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

1 Introduction .................................. 1
   1.1 Motivation .................................. 2
   1.2 Choice of applications ......................... 2
   1.3 Aims and objectives ............................ 4
   1.4 Roadmap ................................... 5

2 Complex networks .............................. 7
   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

3 Methodology using random graphs ................. 29
   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
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
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Network Datsets</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>Common Metrics on Complex Networks</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Power-laws in Degree Distributions</td>
<td>13</td>
</tr>
<tr>
<td>5.1</td>
<td>Vertex cover numbers</td>
<td>83</td>
</tr>
<tr>
<td>5.2</td>
<td>Cluster edge deletion numbers</td>
<td>86</td>
</tr>
<tr>
<td>5.3</td>
<td>Feedback edge numbers</td>
<td>87</td>
</tr>
<tr>
<td>5.4</td>
<td>Treewidth bounds</td>
<td>88</td>
</tr>
</tbody>
</table>