nodes that stored significant amounts of energy, when all nodes can be affected by disturbances equally well. This increases potential for innovation and for discovery of more stable configurations.

B. Preference for efficient connectivity patterns:

Another consistently emerging pattern is that low values of edge density D strongly correlate with high numbers C component nodes and high amounts E of energy in the system: A small number of connections is enough to efficiently distribute the energy across the components and additional edges lead to excessive energy expenditure due to unnecessary competition and flow friction along the edges (Figure 12.1).

C. Feedback between network efficiency and edge density causes connectivity phase transitions:

In typical runs with or without structural modifications, the maximum present node age B is normally low (< 1000), indicating internal instability. Over time, robust network configurations are discovered, signified by a growing value of B ($\gg 1000$). Edge density in these stable configurations grows, making them less efficient and more susceptible to catastrophic change. Eventually, the amount E of energy stored in the system reaches a very low value and the stable configurations collapse leading to the next variation phase of low connectivity (Figure 12.1). The low-connectivity variation phase is typically long, indicating that it takes many iterations until stable structures that can support growing connectivity are discovered. This may be due to the absence of higher order stabilising control mechanisms such as selection between network configurations.

12.2.3. Conclusion

The preliminary simulation results presented here indicate the presence of connectivity phase changes and cycles in system dynamics. Such behavioural cycles are predicted by DPE and the Adaptive Cycles theory (section 14.2). The current abstract model of resource flow in a network may represent a variety of CAS, however further research on the model is necessary to adequately describe the key mechanisms responsible for the emergent phase changes. Here, I used the model as an example of how recurrent connectivity phases can result from internal feedbacks between system configurations, and not be immediately caused by external disturbances. In the next section, I present another model as a further example for emergence of recurrent connectivity transitions that are not facilitated by disturbances.

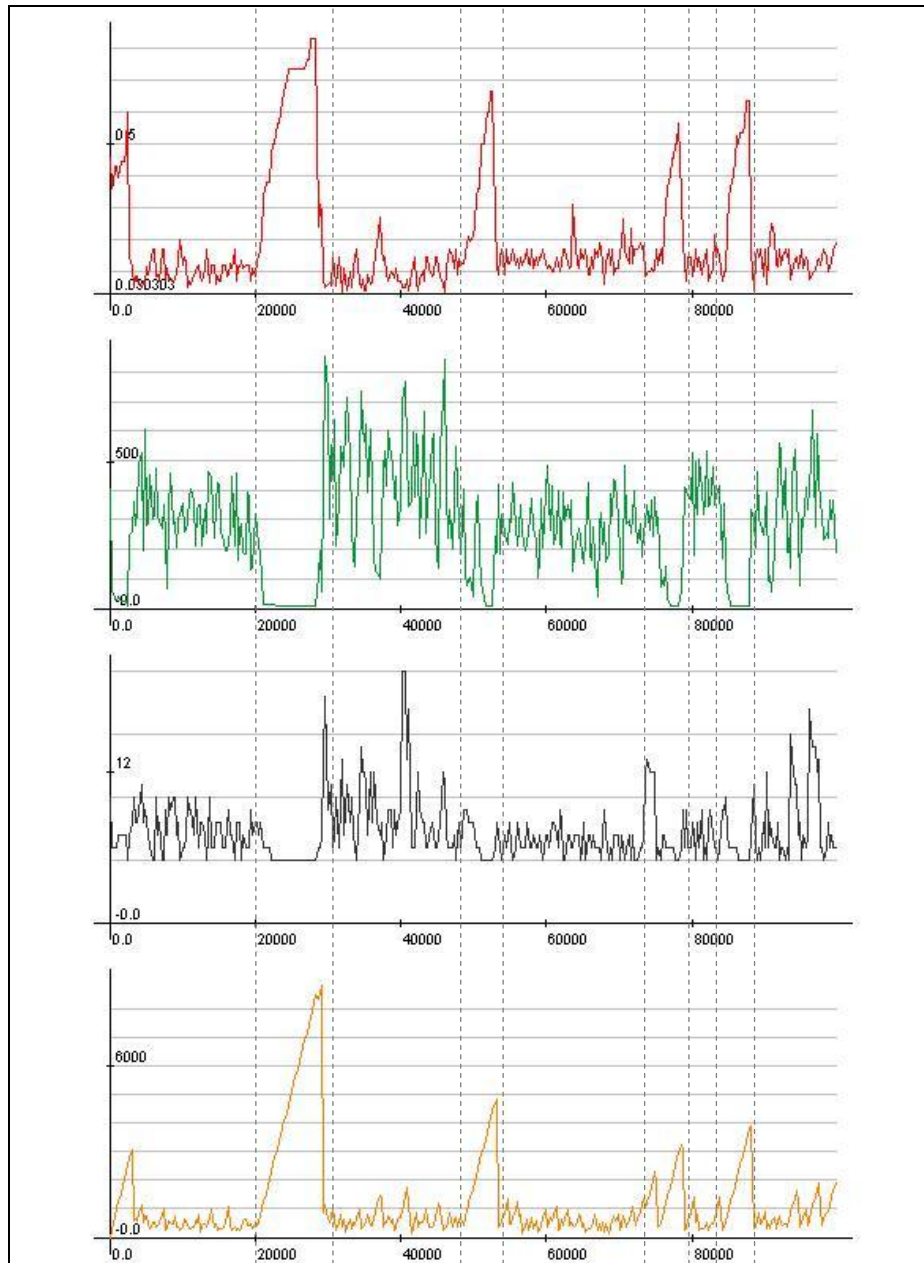


Figure 12.1. A typical simulation run. Shown are (from top to bottom): edge density D (red), total stored energy E (green), number of component nodes C (black), oldest node age B (yellow). Mean node age (not shown) strongly correlates with B . The x-axes represent iterations. The vertical dashed lines are a visual aid to stress apparent phase changes.

12.3. Connectivity phase transitions in a copycat network model

In this section I present a simple simulation model that demonstrates how recurrent transitions between phases of high and low connectivity can emerge in networks without facilitation by external disturbances. The model is based within a context of social groups. However, the simple mechanisms that drive behaviour of this model are also present in other systems, making the results applicable to a wider range of CAS.

The central idea of the current simulation is that individuals tend to behave similarly to their peers and copy their behaviour. Consider a network where nodes represent individuals and links represent friendship or acquaintance relationships. Over time, individuals forge new relationships, which is represented by a steady addition of links. At rare occasions, individuals leave the social group, for instance by moving to another city. In this model, all relationship links of individuals that leave the group are destroyed. For simplicity I assume that the group has a constant size, so whenever an individual leaves the group, a new individual joins (new individuals have initially no acquaintances). Relationships strengthen with time. Old relationships may become so important that individuals begin to copy the behaviour of peers to whom they have such important relationships. If an individual is related in this way to another individual who leaves the group, they follow and also leave. In principle, this can result in large numbers of individuals leaving at once.

Results show that the density of relationships in this model undergoes recurrent transitions between well connected and poorly connected phases. The mechanism leading to such transitions is the slow accumulation of links that cause individuals to copy each other's behaviour (*copycat links*). When a certain copycat link density is reached, the system undergoes a critical transition towards different dynamics. The system returns to original dynamics once a number of individuals leave.

12.3.1. Model details

Individuals:

Model individuals are represented by network nodes. In the experimental results reported here, the network consists of $N = 200$ nodes.

Relationships:

Common relationships between individuals are represented through *normal links* in the model. To account for the fact that individuals tend to forge new relationships over time, new links between uniformly randomly selected nodes are added to the network at a rate of one link per model iteration.

Copycat relationships:

When relationships between individuals mature, they are likely to begin copying each other's behaviour. In this model, relationship links that cause such copying are called *copycat links*. At each iteration, a link e is randomly selected. If that link is a copycat link, it remains a copycat link. If e is a normal link, it becomes a copycat link with a probability $p_c = 1 - (F_C)^{a_e}$ where a_e is the age of link e counted in model iterations, and F_C is a parameter such that $F_C = 1 - \varepsilon$ for some small ε . As a result, recently created normal links are unlikely to become copycat links, but as links become older, their probability to turn copycat increases. Below, I report on the model behaviour for different values of F_C .

Group dynamics:

At rare occasions, individuals leave the social group. To model this, nodes are removed from the network at a rate of 10^{-3} per iteration. When a node leaves the network, all its links are also removed. For simplicity I assume here that the social group is of constant size. Thus, when an individual leaves the group, another individual is added. The newly added individual has initially no relationships, but such relationships can be forged over time (see *relationships* above).

Copypat behaviour:

If an individual node that leaves the group has copypat relationships to other individuals, they also leave the group. In the model, *all* nodes that are connected to a departing node through copypat links, also leave the network. In turn, all further nodes connected by copypat relationships to any of the leaving individuals, leave as well. Thus, a large number of nodes may depart at once.

12.3.2. Results

I simulate the above model for different values of F_C . For each F_C I perform 25 independent model runs. The results are summarised in [Figure 12.2](#) and [Figure 12.3](#). For lower values of F_C (≤ 0.999), the number of links in the model network varies hectically, but remains low. This is signified by a relatively small variation in link counts over the model runs. This is due to the fact that links are likely to turn copypat soon after their creation. When nodes leave the network, large numbers of nodes are frequently drawn along, breaking their links to remaining nodes and keeping link density low.

For large values of F_C (≥ 0.9999999), the network becomes saturated with links, i.e. every node interacts with almost every other node. This behaviour is consistent across runs as shown by the small variation in link numbers. This effect occurs because links are unlikely to turn copypat even after a large number of iterations. Thus, when nodes leave the network, their interaction partners are not drawn along. Links lost due to single nodes leaving the system are quickly restored when newly arriving replacement nodes become connected to the network.

Interesting behaviour from the perspective of the current discussion occurs for intermediate values of F_C between 0.9999 and 0.9999999, with most pronounced effects occurring for $F_C = 0.9999$. There, the network undergoes recurrent transitions between phases with high and low numbers of links relative to each other (i.e. high and low connectivity density). The transitions are strong and occur stochastically, as shown by the large variation in link numbers over the model runs. As a result, the network remains in each connectivity phase for a prolonged period of time. This behaviour is due to the non-linear effects introduced by the copypat links: In a network with predominantly young links, the probability for links turning copypat is small, which allows the overall link density to increase. As links age, the rate at which they become more likely to turn copypat increases. Eventually, the average probability of links becoming copypat reaches a value for which a significant proportion of links are likely to turn copypat. This causes the system to transition to a state where a single node departure can propagate to a large number of nodes, drawing them along and thus significantly reducing link density. Both, normal and copypat links are removed during such departure events, but they are replaced with normal links of young age. Thus the model arrives back in a state dominated by young normal links and a new phase-transition cycle begins.

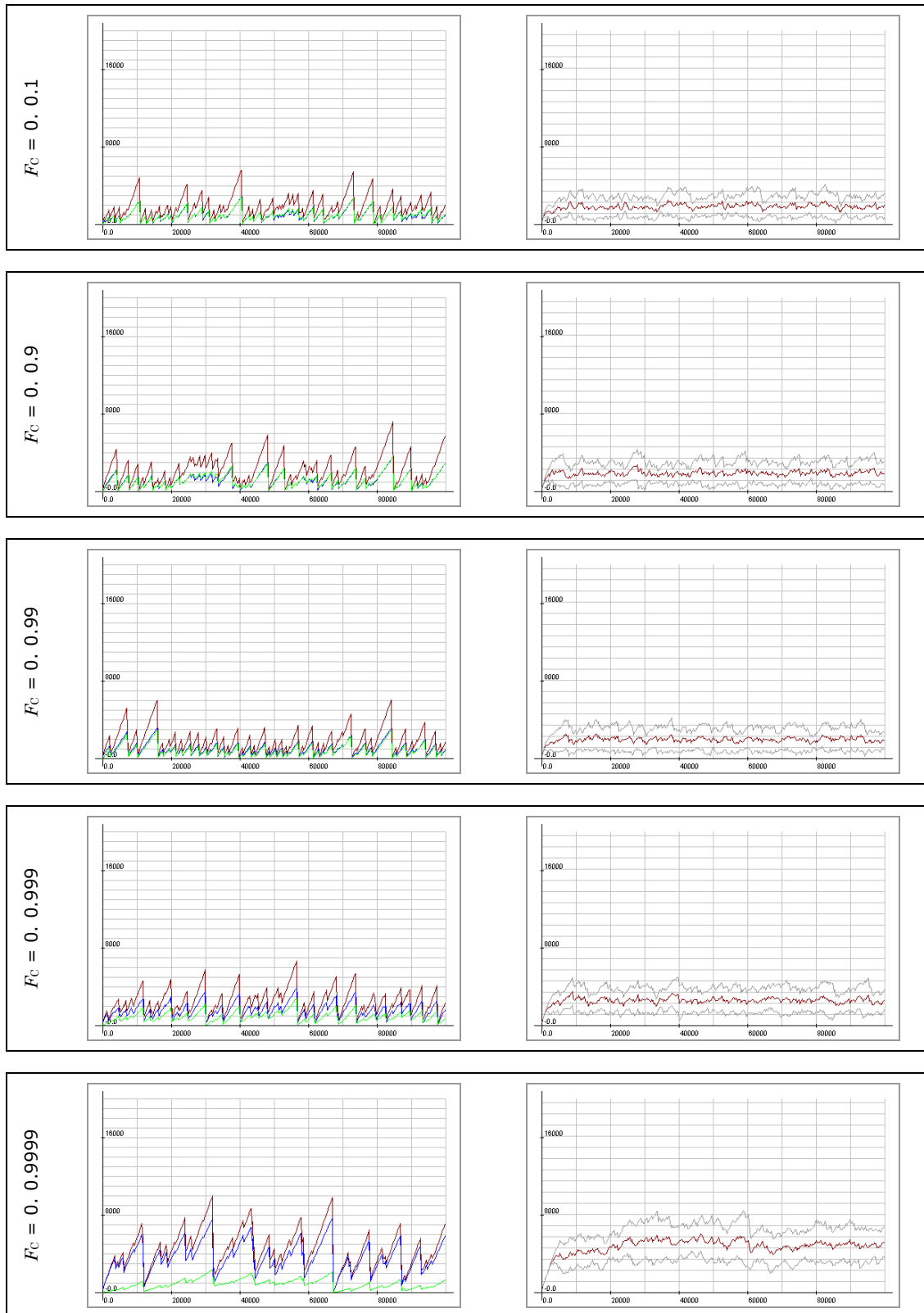


Figure 12.2. Number of edges in the copycat model (part A). Shown are typical runs (**left**) and averages over 25 independent runs (**right**) for different values of F_C between 0.1 and 0.9999. See the description of [Figure 12.3](#) on page 151 for more information.

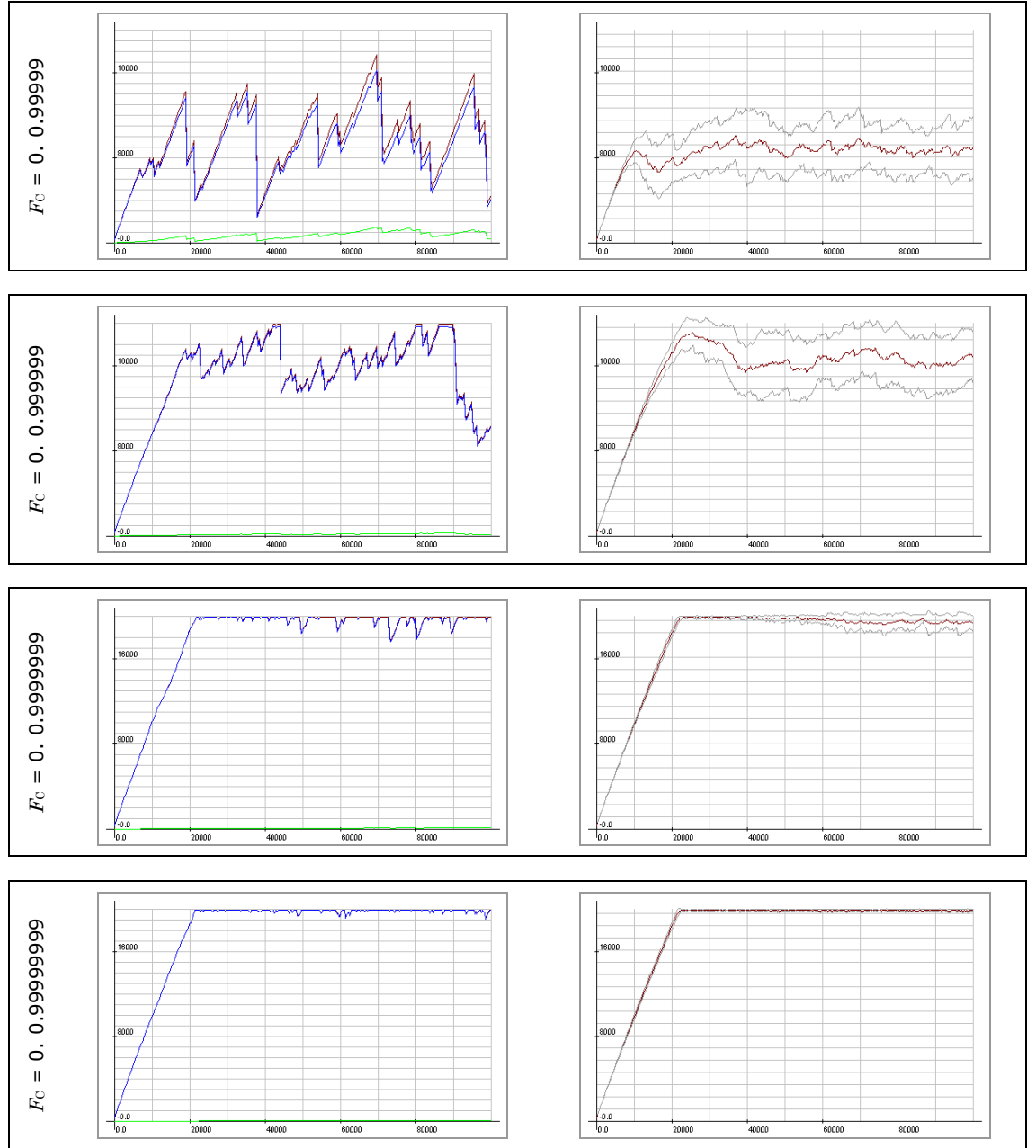


Figure 12.3. Number of edges in the copycat model (part B). Shown are typical runs (left) and averages over 25 independent runs (right) for different values of F_C between 0.99999 and 0.99999999. (This figure supplements Figure 12.2 on page 150 that shows data for F_C -values between 0.1 and 0.9999.)

Left (typical model run): Total number of links (dark red), number of normal links (blue) and number of copycat links (green).

Right (average of 25 independent model runs): Average total number of links (dark red) with grey lines showing the range of $\pm$ one standard deviation in the total number of links over the 25 runs.

For small F_C (≤ 0.999), link density / number of links remains small and the standard deviation is not significant. For large F_C (≥ 0.9999999), the network is saturated with links and the variance of link density is very small. Interesting behaviour occurs for intermediate values of F_C ($0.9999 \leq F_C \leq 0.999999$). For such F_C , the network continuously undergoes significant transitions in links density with prolonged phases with higher and lower connectivity relative to each other. This is signified by large variation in link numbers over the independent runs. This effect is particularly pronounced for $F_C = 0.99999$.

12.3.3. Conclusion

In this section I discussed a simple abstract social network model that exhibits recurrent connectivity phase transitions that emerge due to slow accumulation of particular structural elements (links of type *copycat*). This behaviour occurs only for a specific range of the F_C -parameter, however, this range spans three orders of magnitude.

The current model was motivated within the context of social groups. However, the simple principle governing the model dynamics also occur in other systems. For instance, older (in evolutionary terms) genetic interactions are thought to incur higher dependence between genes [Kauffman, 1993]. This implies that the principles exemplified here may have wider applications. A detailed discussion of parallels between the current model and particular natural systems may result in useful insights, however, it is beyond the scope of the current discussion. Here, the key objective was to demonstrate simple principles for emergence of connectivity phase transitions without facilitation by external disturbances. In the next chapter I consider a number of generative models for key network topologies based on both, emergent and explicitly facilitated transitions between high and low interaction densities.

Chapter 13:

Generative models for important network topologies

Large-scale network topology influences systems behaviour. Thus, understanding generative mechanisms for important topological patterns is essential for understanding system dynamics. Several of such generative models are considered. A social network model shows that both, modular structures, and circuit and chain structures found in real social networks, emerge according to different parameter values. A network model of an economy demonstrates the emergence of multi-level modules organised in complex hierarchies. A model of interactions in human and animal groups demonstrates a hitherto unknown, natural generative mechanism for scale-free networks that do not grow in size.

Some of the material presented in this chapter is based on the following publications:

-  G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
-  G. Paperin, D. G. Green, T. G. Leishman (2008):
Dual Phase Evolution and Self-Organisation in Networks.
7th International Conference on Simulated Evolution And Learning (SEAL'08). Ref: [Paperin, Green, et al., 2008].
-  T. G. Leishman, D. G. Green, G. Paperin (2007):
Dual Phase Evolution – a mechanism for self-organisation and optimisation.
Asia-Pacific Symposium on Intelligent and Evolutionary Systems 2007 (APSIES'07). Ref: [Leishman et al., 2007].
-  G. Paperin (2008):
Dual Phase Evolution in Natural and Artificial Computational Systems.
Invited talk at the 12th Asia-Pacific Symposium on Intelligent and Evolutionary Systems (IES'08). Ref: [Paperin, 2008a].

13.1. Introduction

Previously I have argued at several occasions that topological properties of networks underlying complex systems are a fundamental factor in facilitating many emergent phenomena. In this chapter I consider generative mechanisms that facilitate the emergence of important large-scale network topologies. I thereby continue the discussion began in the previous chapter, where I used simulation models to demonstrate mechanisms leading to recurrent connectivity phase transitions.

In this chapter, I seek to apply the DPE framework where possible by keeping the role of local and global interactions and the role of different connectivity phases in focus.

Several large-scale network topologies have received particular attention from researchers over the last decade. Recall briefly the review in [chapter 5](#):

- **Random networks:** The “classic” network model is obtained by adding edges at random to a set of nodes. Much of the intuition and many analytical results in network theory are obtained from this model, however, random networks (also: Poisson graphs) are not thought to be a good model of most naturally occurring networks, which exhibit other edge topologies.
- **Trees:** A network that has no loops or circuits is a tree. Trees are naturally associated with hierarchies and taxonomies. Tree structures are ubiquitous in human analytical thinking; they lie as the basis of concepts as diverse as biological phylogenetics, military command structures, linguistic grammars and the organisation of XML and HTML documents on the web. However, tree structures are not often found in natural

interaction networks. Complex systems rarely adhere to strict taxonomic rules and although modular organisations are common, loops and feedbacks occur frequently.

- **Modular networks:** A modular network consist of groups of nodes (modules) where the internal connectivity is significantly higher than the connectivity between the module and the rest of the network (Figure 5.5, C). Modular structures are a common way by which natural and artificial systems reduce complexity through functional encapsulation and decoupling. The abundance of modular topologies in complex natural networks implies that mechanisms behind self-organisation into modular structures must be common. However, such mechanisms are not well understood. Section 13.2 discusses some possible mechanisms behind the emergence of modular network topologies.
- **Scale-free networks:** Scale-free networks consist of many nodes with a few edges and a few nodes with many edges (Figure 5.5, A). More precisely, the node degree distribution in scale-free networks follows a power-law. Scale-free interaction topologies have been discovered in a variety of systems as diverse as scientific collaborations, the World Wide Web and inter-protein interactions in cells. Scale-free networks are thought to play a role in systems that exhibit self-organised criticality (section 14.1), they have been extensively studied, and several generative mechanisms are known. However, in section 13.3 I discuss hitherto unknown generative models for scale-free networks that can produce some scale-free structures that do not arise from other known models.
- **Small-world networks:** Networks exhibit the small-world property if the length of the shortest paths between nodes increases slowly as the network size grows. A typical small-world network is dominated by local links with a few long-range connections (Figure 5.5, B), however, other topologies, including scale-free as well as random Poisson networks, typically exhibit this property. Examples range from many biological within-cell interactions to modern data-communication networks to the famous *six degrees of separation* principle that implies that any two people on earth (presently alive) are connected to each other through a chain of acquaintances that is not longer than 6 steps. However, not all natural networks have small-world topologies. For instance, networks where connectivity is constrained by the location of components in physical space or it time are often not small-world: E.g., the six degrees of separation principle does not apply to individuals that lived during different periods. Gene flow connectivity of stationary plants with small seed dispersal range relies on close proximity. Pre-industrial communication networks were constrained by line-of sight distance between repeater stations (early telegraphs, Indian or Chinese smoke signals) or by the need to physically carry written messages (on foot or on horse).
Small-world networks can easily arise through DPE processes through recurrent transitions between locally constrained and global links-building interactions. For a simple thought experiment, recall the terminology for primary and secondary interactions from chapter 5. Consider a system whose primary structure arises as a result of some secondary interactions (e.g. friendship relationship structure (primary) arising from communicative interactions (secondary)). In such system, the distance (locality) of secondary interactions can be described in terms of the primary structure (e.g. friends communicate – local interaction, non-fiends communicate – global interaction). Recurrent transitions between local and global interaction phases in such system will result in a primary network topology with local links supplemented with some global links (local phase, low connectivity – friends talk to friends only; global phase, high connectivity – everyone talks to anyone; result: new members are added to a person’s friendship network leading to a small-world network of friendships).

In the following sections I discuss some generative mechanisms for modular and scale-free networks.

13.2. Modular networks

Some of the material presented in this section is based on the following publications:

-  G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
-  G. Paperin, D. G. Green, T. G. Leishman (2008):
Dual Phase Evolution and Self-Organisation in Networks.
7th International Conference on Simulated Evolution And Learning (SEAL'08). Ref: [Paperin, Green, et al., 2008].
-  T. G. Leishman, D. G. Green, G. Paperin (2007):
Dual Phase Evolution – a mechanism for self-organisation and optimisation.
Asia-Pacific Symposium on Intelligent and Evolutionary Systems 2007 (APSIES'07). Ref: [Leishman et al., 2007].
-  G. Paperin (2008):
Dual Phase Evolution in Natural and Artificial Computational Systems.
Invited talk at the 12th Asia-Pacific Symposium on Intelligent and Evolutionary Systems (IES'08). Ref: [Paperin, 2008a].

In this section I discuss generative models that can lead to emergence of modular network structures. The proposed models are informed by the DPE framework – they are based on a succession of phases dominated by local or by global interactions. The models discussed here act on abstract networks, however, I show how the described processes correspond to processes that occur in natural biological or social systems.

13.2.1. Social network structures resulting from explicit dual phase processes

In this sub-section I discuss a simple abstract model of a social network. Analysis shows that for different parameters, the model produces different patterns observed in real social networks.

The main process captured by the model in this section is the succession of interactions within the immediate social field, and interactions with larger groups that occur for most people. Normally, people interact with other people with whom they are already acquainted or with likeminded people. Mutual relationships are strengthened through these interactions. On rare occasions, such as social or business meetings or conventions, people interact with anyone who they encounter during the event, thus forming new links. When a personal relationship is not maintained through repeated contact, it slowly decays and eventually disappears.

13.2.1.1. Model details

The above process can be described by an abstract network model, where nodes represent people and links represent friendship or acquaintanceship relations.

Individuals:

As the focus is on the network organisation, the representation of individuals (nodes) is as simple as possible. Each node possesses a set of L attributes, with each attribute having A possible discrete states.

Relationships:

Social relationships (edges) between individuals have a weight $w \in \mathbb{R}$, indicating the strength of the relationship. Each interaction reinforces the connection by an amount Δ_+ ; but when not reinforced, the strength of a connection decreases by an amount Δ_- at each time step. Connections are broken when their weight falls to zero. When two unconnected individuals interact, a link with an initial weight w_0 is created between these individuals.

Global phase:

Individuals interact by a series of encounters between pairs of individuals. During the global phase all individuals interact randomly. Per iteration of the global phase, each individual is chosen once as the 1st partner, and a 2nd interaction partner is chosen randomly and independently from all other individuals. The global phase occurs at a rate of f_G iterations.

Local phase:

During this phase the individuals interact with other individuals to whom they are connected or with the individuals connected to their neighbours. In addition, individuals also interact with other unlinked individuals that have similar attribute values. The local phase occurs at a rate of f_L iterations.

Experiments:

The model is run with the parameters listed in Table 13.1 for 3 different scenarios: local interactions only, global interactions only, global and local interaction phases as defined by f_G and f_L .

Parameter	Notation	Value
Network size in nodes	N	50
Initial edge density	D_0	0.1
Frequency of global interactions	f_G	0.05
Frequency of local interactions	f_L	0.95
Number of node attributes	L	1
Number of distinct values per attribute	A	5
Weight for newly created edges	w_0	5
Weight increase for reinforces relationships	Δ_+	5
Weight decrease for not reinforced relationships	Δ_-	1

Table 13.1. Parameter values used in the experiments in section 13.2.1 resulting in modular network topologies.

13.2.1.2. Results

The results of the above experiments show very distinct outcomes from one scenario to another (Figure 13.1).

In networks with local interactions only, existing connections rapidly disappear. Only connections between agents with the same attributes remain, and no new connections can be formed between individuals with different attribute states. As a result, connectivity remains low and network components held together by the same attribute value remain disconnected. The network falls into a stable structure in which it remains.

In networks with global interactions only, new connections form continually, so the edge density increases rapidly. As connections are formed faster than they decay, the network eventually becomes fully connected.

In contrast, the DPE scenario of recurring local and global interactions leads to formation of modules in an initially random network. The continuous reinforcement of local links and the decay of intra-modular links during local phases maintain a stable structure within the modules. Simultaneously, modules become partially or completely disconnected from each other. During the subsequent global phase the modules become connected in new ways forming novel and complex structures ().

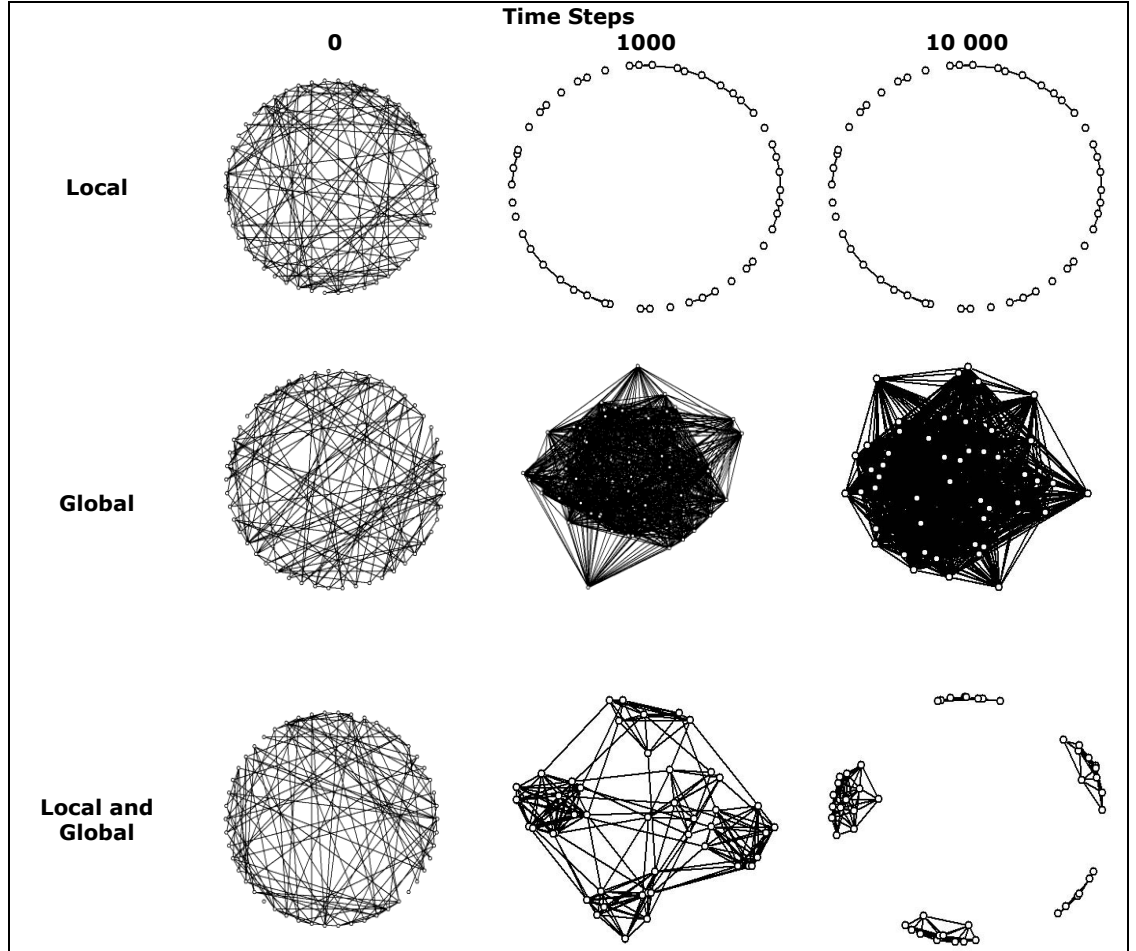


Figure 13.1. The effect of local and global interactions on the formation of network structure. Each row shows the evolution of a typical network under different conditions. In the top row, individuals interact only locally, with individuals to whom they are already connected. In the centre row, all interactions are randomly selected from the entire network. In the bottom row, individuals normally interact locally, but global interaction events occur at regular intervals.

Sensitivity analysis:

To investigate the effect of the assumptions made about the parameter values, sensitivity analysis on two parameters is performed: the frequency of global interaction phases and the number of different values that can be assigned to the node attribute.

The analyses of both parameters investigate the effect on modularity, measured by compartmentalisation as suggested by Pimm [1982]. This measure M' is defined as the average of the compartmentalisation ratio $S_{j,k}$ that is described for each pair of nodes j and k in a network by the quotient of the number of neighbours j and k share in common over the total number of neighbours of j and k :

$$M' = \frac{1}{N \cdot (N-1)} \sum_{k,j=1; j \neq k}^N (S_{j,k})$$

where

$$S_{j,k} = \frac{|\mathcal{N}_j \cap \mathcal{N}_k|}{|\mathcal{N}_j| + |\mathcal{N}_k| - |\mathcal{N}_j \cap \mathcal{N}_k|}$$

and $\mathcal{N}_x$ is the set of neighbours of a node x . The value of M' ranges between 0 and 1 with higher values implying higher modularity.

The analysis shows that the modularity M' of the networks produced is greater when global phase events are more frequent (Figure 13.3, Left). Over the range tested, M' decreases rapidly at first, as the interval between events increases, then in approximately linear fashion. The presence of high M' values for small intervals between global events is a result of high edge density. The smaller rate of decrease for greater intervals reflects the shifting balance between edge formation and edge removal during the two phases.

Increases in the number of attribute states led to a decrease in modularity (Figure 13.3, Right), but the decrease was at a much faster rate, following a power law within the tested range.

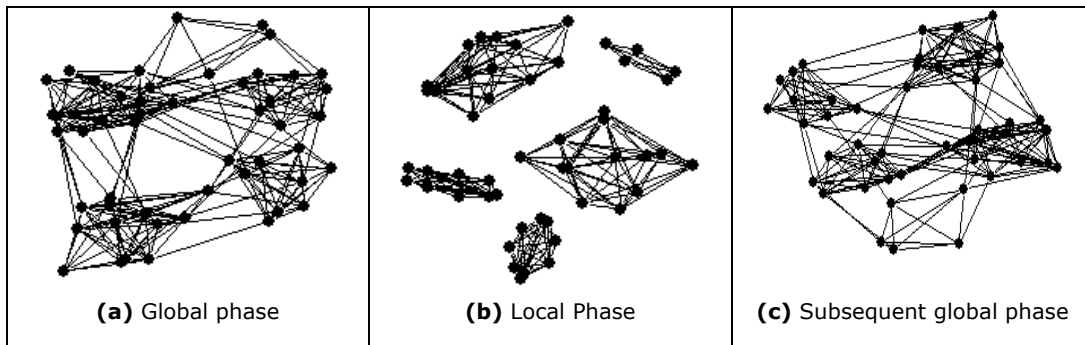


Figure 13.2. DPE process leading to a modular network topology. The continuous reinforcement of neighbourhood edges and the decay of intra-modular edges during the local phase maintain a stable local structure and may cause the modules to become poorly connected or disconnected from each other over time. During the global interaction phases the modules become connected in new ways.

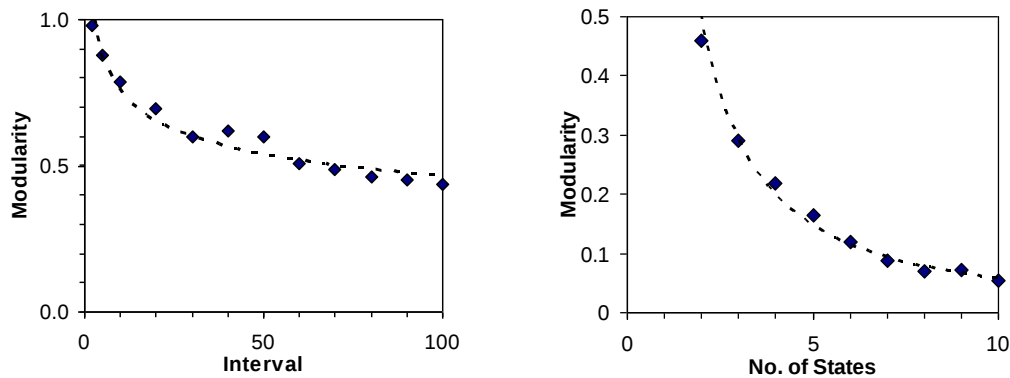


Figure 13.3. Sensitivity analysis of modularity in the network after 5000 simulation iterations.

Left: Decrease in network modularity M' as the interval between global events increases. The interval directly relates to the rate f_G of occurrence of a global phase, e.g.: $f_G = 0.05 = 1/20 \Rightarrow \text{interval} = 20$ iterations; $f_G = 0.02 = 1/50 \Rightarrow \text{interval} = 50$ iterations.

Right: Decrease in modularity as the number of attribute states (possible attribute values) increases.

Parameter	Notation	Value
Network size in nodes	N	50
Initial edge density	D_0	0.1
Frequency of global interactions	f_G	0.05
Frequency of local interactions	f_L	0.95
Number of node attributes	L	5 *
Number of distinct values per attribute	A	1 *
Weight for newly created edges	w_0	10 *
Weight increase for reinforces relationships	Δ_+	10 *
Weight decrease for not reinforced relationships	Δ_-	1

Table 13.2. Parameter values used in the experiments in section 13.2.1 resulting in circuit and chain-dominated network topologies. Compare these values to Table 13.1. Values that differ in this table are highlighted.

13.2.1.3. Parameter configurations leading to other network topologies

Seeking to investigate the sensitivity of the network towards the number and the type of node attributes, the experiments are repeated with the values described in Table 13.2. The main change is that instead of a single attribute that can assume several values, several binary attributes are now used.

The resulting networks show a different topology. They are dominated by chain, tree and circuit sub-structures. The common features between the resulting networks are difficult to quantify, however, a visual inspection makes the similarities apparent (Figure 13.4). There is an intriguing parallel between the networks resulting from this experiment and some social networks occurring in real societies. Recently, a research paper [Bearman et al., 2004] attracted much attention, in which the romantic relationships between students at a US high school were mapped and represented as a network. The similarity between that network and the current simulation results is striking (Figure 13.5). The unexpected emergence of this similarity patterns suggests that the current simple and abstract model captures some essential aspects of the formation of some real social networks.

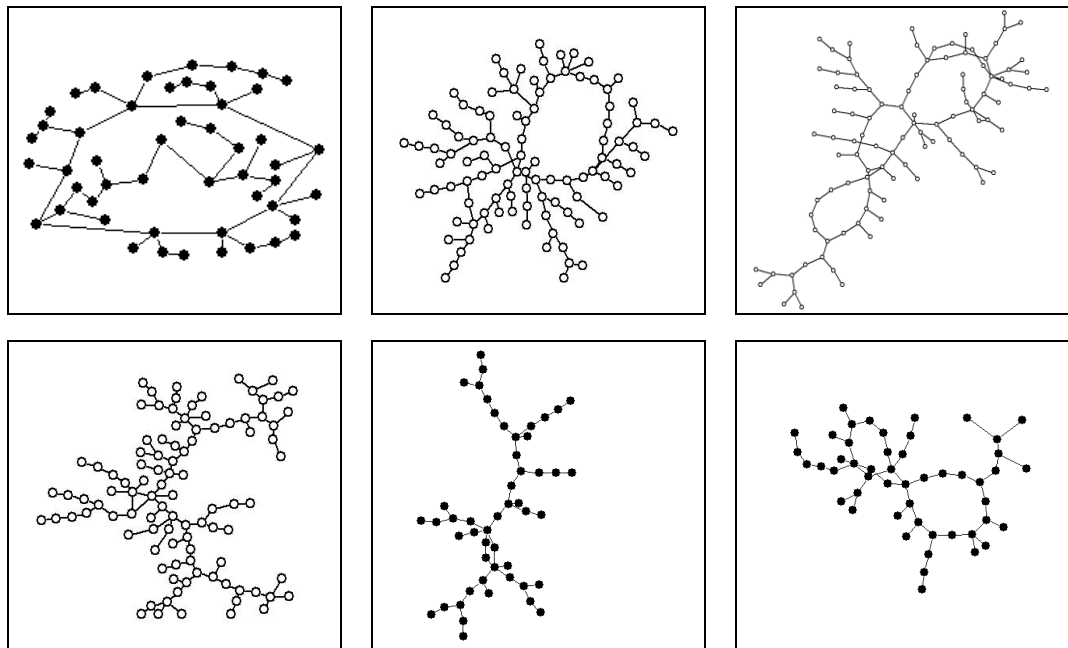


Figure 13.4. Typical networks resulting from the model with several binary attributes per node.

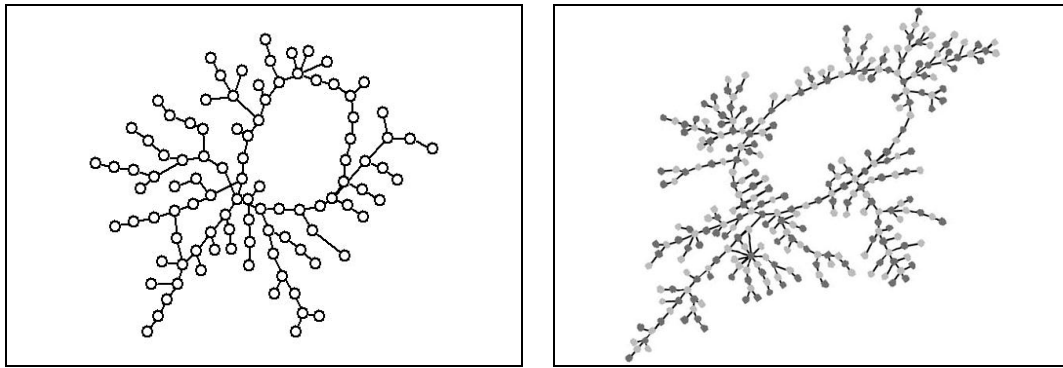


Figure 13.5. Typical network resulting from the simulation model (left) and an empirical social network of romantic relationships in a high school (right). The similarity of the network is intriguing. The right figure is reproduced from Bearman et al. [2004] with permission by Prof. P. S. Bearman, Columbia University, USA.

13.2.2. Modular networks resulting from implicit dual phase processes

The model presented in the previous section was based on an explicit dual phase process: The mediation between the global and local interaction phases was enforced by a global controller, in that case – the model algorithm. In this section, I consider another generative model for modular networks. In this model, both, the recurrent fluctuations in connectivity density, as well as the modular topology, emerge from local interaction rules.

Recall the discussion of the DPE framework (chapter 7). According to DPE, underlying system networks flip between selection-dominated balance phases (high connectivity) and variation-dominated exploration phases (low connectivity). Subsystems that remain internally stable during a variation phase – either by chance or though some stabilising force – can form new interactions to other subsystems and assume new roles in a changing system. In this way, they become functional components during subsequent variation phases. This can lead to the emergence of modular structures. When emerging functional components already exhibit modular structure internally, the above mechanism can lead to hierarchical organisation.

The model presented here demonstrates the above mechanism. Consider a network of interacting components, for instance agents in an economy (e.g. companies, people, organisations), connected by edges representing business relations or trade co-operations. Organisations forge co-operations at random. At occasions, particular organisations are struck by crises, resulting in a loss of co-operation links. Such crises can propagate through the system affecting all connected organisations. However some co-operative links between organisations with particular common interests can withstand crises, and the more crises a link survives, the stronger it becomes. Results show that over time, stable co-operations self-organise into well connected market segments (modules) with weaker links to other segments. In detail:

Agents:

The model consists of N agents (business organisations) represented by nodes. Every node has an attribute that can assume one of A discrete values (e.g. a particular attribute value can represent a type of particular trade interests an organisation has). In the basis scenario, attributes can assume one of $A = 5$ different values.

Relationships:

Relationships (business dependencies or co-operations) are represented by links (edges). At each discrete model iteration, a new link is forged between two randomly selected nodes.

Crises:

Organisations can be struck by crises. Crises occur at a rate $C_R = 0.05$ per iteration. When a crisis occurs, a node is randomly selected and all the links between that node and its neighbours are removed. Unless certain conditions occur (see below under *crisis weathering*), the crisis also affects all of the node's neighbours who also lose their links, and, in turn, their neighbours, and so on. This way, a crisis can propagate through large parts of the network recursively removing links.

Crisis weathering by links:

If organisations have strong mutual business interests, a co-operative link may withstand a crisis. In the model, if two nodes share the same attribute value, the link between the two nodes may withstand a crisis with probability $s = 1 - d_I$ (where $d_I = 0.9$ is the "link death probability"). If a link survives, it is not removed and the crisis does not propagate through that link. Relationships that weather crises become stronger: if a link survives a crisis, the probability s that it will survive future crises grows: $s = d_I \cdot (d_F)^C$, where C is the number of crises survived by the link and $d_F = 0.75$, regardless of the respective node attribute values.

13.2.2.1. Basis scenario

In the basis scenario I perform 25 independent runs of the above model, each run lasts for 1,000,000 iterations. The parameters used are summarised in [Table 13.3](#).

Through continuous addition of new links, the network connectivity grows steadily with intermittent collapses caused by propagating crises. However, on contrary to initial intuition, the edge density does not settle close to the critical percolation density. Instead, it fluctuates with an amplitude of several hundred edges ([Figure 13.6, top](#)). This effect can probably be accounted for through link fixation (see also below).

Over time, some links survive crises co-incidentally and their stability is increased. As a result, the crisis propagation becomes more constrained and the edge density increases. However, the magnitude of density fluctuations remains high ([Figure 13.6, bottom](#)). This shows that the emergent modularity (see discussion below) is not merely due to link fixation.

Modularity in this model can be measured as follows: Let $\mathcal{M}_a$ be the set of all nodes that have the attribute value a . Then, modularity M is the proportion of all network edges that begin and end within the same set $\mathcal{M}_a$, for all $a = 1 \dots A$, out of all edges:

$$M = \frac{\sum_{e \in \mathcal{E}} \text{inner}(e)}{|\mathcal{E}|}$$

where $\mathcal{E}$ is the set of all edges, $|\mathcal{E}|$ is the number of all edges, and $\text{inner}(e) = 1$ iff both nodes connected by $e \in \mathcal{E}$ have the same attribute value and 0 otherwise. In a random network, the expected value of M is $1/A$; in this case, $M = 1/5 = 0.2$ is expected. Higher values of M imply a modular network. The value $M = 1$ shows that the network modules are not connected to each other, as there are no cross-module links.

(Note that the above estimation M in a random network applies in cases where the expected module size is the same for all modules. In a general case, M must be computed for each module separately, and combined in an average that is weighted by the module sizes relative to the size of the entire network. However, here, the modules are equally large on average, I thus choose the simpler approach. Moreover, the value $M = 1$ also occurs when the network consists of a single module. However, as $A > 1$, I disregard this special case.)

[Figure 13.7 \(left\)](#) shows the average modularity obtained over 25 independent model runs. High M values show that modular structures emerge within the model. These structures are apparent and can be recognised by a simple visual inspection ([Figure 13.7, right](#)).

The modules reflect groups of organisations that share common interests (same attribute values). However, [Figure 13.7 \(right\)](#) shows that the resulting structure of the network is not constrained to segregation of nodes according to attribute values. Instead, organisations self-organise into a complex modular networks with dense interactions within market segments, but also with significant interactions across segments. In [section 13.2.2.3](#) I consider variations on this model to investigate how changes of link-fixation rules affect predictions about the topology of the emergent economy-network.

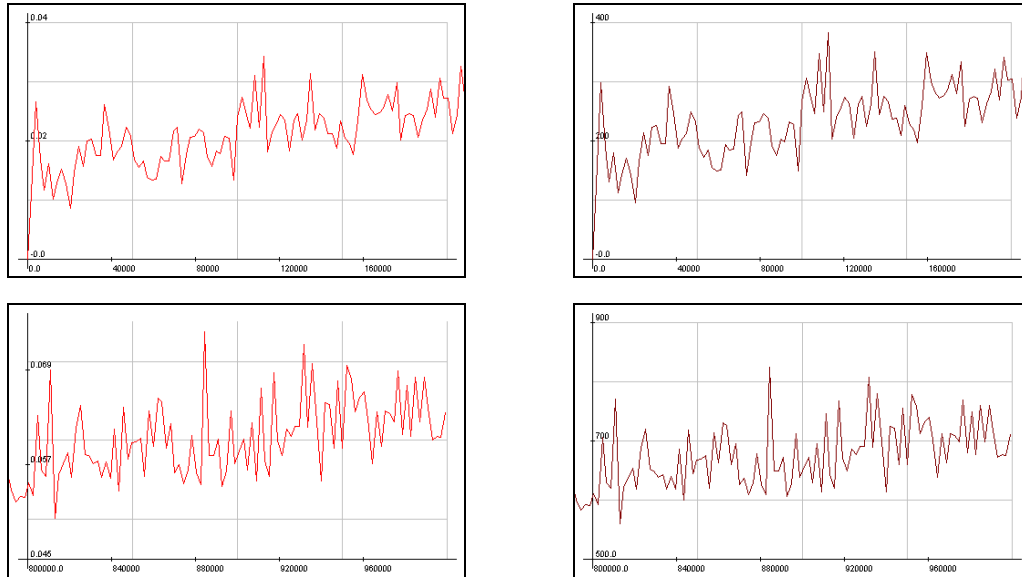


Figure 13.6. Edge density and number of nodes. Shown is the edge density (left, red) and the number of edges (right, dark red) for the first 200,000 (**top**) and the last 200,000 (**bottom**) iterations of a typical model run with 1,000,000 total iterations. Over the runtime of the model, the edge density increases approximately by a factor of 3. Fluctuations of several hundred edges occur throughout the entire model run.

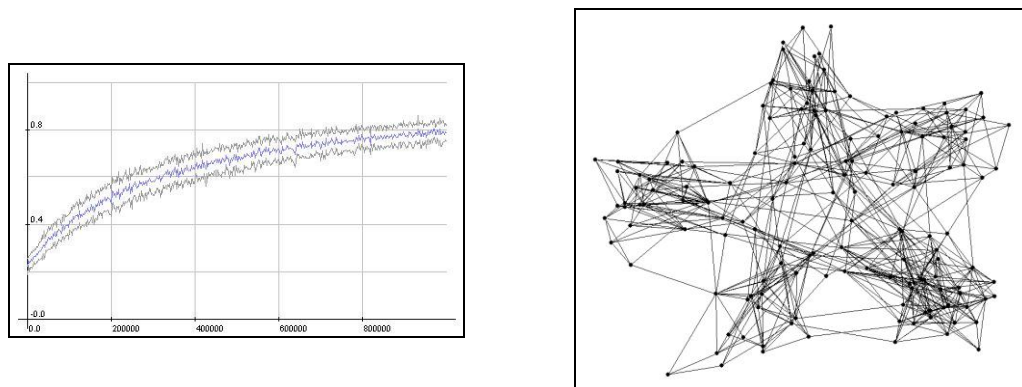


Figure 13.7. Emergent modularity in model network (basis scenario). Shown is the modularity throughout a model run (left), averaged over 25 independent model runs (the grey lines show the range of $\pm$ one standard deviation in modularity over the 25 runs). On the right: a typical network resulting at the end of a model run, automatically laid out with the algorithm introduced by Fruchterman and Reingold [1991], as implemented in the Pajek software [Batagelj & Mrvar, 1996].

Parameter	Notation	Value
Number of nodes	N	150
Number of attributes per node	L	1
Number of values per attribute	A	5
Initial edge density	D_0	0
Crises frequency	C_R	0.05
Probability of edge "death"	d_I	0.9
Edge stability increase factor base after survivals	d_F	0.75

Table 13.3. Parameter values used in the basis scenario experiments in section 13.2.2.

13.2.2.2. Sensitivity analysis

The model is not sensitive to specific attribute values. In particular, sensitivity tests reveal:

Number of nodes:

A network size $N > 30$ has no significant effect on model behaviour. (However, it influences the runtime of the simulation.) Very small values of N do not provide enough nodes, and thus links, for statistically representative measurements of modularity.

Number of modules:

Due to the way M is defined above, the number of modules corresponds to the number A of different attribute values. Varying the number of A between 2 and 15 does not affect the model behaviour significantly. It is expected, that if N is increased to allow a significant number of nodes per module, higher values of A will not affect the results either.

Initial edge density:

Due to frequent propagating crises the model does not appear to be dependent on the initial edge density D_0 .

Crises frequency:

The value of C_R controls the speed of model dynamics, but not its qualitative behaviour. Lower or higher values of C_R result in the same model behaviour, but after a larger or a smaller number of iterations respectively. The value of C_R also affects the magnitude of the fluctuations of edge density with higher values of C_R resulting in smaller fluctuations. However, tests with C_R values ranging between 0.001 and 0.2 (not shown) reveal that that relationship is slower than linear.

Link stability:

The probability of the initial random survival of links between same-attribute nodes affects the rate of increase in link density, but not the quality of the model behaviour. This is expected since lower values of d_I result in more links being stabilised per time period, and vice versa. The value of d_F defines how fast the probability that a link survives a crisis grows as a function of the number of previously survived crises by that link. Tests reveal that d_F values close to 1 (slow stabilisation rate) fail to produce the above model dynamics. This is because links that survived several crises are then hardly more likely to survive a future crisis than "inexperienced" links. Very low values of d_F lead to rapidly increasing link density as a few random survivals make it unlikely that a link will fall victim to a crisis in future.

13.2.2.3. Complex and hierarchical modules

In the next series of experiments I consider the capability of the model to produce more complex modular structures. In the basis scenario, the modularity resulted from a higher probability to withstand a crisis for links between organisations sharing common business interests (nodes sharing same attribute values). However, in reality an organisation may have more than a single business interest overall. To account for this in the model, I extend the number L of attributes per node. In the experiments reported below, I set $L = 3$.

13.2.2.3.1. Scenario 1

Initially, I assume that organisations prefer to concentrate on a particular business interest at any one time. For this, I introduce a parameter $\alpha \in \{1, \dots, L\}$ that describes which attribute is active at the current time (i.e., here $\alpha \in \{1, 2, 3\}$). At the beginning of the simulation I set $\alpha = 1$, and then change the value of α every α_R generations.

In [section 13.2.2.1 above](#), modularity was measured in respect to segmentations implied by node attribute values. Here, there are 3 attribute values, and I thus measure M in respect to each of the respective segmentations. M_α then denotes the modularity measured in respect to the currently active attribute.

Results:

[Figure 13.8](#) shows results of a model run for 10,000,000 generations with $\alpha_R = 500,000$. Similar to the basis scenario, the overall link density steadily increases. At any one time, M_α grows, while the modularity $M_{k \neq \alpha}$ in respect to other attributes declines. This is expected: Previous experiments ([section 13.2.2.1](#)) established that M_α increases over time, and as module boundaries in respect to different attributes likely do not coincide, increase in inner-module links density for attribute α likely results in an increase in cross-module links in respect to other attributes.

During every cycle of α , modularity in respect to each attribute rises to a smaller maximum value than in the previous cycle. Similarly, the modularity of non-active attributes falls to a higher minimum value than previously ([Figure 13.8](#)). This may be due to fixation of links over time: If during a period with any specific α value a link happened to survive several crises, it is likely to continue wreathing future crises during subsequent periods of other α -values. As a result, its ability to survive further crises increases exponentially. Links fixed in this way are likely to be cross-modular in respect to periods of most α -values except the α -value during which they were fixed initially. This process imposes an increasing lower limit of cross-modular links, thus providing an upper limit on M_α . At the same time, fixed links also provide a lower limit on $M_{k \neq \alpha}$ as such links are not removed.

Notable is the attractor value M_S for modularity in respect to all attributes. For the parameter values used here, $M_S \approx 0.42$. This implies a complex modular topology: For each attribute k , the proportion of inner-modular links is lower than 50% ($\forall k \in \{1, \dots, L\}: M_k < 0.5$), however, significantly higher than would be expected if links were random ($\forall k \in \{1, \dots, L\}: M_k \gg 1/A = 0.2$). Initial tests suggest that the speed of convergence of upper and lower limits on M towards M_S follows a power-law, however, a detailed investigation of this behaviour for different parameter values remains an open question for future research. Also, the dependence of M_S on specific parameters, and the underlying causal mechanism present an interesting open question.

Sensitivity:

A detailed quantitative investigation of the convergence behaviour of M , and the dependence of M_S on parameter values is already identified as an open question for future research. The *qualitative* model behaviour, however, is investigated here and found not to be sensitive to particular parameter values:

Experiments with L -values between 2 and 5 show that the general behaviour pattern of the model is not sensitive to the number of attributes. For larger values of L , the results are less pronounced. However, when the number of nodes is increased and the model is run for a

larger number of generations with a higher value of α_R , the dynamics are again similar to the case $L = 3$ that is reported here.

The point of time when α is changed is not probabilistic in the above experiments. Instead, α is changed every α_R generations. In addition, when α is changed, it is not chosen randomly, but updated according to $\alpha_{\text{new}} = (\alpha_{\text{old}} + 1) \bmod 3$. This strategy is chosen mainly in order to achieve results that can be conveniently averaged over several runs and presented graphically (e.g. see Figure 13.8). When the value of α or the point of time when it is changed are chosen randomly, trajectories for M_k differ widely between runs and averaged plots are only meaningful to investigate the asymptotic behaviour. The general asymptotic behaviour and the convergence of all modularity measures on M_S is not affected. In particular, a range of sensitivity tests show that:

- When changes of α remain deterministic, but α_R is increased (decreased), the speed of convergence on M_S becomes slower (faster) at the beginning of the simulation and faster (slower) towards the end. I.e., higher α_R makes the convergence curve steeper.
- When the point in time when α changes remains deterministic, but the next value of α is chosen randomly, the convergent oscillations for each of the different M_k are less regular, but the overall convergence on M_S is not affected.
- When instead of regular changes of α every α_R generations, α can change at every generation with the probability $1 / \alpha_R$, the convergence curves become further irregular, but the overall convergence behaviour is not affected.

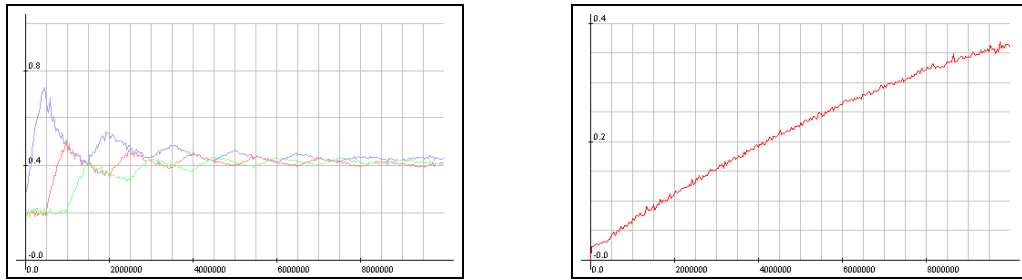


Figure 13.8. Model behaviour with $L = 3$ attributes with one attribute active at any given time (scenario 1). Shown are modularity values in respect to each of the 3 attributes (left) and the average edge density (right, red), averaged over all runs. Standard deviation is small and not shown for clarity.

13.2.2.3.2. Scenario 2

In the previous scenario the overall edge density in the model increased steadily. However, in the real world, the density of relationships between organisations never saturates: everyone does not co-operate with everybody directly. Even old relationships that weathered many crises may come to an end. To account for this, I alter the model:

At every generation, a single link is removed from the network with a probability p_k that does not depend on the number of crises withstood by that link. The link is selected randomly and uniformly from the network without regard for the attributes of the nodes it connects. The probability p_k is:

$$p_k = p_{k_{\min}} + \frac{p_{k_{\max}}}{1 + e^{-v \cdot (D - D_9)}}$$

where $p_{k_{\min}}$ and $p_{k_{\max}}$ are a bounds resulting in $p_{k_{\min}} < p_k < (p_{k_{\max}} + p_{k_{\min}})$, D is the current overall network edge density, and $D_9 = 0.1$ and $v = 50$ are parameters.

The consequence of this approach is that p_k transitions sharply from very low values (close to $p_{k_{min}}$) to very high values (close to $p_{k_{max}}$ for small $p_{k_{min}}$) when the link density D grows above D_9 . The parameter v defines the sharpness of this transition. Here, I set $p_{k_{min}} = 10^{-4}$ and $p_{k_{max}} = 10^{-2}$ meaning that when link density is low, the probability of forced link removal is near zero; and when link density grows larger than $D_9 = 0.1$, the probability of forced link removal transitions to approximately 0.01 per iteration.

Results:

I perform 25 independent runs of this scenario, each run lasts 10,000,000 model iterations. α_R is set to 750,000. The averaged results are shown in Figure 13.9. In this scenario the effect of link fixation discussed in section 13.2.2.3.1 is neglected by the forced removal of links that does not depend on previous history of a link. As a result, the modularity computed in respect to non-active attributes falls quicker than in the previous scenario and almost reaches 0.2 (Figure 13.9, left) – a value that indicates no modularity compared to a random network.

The implications of these results are twofold. First, they confirm that the emergence of complex market modules in scenario 1 (section 13.2.2.3.1) was facilitated through the ability of co-operative links to mature and stabilise by weathering external pressures. Second, they show that when relations remain transient despite relative stabilisation through successful crisis weathering, market segments become more dynamic. When organisations loose interest in particular areas, previously forged relationships wear off slowly, and market segments eventually dissolve.

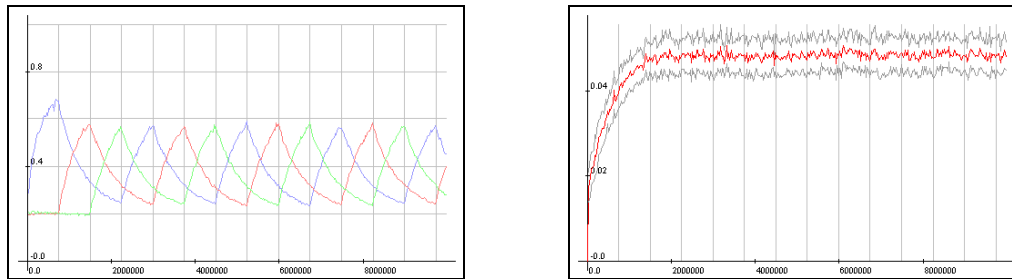


Figure 13.9. Model behaviour with forced edge removal (scenario 2). Shown are modularity values in respect to each of the 3 attributes (left) and the average edge density (right, red) with the grey lines showing the range of $\pm$ one standard deviation.

Sensitivity:

Tests show that the sensitivity of the model discussed here in respect to α -related parameters follows the same pattern as in the previous scenario (section 13.2.2.3.1). Here, I concentrate on model sensitivity towards parameters related to forced link removal that was introduces in this sub-section:

- Sensitivity analysis reveals that the key model dynamics are not significantly affected by changes of $p_{k_{max}}$ and v , as long as these values are sufficiently large. Small values of these variables neglect the effect of the logistic function that controls p_k , the probability of forced link removal.
- Large values of $p_{k_{min}}$ result in a significant probability of link removal even at very low densities. In such case, a sufficient number of links necessary to obtain meaningful modularity measures does not occur.
- The value of D_9 influences the overall edge density around which the network remains. If D_9 is high, the model dynamics are similar to the previous scenario 1; if D_9 is low, not sufficiently many links can develop for complex modules to emerge.

The “interesting” value range of D_{g} depends on network size N , the number L of attributes and the number A of values per attribute. This is because these three parameters influence the mean number of nodes per module. For the parameter values used in the above experiments ($N = 150$, $L = 3$, $A = 5$), the model has been verified not to be sensitive to values of D_{g} in range $[0.8, 0.15]$. For comparison: for parameters $N = 500$, $L = 3$ and $A = 5$, a similar behaviour was found at least in the range $D_{\text{g}} \in [0.02, 0.05]$.

13.2.2.3.3. Scenario 3

In the above experiments I assumed that organisations concentrate at a single business interest at any one time. In the model, this was reflected through a parameter α that specifies the “active” node attribute. However, real organisations typically pursue multiple business interests at any one time. To reflect this, I modify the model to allow all attributes to be active simultaneously at all times. The ability of a link to persist through crises is then based on the number of corresponding attribute values shared by the organisations connected by that link. In detail:

Previously, a link “survived” a crisis either if it had already survived one or more crises, or if it connected two nodes that had the same value for the currently active attribute. In the latter case, the survival-probability was $s = 1 - d_1$, where $d_1 = 0.9$ was the “link death probability”. In this scenario, the “link death probability” for links that have not yet withstood one or more crises is changed to:

$$d_1 = \begin{cases} 1 & , \text{ if } \varsigma = 0 \\ 0.999 & , \text{ if } \varsigma = 1 \\ 0.99 & , \text{ if } \varsigma = 2 \\ 0.9 & , \text{ if } \varsigma = 3 \end{cases}$$

where ς is the number of attributes for which the nodes connected by that link have the same value in common.

Results:

I perform 25 independent model runs with and without edge removal that was introduced in the previous scenario. The results (Figure 13.10) show that modularity M measured in respect to each of the 3 attributes grows simultaneously (the graphs largely overlap). With forced edge removal, M is constrained by the edge density (Figure 13.10, top); without edge removal, M grows steadily (Figure 13.10, bottom). In both cases, M is significantly higher than 0.2, which is the expected value in a random network with $A = 5$.

A simple consideration of various attribute permutations involved hints at the complexity of possible emerging topologies: Attribute values are assigned randomly and independently, so for $L = 3$, the network contains A^3 node groups of approximately equal size that share all attribute values. Each of these groups represents a unique possible permutation of attribute values. Each such group shares 2 attributes with $3 \cdot (A - 1)$ other distinct attribute permutations, resulting in $A^3 \cdot 3 \cdot (A - 1) / 2$ possible pairs. Similarly, each unique attribute permutation shares a single attribute with $3 \cdot (A - 1)^2$ other distinct permutations, resulting in $A^3 \cdot 3 \cdot (A - 1)^2 / 2$ pairs. Groups sharing attributes exhibit higher internal connectivity compared to the overall connectivity of the network, and the density of the internal connectivity depends on the number of shared attribute values. Groups that share 3 attributes are more dense and can participate in several modules that are based on 2 common attributes. In turn, groups that share 2 attributes can participate in several modules based on a single common attribute. The resulting topology parallels real economies where agents self-organise into higher order structures that play specific market roles, while the agents still continue behaving according to a local world-view without a global control function.

Due to a large number of nodes and edges, networks resulting from the experiments described here are difficult to visualise. Thus I repeated the experiments for $A = 2$, $L = 3$ and $N = 75$. While the overall dynamics are similar to Figure 13.10, the networks were small

enough to represent graphically. [Figure 13.11](#) shows a typical network resulting from the model with some of emerging modules of different hierarchy levels highlighted.

13.2.2.4. Summary

In the models discussed in this section, modules emerged through an implicit DPE-process, where the transitions between poorly connected and well connected network phases resulted from a continuous addition of links and intermittent propagation of crises. However, due to the non-linear effects introduced through stabilisation and fixation of links that survived several crises, link density fluctuated strongly and deviated far from the critical propagation threshold, resulting in prolonged connectivity phases. This emergent DPE behaviour is in contrast to the models discussed in [section 13.2.1](#), where phases of predominantly low and local connectivity and global phases where facilitated explicitly.

In the basis scenario, modularity emerges through the tendency of nodes sharing attribute values to continue interacting despite crises. The modules thus correspond to segments defined by node sets with the same attribute values and modular structure incorporates a single level of hierarchy.

When multiple attributes are present and act sequentially, multiple and overlapping modular structures emerge. Depending on whether link fixation is immutable (scenario 1, [section 13.2.2.3.1](#)) or whether stable links can still fail randomly (scenario 2, [section 13.2.2.3.2](#)), the modular structures are either locked in a complex overlapping topology (scenario 1), or dynamically transition between periods where one segregation dominates over others and periods where several segregations create a complex web of more and less dense network regions.

When multiple attributes act simultaneously in an additive fashion (scenario 2, [section 13.2.2.3.3](#)), network topology develops overlapping modular structures of several hierarchical levels, with small and dense modules participating in multiple larger, less dense modules of higher hierarchy levels. Networks obtained in the experiments reported here contain modular structures of at least three hierarchy levels.

Simple assumptions about network components and the nature of their interactions lead to the emergence of complex modular structures within the simulation models discussed in this chapter. The models were developed in the context of agents acting in economies. Clearly, the models present strong simplifications and lack many features of real economies. Nevertheless, they capture several features of self-organisation in real markets. The abstract nature of the models means that the insights obtained from this work may also apply to other areas. For instance, preferential stabilisation of interactions between entities that share common features and further stabilisation of interactions that weather disturbances can be found in systems such as genetic regulatory systems, protein-protein interactions, and other bio-chemical networks [Bower & Bolouri, 2004; Golemis & Adams, 2005]. The work discussed here thus presents a step towards a better understanding of mechanisms that lead to complex modular self-organisation in natural and social systems.

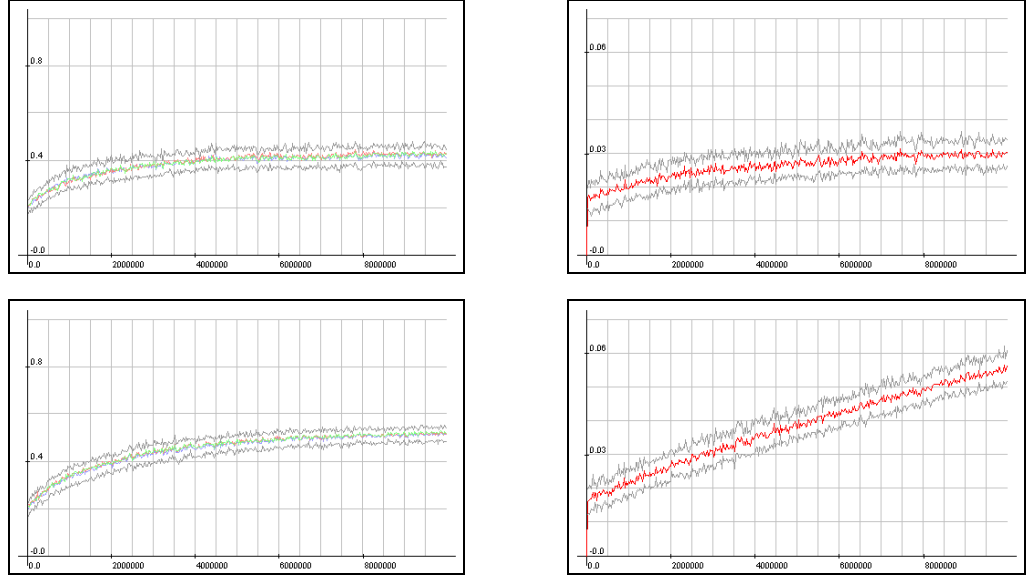


Figure 13.10. Model behaviour with simultaneous effects of all node attributes (scenario 3) with forced edge removal (top) and without forced edge removal (bottom). Shown are (overlapping) modularity values in respect to each of the 3 attributes (left) and the average edge density (right, red). The grey lines show the range of $\pm$ one standard deviation.

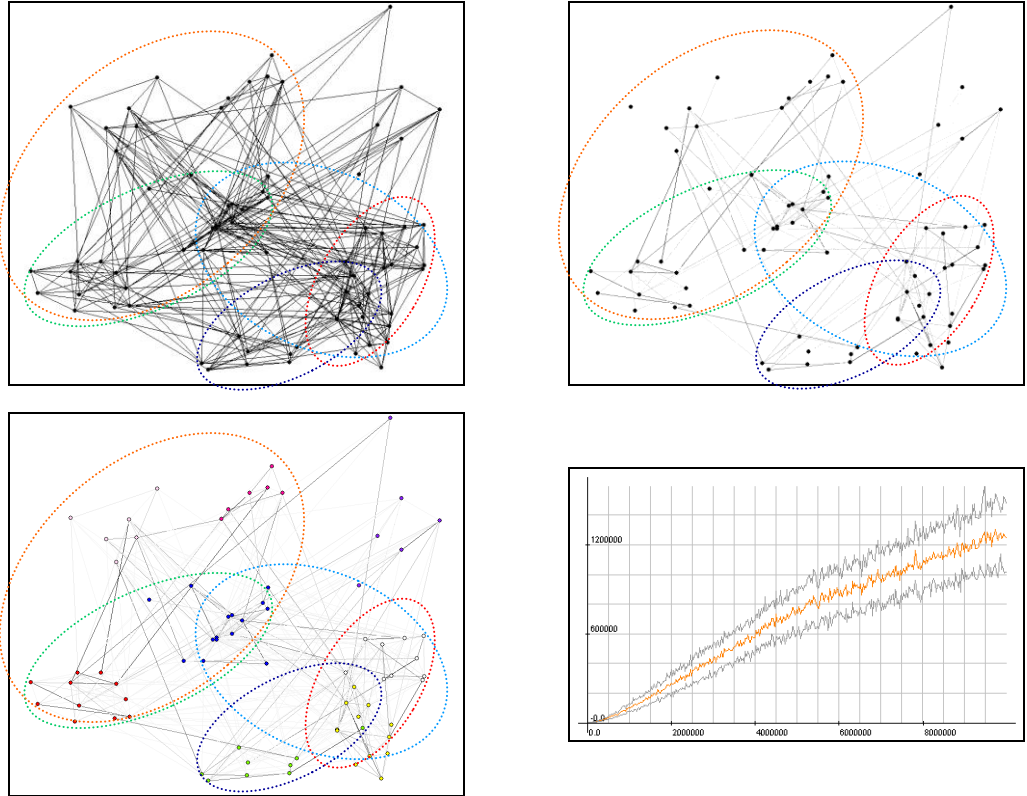


Figure 13.11. Typical network resulting from a model with simultaneous effects of all node attributes (scenario 3). The parameters used here are $N = 75$, $A = 2$ and $L = 3$. The same network is shown with all nodes and links in black (top, left); nodes in black and links in shades of grey in proportion to the age of the link counted in iterations, older links are darker (top, right); links in shades of grey in proportion to age and nodes coloured according to the unique permutation of all attribute values (bottom, left). The graph on the bottom right shows the average age of all links in the network throughout the model life-time with grey lines showing the range of $\pm$ one standard deviation over 25 runs.

13.3. Scale-free networks

Some of the material presented in this chapter is based on the following publications:

- 📄 G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
- 📄 G. Paperin, D. G. Green, T. G. Leishman (2008):
Dual Phase Evolution and Self-Organisation in Networks.
7th International Conference on Simulated Evolution And Learning (SEAL'08). Ref: [Paperin, Green, et al., 2008].
- 📄 G. Paperin (2008):
Dual Phase Evolution in Natural and Artificial Computational Systems.
Invited talk at the 12th Asia-Pacific Symposium on Intelligent and Evolutionary Systems (IES'08). Ref: [Paperin, 2008a].

Scale-free networks have received much attention by researchers due to their wide presence in nature and due to their relationship to self-organised critical systems implied by the power-law node degree distribution that signifies scale-free networks ([chapter 5](#)).

Traditional generative models for scale-free networks are based on preferential attachment in growing networks [Barabasi & Albert, 1999], a process where new nodes are added to the network and connected to existing nodes with a probability proportional to their degree. However, scale-free structures also occur in many natural networks of fixed size. For instance, it has been found [Eguiluz et al., 2005] that functional networks in brains exhibit scale-free structures. The vertices of such networks (neurons) develop during fetal development, but the connections mature at latter stages when the number of neurons is largely constant. Some generative models for scale-free networks of fixed size have been proposed [e.g. Dangalchev, 2004; Catanzaro et al., 2005], however, these models are highly synthetic and unlikely to correspond to any natural process.

Here I propose and examine two similar generative models for scale-free networks of constant size, and argue that these models may correspond to processes occurring in nature.

13.3.1. Model overview

A primary network structure is generated through a process of alternating local and global interactions. During local interactions, nodes selected randomly with a probability proportional to their degree interact with a neighbour selected randomly with a probability inversely proportional to its degree. A local interaction breaks edges of the primary structure. During global interactions, low degree nodes interact preferentially with high degree nodes anywhere in the network. A global interaction results in a new structural edge. Computer simulations discussed here show that over time the node degree distribution approaches a power-law, and that edge density converges on a value that depends on the fixed network size.

A key feature of this abstract model is that it corresponds to processes occurring in nature. Recall, for instance, that many social interaction networks exhibit scale-free structures. In many societies (human or animal), individuals are under social selection for the ability to acquire social links with other individuals whose co-operation may be valuable. Determining the actual social value of an individual is a resource intensive task and individuals often use the assessment by others as a heuristic; i.e., the number of social links an individual already has is used by others as an estimate of that individual's "attractiveness" [Eder, 1985]. In a growing social group this mechanism corresponds to preferential attachment [Barabasi & Albert, 1999], however, most human and animal societies tend to keep an approximately constant size over time [Dunbar, 1993]. Due to selection pressure, individuals attempt to forge cooperative connections to popular individuals and tend to break connections with less popular group members that are not useful to them. This corresponds to the above abstract model: poorly connected entities seek to create edges to well connected entities, while well connected entities tend to break edges to poorly connected ones.

Other examples of natural processes bearing resemblance to the current model are possible. A key observation is that the interactions that create or destroy links of the primary network structure are not directed or organised by a global controller that aims to arrive at a power-law degree distribution. Instead, such interactions occur according to local probabilistic rules and the scale-free structures are emergent. This is a fundamental difference between the current and previously known generative models [e.g. Dangalchev, 2004; Catanzaro et al., 2005] for constant-size scale-free networks.

13.3.2. Initial investigation

For a preliminary investigation of the above model architecture I create a computer simulation consisting of a network of $N=1000$ nodes. The network is initialised with random connections. (Sensitivity analysis of this and other experiments described in this section has shown that the initial edge density $\neq 0$ does not affect the result, except for the number of iterations required to approach the attractor edge density.) Thereafter, the network is subjected to structural modifications as follows:

Local interactions:

High-degree nodes preferentially interact with their low-degree neighbours. Local interactions are destructive.

First, a random node a is selected with a probability $p_{L,s}$ proportional to the number of edges connected to a :

$$p_{L,s}(a) = \frac{\text{degree}(a)}{1 + \sum_{i=1}^N (\text{degree}(n_i) + 1)}$$

where $p_{L,s}(a)$ is the probability of the node a being selected and $\text{degree}(n_i)$ is the number of edges connected to a node n_i .

Then, if a has any connected neighbours, a neighbour b of a is randomly selected with a probability $p_{L,e}$ anti-proportional to the number of edges connected to b :

$$p_{L,e}(b) = 1 - \frac{\text{degree}(b)}{1 + \sum_{k_j \in \mathcal{N}_a} (\text{degree}(k_j) + 1)}$$

where $\mathcal{N}_a$ is the set of direct neighbours of a .

Thereafter, the edge (a, b) is removed.

Global interactions:

Low-degree nodes preferentially interact with high-degree nodes. Global interactions are constructive.

First, a node c is randomly selected with a probability $p_{G,s}$ anti-proportional to the number of connected edges:

$$p_{G,s}(c) = 1 - \frac{\text{degree}(c)}{1 + \sum_{i=1}^N (\text{degree}(n_i) + 1)}$$

Then a second node d is chosen randomly and independently of c with a probability $p_{G,e}$ proportional to the number of edges connected to d :

$$p_{G,e}(d) = \frac{\text{degree}(d)}{1 + \sum_{i=1}^N (\text{degree}(n_i) + 1)}$$

If there is no edge (c, d) then such edge is added to the network. Self-connections are not allowed.

Initial results:

I run this model with $N = 1000$ nodes for several hundred thousand iterations with a phase change between global and local interactions every 10 iterations, and measure the edge density D and the distribution of node degrees at every iteration.

To verify that the model produces a scale-free network, I perform a linear regression of the log of the degree distribution to a linear model and determine the coefficient of determination R^2 (i.e. the square of the linear correlation coefficient between the log degree distribution and a linear model). A value of R^2 close to 1 indicates a good fit between a power-law distribution and the distribution of node degrees in the model.

The results show (Figure 13.12) that the edge density D converges on a small value D_s . The initial random network is a common Poisson graph, and R^2 is close to zero. As D approaches D_s , R^2 experiences a critical transition to a value close to 1. This indicates that the proposed generative process may indeed result in a scale-free network topology. To further verify this, as well as to investigate the relationship between network size N and the attractor edge density D_s , at which the network becomes scale-free, I perform a series of further experiments.

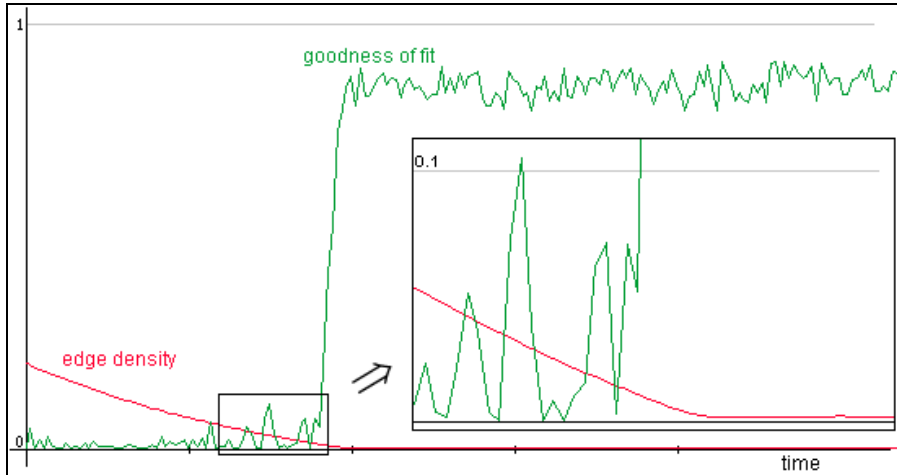


Figure 13.12. Critical transition of the node degree distribution from random to scale-free. Depicted is a typical run for a network of $N=1000$ nodes.

13.3.3. Systematic investigation

13.3.3.1. Relative length of local and global interaction phases

A sensitivity analysis performed on this as well as on other models discussed later in this section shows that the models do not depend on the duration of phases of global and local interactions. Phase lengths between 1 and 100 iterations did not affect the general results, and only affected the speed of convergence (however, in most cases – by a statistically not significant amount). Long phases in smaller networks can lead to a breakdown in the model dynamics because an on-going local phase can lead to the removal of all edges from the network making preferential selection meaningless.

When the phases are of different length, the model dynamics break down. Longer global phases lead to ever-increasing edge densities; longer local phases lead to an eventual loss of all edges. Thus, in all following experiments, the phase lengths are set to 1, i.e. local and global interactions alternate.

13.3.3.2. Power-law goodness of fit and estimated exponent

In order to thoroughly verify the initial results I repeat the initial experiment for different network sizes of $N = 100, 200, 500$ and 1000 nodes. For each network size I perform 35 independent simulation runs and average the results (Figure 13.13). The R^2 value between 0.6 and 0.8 suggest that the distribution of node degrees resembles a power-law, however, better fits can be expected of truly scale-free networks. Particularly, for large networks ($N=1000$) a rough fit cannot be explained by a small number of sample nodes. Here, however, the R^2 value rather slightly deteriorates for larger N . Figure 13.14 shows the distributions obtained in the last iteration of each run.

The results in the right column of Figure 13.14 show that although there is a good agreement of larger node degree frequencies to a power-law, smaller-degree nodes are clearly under-represented relative to what can be expected in a scale-free network. A visual inspection of the left column in Figure 13.14 suggests a Poisson distribution of node degrees that is skewed towards smaller degrees in comparison to a classical random Poisson graph.

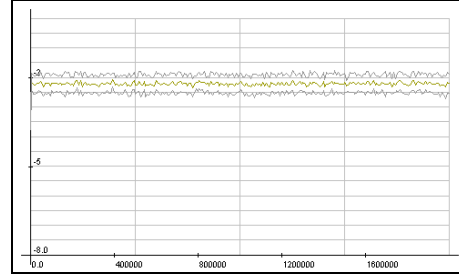
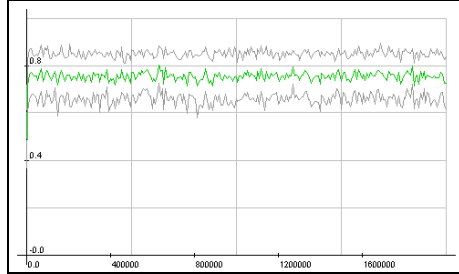
Seeking to improve the generative model I repeated the series of experiments while allowing self-connections for nodes. The results (not provided for conciseness) show a better agreement between the resulting network structure and a scale-free topology: the resulting node degree frequencies do not fall for low degree values as above, and the R^2 values outperform the above values by about 0.1. However, the low-degree node frequencies in the allow-self-connections scenario are still not as high as can be expected from a true power-law distribution. Moreover, recall the real-world scenario that motivated the current model: poorly connected entities seek to associate with well connected entities, and well connected entities seek to avoid such associations. A self-connection does not have a natural interpretation within this scenario. In the next sub-section I describe a different modification to the current model that results in both, a better agreement with a power-law, and a wider applicable interpretation within natural systems.

Net size

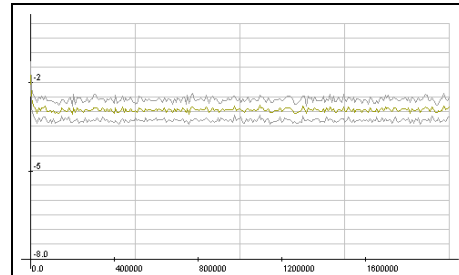
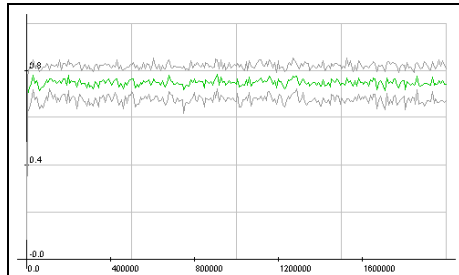
Goodness of fit (R^2)

Estimated exponent

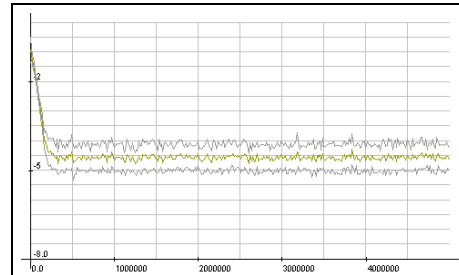
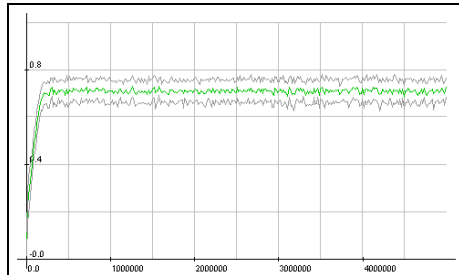
$N = 100$



$N = 200$



$N = 500$



$N = 1000$

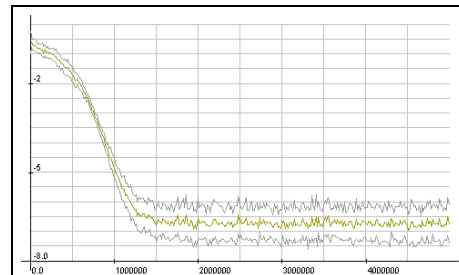
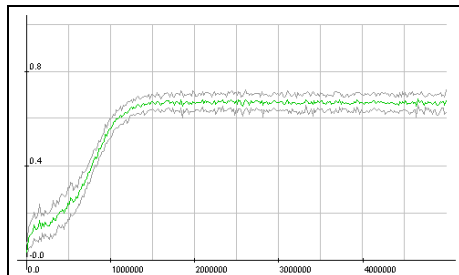


Figure 13.13. Goodness of fit and estimated exponent of the node degree distribution to a power-law. The goodness of fit (green) is measured as a linear correlation coefficient R^2 of the log degree distribution. The estimated exponent (dark yellow) is obtained on the basis of a least squares fit of the log degree distribution to a linear model. For each network size, 35 independent runs are performed and the results are averaged. The grey lines show the range of $\pm$ one standard deviation of each value over the 35 runs. The x-axis shows the iteration number.

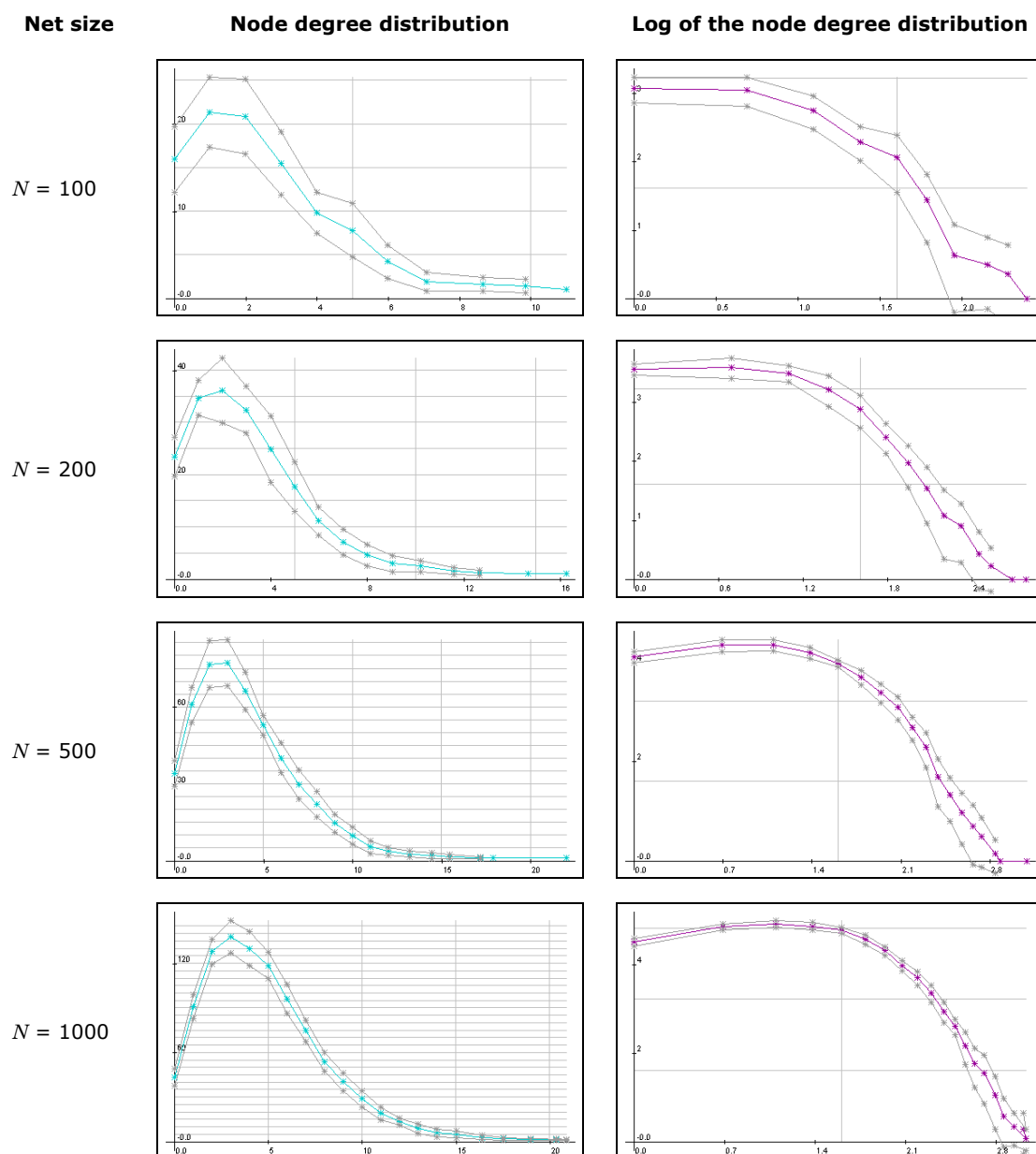


Figure 13.14. Actual node degree distribution and its logarithm. Shown is the frequency distribution of node degrees (cyan) at the end of the simulation, averaged over 35 independent runs for each network size. The grey lines show the range of $\pm$ one standard deviation in frequency for each observed node degree over the 35 runs. The right column shows node degree distribution on a log-log scale (bordeaux). All values in the right column are obtained by taking the natural logarithm of the corresponding values in the left column.

13.3.4. Improved generative model

In the generative model described in [section 13.3.2 above](#), low degree nodes seek to forge links to high degree nodes. This behaviour was motivated with natural scenarios where poorly connected entities gain advantage from interactions with well connected entities. However, it is not necessarily clear that poorly connected entities should gain a higher advantage from interactions with well connected nodes, than other entities. This prompts for a simplification of the model: Any node, regardless of its degree seeks to be connected to well connected nodes.

Similarly, the previous model assumed that well connected entities are more likely to break links to poorly connected nodes, then other entities. This assumption can also be simplified for equivalent reasons: Any node, regardless of its degree seeks to break away from poorly connected nodes. The updated model is summarised in [Table 10.7](#).

As with the initial model, I simulate the updated process for network sizes $N = 100, 200, 500$ and 1000 nodes, performing 35 independent simulation runs for each N , and average the measurements.

The results of the updated model ([Figure 13.15](#) and [Figure 13.16.](#)) show clearly that networks produced by that model are much closer to a scale-free topology than the networks produced by the original model, although small-degree nodes are still underrepresented. On contrary to the previous model, the goodness of fit improves slightly for larger networks. This is expected, as small networks do not have enough nodes for a smooth power-law distribution of degree frequencies.

Node type		Initial model	Updated model
Local interaction (breaking links)	start node	$p_{L,s}(a) = \frac{\text{degree}(a)}{1 + \sum_{i=1}^N (\text{degree}(n_i) + 1)}$	$p_{L,s}(a) = \frac{1}{N}$
	end node	$p_{L,e}(b) = 1 - \frac{\text{degree}(b)}{1 + \sum_{k_j \in \mathcal{N}_a} (\text{degree}(k_j) + 1)}$	$p_{L,e}(b) = 1 - \frac{\text{degree}(b)}{1 + \sum_{k_j \in \mathcal{N}_a} (\text{degree}(k_j) + 1)}$
Global interaction (making links)	start node	$p_{G,s}(c) = 1 - \frac{\text{degree}(c)}{1 + \sum_{i=1}^N (\text{degree}(n_i) + 1)}$	$p_{G,s}(c) = \frac{1}{N}$
	end node	$p_{G,e}(d) = \frac{\text{degree}(d)}{1 + \sum_{i=1}^N (\text{degree}(n_i) + 1)}$	$p_{G,e}(d) = \frac{\text{degree}(d)}{1 + \sum_{i=1}^N (\text{degree}(n_i) + 1)}$

Table 13.4. Probabilities of nodes to be selected as start or end nodes during local and global interactions in the original and the updated models. Local interactions: in the initial model, high-degree nodes interact with their low degree neighbours; in the updated model, any nodes interact with their low-degree neighbours. Global interactions: in the initial model, low-degree nodes interact with high-degree nodes; in the updated model, any nodes interact with high-degree nodes. Refer to [section 13.3.2](#) on [page 171](#) for variable notation.

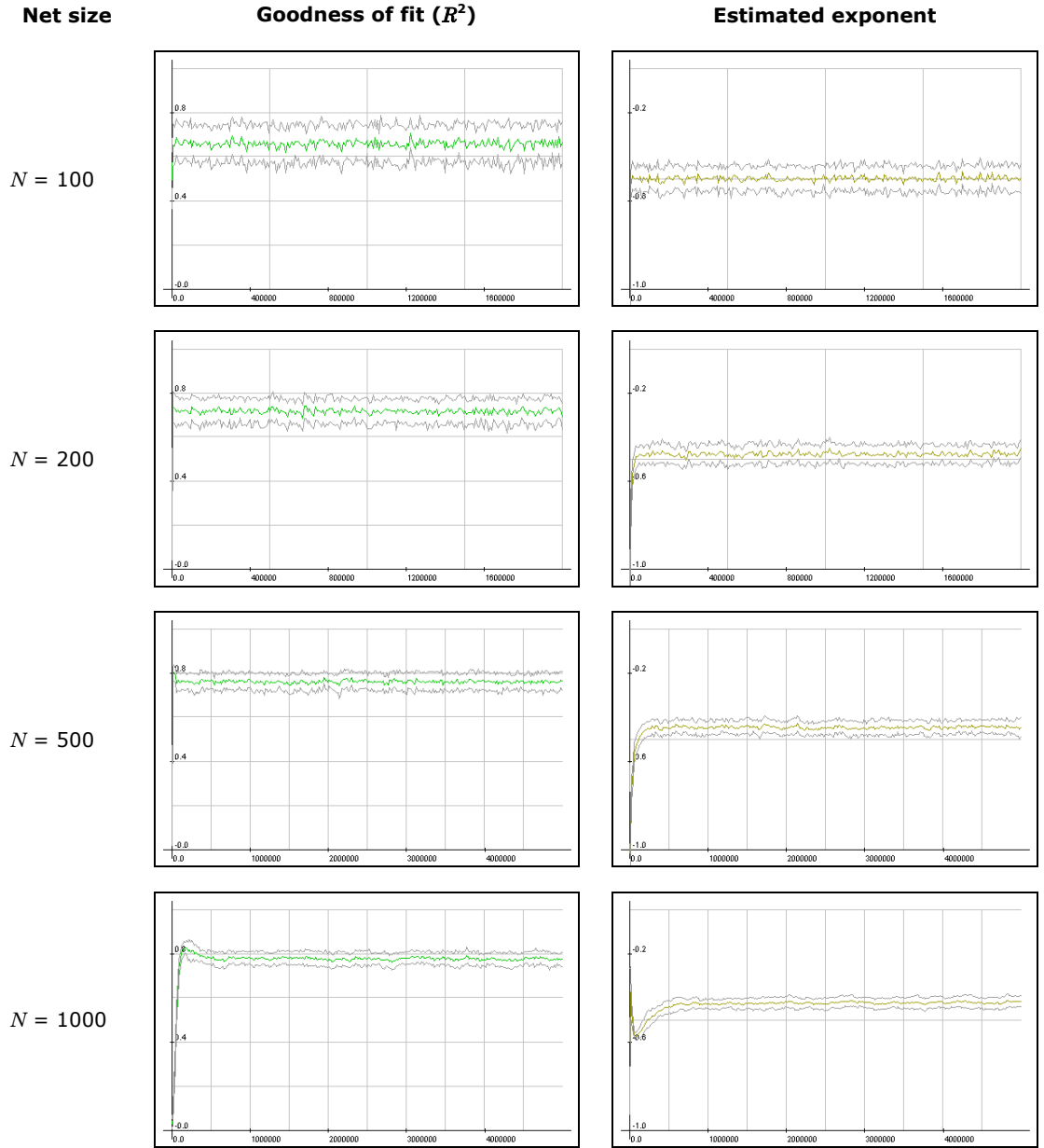


Figure 13.15. Goodness of fit and estimated exponent of the node degree distribution to a power-law in the updated model. The goodness of fit (green) is measured as a linear correlation coefficient R^2 of the log degree distribution. The estimated exponent (dark yellow) is obtained on the basis of a least squares fit of the log degree distribution to a linear model. For each network size, 35 independent runs are performed and the results are averaged. The grey lines show the range of $\pm$ one standard deviation of each value over the 35 runs. The x-axis shows the iteration number.

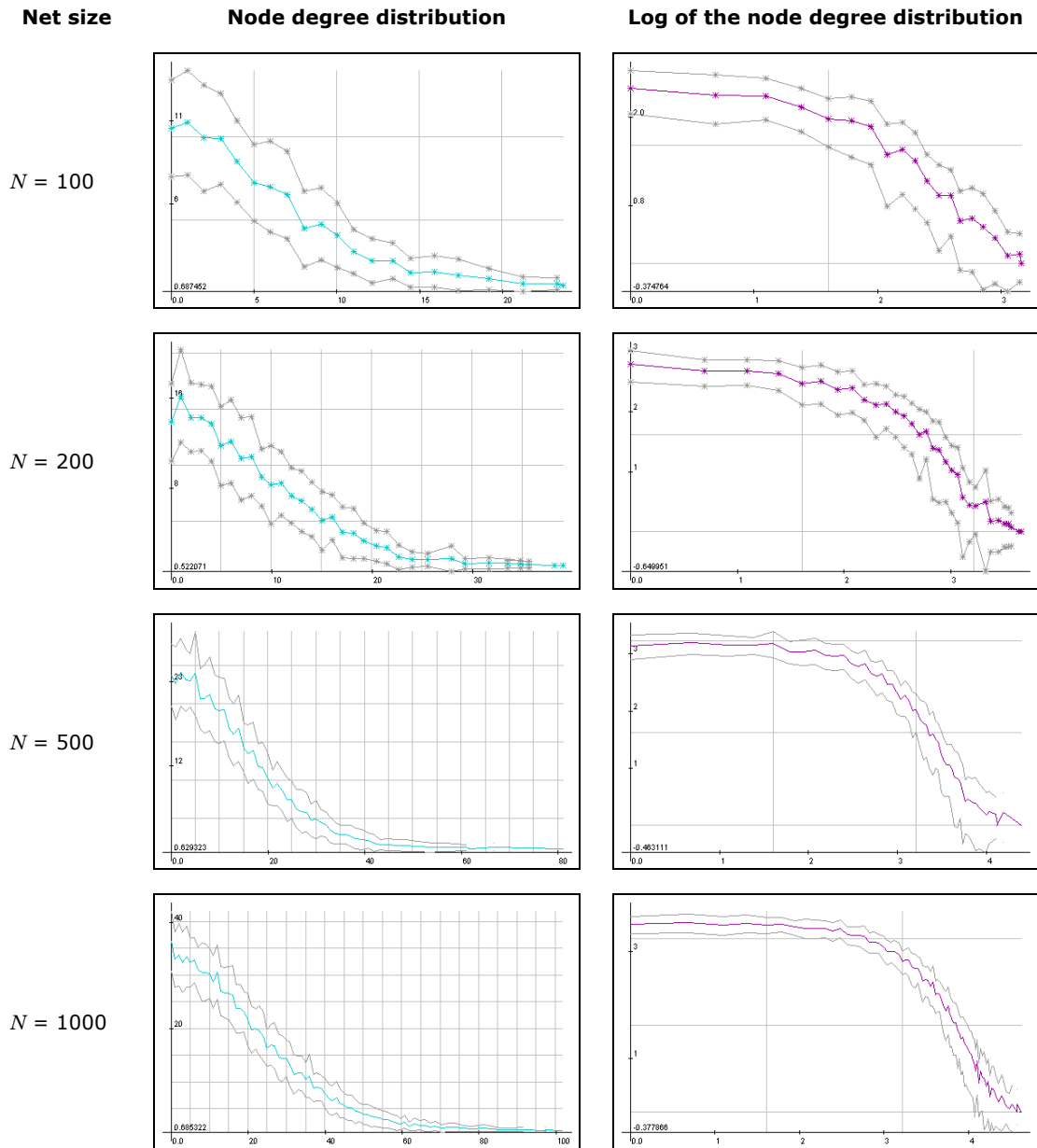


Figure 13.16. Actual node degree distribution and its logarithm in the updated model. Shown is the frequency distribution of node degrees (**cyan**) at the end of the simulation, averaged over 35 independent runs for each network size. The **grey** lines show the range of $\pm$ one standard deviation in frequency for each observed node degree over the 35 runs. The right column shows node degree distribution on a log-log scale (**bordeaux**). All values in the right column are obtained by taking the natural logarithm of the corresponding values in the left column.

13.3.4.1. Emergent two-phase dynamics in the updated model

Interesting behaviour is observed for $N = 1000$ in Figure 13.15: The measured quantities approach an attractor value, but initially “overshoot” and then return to the attractor. This behaviour is particularly apparent in the edge density of the network (Figure 13.17, left). A closer examination of the actual number of edges in the network (Figure 13.17, right) reveals that the point at which the system behaviour reverses (edge density stops falling and begins to grow towards the final attractor) is by several hundred edges smaller than the attractor. Moreover, it takes over 50,000 iterations for the edge density to reach the attractor value from the saddle point. The saddle point coincides in time with the transition of R^2 from low to high values. This shows that this effect is not due to overshooting, but results from qualitatively different model dynamics that depend on the network structure at any given time. Close to the beginning of the simulation, the edge distribution is largely random and independent. In this configuration, the rewiring rules defined by local and global interactions rapidly remove edges. This process does not stop when the eventual attractor density is reached and continues until the edge density is significantly lower. Only when the network approaches a scale-free topology, the model dynamics change, which results in slow addition of edges. These changed dynamics persist at edge densities during which edges were removed previously. Once a certain attractor edge density is reached, the network begins to fluctuate around this density. Figure 13.18 verifies that the initial dynamics are not caused by high edge density, but indeed by the random network configuration at the beginning of the simulation: Even when the initial edge density is below the eventual attractor value, the edges are removed until the network enters an approximately scale-free configuration. Then the dynamics change and edges are added again. Notably, this complex dual-phase behaviour emerges from fixed simple rules that are based only on local node information.

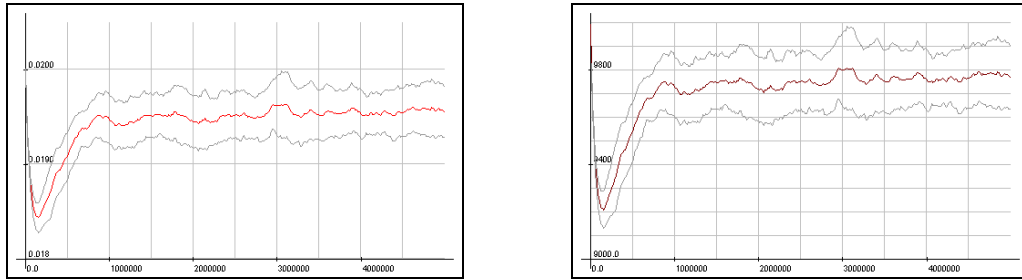


Figure 13.17. Edge density and number of nodes. Shown is the edge density (left, red) and the number of edges (right, dark red) for networks of size $N = 1000$, averaged over 35 independent runs of the updated model. The grey lines show the range of $\pm$ one standard deviation over the 35 runs.

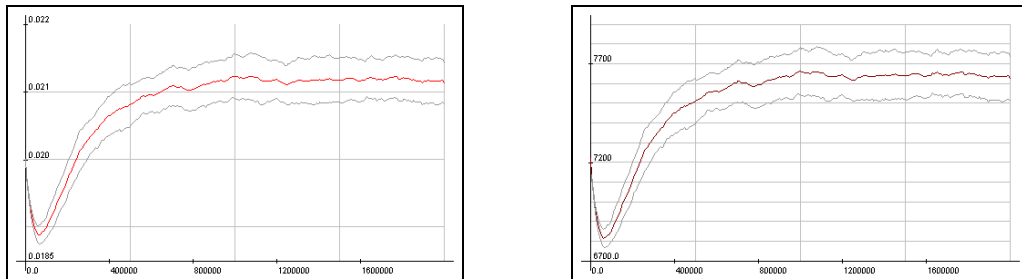


Figure 13.18. The two different classes of dynamical behaviour do not depend on the initial edge density, but only on the network topology. Shown is the edge density (left, red) and the number of edges (right, dark red) for networks of size $N = 850$. Although the initial edge density is below the eventual attractor value, the edges are removed until the network enters an approximately scale-free configuration, then edges are added again.

13.3.5. Dependence of attractor edge density on network size

In all of the above experiments the network settles at an approximately constant edge density D_s once an approximately scale-free configuration is reached. Sensitivity analysis shows that this attractor depends only on the network size measured as the number of nodes N .

To investigate this relationship in more detail I repeat the simulations for both generative models for network sizes between $N = 40$ and $N = 1100$ in 50 nodes increments. For each N , I perform 25 independent runs for 2,000,000 iterations each. The results are averaged. Then, for each N , the mean edge density during the last 500,000 iterations is determined. The results are shown in Figure 13.19 (initial model) and Figure 13.20 (updated model).

The relationship between N and D_s is clearly a power-law. For both models a power-law fit results in a value $R^2 > 0.99$. For the initial model, the relationship within the measured range is given by $D_s = 0.0416 \cdot N^{-0.7317}$; for the updated model, by $D_s = 0.0808 \cdot N^{-0.484}$.

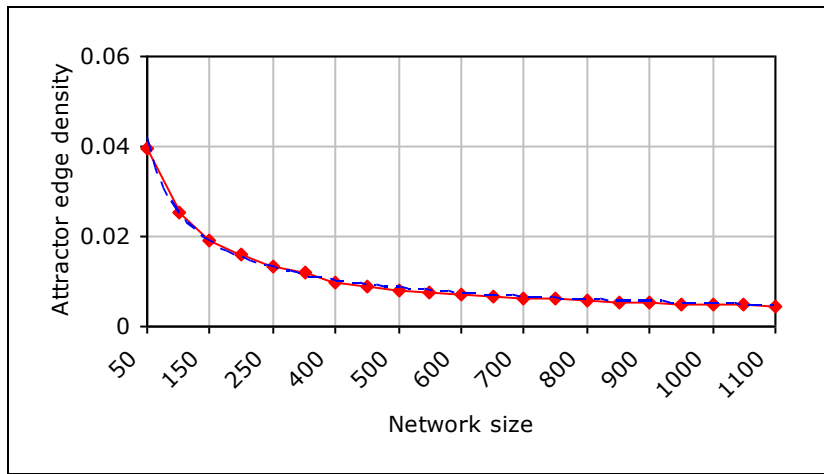


Figure 13.19. The dependence of the attractor edge density D_s on the network size N in the initial model. Shown are the measured values of D_s (solid red line) and a power-law fit (punctuated blue line).

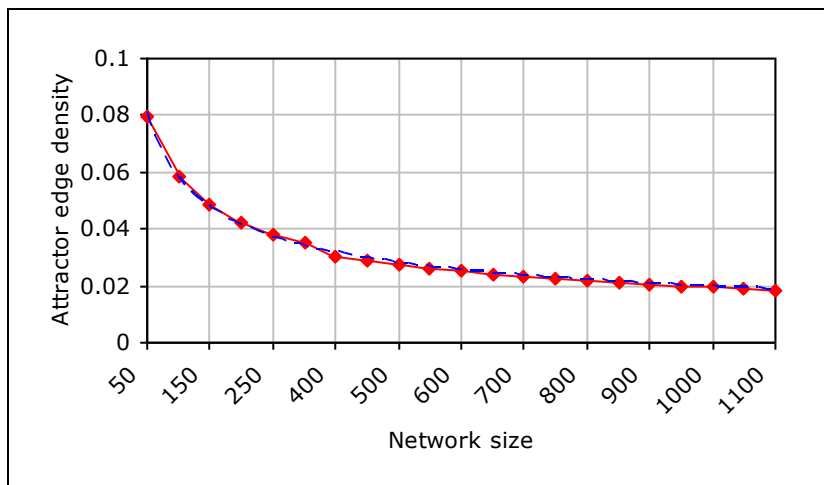


Figure 13.20. The dependence of the attractor edge density D_s on the network size N in the updated model. Shown are the measured values of D_s (solid red line) and a power-law fit (punctuated blue line).

13.3.6. Summary

In this section I analysed two similar generative model for scale-free networks of constant size. The first model (section 13.3.2) shows a good fit to a power-law distribution of node degrees for large degrees, however, the frequency of small-degree nodes does not follow a power-law. The second model (section 13.3.4) produces networks with a node degree distribution that is close to a power-law, with an exponential cut-off for *small* degrees (see Figure 13.21 for a typical network produced by this model). Explorative experiments suggest that if self-connections are allowed, a better power-law fit for low degrees may result. Arguably, self-connections do not have a meaningful interpretation within the natural processes that motivate the model. However, one possibility to interpret a network of co-operations between social animals is to consider connections to represent investment of resources. In that case, investment in both, other individuals and oneself is meaningful. A further investigation of the current models with the presence of self-connections may provide useful results.

When applied to larger networks, the second model produces topologies that are closer to truly scale-free structures, then when applied to smaller networks. This suggests that a large number of nodes may be necessary to produce a better power-law fit. In order to comprehensively describe the topologies produced by the models studied here, an analytical analysis of the generative processes is necessary. I would be excited to see this work extended in such way in future.

Networks with power-law degree distributions are ubiquitous in many complex systems and may be responsible for various scale-free phenomena in nature. The work discussed in this chapter presents a step toward a better understanding of processes behind the appearance of this important class of network topologies.

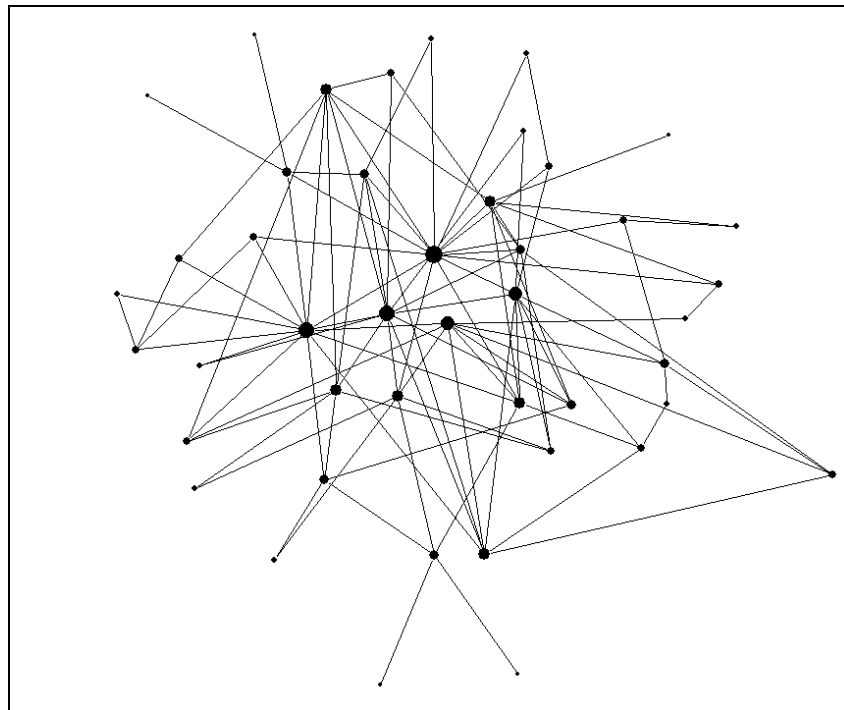


Figure 13.21. Example of a network obtained through the updated model. Shown is a typical result for $N = 50$. The sizes of the depicted nodes are chosen according to their degree: larger node size denotes higher node degree.

Part VI

The Big Picture

This part puts the research presented in this thesis into a wider context. Key findings are cross-referenced, and the relationship to established frameworks for reasoning about complex systems is discussed. The implications for future research are considered and specific potential directions for future studies are laid out.

Chapter 14: DPE and other frameworks for reasoning about complex systems.

DPE relates to other frameworks for reasoning about CAS. For instance, Self-Organised Criticality (SOC) implies that CAS tend to self-organise towards a critical attractor, and DPE proposes that systems develop towards a simple attractor, but repeatedly cross the critical area when perturbed. Both mechanisms can result in similar observations. Another well-known framework, the Adaptive Cycle (AC) describes recurrent transitions in dynamics of socio-economic systems. DPE extends this concept to a wider range of systems.

Chapter 15: Conclusions.

This thesis identified some mechanisms by which underlying networks facilitate emergence and self-organisation in complex systems, distilled these mechanisms into the DPE-framework, and applied the findings to elicit new insights into concrete systems. Several questions remain. Applications and extensions of introduced simulation techniques can provide further insights into biological evolution. The DPE framework itself can be extended and further formalised. In this last chapter, I summarise the main research results on this thesis and discuss some future research directions in detail.

Chapter 14:

DPE and other frameworks for reasoning about complex systems

DPE relates to other frameworks for reasoning about CAS, such as Self-Organised Criticality (SOC) and the Adaptive Cycle (AC). SOC implies that CAS tend to self-organise towards a critical attractor. On contrary, DPE proposes that systems develop towards a simple attractor, and are repeatedly perturbed away by internal and external disturbances, thereby repeatedly crossing the critical threshold. In many instances these two cases may result in equivalent observations. The AC describes recurrent transitions in dynamics of socio-economic systems between exploration of new development directions and exploitation of available resources. DPE generalises this concept in terms of network-theory and applies it to a wider range of systems.

Material presented in all sections of this chapter is based on the following publications:

- 📄 G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
- 📄 G. Paperin, S. Sadedin (2009):
Towards formalising the theory of Dual Phase Evolution.
10th European Conference on Artificial Life (ECAL'09). Ref: [Paperin & Sadedin, 2009b].
- 📄 G. Paperin, S. Sadedin (2009):
The Dual Phase Evolution Framework for Understanding Evolutionary Dynamics in Complex Adaptive Systems.
2009 International Conference on Evolutionary Computation (ICEC'09). Ref: [Paperin & Sadedin, 2009a].

Throughout these thesis I focused on network-theoretical considerations to reason about emergence and self-organisation in CAS. I showed that the DPE framework that concentrates on the role of connectivity phases in underlying networks is useful in uncovering common mechanisms in seemingly unrelated systems and for explaining mechanisms behind a range of complex phenomena. In this chapter I relate the DPE framework to two other frameworks that are widely applied for reasoning about CAS: Self-Organised Criticality (SOC) and the Adaptive Cycle (AC). By emphasising network-theoretical approaches and empirical observations from a variety of adaptive and non-evolutionary complex systems, I show how DPE complements both ideas.

14.1. Self-Organised Criticality

Self-Organised Criticality (SOC) [Bak et al., 1988; Bak, 1999] implies that CAS self-organise to a critical state, in which the complexity of system responses to external stimuli emerges through a propagation of the stimuli through an interaction network with thresholds at each component. The thresholds effectively control the propagation connectivity near the critical threshold between the connectivity phases, resulting in propagation avalanches of different sizes. In this sense, SOC suggests that CAS evolve to reside in a transitional state at the boundary between stasis and chaos.

Langton coined the term *edge-of-chaos* to describe the above critical attractor state [Langton 1990; 1991] in the context of studying cellular automata (for a comprehensive review of cellular automata, see Wolfram [2002]). He found that cellular automata of a certain class, termed *Wolfram class 4* [Wolfram, 1984], exhibited dynamics similar to systems that are thought to be dominated by SOC processes. Langton conjectured [Langton 1991] that computationally universal and various other “interesting” cellular automata are located within that class. This conjecture was supported by some studies [e.g. Packard, 1988], but later evidence [Mitchell et al., 1993] has shown that the arguments were not conclusive.

The size of propagation avalanches in SOC systems follows a power-law distribution. The presence of this distribution in data is often taken to indicate that a system may self-organise to a critical state. A number of models [Bak, 1999] led to suggestions that various complex systems, including many CAS, exhibit SOC dynamics. For instance, it has been suggested [Bak & Sneppen, 1993] that since the sizes of extinction events observed in the fossil record follow a power law distribution, punctuated equilibria [Eldredge & Gould, 1972] may be a result of the biosphere's self-organisation to a critical state. However, the extent to which SOC presents a general form for organisation of CAS remains unclear. Many SOC models concern complex systems that are not subject to selection, and in many cases there are several processes which may lead to power-law distributed data. For instance, Newman [1997] demonstrates a non-critical extinction model without any species interactions. That model yields a power-law with an exponent that is closer [Lenton & Van Oijen, 2002] to the empirical punctuated equilibria data than the SOC model [Bak & Sneppen, 1993]. Several authors [e.g. Sornette et al., 1995; De Carvalho & Prado, 2000] suggest that critical behaviour requires fine-tuning of an order parameter. Furthermore, it remains unclear whether SOC occurs in non-conservative systems [Kinouchi & Prado, 1999; De Carvalho & Prado, 2000]. The notion of *nearly-critical behaviour* avoids some of the problems related to SOC and can be applied to a wider range of systems [e.g. Kinouchi & Prado, 1999].

SOC has been associated with scale-free networks due to the apparent parallel between the power-laws present in such networks and in self-organised critical systems. Notably, some *non-critical* systems exhibit power-laws in system output sizes due to the scale-free topology of underlying networks. For instance, the distribution of sizes of random-fault affected subnets in some telecom networks follows a power law due to the scale-free nature of the networks and not due to any critical interactions [Albert et al., 2000].

Recently Eliazar and Klafter [2009] have shown analytically that 1-on-f noises, traditionally thought to indicate SOC dynamics, can generally emerge from a superimposition of many independent stochastic signals. In summary, the presence of power-law data may not always imply self-organisation towards critical dynamics.

DPE mechanisms may be involved in non-critical systems that produce power-law data. Throughout this thesis I discussed and analysed a range of systems that tend to self-stabilise, while external stimuli or internal phase changes perturb them away from their attractors and increase the connectivity of state transition networks. To describe this using SOC-vocabulary: Under SOC, systems self-organise *towards a critical region*; under DPE, systems rather develop towards a settled, balanced state, and are repeatedly pushed *across the critical region*, towards a chaotic state (in the sense that responses to random stimuli and variations are unpredictable), from which systems return to a new settled state, accumulating order and complexity on the way (Figure 14.1).

Notably, the DPE and the SOC paradigms are complementary, not contradictory. The magnitude DPE-oscillations (i.e. how far underlying networks and system dynamics deviate from critical during different phases) may vary significantly according to specific circumstances. Thus, recognising whether a particular phenomenon is caused by a truly critical attractor or by a repeatedly unsettled non-critical attractor may be difficult.

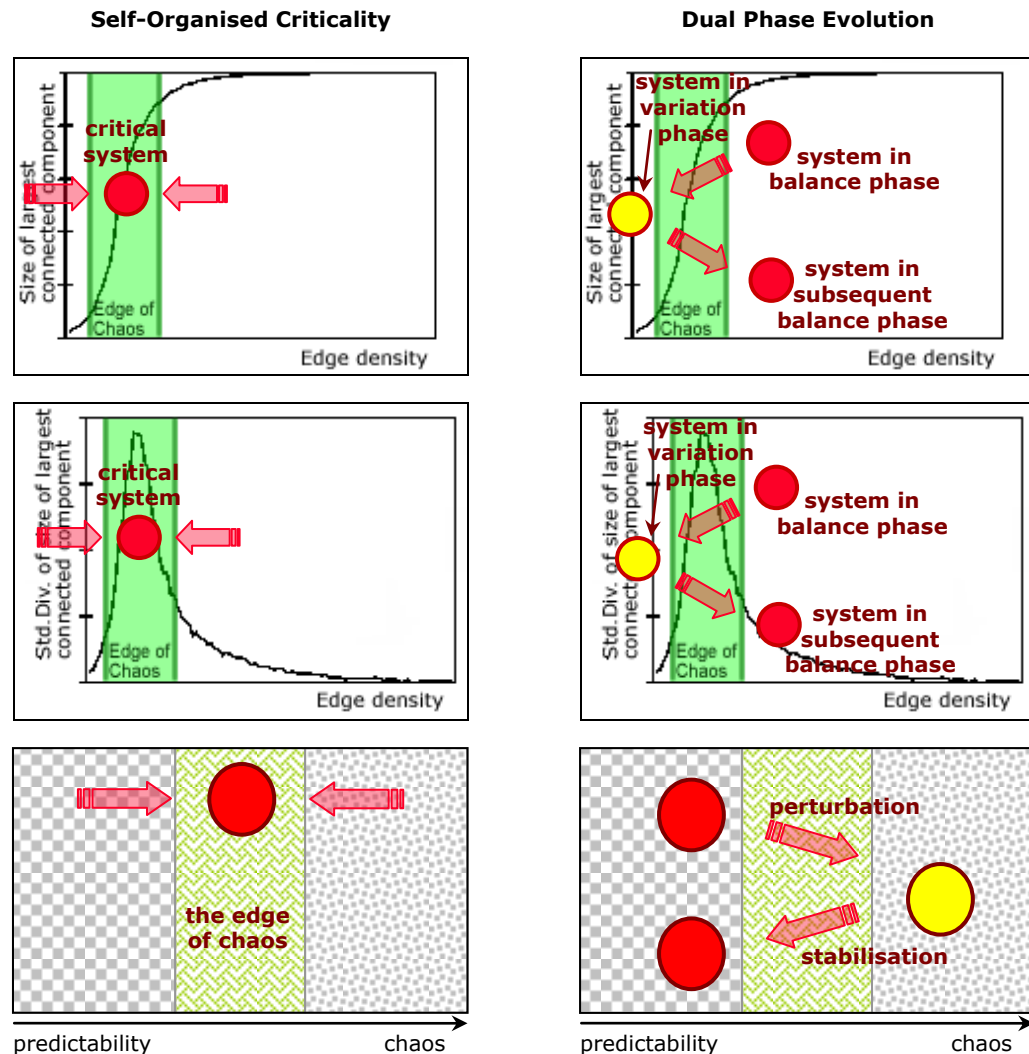


Figure 14.1. Self-Organised Criticality vs. Dual Phase Evolution.

Top two rows: Contrast between SOC and DPE in terms of network connectivity. SOC-theory (left) suggests that networks that underlie CAS self-organise towards the critical percolation threshold at which the size of the giant connected component undergoes a critical transition (top row), during which it exhibits large variation (middle row). On contrary, according to DPE (right), underlying networks repeatedly flip between poorly connected and well connected phases.

Bottom row: Contrast between SOC and DPE in terms of system dynamics. SOC-theory (left) suggests that CAS self-organise towards a critical transition state between the general stasis of equilibrium systems and the random behaviour of chaotic systems. According to DPE (right), CAS are repeatedly pushed from a balance-phase to a variation-phase and are thereafter driven to restore into a new balance phase.

14.2. The Adaptive Cycle

Another influential concept in CAS theory is the adaptive cycle (AC) [Gunderson & Holling, 2002]. The AC extends the idea of ecological succession [Gleason, 1927], and is predominantly applied to ecological and socio-ecological systems, especially with reference to ecosystem management.

The AC is closely related to the concept of *panarchies* [Holling, 2001; Gunderson & Holling, 2002] – a metaphor that describes the hierarchical relationships in dynamical systems of different scales. The AC is a metaphoric concept [Walker et al., 2004] suggesting that the behaviour of many socio-ecological systems exhibit cycles consisting of four qualitative phases:

- A growth and exploitation phase (designated r), in which new or freed-up areas and niches are rapidly populated by opportunistic organisms.
- A conservation phase (K) signified by competition, selection and resource accumulation.
- A collapse or release phase (Ω), in which accumulated resources are catastrophically released, often mediated by disturbances.
- A reorganisation phase (α) in which the remains of an Ω -collapse are reorganised and restructured.

The bulk of the literature on ACs [Resilience Alliance, 2004] is concerned with empirical studies [e.g. Cumming, 2004; Taylor & Krawchuk, 2005], secondary data analysis [e.g. Anderies, 2005; Allen, 2006], and applications to human-managed systems [e.g. Folke et al., 1994; Crema et al., 2000; Carpenter & Gunderson, 2001]. However, some work has been undertaken to provide a causal explanation for the AC (see Gunderson & Holling [2002] for an overview). Most of this work concentrates on ecosystem resilience and threshold dynamics [Holling, 2001; Gunderson & Holling, 2002; Scheffer & Carpenter, 2003]. Empirical studies attribute typical CAS properties to each phase of the AC. For instance, resilience against external forcing is expected to be high during the r and α phases but low during K . Resource availability is high during α and K phases, but low during r and Ω . Connectedness of control variables is maximal near the end of a K -phase.

The AC and DPE concepts are complementary. The DPE framework aims to incorporate a wider range of CAS from different subject domains and has important parallels to the AC:

- The balance (high connectivity) phase in DPE loosely corresponds to the $r - K$ transition in AC. This phase is signified by stabilising selection, increasing connectivity, and growing potential for propagation of disturbances.
- The variation (low connectivity) phase in DPE loosely corresponds to the $\Omega - \alpha - r$ transition in AC. This is a phase of innovation and re-organisation of underlying networks.

Notably, connectedness in AC refers to the richness of interactions of control variables [Holling, 2001; Gunderson & Holling, 2002]. In fact, however, there may be several interaction networks with different connectivity regimes within a system at any one time. For instance, species in food webs and populations in landscapes form interaction networks that act simultaneously on the same groups but may have very different topologies. The network of control variable interactions may thus have very different properties from the interaction network of specific system components.

In summary, the DPE framework encapsulates many ideas from SOC and AC. It adds to their explanatory power, for instance by providing network-theoretically grounded explanation for power-laws in some non-critical systems and by extending the applicability of the balance-variation cycles concept beyond socio-ecological systems.

Chapter 15:

Conclusions

This thesis has identified common mechanisms in which connectivity of underlying networks facilitates emergence and self-organisation in complex systems. It has distilled these mechanisms into a theoretical framework, DPE, and applied the findings to elicit new insights into concrete systems including landscape ecosystems and social networks. Several directions for future research arise from these findings. New applications and extensions of introduced simulation techniques can provide further insights into biological evolutionary dynamics. The DPE framework itself can be extended to a wider class of networks and further formalised. In this chapter, I summarise the main research results of this thesis and discuss some future research directions in detail.

The material this chapter is based on the following publications:

-  G. Paperin, D. Green, S. Sadedin (2010):
Dual Phase Evolution in Complex Adaptive Systems.
Journal of the Royal Society Interface. Accepted for review. Ref: [Paperin et al., 2010].
-  G. Paperin, S. Sadedin (2009):
Towards formalising the theory of Dual Phase Evolution.
10th European Conference on Artificial Life (ECAL'09). Ref: [Paperin & Sadedin, 2009b].

In this thesis I set out to investigate the role played by networks in natural and artificial systems. Networks are inherent in all complex systems. Patterns and processes that occur within these underlying networks play a major role in the emergence of system wide features ([chapter 5](#)). Perhaps the most widely felt network processes are changes in edge density, especially the connectivity avalanche ([section 5.3](#)). As discussed in [chapter 8](#), phase changes occur in complex systems of virtually every kind. Crucially, those phase changes can typically be traced to connectivity avalanches within underlying networks.

Several important conclusions stem from this thesis: Recurrent phase changes in network connectivity play a key role in emergence and self-organisation within many kinds of systems. This thesis reveals several mechanisms that can lead to connectivity phase transitions. These mechanisms include external disturbances, slow forcing resulting in crossing of attractor basin boundaries, as well as internal feedback. The two connectivity phases usually display the features that characterise Dual Phase Evolution: a phase of apparent balance and stabilising selection, and a phase of vivid variation and exploration.

Interpreting processes as DPE reveals new, useful insights. For instance, studies in [part IV](#) reveal mechanisms that may be responsible for the punctuated nature and continuity of landscape evolution.

DPE also suggests new methodologies. For instance, the generative models in [part V](#) are a result of applying abstract DPE processes to concrete problems. These models are capable of generating networks that resemble several hitherto poorly understood topologies found in social and economical systems.

An extended summary of the key contributions of this thesis is already provided in [section 1.3](#). Here, I concentrate on the implications of these findings and on the directions for future research they imply.

15.1. Directions for future research

15.1.1. Biological modelling

In response to an analysis of simulation results in [chapter 10](#), in [chapter 11](#) I took a detour from the main network-theoretical theme of this thesis and proposed a genetic model that combines some advantages of previous models. This genetic model was applied in an individual-based simulation that investigates the maintenance of biological diversity under migration ([section 11.4](#)).

Simulation results implied that HFL genetics not only represent a strong factor in maintenance of reproductive isolation, but also that HFL genetics may facilitate the evolution of social / sexual selection through reinforcement. An extension of the simulation model that incorporates mechanisms that would permit such evolution (e.g. capability for selective mate choice) can test this hypothesis.

Another key observation in [section 11.4](#) was emergence of reproductively isolated hybrid populations. Such populations did not persist due to low population numbers in comparison to the main model populations, however extending the model by including habitat niches that may be occupied by hybrids can allow hybrid populations to grow. Such an extension will permit to explore scenarios of speciation through hybridisation.

Another straight-forward application of the genetic model is to incorporate it into the larger-scale simulation discussed in [chapter 10](#) in order to test the earlier hypothesis that HFL genetics will reduce the dependence of emerging diversity on directional selection caused by environmental changes in disturbed areas.

More generally, the proposed genetic model may be used to explore a wide range of biological simulation scenarios. An analysis of the FL of the genetic model of [section 11.3](#) provides detailed information about the FL structure and properties and can thus inform such research.

15.1.2. Evolutionary computation

In [section 11.5](#) I speculated on methods for applying the computational HFL model developed in [chapter 11](#) to approach the problem of premature convergence in genetic and other evolutionary algorithms. I would be excited to see an application of those ideas to concrete optimisation problems.

The DPE framework implies some broad considerations in design of intelligent search algorithms. In [section 8.8](#) I argued that many algorithms implicitly use connectivity phase transitions when mitigating between local and global search, however, most approaches perform this transition only once or a few times. In nature, however, such transitions are ongoing. Some algorithms explicitly use ongoing phase transitions to achieve better results ([section 8.8](#)). Further explicit application of DPE in future algorithm design may lead to improvements in optimisation techniques, particularly in areas where stable and adaptive approaches are more important than speed and precision, such as cases where objectives or constraints change over time.

15.1.3. Network models

In [section 12.2](#) I presented only preliminary results of a simulation model of resource flow that applied to a diverse range of systems such as economies or trophic food-webs. Those results exemplify the emergence of connectivity phase transitions from internal feedbacks. The presence of these complex patterns in the model indicates that further work on that

model may likely result in new general insights about self-organisation in systems such as the above.

In [section 13.2.1](#) I discussed an abstract social network model. Simple changes in model architecture (number of attributes per individual / node and number of possible values per attribute) resulted in different network topologies, and all such topologies can also found in real social networks. Research relating empirical studies of communities to further experiments with similar models may reveal insights into the relationship between different attributes of human behaviour and different social interaction structures. Such insight may allow predictions about group dynamics on large scales. For instance, what social settings lead to what kind of group structures?

[Section 13.2.2](#) concentrated on the emergence of modular structures in an abstract network model of business interactions and revealed qualitative insights into the mechanisms behind the emergence of modularity. A comparison between model results and empirical data may allow to extend the model to allow qualitative predictions about formation of co-operative communities.

The work presented in [section 13.3](#) introduces generative models for scale-free networks of constant size that may correspond to processes found in nature. Results show that in the emerging structures, low-degree nodes are under-represented compared to an ideal scale-free topology. However, results indicate two directions for future investigation. First, preliminary experiments indicate that allowing self-connections in the model increase the number of low-degree nodes relative to average-degree nodes, thus improving the power-law fit. Although self-connections do not have an immediate interpretation within the natural processes that inspire the model, a thorough investigation of self-connections in the model may reveal useful insights. Second, results of the enhanced model ([section 13.3.4](#)) show that the fit to a power-law distribution improves for larger network sizes. This prompts for an investigation of the enhanced model for even larger network sizes of many thousands of nodes.

Finally, an analytic examination of the generative model from [section 13.3](#) can assert the true distribution of node degrees in the resulting network in the limit of a large network size. Such investigation may also provide insights into the relationship between the rules for choice of interaction partners and the resulting network topology.

15.1.4. DPE theory

15.1.4.1. DPE-interactions with external system environments

So far I have concentrated on discussing connectivity phase transitions in networks of interactions *within* systems. How can DPE be useful in considering environmental changes that take place *outside* the system?

Processes can affect system behaviour by affecting the system directly, or by affecting its environment. Notably, the distinction between a system and its environment is not always obvious. Since complex systems are not closed ([section 4.3](#)) and constantly interact with their environment, the system boundaries are often not an intrinsic property, but an attribute assigned for the purpose of a particular theory or study. For instance, ecosystems can either be studied in isolation or as an integral part of the wider environment (e.g. a pond is often treated as a closed system, but there is always some interaction with the outside world – birds flying in, water-borne nutrients etc.; see also [section 4.3](#)). Nevertheless, assuming some clear distinction between the studied system and its environment is an overwhelmingly useful methodology [Allen & Starr, 1982; Coveney & Highfield, 1996].

An obvious approach is to consider the density of the interaction network between a system and its environment. However, this effectively results in widening the system boundaries and applying the previous methodology. This may not always be possible, for instance when information about the components of a system environment and their interactions are not sufficiently available.

Another approach, however, is to extend the ideas about network connectivity beyond component interactions. In particular, while topological properties of underlying *interaction networks* can drive system dynamics, the topologies of *state transition networks* are an emergent attribute of a system. Similarly to system boundaries, state transition networks are not actually existing tangible system properties, but abstract constructs created by humans to describe and understand system behaviour (section 5.2). In other words, topology of intrinsic *component interaction networks* influences processes that drive system dynamics, and *state transition networks* describe these emergent system dynamics.

For example, consider the landscape model discussed in chapter 10. There, when landscape connectivity is high (i.e. the density of gene flow *interactions* is high), changes in ecosystem dynamics are infrequent (i.e. the connectivity of the *state transition* network is low). Conversely, disruptions in connectivity of gene flow interactions coincide with frequent changes in model ecosystem development (i.e. high connectivity of the state transition network).

More generally, DPE relates some topological properties of interaction networks to characteristic system dynamics, and thus to characteristic properties of state transition networks. For instance, phases of high connectivity of component interactions often correspond to balance, uniform dynamics and thus to sparser state transition networks (*balance phase*). Conversely, DPE phases of local interactions and enhanced evolutionary exploration often correspond to phases of dense connectivity in state networks with increased transition possibilities between attractor basins (*variation phase*).

Understanding how internal system processes as well as external environmental changes affect the connectivity of state transition networks of systems can help to describe, model and understand system behaviour. Thus, the future research on application of the DPE framework to state transition networks may likely result in new insights into the general patterns of emergence and self-organisation.

When using state transition networks to model systems and understand their behaviour, it is important to consider the origins of changes in state networks. In particular, the way in which a state transition network of a given system is affected by internal system changes versus changes of the external environment. A state transition network describes the possible system states (the nodes) and the system potential to change between the states (the links). Thus, when a system changes, it “traverses” one or more links of the state network to enter a new state (section 5.7). In doing so, the system may “arrive” at a region of the state network with local connectivity properties that are different from the region where the previous state was located. Conversely, when an environmental change does not directly affect the system state, it may arguably be inappropriate to say that a system traversed a link in the state network. Such change can affect system dynamics by manipulating the state transition network directly and adding or removing some links or nodes thus shaping the topology of possible future state changes. Both interpretations may have similar consequences for system behaviour, however, explicitly differentiating these cases may be useful for uncovering underlying mechanisms behind complex phenomena. Crucially, state transition networks are a modelling methodology and not a naturally occurring phenomenon. Thus, the most appropriate approach of relating changes in system dynamics to the topology of a state network and the system location in it is dictated by a particular theory or study.

15.1.4.2. Other directions

Several other aspects of the DPE framework call for further investigation. For instance, should DPE be seen as a single underlying process or as a family of related classes of processes? An important practical problem is how to identify DPE processes from empirical observations. In particular, how can they be distinguished from other phenomena that can drive the emergence of complexity, such as criticality? A necessity to observe two distinct phases of system dynamics can be a starting point for such investigation, but detailed research is necessary to address this issue.

Another important matter is the role of distinct phases dominated by local or global interactions. What phenomena depend on such phases with a well-defined transition, and what can also be caused by stochastic occurrences of local and global interactions, such as in

small-world networks? The network models in [part V](#) showed that in some cases, the relative duration and frequency of the two phases can affect the emergent behaviour. Can this dependency be formulated in general terms for a wider class of systems?

The landscape model in [chapter 10](#) implicitly included at least two interaction networks (the network of reproductive interactions and the network of free migration paths). However, these networks were essentially inverse of each other and thus closely related. Other models in this thesis concentrated on a single interaction network. However, real systems usually consist of many networks representing different kinds of interactions ([section 7.3](#)). Models explicitly investigating effects of DPE processes acting on several networks within the same system are necessary to shed light on feedbacks and other interactions between multiple underlying networks in systems.

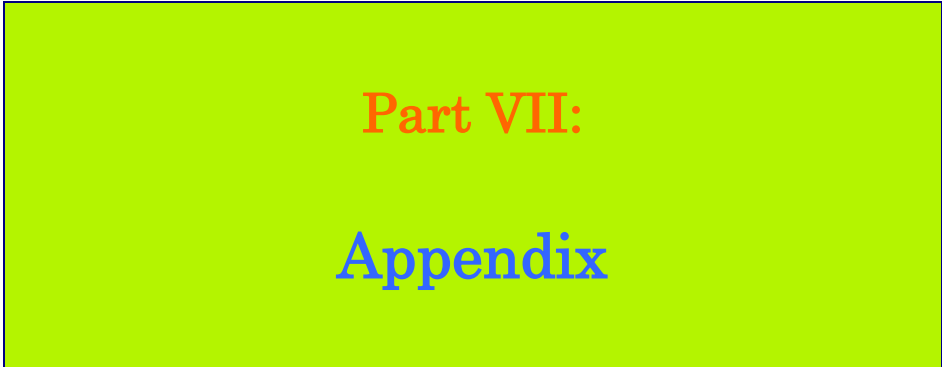
The term *Dual Phase Evolution* has been initially coined by Green [e.g. Green, 2000; Green et al., 2000; Green & Sadedin, 2005; Green et al., 2006] to generally refer to recurrent transitions between balance and variation, exploitation and exploration, connectedness and separation in the development of systems. Here, I have re-introduced the concept as a well-defined framework rooted within network theory. However, this is only a first step towards a formal theory of DPE. Further theoretical, empirical and simulation studies are necessary to provide a formal theory with analytic, predictive power. I will be extremely excited to see future research extending my work in this way.

15.2. Verdict

There are a number of key questions that are of wide-spread interest to researchers in complex systems [Bedau et al., 2000]. What are the mechanisms behind emergence and self-organisation? Why are scale-free structures ubiquitous, even in non-critical systems? How does perpetual novelty arise in adaptive systems? This thesis shows that DPE-phase-transitions present one mechanism involved in these complex phenomena.

Explicitly characterising common mechanisms aids cross-fertilisation between disciplines. Network theory and related methods from statistical physics found useful applications in evolutionary biology ([part IV](#)) and social sciences ([part V](#)). Other areas where DPE phenomena appear to occur are as diverse as cognitive neuroscience ([section 8.4](#)), geomorphology, cosmology ([section 8.1](#)) and computer science ([sections 8.7 and 8.8](#)). Such areas may also benefit from network-theoretical techniques.

As a science, complexity theory still suffers from its recent origins. Related phenomena are still interpreted using a wide variety of terminology, concepts and methods that are usually specific to particular fields of research: “*Perhaps there can be a general theory of complex systems, but it is clear we don’t have one yet.*” [Gordon, 2007]. Identifying common processes, such as Dual Phase Evolution, describing them in real systems, and generalising them in models, is an essential step towards the goal of understanding complexity, and the major contribution of this thesis.



Part VII:

Appendix

The appendices provide additional technical details, indices for quick reference and a list of cited literature.

Appendix A: Technical details of the simulation model of adaptive radiation in landscapes.

Additional details of the simulation model presented in [chapter 10](#) are provided. The reader is advised to consult [chapter 10](#), particularly [sections 10.1.1](#) and [10.1.2](#) before reading this appendix.

Appendix B: List of Figures.

This appendix provides a list of figures for quick reference.

Appendix C: List of Tables.

This appendix provides a list of tables for quick reference.

Appendix D: Index.

This appendix provides an alphabetical index of terms and abbreviations for the entire thesis.

Appendix E: References.

This appendix provides a complete list of cited literature.

Appendix A:

Technical details of the simulation model of adaptive radiation in landscapes

In this appendix I discuss additional details of the simulation model presented in [section 10.1](#). Where applicable, I provide results of the sensitivity analysis. The reader is advised to consult [sections 10.1.1 and 10.1.2](#) before reading this appendix, as those sections provide an essential context.

In this appendix I discuss some additional details of the simulation model presented in [section 10.1](#). The reader is advised to consult [sections 10.1.1 and 10.1.2](#) before reading this appendix, as those sections provide an essential context. Where applicable, I provide the results of the sensitivity analysis. Additional details on model sensitivity are discussed in [sections 10.1.2.3 and 10.1.4](#), where they provide a clearer context for the experimental results. In cases where qualitative model behaviour was not found to depend on specific parameter values, parameter were chosen as large as possible (to investigate dynamics of a larger system) while still considering the runtime performance of the simulation.

A.1. Model configuration

A.1.1. Landscape

The model environment is a landscape represented by a 2-dimensional, rectangular grid of 100×100 cells. The shape of the model universe is a torus. This means that the top row of cells is adjacent to the bottom row, and the most left column of cells is adjacent to the most right column. This approach helps avoiding any boundary effects, as there are no boundary cells and each cell has the same number of neighbours (8).

Each grid cell has a carrying capacity of 4 individuals. Carrying capacity is chosen as large as possible to allow for larger interbreeding groups (see [section A.1.5](#), Reproduction). However, larger carrying capacity significantly slows down the runtime performance of the model. Tests have shown that varying the carrying capacity between 5 and 8 does not affect the qualitative model dynamics. Larger carrying capacity was not possible without reducing the size of the landscape grid, which, however, must be large to allow disaster scenarios to clear sufficiently spacious areas (see [section 10.1.4](#)). As hardware becomes faster, it would be interesting to compare the current model results with a similar simulation that uses larger carrying capacities of tens and hundreds of individuals per cell.

There are 60 habitat niches equally present in each cell. A niche is represented by a string of $L = 20$ bits and a *condition* $\in [0, 1]$ parameter. Each bit represents the requirement, that a particular trait must be present (1) or absent (0) in an individual in order to be well adapted to that niche. The niche condition describes the maximum fitness that an individual can obtain by being fully adapted to that niche.

A.1.2. Individuals

The model landscape is occupied by haploid individuals. An individual is represented by its genotype and there is no genotype-phenotype distinction in this model. Previous similar models (i.e. [Gavrilets & Vose, 2005] as well as several described in [Gavrilets, 2004]) use an additive quantitative genetic trait model that incorporates a genotype-phenotype distinction. In [subsection 10.2.2](#) I show that for the current purposes, the employed approach is fully equivalent, and computationally more efficient, to the additive quantitative

trait model used by Gavrillets and Vose [2005]. The equivalence holds as long as the number of traits in the current model equals to *number of traits* $\times$ *number of loci per trait* in their model.

An individual's genotype is represented by a string of $L = 20$ bits. Each bit denotes the presence (1) or absence (0) of a certain traits. The traits represented in genotypes correspond to the traits represented in niches at the respective positions.

The degree of adaptation of an individual λ to any given niche M is defined by the hamming distance of the respective bit strings. It is assumed that individuals exist within a niche to which they are best adapted. Thus, the *habitat niche* M_λ of an individual λ is:

$$M_\lambda = \arg \min_{M_i} \text{dist}(\lambda, M_i)$$

where M_i are all available the habitat niches and $\text{dist}(\lambda, M_i)$ is the hamming distance between the bit-strings representing the individual λ and the niche M_i respectively.

The raw fitness $\text{fitness}_{\text{raw}}(\lambda)$ of an individual λ is based on that individuals adaptation to its habitat niche linearly scaled into the unit interval such that $\text{fitness}_{\text{raw}}(\lambda) = 1$ denotes perfect adaptation:

$$\text{fitness}_{\text{raw}}(\lambda) = 1 - \frac{\text{dist}(\lambda, M_\lambda)}{L}$$

Note that an individual may be equally well adapted to more than one niche. This, however, does not affect the raw fitness of an individual. Other consequences arising from the possible existence of multiple equi-distant habitat niches are discussed in [section 10.2.4.5](#) in more detail. In short, this ambiguity is resolved such that a niche is chosen arbitrary, but always the same niche is chosen for the same genotype.

Recall that every niche M_i has a parameter condition_{M_i} that describes the living conditions offered to the individuals adapted to that niche. Thus, the actual fitness fitness_λ of an individual λ is:

$$\text{fitness}_\lambda = \text{fitness}_{\text{raw}}(\lambda) \cdot \text{condition}_{M_\lambda} = \left(1 - \frac{\text{dist}(\lambda, M_\lambda)}{L}\right) \cdot \text{condition}_{M_\lambda}$$

Note that the habitat niche of an individual is chosen to optimise the individual's adaptation to its habitat niche, with disregard to its condition. If a habitat niche has a poor condition, the actual fitness of an organism may be lower than if the individual existed in a different niche with better conditions. However, in nature, the individuals exist in habitats to which they are adapted, despite the fact that other habitats may in theory provide better living conditions. Over long periods, selection will drive the future populations towards niches with better conditions.

A.1.3. Lifecycle

The overall model lifecycle is as follows:

Initialisation: initialise landscape, niches, and initial population.

Repeat for 40000 generations {

Measurement: collect measurements and statistics.

Environment update: update niche conditions, simulate disasters if applicable.

Reproduction: perform mating, replace old generation with the new generation.

Selection: ensure cell carrying capacity by selecting survivors according to fitness.

Dispersal: probabilistically migrate individuals to neighbouring cells.

}

Analysis has shown that running the model for longer than 40000 generations does not produce any new model dynamics. The details of the lifecycle steps are considered in more details below.

A.1.4. Initialisation

The model is initialised with an empty homogeneous landscape grid of size 100×100 and with 60 niches (30 “normal” + 30 “disaster”). In the bit-strings representing the niches, each bit is independently set to 0 or 1 with equal probability. For normal niches, *condition* is set to 1; for disaster-niches, *condition* is set to 0.2. The qualitative model behaviour was not found to be sensitive to these values as long as $1 \geq \text{condition} \geq 0.8$ for normal niches and $0.05 \leq \text{condition} \leq 0.3$ for disaster niches. Section 10.2 contains a discussion of how the number of niches may affect model dynamics.

Two random founder individuals are created by setting each bit in the bit-strings representing the genotypes independently to 0 or 1 with equal probability. Several clones of these founder individuals are placed at randomly and independently selected cells. When a single founder individual is used, the population fails to speciate before the entire grid is populated in over 40% of runs due to lack of genetic diversity. In cases where speciation does occur early, model dynamics are largely similar to scenarios with two founder individuals.

A.1.5. Reproduction

Any individual can mate with any individual in the same cell and the same cell only, and mating in each landscape cell occurs independently from all other cells. Each individual in the cell is selected as a *mother* exactly once. If there is only one organism in a cell, it engages in hermaphroditism acting as *mother* and *father* at the same time. If there are more than one individual in a cell, for each *mother*, a *father* is selected with equal probability out of all remaining individuals. Generations are non-overlapping, i.e. offspring do not mate with parents and the new population replaces the old population once all mating in a cell is complete.

Inline with the model of Gavrillets and Vose [2005], the number of offspring for each mating pair is selected randomly from a Poisson distribution with the mean $\lambda = 5$. Sensitivity tests have shown that if λ is very small, new generations may have consistently fewer individuals than previous generations, leading to frequent extinctions. If λ is large dynamics are not affected, but the runtime performance deteriorates as selection must be performed on a large number of individuals. The value 5 ensures sufficient diversity in offspring with acceptable runtime performance.

Each of the offspring is generated through free recombination of the parents genotypes. This means each bit (trait) in the offspring genotype is inherited from *mother* or *father* with equal probability. The offspring genotype mutates with a probability of 10^{-5} per bit. The mutation rate in natural genotypes lies in the range 10^{-4} to 10^{-6} per locus depending on the type of locus [Russell, 1994]. The model was found not to be sensitive to variations of the mutation rate within this range.

Individuals mate regardless of their habitat niche. The habitat niche of every offspring is chosen on the basis of its genotype as described in section A.1.2 without consideration of the parents. As the number of possible niches is large ($2^L = 2^{60} \approx 1.15 \times 10^{18}$) and the number of actual niches is small (60), the density of well-adapted genotypes is low. Thus, offspring of two individuals well adapted to two different niches are expected to be poorly adapted to either niche. Such offspring will have a poor fitness and will unlikely survive under selection. Therefore, populations adapted to different niches likely exhibit postzygotic reproductive isolation. Thus, in this model, populations adapted to different niches are assumed to correspond to different species. Note, however, that this assumption is critically considered in section 10.2 (in particular, sub-section 10.2.4.6).

A.1.6. Selection

If after the reproduction stage the number of individuals in a cell does not exceed the carrying capacity of 4 individuals, all individuals survive. Otherwise 4 survivors are probabilistically selected according to their fitness and the rest of the individuals is removed. Formally, selection is performed as follows (note, this algorithm is optimised in the actual model implementation to handle small values of $fitness_\lambda$ more efficiently):

```
selection(cell) {  
  
    maxCellCapacity = 4;  
    oldPopulation = set of all individuals in cell;  
    survivorPopulation = empty set of individuals;  
  
    while (sizeof(survivorPopulation) < maxCellCapacity) {  
        choose an individual  $\lambda$  from oldPopulation with equal probability;  
        choose a random number  $R$  uniformly distributed in  $[0, 1[$ ;  
        if ( $R < fitness_\lambda$ ) {  
            remove  $\lambda$  from oldPopulation and add it to survivorPopulation;  
        }  
    }  
    replace all individuals in cell with survivorPopulation;  
}
```

Note that this approach where individuals are *selected for survival* is different from the approach commonly used in genetic algorithms, where all individuals survive and then compete to be *selected for reproduction*.

A.1.7. Dispersal

Each individual may migrate to one of the 8 neighbouring cells with the probability 0.1. An individual does not migrate more than once per generation. If an individual migrates, then one of the 8 neighbouring cells is chosen as destination with equal probability. All individuals migrate concurrently. Lower migration rates reduce interbreeding between cell populations and lead to slower model dynamics. High migration rates increase interbreeding and gene flow across the landscape. This reduces speciation and diversity in the model, as populations of high-fitness individuals tend to take over the landscape before other populations can adapt to their niches and increase their own fitness.

A.2. Analysis of the basis run scenario

The initial founder genotypes are random and not yet adapted to any niche. Their mean fitness and the fitness of their offspring are similar on average, implying no significant reproductive isolation between niche populations. The habitat niche of the early offspring generations has little correlation to the ancestors' niches which results in the observed initial burst of adaptive radiation. While the landscape is not fully populated, selection is weak, which further facilitates adaptive radiation, as even deleterious mutations can spread stochastically.

Once the complete grid has been populated, selection pressure increases. Niche populations that by chance adapted quicker and achieved higher average fitness than other populations begin taking over the landscape until most niche proportions fall to insignificant levels. The remaining "dominant" niche populations occupy large continuous areas of the landscape. In some model runs the average fitness of the dominant species is similar, but in a number of runs there are fitness differences. In such cases, the dominant species with the higher fitness is likely to eventually take over the entire landscape (Figure A.1). However, on some occasions, once lower-fitness populations acquire a large stable habitat, they appear to maintain it. This somewhat unintuitive phenomenon is known to occur in nature [Green, 1989]. Although the *average* fitness of a model species may be consistently higher than the average fitness of another species, the variance in the *actual* fitness of specific individuals may result in similar fitness values of individuals from different species at a particular location. Because selection is local, competition is limited to small border regions between habitats. As a result, significant fitness advantage at habitat borders may fail to occur, resulting in roughly stable habitats.

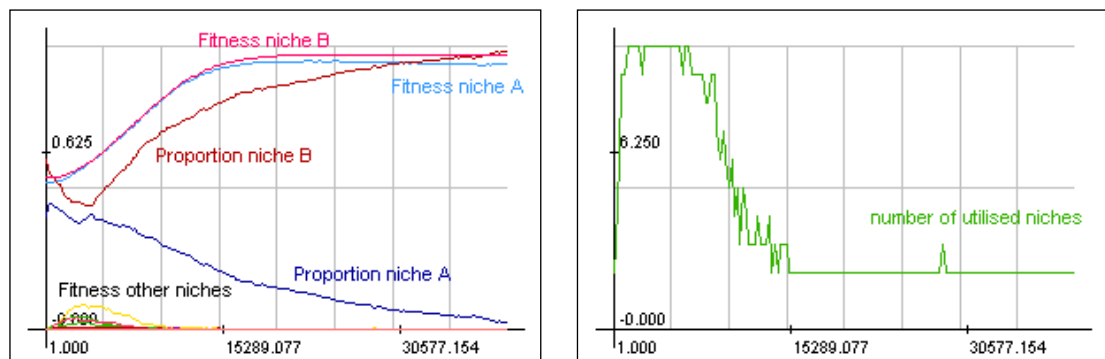


Figure A.1. Typical simulation run of the basis scenario. An initial burst of adaptive radiation is followed by two species quickly taking over the entire population. One of the two dominant species (blue) then slowly declines. After 40000 generations it is about to die out.

Left: X-axis: generation number. Y-axis: average niche fitness and niche proportion (brown graph: proportion of the 1st dominant species; red graph: average fitness of the 1st species; dark blue graph: proportion of the 2nd dominant species; light blue graph: average fitness of the 2nd species).

Right: X-axis: generation number. Y-axis: number utilised niches.

(The rounding artefacts on axes labels are due to an issue in an early version of LiveGraph that was used to render this figure (chapter 3).)

While the number of dominant species (i.e. occupied niches) typically ranges between 1 and 3, the total number of species tends to remain higher. Analysis shows that these "weak" species typically occupy the border regions between dominant species, which implies that they are likely to be hybrids. Organisms that arise through recombination of 2 parents from different species may have a genotype that is closer to a distinct third niche. A similar effect is observed in nature [Spaak & Hoekstra, 1997; Rosenthal et al., 2002]. In many model runs, hybrid individuals inhabiting a niche different from both their parents occur sporadically and do not manage to establish a stable population. However, during some runs, stable hybrid zones occur. Such zones are regions along the habitat borders of the parent species occupied by hybrid individuals that are adapted to an own unique niche (Figure A.2). As a result, hybrid zones effectively separate the habitats of the main species.

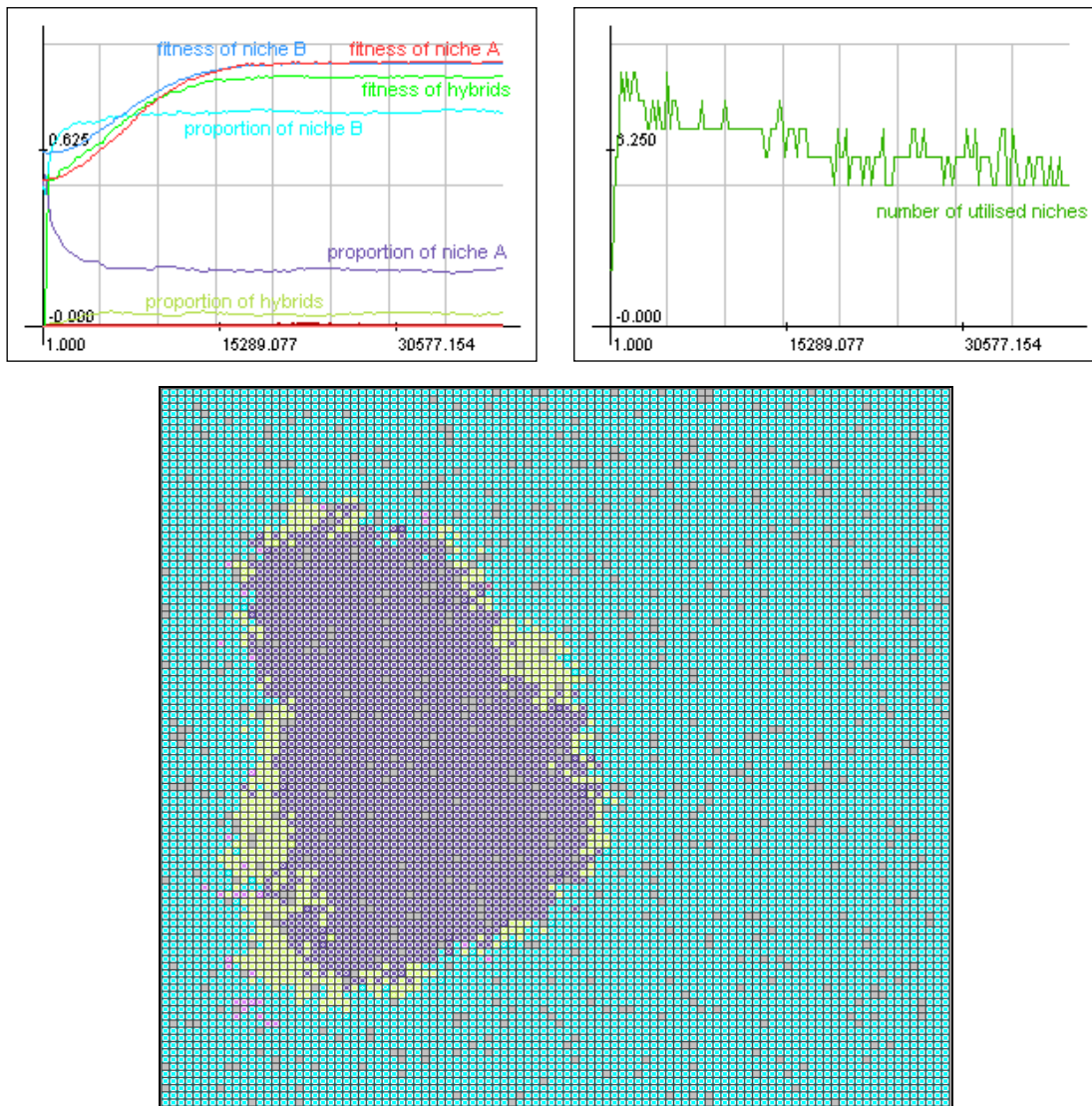


Figure A.2. Typical basis run with a developed tension zone. Each cell is coloured according to the majority-species in that cell. Two dominant species control the population. The proportion of the **cyan** species is stable around 0.75; the proportion of the **violet** species is stable around 0.2. Both their fitness values (**light blue** and **red** graphs respectively) are very similar and relatively high. The **green** hybrid species retains a stable proportion around 0.04, although its **fitness** is significantly lower. Other hybrid species sporadically emerge, but do not manage to establish a stable population.

Left: X-axis: generation number. Y-axis: average niche fitness and niche proportion.

Right: X-axis: generation number. Y-axis: number utilised niches.

Bottom: A landscape snapshot of this simulation run at generation 30000. The tension zone is clearly visible.

(The rounding artefacts on axes labels are due to an issue in an early version of LiveGraph that was used to render this figure ([chapter 3](#)).)

In nature, there are two main mechanisms behind stable hybrid zones. The first occurs when hybrids adapt to some environmental condition that is local to the hybrid zone, thus gaining a fitness advantage to the parent species in that geographic area. The second mechanism occurs when the hybrids have a lower fitness than the two hybridising species and the hybrid zone is maintained by a balance between dispersal and selection against hybrids [Barton & Hewitt, 1985; Szymura & Barton, 1986; Stein & Barrington, 1990]. Hybrid zones of the latter kind are termed *tension zones*. There are indications that most hybrid zones in nature are of this type [Barton & Hewitt, 1985].

As the model environment is homogeneous, there cannot be any local adaptation, and the hybrid zones in the model are tension zones. This conclusion is confirmed by fitness measurements: the average fitness of the hybrid species is typically lower than the average fitness of the hybridising species. The proportion of tension zone populations (when they occur) relative to total modal population is small but stable around 0.1 to 0.05.

The observation of stable habitats of lower-fitness populations and of tension zones in the model is particularly interesting because these patterns were not originally expected. As similar emergent patterns are found in nature, these observations indicate that the model appropriately captures relevant dynamics of natural landscapes and supports the validity of experiments performed with the model (chapter 2 and [Grimm et al., 2005]).

Appendix B:

List of Figures

Figure 1.1. Thesis structure.....	16
Figure 3.1. LiveGraph framework for exploratory data analysis.	29
Figure 3.2. LiveGraph download statistics.	31
Figure 5.1. Interaction and information networks.	46
Figure 5.2. Critical connectivity phase transition in networks.	47
Figure 5.3. Unpredictable dynamics near critical threshold.	47
Figure 5.4. Connectivity avalanche in networks.	48
Figure 5.5. Some common structures in networks.	50
Figure 6.1. Adaptive fitness landscape metaphor.	56
Figure 6.2. Local maxima can become saddle points in higher dimensions.....	58
Figure 6.3. Percolation on a square lattice.....	60
Figure 6.4. Holey fitness landscape (HFL) concept.....	61
Figure 7.1. The mechanism of Dual Phase Evolution.	67
Figure 9.1. DPE phase transition cycles from the perspective of CAS dynamics.	86
Figure 10.1. Typical model runs.	92
Figure 10.2. Measurement results of typical runs.	93
Figure 10.3. Niche turnover and diversity measured at the end of each run.....	94
Figure 10.4. Schematic representation of different disaster sizes.	96
Figure 10.5. Frequency distribution of the closest-goal distance.....	104
Figure 10.6. Proportion of hybrids with lower fitness than both parents.	108
Figure 11.1. Approach to independent random selection of viable genotypes for a HFL suitable for computer modelling.....	117
Figure 11.2. Phase 1 of algorithm for selection of viable genotypes for a HFL computer model.....	119
Figure 11.3. Phase 2 of algorithm for selection of viable genotypes for a HFL computer model.....	119
Figure 11.4. An example of a model genotype.	128
Figure 11.5. Using Markov Clustering (MCL) for determining reproductively isolated groups.	130
Figure 11.6. Evolution on a 1×2 grid for different migration rates.	132
Figure 11.7. Evolution on a 2×2 grid for different migration rates.	133
Figure 11.8. Average genetic divergence DB at neutral loci between the individuals of the entire population in neutral evolution (without HFL).....	134
Figure 11.9. Reinforcement and evolution of sexual selection as a result of selective pressure to avoid mating with immigrants.....	135
Figure 11.10. Hypothetical approach to using HFLs to combat premature convergence in genetic algorithms.....	138
Figure 12.1. A typical simulation run.	147
Figure 12.2. Number of edges in the copycat model (part A).	150
Figure 12.3. Number of edges in the copycat model (part B).	151
Figure 13.1. The effect of local and global interactions on the formation of network structure.....	157
Figure 13.2. DPE process leading to a modular network topology.	158
Figure 13.3. Sensitivity analysis of modularity in the network after 5000 simulation iterations.....	158
Figure 13.4. Typical networks resulting from the model with several binary attributes per node.....	159

Figure 13.5. Typical network resulting from the simulation model (left) and an empirical social network of romantic relationships in a high school (right).	160
Figure 13.6. Edge density and number of nodes.	162
Figure 13.7. Emergent modularity in model network (basis scenario).	162
Figure 13.8. Model behaviour with $L = 3$ attributes with one attribute active at any given time (scenario 1).	165
Figure 13.9. Model behaviour with forced edge removal (scenario 2).	166
Figure 13.10. Model behaviour with simultaneous effects of all node attributes (scenario 3) with forced edge removal (top) and without forced edge removal (bottom).	169
Figure 13.11. Typical network resulting from a model with simultaneous effects of all node attributes (scenario 3).	169
Figure 13.12. Critical transition of the node degree distribution from random to scale-free.	172
Figure 13.13. Goodness of fit and estimated exponent of the node degree distribution to a power-law.	174
Figure 13.14. Actual node degree distribution and its logarithm.	175
Figure 13.15. Goodness of fit and estimated exponent of the node degree distribution to a power-law in the updated model.	177
Figure 13.16. Actual node degree distribution and its logarithm in the updated model.	178
Figure 13.17. Edge density and number of nodes.	179
Figure 13.18. The two different classes of dynamical behaviour do not depend on the initial edge density, but only on the network topology.	179
Figure 13.19. The dependence of the attractor edge density D_s on the network size N in the initial model.	180
Figure 13.20. The dependence of the attractor edge density D_s on the network size N in the updated model.	180
Figure 13.21. Example of a network obtained through the updated model.	181
Figure 14.1. Self-Organised Criticality vs. Dual Phase Evolution.	187
Figure A.1. Typical simulation run of the basis scenario.	201
Figure A.2. Typical basis run with a developed tension zone.	202

Appendix C:

List of Tables

Table 3.1. Agent-based modelling platforms.....	27
Table 10.1. Niche turnover and diversity measured at the end of each run.....	94
Table 10.2. Frequency distribution of the closest-goal distance d_{min} on the model fitness landscape for $L = 10$ loci, averaged over 20 runs.	103
Table 10.3. Frequency distribution of the closest-goal distance d_{min} on the model fitness landscape for $L = 15$ loci, averaged over 20 runs.	103
Table 10.4. Frequency distribution of the closest-goal distance d_{min} on the model fitness landscape for $L = 20$ loci, averaged over 20 runs.	104
Table 10.5. Frequency distribution of 5000 random genotypes according to the corresponding reach cluster sizes.....	105
Table 10.6. Frequency distribution of 5000 random genotype pairs according to the length of the shortest adaptive walk between the genotypes that contains non-deleterious mutations only.....	106
Table 10.7. Frequency distribution of genotypes according to the number of better fitness neighbours.	106
Table 10.8. Frequency distribution of genotypes according to mG – the number of equidistant closest-goal sequences.	107
Table 11.1. Percentage of local optima genotypes in respect to all viable genotypes.....	121
Table 11.2. Number of possible adaptive mutations per genotype.	122
Table 11.3. Measures of possible adaptation trajectories.....	122
Table 11.4. Proportions of differentiated and undifferentiated genotypes.	124
Table 11.5. Average closest-niche distances of genotypes.....	124
Table 11.6. Proportion of viable offspring of parent genotypes from the name niche.....	125
Table 11.7. Proportion of viable offspring of parent genotypes from different niches.....	125
Table 13.1. Parameter values used in the experiments in section 13.2.1 resulting in modular network topologies.	156
Table 13.2. Parameter values used in the experiments in section 13.2.1 resulting in circuit and chain-dominated network topologies.....	159
Table 13.3. Parameter values used in the basis scenario experiments in section 13.2.2.	163
Table 13.4. Probabilities of nodes to be selected as start or end nodes during local and global interactions in the original and the updated models.	176

Appendix D:

Index of terms and abbreviations

A

A (variable) 155, 160
AC See adaptive cycle
accommodation 74
adaptive cycle 185, 188
adaptive landscape .. See fitness landscape
adaptive radiation 54, 73, 74
additive quantitative trait model 98
additive trait model See additive
quantitative trait model
aggregate complexity 36
algorithmic complexity 35
allele 56
allopatric speciation 54
annealing 72
Artificial Life 38
assimilation 74
atomic mutation 57
average niche fitness 91

B

backpropagation algorithm 78
backswamp 71
balance phase. See connectivity phase: well
connected
balance vs. variation phase 85
Baldwin effect 79
ballance phase. See connectivity phase: well
connected
bank transaction 75
bifurcation 37
braided river channel 72
Butterfly Effect 35

C

capitalism 75
CAS See complex adaptive system
catastrophe theory 37
cellular automation See cellular automaton
cellular automaton 185
cellular automaton 38
climate change 73
closest-goal See closest-niche
closest-goal distance See closest-niche
distance
closest-niche 99
closest-niche distance 99
co-authorship network 76
co-evolution 85
cognition 74
communication network 76
complex adaptive system 36
complex system 33, 36
complexity 35
complexity science 35

component interaction network See
interaction network
connected phase See connectivity
phase: well connected
connectivity phase
poorly connected 66, 192
well connected 66, 192
consciousness 75
copycat 148
critical phase transition 38
critical threshold 59
critical transition 148
cybernetics 38

D

deterministic chaos 35, 37
deterministic complexity 35
dimensionality
of fitness landscape 57
of genotype space 57, 60
diploid 56
directional selection 66, 77
disconnected phase See connectivity
phase: poorly connected
dissipative
system 37
dist (function) 98, 99, 113
disturbance 85
external 68, 85
diversification 73, 74
 $d_{\min}$ (function) 99
DNA 73
DPE See Dual Phase Evolution
empirical evidence 71
DPE phase See connectivity phase
dreaming 75
Dual Phase Evolution 65, 66, 86

E

ecological niche See niche
economics 75
economy 75
edge of chaos 38
edge-of-chaos 185
emergent
behaviour 36
emergent phase transitions 76
energy flow 144
environmental change 77
environmental condition 77
epistasis 59
equilibration 74
equilibrium theory 75
evolution by selection
theory of See evolutionary theory
evolutionary computation 38

evolutionary theory..... 37
 external disturbance See
 disturbance:external
 extinction 73

F

feedback process.....68, 76
 fitness 57
 fitness function 57, 114
 fitness landscape..... 37, 38, 55, 57, 77
 high dimensional 58
 low dimensional 58
 metaphor 58
 neutral..... 58
 N-k See N-k fitness landscape
 rugged..... 57
 fitness peak 58
 fitness valley 58
 fixation index 130
 forcing
 external 68
 forrest fires 73
 fractal.....35, 37

G

$\mathbb{G}$ (set) 56
G (variable) 56, 113
 game theory 75
gdist (function)..... 57
 general equilibrium theory See equilibrium
 theory
 general systems theory 38
 genetic algorithm 38, 136
 cellular 78
 genetic distance See genotype distance
 genetic diversity 73
 genetic drift 54, 58
 genetic programming 78
 genotype..... 56
 differentiated 123
 generalist See genotype:undifferentiated
 hybrid 123
 inviable 60, 114
 undifferentiated 109, 114, 123
 viable 60, 114
 genotype distance 56, 57
 genotype space 57
 geomorphology 71
 ghetto 76
 giant connected component .. 60, 116, 117
 glacial cycles 73
 global phase .. See connectivity phase:well
 connected
 globalisation 75
 goal sequence 99
 graph colouring 78
 great deluge algorithm 78

H

habitat
 connected 73
 fragmented 73
 habitat connectivity 73

habitat network..... 73
 habitat niche.....74, See niche
 haploid..... 56
 HFL See holey fitness landscape
 hierarchy..... 168
 holey fitness landscape..... 59, 61, 114
 hybrid 107
 hybrid zone 91, 202

I

information theory 38
 interaction network..... 192

K

keystone species 68
 k-SAT 78

L

L (variable) 56, 99, 113, 120, 155, 164
 Lamarckian evolution 79
 landscape change 72
 landscape connectivity 72
 largest connected component See giant
 connected component
 local optimum 78, 85
 local phase . See connectivity phase:poorly
 connected
 loci See locus
 locus 56

M

market..... 75
 market percolation..... 76
 Markov clustering algorithm 79
 Markov Clustering algorithm..... 129
 mass extinction 74
 MCL See Markov Clustering algorithm, See
 Markov clustering algorithm
 memory 75
 modular network 154, 155, 160
 modularity..... 77, 157, 161, 168
 molecular bonding network 73

N

nearly neutral network 59
 nearly-critical behaviour 186
neighbours (function) 57
 network topology..... 143
 neural network 38
 artificial..... 78
 niche 90, 197
 disaster..... 90
 normal 90
 niche condition 90, 197
 niche proportion 91
 N-k fitness landscape 59
 N-k landscape ... See N-k fitness landscape
 novelty 85
 NP-completeness 78
 number of utilised niches 91

O

one-on-f noise..... 186

one-point mutation... See atomic mutation
 optimisation algorithm 78
 optimisation problem 78, 137

P

panarchy..... 188
 parapatric speciation..... 54
 particle swarm optimisation 78
 PCR..... See Polymerase Chain Reaction
 perception..... 74
 percolation 59, 60
 percolation threshold 59, 67
 perturbation See disturbance
 phase
 change See phase transition
 connectivity..... See connectivity phase
 DPE..... See connectivity phase
 transition 66, 68
 phase transition
 mechanisms for 143
 plant network 73
 Poisson graph 153
 politics..... 75
 Polymerase Chain Reaction 73
 poorly connected phase... See connectivity
 phase:poorly connected
 postzygotic reproductive isolation . 54, 135
 power law 37
 power-law 186
 power-law distribution 76
 preferential attachment..... 170
 premature convergence..... 136
 prezygotic reproductive isolation .. 54, 135
 punctuated equilibrium 37, 74, 186

Q

quantitative trait model..... See additive
 quantitative trait model

R

Random network..... 153
 random network model 38
 repeated annealing 72
 reproductive isolation..... 53, 54, 97, 109,
 127, 135
 postzygotic..... See postzygotic
 reproductive isolation
 prezygotic . See prezygotic reproductive
 isolation
 reproductively isolated group 129
 resource flow 144
 revolution 76
 RI group See reproductively isolated group

S

S (variable) 99, 114, 120
 scale free topology See scale-free network
 scale-free network..... 76, 154, 170, 186
 generative model for 170

 of constant size..... 170
 schema 74
 search
 global 78
 local 78
 search problem 77
 search space..... 77, 78
 second-order cybernetics..... 38
 sediment transportation 72
 selection 85
 selection pressure 77
 self-organisation 37, 38
 self-organised criticality..... 37, 185
 self-regulation 38
 sexual selection 135
 simulated annealing..... 78
 six degrees of separation principle 154
 sleep 75
 sleep-wake cycle 75
 small-world network 154
 small-world property..... 154
 SOC See self-organised criticality
 social consensus..... 76
 social network..... 76, 155
 social sciences 75
 social system 76
 speciation..... 53, 73, 74
 allopatric..... See allopatric speciation
 parapatric See parapatric speciation
 sympatric..... See sympatric speciation
 species 53, 109
 species proportion See niche proportion
 stabilising selection..... 66, 77
 star formation..... 72
 state network See state transition network
 state transition network 77, 192
 support vector machine 78
 sympatric speciation 54

T

tension zone..... 202
 Trees (network structure)..... 153

V

V (set)..... 60, 99, 113, 114
 variation 85
 variation phase See connectivity
 phase:poorly connected
 variation phase (in simulation model) ... 95
 variation vs. ballance phase 85
 varying environment..... 77
 viable (function) 115

W

well connected phase See connectivity
 phase:well connected
 Wolfram class 185

Appendix E:

References

- D. Achlioptas, A. Naor and Y. Peres (2005): "Rigorous Location of Phase Transitions in Hard Optimization Problems". *Nature*. 435: pp. 759-764.
- C. Adami and C. T. Brown (1994). "Evolutionary Learning in the 2d Artificial Life System 'Avida'", *Proceedings of 4th International Conference on the Synthesis and Simulation of Living Systems (ALife IV)*, pp. 377-381. MIT Press.
- E. Alba and B. Dorronsoro: (2008): "Cellular Genetic Algorithms". *Operations Research / Computer Science Interfaces*. Vol. 42. Springer. ISBN 978-0-387-77609-5.
- R. Albert and A. L. Barabasi (2002): "Statistical Mechanics of Complex Networks". *Reviews of Modern Physics*. 74 (1): pp. 47-97.
- R. Albert and A. L. Barabási (2000): "Topology of Evolving Networks: Local Events and Universality". *Physical Review Letters*. 85 (24): pp. 5234-5237.
- R. Albert, H. Jeong and A. L. Barabasi (2000): "Error and Attack Tolerance of Complex Networks". *Nature*. 406 (6794): pp. 378-382.
- C. R. Allen (2006): "Predictors of Introduction Success in the South Florida Avifauna". *Biological Invasions*. 8 (3): pp. 491-500.
- T. F. H. Allen and T. B. Starr: (1982): "Hierarchy: Perspectives for Ecological Complexity". University of Chicago Press. ISBN 978-0226014326.
- U. Alon (2003): "Biological Networks: The Tinkerer as an Engineer". *Science*. 301 (5641): pp. 1866-1867.
- L. W. Alvarez, W. Alvarez, F. Asaro and H. V. Michel (1980): "Extraterrestrial Cause for the Cretaceous-Tertiary Extinction". *Science*. 208 (4448): p. 1095.
- J. M. Anderies (2005): "Minimal Models and Agroecological Policy at the Regional Scale: An Application to Salinity Problems in Southeastern Australia". *Regional Environmental Change*. 5 (1): pp. 1-17.
- P. Anderson (1972): "More Is Different". *Science*. 177 (4047): pp. 393-396.
- H. Andren (1994): "Effects of Habitat Fragmentation on Birds and Mammals in Landscapes with Different Proportions of Suitable Habitat: A Review". *Oikos*: pp. 355-366.
- S. J. Arnold, M. E. Pfrender and A. G. Jones (2001): "The Adaptive Landscape as a Conceptual Bridge between Micro- and Macroevolution". *Genetica*. 112-113 (1): pp. 9-32.
- W. R. Ashby: (1964): "An Introduction to Cybernetics". Methuen: London. ISBN.
- J. Avise and D. Walker (1998): "Pleistocene Phylogeographic Effects on Avian Populations and the Speciation Process". *Proceedings of the Royal Society B: Biological Sciences*. 265 (1395): pp. 457-463.
- R. Axelrod and W. Hamilton (1981): "The Evolution of Cooperation". *Science*. 211 (4489): p. 1390.
- R. Axtell, R. Axelrod, J. M. Epstein and M. D. Cohen (1996): "Aligning Simulation Models: A Case Study and Results". *Computational & Mathematical Organization Theory*. 1 (2): pp. 123-141.

-
- P. Bak: (1999): "How Nature Works: The Science of Self-Organized Criticality". Springer-Verlag Telos; Reprint edition. ISBN 978-0387987385.
- P. Bak and K. Sneppen (1993): "Punctuated Equilibrium and Criticality in a Simple Model of Evolution". *Physical Review Letters*. 71 (24): p. 4083.
- P. Bak, C. Tang and K. Wiesenfeld (1988): "Self-Organized Criticality". *Physical Review A*. 38 (1): pp. 364-374.
- J. M. Baldwin (1896): "A New Factor in Evolution". *The American naturalist*. 30 (354): pp. 441-451.
- Y. Bar-Yam: (1997): "Dynamics of Complex Systems". Westview Press. ISBN 978-0201557480.
- A. L. Barabasi: (2003): "Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life". Plume Books. ISBN 978-0452284395.
- A. L. Barabasi and R. Albert (1999): "Emergence of Scaling in Random Networks". *Science*. 286 (5439): p. 509.
- A. L. Barabasi, R. Albert and H. Jeong (2000): "Scale-Free Characteristics of Random Networks: The Topology of the World-Wide Web". *Physica A: Statistical Mechanics and its Applications*. 281 (1-4): pp. 69-77.
- N. H. Barton and G. M. Hewitt (1985): "Analysis of Hybrid Zones". *Annual Review of Ecology and Systematics*. 16: pp. 113-148.
- V. Batagelj and A. Mrvar (1996): "Pajek – Program for Large Network Analysis". Retrieved on 02/02/2010 from: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.
- G. Bateson: (1972): "Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology". University Of Chicago Press. ISBN 0-226-03905-6.
- P. S. Bearman, J. Moody and K. Stovel (2004): "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks". *American Journal of Sociology*. 110: pp. 44-91.
- M. Bedau (2002): "The Scientific and Philosophical Scope of Artificial Life". *Leonardo*. 35 (4): pp. 395-400.
- M. A. Bedau (1998): "Philosophical Content and Method of Artificial Life" in *The Digital Phoenix: How Computers Are Changing Philosophy*, T. W. Bynum and J. H. Moor, Editors. Wiley-Blackwell. pp. 135-152. ISBN 978-0631203520.
- M. A. Bedau (1999). "Can Unrealistic Computer Models Illuminate Theoretical Biology?". In: A. S. Wu (ed.), *Proceedings of 1999 Genetic and Evolutionary Computation Conference Workshop Program*, pp. 20-23. Morgan Kaufmann.
- M. A. Bedau, J. S. McCaskill, N. H. Packard, S. Rasmussen, C. Adami, D. G. Green, T. Ikegami, K. Kaneko and T. S. Ray (2000): "Open Problems in Artificial Life". *Artificial Life*. 6 (4): pp. 363-376.
- K. D. Bennett (2004): "Continuing the Debate on the Role of Quaternary Environmental Change for Macroevolution". *Philosophical Transactions: Biological Sciences*. 359 (1442): pp. 295-303.
- G. Bianconi, M. Marsili and F. Vega-Redondo (2005): "On the Non-Trivial Dynamics of Complex Networks". *Physica A*. 346: pp. 116-122.
- N. Biggs, E. K. Lloyd and R. J. Wilson: (1986): "Graph Theory". Oxford University Press, USA. ISBN.
-

- C. M. Bishop: (1995): "Neural Networks for Pattern Recognition". Oxford University Press. ISBN 0198538642.
- M. J. Blechner: (2001): "The Dream Frontier". Analytic Press. ISBN 978-0881632248.
- G. Booch, R. Maksimchuk, M. Engle, B. Young, J. Conallen and K. Houston: (2007): "Object-Oriented Analysis and Design with Applications, Third Edition". Addison-Wesley Professional. ISBN 9780201895513.
- J. M. Bower and H. Bolouri: (2004): "Computational Modeling of Genetic and Biochemical Networks". MIT Press. ISBN 978-0262524230.
- G. Brierley, K. Fryirs and V. Jain (2006): "Landscape Connectivity: The Geographic Basis of Geomorphic Applications". *Area*. 38 (2): pp. 165-174.
- R. Butlin (1987): "Speciation by Reinforcement". *Trends in Ecology & Evolution*. 2 (1): pp. 8-13.
- R. K. Butlin, C. Walton, K. A. Monk and J. R. Bridle (1998): "Biogeography of Sulawesi Grasshoppers, Genus Chitaura, Using DNA Sequence Data". *Biogeography and geological evolution of Southeast Asia*. Backhuys Publishers, Leiden, The Netherlands: pp. 355-359.
- E. Cantú-Paz (1998): "A Survey of Parallel Genetic Algorithms". *Calculateurs Paralleles, Reseaux et Systems Repartis*. 10 (2): pp. 141-171.
- S. R. Carpenter and L. H. Gunderson (2001): "Coping with Collapse: Ecological and Social Dynamics in Ecosystem Management". *BioScience*. 51 (6): pp. 451-457.
- S. Casper (2007): "How Do Technology Clusters Emerge and Become Sustainable? Social Network Formation and Inter-Firm Mobility within the San Diego Biotechnology Cluster". *Research Policy*. 36 (4): pp. 438-455.
- C. Castellano, M. Marsili and A. Vespignani (2000): "Nonequilibrium Phase Transition in a Model for Social Influence". *Physical Review Letters*. 85 (16): pp. 3536-3539.
- J. L. Casti (2009): "Complexity", in "*Encyclopedia Britannica Online*". Encyclopædia Britannica, Inc. ISBN.
- F. Castiglione (2006): "Agent-Based Modeling", in "*Scholarpedia*". p. 1562. ISBN.
- M. Catanzaro, M. Boguñá and R. Pastor-Satorras (2005): "Generation of Uncorrelated Random Scale-Free Networks". *Physical Review E*. 71 (2).
- V. Cerný (1985): "Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm". *Journal of Optimization Theory and Applications*. 45 (1): pp. 41-51.
- G. J. Chaitin (1969): "On the Length of Programs for Computing Finite Binary Sequences: Statistical Considerations". *Journal of the ACM*. 16 (1): pp. 145-159.
- S. C. Cobb (1999): "The Role of Corporate, Professional, and Personal Networks in the Provision of Offshore Financial Services". *Environment and Planning A*. 31: pp. 1877-1892.
- S. A. Cook (1971). "The Complexity of Theorem-Proving Procedures", *Proceedings of 3rd annual ACM symposium on Theory of computing*, pp. 151-158. ACM.
- G. R. Coope (2004): "Several Million Years of Stability among Insect Species Because of, or in Spite of, Ice Age Climatic Instability?". *Philosophical Transactions: Biological Sciences*. 359 (1442): pp. 209-214.
- O. Cordon, F. Moya and C. Zarco (2002): "A New Evolutionary Algorithm Combining Simulated Annealing and Genetic Programming for Relevance Feedback in Fuzzy Information Retrieval Systems". *Soft Computing*. 6 (5): pp. 308-319.

-
- R. Coutts (2008): "Dreams as Modifiers and Tests of Mental Schemas: An Emotional Selection Hypothesis". *Psychological reports*. 102 (2): p. 561.
- P. Coveney and R. Highfield: (1996): "Frontiers of Complexity: The Search for Order in a Chaotic World". Ballantine Books. ISBN 978-0449910818.
- S. A. Cowling, M. A. Maslin and M. T. Sykes (2001): "Paleovegetation Simulations of Lowland Amazonia and Implications for Neotropical Allopatry and Speciation". *Quaternary Research*. 55 (2): pp. 140-149.
- J. Crane (1991): "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping out and Teenage Childbearing". *American Journal of Sociology*. 96 (5): p. 1226.
- R. Crema, D. Prevedelli, A. Valentini and A. Castelli (2000): "Recovery of the Macrozoobenthic Community of the Comacchio Lagoon System (Northern Adriatic Sea)". *Ophelia*. 52 (2): pp. 143-152.
- F. Crick and G. Mitchison (1983): "The Function of Dream Sleep". *Nature*. 304 (5922): pp. 111-114.
- J. Crutchfield, J. D. Farmer, N. Packard and R. Shaw (1986): "Chaos". *Scientific American*. 255 (6): pp. 46-57.
- G. S. Cumming (2004): "The Impact of Low-Head Dams on Fish Species Richness in Wisconsin, USA". *Ecological Applications*. 14 (5): pp. 1495-1506.
- C. Dangalchev (2004): "Generation Models for Scale-Free Networks". *Physica A: Statistical Mechanics and its Applications*. 338 (3-4): pp. 659-671.
- C. Darwin: (1859): "On the Origin of Species". John Murray, London. ISBN.
- J. Davidsen, H. Ebel and S. Bornholdt (2002): "Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks". *Physical Review Letters*. 88 (12): p. 128701.
- J. B. Davis (2007): "Complexity Theory's Network Conception of the Individual" in *Money and Markets: A Doctrinal Approach*, A. Giacomini and M. C. Marcuzzo, Editors. Routledge. pp. 30-47. ISBN 978-0415384032.
- M. B. Davis (1979): "Pleistocene Biogeography of Temperate Deciduous Forests". *Geoscience and Man*. 13: pp. 13-26.
- J. X. De Carvalho and C. P. C. Prado (2000): "Self-Organized Criticality in the Olami-Feder-Christensen Model". *Physical Review Letters*. 84 (17): p. 4006.
- K. A. De Jong (1975): "Analysis of the Behavior of a Class of Genetic Adaptive Systems". University of Michigan. ISBN.
- K. De Queiroz (1998): "The General Lineage Concept of Species, Species Criteria, and the Process of Speciation: A Conceptual Unification and Terminological Recommendations" in *Endless Forms: Species and Speciation*, D. J. Howard and S. H. Berlocher, Editors. Oxford University Press: New York. pp. 57-75. ISBN.
- D. J. De Solla Price (1965): "Networks of Scientific Papers". *Science*. 149 (3683): pp. 510-515.
- E. B. Dent (1999): "Complexity Science: A Worldview Shift". *Emergence*. 1 (4): pp. 5-19.
- E. A. Di Paolo, J. Noble and S. Bullock (2000). "Simulation Models as Opaque Thought Experiments". In: M. Bedau, J. S. McCaskill, N. H. Packard, S. Rasmussen, J. McCaskill and N. Packard (eds.), *Proceedings of 7th International Conference on the Synthesis and Simulation of Living Systems (ALife VII)*, pp. 497-506.
- J. Diesner and K. M. Carley (2005). "Exploration of Communication Networks from the Enron Email Corpus", pp. 21-23. Citeseer.
-

- K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas and H. S. Chan (1995): "Principles of Protein Folding - a Perspective from Simple Exact Models". *Protein Science*. 4 (4): pp. 561-602.
- T. G. Dobzhansky: (1937): "Genetics and the Origin of Species". Columbia University Press: New York. ISBN.
- W. F. Doolittle (1981): "Is Nature Really Motherly?". *Co-Evolution Quarterly*. 29: pp. 58-63.
- G. Dueck (1993): "New Optimization Heuristics: The Great Deluge Algorithm and the Record-to-Record Travel". *Journal of Computational Physics*. 104 (1): pp. 86-92.
- R. I. M. Dunbar (1993): "Co-Evolution of Neocortex Size, Group Size and Language in Humans". *Behavioral and brain sciences*. 16 (4): pp. 681-735.
- D. Eder (1985): "The Cycle of Popularity: Interpersonal Relations among Female Adolescents". *Sociology of Education*. 58: pp. 154-165.
- M. M. Edson, T. C. Foin and C. M. Knapp (1981): "Emergent Properties and Ecological Research". *American Naturalist*. 118 (4): pp. 593-596.
- V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki and A. V. Apkarian (2005): "Scale-Free Brain Functional Networks". *Physical Review Letters*. 94 (1): p. 18102.
- G. C. M. A. Ehrhardt, M. Marsili and F. Vega-Redondo (2005). "On the Rise and Fall of Networked Societies", Proceedings of *Modeling Cooperative Behaviour in the Social Sciences*. American Institute of Physics.
- M. Eigen and P. Schuster (1979): "The Hypercycle - a Principle of Natural Self-Organization".
- N. Eldredge and S. J. Gould: (1972): "Punctuated Equilibria: An Alternative to Phyletic Gradualism". "*Models in Paleobiology*", ed. T. J. M. Schopf. Freeman Cooper: San Francisco. ISBN.
- I. Eliazar and J. Klafter (2009): "A Unified and Universal Explanation for Lévy Laws and 1/F Noises". *Proceedings of the National Academy of Sciences*. 106 (30): pp. 12251-12254.
- B. G. Elmegreen (1991). "Triggered Star Formation". In: G. Tenorio-Tagle, M. Prieto and F. Sanchez (eds.), *Proceedings of 3rd Canary Islands Winter School*. Cambridge University Press.
- J. A. Endler: (1977): "Geographic Variation, Speciation and Clines". "*Mpb*". Vol. 10. Princeton University Press. ISBN.
- A. J. Enright, S. Van Dongen and C. A. Ouzounis (2002): "An Efficient Algorithm for Large-Scale Detection of Protein Families". *Nucleic Acids Research*. 30 (7): pp. 1575-1584.
- P. Erdős and A. Rényi (1960): "On the Evolution of Random Graphs". *Matematikai Kutató Intézetének Közleményei*. 5: pp. 17-61.
- C. I. M. Escobar (2007): "Herramienta Autoconfigurable Para El Desarrollo De Aplicaciones Distribuidas De Tiempo Real Flexibles Y Dinámicas (in Spanish, Approach to Developing Flexible and Dynamic Self-Configuring Real-Time Distributed Applications)". Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Cinvestav, Mexico. ISBN.
- L. Euler (1741): "Solutio Problematis Ad Geometriam Situs Pertinentis". *Commentarii Academiae Scientiarum Imperialis Petropolitanae*. 8: pp. 128-140.
- M. J. Feigenbaum (1983): "Universal Behavior in Nonlinear Systems". *Physica D: Nonlinear Phenomena*. 7 (1-3): pp. 16-39.

-
- K. Filchak, J. Roethele and J. Feder (2000): "Natural Selection and Sympatric Divergence in the Apple Maggot Rhagoletis Pomonella". *Nature*. 407 (6805): pp. 739-742.
- R. A. Fisher: (1930): "The Genetical Theory of Natural Selection". The Clarendon Press: Oxford. ISBN.
- C. Flamm, I. L. Hofacker and P. F. Stadler (1999): "Rna in Silico the Computational Biology of Rna Secondary Structures". *Advances in Complex Systems*. 2 (1): pp. 65-90.
- C. Folke, N. Kautsky and M. Troell (1994): "The Costs of Eutrophication from Salmon Farming: Implications for Policy". *Journal of Environmental Management*. 40 (2): pp. 173-182.
- W. J. Freeman: (1975): "Mass Action in the Nervous System: Examination of Neurophysiological Basis of Adoptive Behavior through the Eeg". Academic Press: New York. ISBN 978-0122671500.
- W. J. Freeman (1991): "The Physiology of Perception". *Scientific American*. 264 (2): pp. 78-85.
- W. J. Freeman (1992): "Tutorial on Neurobiology: From Single Neurons to Brain Chaos". *International Journal of Bifurcation and Chaos*. 2 (3): pp. 451-482.
- T. M. J. Fruchterman and E. M. Reingold (1991): "Graph Drawing by Force-Directed Placement". *Software: Practice and Experience*. 21 (11): pp. 1129-1164.
- K. A. Fryirs, G. J. Brierley, N. J. Preston and M. Kasai (2007): "Buffers, Barriers and Blankets: The (Dis)Connectivity of Catchment-Scale Sediment Cascades". *Catena*. 70 (1): pp. 49-67.
- D. J. Futujama (1987): "On the Role of Species in Anagenesis". *The American naturalist*. 130: pp. 465-473.
- D. Garlaschelli, A. Capocci and G. Caldarelli (2007): "Self-Organized Network Evolution Coupled to Extremal Dynamics". *Nature Physics*. 3 (11): pp. 813-817.
- S. Gavrillets (1997): "Hybrid Zones with Dobzhansky-Type Epistatic Selection". *Evolution*. 51 (4): pp. 1027-1035.
- S. Gavrillets (2003): "Models of Speciation: What Have We Learned in 40 Years?". *Evolution*. 57 (10): pp. 2197-2215.
- S. Gavrillets: (2004): "Fitness Landscapes and the Origin of Species". "*Monographs in Population Biology*". Princeton University Press: Princeton / Oxford. ISBN.
- S. Gavrillets and J. Gravner (1997): "Percolation on the Fitness Hypercube and the Evolution of Reproductive Isolation". *Journal of Theoretical Biology*. 184 (1): pp. 51-64.
- S. Gavrillets and A. Vose (2005): "Dynamic Patterns of Adaptive Radiation". *Proceedings of the National Academy of Sciences USA*. 102 (50): pp. 18040-18045.
- M. Gen and R. Cheng: (1997): "Genetic Algorithms and Engineering Design". Wiley-Interscience. ISBN 978-0471127413.
- D. Gfeller, D. M. De Lachapelle, P. De Los Rios, G. Caldarelli and F. Rao (2007): "Uncovering the Topology of Configuration Space Networks". *Physical Review E*. 76 (2): p. 26113.
- M. Girvan and M. E. J. Newman (2002): "Community Structure in Social and Biological Networks". *Proceedings of the National Academy of Sciences*. 99 (12): p. 7821.
- H. A. Gleason (1927): "Further Views on the Succession-Concept". *Ecology*. 8: pp. 299-326.
- M. Gligor and M. Ignat (2001): "Some Demographic Crashes Seen as Phase Transitions". *Physica A: Statistical Mechanics and its Applications*. 301 (1-4): pp. 535-544.
-

- K. Gödel (1931): "Über Formal Unentscheidbare Sätze Der Principia Mathematica Und Verwandter Systeme". *Monatshefte für Mathematik und Physik*. 38: pp. 173-198.
- D. E. Goldberg and J. Richardson (1987). "Genetic Algorithms with Sharing for Multimodal Function Optimization", *Proceedings of 2nd International Conference on Genetic Algorithms and their application*, pp. 41-49. Lawrence Erlbaum Associates Inc., Mahwah, NJ, USA.
- J. Goldstein (1999): "Emergence as a Construct: History and Issues". *Emergence*. 1 (1): pp. 49-72.
- E. Golemis and P. D. Adams: (2005): "Protein-Protein Interactions: A Molecular Cloning Manual". Cold Spring Harbor Laboratory Press. ISBN 978-0879697235.
- D. M. Gordon (2007): "Control without Hierarchy". *Nature*. 446 (7132): p. 143.
- S. J. Gould: (1990): "Wonderful Life: The Burgess Shale and the Nature of History". W. W. Norton & Company. ISBN 978-0393307009.
- S. J. Gould: (2002): "The Structure of Evolutionary Theory". Belknap Press: Harvard. ISBN.
- D. G. Green (1982): "Fire and Stability in the Postglacial Forests of Southwest Nova Scotia". *Journal of Biogeography*. 9 (1): pp. 29-40.
- D. G. Green (1989): "Simulated Effects of Fire, Dispersal and Spatial Pattern on Competition within Forest Mosaics". *Plant Ecology*. 82 (2): pp. 139-153.
- D. G. Green (1993): "Emergent Behaviour in Biological Systems" in *Complex Systems: From Biology to Computation*, D. G. Green and T. R. J. Bossomaier, Editors. IOS Press. pp. 24-33. ISBN.
- D. G. Green (2000): "Self-Organization in Complex Systems" in *Complex Systems*, T. R. J. Bossomaier and D. G. Green, Editors. Cambridge University Press. pp. 7-41. ISBN.
- D. G. Green, T. G. Leishman and S. Sadedin (2006): "Dual Phase Evolution: A Mechanism for Self-Organization in Complex Systems". *International Journal Complex Systems*.
- D. G. Green and D. Newth (2001): "Towards a Theory of Everything? – Grand Challenges in Complexity and Informatics". *Complexity international*. 8: pp. 1-10.
- D. G. Green, D. Newth and M. G. Kirley (2000). "Connectivity and Catastrophe - Towards a General Theory of Evolution". In: M. Bedau, J. S. McCaskill, N. H. Packard, S. Rasmussen, J. McCaskill and N. Packard (eds.), *Proceedings of 7th International Conference on the Synthesis and Simulation of Living Systems (ALife VII)*.
- D. G. Green and S. Sadedin (2005): "Interactions Matter - Complexity in Landscapes and Ecosystems". *Ecological Complexity*. 2 (2): pp. 117-130.
- V. Grimm and S. F. Railsback: (2005): "Individual-Based Modeling and Ecology". Princeton University Press. ISBN 978-0691096667.
- V. Grimm, E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand and D. L. Deangelis (2005): "Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology". *Science*. 310: pp. 987-991.
- G. R. Grimmett: (1999): "Percolation". Springer. ISBN 3540649026.
- J. L. Gross and J. Yellen: (2003): "Handbook of Graph Theory". "*Discrete Mathematics and Its Applications*". Vol. 25. CRC Press. ISBN 1584880902.
- L. H. Gunderson and C. S. Holling: (2002): "Panarchy: Understanding Transformations in Human and Natural Systems". Island Press. ISBN.

-
- R. G. Harrison (1998): "Linking Evolutionary Pattern and Process: The Relevance of Species Concepts for the Study of Evolution" in *Endless Forms: Species and Speciation*, D. J. Howard and S. H. Berlocher, Editors. Oxford University Press: New York. pp. 19-31. ISBN.
- J. Hartmanis and R. E. Stearns (1965): "On the Computational Complexity of Algorithms". *Transactions of the American Mathematical Society*. 117: pp. 285-306.
- A. K. Hartmann and M. Weigt: (2006): "Phase Transitions in Combinatorial Optimization Problems: Basics, Algorithms and Statistical Mechanics". Wiley-VCH. ISBN 978-3527404735.
- F. Hazel (2008): "Connectivity as a Crucial Determinant of Badland Morphology and Evolution". *Geomorphology*. 100 (1-2): pp. 91-103.
- B. H. Heede, M. D. Harvey and J. R. Laird (1988): "Sediment Delivery Linkages in a Chaparral Watershed Following a Wildfire". *Environmental Management*. 12 (3): pp. 349-358.
- N. W. Henley: (1916): "Henley's Twentieth Century Formulas, Recipes and Processes". Vol. 3. The Norman W. Henley Publishing Company. ISBN.
- G. Hewitt (1996): "Some Genetic Consequences of Ice Ages, and Their Role, in Divergence and Speciation". *Biological Journal of the Linnean Society*. 58 (3): pp. 247-276.
- G. M. Hewitt (2004): "Genetic Consequences of Climatic Oscillations in the Quaternary". *Philosophical Transactions of the Royal Society, B: Biological Sciences*. 359 (1442): pp. 183-195.
- F. Heylighen (1997): "Classic Publications on Complex, Evolving Systems: A Citation-Based Survey". Retrieved on 12.06.2007 from: <http://pespmc1.vub.ac.be/Papers/PublicationsComplexity.html>.
- G. E. Hinton and S. J. Nowlan (1987): "How Learning Can Guide Evolution". *Complex Systems*. 1: pp. 492-502.
- B. Hjørland and K. Pedersen (2005): "A Substantive Theory of Classification for Information Retrieval". *Journal of documentation*. 61 (5): pp. 582-597.
- J. H. Holland: (1992): "Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence". MIT Press. ISBN 0262581116.
- J. H. Holland: (1995): "Hidden Order: How Adaptation Builds Complexity". Perseus Books. ISBN 0201407930.
- C. S. Holling (2001): "Understanding the Complexity of Economic, Ecological, and Social Systems". *Ecosystems*. 4 (5): pp. 390-405.
- I. Hors and F. Lordon (1997): "About Some Formalisms of Interaction Phase Transition Models in Economics?". *Journal of Evolutionary Economics*. 7 (4): pp. 355-373.
- S. P. Hubbell: (2001): "The Unified Neutral Theory of Biodiversity and Biogeography". Princeton University Press. ISBN.
- P. F. Hudson (2005): "Natural Levees" in *Encyclopedia of Water Science*, S. W. Trimble, B. A. Stewart and T. A. Howell, Editors. Taylor & Francis. ISBN 978-0-8493-9627-4.
- R. R. Hudson, M. Slatkin and W. P. Maddison (1992): "Estimation of Levels of Gene Flow from DNA Sequence Data". *Genetics*. 132 (2): pp. 583-589.
- G. Iori, G. De Masi, O. V. Precup, G. Gabbi and G. Caldarelli (2008): "A Network Analysis of the Italian Overnight Money Market". *Journal of Economic Dynamics and Control*. 32 (1): pp. 259-278.
-

- H. Ishibuchi, S. Kaige and K. Narukawa (2005). "Comparison between Lamarckian and Baldwinian Repair on Multiobjective 0/1 Knapsack Problems". In: C. A. C. Coello, A. H. Aguirre and E. Zitzler (eds.), *Proceedings of 3rd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, pp. 370-385. Springer.
- M. B. Isichenko (1992): "Percolation, Statistical Topography, and Transport in Random Media". *Reviews of Modern Physics*. 64 (4): p. 961.
- S. Itzkovitz, R. Levitt, N. Kashtan, R. Milo, M. Itzkovitz and U. Alon (2005): "Coarse-Graining and Self-Dissimilarity of Complex Networks". *Physical Review E*. 71 (1): p. 16127.
- H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabasi (2000): "The Large-Scale Organization of Metabolic Networks". *Nature*. 407 (6804): pp. 651-654.
- K. Johannesson (2001): "Parallel Speciation: A Key to Sympatric Divergence". *Trends in Ecology & Evolution*. 16 (3): pp. 148-153.
- S. E. Jørgensen and M. Straskraba (2000): "Ecosystems as Cybernetic Systems" in *Handbook of Ecosystem Theories and Management*, S. E. Jørgensen and F. Müller, Editors. CRC Press. pp. 249-264. ISBN 978-1566702539.
- Y. J. Jung and D. H. Lim (2010). "Measuring Responsiveness of Linux Kernel on Embedded System", *Proceedings of CELF Embedded Linux Conference*.
- N. Kashtan and U. Alon (2005): "Spontaneous Evolution of Modularity and Network Motifs". *Proceedings of the National Academy of Sciences*. 102 (39): pp. 13773-13778.
- N. Kashtan, S. Itzkovitz, R. Milo and U. Alon (2004): "Topological Generalizations of Network Motifs". *Physical Review E*. 70 (3): p. 31909.
- N. Kashtan, E. Noor and U. Alon (2007): "Varying Environments Can Speed up Evolution". *Proceedings of the National Academy of Sciences*. 104 (34): pp. 13711-13716.
- N. Kashtan, M. Parter, E. Dekel, A. E. Mayo and U. Alon (2009): "Extinctions in Heterogeneous Environments and the Evolution of Modularity". *Evolution*. 63 (8): pp. 1964-1975.
- J. S. Katz (1994): "Geographical Proximity and Scientific Collaboration". *Scientometrics*. 31 (1): pp. 31-43.
- S. A. Kauffman: (1993): "The Origins of Order: Self-Organization and Selection in Evolution". Oxford University Press: USA. ISBN 0195079515.
- M. Kawata and Y. Toquenaga (1994): "From Artificial Individuals to Global Patterns". *Trends in Ecology & Evolution*. 9 (11): pp. 417-421.
- A. P. Kershaw, J. S. Clark, A. M. Gill and D. M. D'costa (2002): "A History of Fire in Australia" in *Flammable Australia: The Fire Regimes and Biodiversity of a Continent*, R. A. Bradstock, J. E. Williams and A. M. Gill, Editors. Cambridge University Press. pp. 3-25. ISBN 978-0521805919.
- M. Kimura: (1983): "The Neutral Theory of Molecular Evolution". Cambridge University Press: Cambridge, New York. ISBN 0521231094.
- O. Kinouchi and C. P. C. Prado (1999): "Robustness of Scale Invariance in Models with Self-Organized Criticality". *Physical Review E*. 59 (5): p. 4964.
- S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi (1983): "Optimization by Simulated Annealing". *Science*. 220 (4598): pp. 671-680.
- M. G. Kirley (2002): "A Cellular Genetic Algorithm with Disturbances: Optimisation Using Dynamic Spatial Interactions". *Journal of Heuristics*. 8 (3): pp. 321-242.

-
- M. G. Kirley, X. Li and D. G. Green (1998). "Investigation of a Cellular Genetic Algorithm That Mimics Landscape Ecology". In: M. e. al. (ed.), *Proceedings of Simulated Evolution and Learning (SEAL'98)*, pp. 90-97. Springer.
- A. Kirman (1997): "The Economy as an Evolving Network". *Journal of Evolutionary Economics*. 7 (4): pp. 339-353.
- H. Kitano (2004): "Biological Robustness". *Nature Reviews Genetics*. 5 (11): pp. 826-837.
- H. Kitano, S. Hamahashi, J. Kitazawa, K. Takao and S. Imai (1997). "The Virtual Biology Laboratories: A New Approach of Computational Biology". In: P. Husbands and I. Harvey (eds.), *Proceedings of 4th European Conference on Artificial Life*, pp. 274-283. MIT Press.
- D. E. Knuth: (1998): "Art of Computer Programming". Vol. 1-3. Addison-Wesley Professional. ISBN 978-0201485417.
- K. Kollman, J. H. Miller and S. E. Page (1992): "Adaptive Parties in Spatial Elections". *The American Political Science Review*. 86 (4): pp. 929-937.
- A. N. Kolmogorov (1965): "Three Approaches to the Quantitative Definition of Information". *Problems of Information and Transmission*. 1 (1): pp. 1-7.
- I. Kornfield and P. F. Smith (2000): "African Cichlid Fishes: Model Systems for Evolutionary Biology". *Annual Review of Ecology and Systematics*. 31: pp. 163-196.
- G. Kossinets and D. J. Watts (2006): "Empirical Analysis of an Evolving Social Network". *Science*. 311 (5757): p. 88.
- A. Kruess and T. Tschardt (1994): "Habitat Fragmentation, Species Loss, and Biological Control". *Science*. 264 (5165): pp. 1581-1584.
- T. S. Kuhn (1977): "A Function for Thought Experiments" in *The Essential Tension: Selected Studies in Scientific Tradition and Change*. University Of Chicago Press. pp. 240-265. ISBN 978-0226458069.
- T. Kuran (1989): "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution". *Public Choice*. 61 (1): pp. 41-74.
- J. B. Lamarck: (1815): "Histoire Naturelle Des Animaux Sans Vertebres". ISBN.
- R. Lande (1981): "Models of Speciation by Sexual Selection on Polygenic Traits". *Proceedings of the National Academy of Sciences*. 78 (6): pp. 3721-3725.
- C. G. Langton (1986): "Studying Artificial Life with Cellular Automata". *Physica D*. 2 (1-3): pp. 120-149.
- C. G. Langton (1990): "Computation at the Edge of Chaos: Phase Transitions and Emergent Computation". *Physica D: Nonlinear Phenomena*. 42 (1-3): pp. 13-37.
- C. G. Langton (1991). "Life at the Edge of Chaos". In: C. G. Langton , C. Taylor, J. D. Farmer and S. Rasmussen (eds.), *Proceedings of 2nd International Conference on the Synthesis and Simulation of Living Systems (ALife II)*. Addison Wesley.
- T. G. Leishman, D. G. Green and G. Paperin (2007). "Dual Phase Evolution-a Mechanism for Self-Organisation and Optimisation". In: A. Namatame (ed.), *Proceedings of 11th Asia-Pacific Workshop on Intelligent and Evolutionary Systems*.
- T. M. Lenton and M. Van Oijen (2002): "Gaia as a Complex Adaptive System". *Philosophical Transactions of the Royal Society: Biological Sciences*. 357 (1421): pp. 683-695.
- L. B. Leopold, M. G. Wolman and J. P. Miller: (1995): "Fluvial Processes in Geomorphology". Dover Publications. ISBN 978-0486685885.
-

- S. A. Levin (1998): "Ecosystems and the Biosphere as Complex Adaptive Systems". *Ecosystems*. 1 (5): pp. 431-436.
- M. Levy (2005): "Social Phase Transitions". *Journal of Economic Behavior and Organization*. 57 (1): pp. 71-87.
- R. Lewin: (1992): "Complexity: Life at the Edge of Chaos". Macmillan. ISBN 978-0025704855.
- S. W. Lin, Z. J. Lee, S. C. Chen and T. Y. Tseng (2008): "Parameter Determination of Support Vector Machine and Feature Selection Using Simulated Annealing Approach". *Applied Soft Computing Journal*. 8 (4): pp. 1505-1512.
- H. Lipson, J. B. Pollack and N. P. Suh (2002): "On the Origin of Modular Variation". *Evolution*. 56 (8): pp. 1549-1556.
- J. Liu, Z. Cai and J. Liu (2000). "Premature Convergence in Genetic Algorithm: Analysis and Prevention Based on Chaos Operator", *Proceedings of the 3rd World Congress on Intelligent Control and Automation*, pp. 495-499.
- B. Liua, L. Wanga, Y.-H. Jina, F. Tangb and D.-X. Huang (2005): "Improved Particle Swarm Optimization Combined with Chaos". *Chaos, Solitons & Fractals*. 25 (5): pp. 1261-1271.
- Livegraph Open Source Project (2008): "Livegraph - a Framework for Real-Time Data Visualisation, Analysis and Logging". Retrieved on 02/02/2010 from: <http://www.live-graph.org>.
- J. E. Lovelock (1972): "Gaia as Seen through the Atmosphere". *Atmospheric Environment*. 6 (8): pp. 579-580.
- W. Ma, L. Lai, Q. Ouyang and C. Tang (2006): "Robustness and Modular Design of the Drosophila Segment Polarity Network". *Molecular Systems Biology*. 2 (1).
- S. W. Mahfoud (1995): "Niching Methods for Genetic Algorithms", in "*Urbana*". Iniversity of Illinois: Urbana-Champaign. p. 61801. ISBN.
- J. Mallet (1995): "A Species Definition for the Modern Synthesis". *Trends in Ecology and Evolution*. 10 (7): pp. 294-299.
- B. B. Mandelbrot: (1983): "The Fractal Geometry of Nature". Freeman: New York. ISBN.
- M. Mann (1997): "Has Globalization Ended the Rise and Rise of the Nation-State?". *Review of International Political Economy*. 4 (3): pp. 472-496.
- S. M. Manson (2001): "Simplifying Complexity: A Review of Complexity Theory". *Geoforum*. 32: pp. 405-414.
- L. Marengo and M. Willinger (1997): "Alternative Methodologies for Modelling Evolutionary Dynamics: Introduction". *Journal of Evolutionary Economics*. 7 (4): pp. 331-338.
- A. A. Markov (1906): "Распространение Закона Больших Чисел На Величины, Зависящие Друг От Друга". *Известия Физико-математического общества при Казанском университете*. 15 (2): pp. 135-156.
- M. Marsili, F. Vega-Redondo and F. Slanina (2004): "The Rise and Fall of a Networked Society: A Formal Model". *Proceedings of the National Academy of Sciences*. 101 (6): p. 1439.
- O. C. Martin, R. Monasson and R. Zecchina (2001): "Statistical Mechanics Methods and Phase Transitions in Optimization Problems". *Theoretical computer science*. 265 (1-2): pp. 3-67.
- J. Maynard Smith: (1958): "The Theory of Evolution". Penguin Books, London. ISBN.

J. Maynard Smith: (1974): "Models in Ecology". Cambridge University Press. ISBN 978-0521294409.

E. Mayr: (1942): "Systematics and the Origin of Species". Columbia University Press: New York. ISBN.

S. Meyn and R. L. Tweedie: (2009): "Markov Chains and Stochastic Stability". Cambridge University Press: New York. ISBN.

M. Mezard, G. Parisi and M. Virasoro: (1987): "Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications". "World Scientific Lecture Notes in Physics". Vol. 9. World Scientific: Singapore. ISBN.

P. G. Mezey: (1987): "Potential Energy Hypersurfaces". "Studies in Physical and Theoretical Chemistry ; 53". Elsevier: Amsterdam ; Oxford. ISBN 0444416994 (series) 0444428879 : 195.52.

R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer and U. Alon (2004): "Superfamilies of Evolved and Designed Networks". American Association for the Advancement of Science. pp. 1538-1542. ISBN.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon (2002): "Network Motifs: Simple Building Blocks of Complex Networks". *Science*. 298 (5594): pp. 824-827.

M. Mitchell, P. T. Hraber and J. P. Crutchfield (1993): "Revisiting the Edge of Chaos: Evolving Cellular Automata to Perform Computations". *Complex Systems*. 7: pp. 89-130.

B. Morel and R. Ramanujam (1999): "Through the Looking Glass of Complexity: The Dynamics of Organizations as Adaptive and Evolving Systems". *Organization Science*: pp. 278-293.

G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson (1998): "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function". *Journal of Computational Chemistry*. 19 (14): pp. 1639-1662.

K. B. Mullis (1990): "The Unusual Origin of the Polymerase Chain Reaction". *Scientific American*. 262 (4): pp. 56-65.

M. E. J. Newman (1997): "A Model of Mass Extinction". *Journal of Theoretical Biology*. 189: pp. 235-252.

M. E. J. Newman (2001a): "Clustering and Preferential Attachment in Growing Networks". *Physical Review E*. 64 (2): p. 25102.

M. E. J. Newman (2001b): "The Structure of Scientific Collaboration Networks". *Proceedings of the National Academy of Sciences*. 98 (2): p. 404.

M. E. J. Newman (2003): "The Structure and Function of Complex Networks". *SIAM Review*. 45: pp. 167-256.

M. E. J. Newman and M. Girvan (2004): "Finding and Evaluating Community Structure in Networks". *Physical Review E*. 69 (2): p. 26113.

M. E. J. Newman and R. M. Ziff (2000): "Efficient Monte Carlo Algorithm and High-Precision Results for Percolation". *Physical Review Letters*. 85 (19): pp. 4104-4107.

G. Nicolis and I. Prigogine: (1977a): "Self-Organisation in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations". John Wiley & Sons: New York. ISBN 978-0471024019.

G. Nicolis and I. Prigogine: (1977b): "Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations". John Wiley & Sons ISBN 978-0471024019.

- T. Ohta and M. Kimura (1973): "A Model of Mutation Appropriate to Estimate the Number of Electrophoretically Detectable Alleles in a Finite Population". *Genetical Research*. 22 (2): pp. 201-204.
- H. A. Orr (1995): "The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities". *Genetics*. 139 (4): pp. 1805-1813.
- N. H. Packard (1988): "Adaptation Towards the Edge of Chaos" in *Dynamic Patterns in Complex Systems*, J. A. S. Kelso, A. J. Mandell and M. F. Shlesinger, Editors. World Scientific. pp. 293-301. ISBN 9971504855.
- R. T. Paine (1969): "A Note on Trophic Complexity and Community Stability". *The American naturalist*. 103: pp. 91-93.
- R. T. Paine (1995): "A Conversation on Refining the Concept of Keystone Species". *Conservation Biology*. 9 (4): pp. 962-964.
- R. G. Palmer, B. Arthur, J. H. Holland, B. Lebaron and P. Tayler (1994): "Artificial Economic Life: A Simple Model of a Stockmarket". *Physica D: Nonlinear Phenomena*. 75 (1-3): pp. 264-274.
- G. Paperin (2007): "Livegraph Software Demonstration". At *8th Asia-Pacific Complex Systems Conference (Complex'07)*. Gold Coast, Australia.
- G. Paperin (2008a): "Dual Phase Evolution in Natural and Artificial Computational Systems (Invited Talk)". At *12th Asia-Pacific Symposium on Intelligent and Evolutionary Systems (IES'08)*. Melbourne, Australia.
- G. Paperin (2008b): "Using Holey Fitness Landscapes to Counteract Premature Convergence in Evolutionary Algorithms". In: S. Gustafson (ed.), *Proceedings of Graduate Student Workshop at the Genetic and Evolutionary Computation Conference 2008 (GECCO'08)*, pp. 1815-1818. ACM.
- G. Paperin (2010): "The Role of Networks in Computational Complex Systems". *International Transactions on Systems Science and Applications*. In Press.
- G. Paperin, D. G. Green and A. Dorin (2007). "Fitness Landscapes in Individual-Based Simulation Models of Adaptive Radiation". In: T. D. Pham and X. Zhou (eds.), *Proceedings of 2007 International Symposium on Computational Models for Life Science (CMLS'07)*, pp. 268-278. American Institute of Physics.
- G. Paperin, D. G. Green and T. G. Leishman (2008). "Dual Phase Evolution and Self-Organisation in Networks", *Proceedings of 7th International Conference on Simulated Evolution and Learning (SEAL'08)*, pp. 575-584. Springer.
- G. Paperin, D. G. Green and S. Sadedin (2010): "Dual Phase Evolution in Complex Adaptive Systems". *Under Review by Royal Society Journal Interface*.
- G. Paperin, D. G. Green, S. Sadedin and T. G. Leishman (2007a): "Complexity in Speciation: Effects of Disasters on Adaptive Radiation in a Dual Phase Evolution Model". At *8th Asia-Pacific Complex Systems Conference*. Gold Coast, Australia.
- G. Paperin, D. G. Green, S. Sadedin and T. G. Leishman (2007b). "A Dual Phase Evolution Model of Adaptive Radiation in Landscapes". In: M. Randall, H. A. Abbass and J. Wiles (eds.), *Proceedings of 3rd Australian Conference on Artificial Life (ACAL'07)*, pp. 131-143. Springer.
- G. Paperin and S. Sadedin (2009a). "The Dual Phase Evolution Framework for Understanding Evolutionary Dynamics in Complex Adaptive Systems", *Proceedings of 2009 International Conference on Evolutionary Computation (ICEC'09)*.
- G. Paperin and S. Sadedin (2009b): "Towards Formalising the Theory of Dual Phase Evolution". At *10th European Conference on Artificial Life (ECAL'09)*.

G. Paperin, S. Sadedin, D. G. Green and A. Dorin (2008). "Holey Fitness Landscapes and the Maintenance of Evolutionary Diversity". In: S. Bullock, J. Noble, R. Watson and M. A. Bedau (eds.), *Proceedings of 11th International Conference on the Synthesis and Simulation of Living Systems (ALife XI)*, pp. 450-457. MIT Press.

A. Parravano (1988): "Self-Regulating Star Formation in Isolated Galaxies - Thermal Instabilities in the Interstellar Medium". *Astronomy and Astrophysics*. 205 (1-2): pp. 71-76.

A. Parravano, P. Rosenzweig and M. Teran (1990): "Galactic Evolution with Self-Regulated Star Formation- Stability of a Simple One-Zone Model". *Astrophysical Journal*. 356 (Part 1): pp. 100-109.

R. Pastor-Satorras and A. Vespignani (2001): "Epidemic Spreading in Scale-Free Networks". *Physical Review Letters*. 86 (14): pp. 3200-3203.

J. D. Phillips (1993): "Biophysical Feedbacks and the Risks of Desertification". *Annals of the Association of American Geographers*. 83 (4): pp. 630-640.

J. Piaget, T. Brown and K. J. Thampy: (1985): "The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development". University of Chicago Press Chicago. ISBN 978-0226667812.

J. Piaget and B. Inhelder: (2000): "The Psychology of the Child". Basic Books. ISBN 978-0465095001.

S. L. Pimm: (1982): "Food Webs". Chapman and Hall. ISBN 041223100X.

J. H. Poincaré (1890): "Sur Le Problème Des Trois Corps Et Les Équations De La Dynamique. Divergence Des Séries De M. Lindstedt.". *Acta Mathematica*. 13: pp. 1-270.

I. Prigogine: (1980): "From Being to Becoming: Time and Complexity in the Physical Sciences". W H Freeman & Co. ISBN 978-0716711087.

W. B. Provine: (1986): "Sewall Wright and Evolutionary Biology". University of Chicago Press. ISBN 0226684733.

A. Purvis and A. Hector (2000): "Getting the Measure of Biodiversity". *Nature*. 405: pp. 212-219.

C. V. Ramamoorthy and S. Shekhar (1989). "Stochastic Backpropagation: A Learning Algorithm for Generalizationproblems", *Proceedings of 13th Annual International Computer Software and Applications Conference 1989 (COMPSAC'89)*, pp. 664-671.

T. S. Ray (1991). "An Approach to the Synthesis of Life". In: C. G. Langton , C. Taylor, J. D. Farmer and S. Rasmussen (eds.), *Proceedings of 2nd International Conference on the Synthesis and Simulation of Living Systems (ALife II)*, pp. 371-408. Addison Wesley.

C. M. Reidys (1997): "Random Induced Subgraphs of Generalized N-Cubes". *Advances in Applied Mathematics*. 13 (3): pp. 360 - 377.

I. Reinfelds and G. Nanson (1993): "Formation of Braided River Floodplains, Waimakariri River, New Zealand". *Sedimentology*. 40 (6): pp. 1113-1128.

Resilience Alliance (2004): "Resilience Alliance Bibliography Database on Thresholds and Alternate States in Ecological and Social-Ecological Systems". Retrieved on 07/08/2009 from: <http://www.resalliance.org/2.php>.

M. E. Ritter (2006): "The Physical Environment: An Introduction to Physical Geography". Retrieved on 20/02/2010 from: http://www.uwsp.edu/geo/faculty/ritter/geogl01/textbook/title_page.html.

- B. M. Roehner (2005): "A Bridge between Liquids and Socio-Economic Systems: The Key Role of Interaction Strengths". *Physica A: Statistical Mechanics and its Applications*. 348: pp. 659-682.
- M. G. S. Rojas (2004): "Measures of Diversity: A Comparison of Spatial Patterns in a Marine Fouling Community", in "*Marine Ecology*". Göteborg University: Göteborg ISBN.
- D. M. Rosenthal, A. E. Schwarzbach, L. A. Donovan, O. Raymond and L. H. Rieseberg (2002): "Phenotypic Differentiation between Three Ancient Hybrid Taxa and Their Parental Species". *International Journal of Plant Sciences*. 163 pp. 387-398.
- H. Rundle, L. Nagel, J. Boughman and D. Schluter (2000): "Natural Selection and Parallel Speciation in Sympatric Sticklebacks". *Science*. 287 (5451): p. 306.
- P. J. Russell: (1994): "Fundamentals of Genetics". HarperCollinsCollege Publishers: New York. ISBN 0065006402.
- S. Sadedin and G. Paperin (2009). "Implications of the Social Brain Hypothesis for Evolving Human-Like Cognition in Digital Organisms", Proceedings of *10th European Conference on Artificial Life (ECAL'09)*.
- M. Sain, H. J. Lee and W.-Y. Chung (2009a). "Middleware for Ubiquitous Healthcare Information System", Proceedings of *11th international conference on Advanced Communication Technology*.
- M. Sain, H. J. Lee and W.-Y. Chung (2009b). "Muhis: A Middleware Approach Using Livegraph", Proceedings of *Multimedia, Signal Processing and Communication Technologies, 2009 (IMPACT '09)*, pp. 197-200.
- D. Saunders, R. Hobbs and C. Margules (1991): "Biological Consequences of Ecosystem Fragmentation: A Review". *Conservation Biology*: pp. 18-32.
- M. Scheffer and S. R. Carpenter (2003): "Catastrophic Regime Shifts in Ecosystems: Linking Theory to Observation". *Trends in Ecology & Evolution*. 18 (12): pp. 648-656.
- M. Schilthuizen: (2001): "Frogs, Flies, and Dandelions: Speciation - the Evolution of New Species". Oxford University Press, USA. ISBN.
- D. Schluter: (2000): "The Ecology of Adaptive Radiation". Oxford University Press, USA. ISBN.
- M. Schroeder: (1991): "Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise". Freeman: New York. ISBN 0-7167-2136-8.
- P. E. Seiden and L. S. Schulman (1986): "Percolation and Galaxies". *Science*. 233: pp. 425-431.
- P. E. Seiden and L. S. Schulman (1990): "Percolation Model of Galactic Structure". *Advances in Physics*. 39 (1): pp. 1-54.
- C. Sergent and S. Dehaene (2004): "Neural Processes Underlying Conscious Perception: Experimental Findings and a Global Neuronal Workspace Framework". *Journal of Physiology-Paris*. 98 (4-6): pp. 374-384.
- C. E. Shannon: (1993): "Collected Papers", ed. N. J. A. Sloane and A. D. Wyner. Wiley-IEEE Press. ISBN 978-0780304345.
- K. L. Shaw (1998): "Species and the Diversity of Natural Groups" in *Endless Forms: Species and Speciation*, D. J. Howard and S. H. Berlocher, Editors. Oxford University Press: New York. pp. 44-56. ISBN.
- G. Singh and E. A. Geissler (1985): "Late Cainozoic History of Vegetation, Fire, Lake Levels and Climate, at Lake George, New South Wales, Australia". *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 311 (1151): pp. 379-447.

-
- L. Smolin (1995): "Cosmology as a Problem in Critical Phenomena" in *Complex Systems and Binary Networks*. Springer: Berlin / Heidelberg. pp. 184-223. ISBN 978-3-540-60339-9.
- S. Solomon, G. Weisbuch, L. De Arcangelis, N. Jan and D. Stauffer (2000): "Social Percolation Models". *Physica A: Statistical Mechanics and its Applications*. 277 (1-2): pp. 239-247.
- R. J. Solomonoff (1964): "A Formal Theory of Inductive Inference". *Information and control*. 7 (2): pp. 224-254.
- D. Sornette, A. Johansen and I. Dornic (1995): "Mapping Self-Organized Criticality onto Criticality". *Journal de Physique I*. 5: pp. 325-335.
- P. Spaak and J. R. Hoekstra (1997): "Fish Predation on a Daphnia Hybrid Species Complex: A Factor Explaining Species Coexistence?". *Limnology and Oceanography*. 42 (4): pp. 753-762.
- H. G. Spencer, B. H. Mcardle and D. M. Lambert (1986): "A Theoretical Investigation of Speciation by Reinforcement". *The American naturalist*. 128 (2): pp. 241-262.
- P. F. Stadler (2002): "Fitness Landscapes" in *Lecture Notes in Physics*. Springer Berlin / Heidelberg. pp. 183-. ISBN.
- L. Steels and R. A. Brooks: (1995): "The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents". Psychology Press. ISBN 978-0805815184.
- D. B. Stein and D. S. Barrington (1990): "Recurring Hybrid Formation in a Population of Polystichum X Potteri: Evidence from Chloroplast DNA Comparisons". *Annals of the Missouri Botanical Garden*. 77: pp. 334-339.
- R. Stickgold, J. A. Hobson, R. Fosse and M. Fosse (2001): "Sleep, Learning, and Dreams: Off-Line Memory Reprocessing". pp. 1052-1057. ISBN.
- C. Sturmbauer (1998): "Explosive Speciation in Cichlid Fishes of the African Great Lakes: A Dynamic Model of Adaptive Radiation". *Journal of Fish Biology*. 53: pp. 18-36.
- C. Sturmbauer and A. Meyer (1992): "Genetic Divergence, Speciation and Morphological Stasis in a Lineage of African Cichlid Fishes". *Nature*. 358 (6387): pp. 578-581.
- M. A. Summerfield: (1991): "Global Geomorphology". Prentice Hall. ISBN 978-0582301566.
- F. Sun and M. Sun (2005). "Transductive Support Vector Machines Using Simulated Annealing". In: Y. Hao, J. Liu, Y. Wang, Y.-m. Cheung, H. Yin, L. Jiao, J. Ma and Y.-C. Jiao (eds.), *Proceedings of 2005 International Conference Computational Intelligence and Security (CIS'2005)*, pp. 536-543. Springer.
- N. Swenson and D. Howard (2005): "Clustering of Contact Zones, Hybrid Zones, and Phylogeographic Breaks in North America". *The American naturalist*. 166 (5): pp. 581-591.
- J. J. Sylvester (1878): "Chemistry and Algebra". *Nature*. 17 (432): p. 284.
- R. Szostak (2008): "Classification, Interdisciplinarity, and the Study of Science". *Journal of documentation*. 64 (3): p. 319.
- J. M. Szymura and N. H. Barton (1986): "Genetic Analysis of a Hybrid Zone between the Fire-Bellied Toads, Bombina Bombina and B. Variegata, near Cracow in Southern Poland". *Evolution*. 40 (6): pp. 1141-1159.
- T. S. Tadele (2009): "Development of Oromacs Browser", in "*Control Engineering*". University of Twente, Netherlands. ISBN 021CE2009.
- C. Taylor and D. Jefferson (1995): "Artificial Life as a Tool for Biological Enquiry" in *Artificial Life - an Overview*, C. G. Langton Editor. MIT Press. pp. 1-13. ISBN 978-0262621120.
-

- C. E. Taylor (2003). "Fleshing out Artificial Life II". In: C. G. Langton , C. Taylor, J. D. Farmer and S. Rasmussen (eds.), *Proceedings of 2nd International Conference on the Synthesis and Simulation of Living Systems (ALife II)*, pp. 25--. Addison Wesley.
- P. D. Taylor and M. A. Krawchuk (2005): "Scale and Sensitivity of Songbird Occurrence to Landscape Structure in a Harvested Boreal Forest". *Avian Conservation and Ecology*. 1 (1).
- A. R. Templeton (1998): "Species and Speciation: Geography, Population Structure, Ecology and Gene Trees" in *Endless Forms: Species and Speciation*, D. J. Howard and S. H. Berlocher, Editors. Oxford University Press: New York. pp. 32-43. ISBN.
- R. Thom: (1975): "Structural Stability and Morphogenesis". Mass: Reading. ISBN.
- M. F. Thomas (2001): "Landscape Sensitivity in Time and Space—an Introduction". *Catena*. 42 (2-4): pp. 83-98.
- W. Tinner, P. Hubschmid, M. Wehrli, B. Ammann and M. Conedera (1999): "Long-Term Forest Fire Ecology and Dynamics in Southern Switzerland". *Journal of Ecology*. 87: pp. 273-289.
- S. Van Dongen (2000a): "A Cluster Algorithm for Graphs". National Research Institute for Mathematics and Computer Science in the Netherlands: Amsterdam. ISBN INS-R0010.
- S. Van Dongen (2000b): "Graph Clustering by Flow Simulation". University of Utrecht: Utrecht. ISBN.
- S. Van Dongen (2000c): "Performance Criteria for Graph Clustering and Markov Cluster Experiments". National Research Institute for Mathematics and Computer Science in the Netherlands: Amsterdam. ISBN INS-R0012.
- S. Van Dongen (2000d): "A Stochastic Uncoupling Process for Graphs". National Research Institute for Mathematics and Computer Science in the Netherlands: Amsterdam. ISBN INS-R0011.
- L. H. Van Vlack: (1980): "Elements of Materials Science and Engineering". 6 edition. Prentice Hall. ISBN 978-0201093148.
- F. Varela: (1979): "Principles of Biological Autonomy". North Holland: New York. ISBN 978-0135009505.
- E. A. Variano, J. H. McCoy and H. Lipson (2004): "Networks, Dynamics, and Modularity". *Physical Review Letters*. 92 (18): pp. 188701-188701.
- H. Von Foerster (1960): "On Self-Organising Systems and Their Environments" in *On Self-Organising Systems and Their Environments*, M. C. Yovits and S. Cameron, Editors. Pergamon Press: London. pp. 30-50. ISBN.
- H. Von Foerster: (1981): "Observing Systems: Selected Papers of Heinz Von Foerster". Intersystems Publications: Seaside. ISBN 978-0914105190.
- J. Von Neumann: (1966): "Theory of Self-Reproducing Automata", ed. A. W. Burks. University of Illinois Press: Urbana. ISBN.
- A. Wagner and D. A. Fell (2001): "The Small World inside Large Metabolic Networks". *Proceedings of the Royal Society: Biological Sciences*. 268 (1478): pp. 1803-1810.
- G. P. Wagner (1996): "Homologues, Natural Kinds and the Evolution of Modularity". *Integrative and Comparative Biology*. 36 (1): pp. 36-43.
- G. P. Wagner and L. Altenberg (1996): "Complex Adaptations and the Evolution of Evolvability". *Evolution*. 50 (3): pp. 967-976.

-
- B. Walker, C. S. Holling, S. R. Carpenter and A. Kinzig (2004): "Resilience, Adaptability and Transformability in Social and Ecological Systems". *Ecology and Society*. 9 (2): p. 5.
- C. S. Wallace and D. M. Boulton (1968): "An Information Measure for Classification". *The Computer Journal*. 11 (2): pp. 185-194.
- X. F. Wang and G. Chen (2003): "Complex Networks: Small-World, Scale-Free and Beyond". *IEEE circuits and systems magazine*. 3 (1): pp. 6-20.
- X. H. Wang and J. J. Li (2004). "Hybrid Particle Swarm Optimization with Simulated Annealing", Proceedings of 2004 International Conference on Machine Learning and Cybernetics, pp. 2402- 2405.
- J. Warburton (1994): "Channel Change in Relation to Meltwater Flooding, Bas Glacier D'arolla, Switzerland". *Geomorphology*. 11 (2): pp. 141-149.
- A. J. Watson and J. E. Lovelock (1983): "Biological Homeostasis of the Global Environment: The Parable of Daisyworld". *Tellus B*. 35 (4): pp. 284-289.
- D. J. Watts and S. H. Strogatz (1998): "Collective Dynamics of 'Small-World' Networks". *Nature*. 393 (6684): pp. 409-410.
- S. L. Weber (2001): "On Homeostasis in Daisyworld". *Climatic Change*. 48 (2-3): pp. 465-485.
- D. Whitley, S. Rana and R. B. Heckendorn (1999): "The Island Model Genetic Algorithm: On Separability, Population Size and Convergence". *Journal of Computing and Information Technology*. 7 (1): pp. 33-47.
- L. D. Whitley (1993). "Cellular Genetic Algorithms", Proceedings of 5th International Conference on Genetic Algorithms. Morgan Kaufmann.
- K. Willis, K. Bennett and D. Walker (2004): "The Evolutionary Legacy of the Ice Ages-Introduction". *Phil. Trans. R. Soc. Lond. B*. 359: pp. 157-158.
- D. S. Wilson (1988): "Holism and Reductionism in Evolutionary Ecology". *Oikos*. 53 (2): pp. 269-273.
- K. With (2002): "The Landscape Ecology of Invasive Spread". *Conservation Biology*. 16 (5): pp. 1192-1203.
- S. Wolfram (1984): "Universality and Complexity in Cellular Automata". *Physica D: Nonlinear Phenomena*. 10 (1-2): pp. 1-35.
- S. Wolfram: (2002): "A New Kind of Science". "Applied Mechanics Reviews". Wolfram Media. ISBN 978-1579550080.
- A. Wood, G. Ackland, J. Dyke, H. Williams and T. Lenton (2008): "Daisyworld: A Review". *Reviews of Geophysics*. 46 (1): pp. RG1001-1023.
- S. Wright (1932). "The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution". In: D. F. Jones (ed.), Proceedings of 6th International Congress of Genetics, pp. 256-366.
- S. Wuchty (2001): "Scale-Free Behavior in Protein Domain Networks". *Molecular biology and evolution*. 18 (9): pp. 1694-1702.
- A. S. Yang (2001): "Modularity, Evolvability, and Adaptive Radiations: A Comparison of the Hemi-and Holometabolous Insects". *Evolution & development*. 3 (2): pp. 59-72.
- A. Young, T. Boyle and T. Brown (1996): "The Population Genetic Consequences of Habitat Fragmentation for Plants". *Trends in Ecology & Evolution*. 11 (10): pp. 413-418.
-