

# Experimental Investigations into Graph Grammar Evolution

by

Martin Holger Luerssen, *B.Sc. (Hons)*  
School of Informatics and Engineering,  
Faculty of Science and Engineering

May 29, 2006

A thesis presented to the  
Flinders University of South Australia  
in fulfillment of the requirements for the degree of  
Doctor of Philosophy



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Conceptual Background . . . . .	1
1.2	Research Strategy . . . . .	3
1.3	Contributions . . . . .	3
1.4	Overview . . . . .	4
<b>2</b>	<b>Network Adaptation</b>	<b>7</b>
2.1	A Brief Note on Graphs & Networks . . . . .	7
2.2	Automata Networks . . . . .	8
2.2.1	Cellular Automata . . . . .	9
2.2.2	Threshold Automata . . . . .	11
2.2.3	Neural Networks . . . . .	12
2.2.4	Random Boolean Networks . . . . .	13
2.3	Complex Networks . . . . .	14
2.4	Network Learning . . . . .	16
2.4.1	Weight Learning . . . . .	17
2.4.2	Structure Learning . . . . .	18
2.5	Evolutionary Computation . . . . .	18
2.5.1	Exploring the Fitness Landscape . . . . .	20
2.5.2	Genetic Algorithms . . . . .	21
2.5.3	Genetic Programming . . . . .	22
2.5.4	Learning vs. Evolution . . . . .	22
2.6	Offspring Generation . . . . .	23
2.6.1	Mutation . . . . .	24

---

2.6.2	Recombination . . . . .	25
2.6.3	Estimation of Distribution Algorithms . . . . .	27
2.7	Network Evolution . . . . .	27
2.7.1	Weight Evolution . . . . .	28
2.7.2	Structure Evolution . . . . .	29
2.7.3	Cartesian Genetic Programming . . . . .	30
2.8	Summary . . . . .	31
<b>3</b>	<b>From Embryogeny to Graph Grammar</b>	<b>33</b>
3.1	Embryogeny . . . . .	34
3.2	Modularity . . . . .	35
3.3	Neutrality . . . . .	37
3.4	Categories of Artificial Embryogeny . . . . .	38
3.4.1	A Survey of Artificial Embryogeny . . . . .	39
3.5	Grammar Evolution . . . . .	42
3.5.1	Generative Grammars . . . . .	42
3.5.2	Lindenmayer-Systems . . . . .	43
3.5.3	Cellular Encoding . . . . .	48
3.6	Graph Grammars . . . . .	50
3.6.1	Hyperedge Replacement Systems . . . . .	51
3.7	Grammar Model-Building . . . . .	53
3.8	Summary . . . . .	55
<b>4</b>	<b>Cellular Graph Grammar Evolution</b>	<b>57</b>
4.1	Cellular Graphs . . . . .	58
4.2	Unified Grammar Model . . . . .	63
4.3	Evolving the Grammar . . . . .	66
4.3.1	Variation Operators . . . . .	68
4.3.2	Size Control . . . . .	70
4.3.3	Multi-objective Optimisation . . . . .	71
4.4	Evaluating the Network . . . . .	72
4.5	The G/GRADE System . . . . .	75

---

4.6	Applications . . . . .	76
4.6.1	Symbolic Regression . . . . .	76
4.6.2	Neural Networks . . . . .	77
4.6.3	Circuit Design . . . . .	78
4.6.4	Telecommunications . . . . .	80
4.7	Summary . . . . .	83
<b>5</b>	<b>Methods of Graph Construction</b>	<b>85</b>
5.1	Graph Gluing Models . . . . .	85
5.1.1	Strict Matching . . . . .	86
5.1.2	Soft Matching . . . . .	89
5.1.3	Method . . . . .	91
5.1.4	Results & Discussion . . . . .	95
5.2	Graph Modularity . . . . .	99
5.2.1	Method . . . . .	102
5.2.2	Results & Discussion . . . . .	103
5.3	Implicit and Explicit Gluing Models . . . . .	104
5.3.1	Feedforward Networks . . . . .	105
5.3.2	Method . . . . .	106
5.3.3	Results & Discussion . . . . .	107
5.4	Enforcing Simple Graphs . . . . .	109
5.4.1	Method . . . . .	110
5.4.2	Results & Discussion . . . . .	110
5.5	Summary . . . . .	112
<b>6</b>	<b>A Case for Phenotypic Diversity</b>	<b>113</b>
6.1	Diversity Objectives . . . . .	114
6.1.1	Measures of Phenotypic Diversity . . . . .	115
6.1.2	Age as a Diversity Heuristic . . . . .	117
6.1.3	Method . . . . .	118
6.1.4	Results & Discussion . . . . .	119
6.2	Island Models . . . . .	124

---

6.2.1	Graph Grammar Islands . . . . .	125
6.2.2	Method . . . . .	127
6.2.3	Results & Discussion . . . . .	128
6.3	Summary . . . . .	131
<b>7</b>	<b>Convergence Outcomes and Analysis</b>	<b>133</b>
7.1	Rates of Change . . . . .	134
7.1.1	Method . . . . .	135
7.1.2	Results & Discussion . . . . .	136
7.2	Evaluation Length . . . . .	137
7.2.1	Method . . . . .	138
7.2.2	Results & Discussion . . . . .	138
7.3	A Basic Convergence Proof . . . . .	141
7.4	Adaptive Search . . . . .	142
7.4.1	Swarm Intelligence . . . . .	143
7.4.2	Adaptive Production Search . . . . .	145
7.4.3	Method . . . . .	147
7.4.4	Results & Discussion . . . . .	148
7.5	Balancing Bloat against Performance . . . . .	149
7.5.1	Method . . . . .	150
7.5.2	Results & Discussion . . . . .	150
7.6	Observations on Evolved Graph Grammars . . . . .	154
7.6.1	Cellular Graph Components . . . . .	155
7.6.2	Production Reuse . . . . .	156
7.6.3	Production Copying . . . . .	159
7.6.4	Generational Development . . . . .	160
7.7	Comparisons against Other Systems . . . . .	166
7.8	Summary . . . . .	167

---

<b>8</b>	<b>Conclusions</b>	<b>169</b>
8.1	Contributions . . . . .	170
8.2	Limitations . . . . .	172
8.3	Future Work . . . . .	173
8.3.1	Strong Graph Bias . . . . .	173
8.3.2	Real Space Evolution . . . . .	173
8.3.3	Production Inheritance . . . . .	174
8.3.4	Grammar-Guided Search . . . . .	175
<b>A</b>	<b>Complete Tables of Results</b>	<b>177</b>
A.1	Binomial-3 Regression . . . . .	177
A.2	6th-order Polynomial Regression . . . . .	179
A.3	Random Bit Sequence . . . . .	179
A.4	6-bit Multiplexer . . . . .	181
A.5	Backpropagation MLP . . . . .	181
A.6	Pole Balancing . . . . .	182
A.7	Computer Network Topology . . . . .	183
	<b>Bibliography</b>	<b>185</b>





# List of Figures

2.1	Graph concepts (simple graphs, multigraphs, pseudographs) . . .	8
2.2	Moore and Von Neumann neighbourhoods on a Euclidean lattice .	10
2.3	Examples of automata networks (cellular automata, neural networks)	12
2.4	The canonical evolutionary algorithm . . . . .	19
2.5	A population on a fitness landscape . . . . .	20
2.6	Cartesian Genetic Programming . . . . .	31
3.1	Structural modularity . . . . .	37
3.2	String to graph translation with the G2L-system . . . . .	46
3.3	Cellular Encoding . . . . .	49
3.4	Hyperedges . . . . .	51
4.1	Position independence in hyperedge replacement . . . . .	59
4.2	Two labelled hypergraphs for illustrating the cellular graph representation . . . . .	60
4.3	A cellular production . . . . .	63
4.4	A unified grammar as an alternative to multiple grammars . . . .	64
4.5	An algorithm for cellular graph grammar evolution . . . . .	65
4.6	Evolution of a cellular graph grammar over 6 generations . . . . .	67
4.7	The Pareto frontier . . . . .	72
4.8	Example derivation of a network from a cellular graph grammar .	73
4.9	Screenshot: G/GRADE software suite . . . . .	75
4.10	The pole balancing task . . . . .	77
4.11	Various solution networks evolved by G/GRADE . . . . .	79
4.12	Design problem: data streaming computer network . . . . .	81
4.13	Computer network topologies evolved by G/GRADE . . . . .	83

---

5.1	How target labels match source labels . . . . .	86
5.2	Strict matching probability graph . . . . .	88
5.3	Strict matching versus soft matching . . . . .	90
5.4	Box plot tutorial . . . . .	96
5.5	Performance box plot: different configurations of strict matching compared to soft matching . . . . .	97
5.6	Performance box plot: different offset schemes for soft matching .	98
5.7	The scope of label matching in a cellular production . . . . .	100
5.8	Label offsets and nonmodular productions . . . . .	102
5.9	Performance box plot: selective modularity and global tentacles .	103
5.10	Implicit and explicit gluing models . . . . .	105
5.11	Performance box plot: gluing models for vertices . . . . .	107
5.12	Performance box plot: gluing models for hyperedges . . . . .	108
5.13	Performance box plot: constraining the search to simple graphs .	111
6.1	Performance box plot: diversity objectives (I) . . . . .	120
6.2	Performance box plot: diversity objectives (II) . . . . .	121
6.3	An island model for graph grammar evolution . . . . .	126
6.4	Performance box plot: island models . . . . .	128
6.5	Pareto frontier for the default configuration . . . . .	129
6.6	Pareto frontier for the 5 island model . . . . .	130
7.1	Performance box plot: different numbers of mutations for each production . . . . .	135
7.2	Performance box plot: different numbers of mutated productions .	136
7.3	Performance box plot: population size against generations . . . .	139
7.4	Performance plot over 50000 generations for 30 individual runs . .	140
7.5	A model of adaptive production search . . . . .	146
7.6	Performance box plot: adaptive productions swarms . . . . .	148
7.7	Performance box plot: different size objectives . . . . .	151
7.8	Pareto frontier for evolution without a primary size objective . . .	152
7.9	A very large solution for the Binomial-3 regression . . . . .	154
7.10	Histogram: component numbers in cellular graphs . . . . .	155

---

7.11	Histogram: usage and descendants of productions . . . . .	157
7.12	Verbosity box plot for main different configurations . . . . .	158
7.13	Copy ratio box plot for main different configurations . . . . .	159
7.14	Generational development of performance, size, and diversity statistics for the Binomial-3 regression problem . . . . .	162
7.15	Generational development of performance, size, and diversity parameters for the RBS circuit problem . . . . .	163
7.16	Generational development of performance, size, and diversity parameters for the pole balancing problem . . . . .	164
7.17	Generational development of performance, size, and diversity parameters for the CNT design problem . . . . .	165
8.1	A model of production inheritance . . . . .	175



# List of Tables

A.1	Results for the Binomial-3 Regression problem. . . . .	179
A.2	Results for the 6th-order Polynomial Regression problem. . . . .	179
A.3	Results for the Random Bit Sequence circuit design problem. . . .	181
A.4	Results for the 6-bit Multiplexer problem. . . . .	181
A.5	Results for the Backpropagation MLP (Iris dataset) problem. . . .	182
A.6	Results for the double Pole Balancing problem. . . . .	183
A.7	Results for the Computer Network Topology design problem. . . .	184



# Abstract

Artificial and natural instances of networks are ubiquitous, and many problems of practical interest may be formulated as questions about networks. Determining the optimal topology of a network is pertinent to many domains. Evolutionary algorithms constitute a well-established optimisation method, but they scale poorly if applied to the combinatorial explosion of possible network topologies. Generative representation schemes aim to overcome this by facilitating the discovery and reuse of design dependencies and allowing for adaptable exploration strategies. Biological embryogenesis is a strong inspiration for many such schemes, but the associated complexities of modelling lead to impractical simulation times and poor conceptual understanding. Existing research also predominantly focuses on specific design domains such as neural networks.

This thesis seeks to define a simple yet universally applicable and scalable method for evolving graphs and networks. A number of contributions are made in this regard. We establish the notion of directly evolving a graph grammar from which a population of networks can be derived. Compact cellular productions that form a hypergraph grammar are optimised by a novel multi-objective evolutionary design system called G/GRADE. A series of empirical investigations are then carried out to gain a better understanding of graph grammar evolution. G/GRADE is applied to four domains: symbolic regression, circuit design, neural networks, and telecommunications. We compare different strategies for composing graphs from randomly mutated productions and examine the relationship between graph grammar diversity and fitness, presenting both the use of phenotypic diversity objectives and an island model to improve this. Additionally, we address the issue of bloat and demonstrate how concepts from swarm intelligence can be applied to production selection and mutation to improve grammatical convergence. The results of this thesis are relevant to evolutionary research into networks and grammars, and the wide applicability and potential of graph grammar evolution is expected to inspire further study.





# Certification

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

As requested under Clause 14 of Appendix D of the *Flinders University Research Higher Degree Student Information Manual* I hereby agree to waive the conditions referred to in Clause 13(b) and (c), and thus

- Flinders University may lend this thesis to other institutions or individuals for the purpose of scholarly research;
- Flinders University may reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signed

Dated

Martin Holger Luerssen



# Acknowledgements

Although this thesis has only a single name on its cover, it would not have been accomplished without the support, academic and otherwise, of other people. I therefore wish to extend my sincerest thanks to the following individuals who have facilitated this work:

- My supervisor, Dr. David M. W. Powers, for his many creative, inspiring, sometimes impossible ideas, and who still let me have my own (stubborn) way
- My office roommate, Dr. Trent Lewis, for putting up with my many eccentric ideas and not cleaning them off the whiteboard
- My fellow postgraduates in the AI group and beyond, for being good friends as well as good intellectual sounding boards: Aaron Ceglar, Denise de Vries, Kirsty Kitto, Richard Leibbrandt, Takeshi Matsumoto, Darius Pfitzner, and Dongqiang Yang
- My anonymous paper reviewers, whoever they are, and innumerable fellow academics I've bumped into at conferences, for all their constructive feedback and interesting conversations
- My parents, Dr. Holger Luerssen and Maike-Vogt Luerssen, for the support and happiness they provided to me, unconditionally, over these tough years

