

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

STRUCTURE AND RANDOMNESS IN  
COMPLEX NETWORKS APPLIED  
TO THE TARGET SET SELECTION  
PROBLEM

A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF  
DOCTOR OF PHILOSOPHY  
IN  
COMPUTER SCIENCE  
AT MASSEY UNIVERSITY, MANAWATU,  
NEW ZEALAND.

Callum William Scudamore Lowcay

2014



## Abstract

Advances in technology have enabled the empirical study of large, so-called ‘complex’ networks with tens of thousands to millions of vertices, such as social networks and large communications networks. It has been discovered that these networks share a non-random topology characterised mainly by highly skewed, heavy-tailed degree distributions and small average distances between vertices. The work of this thesis is to attempt to leverage the well-known topological properties of complex networks to efficiently solve difficult NP-complete problems, with the aim of obtaining better or faster solutions than would be possible for general graphs.

Two related NP-complete problems are selected for study: the minimum target set problem, and the maximum activation set problem. Both problems relate to finding a ‘target set’ of vertices which is capable of initiating a spreading process (such as the spread of a rumour) that reaches a large proportion of the network. This thesis introduces several novel heuristics for these two problems inspired by the topology of complex networks. It is discovered that in many (but not all) cases it is possible to make relatively small alterations to the network that enable the computation of a considerably smaller target set than would be possible on general graphs.

The evaluation of the various heuristics is entirely experimental, which required the development of procedures to generate ‘random’ networks that can be used as experimental controls. This thesis includes a survey of several popular techniques for generating random networks and finds all but one (random rewiring) to be unsuitable as controls. The validity of random rewiring relies on a somewhat obscure theorem. Although a proof of the theorem (essentially an existence proof) is already known, this thesis offers a constructive algorithmic proof. The new proof advances on the old by providing an upper bound on the maximum number of rewiring operations required to transform between networks of the same degree-sequence, whereas an upper bound could not be determined under the old proof.



# Acknowledgements

I would like to thank my two supervisors Dr. Catherine McCartin and Professor Stephen Marsland. The timely completion of this thesis would not have been possible without their efforts.

I would also like to thank Professor Mike Langston and his research group for hosting and welcoming me at the University of Tennessee in Knoxville, where we collaborated on the minimum vertex cover problem.

I was supported for the duration of my PhD research by the Massey University Vice Chancellor's Doctoral Scholarship. I was also supported by Stephen Marsland's Marsden grant MAU0908.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	2
1.2	Choice of applications . . . . .	2
1.3	Aims and objectives . . . . .	4
1.4	Roadmap . . . . .	5
<b>2</b>	<b>Complex networks</b>	<b>7</b>
2.1	Network datasets . . . . .	8
2.1.1	Social networks . . . . .	9
2.1.2	Communication networks . . . . .	10
2.1.3	Biological networks . . . . .	11
2.1.4	A note on density . . . . .	12
2.2	Properties of complex networks . . . . .	12
2.2.1	Degree distribution . . . . .	12
2.2.2	Degree correlations . . . . .	16
2.2.3	Extension to dK distributions . . . . .	17
2.2.4	The small-world effect . . . . .	18
2.2.5	$k$ -Core and $k$ -shell decomposition . . . . .	21
2.2.6	Centrality metrics . . . . .	25
2.2.7	Community structure . . . . .	26
2.3	Summary . . . . .	27
<b>3</b>	<b>Methodology using random graphs</b>	<b>29</b>
3.1	Random graphs as null models . . . . .	30
3.1.1	Comparing random graph models: methodology . . . . .	32
3.2	Generative models . . . . .	34
3.2.1	Erdős-Rényi random graphs . . . . .	34
3.2.2	Configuration model . . . . .	36
3.2.3	Havel-Hakimi procedure . . . . .	40
3.2.4	Preferential attachment . . . . .	43



3.3	Randomizing models . . . . .	46
3.3.1	OK randomization . . . . .	48
3.3.2	Degree preserving rewiring . . . . .	49
3.3.3	Targeted rewiring . . . . .	52
3.4	Summary . . . . .	55
<b>4</b>	<b>Proofs of rewiring theorems</b>	<b>57</b>
4.1	Preliminaries . . . . .	58
4.1.1	An example . . . . .	59
4.2	Proof for simple (unconnected) graphs . . . . .	61
4.2.1	Alternating cycle construction . . . . .	61
4.2.2	Rewiring alternating cycles . . . . .	61
4.2.3	A special case . . . . .	63
4.3	Extension to simple connected graphs . . . . .	65
4.3.1	4-cycles . . . . .	65
4.3.2	The $C'$ construction . . . . .	65
4.3.3	Ordering the reductions . . . . .	67
4.4	Bounds . . . . .	69
4.5	Summary . . . . .	70
<b>5</b>	<b>The target set selection problem</b>	<b>73</b>
5.1	Minimum target set . . . . .	74
5.2	Maximum activation set . . . . .	76
5.3	Parametrization of minimum target set . . . . .	77
5.3.1	Vertex cover number . . . . .	79
5.3.2	Cluster edge deletion number . . . . .	84
5.3.3	Feedback edge set number . . . . .	86
5.3.4	Treewidth . . . . .	87
5.4	Summary . . . . .	88
<b>6</b>	<b>Heuristics for target set selection</b>	<b>91</b>
6.1	Greedy heuristics . . . . .	91
6.1.1	Hubs first . . . . .	92
6.1.2	Marginal gain . . . . .	93
6.1.3	The Shakarian-Paulo-Reichman algorithm . . . . .	94
6.1.4	Experimental comparison . . . . .	95
6.2	A distributed heuristic algorithm . . . . .	96
6.2.1	Experimental results . . . . .	100
6.3	Combining heuristics with parameters . . . . .	103

6.4	Summary . . . . .	112
<b>7</b>	<b>Shrinking a target set by edge augmentation</b>	<b>115</b>
7.1	Augmentation for minimum target set selection . . . . .	116
7.1.1	Experimental results . . . . .	117
7.2	Augmentation for maximum activation set . . . . .	122
7.2.1	Experimental evaluation . . . . .	124
7.2.2	Interpretation of results . . . . .	127
7.3	Summary . . . . .	128
<b>8</b>	<b>Conclusions</b>	<b>131</b>
8.1	Summary . . . . .	131
8.1.1	Target sets for complex networks . . . . .	133
8.2	Main findings . . . . .	134
8.3	Future work . . . . .	135
	<b>Bibliography</b>	<b>136</b>
<b>A</b>	<b>Description of software tools</b>	<b>145</b>
A.1	Architecture . . . . .	146
A.2	Job specification DSL . . . . .	147
A.2.1	An example . . . . .	148
A.3	Implementation . . . . .	149
A.4	Limitations and future work . . . . .	150

# List of Figures

2.1	Degree distribution of Physicists 1 . . . . .	14
2.2	K-core decomposition . . . . .	21
2.3	$k$ -shell decompositions of empirical networks . . . . .	24
3.1	Power-law fit for the projected configuration model . . . . .	38
3.2	Size of the largest component in the projected configuration model . . . . .	39
3.3	Assortativity in the projected configuration model . . . . .	40
3.4	Clustering in the projected configuration model . . . . .	41
3.5	Missing edges in Havel-Hakimi random graphs . . . . .	44
3.6	Assortativity in Havel-Hakimi random graphs . . . . .	44
3.7	Clustering in Havel-Hakimi random graphs . . . . .	45
3.8	Power-law fit in Barabási-Albert random graphs . . . . .	47
3.9	Assortativity in Barabási-Albert random graphs . . . . .	47
3.10	Clustering in Barabási-Albert random graphs . . . . .	48
3.11	The degree-preserving rewiring operation . . . . .	50
3.12	Assortativity following degree-preserving rewiring . . . . .	52
3.13	Clustering following degree-preserving rewiring . . . . .	53
3.14	Assortativity of 1K random graphs . . . . .	53
3.15	Clustering of 1K random graphs . . . . .	54
3.16	Power-law fit following targeted rewiring . . . . .	55
4.1	Notation . . . . .	58
4.2	A rewiring example . . . . .	60
4.3	Example sequence of rewiring operations . . . . .	60
4.4	Rewiring a 4-cycle . . . . .	62
4.5	Choosing four distinct consecutive vertices . . . . .	62
4.6	Reducing an alternating cycle . . . . .	63
4.7	The edge $ad$ is included in $X$ . . . . .	64
4.8	A 4-cycle that cuts the graph . . . . .	66
4.9	The $C'$ construction . . . . .	66
4.10	An impossible $C'$ construction . . . . .	68

4.11	Rewiring two 4-cycles . . . . .	68
4.12	4-cycles do not overlap . . . . .	68
4.13	4-cycle cases . . . . .	70
4.14	4-cycle cases for reducing alternating cycles . . . . .	71
5.1	The tipping model . . . . .	74
6.1	Comparison of heuristics for minimum target set selection with proportional thresholds . . . . .	97
6.2	Comparison of heuristics for minimum target set selection with constant thresholds . . . . .	98
6.3	A distributed heuristic for minimum target set . . . . .	99
6.4	Distributed target set algorithm with limited rounds . . . . .	101
6.5	Size of minimum feedback edge set as hubs only are removed . . . . .	106
6.6	Upper bound on vertex cover number as hubs only are removed . . . . .	107
6.7	Size of largest component as hubs only are removed . . . . .	108
6.8	Size of minimum feedback edge set as hubs and activation sets are removed	109
6.9	Upper bound on vertex cover number as hubs and activation sets are removed . . . . .	110
6.10	Size of largest component as hubs and activation sets are removed . . . . .	111
7.1	Augment by connecting neighbours of neighbours . . . . .	118
7.2	Augment by connecting distant vertices . . . . .	119
7.3	Augment within a known target set . . . . .	120
7.4	Edge augmentation for Maximum activation set . . . . .	126

# List of Tables

2.1	Network Datasets . . . . .	8
2.2	Common Metrics on Complex Networks . . . . .	13
2.3	Power-laws in Degree Distributions . . . . .	13
5.1	Vertex cover numbers . . . . .	83
5.2	Cluster edge deletion numbers . . . . .	86
5.3	Feedback edge numbers . . . . .	87
5.4	Treewidth bounds . . . . .	88