

Robustness of hierarchical spatial critical infrastructure networks

Craig Robson

School of Civil Engineering and Geosciences

Newcastle University



Thesis submitted for the Degree of Doctor of Philosophy

April 2017

Abstract

The economic state and wellbeing of a nation is dependent upon the critical infrastructure networks that deliver resources, goods and services. However, these are increasingly exposed to a number of hazards, both natural and man-made, which threaten to disrupt their ability to function. It is essential that in order to develop long-term strategic plans of infrastructure provision we are able to understand their current robustness to such hazards.

The robustness of critical infrastructure networks has typically been investigated from a topological perspective as a means of simplifying the complexities associated with their analysis. Such work has led to many studies suggesting critical infrastructures exhibit a topological structure, from random to exponential degree distributions. However, often such analysis ignores the explicit spatial characteristics of the node and edge assets. Furthermore, the very nature of topological analysis means that flows/movements that take place over such networks cannot be considered.

This work addresses these weaknesses by extending traditional topological analysis to consider emergent properties critical infrastructure networks exhibit when considering higher-order connectivity and flows. An analysis of a suite of synthetic networks with a spectrum of topologies alongside real infrastructure spatial networks, in terms of their basic topology and high-order connectivity, shows that a number of critical infrastructure networks seem to be better characterised as hierarchical networks. Subsequent failure modelling reveals that such hierarchical networks responded in a dramatically different manner to perturbations; complete failure occurring approximately 19 and 34 percent sooner for random and targeted failures compared to random networks. Such poor robustness is further exacerbated when flow simulation modelling over the resulting hierarchical networks is undertaken, revealing particular sensitivity to cascading failures from spatial hazards. In light of these results, it is suggested that it is essential to improve the robustness of critical infrastructure networks that exhibit a hierarchical spatial organisation.

Acknowledgements

I give thanks to the support of family and friends during the period of study of which without this work would have not been possible.

I hereby acknowledge the help, assistance and guidance given by my supervisory team of Stuart Barr, Philip James and Alistair Ford.

Finally, I would like to thank the School of Civil Engineering and Geosciences, Newcastle University, for giving me the opportunity to study for a PhD through funding this work.

List of Publications

Conference Paper Publications

Robson C, Barr SL, James P, Ford A. Resilience of Hierarchical Critical Infrastructure Networks. In: *International Symposium on Next Generation Infrastructures*. 2014, Vienna, Austria.

Robson C, Barr SL, James P, Ford A. Exploring the vulnerability of spatial infrastructure networks compared to aspatial specimens with a focus on hierarchical organisation. In: *Student Conference on Complex Systems*. 2014, University of Sussex, Brighton, UK.

Robson C, Barr SL, James P, Ford A. Identifying the presence of hierarchical structure in infrastructure networks. In: *ITRC Early Career Researchers Conference: Infrastructure Delivery in an Uncertain Future*. 2012, Clare College, Cambridge, UK: Infrastructure Transitions Research Consortium UK (ITRC).

Table of Contents

ABSTRACT.....	I
ACKNOWLEDGEMENTS.....	III
LIST OF PUBLICATIONS.....	IV
TABLE OF CONTENTS.....	V
LIST OF KEY FIGURES.....	IX
LIST OF KEY TABLES.....	XII
LIST OF ABBREVIATIONS AND NOTATIONS.....	XIV
GLOSSARY OF NOTATION.....	XV
CHAPTER 1: INTRODUCTION.....	1
1.1 RESEARCH INTRODUCTION	1
1.2 AIMS AND OBJECTIVES.....	4
1.3 THESIS STRUCTURE	5
CHAPTER 2: INFRASTRUCTURE NETWORK MODELLING.....	6
2.1 INFRASTRUCTURE SYSTEMS AND NETWORKS	6
2.2 INFRASTRUCTURE ROBUSTNESS AND RESILIENCE	7
2.3 MODELS OF INFRASTRUCTURE NETWORKS.....	11
2.3.1 <i>Graph theory</i>	13
2.3.2 <i>Measuring network characteristics</i>	15
2.3.3 <i>Graph models</i>	18
2.3.4 <i>Hierarchical networks and models</i>	22
2.4 STUDIES OF ROBUSTNESS AND RESILIENCE.....	25
2.4.1 <i>Robustness of graph models</i>	26
2.4.2 <i>Robustness of infrastructure networks</i>	28
2.4.3 <i>Spatial infrastructure robustness</i>	33
2.5 DISCUSSION AND RESEARCH CHALLENGES.....	38
2.6 CONCLUSION.....	40
CHAPTER 3: METHODOLOGICAL FRAMEWORK.....	41
3.1 INTRODUCTION	41
3.2 OVERALL EXPERIMENTAL DESIGN.....	41
3.3 SYNTHETIC SUITE OF GRAPHS	42
3.3.1 <i>Erdos-Renyi graph model</i>	46

3.3.2	<i>GNM random graph model</i>	46
3.3.3	<i>Watts-Strogatz small-world graph model</i>	46
3.3.4	<i>Barabasi-Albert scale-free graph model</i>	48
3.3.5	<i>Hierarchical random</i>	48
3.3.6	<i>Hierarchical random +</i>	49
3.3.7	<i>Hierarchical communities</i>	51
3.3.8	<i>Tree</i>	52
3.4	STATISTICAL COMPARISON OF GRAPH TYPES	53
3.4.1	<i>Degree distributions</i>	53
3.4.2	<i>Metrics</i>	53
3.4.3	<i>Multivariate metric analysis</i>	58
3.4.4	<i>Transformed divergence</i>	58
3.5	TOPOLOGICAL FAILURES	59
3.5.1	<i>Methods of node selection</i>	61
3.5.2	<i>Recording failure behaviour</i>	63
3.6	REAL-WORLD SPATIAL INFRASTRUCTURE DATA.....	64
3.6.1	<i>Air networks</i>	65
3.6.2	<i>Communication networks</i>	65
3.6.3	<i>Energy networks</i>	66
3.6.4	<i>Rail networks</i>	66
3.6.5	<i>River networks</i>	67
3.6.6	<i>Road networks</i>	67
3.7	IDENTIFYING HIERARCHICAL INFRASTRUCTURE NETWORKS.....	69
3.8	ENHANCED NETWORK REPRESENTATION.....	69
3.9	CAPACITY CONSTRAINED FAILURE MODELLING	71
3.9.1	<i>Developed failure model</i>	72
3.9.2	<i>Triggering cascading failures</i>	76
3.9.3	<i>Recording failure behaviour</i>	77
3.9.4	<i>Analysis scenarios</i>	78
3.10	HIERARCHICAL FLOW ROBUSTNESS MODELLING	80
3.10.1	<i>Hierarchical connectivity modelling</i>	82
3.10.2	<i>Spatial hazard modelling</i>	83
3.11	SOFTWARE STACK	84
3.11.1	<i>Framework</i>	84
3.11.2	<i>The network database schema</i>	86
3.11.3	<i>The analysis database schema</i>	89
3.11.4	<i>Developed modules</i>	92
3.11.5	<i>Graphical User Interface (GUI)</i>	97

3.12	CONCLUSIONS	99
CHAPTER 4: RESULTS.....		101
4.1	INTRODUCTION	101
4.2	CHARACTERISTICS OF HIERARCHICAL GRAPHS	101
4.2.1	<i>Introduction.....</i>	<i>101</i>
4.2.2	<i>Graph degree distributions.....</i>	<i>102</i>
4.2.3	<i>Assessment of graph metrics</i>	<i>104</i>
4.2.4	<i>Statistical graph similarity</i>	<i>108</i>
4.2.5	<i>Topological hierarchical graph robustness</i>	<i>111</i>
4.2.6	<i>Failure characteristics of hierarchical graphs</i>	<i>115</i>
4.3	THE HIERARCHICAL CHARACTERISTICS OF INFRASTRUCTURE NETWORKS	118
4.3.1	<i>Introduction.....</i>	<i>118</i>
4.3.2	<i>Degree distributions of critical infrastructure networks.....</i>	<i>119</i>
4.3.3	<i>Assessment of critical infrastructure network metrics</i>	<i>123</i>
4.3.4	<i>Topological robustness of infrastructure networks.....</i>	<i>135</i>
4.3.5	<i>Failure characteristics of infrastructure networks.....</i>	<i>138</i>
4.4	CAPACITY CONSTRAINED CASCADING FAILURES ON HIERARCHICAL GRAPHS	142
4.4.1	<i>Introduction.....</i>	<i>142</i>
4.4.2	<i>Scenarios</i>	<i>145</i>
4.4.3	<i>Cascading failures results</i>	<i>148</i>
4.5	ROBUSTNESS OF A HIERARCHICAL CRITICAL INFRASTRUCTURE NETWORK	157
4.5.1	<i>Introduction.....</i>	<i>157</i>
4.5.2	<i>Hierarchical dependency of the electricity network.....</i>	<i>157</i>
4.5.3	<i>Robustness to geographic hazards.....</i>	<i>160</i>
4.6	SUMMARY	177
CHAPTER 5: DISCUSSION.....		179
5.1	INTRODUCTION	179
5.2	CHARACTERISTICS OF HIERARCHICAL GRAPHS AND NETWORKS.....	179
5.2.1	<i>Graph models and characteristics.....</i>	<i>179</i>
5.2.2	<i>Critical spatial infrastructure networks</i>	<i>182</i>
5.2.3	<i>Robustness analysis.....</i>	<i>186</i>
5.3	CASCADING FAILURE ANALYSIS	189
5.4	ANALYSIS OF A HIERARCHICAL ELECTRICITY NETWORK	191
5.5	FUTURE ANALYSIS OPPORTUNITIES	193
5.6	CONCLUSION	195
CHAPTER 6: CONCLUSION.....		196

6.1	INTRODUCTION	196
6.2	MAIN FINDINGS	196
6.2.1	<i>Hierarchical networks and critical infrastructures</i>	196
6.2.2	<i>Hierarchical graphs</i>	197
6.2.3	<i>Hierarchical infrastructure networks</i>	197
6.2.4	<i>Robustness of hierarchical infrastructure networks</i>	198
6.3	FUTURE WORK.....	198
6.3.1	<i>Geographic and spatial graph models</i>	198
6.3.2	<i>Cascading failures on hierarchical graphs</i>	199
6.3.3	<i>Modelling of infrastructure flows</i>	200
6.3.4	<i>Spatial hazard modelling</i>	200
6.3.5	<i>Improving the robustness of hierarchical networks</i>	201
6.3.6	<i>Robustness to dependencies in hierarchical infrastructures</i>	202
6.3.7	<i>Software framework</i>	203
6.4	KEY FINDINGS AND IMPLICATIONS.....	203
	BIBLIOGRAPHY	206
	APPENDIX A: SUITE OF SYNTHETIC NETWORKS.....	225
	APPENDIX B: SUITE OF CRITICAL INFRASTRUCTURE NETWORKS	232
	APPENDIX C: DETAILS ON UTILISING DEVELOPED SOFTWARE	250
	APPENDIX D: SYNTHETIC GRAPHS ANALYSIS RESULTS.....	257
	APPENDIX E: INFRASTRUCTURE ANALYSIS RESULTS.....	293
	APPENDIX F: NX_PGNET_ATTTS DOCUMENTATION	328
	APPENDIX G: GRAPHICAL USER INTERFACE (GUI) DOCUMENTATION.....	338

List of Key Figures

Chapter 2

Figure(s)	Showing	Page(s)
Figure 2.1	An illustration of the properties of resilience (McDaniels <i>et al.</i> , 2008).	9
Figure 2.2	Diagrammatic definition of resilience (Bruneau <i>et al.</i> , 2003).	9
Figure 2.8	Rewiring a regular lattice (a) to create the small-world network (b) and (c) where edges have been added instead of existing edges being re-wired (Newman, 2003b).	21
Figure 2.10	Two hierarchical networks; (a) a tree with no cycles, and (b) a tree with increased redundancy creating cycles.	24
Figure 2.11	The hierarchical community model. (a) Shows the base level community of nodes and (b) shows how these are combined. (c) shows the third level where the community in (b) is used to generate a much larger network (Ravasz and Barabasi, 2003).	25

Chapter 3

Figure(s)	Showing	Page(s)
Figure 3.1	Employed graph spectrum.	43
Figure 3.2	The spectrum of graph models through a network diagram and a degree distribution plot, where $P(k)$ is the fraction of nodes with degree k .	45
Figure 3.11	Process diagram of the developed topological failure model.	61
Figure 3.12	Exemplifying the three different node selection methods during a topological failure model using the same example network at three selected epochs, 1, 2 and 6.	62
Figure 3.22	Developed capacity constrained cascading failure model.	73
Figure 3.25	Capacity constrained cascading failure example where node 5 has a demand of five and node 1 has a supply of five, with each node and edge having a capacity of two. (a) shows the network at $T(0)$, (b) after the rerouting of flows after the removal of edge (1,3) as a trigger, (c) shows the failure of edge (1,5) as a result of being over capacity and the flow on each edge after the re-routing of flows again, and (d) shows the results of the failure of those edges over capacity, with no route from the supply node, node 1.	76
Figure 3.27	Electricity transmission and distribution network for England and Wales (ITRC, 2013).	82
Figure 3.28	Developed software framework.	84
Figure 3.34	Database schema as employed for the research.	90

Chapter 4

Figure(s)	Showing	Page(s)
Figure 4.1	Example degree distribution plots for the eight graph models.	103
Figure 4.2/ Figure 4.3/ Figure 4.4	The standard deviation ellipse for the three multivariate metric comparisons for the eight graph models; assortativity coefficient and maximum betweenness centrality, assortativity coefficient and cycle basis per node, and maximum betweenness centrality and cycle basis per node.	105/107/ 108
Figure 4.5	Average percentage of nodes removed for each graph to become empty for (a) random node selection, (b) degree based node selection and (c) betweenness centrality based node selection.	114
Figure 4.6/ Figure 4.7	The response of the (i) non-hierarchical graph models and the (ii) hierarchical graph models.	117/118
Figure 4.8/ Figure 4.9/ Figure 4.10/ Figure 4.11/ Figure 4.12/ Figure 4.13	Degree distribution plots for selected infrastructure networks; rail, air, river, national roads, regional roads and energy networks.	120/121/ 121/122/ 123/123
Figure 4.14/ Figure 4.23/ Figure 4.24	The multivariate metric results for the infrastructure networks with the standard deviation ellipses for the synthetic graph models; assortativity coefficient and maximum betweenness centrality, assortativity coefficient and cycle basis per node, and maximum betweenness centrality and cycle basis per node.	124/133/ 134
Figure 4.25	Plots showing the average response across the infrastructure groups to the three failure models. Failure plots for each infrastructure network can be found in Appendix E Section E.3.	136
Figure 4.26/ Figure 4.27/ Figure 4.28/ Figure 4.29/ Figure 4.30	Example responses from individual infrastructure networks to the topological failure model including air, river, energy, rail and road networks.	139/139/ 140/141/ 142
Figure 4.32	Results from cascading failure simulation (scenario (i)) over the synthetic network exemplars with a single trigger edge, a demand of four and capacities of four.	149
Figure 4.33	Results of the cascading failure simulation (scenario (ii)) over the selected synthetic networks where the demand was four, the capacities were based on structure and a single trigger edge was removed.	150
Figure 4.34	Result of the cascading failure simulations (scenario (iii)) on the selected synthetic networks where a supply of four has been used, with capacities set to four and multiple trigger edges removed.	151
Figure 4.35	Percentage of edges removed as trigger edges for the analysed synthetic graphs for scenario (iii).	152

Figure 4.36	The results from the cascading failure simulations (scenario iv) on the selected synthetic networks where a single demand of four was used, capacities were based on the graph structure and multiple trigger edges were removed.	153
Figure 4.37	Percentage of edges removed as trigger edges for the cascading failure scenario (iv).	153
Figure 4.38	The results from the cascading failure simulations (scenario (v)) over the selected synthetic networks where the demand is six, the capacities are four, and a single trigger edge is removed.	155
Figure 4.39	The results from the cascading failure simulation (scenario (vi)) over the selected synthetic networks where the edge capacities were set to four, the demand six and multiple trigger edges were removed.	156
Figure 4.45	Second-order failures for realisations in scenario set A.	166
Figure 4.48	Second-order substation and edge failures for the five realisations in scenario set B.	171
Figure 4.51	Second-order asset failures for the five hazard realisations in scenario set C.	176

List of Key Tables

Chapter 3

Table(s)	Showing	Page(s)
Table 3.1	Graph models and the parameters required for generation of the graphs. More details on the parameters are given in the model specific sub-sections, 3.3.1 to 3.3.8.	45
Table 3.3	Details of the graph based assignment of node and edge capacities.	79

Chapter 4

Table(s)	Showing	Page(s)
Table 4.2	The mean values for each of the eight graph types across the graph metrics computed.	106
Table 4.3	The transformed divergence analysis of the distribution of the metric values for each graph type. * Results cannot be computed as the TREE model has no cycles.	109
Table 4.4	Mean percentage of nodes removed for the node removal options for the hierarchical and non-hierarchical network groups and the percentage difference between these.	113
Table 4.9	Average metric values for the infrastructure network groups.	135
Table 4.11	A breakdown of the nodes and edges in the electricity transmission and distribution network for England and Wales.	159
Table 4.12	First-order failure counts for the realisations in scenario set A. Figure 4.44 shows the location of the each hazard areas.	163
Table 4.13	Second-order asset failure counts for the realisations in scenario set A. Mapped in Figure 4.45.	165
Table 4.14	Average and maximum distance of the second-order substation failures from the hazard of the realisations in scenario set A.	165
Table 4.15	First-order asset failure counts for the five realisations in scenario set B. Figure 4.46 shows the hazard areas for the five realisations.	168
Table 4.16	Second-order asset failures for the five realisations in scenario set B, also mapped in Figure 4.48.	169
Table 4.17	Average and maximum distance of the second-order substation failures from the hazard areas for each realisation in scenario set B.	170

Table 4.18	First-order network asset failures for each realisation in scenario set C. The location of the hazards for these realisations are shown in Figure 4.49.	172
Table 4.19	Second-order network asset failure counts for each realisation in scenario set C.	174
Table 4.20	Average and maximum distance from hazard areas of second-order substation failures for the five realisations in scenario set C.	177

List of Abbreviations and Notations

Abbreviations

GIS	Geographic Information System
ITRC	Infrastructure Transitions Research Consortium
Kv	Kilo volts
OS	Ordnance Survey
OSM	Open Street Map

Graph models

ER	Erdos-Renyi
BA	Barabasi-Albert
WS	Watts-Strogatz
HR	Hierarchical Random
HR+	Hierarchical Random +
HC	Hierarchical Community

Graph/network metrics

AC	Assortativity coefficient
MBC	Maximum betweenness centrality
CB	Cycle basis per node

Glossary of Notation

Notation	Description
Graph notation	
G	Graph/network
N	Node set
E	Edge set
n	Node
e	Edge
Graph generation models	
E_{new}	New edge
$E_{newadded}$	Number of edges added
G_{edges}	Edges in G
$N1, N2$	Node 1, Node 2
$N1edges$	Edges incident on Node 1
p	Probability
Transformed divergence statistic	
C_i	Covariance matrix for i
C_i^{-1}	Inverse of the covariance matrix for i
D_{ij}	Divergence between nodes i and j
tr	Trace function
μ_i	Mean vector of the metric values for i
Enhanced network representation	
F, n_F, e_F	Flow; node flow and edge flow
FC, n_{FC}, e_{FC}	Flow capacity; node flow capacity and edge flow capacity
W, n_W, e_W	Weight; node weight and edge weight
B, n_B	Buffering; node buffering
BC, n_{BC}	Buffering capacity; node buffer capacity
L, n_L	Latency; node latency
D, e_D	Length; edge length
S, e_S	Stacking (queuing); edge queue

R, n_R	Role; node role
Capacity constrained cascading model	
C	Capacity
K	Trigger edge
T	Epoch

Chapter 1: Introduction

1.1 Research introduction

Critical infrastructure networks are vital to the functioning of societies (Boin and McConnell, 2007) and “if lost would lead to severe economic or social consequences or to loss of life” (Cabinet Office, 2010). Infrastructure networks provide services which are used and relied upon in nearly all aspects of life (Sterbenz *et al.*, 2011), making their security vital in order to avoid failure or disruption that impacts on national security, economic security and public health and wellbeing (107th Congress, 2001; Rinaldi, 2004; Schulman and Roe, 2007; HM Treasury, 2010). Critical infrastructures are spread over nine sectors including energy, food, water, transportation, communications, emergency services, health care, financial services and government (Cabinet Office, 2010). More concisely, networks such as those for electric, gas and water distribution, roads, rail and air as well as telecommunications are generally regarded as those physical infrastructure networks which are critical (Rinaldi, 2004; Ulieru, 2007; Doglioni *et al.*, 2009).

Recent failures of critical infrastructure networks, such as energy distribution and telecommunications (Rinaldi *et al.*, 2001; Andersson *et al.*, 2005) have shown the extent to which modern societies rely upon them and to how vulnerable these networks can be when exposed to hazard events or smaller asset failures. For example, a electricity blackout of Italy in 2003 lasted 19 hours (power started being restored after 1.5 hours) and is estimated to have had an economic cost of €1,182 million (Royal Academy of Engineering, 2014). Four deaths were reported, along with severe transport impacts, with train services disrupted and flights cancelled. The extent of the impacts was lessened with the blackout occurring during the night, with services gradually restored across Italy. The blackout was triggered following a fault on the power network in Switzerland which fed transmission lines to Italy, causing the failure of the lines Italy relied upon as an importer of Swiss electric, triggering a cascading failure (Berizzi, 2004; Royal Academy of Engineering, 2014). Similarly, the failure of a high-voltage line, or number of, in Ohio triggered a blackout which is estimated to cost the US \$6billion economically and Canada lost 18.9million work hours in total (Royal Academy of Engineering, 2014). The blackout shutdown oil refineries and pipelines, transport systems and key manufacturing industries for on average for over 24 hours, all ultimately due to the failure of a single utility company to ensure trees were properly cut back from transmission lines and a lack

of built in resilience in the electric supply network (Electricity Consumers Resource Council, 2004).

Large natural hazards have caused significant impacts on critical infrastructure networks, with the disruption and destruction caused by events such as earthquakes, flooding and ice storms impacting on the ability for infrastructure networks to continue to function. The 1998 ice storm in Canada demonstrated how dependent and vulnerable it is to power outages (Purcell and Fyfe, 1998) with 16% of the population left without power for up to three weeks at an economic cost in excess of \$4.7billion (Chang *et al.*, 2007). More recently the 2011 earthquake off the coast of Japan has highlighted the vulnerability of network assets to hazards with the water inundation at Fukushima nuclear power plant, the subsequent failure of the plant, which in response led to the closure of similar plants, leading to rolling blackouts due to a reduction of 30% in electricity capacity (Royal Academy of Engineering, 2014). Compounding this, the transmission networks across mainland Japan, between the East and West are effectively separate networks meaning shortfalls could not be compensated by areas where excess was available (Scawthorn *et al.*, 2011). The lack of capacity and redundancy in the transmission network, and the loss in confidence of Nuclear generation led to significant economic impacts, though not quantifiable due to the disaster itself, though industrial production was estimated to be down 15% during the month of the disaster and subsequent period of blackouts (Royal Academy of Engineering, 2014). Hazards such as earthquakes and floods can affect multiple critical infrastructures directly (Little, 2003), rather than just one in the case of an asset fault, exacerbating the potential impacts on the infrastructure networks.

Extreme weather events such as the ice storm mentioned previously affect geographic areas with the critical infrastructures within the area exposed to the hazard. These events extend to hazards such as flooding, forest fires, drought and storms, many a result directly or indirectly, of extreme weather. The impacts of such events on critical infrastructure networks can be significant, as exemplified by the 1998 ice storm in Canada (Chang *et al.*, 2007), as well as the flooding in the UK in 2007 (Cabinet Office, 2008) and hurricane Katrina in 2005 (Leavitt and Kiefer, 2006). Such events were a result of extreme weather, and with the frequency and scale of such weather expected to increase as a result of a changing climate (Royal Academy of Engineering, 2011) the need for an increasing robustness and resilience to these events is becoming critical.

Understanding how extreme weather events as well as asset faults and failures impact on infrastructure networks is critical to reduce the impacts on the networks and services which they provide. However infrastructure networks are complex systems when considered as stand-

alone entities, though do not exist as such, and instead are interconnected, with networks relying on each other (Rinaldi, 2004), forming an interconnected web of networks (Figure 1.1) which deliver the services modern societies depend upon. These interdependencies between networks have developed through increasing demands, technological advances and the drive for greater efficiency (Chiaradonna *et al.*, 2009) with societies now relying upon these (Rinaldi, 2004).

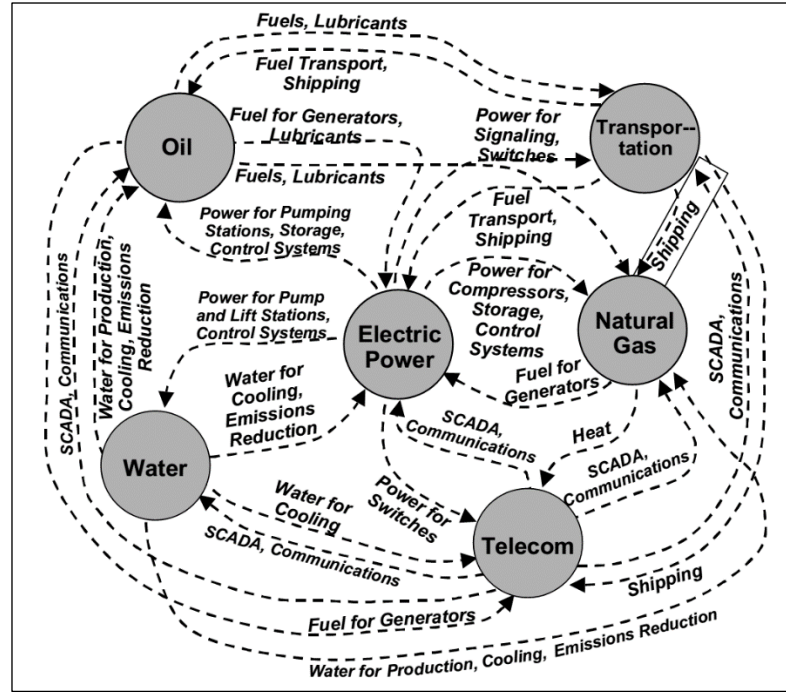


Figure 1.1: The dependencies and interdependencies between a number of critical infrastructures (Little, 2003).

In order to try and understand the functionality of critical infrastructure networks modelling and simulation is increasingly employed (Amaral and Ottino, 2004; Rinaldi, 2004). In particular, graph theory/models have been used for the analysis and simulation of networks (Newman, 2003b; Amaral and Ottino, 2004; Jungnickel, 2004; Boccaletti *et al.*, 2006), since the original ‘random’ graph model of Erdos and Renyi (1959). Since this work, graph theory and models have been extensively developed and applied for the analysis and simulation modelling of networks (Newman, 2003b), including critical infrastructures (Amaral and Ottino, 2004). In particular, since the first random models (Erdos and Renyi, 1959) small-world (Watts and Strogatz, 1998) and scale-free models (Barabasi and Albert, 1999) have been extensively investigated. These graph models have been associated with real-world critical infrastructure networks (Albert and Barabasi, 2002; Newman, 2003b), such as the internet and electricity distribution networks (Watts and Strogatz, 1998; Barabasi and Albert, 1999; Cohen *et al.*, 2001).

However, there is a growing literature around networks having a hierarchical structure (Ravasz *et al.*, 2002; Ravasz and Barabasi, 2003; Trusina *et al.*, 2004; Clauset *et al.*, 2008; Lancichinetti *et al.*, 2009). Analysis has shown hierarchical structures existing in biological networks (Costa *et al.*, 2008) and in particular in metabolic networks (Ravasz *et al.*, 2002; Holme *et al.*, 2003; Ravasz and Barabasi, 2003), as well as social networks (Watts *et al.*, 2002; Clauset *et al.*, 2008). However, analysis of hierarchical organisation in critical infrastructure networks is less strong (Costa and Silva, 2006). Research such as that done by Yerra and Levinson (2005) has suggested hierarchies exist in road networks, and some research has indicated that a hierarchical organisation exists in airline networks (Bagler, 2008a), though Ravasz and Barabasi (2003) have suggested that networks which a geographical organisation do not have a modular hierarchical organisation as found in some networks. This apparent contradiction may suggest that more than one form of hierarchy exists, with the modular organisation potentially differing in characteristics from that found by Yerra and Levinson (2005) and Bagler (2008a) in spatial infrastructure networks. The underlying reasons for the difference in the identified hierarchies could vary, from differences in network evolution for the best structure to meet the purpose of the network, to possible constraints enforced by the network being inherently spatial. These could include the cost of new links for infrastructure networks and the constraints imposed by geographic boundaries, be that physical or political.

Hierarchical networks have been suggested to be vulnerable to failures, but especially to the failure of the most critical nodes (Helbing *et al.*, 2006a; Wuellner *et al.*, 2010). With infrastructure networks being vital to so many aspects of modern societies (Rinaldi, 2004) it is considered imperative that these are robust to all forms of hazards, whether natural hazards, asset breakdowns or targeted attacks (Little, 2003). However, despite an emerging set of literature on the characteristics of hierarchical networks (Barabasi *et al.*, 2003; Clauset *et al.*, 2008), little of this has focused or been directed towards critical infrastructure networks, with only a few studies approaching this potential field of research (Yerra and Levinson, 2005).

1.2 Aims and objectives

The research aim of this thesis is to identify the hierarchical organisation of critical spatial infrastructure networks and the robustness of these to a range of failure scenarios. To address this aim, four objectives have been defined:

1. Review the research field pertaining to hierarchical networks and graph models and their application in the analysis of critical spatial infrastructure networks.

2. Investigate the properties of hierarchical graphs to identify the characteristics which makes them recognisable from non-hierarchical graphs.
3. Identify examples of hierarchically organised critical spatial infrastructure networks using the outcomes from objective 2.
4. Explore the robustness of hierarchical infrastructure networks to perturbations and the reasons why such networks behave differently to those of other topological structures.

This research will investigate the extent of a hierarchical organisation in critical spatial infrastructure networks and the effect this has on the robustness of these infrastructure networks to perturbations. To achieve this the characteristics of hierarchically organised graphs must be investigated to identify how these can be recognised from those without a hierarchical organisation. This then allows for the identification of hierarchically organised critical infrastructure networks and the effect the hierarchical organisation has on the robustness of the infrastructure networks and their ability to continue to deliver services/function as designed.

1.3 Thesis structure

The remainder of this thesis addresses the aims and objectives as set out above in Section 1.2 and is split into a further five chapters. The following chapter, Chapter 2, reviews the previous research which has been undertaken with regard to critical spatial infrastructure networks, the characterisation of these networks and the analysis undertaken to establish and improve the robustness of infrastructure networks to hazards and failures. Chapter 3 presents the methodological approach used to address the aims and objectives of this research using the knowledge acquired from Chapter 2. The suite of software, including tools, modules and database schema's developed for the research are also detailed in Chapter 3. The results from the analysis set out in Chapter 3 are then presented in Chapter 4. Chapter 5 then discusses the results and the major findings from the analysis undertaken as well as critiquing the employed methods. Chapter 6 then finally presents the conclusions of the research including a summary of the results along with potential future areas of research.

Chapter 2: Infrastructure network modelling

2.1 Infrastructure systems and networks

Infrastructure systems are those systems which provide the critical services we depend upon including energy, water, telecommunication and transportation (Rinaldi, 2004; Ulieru, 2007; Doglioni *et al.*, 2009). These systems are critical, influencing both the economic and social wellbeing of society through the services and commodities they provide (107th Congress, 2001; Rinaldi, 2004; Boin and McConnell, 2007; Schulman and Roe, 2007; HM Treasury, 2010; Sterbenz *et al.*, 2011).

The term infrastructure system describes the functioning of the infrastructure, or multiple infrastructures, from the physical assets to the operational control of the infrastructure and the behaviour of it (Rinaldi *et al.*, 2001; Egan, 2007; Richards *et al.*, 2007), as well as the interactions humans have with the infrastructure (Egan, 2007). At all aspects of the infrastructure system failures/errors can occur which affect the ability of the infrastructure to function (Rinaldi *et al.*, 2001; Little, 2003), from the breakdown of individual components to the failure of control systems. However, it is the failure of physical assets that has triggered some of the worst recorded infrastructures failures, such as the blackouts in Europe and North America in 2003 (Andersson *et al.*, 2005). Physical network assets are exposed to a range of known hazards including natural events like earthquakes and weather related events such as wind storms and flooding (Rinaldi *et al.*, 2001; Rinaldi, 2004).

The control systems for infrastructure systems are becoming increasingly computerised (Little, 2003), including SCADA (Supervisory Control And Data Acquisition) systems for the electric distribution networks for example (Rinaldi *et al.*, 2001; Bobbio *et al.*, 2010). These systems are also under threat, and increasingly so, from cyber threats which target the control systems aiming to disrupt the operation of infrastructure systems (Ten *et al.*, 2008; Bronk, 2015; Genge *et al.*, 2015). An increasing number of SCADA systems rely on the internet (Ten *et al.*, 2008), which itself has physical assets which are vulnerable to failures, thus enabling a robustness to the cyber threat is alone not enough.

Physical assets are those assets which constitute the network which delivers the service, which for an electricity network for example, includes, but is not limited to the power stations, substations as well as transmission and distribution lines (Wang *et al.*, 2012). The network itself is considered decoupled from the systems used to control, operate and manage it, such as

SCADA systems (Rinaldi, 2004; Zio, 2014), as well as those which influence how it develops, with these part of the wider infrastructure system. Control systems can affect a networks response to perturbations, with these managing the networks in real-time to limit the impact on the networks functionality (Rinaldi *et al.*, 2001; Eusgeld *et al.*, 2011; Merabti *et al.*, 2011). The control systems also can cause disruptions to the systems themselves, with these being vulnerable to failures, preventing the system from functioning (Ulieru, 2007; Velykiene and Jones, 2011). By excluding this aspect of infrastructure systems, as well as the possible human interventions associated with the control of the systems from any analysis the operational behaviour of the system is ignored. This therefore limits the ability of any analysis to realistically model the behaviour of the systems and especially their response to perturbations, with the mechanisms whereby interventions may be made to reduce the impact of failures on the networks operational functionality ignored in the analysis. However, by not attempting to model the control systems it allows for subsequent analysis to focus on the built aspect of the network system, the physical infrastructure which exists and how this aspect might be affected by hazards/perturbations. To this end, the physical assets can be modelled using graph theory approaches where the network, G , is modelled using a set of nodes, N and a set of edges, E , $G = \{N, E\}$. Assets such as power stations and substations, those located as a single point form the node set, $n \in N$, and assets such as transmission/distribution lines the edge set, $e \in E$. Graph theory for network modelling is reviewed in much more detail in Section 2.3.

2.2 Infrastructure robustness and resilience

Critical infrastructure networks are exposed to a large range of hazards including those caused by human interventions, mechanical failures of equipment and natural events such as floods and extreme winds (Little, 2002). All such hazards can result in failures which effect the service provided, with what may seem like minor failures potentially having major impacts (Merabti *et al.*, 2011). The ability of infrastructure systems and networks to continue to deliver the services is thus critical. A number of terms have been associated with the measurement of the ability for infrastructure networks to withstand perturbations; resilient (resilience), robustness, reliability and vulnerability, all of which will be discussed in this section in relation to the analysis of critical infrastructure networks.

The term resilience was first defined by Holling (1973) who defined resilience as “determining the persistence of relationships within a system and is a measure of the ability of these systems to absorb changes of state variables, driving variables, and parameters, and still persist”.

Although from the field of ecology, the definition refers to the ability of a system to cope with changes, and appears equally applicable to infrastructure networks/systems as it is ecology. Within the field of ecology Walker *et al.* (2004) have more recently refined the definition of resilience as ‘the capacity of a system to absorb disturbance and reorganize while undergoing change so as to still retain essentially the same function, structure, identity, and feedbacks’. Unlike the definition of Holling (1973), this refers to the ability of the system to not only continue to function while perturbed, but to re-organise and continue to deliver the same level of performance. This feature, the ability of a system to return to its normal operating state following perturbations, is also suggested in the definition by Pimm (1984) who defines resilience as how fast a system returns to equilibrium (its normal state) following a change to its normal operating state.

Within the fields of engineering and infrastructure systems analysis, a range of definitions have also been proposed that differ somewhat from the ecological perspective. McDaniels *et al.* (2008) defines resilience as the capacity to absorb shocks while maintaining function, Leu *et al.* (2010) defines it as ‘the system’s ability to keep focusing on and meeting key objectives when faced with challenges in the surrounding operating environment’, Sterbenz *et al.* (2011) defines it ‘as the ability of a network to provide desired service even when challenged by attacks, large-scale disasters, and other failures’ and Hosseini *et al.* (2016) states that the term resilience implies ‘the ability of an entity or system to return to normal condition after the occurrence of an event that disrupts its state’. These all provide a similar definition to that of Holling (1973), with the focus on the ability to continue to function and provide a service while being perturbed. However, it is worth noting that they have no explicit mention of the ability of the system/infrastructure to recover, as suggested by the definition provided by Pimm (1984). McDaniels *et al.* (2008) does state that there are many aspects of resilience, with robustness and rapidity (the time required for the system to return to its normal state of operation), being important measures. This is demonstrated in Figure 2.1 where the rapidity of the network to respond to a failure, illustrated by a drop in system function, is shown as a measure of time. The robustness is shown to be the amount of the network/system unaffected by the failure. Two dashed lines also highlight the possible alternative effects of a failure if early mitigation methods are employed or if greater post failure adaption is undertaken.

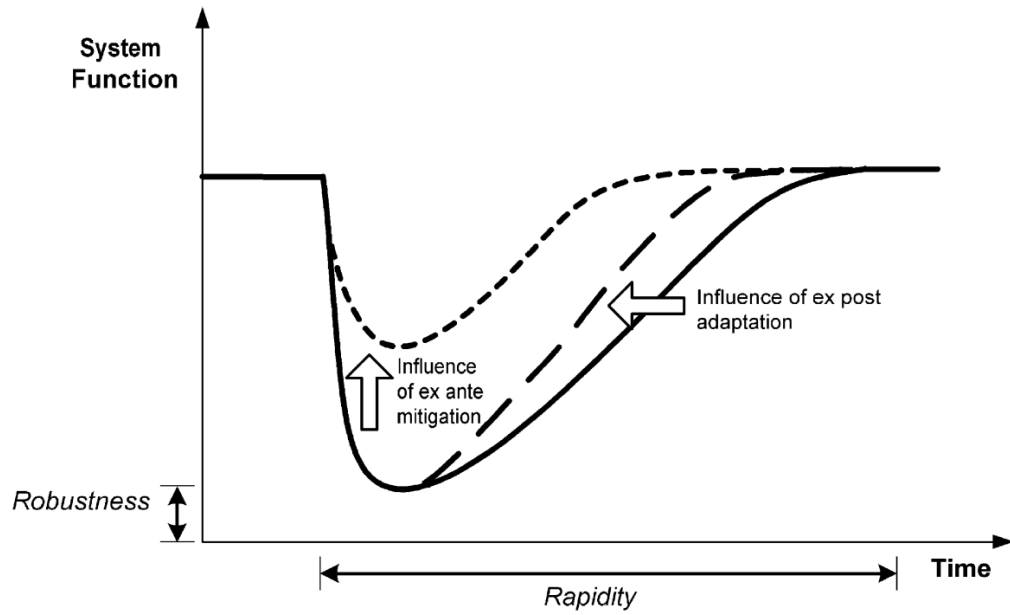


Figure 2.1: An illustration of the properties of resilience (McDaniels *et al.*, 2008).

A number of authors do include the ability to recover within their definitions with regard to engineering and infrastructure analysis. Reed *et al.* (2009) define resilience as ‘the ability to bounce back after a major disruption’, Ouyang *et al.* (2012) it as ‘the ability to resist (prevent and withstand) possible hazards, absorb the initial damage, and recover to normal operation’ and O’Rourke (2007) states that resilience by ‘the loss in quality (of service) over the time to recover’. These three definitions all refer to resilience as some function of the time to recover, as well as the ability to continue to function while perturbed; a more complete assessment of the performance of a network when perturbed. Bruneau *et al.* (2003) also provide a diagram of resilience (Figure 2.2), highlighting that resilience is the ability to function when not at 100% due to an event. This explicit handling of the time components of resilience re-enforces the concept of the time to recover being a critical part of a systems resilience.

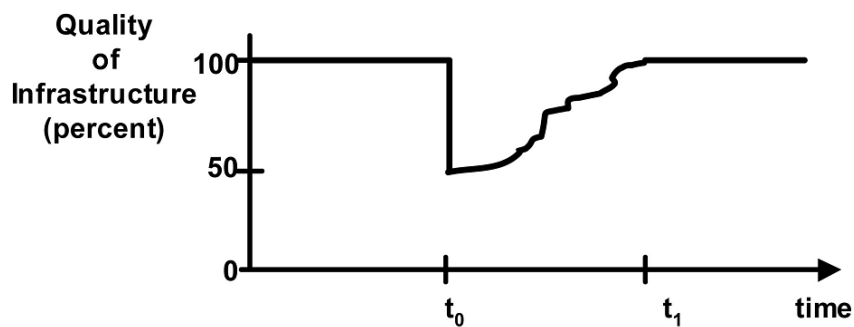


Figure 2.2: Diagrammatic definition of resilience (Bruneau *et al.*, 2003).

The need for critical infrastructures to be able to perform while perturbed has also involved many governments, including the UK and the USA. Resilience was defined by the Pitt Review in the UK as ‘the ability of a system or organisation to withstand and recover from adversity’ (Cabinet Office, 2008) and the Homeland Security Advisory Council (2011) for the USA defines it as the ‘ability to resist, absorb, recover from or successfully adapt to adversity due to a change in conditions’. This highlights the recognition of the resilience of networks being critical from a governance perspective, with the ability to recover from perturbations forming a key part of this.

The definitions of resilience share similarities to other concepts, including robustness (Hosseini *et al.*, 2016), itself a significant characteristics for infrastructure systems (Callaway *et al.*, 2000; Reed *et al.*, 2009; Gao *et al.*, 2011). Bruneau and Reinhorn (2007) define it as the ‘strength, or the ability of elements, systems, and other measures of analysis to withstand a given level of stress or demand without suffering degradation or loss of function’ and McDaniels *et al.* (2008) define it as ‘the extent of system function that is maintained’ (Figure 2.1). These three definitions are similar with all referring to robustness being the ability of a network to perform while perturbed in some manner, though do share some similarities to the definition of resilience, such as those by Leu *et al.* (2010) and Sterbenz *et al.* (2011), with regard to the ability of network to still function when perturbed. These definitions however are distinct from those for resilience, with resilience including the ability of a system to absorb and recover from a hazard/perturbation and the consequences of this on the system (Reed *et al.*, 2009; Ouyang *et al.*, 2012; Hosseini *et al.*, 2016). This clearly differentiates the two concepts of resilience and robustness, of which it is evident in some cases from the literature that the terms are used interchangeably and can cause some confusion. However, for the rest of this work, the focus will be on network robustness, and the ability of a network to withstand perturbations.

Although much of the literature presented refers to the robustness and/or resilience of infrastructure networks as a single entity (Walker *et al.*, 2004; Cabinet Office, 2008; McDaniels *et al.*, 2008; Leu *et al.*, 2010), there are many individual components which form an infrastructure system, with each having its own level of robustness/resilience. Infrastructure systems commonly exist of physical assets/entities, as well as a control system, such as a SCADA system (Zio, 2014). For an infrastructure system to be resilient or robust all aspects must play a part in this, from the individual assets which form the network, to the operational and control aspects, and by ignoring one in a systems analysis results in un-realistic results. When defining robustness Bruneau and Reinhorn (2007), as discussed above, include ‘elements’ as well as ‘systems’ in their definition of robustness, suggesting an acknowledgement that the

robustness of the elements are just as important as that of the system. This work instead of focusing on the robustness of the system focuses on the robustness of the network, the way in which the physical assets are connected, ignoring that of the individual assets as well as the control systems. This allows for the work to focus on the structure as well connectivity of the network, and how this influences the ability of an infrastructure to function when perturbed.

As defined in the previous paragraphs, resilience and robustness have contrasting definitions, and thus effect the design and running of infrastructure networks, with one about the system being able to resist failures and the other the system being able to recover from a failure quickly. A robust network/entity is not necessarily resilient, with the entity being designed to reduce the likelihood of failure and thus there is little need to consider how the entity/network may recover (be resilient). In contrast a resilient network/entity is designed to be able to recover from a failure, with the robustness, the ability not to fail, a lesser concern as some failures are expected. A robust system therefore is not designed with the features which may make a system resilient, such as redundancy (Jenelius, 2010; Royal Academy of Engineering, 2011; Yazdani and Jeffrey, 2012), as these are not expected to be required. A resilient system focuses on having high levels of redundancy, with the expectation that some components will fail and the redundancy will therefore be required if the network as a whole is to not fail. Despite this apparent tension between these two factors, there has been little research into this relationship, and how this affects the design and control of infrastructure networks.

2.3 Models of infrastructure networks

The function, characteristics and behaviour of infrastructure networks has been explored through a range of approaches; from modelling using complex networks (Bagler, 2008a; Wang and Rong, 2009; Wilkinson *et al.*, 2012), agent based models (van Dam and Lukszo, 2006; Oliva *et al.*, 2010) and petri-nets (Gursesli and Desrochers, 2003; Pye and Warren, 2006). This range of approaches has afforded the potential to gain a greater understanding of infrastructure networks, with each method offering the ability to model the networks/systems from different perspectives.

Agent based modelling simulates individual agents which behave based on a set of predefined rules to analyse how systems and models change over time (Bonabeau, 2002; Macal and North, 2010; Helbing, 2012). Such modelling has been used across a range of fields, from economic modelling and population modelling (Helbing, 2012), to the modelling of interdependencies between infrastructure systems (Valeria *et al.*, 2007) and the modelling of flows over networks

(Hoogendoorn and Bovy, 2001). Models are developed with multiple agents used to simulate the network and the processes on it, with domain specific knowledge required to parameterise the agents using rules, conditional or mathematical, in as realistic way as possible (Bonabeau, 2002). For example, the TRANSMIS agent based model has been developed to simulate road networks at a fine scale, with the model using individual agents moving over a regional road network (Nagel *et al.*, 1997; Nagel and Rickert, 2001), and thus rules are developed to describe how they make decisions about moving over the network. This micro-simulation agent based model has been deployed to simulate traffic behaviour including queuing patterns, on road networks in the cities of Dallas and Portland (USA) (Helbing, 2012), aiding traffic management strategies and future planning. At a broader scale Valeria *et al.* (2007) have developed an agent based model to analyse the impact of perturbations of one infrastructure on those which are dependent on it. The agents within the model represent different sectors and the components within them and are developed with rules about how they communicate and interact with each other, allowing the simulation to help improve the understanding of how the system of infrastructure systems are affected by just a single failure. As well as highlighting why domain specific knowledge is required per agent based model, it also exemplifies the different scales at which agent based models can be used to learn and help understand infrastructure systems.

With the ability of agent based models to simulate systems at a vast range of scales, the number of agents used in the models can vary. As the number of agents does increase, the computational overheads associated with simulating the behaviour of each agent, and it's interactions with other agents as well as the environment/infrastructure, can make the application of agent based modelling less conducive in some situations (Helbing, 2012). However, as a method it can provide detailed information on modelled systems as the behaviour of agents (people or components) can be modelled explicitly providing a wealth of detailed information on the processes, flows and interactions.

Graph theory has been widely adopted as an approach to modelling and understanding networks from the social sciences (Girvan and Newman, 2002), biology (Ravasz *et al.*, 2002; Costa *et al.*, 2008) through to communication/technology networks (Cohen *et al.*, 2001; Doyle *et al.*, 2005). As a method of modelling graphs and networks it has been used widely in helping to understand the structure of real-world networks, aiding in improving the research communities knowledge on how networks form/evolve and their structural properties. These have helped to develop better models which offer a more realistic view of networks where the real network data is unavailable, allowing further research to be undertaken in areas such as a networks resilience to failures. However, modelling the behaviour of processes on networks and behaviours of

assets in the networks is not a native application of the complex networks field with it better suited to analysis based around the topological connectivity of the networks allowing insights into the structure and form of the networks (Newman, 2003b; Amaral and Ottino, 2004).

Petri-Nets are an alternative to the complex networks methods discussed above, based on a similar theory where the systems/networks are modelled using graph representation (Pye and Warren, 2006). However, Petri-Net methods focus on the processes, states and conditions of the nodes and edges within the graph representation, enabling an analysis of both the physical and operational structure of the infrastructure system (Pye and Warren, 2006). The graphs are constructed as directed, weighted and bi-partite graphs (Murata, 1989), which more easily facilitate the modelling of processes, but also makes them less conducive to fundamental network analysis possible through complex network methods discussed previously. This added complexity within these graph representations of systems also greatly increases the computational requirements, limiting the size of systems/networks which can realistically be modelled (Murata, 1989; Ng *et al.*, 2013). Petri-Nets are thus better suited to analysing systems to understand their operational states and the process on the networks, exemplified by Laprie and Kanoun (2007), who analysed how the interdependency between an energy system and an information system affected the state of each when either network was perturbed. The analysis used a series of defined states to describe the effect of failures on both networks, while the processes between the two networks were analysed to identify the effects of failures.

The merits and downfalls of the three methods looked at for the analysis of graphs/networks, agent based modelling, complex network theory and Petri-Nets, have been briefly discussed. Each has been developed for a specific analysis purpose, though cross-over does exist between the methods. Of the three methods complex network theory is the most applicable for large ensemble analysis for investigating the characteristics of graphs and networks from a topological and structural perspective. This is highlighted by the wide adoption of this in the analysis of graph models and networks (Dunn *et al.*, 2013), with it being used since the development of the first graph models (Erdos and Renyi, 1959). The remainder of this section provides more in depth review of the key areas for the modelling of complex networks with network theory with a focus on critical infrastructure networks.

2.3.1 Graph theory

Graph theory is the mathematical study of graph structures which consist of a set of nodes (points/vertices) and edges (lines) (Dolan and Aldous, 1993; Newman, 2004). Figure 2.3 shows

a graph, G , consisting of a set of seven nodes (N) and nine edges (E), $G = \{N, E\}$, where $N \neq 0$ and $E = \{e_1, e_2, \dots, e_n\}$. The edges connect nodes, forming links between the nodes, such as sections of roads between junctions for example. G can exist where $E = 0$, though $N > 0$, and where edges do exist, they must connect nodes. Where a node exists, but is not connected, such as node 7 in Figure 2.3(a), the node is termed *isolated* (Erdos and Renyi, 1959; Albert and Barabasi, 2002; Dueñas-Osorio, 2005).

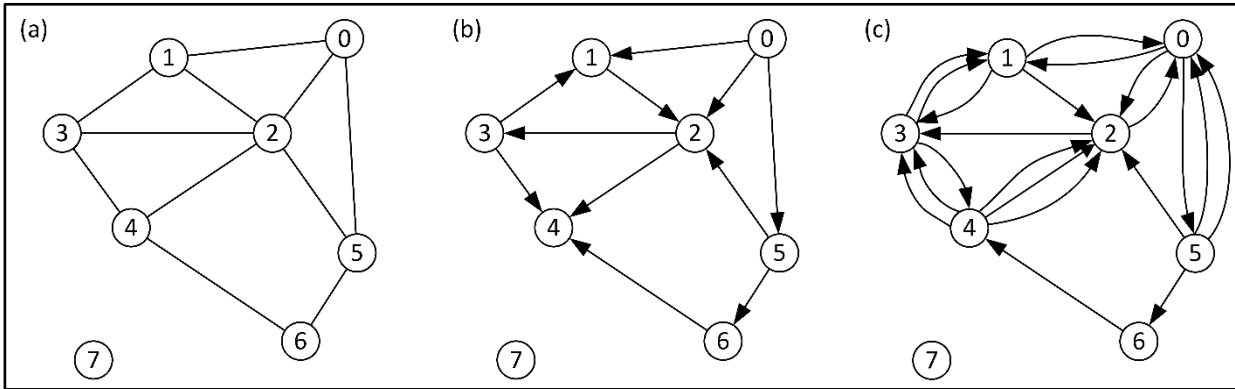


Figure 2.3: Two graphs, (a) a simple graph with an isolated node, (b) a digraph of the same form as (a) and (c) which shows a directed multigraph.

Each $n \in N$ and $e \in E$, can have characteristics which can define how a graph functions. e (edges) can have directions assigned to them, making G a *digraph* (Dolan and Aldous, 1993; Newman, 2003a; Boccaletti *et al.*, 2006), as in Figure 2.3(b) and (c). These can be used to model a network where flows in opposite directions between nodes have different attributes (e.g. time to traverse), or where multiple routes exist (Fortunato, 2010).

Each $n \in N$ has $\deg(n)$, defined as the number of edges incident on a node (Callaway *et al.*, 2000; Shi *et al.*, 2008). Table 2.1 shows the degree of each node for G in Figure 2.3(a), along with the in and out-degree of G in Figure 2.3 (b), where the in-degree is edges directed to a node, and the out-degree those directed away from a node (Newman, 2003b).

Node	Degree	In-degree	Out-degree
0	3	0	3
1	3	2	1
2	5	3	2
3	3	1	2
4	3	3	0
5	3	1	2
6	2	1	1
7	0	0	0

Table 2.1: Degree statistics for graph in Figure 2.3, including the degree of nodes in (a) and the in-degree and out-degree of nodes in (b).

2.3.2 Measuring network characteristics

The metrics and methods covered in the following are only those which are frequently used in the characterisation of graph/networks. A more comprehensive presentation of the field of graph theory for the analysis of complex networks is given in review papers such as those by Albert and Barabasi (2002), Newman (2003b) and Costa *et al.* (2007) and is deemed beyond the scope of this chapter.

The degree of node can suggest it's importance in the graph, with those with proportionally high degrees being hub nodes, vital to the functioning and structure of the graph (Barabasi *et al.*, 2003; Ravasz and Barabasi, 2003). This is shown in Figure 2.3(a) where node 2 has a degree 5 and the others with degree 3 or lower. Real-world examples of such hubs include major international air ports such as Heathrow (London) and Charles de Gaulle (Paris) which are viewed as long-haul hubs for Europe (Dennis, 2005; Grubestic *et al.*, 2009).

The degree distribution, the probability ($P(k)$) that a node selected randomly from the graph will have degree k (Albert and Barabasi, 2002; Newman, 2003b), it becomes possible to compare the topological structure of graphs (Albert and Barabasi, 2002). Through analysing the distributions of graphs, new models have been developed such as the development of the scale-free Barabasi-Model (Barabasi and Albert, 1999) where the degree distribution is scale-free, following a power law (Figure 2.4).

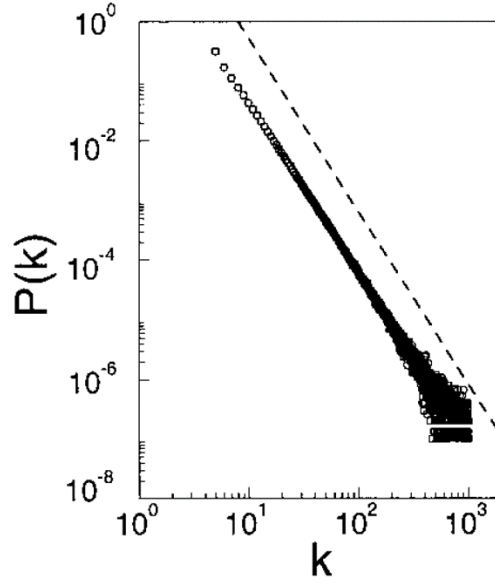


Figure 2.4: A scale-free degree distribution (Barabasi and Albert, 1999).

The degree distribution is one method of characterising a network. However, many other methods exist, including the average (geodesic) path length, the clustering coefficient and centrality metrics (Albert and Barabasi, 2002; Boccaletti *et al.*, 2006; Costa *et al.*, 2008). The shortest path, d_{ij} , or the geodesic path, is measured by the number edges between the two nodes, i and j . In Figure 2.5 the shortest path between nodes 1 and 6 is 2, passing over edges (1, 4) and (4, 6). The weighted shortest path between nodes 1 and 6 in Figure 2.5, where the weight of edges is denoted by L , is edges (1, 2), (2, 3), (3, 6), different to the geodesic shortest path.

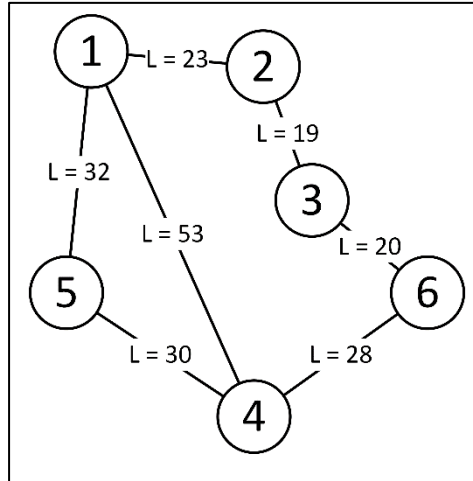


Figure 2.5: A graph where the geodesic shortest between node 1 and 6 is (1-4-6), measured in the term of edges. The shortest weighted, using edge lengths, L , is (1-2-3-6).

The average path length is the average of the sum of all the shortest paths between all node pairs (Albert and Barabasi, 2002; Dueñas-Osorio *et al.*, 2007a). This describes the ease with which the network can be traversed. It is defined as (Newman, 2003b):

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \quad (\text{Equation 2.1})$$

where l is the average path length of G (the network), d_{ij} is the shortest path between nodes i and j and n is the number of nodes in G . In some cases Equation 2.3 may also be referred to as the *characteristic path length* (Boccaletti *et al.*, 2006). The measure is used to characterise the structure of networks with a shorter l suggesting a better connected network with few long paths between pairs of nodes (Newman, 2003b). An alternative to the average path length metric is the diameter, defined by Newman (2003b) as the longest geodesic (shortest) path between any two pairs of nodes. Albert *et al.* (2000) uses the diameter of the network to record the behaviour of a network while its being perturbed while Gastner and Newman (2006) use the diameter metric to help characterise the structure of networks, finding the diameter of graphs varies per graph model, with low values expected for better connected graphs.

The betweenness centrality measures the number of shortest paths which pass through a node when shortest paths are computed between each pair of nodes in the graph. Girvan and Newman (2002) define it as the number of geodesic (shortest) paths between all vertex pairs which pass through a vertex (or edge). The value can then be normalised by dividing the betweenness value by the number of total node pairs in the graph, excluding the node of interest. The normalised betweenness centrality of node i in graph G is defined as (Crucitti *et al.*, 2006):

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j,k \in G, j \neq k \neq i} n_{jk}(i)/n_{jk} \quad (\text{Equation 2.2})$$

where N is the number of nodes in G , n_{jk} is the count of geodesic paths between node j and node k and $n_{jk}(i)$ is the count of geodesic (shortest, Figure 2.5) paths through node i between nodes j and k . This results in those nodes with the greatest values being those which have the greatest proportion of shortest paths passing through them, an indicator of their criticality in the network. The size of the maximum value also gives an indication as to how critical the node is,

with the greater the value, the greater the importance in the network (Girvan and Newman, 2002; Luca *et al.*, 2006).

Further to path based metrics the clustering coefficient has been used by many authors to characterise the topological structure of networks (Albert and Barabasi, 2002; Newman, 2003b; Bagler, 2008a). The clustering coefficient for node i in network G is defined as (Albert and Barabasi, 2002):

$$C_i = \frac{2E_i}{K_i(K_i - 1)} \quad (\text{Equation 2.3})$$

where E_i is the set of edges in G and K_i is the number of edges incident on node i . The value for node i refers to the likelihood of the nodes which it is connected to also being connected to each other (Newman, 2003b). As such it forms a local measure of network structure (Girvan and Newman, 2002; Newman, 2003b), but can be calculated for G (Newman, 2003b):

$$C = \frac{1}{N} \sum_i C_i \quad (\text{Equation 2.4})$$

where C is the clustering coefficient for G , C_i is the clustering coefficient for n_i and N is the number of nodes in G . This allows for the assessment of the connectivity of the entire graph allowing graphs to be compared to one another (Albert and Barabasi, 2002).

2.3.3 Graph models

Graphs have been used to replicate and model real-world infrastructure networks allowing an analysis of their topological structure and characteristics (Newman, 2003b). The first model was developed by Erdos and Renyi (1959), before further developments led to the small-world model developed by Watts and Strogatz (1998) and then the scale-free model proposed by Barabasi and Albert (1999). These models have been used extensively across a range of research fields for comparing the structure of networks in social sciences such as human friendship/sexual contact networks (Liljeros *et al.*, 2001; Newman *et al.*, 2002) and biology (Barabasi and Oltvai, 2004; Costa *et al.*, 2008) to the study of critical infrastructure networks

such as the internet (Cohen *et al.*, 2001) and water distribution networks (Shuang *et al.*, 2014). Review papers such as those by Albert and Barabasi (2002), Newman (2003b) and Boccaletti *et al.* (2006) provide multiple references to the use of and association of graph models to infrastructure networks from a spectrum of sectors.

The random model (Erdos and Renyi, 1959), known as the Erdos-Renyi model, is defined as N nodes connected with E edges where each $e \in E$ is added at random, with the number of $e \in E$ calculated using (Albert and Barabasi, 2002):

$$E = p \left[\frac{N(N-1)}{2} \right] \quad (\text{Equation. 2.5})$$

where p is the probability value and N is the number of nodes. Equation 2.1 presumes the graph is undirected, allowing the maximum number edges such that each node is connect to each other node through one edge only. $E = 0$ where $p = 0$ (Figure 2.6), with E increasing as p increases, with a complete graph (all possible edges) forming where $p = 1$.

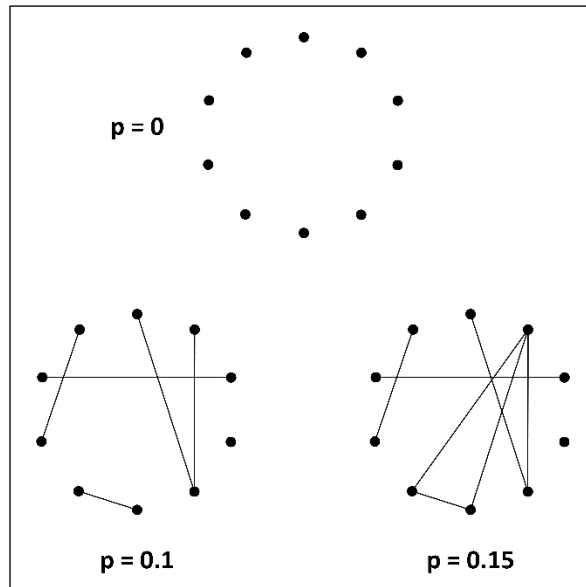


Figure 2.6: The generation of the Erdos-Renyi random graph, showing the difference the probability (p) value makes (Albert and Barabasi, 2002).

The typical degree distribution of the Erdos-Renyi random graph exhibits a Poisson distribution (Newman, 2003b; Birmelé, 2009) (Figure 2.7), and typically has a low clustering coefficient, C , a measure of the connectivity of a node and its neighbours. This model has been used for analysis of networks and a benchmark for many studies (Callaway *et al.*, 2000; Albert and Barabasi, 2002; Newman *et al.*, 2002; Palla *et al.*, 2005). However, it has been found to be inadequate for modelling all real-world networks by Albert and Barabasi (2002) who showed the C (clustering coefficient) of the Erdos-Renyi model was lower than that found for many real-world networks, such as the power grid for Western USA which instead returned a value more closely to those for regular grids (Watts and Strogatz, 1998).

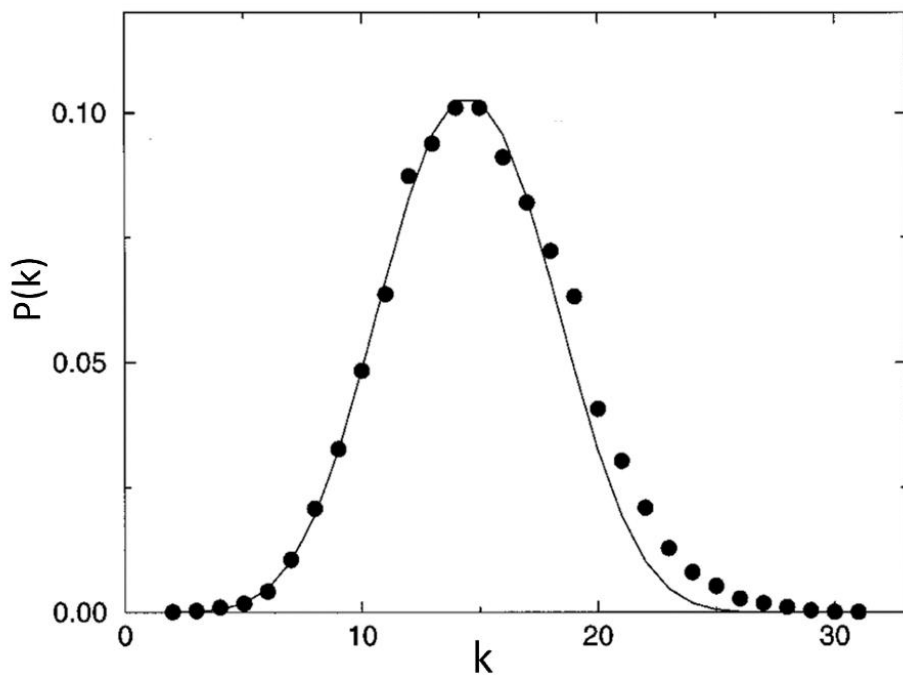


Figure 2.7: Degree distribution of a random network (Albert and Barabasi, 2002).

Watts and Strogatz (1998) were the first to successfully generate a network which has a higher C , than that for the Erdos-Renyi model (where N is equal for both models), typical of real-world networks (Barrat and Weigt, 2000) such as a power grid and a graph of film actors (Watts and Strogatz, 1998). This is known as the Watts-Strogatz model (WS) and is said to have a small-world topology (Watts and Strogatz, 1998) as $n \in N$ are well connected locally, but not globally, hence the high C while having a large average path length (Barrat and Weigt, 2000). The graph, G , is based on a regular lattice where each node, $n \in N$, is connected to a set number of its nearest neighbours, resulting in G having a high average path length and C (Watts and Strogatz, 1998). The small-world model is then created by rewiring a proportion of the $e \in E$

randomly to create the randomness of the graph, while in turn lowering the average path length by creating shortcuts across the graph (Watts and Strogatz, 1998; Newman, 2000; Newman, 2003b), as illustrated in Figure 2.8. Albert and Barabasi (2002) have found that just by rewiring a relatively low fraction of edges the average path length is reduced drastically, bringing it closer to that observed in some networks such as power grids and social friendship networks.

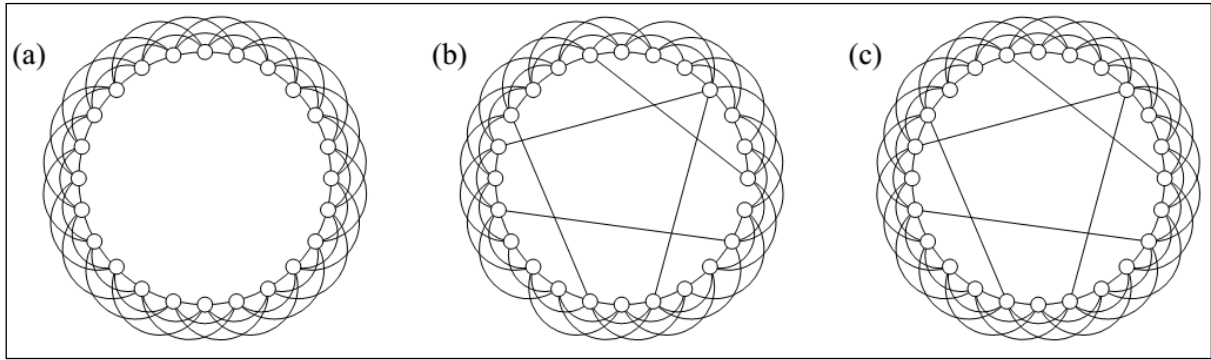


Figure 2.8: Rewiring a regular lattice (a) to create the small-world network (b) and (c) where edges have been added instead of existing edges being re-wired (Newman, 2003b).

More success in replicating the characteristics of real-world networks was found by modelling the evolution of networks using a growth based approach, modelling a preferential attachment methodology, where new nodes are most likely to be connected to already well connected nodes, as seen in some real-world networks (Barabasi *et al.*, 2003; Dueñas-Osorio *et al.*, 2004). This produces a network with a power-law degree distribution (Figure 2.4), which is scale-free, being independent of the number of nodes in the graph. This also results in the formation of hub nodes, nodes with proportionally higher degrees, a result of the preferential attachment criteria in the evolution of the network (Ahmed *et al.*, 2005). The generated graphs have a low average path length compared to the Watts-Strogatz (WS) model, and C , clustering coefficient values similar to the WS model which can be five times higher than values found in random graphs (Albert and Barabasi, 2002). These values are also much closer to those exhibited by some real-world networks, such as worldwide flights (Barrat *et al.*, 2004), the worldwide web (Albert *et al.*, 1999) and the web of sexual contacts (Liljeros *et al.*, 2001).

The first scale-free model to achieve this was developed by Barabasi and Albert (1999), where an initial ensemble of nodes is used then new nodes are added with a number of new edges using a preferential attachment rule:

$$p_i = \frac{K_i}{\sum_j K_j} \quad (\text{Equation. 2.6})$$

where p_i is the probability of the new node connecting to i , K_i is the degree of n_i , and K_j is the degree of n_j . Therefore, the probability, p_i , of a new node connecting to n_i is dependent of the degree of n_i , with the greater the degree, K_i , the greater the probability of the node connecting to i .

2.3.4 Hierarchical networks and models

Hierarchical organisation is a common feature of many complex systems; the organisation of large companies (Trusina *et al.*, 2004) to the modular and hierarchically organised metabolic networks (Costa *et al.*, 2008), hierarchically organised transport networks (Yerra and Levinson, 2005) and the hierarchical organisation of the internet (Pastor-Satorras *et al.*, 2004). These are all hierarchical through a number of ‘levels’ within the networks, which in the case of metabolic and social friendship networks translates as the graphs has a whole subdividing into smaller modules/communities, which themselves sub-divided into smaller modules/communities, each forming levels within the graph (Sales-Pardo *et al.*, 2007; Clauset *et al.*, 2008). Road networks on the other hand have a hierarchical structure based on the quantity of flow which each node and edge carries, with those carrying the most at the top of the hierarchy (Yerra and Levinson, 2005).

The hierarchical structure of networks can be represented through a dendrogram (Clauset *et al.*, 2008), Figure 2.9, which in this case shows the relationship between each of the nodes (coloured) along the bottom of diagram. This highlights the criticality of the two black nodes in Figure 2.9 which without, the network would be disconnected, with two/three distinct subgraphs/components forming. These two nodes are just as important as hub nodes, nodes which are highly connected (Ravasz and Barabasi, 2003), as these connect otherwise disparate parts of the network together. Such features are common in hierarchical networks such as in the internet (Pastor-Satorras *et al.*, 2004), air networks (Grubestic *et al.*, 2009), metabolic networks (Ravasz *et al.*, 2002) and social networks (Clauset *et al.*, 2008).

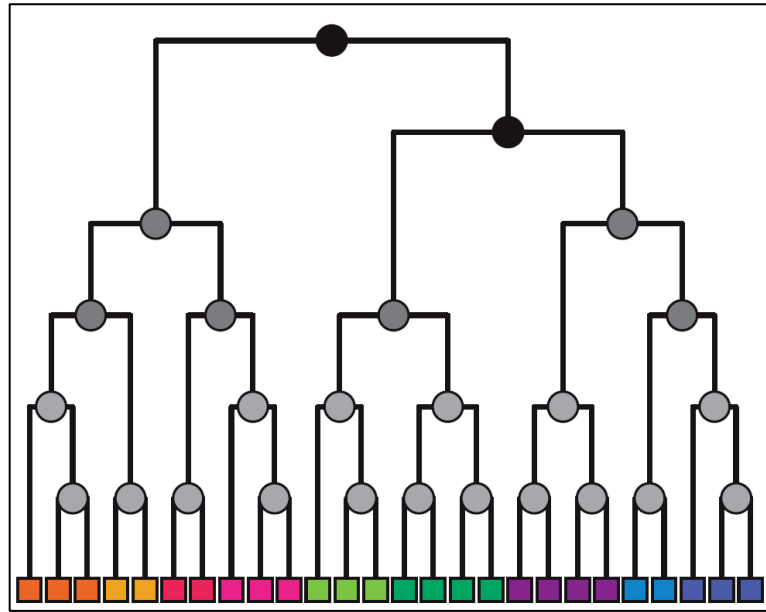


Figure 2.9: A hierarchical network shown as a dendrogram (Clauset *et al.*, 2008).

The clearest example of a hierarchical network is river networks, which form an explicit tree like structure, with water flowing from multiple points of source which all eventually join together by the river mouth (Katifori *et al.*, 2010; Barthelemy, 2011). The internet also has a hierarchical structure, realised through the volume of data/traffic using the links, where a number of links and nodes form the ‘backbone’ for the internet carrying a large fraction of all traffic (Pastor-Satorras *et al.*, 2004). These backbone links and nodes allow data to be transferred between weakly connected parts of the internet facilitating the functioning of the network. It has also been suggested that road networks possess a hierarchical structure based, like the internet, on the volume of traffic which flows over each node/edge (Yerra and Levinson, 2005). This creates the hierarchical structure of road classifications, with the most extreme being highways/motorways, with the length of roads at each level of the hierarchy increasing as they move away from the top level. Yerra and Levinson (2005) also showed that this structure was not designed, but instead emerged as the network has developed over time due to constraints such as geography and monetary considerations. Airline networks have also been shown to be hierarchical (Bagler, 2008b), with the topology showing the presence of hub nodes with traffic accumulated into interconnected communities of airports, connected through these hub airports. This results in a large number of airports with a small number of flights, with a few highly connected.

The hierarchical structure of some networks, such as rivers (Katifori *et al.*, 2010; Barthelemy, 2011), can be characterised through a tree graph (Figure 2.10(a)). This structure contains the

minimum number of possible edges to connect all the nodes in the graph, $E = N - 1$ (Barthelemy, 2011), making it the most efficient graph topology for N where each edge has a cost associated with it (Katifori *et al.*, 2010). However, this results in there being no loops (a closed path of edges (Newman, 2003b)), making the network potentially vulnerable to perturbations, failures of nodes and edges, with a higher density of loops inducing a greater resilience (Barthelemy, 2011). The network can however be augmented through the addition of edges, creating cycles and thus increasing redundancy (Helbing *et al.*, 2006b), Figure 2.10(b).

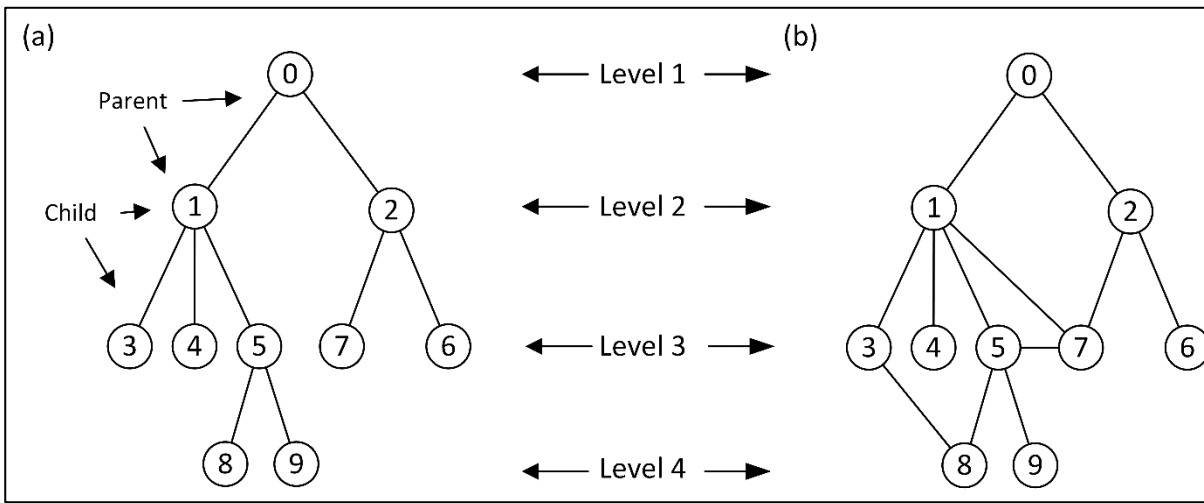


Figure 2.10: Two hierarchical networks; (a) a tree with no cycles, and (b) a tree with increased redundancy creating cycles.

An alternative hierarchical structure to the tree topology has been suggested based around a more modular structure (Ravasz *et al.*, 2002; Ravasz and Barabasi, 2003) (Figure 2.11). The graph uses a small community of nodes (Figure 2.11(a)), where a community is defined as a collection of nodes where the density of edges is greater between some nodes than between other nodes (Boccaletti *et al.*, 2006). This base community is then built upon as the level of the hierarchical graph is increased. Figure 2.11 (b) shows the model with a level of 1, which combines N (the number of nodes in the base model (5)), number of level 0 graphs (Figure 2.11(a)) to form the level 1 graph. This process is repeated for a level 2 graph, with N level 1 graphs used to form the graph. This results in a graph with a hierarchical structure, with all communities forming part of larger communities. Two versions have been used, with Ravasz *et al.* (2002) using a graph where 4 nodes are in the base community, and (Ravasz and Barabasi, 2003) using a graph where 5 nodes are in the base community as presented in Figure 2.11. With

each community being well connected, there is redundancy within these communities of nodes, though globally the redundancy in the network may be poor as four of the five communities in Figure 2.11(c) are only connected through the central community of nodes, a result of the underlying hierarchical nature of the model.

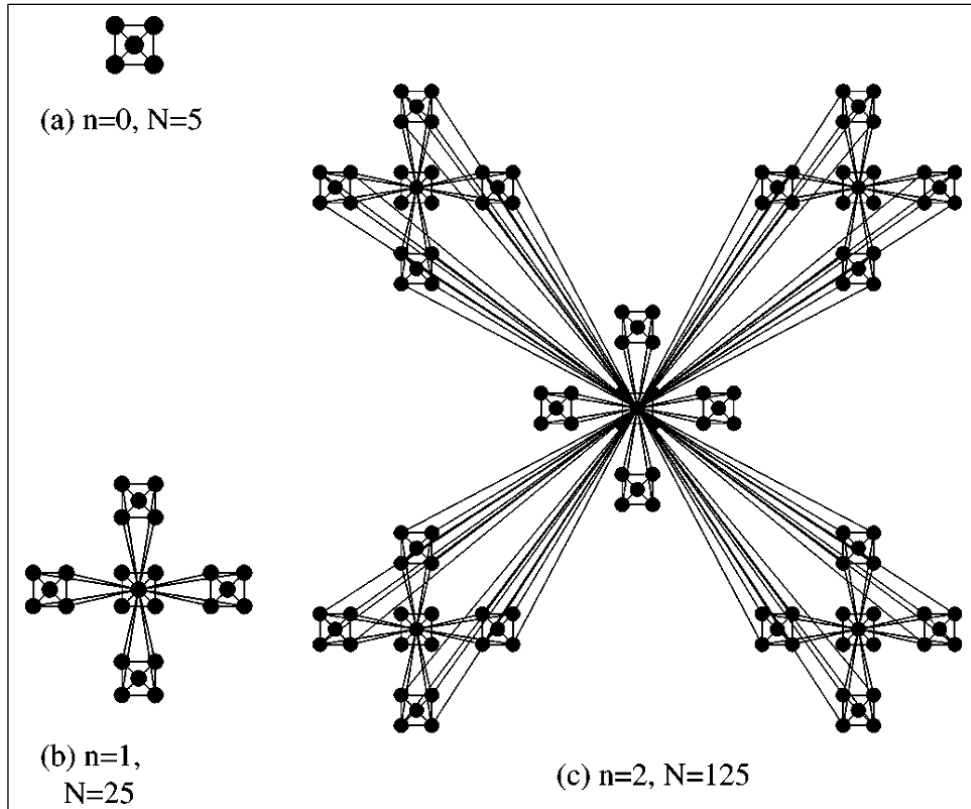


Figure 2.11: The hierarchical community model. (a) Shows the base level community of nodes and (b) shows how these are combined. (c) shows the third level where the community in (b) is used to generate a much larger network (Ravasz and Barabasi, 2003).

2.4 Studies of robustness and resilience

This section presents a review of the models employed for assessing the resilience and robustness of graphs and networks. The first section, Section 2.4.1, reviews the methods employed when assessing the robustness of graphs generated by graph models, with a number of papers presented where methods have been detailed. The second section, Section 2.4.2, reviews the methods used to assess critical infrastructure networks. The final section, Section 2.4.3, reviews the methods which have been presented for the explicit analysis of the robustness of spatial critical infrastructure networks to spatial hazards.

2.4.1 Robustness of graph models

Albert *et al.* (2000) has studied the topological response of the scale-free (Barabasi-Albert) graph model to two forms of topological failures, one where at each time step another random node is removed and the other simulates a targeted attack through the removal of the node with the greatest node degree at each step. Three variants of the graph are used each with different sized node sets; $N = 1,000$, $N = 5,000$ and $N = 20,000$. This found that the diameter of the graph, the longest shortest path between node pairs, increased more slowly through the removal of random nodes then compared to the targeted method. Further to this Albert *et al.* (2000) also found that the rate of change in the diameter of the networks during perturbations was independent of the number of nodes in the networks, indicating the exhibited behaviour was a result of the graph structure alone, and the way in which the available edges are used to connect nodes across the networks..

Holme *et al.* (2002) has employed similar methods to Albert *et al.* (2000) to compare the robustness of a Erdos-Renyi model generated random graph, a Watts-Strogatz model generated small-world graph and a scale-free Barabasi-Albert model generated graph. These were used as comparators for two real networks, a network of scientific collaborators and a computer network. Four node removal strategies were used; initial node degree, initial node betweenness, recalculated degree and recalculated betweenness, where in the recalculated methods the values are recalculated after the removal of each node. The recalculated betweenness method was the most effective at disrupting the graphs, with recalculated degree the second most effective. However, the initial degree method was more effective than initial betweenness indicating the distribution of node betweenness values changes more during the failure process. This is due to betweenness being calculated over the global network whereas the degree of each node is dependent on its neighbours and thus a more local metric.

Bassett and Bullmore (2006) have also used the random, scale-free and small-world graphs to compare the robustness of the latter, used to represent a brain network, to the other models. Again a topological failure model has been used where the graphs were perturbed using the random and maximum node degree removal strategies, with the same underlying methodology as used in the previous two studies reviewed. The effectiveness of the node removal strategies has been measured using the largest component size against the fraction of nodes removed. The results show, Figure 2.12, that the scale-free graph is much less robust than the random graph with the largest component size decreasing much faster for the targeted (node degree) removal strategy than for the random strategy. The small-world (brain) graph exhibits a greater robustness than the scale-free graph, but is clearly less robust than the random graph. This poor

resilience is explained through the presence of hub nodes in both the scale-free and small-world graphs, though due to the lower frequency of these, and the stronger local connectivity in the small-world graph, they display different behaviours.

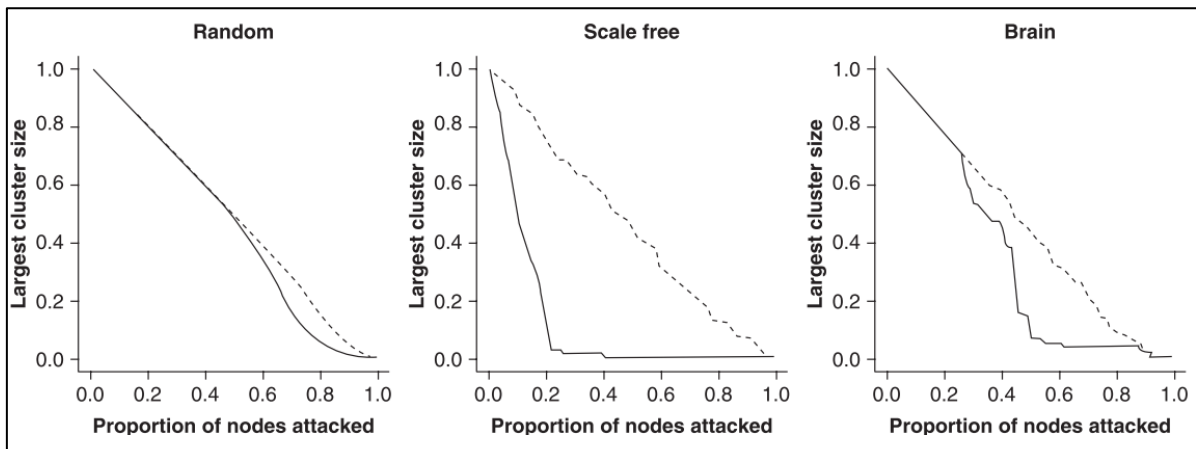


Figure 2.12: The results from the removal of random nodes (dashed) and nodes with the highest degrees (solid) for a random, scale-free and human Brain network (Bassett and Bullmore, 2006).

Shi *et al.* (2008) has examined the robustness of a Barabasi-Albert scale-free graph to failures using four scenarios, varying from those used in the previous studies reviewed. The first removes the node with the greatest degree at each epoch and the second removes the nodes based on degree but does not remove the node with the greatest degree. The final two methods remove nodes based on degree, but leaves 90% and 50% respectively of the nodes which are in the top 2-5% when the nodes are ranked by their degrees. Through this combination of targeted strategies, the robustness of the graphs when different proportions of the most connected nodes are removed is explored. The results show that removing nodes with the greatest degree at each step has the greatest impact reducing the size of the greatest cluster quicker than the other three strategies, with the graph failing after only 15% of nodes have been removed. The second strategy, missing the most connected node, and the 90%, strategy, leaving 90% of the top 2-5%, both result in a similar behaviour from the graph which results showing only a slightly greater robustness than the first method, the graph failing once closer to 20% of nodes have been removed. The plot for the 50% methodology exhibits a much better robustness, only failing once approximately 25% of nodes have been removed.

Cohen *et al.* (2001) has also analysed the robustness of the scale-free graph to perturbations by again using a random node removal strategy and a targeted strategy using node degree, where a node is removed at each epoch until the network fails. Three variations of the scale-free graph are generated using different power law values, 2.5, 2.8 and 3.3. The results highlight the

sensitivity of the scale-free networks to targeted attacks, with the presence of a small number of highly connected nodes which are critical to the connectivity of the networks making them vulnerable.

The reviewed methods have all employed the same underlying failure process whereby at each epoch/iteration another node is removed from the graph and the status of the graph assessed. However, the failure methods then vary as to how the nodes to remove are selected, though a number employ methods based around node degree and betweenness as well as random selection. A number of methods were also used for measuring the response of the graphs to perturbations. In most cases these centre around size of the giant component, whether measured using the diameter or the number of nodes.

2.4.2 Robustness of infrastructure networks

The resilience and robustness of critical infrastructure networks is important due to societies dependence on them (Little, 2003), and thus this has been a point researched in some studies. The analysis methodologies which have been employed in some of the studies is reviewed in the following paragraphs.

Some of the most extensive analysis has been undertaken on electricity networks, the backbone of modern societies (Dueñas-Osorio, 2005), with examples including work done by Crucitti *et al.* (2004b), Albert *et al.* (2004), Dueñas-Osorio and Vemuru (2009) and Wang and Rong (2011). Crucitti *et al.* (2004b) focused on analysing the Italian power grid (220 and 380Kv assets only), where $N = 341$ and $E = 517$, identifying its vulnerabilities to node failures through the use of model where a graph metric, betweenness (defined in Section 2.3.2), was used to simulate the flows over the network. The robustness of the network was examined by removing those nodes with the highest load and, showing that the network is vulnerable to the failures of these nodes but also vulnerable to the random removal of nodes. A similar method has been used by Albert *et al.* (2004) for the American power grid, where $N = 14,099$ and $E = 19,657$, with the robustness to failures examined by removing nodes with greatest loads, simulated using the betweenness values. As with Crucitti *et al.* (2004b) the results showed that it was more vulnerable to the higher load nodes being removed than random failures and a node degree based strategy, where nodes with the highest degree are removed. Although the use of betweenness as a proxy for the load on network assets is used in both studies, it is a “heavy simplification of what happens in a real electric power grid” (Crucitti *et al.*, 2004b), with this

measure presuming shortest paths from generators and no asset capacities are considered when computing the values (Albert *et al.*, 2004; Crucitti *et al.*, 2004b).

The robustness of transport networks to perturbations has also be analysed. Lordan *et al.* (2014) have analysed the global air traffic network (ATN) robustness to the failure of the most critical airports through five strategies; node degree, betweenness, modal analysis, damage and Bonacich power. The modal analysis ranks nodes based on their busyness (Petreska *et al.*, 2010), the damage method ranks nodes on the proportion of the giant component affected when its removed and the Bonacich method removes nodes based on a combination of centrality measures. Each method was used to perturb the network with results (Figure 2.13) showing that the betweenness method was the most effective at disrupting the network and causing the largest connected component, the giant component, to reduce in size the fastest. In contrast, Wuellner *et al.* (2010) analysed the networks of individual airline carries in the US through a four strategies, random edge failures, random node failures and targeted node failures through both node degree and betweenness. The betweenness method was shown to have the greatest impact on the largest connected component, with it staying connected until more than 50% of nodes removed when using node degree, and only 30% when using betweenness. Although the results varied per network, this highlighted that those airlines which route flights through a smaller number of hub airports are more vulnerable to failures than those which use more hubs for flights as they cannot route flights through other hubs following the failure of a small number of such airports.

Duan and Lu (2014) compared the robustness of road networks in six cities (Paris, London, San Francisco, Toronto, Singapore and Beijing) to random, degree and betweenness node failure methods. Through the successive removal of nodes, the robustness of the networks was analysed finding again that that degree and betweenness based methods had a greater impact on all networks than the random node failure method, with the average path length increasing quicker before falling as the network decreased in size following the removal of nodes.

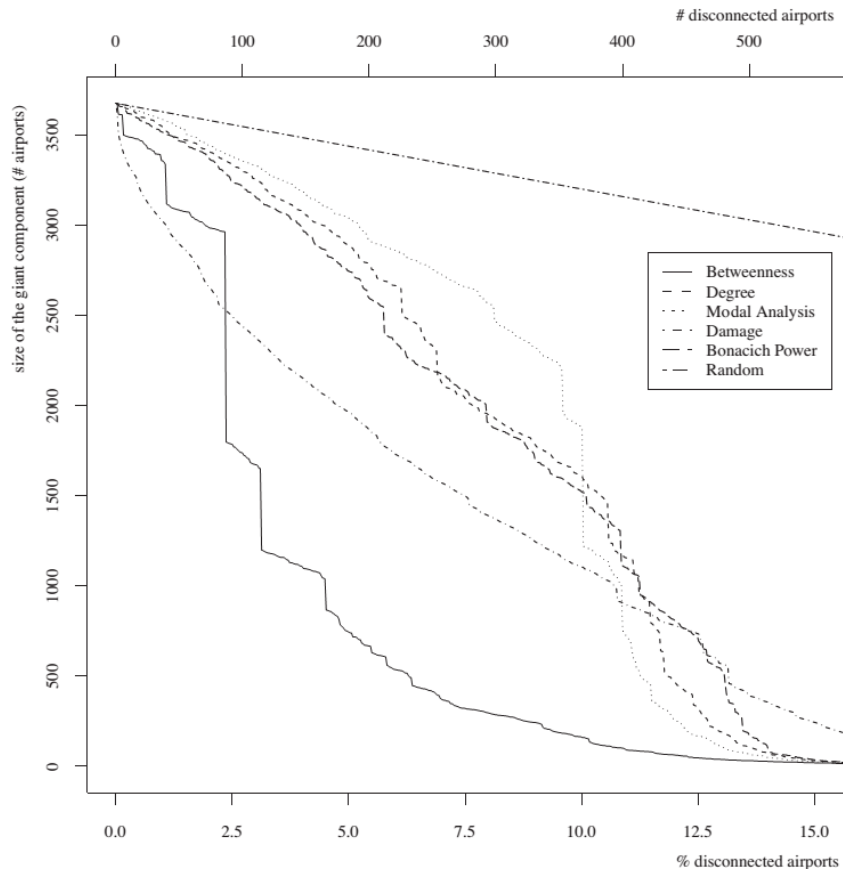


Figure 2.13: Failure analysis of the global airline network to different node removal strategies (Lordan *et al.*, 2014).

Communication infrastructure networks have also received attention due to other critical infrastructure networks and systems relying on them (Rinaldi *et al.*, 2001). Doyle *et al.* (2005) analysed the Abilene network, the high speed internet network between universities in the USA, analysing its robustness to the removal of the most connected routers, those with the highest degree. Through measuring the effect on the original amount of traffic which can still use the network, after being re-routed while considering bandwidth (capacity) constraints, it was shown that the performance degrades with the more nodes which are removed, though the effect is not as fast when nodes are selected at random. Albert *et al.* (2000) also analysed the structure of the internet ($N = 6,209$, $E = 12,200$) with regard to its robustness to the failure of nodes, with both random and node degree methods employed. As seen in the previous results, the random failure method has less effect on the network when compared to the node degree method, with the effect measured using the network diameter (Figure 2.14). The diameter of the network increased quickly under the targeted attack, with by the time 2% of the nodes had been removed the diameter having increased from approximately 4 to 12.5. In the random failure scenario, no noticeable increase was measured, highlighting the different impact these two failure

mechanisms can have and the significant impact targeted methods can have in infrastructure networks.

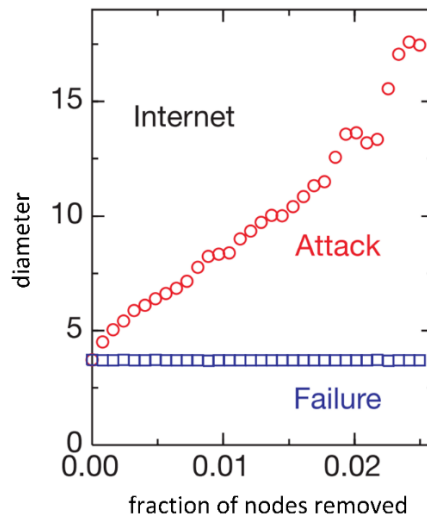


Figure 2.14: The diameter of the internet network ($N = 6,209$, $E = 12,200$) as the fraction of nodes removed increases under the random (failure) and highest degree node removal strategies (Albert *et al.*, 2000).

The robustness analysis examples presented thus far all perform the analysis where the connectivity of the network is explored, with few considering the actual dynamics of the flows over the networks and the effect of the changes in these on the response of the network to failures. Cascading failures model the behaviour of the flows over a network, with node and edge assets failing when they are over capacity, normally as a result of the redistribution of flows following the failure of other nodes or edges (Motter and Lai, 2002; Crucitti *et al.*, 2004a). Such failures occur in many infrastructure networks, but are common in both communication and transport networks (Crucitti *et al.*, 2004a) as well as electricity networks (Motter and Lai, 2002).

Wang and Rong (2011) has performed an analysis of the Western US power grid ($N = 4,941$, $E = 6,594$), using the betweenness metric as a proxy for load, and using a cascading failure model to examine the impact of removing edges from the network. The ‘avalanche’ of failures, where edges are over capacity, which occur following the redistribution of the load (recalculation of betweenness), are recorded to monitor the effect. They found that the cascading failures triggered by removing the edges with the highest load were more effective at disrupting the network, with a greater proportion of failed edges when the cascading failure stopped. Crucitti *et al.* (2004a) has used a developed cascading failure model to compare the robustness of the

electric grid in Western USA ($N = 6,491$, $E = 6,592$) to both targeted and random failure triggered cascading failures. Again the load is simulated using betweenness with insufficient data available to physically model the flow of electric and its behaviour when redistributed. A tolerance parameter was also used to model the likelihood of an edge asset failing when overcapacity, allowing for a tolerance in assets. However, the analysis shows that through targeting the node with highest load the effect is much greater than the random selection of a node to trigger cascading failures, though as very few nodes have high loads there was a low probability of the more catastrophic failure occurring. Bao *et al.* (2009b), although using a simulated electric network ($N = 300$, $E = 600$), uses a physical model, where the flow of electric is actually modelled, to examine the robustness to cascading failures of the electric system on the generated network. A tolerance parameter is used for the likelihood of edges failing when above or at capacity, and varied per simulation. Four simulations are run with a different set of trigger (failed) assets initially, with the resulting cascading failures recorded. The results show that a derived metric, power flow entropy, based around the ratio of load to capacity on edges, can be used to help prevent large scale blackouts through the value increasing as a network becomes more stressed.

Dueñas-Orsorio and Vemuru (2009) has explored the effect of cascading failures on a well-used example electric network. Again betweenness is used as a proxy for load, with capacities assigned to assets as a proportion of the initial load, with this extra capacity viewed as the tolerance to redistribution of flows. Trigger strategies include the nodes with the greatest load and the random selection of nodes, with both used to initiate cascading failures with each removing 1% of the elements as a trigger. Alternative trigger strategies are also used including those assets which through history have had the greatest exposure to lightning and those which lie in areas where seismic activity (earthquakes) is most likely. The results show that the strategy using the nodes with the highest loads has the greatest impact on the network, with the loss of connectivity greater than any of the other scenarios when no allowance for extra capacity was allowed. When the capacity was doubled to that of the initial load the results were still the fourth worst in terms of the loss of connectivity. The lightning and earthquake scenarios have a lesser impact on the connectivity of the network, but still have an effect greater than the random selection method.

The internet has also been analysed using cascading failure models, with Crucitti *et al.* (2004a) analysing a network of the internet ($N = 6,474$, $E = 12,567$), using the same cascading flow model with betweenness as a proxy for flows as used for the analysis of the electricity network reviewed earlier. As with the earlier results, the network was more vulnerable to the targeted

removal of nodes than compared to the random removal, with the internet less vulnerable than the electricity network. Motter and Lai (2002) has undertaken a similar analysis using the same two networks as used by Crucitti *et al.* (2004a), analysing the robustness of the internet and the electricity network for Western USA. Three trigger scenarios were used, with the random node selection scenario using 50 nodes, while a load (betweenness) method and a node degree method used 5 nodes. In the internet network the results showed that the random method had less of an effect than the degree and load methods which returned similar results. However, for the electric network the load method had much greater effect than the random and degree methods which both resulted in a similar effect on the network.

Shuang *et al.* (2014) has developed a physically based model for the analysis of water distribution networks, though has used a simulated network to exemplify the method. The model evaluates the vulnerability of all nodes in the network, with nodes failing if the pressure (of water) at the node is greater than its capacity after the redistribution of loads following an initial failure. Both betweenness and the calculated load were used as methods to trigger cascading failures, with again a tolerance value used to parameterise the point of which network assets fail given the calculated load on them. Both methods identified critical nodes, though these differed, with the most notable difference for the supply node, where through the calculated load it was identified as the most critical, as without it there is no supply to the rest of the network. The betweenness method on the other hand returned the same node as one of the least vulnerable, as the node itself had very few shortest paths passing through it thus returning a low betweenness value.

A range of failure models have been developed for the analysis of critical infrastructure networks, with both topological approaches and cascading approaches employed. The topological approaches have been found to be similar to those used to analyse the robustness of the synthetic graphs. However, attempts have been made to simulate flows over infrastructure networks to examine their robustness to cascading failures.

2.4.3 Spatial infrastructure robustness

The examples presented and discussed thus far all focus on the failures of assets either based on their topological position within the network, or due to the importance in the network as identified through metrics such as betweenness or as in the case of some physical models the actual load on them. However, critical infrastructure networks are embedded in space (Louf *et al.*, 2013; Danziger *et al.*, 2014), which not only affects their layout and development over time

(Huang *et al.*, 2006; Barthelemy, 2011; Louf *et al.*, 2013), but also means they are exposed to events such as natural hazards and weather events, all of which are spatial. Despite this, there has been comparatively little work dedicated to looking at networks in terms of their spatial characteristics including the robustness to spatial failures (Barthelemy, 2011). As Gastner and Newman (2006) note “most previous studies of real-world networks have ignored geography”.

Within many real-world networks long range links tend to connect nodes which are well connected, with shorter links connecting those which are less well connected within their local neighbourhood (Barthelemy, 2003; Barrat *et al.*, 2005); a feature which can only be identified when the geography of the network is considered. Pastor-Satorras *et al.* (2001) and Barrat *et al.* (2005) both suggest that the topology of a network can be affected by the geographic boundary constraints in which a network develops/grows and individual node assets which may limit the capacity of a node or how the network evolves to deliver a service. Further to this, Guimera and Amaral (2004) suggest that as well as geographic constraints such as boundary conditions, and the cost/length of edge features, geo-political constraints effect the development of networks across borders such as air networks and power networks. This was shown through attempting to replicate air networks, and in particular the worldwide air network, with the best representations only arising though the inclusion of geo-political constraints (on top of geographic consideration such as flight distances).

Through considering the geography of a network the efficiency of a network can be assessed by comparing the total length of edges to the straight line distance (Gastner and Newman, 2004; Barrat *et al.*, 2005). Gastner and Newman (2004) have performed this analysis of the subway in Boston (USA) (Figure 2.15(a)), and gas pipelines in Australia using a route factor value, the ratio between the distance along the network edges from each node to the root node and the Euclidean distance from all nodes to the root node (the equivalent of a star graph Figure 2.15(b)). A value of two suggests that the shortest path through the network is twice as long as the Euclidean distance, but values of 1.13 and 1.59 were returned for the two analysed networks. It is clear that both networks had developed a good compromise in structure where instead of an edge going from each node straight to the root node a network has developed which reduces the number of edges required while still being efficient with regard to the length of routes to the root node. This suggest that there is a tendency for networks to develop in an efficient manner with geography being a clear factor in their development otherwise route factor values would be expected to be greater with the length of routes not considered. Figure 2.15(c) shows the minimum spanning tree over the nodes for the Boston rail network, which connects them

all with the minimum number of edges, with a degree of similarity between this and the real network (Figure 2.15(a)).

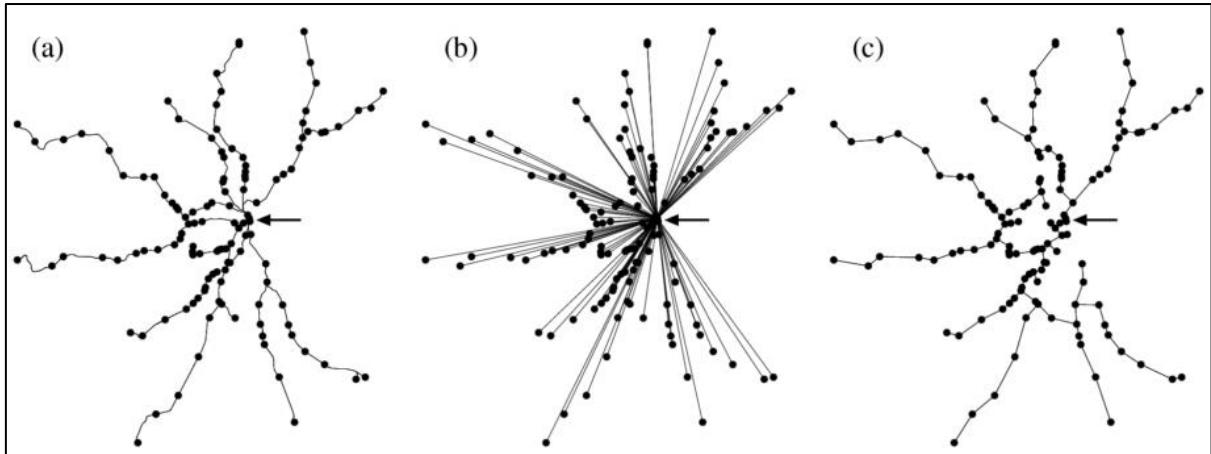


Figure 2.15: (a) the commuter rail in Boston (USA) with the root node indicated with an arrow. (b) the star graph for the same node set (each node connected directly to the root node) and (c) the minimum spanning tree (the network with the minimum set of edges).

Li *et al.* (2016) have presented a methodology for modelling spatial hazards, termed ‘regional failures’, where failures are applied to a network which cover a geographic area. Nodes and edges which lie, either completely or partly, within the geographic area are regarded as failed. The method is used to compare the structure of a series of graphs generated through an employed spatial graph model as developed by Louf *et al.* (2013), the LJB model, where a set of nodes are distributed uniformly across an area with each assigned a weight according to a power law. Using these weights, and the distance between nodes, nodes are connected where the importance of the connection, based on the weights, is considered along with the cost, the distance between the nodes. This generates graphs which can vary in structure based on the geographic distances between nodes. The findings show that the developed graph models generated have a poor robustness to the regional failures, the spatial hazard areas, though as the consideration for the cost of links is reduced, the robustness of the generated graphs improve.

The robustness of the European air network to spatial hazards has been investigated by Wilkinson *et al.* (2012) through using an ensemble of generated spatial models of the network to simulate a range of spatial hazards. A range of scenarios were employed to perturb the network including those to simulate the spread of a hazard over the network from one side (simulating the Eyjafjallajökull volcanic eruption), and random but spatially coherent hazard areas, with both covering the same percentage of airspace allowing for the effects to be compared directly. The results of the simulation, with routes deemed broken if they intersect a hazard area, show that the networks are robust until 10-15% of the network area is affected by

the hazard(s), after which their robustness decreases. This suggests a tipping point within the ability of the network to withstand hazards of different scales, with the network able to withstand those where the affected area is less than 15% of the network space. The worst disruption across all scenarios occurred when random hazards were located near the spatial centre of the network, with catastrophic consequences in the network, as major flight lines passed through this area affecting the connectivity of nodes around the entire network. Such analysis highlights that the robustness of a spatial network needs to consider the geographic distribution of assets, with topological robustness not alone enough to ensure the network is robust to spatial hazards.

The vulnerability of the road network in the City of York to flooding has been assessed by Balijepalli and Oppong (2014) who have considered the impact of flooding on the network. Three scenarios were employed including a 20%, 50% and 100% reduction in capacity for road links prone to flooding, with traffic flows which are simulated using origin and destination pairs redistributed in each scenario. The effects are measured using a developed vulnerability index which considers the change in travel time rather than distance when flows are rerouted. The research found that the effect of a partial reduction in capacity had different effects to a 100% reduction, with edges which had a limited effect when capacity was partially reduced having the greatest effect in the 100% reduction scenario. This is thought to be caused by the redistribution of flows being based on travel time rather than just distance, accounting for congestion on links as well as the availability of alternative routes.

Sterbenz *et al.* (2010) uses a regional hazard to exemplify how a major power failure or storm could be modelled, applying this to the GEANT2 network, the high-bandwidth academic internet for Europe (Figure 2.16). Three scenarios of increasing size from the spatial centre of the network are used showing that as the size of the hazard area grows, the loss of connectivity in the network increases. The largest size results in the network breaking into two components, leading to the suggestion redundancy in the network could be improved, and thus its robustness to such a hazard could be improved through the addition of a link between the UK and Iceland, though there are other options for this as well.

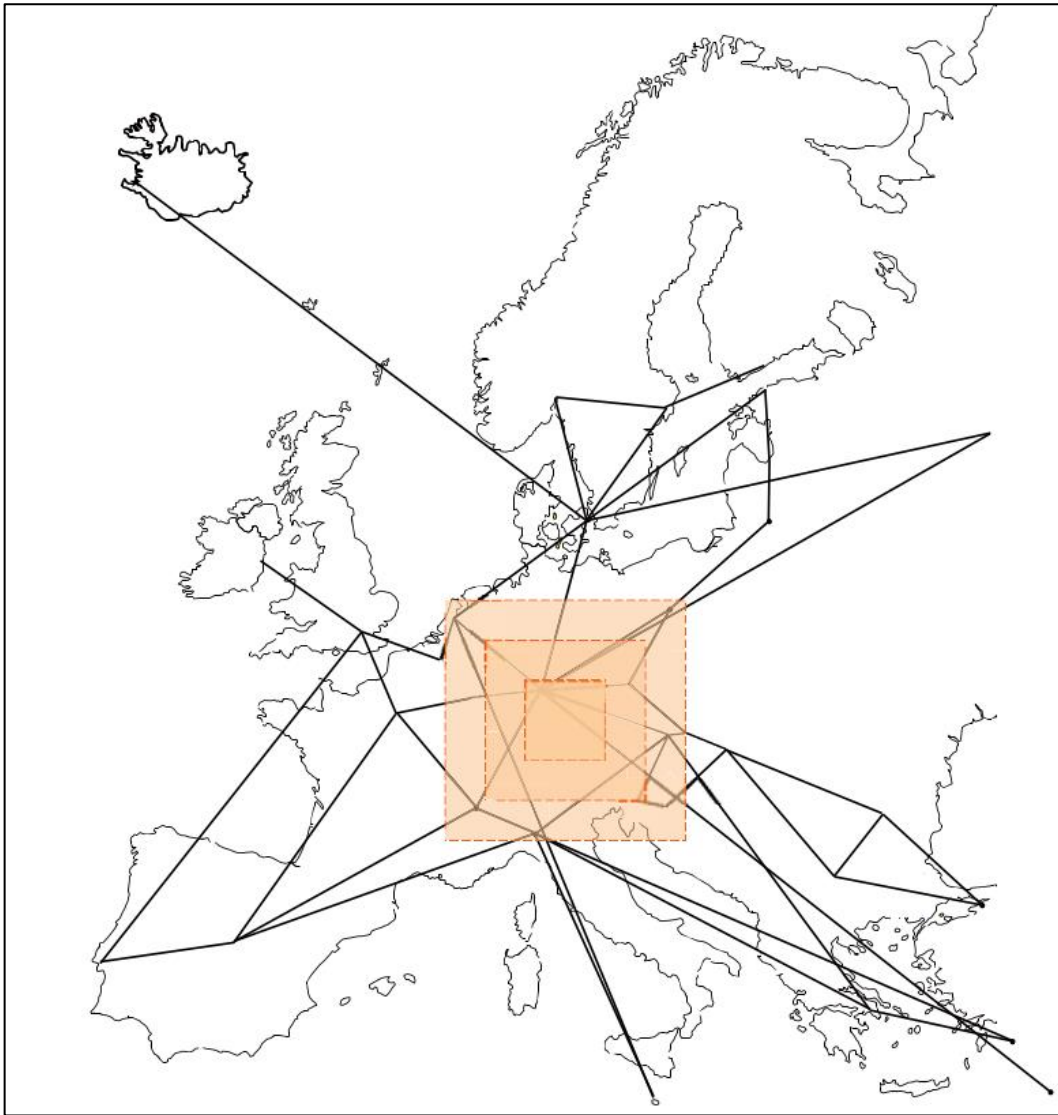


Figure 2.16: GEANT2 high-bandwidth European academic internet network with a spatial hazard at its geographic centre, with three sizes shown (Sterbenz *et al.*, 2010).

Ouyang (2016) has used spatial hazard areas to identify the critical points in infrastructure networks, with the electric and gas networks in Harris County, Texas used as an example (Figure 2.17). Nodes and edges are presumed to fail if they lie within the hazard area, with the electric network dependent on the gas network in the area as well. The analysis has shown that as the hazard area increases in size ($2.5\text{Km} < \text{radius} < 10\text{Km}$), the location with greatest effect on the functionality of the networks moves closer to the centre of the networks, away from the edge where maximum damage was caused when using the smallest hazard size (Figure 2.17).

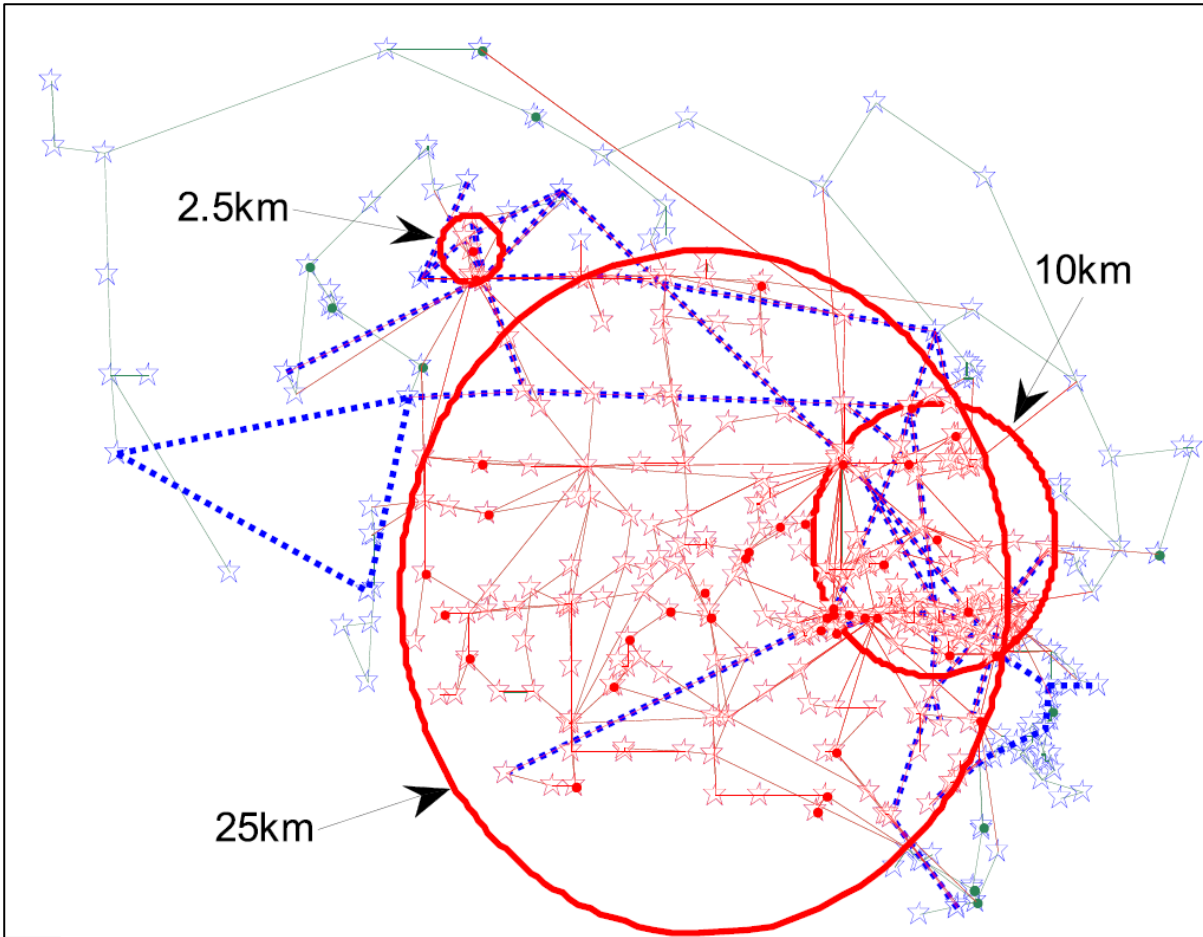


Figure 2.17: Location of the spatial hazards with the greatest impact on the interdependent electricity (solid lines) and gas (dashed lines) networks in Harris County, Texas, with failed assets in red (Ouyang, 2016).

2.5 Discussion and research challenges

The following paragraphs in this section will highlight and outline the research challenges and the recommendations for the research to be undertaken. The previous Sections have presented a review of the key literature in the field of critical infrastructure network analysis including the ability to withstand perturbations and the underlying graph theory methods and models. Section 2.1 has introduced the field of critical infrastructure networks and the extent to which these are relied upon. It has defined the difference between infrastructure systems and infrastructure networks, the physical assets on the ground, and how these can be treated as decoupled parts of the infrastructure system allowing for the physical assets to be treated as a stand-alone network.

Section 2.2 has reviewed the way in which the response of infrastructure networks to perturbations is measured, comparing the use of terminology such as resilience and robustness.

Despite the apparent conflict in the usage of these terms, the definition of each has been reviewed. Therefore, throughout this thesis the term robustness will be used as defined by McDaniels *et al.* (2008), ‘the extent of system function that is maintained’.

A review of the field of models used for the modelling of infrastructure networks and those associated to different infrastructure networks has been presented in Section 2.3.3. This has detailed the existing suite of models which have been associated with infrastructure networks through the characteristics which they exhibit. These have been detailed, with Section 2.3.2 introducing the basic concepts of graph theory required to understand the development and the varying characteristics of the models. Although Section 2.3.3 has presented graph models which have previously been associated with infrastructure networks, Section 2.3.4 reviews the emerging field of hierarchical networks and the properties which these exhibit to suggest hierarchies exist in critical infrastructures, a characteristic not common amongst the original models reviewed. Throughout Section 2.3 a series of graph models have been presented, all of which have been used in the analysis of networks. These will thus be used to form the foundations of a suite of graphs for identifying the characteristics of hierarchical graph models.

A review of the existing literature surrounding the analysis of robustness/resilience of infrastructure networks to hazards has been undertaken to assess the methods which have been employed across multiple infrastructure sectors, Section 2.4. It is clear that a set of common methods have been applied across those studies which have undertaken topological based failure analysis, with these finding the methods sufficient. Based on this the success and breadth of the use of the detailed methods, these should be adopted for the analysis of both infrastructure networks and graphs generated through the earlier reviewed models. A suite of literature has also been reviewed where dynamic failure models, or cascading failure models, have been used to analyse infrastructure networks. These, like the topological methods, all detail a similar failure model whether using a proxy for flows such as betweenness or through modelling them explicitly using physically based models, suggesting a standard which should be adopted for such analysis.

The literature has suggested that the spatial characteristics of the infrastructure networks is critical, and that many of the hazards faced by infrastructure networks, such as natural hazards, are inherently spatial. Yet when reviewed (Section 2.4.3), only a limited number of studies could be found where the effect of spatial hazards on spatial infrastructure networks was analysed. Although a small number of studies were identified, the coverage of these is limited suggesting further work could be done in this area to explore further the effect of spatial hazards on spatial infrastructure networks.

2.6 Conclusion

This chapter has presented a review of the literature in the areas of critical infrastructure analysis and graph theory for the modelling of infrastructure networks. This has highlighted the dependence society has on critical infrastructure networks and the need to ensure these are robust to a range of perturbations. The ways in which complex infrastructure networks can be modelled has been reviewed with both models for infrastructure networks and modelling approaches identified. This includes the growing theory around hierarchical organisations in infrastructure networks and the development of graph models which attempt to capture this characteristic.

The review has defined a number of methods and concepts which can be applied to approach the aims and objectives in Chapter 1 Section 1.2. The following chapter will introduce the methods which will be employed to address the aims and objectives.

Chapter 3: Methodological framework

3.1 Introduction

The previous chapter, Chapter 2, has highlighted gaps and potential areas of study with regard to the analysis of the robustness of critical infrastructure systems. The criticality of infrastructure networks to society has been highlighted along with the importance of these being robust to perturbations from a range of hazards. Previous studies have predominately assessed the topological robustness of infrastructure networks, although few have considered explicitly the spatial aspects of hazards that infrastructure networks can be exposed to. The methods used to model critical infrastructure networks have also been introduced, with recent developments including the emergence of literature suggesting a hierarchical organisation is present in infrastructure networks. However, it has been highlighted that this has yet to be studied extensively with regard to spatial critical infrastructure networks. It is not clear from the existing literature the extent to which hierarchical organisation can be found in critical infrastructure networks, and the effect of this on the robustness of these networks to perturbations. In relation to these points this chapter presents the methods which will be employed to address the aims and objectives presented in Chapter 1 Section 1.2 (page 4).

3.2 Overall experimental design

The robustness of hierarchical critical spatial infrastructure networks will be investigated, exploring how hierarchical networks respond to a range of perturbation scenarios. To understand the hierarchical organisation of networks (objective 2) a suite of graph models will be used to characterise the differences between the hierarchical and non-hierarchical topology. The key properties of hierarchical graph models will be used to characterise a suite of real-world critical spatial infrastructure networks, with networks identified as being hierarchical or not through association to the metric values from the graph models (objective 3).

The robustness of each of the graph models is also explored through topological based failure modelling and capacity constrained cascading failure modelling. These will both provide insights into the robustness of the graph models, including those which are hierarchical and non-hierarchical. Finally, a series of flow based failure simulations will be used to explore the robustness of hierarchical spatial critical infrastructure networks, to both hierarchical failures and spatial hazards (objective 4).

In Section 3.3, a series of graph models (Chapter 2, Section 2.3), for both hierarchical and non-hierarchical graphs, are created and a detailed statistical analysis of the topological structure of these is undertaken to recognise the differences between hierarchical and non-hierarchical topological structures, Section 3.4. Additionally, the robustness of these graphs is explored through a topological failure model developed in Section 3.5. A comparison between the behaviour of each of the graph models, how they respond to perturbations, is used to define the key properties of hierarchical graphs.

Building upon the analysis of the suite of synthetic graphs, the properties and characteristics of a series critical spatial infrastructure networks is investigated, generated in Section 3.6. These are subjected to similar methods employed in Section 3.4 and 3.5 and is presented in Section 3.7. This analysis will support the recognition of hierarchical infrastructure networks and the characteristics of these.

With infrastructure networks supplying commodities/information, the flows over these is an important characteristic (Motter and Lai, 2002; Ash and Newth, 2007). In order to gain a greater understanding of the robustness of hierarchical networks to flows, a capacity constrained cascading failure model has been developed (Section 3.9), along with an adapted network representation model to allow the attributes of nodes to be considered as with the edges, Section 3.8. The failure model is applied to the synthetic graphs in order to understand whether hierarchical networks respond differently to non-hierarchical with regards to flow based cascading failures.

Section 3.10 presents the methods developed to extend the analysis above to investigate the robustness of the England and Wales electricity transmission and distribution network to different configurations of spatial hazards. This analysis is undertaken as it has been recognised that spatial hazards have the potential to cause wide spread disruption in infrastructure networks (Little, 2003). This is followed by details of the framework and software stack developed to facilitate the analysis and research undertaken (Section 3.11), with a final section to conclude the presented methods (Section 3.12).

3.3 Synthetic suite of graphs

A suite of synthetic graphs has been generated covering a spectrum of graph types/topological structures, (Figure 3.1). Figure 3.2 shows the eight graph types employed, that provide a breadth of topological structures. In Figure 3.2 the example graphs have $N = 15$ and while the degree distribution plots have been generated for $N = 781$. For each of the eight models 1,000

realisations unless stated otherwise, are generated to describe the topological structure that exists for a range of possible node degrees. For each graph type a different model has been used to generate the graphs, each with a unique set of parameters, as shown in Table 3.1. However, across these different graph types $0 < N \leq 2000$ and $0 < E \leq 20000$.

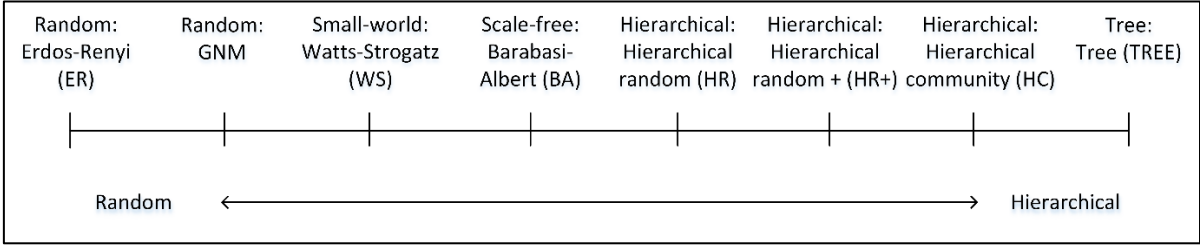
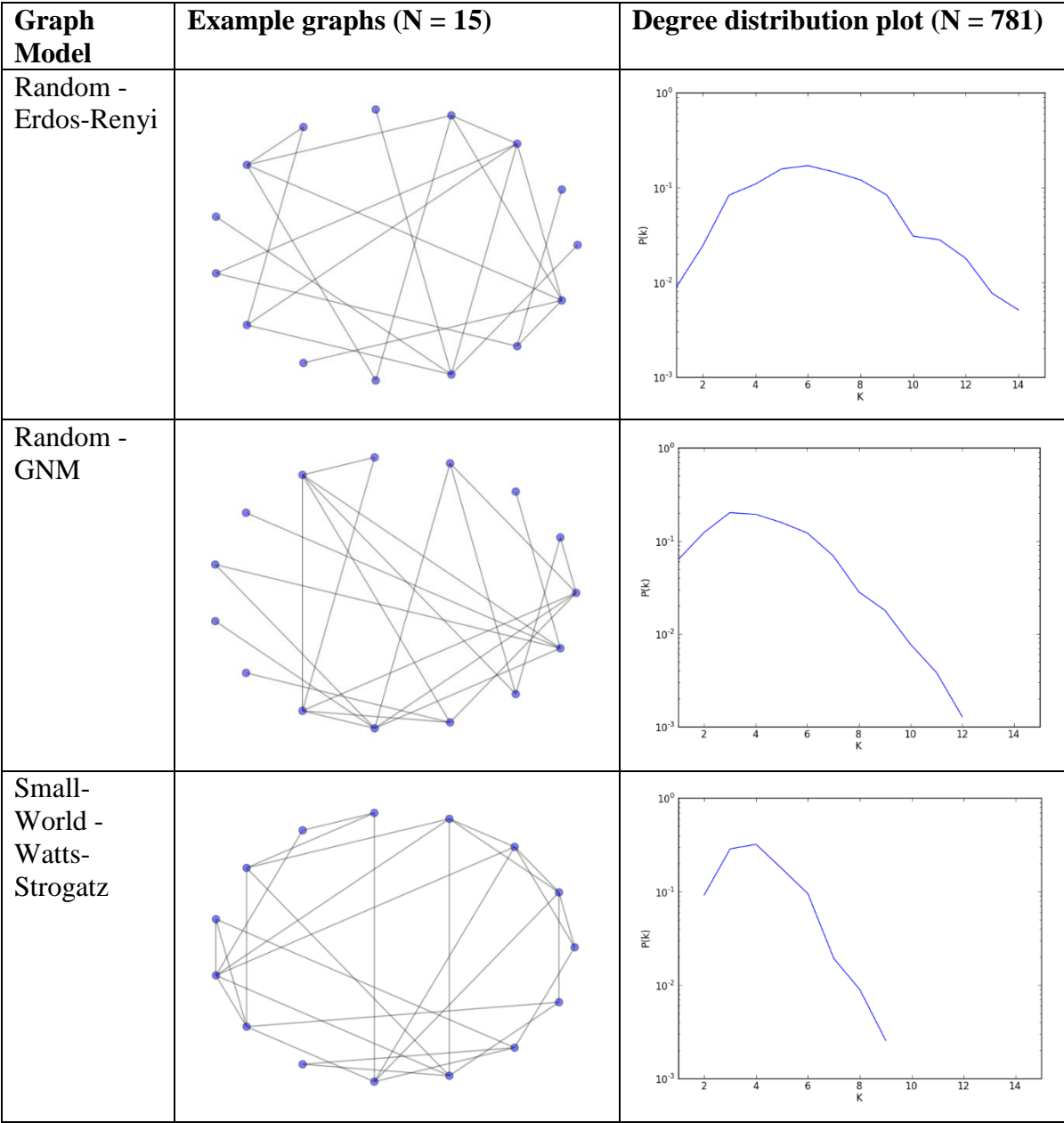
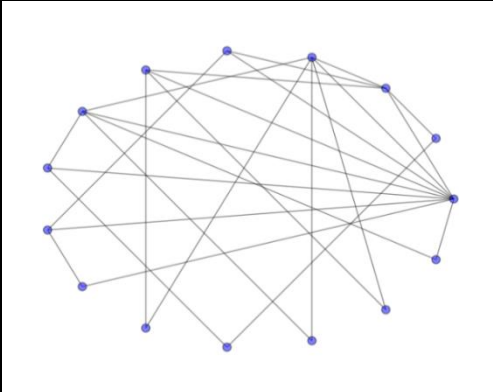
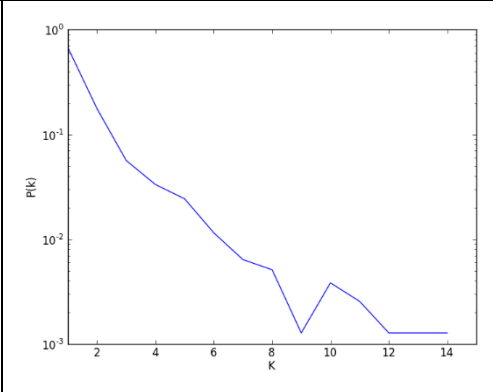
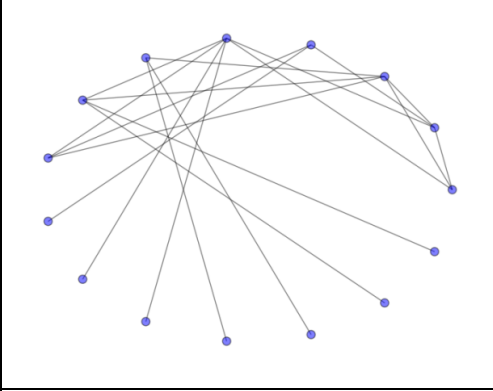
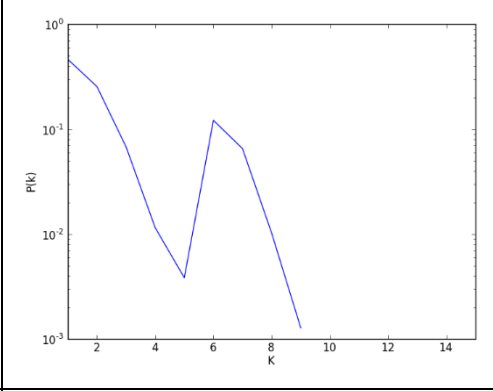
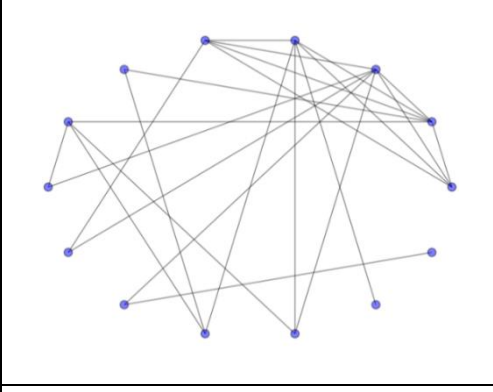
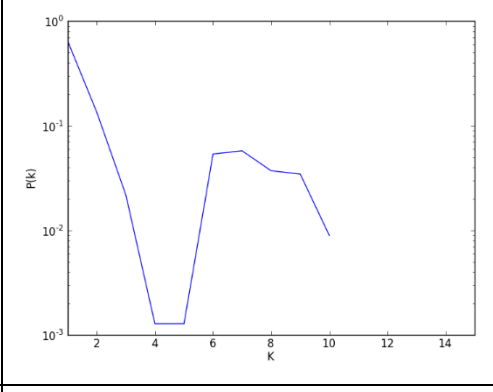
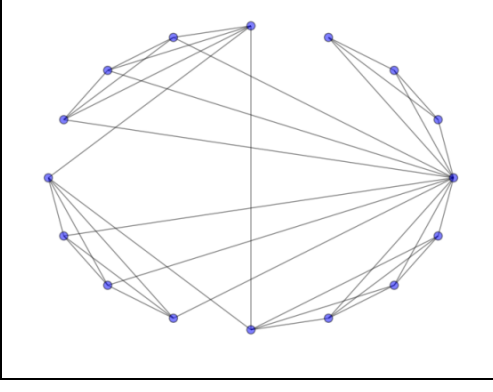
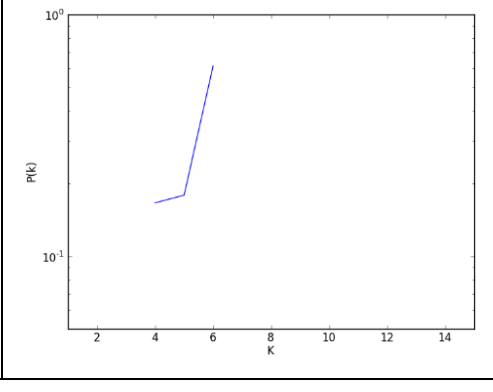


Figure 3.1: Employed graph spectrum.



Scale-free - Barabasi- Albert		
Hierarchical random		
Hierarchical random +		
Hierarchical communities		

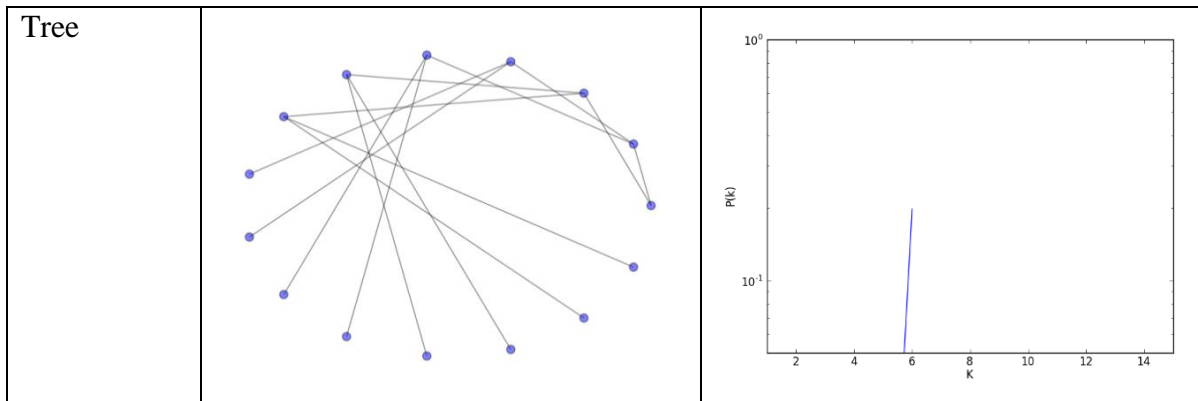


Figure 3.2: The spectrum of graph models through a network diagram and a degree distribution plot, where $P(k)$ is the fraction of nodes with degree k .

Topological structure	Graph model	Parameter(s)	Value range
Random	Erdos-Renyi	Number of nodes	2 - 2000
		Probability value - for the proportion of the total possible edges to be added	0.0 - 1.0
Random	GNM	Number of nodes	2 - 2000
		Number of edges	Minimum required - 2000
Small-World	Watts-Strogatz	Number of nodes	2 - 2000
		Number of closest neighbours connected to	2 - 30
		Probability value – for the proportion of edges to re-wire	0.1 - 1.0
Scale-Free	Barabasi-Albert	Number of nodes	2 - 2000
		Number of edges connected to a new node	1 - 30
Hierarchical	Hierarchical random	Number of levels	2 - 10
		Number of nodes from each parent	2 - 9
		Probability value – for the proportion of new edges to add	0.1 - 1.0
Hierarchical	Hierarchical random +	Number of levels	2 - 10
		Number of nodes from each parent	2 - 9
		Probability value – for the proportion of new edges to add	0.1 - 1.0
Hierarchical community	Hierarchical communities	Triangle or square (0 or 1)	0 - 1
		Number of levels	1 - 4
Tree	Trees	Number of levels	2 - 10
		Number of nodes from each parent	2 - 9

Table 3.1: Graph models and the parameters required for generation of the graphs. More details on the parameters are given in the model specific sub-sections, 3.3.1 to 3.3.8.

3.3.1 Erdos-Renyi graph model

The topological structure of graphs was long presumed to be random (Watts and Strogatz, 1998), consequently leading to the development of the Erdos-Renyi graph model (Erdos and Renyi, 1959). The model is based on a simple premise; connections of a graph develop randomly with no factors affecting which nodes an edge connects (Erdos and Renyi, 1959). The Erdos-Renyi model, also known as $G(n,p)$ (Newman, 2003b; Beygelzimer *et al.*, 2005), generates a graph using two parameters; the number of nodes and a probability value, p ($0 < p < 1$) (Table 3.1). Graphs are generated by creating a node set, N , where the number of nodes is equal to the first parameter. The number of edges to add to the network between nodes is then calculated as a proportion of the total number of possible edges in the graph:

$$\text{number of new edges} = p \times \frac{(N \times (N - 1))}{2} \quad (\text{Equation 3.1})$$

where N is the total number of nodes and p is chosen randomly. Edges are then added at random between different nodes until the calculated number of edges have been added. This creates a graph which has been generated at random giving a unique topological graph structure. These graphs are generated using an existing algorithm available in the NetworkX python library (NetworkX, 2014).

3.3.2 GNM random graph model

The GNM model has also been implemented within the graph suite, generating graphs with a random topology, with the model again using the NetworkX library (NetworkX, 2014). The algorithm generates an instance of a random network for a set number of nodes and edges which are selected at random from a predefined range ($2 \leq N \leq 2000$ and $N-1 \leq E \leq 20000$). The lower bound on the number of edges is set such that a connected network could be potentially generated. As both variables can be set randomly, again 1,000 exemplars are created for this model.

3.3.3 Watts-Strogatz small-world graph model

Increasingly random graph models are considered poor representations of real-world networks such as technological and social networks (Barthelemy and Amaral, 1999). Instead, real-world networks were found to share only some characteristics with random graphs, with Watts and

Strogatz (1998) proposing the small-world model which appeared to share a greater set of characteristics to the real-world networks than the random model.

The Watts-Strogatz model uses three parameters for the generation of the graph; the number of nodes, the number of neighbours each node is connected to and a probability of rewiring ($0 < p < 1$). The first two parameters are used to generate a regular grid of nodes where the total number is equal to the first parameter, with each node then connected to the specified number of neighbouring nodes (the second parameter). The probability value, p , the third parameter, is used to calculate how many of the edges in the regular lattice will be re-wired to create shortcuts across the graph. To ensure only shortcuts are generated, self-loops, where an edge starts and finished at the same node, are not permitted. Where $p = 0$ no edges are rewired and where $p = 1.0$ all edges in the graph are re-wired (Figure 3.3). Therefore graphs generated by the model can vary from those with a regular pattern where the average path length is long as nodes are only connected to a small number of neighbours, to those which have similar characteristics as a random model, such as those produced by the Erdos-Renyi model in Section 3.3.1. Those with a small-world topology will lie within these bounds where only some edges from the regular grid have been re-wired creating the shortcuts across the graph which results in a lower average path length than in a regular grid, though at the same time still leaving groups of nodes connected. Graphs are generated using the available Watts-Strogatz algorithm in the NetworkX python library.

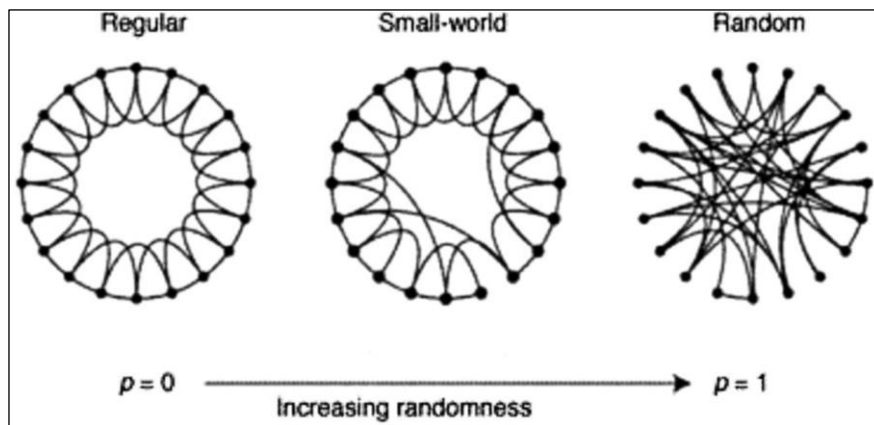


Figure 3.3: The effect of the probability parameter p on the structure of the Watts-Strogatz small-world graph model (Watts and Strogatz, 1998).

3.3.4 Barabasi-Albert scale-free graph model

Barabasi and Albert (1999) developed a further model which better represented some real-world networks, generating graphs with a scale-free topology, where the distribution follows a power law and thus is invariant to graph size (Albert and Barabasi, 2002; Newman, 2003b). Exemplars are created for the suite of synthetic graphs using a model available within the NetworkX python library (NetworkX, 2014), for which two parameters are required. The first is the number of nodes which the graph is to have, again limited to 2,000, and the second the number of edges to add to each new node, set between 1 and 10 so $E \leq 20000$ where $N = 2000$. The graph is generated by adding new edges to each new node added linking this to the existing graph (Figure 3.4), the number of which is set by the second parameter, with a preference for the edges to connect to those nodes which are already well connected (have a high degree). Thus, the larger the number of edges to add for each new node, the greater the likelihood of the presence of highly connected nodes, or ‘hub’ nodes.

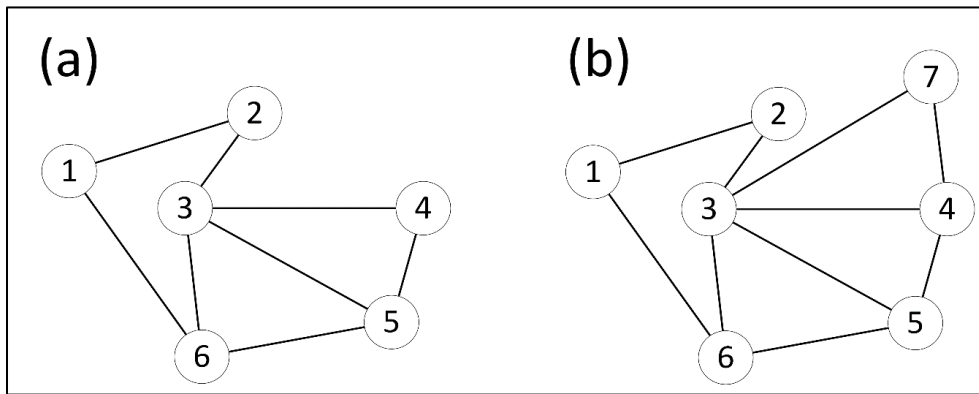


Figure 3.4: Addition of new nodes in the BA model where two edges are added with each new node. (a) shows the base graph, and (b) shows the addition of a new node, 7, and the two new edges (3,7) and (4,7) where the first connects to the node with the highest degree (node 3).

3.3.5 Hierarchical random

To explore the characteristics of hierarchical graphs an example based on the TREE model (Section 3.3.8) has been developed which allows for a range of shortcuts/edges to be added to the graph structure. As with the Watts-Strogatz small-world model (Section 3.3.3), shortcuts are used to make it easier to traverse the graph, improving redundancy (Helbing *et al.*, 2006a). However, unlike the Watts-Strogatz model, these shortcut edges are new to the graph and are not created by re-wiring existing edges, a step performed as the tree model only has the minimum number of edges for it to be connected.

An algorithm (Listing 3.1), has been developed using python and the NetworkX library (NetworkX, 2014) to allow the developed hierarchical random graphs to be generated given the

required set of parameters; the number of levels, the number of branches per node and a probability value, $0 < p < 1$, for the number of new edges to be added. The first two parameters, the number of levels and number of branches, are used to generate the underlying tree graph (detailed in Section 3.3.8). The number of edges to add to the graph is then calculated by multiplying p by E , the number of edges in the graph. Edges are then added to the graph until the number of new edges to add have been successfully added, with edges allowed to connect any two different nodes together, including those in different levels of the tree. p allows for a range of graphs to be generated with different levels of randomness, with a value of 1.0 resulting in a doubling of the number of edges in the graph, with a value of 0 resulting in no new edges being added. For the generation of the graphs, the three parameter values are chosen at random within a set of constraints (Table 3.1).

```

Input:  $h$  :number of levels
Input:  $b$  :number of branches
Input:  $p$  :probability for edges
 $G = \text{balanced tree network}(h,b)$ 
 $E = E(G)$ 
 $N = N(G)$ 
 $E_{old} = |E|$ 
 $E_{new} = |E| \times p$ 
DO
    DO
        iterate = False
         $n_a = \text{random}(N)$ 
         $n_y = \text{random}(N)$ 
        DO
             $n_y = \text{random}(N)$ 
        WHILE  $n_a = n_y$ 
        if  $n_a, n_y$  in  $E_{n_a}$ 
            iterate = True
        WHILE iterate = True
             $E = E + (n_a, n_y)$ 
    WHILE  $|E| \neq E_{old} + E_{new}$ 
RETURN  $G$ 

```

Listing 3.1: Pseudo-code for the HR graph model. More detail is available in Appendix A.

3.3.6 Hierarchical random +

The hierarchical random + model (HR+) uses a similar methodology to the HR model (Section 3.3.5), but with greater constraints on the nodes the new edges can connect to and how the number of edges to add is calculated and assigned throughout the graph. This has been developed additionally to the HR model to provide a further hierarchical model but with much shorter shortcuts, creating a graph which is likely to be more difficult to traverse.

As with the HR model, three parameters are required; the number of levels, the number of nodes and a probability value, $0 < p < 1$, for the calculation of the number of new edges to add. The first two parameters are used to generate the underlying tree graph, the first step in the HR+ algorithm (Listing 3.2). The new edges are then added per level, with the first nodes in each level of the tree graph calculated using the parameters used to generate this. The first set of new edges are added to nodes within the same level of the graph (Figure 3.5(b)), and then the second set between nodes in adjacent levels (Figure 3.5(c)), with each level done separately. The number of edges added for each set varies with the number calculated by multiplying p by the number of level in the graph, and then a value chosen between this and number of nodes in the level. By not allowing edges to connect nodes more than a single level apart the length of these additional edges are less effective as shortcuts compared to those in the HR graphs which can span multiple levels across the graph.

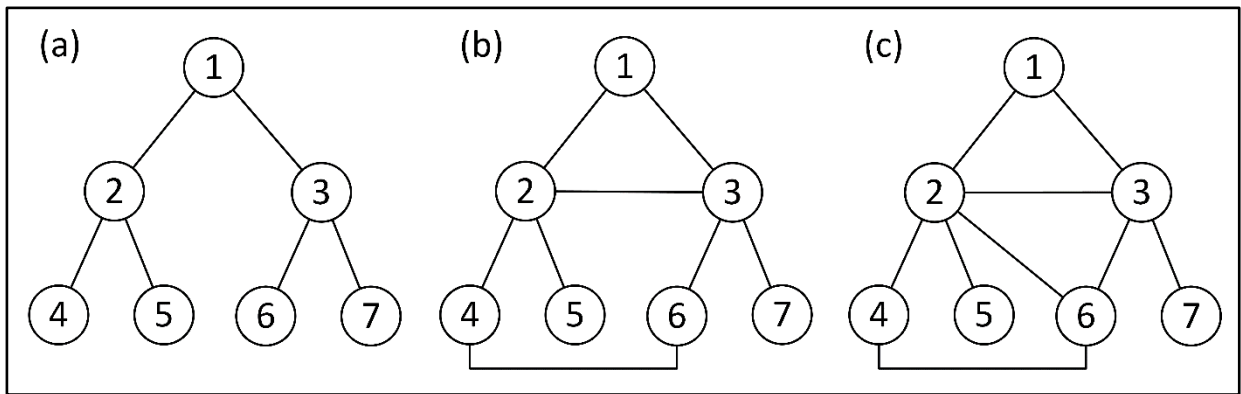


Figure 3.5: The three stages of the HR+ algorithm. (a) a tree network with 2 levels and a branching of 2, (b) the addition of edges within levels (between nodes (2,3) and (4,5)), and (c) the addition of edges between adjacent levels (nodes (2,6)).


```

Input:  $h$  :number of levels
Input:  $b$  :number of branches
Input:  $p$  :probability for edges
 $G = \text{balanced tree network}(h,b)$ 
 $E = E(G)$ 
 $N = N(G)$ 
get_node_levels( $G,h,b$ )
 $i = 1$ 
DO
    Enew = random( $(N_{\text{level } i} \times p) - N_{\text{level } i}$ )
    Enewadded = 0
    DO
        DO
            iterate = False
             $n_a = \text{random}(N_{\text{level } i})$ 
             $n_y = \text{random}(N_{\text{level } i})$ 
            DO
                 $n_y = \text{random}(N_{\text{level } i})$ 
            WHILE  $n_a = n_y$ 
            if  $n_a, n_y \in E_{n_a}$ 
                iterate = True
            WHILE iterate = True
                 $E = E + (n_a, n_y)$ 
                Enewadded += 1
        WHILE Enewadded < Enew

    Enew = random( $(N_{\text{level } i} \times p) - N_{\text{level } i}$ )
    Enewadded = 0
    DO
        DO
            iterate = False
             $n_a = \text{random}(N_{\text{level } i})$ 
             $n_y = \text{random}(N_{\text{level } i+1})$ 
            if  $n_a, n_y \in E_{n_a}$ 
                iterate = True
            WHILE iterate = True
                 $E = E + (n_a, n_y)$ 
                Enewadded += 1
        WHILE Enewadded < Enew
     $i = i + 1$ 
WHILE  $i < h+1$ 
RETURN  $G$ 

```

Listing 3.2: Pseudo-code for the HR+ graph model. More detail is available in Appendix A.

3.3.7 Hierarchical communities

A graph termed hierarchical communities (HC), derived by Ravasz *et al.* (2002) has been included within the suite of synthetic graphs due to the explicit community structure it contains which is similar to those found in social and metabolic networks (Ravasz *et al.*, 2002; Barabasi

et al., 2003), Figure 3.6. An algorithm has been developed to replicate the graphs produced by Ravasz *et al.* (2002) and Ravasz and Barabasi (2003), where the community sizes were set to three and four nodes. The developed algorithm for the hierarchical community graphs is detailed in Appendix A. The function allows the community size to be specified as the first parameter (limited to 3 or 4) as well as the number of levels, currently limited to four due to all graphs in the suite having $N \leq 2000$. Using the community size, the number of communities required is first created, and then the edges between the created communities added to form the HC graph. Due to the $N \leq 2000$ bound only seven instances are generated due to the rigid structure of the model.

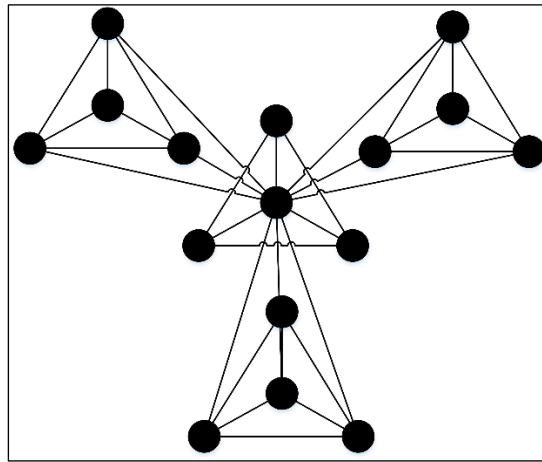


Figure 3.6: Example of a hierarchical community graph (Ravasz *et al.*, 2002).

3.3.8 Tree

The tree graph type is one of the simplest types of graphs (Jungnickel, 2004) and has a hierarchical structure. An example of real-world systems which exhibits this structure are river networks (Barthelemy, 2011). Exemplars for the synthetic suite are created using the balanced tree algorithm available in NetworkX (NetworkX, 2012), which generates a tree network with a symmetrical pattern (Figure 3.7). This algorithm requires two inputs; the number of levels (how deep the network is, e.g., 3 in Figure 3.7), and the number branches (or the number of children for each parent node, e.g., 2 in Figure 3.7). Due to the structure of this network type, only 36 exemplars are created with $N \leq 2000$.

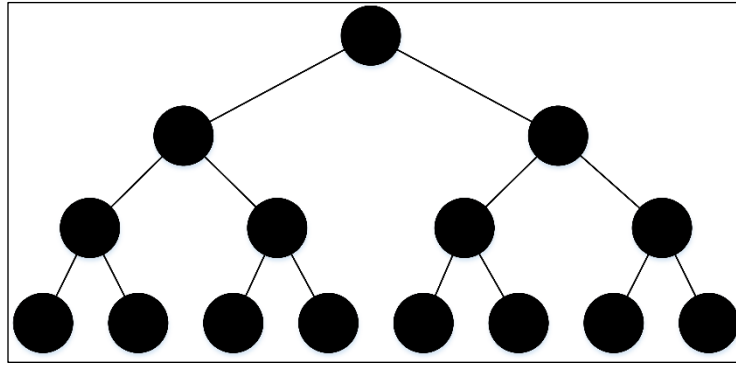


Figure 3.7: Example graph generated by the balanced tree algorithm using input values of 3 (the number of levels) and 2 (the number of branches).

3.4 Statistical Comparison of graph types

The suite of graph models are analysed to establish the key characteristics that allow hierarchical graphs to be recognised from non-hierarchical graphs. This is achieved through first comparing the degree distributions of the graphs, Section 3.4.1, a measure which reports on the topological structure of the graph (Newman, 2003b). In Section 3.4.2 graph metrics are used to characterise the graphs to achieve a characterisation of each of the eight synthetic graph types. The methods used to assess these results are then presented in Section 3.4.3 and 3.4.4.

3.4.1 Degree distributions

One of the most widely used graph descriptors is the degree distribution (e.g. Barabasi and Albert (1999), Amaral *et al.* (2000), Barabasi *et al.* (2000), Jeong *et al.* (2000) and Barrat *et al.* (2005)), the distribution of node degrees in a graph, which describes the connectivity of the nodes by using the probability of selecting a node with a set degree (Newman (2003b)). This allows for a qualitative comparison of graph types, but for large scale analysis of graphs, it can be difficult to use due to the visual interpretation which is required to make an assessment. The degree distribution is calculated over the entire suite of synthetic graphs (up to 1,000 for each type), allowing the topological characteristics of each graph type to be assessed.

3.4.2 Metrics

Typically network/graph characteristics have been identified through metrics including the degree distribution, the average shortest path length and the clustering coefficient (Albert and Barabasi, 2002; Newman, 2003b; Amaral and Ottino, 2004; Boccaletti *et al.*, 2006). These three

metrics offer a global perspective on the topological structure of a graph allowing the connectivity of the nodes to be described and hence can be used to help differentiate between different graph topologies (Newman, 2003b).

As already mentioned the clustering coefficient is widely used as one of three metrics traditionally used in identifying the structure of networks. The metric, as defined in Chapter 2 Section 2.3.1 (page 13), calculates a value per node as to the number of loops of length three which it is part of. Within many networks there are often loops of a higher order found in graphs (Caldarelli *et al.*, 2004; Kim and Kim, 2005), which can be an important characteristic of the graphs being analysed, but are missed by the clustering coefficient (Caldarelli *et al.*, 2004), and this is not an ideal measure to characterise some graphs/networks (Holmgren, 2006). Previous research has hence explored the effectiveness of using other measures which account for longer cycles with these highlighting some new characteristics of graphs (Caldarelli *et al.*, 2004; Boccaletti *et al.*, 2006) or has suggested the use of the number of cycles (Holmgren, 2006).

Along with the clustering coefficient the average path length, defined in Chapter 2 Section 2.3.2, has been used as a metric for identifying the characteristics of graphs (Albert and Barabasi, 2002; Costa *et al.*, 2007), as well as in the assessment of the robustness of graphs and networks to perturbations (Newman, 2003b). The metric provides a measure of how well a graph is connected (Newman, 2003b), with the value an average over all shortest paths between each pair of nodes. This gives an indication of the proximity of each node to all other nodes, but using the average can be problematic (Evans, 2010), with in this case the average being taken over all node pairs, resulting in extreme values being lost, such as those for nodes which are very poorly connected to the rest of the graph. Further to this, if the graph is not connected, there is a node or group of nodes which are not part of the main network, these do not have a path to all nodes, and thus cannot return a shortest path length to be included in the average for the whole graph (Costa *et al.*, 2007). This results in any value returned for the average path length not truly reflecting the structure, and thus the connectivity, of the graph.

The diameter graph metric, defined in Chapter 2 Section 2.3.2, can also be used as a measure for the structure of graphs/networks (Gastner and Newman, 2006), with it reporting on the greatest path length between any two nodes in the graph (Newman, 2003b). It forms a measure of how well a graph is interconnected (Albert *et al.*, 2000), and has been used as a measure for the characterisation of graph structures (Albert *et al.*, 1999; Barabasi *et al.*, 2000; Bagler, 2008a). It has also been used to measure the resilience of graphs to perturbations (Albert *et al.*, 2000), reporting on how perturbations change how well the network is connected, with an increased value suggesting a less connected graph. However, as with the average shortest path

length as discussed in the above paragraph, where a node or group of nodes become disconnected from the rest of the graph, there is no longer paths possible between all node pairs as required for the honest calculation of the metric, and thus any reported values do not reflect the true structure of the graph.

Three commonly used graph metrics have been discussed in the previous paragraphs, including the clustering coefficient, the average shortest path length and the graph diameter. These have previously been employed in the characterisation of graphs, though all have some limitation (Ouyang *et al.*, 2009). A suite of three alternative metrics is thus proposed to be used in this research and are discussed in the following paragraphs. These are, the maximum node betweenness centrality (Section 3.4.2.1), the assortativity coefficient of the graph (Section 3.4.2.2) and the number of cycle basis in the graph (Section 3.4.2.3).

3.4.2.1 Betweenness centrality

Betweenness centrality, as defined in Chapter 2, Section 2.3.1 (page 13), returns centrality values for the vertices which have the greatest number of shortest paths passing through them when the shortest paths are calculated over all node pairs (Girvan and Newman, 2002). The highest values are those nodes with the greatest number of shortest paths and most critical to the ability to traverse the graph (Girvan and Newman, 2002). The removal of these will not only result in metrics such as the average shortest path length increasing, but may also result in the graph fragmenting into a greater number of components (Holme *et al.*, 2002; Luca *et al.*, 2006; Börner *et al.*, 2007) and therefore is considered an import metric for consideration in infrastructure networks as well (Wuellner *et al.*, 2010). The metric is also considered as a proxy for flows through graphs allowing it to be used when simulating the behaviour of graphs when perturbed (Dueñas-Osorio and Vemuru, 2009; Mishkovski *et al.*, 2011). For these reasons, the maximum betweenness centrality value is employed as a measure as it allows for the identification of a graphs dependence on a single critical node, something which might be expected in hierarchical graphs, such as those similar to the tree model (Section 3.3.7).

The ability of the betweenness centrality to identify key nodes which may be critical in a graph staying connected or for the average shortest path length to remain low can be exemplified when looking at hierarchical graphs. For those graphs with an explicit tree structure the node at the top level of the hierarchy connects two distinct parts of the graph together, node 1 in Figure 3.8(a), and thus has a higher betweenness centrality value compared to all other nodes in the

graph. However, by adding more links to the tree structure the betweenness of the nodes at the top decreases, as shown in Figure 3.8(b), as shortest paths no longer use this node.

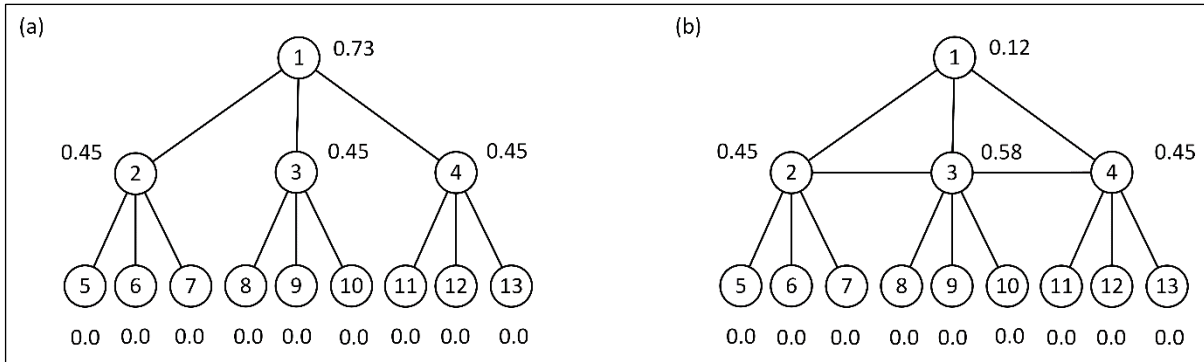


Figure 3.8: The betweenness centrality (normalised) values variations in a tree graph (a) and where two extra edges (between nodes 2 and 3 and between 3 and 4) have been added to a previously identical tree graph (b).

3.4.2.2 Assortativity coefficient

The assortativity coefficient allows for the characterisation of the structure of the graph through reporting on the similarity of the degree of the nodes that each node is connected to, describing the topological correlation between the degree of the nodes in a graph (Barthelemy, 2003). A value close to one suggests an assortatively mixed graph (nodes are connected to nodes with similar degrees (Figure 3.9(b)), and negative one indicates a disassortatively mixed graph (nodes are connected to nodes with different degrees Figure 3.9(a)) (Newman, 2002). It has been suggested that an assortatively mixed graph has no significant dominance of high and/or low connected nodes, a feature which implies a robust graph (Newman, 2002).

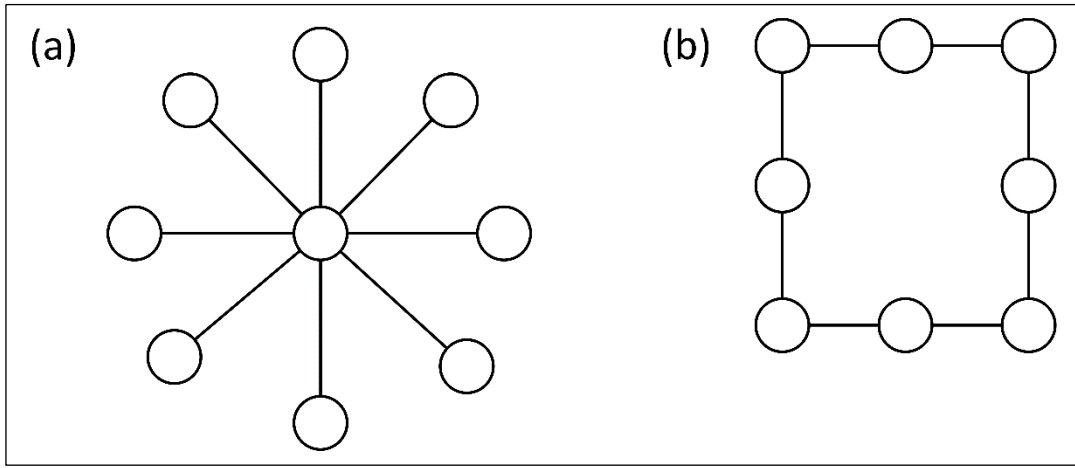


Figure 3.9: A dissassortatively mixed network (a) where nodes with high degree are only connected to nodes with low degree and (b) an assortative network where every node is connected to another with the same degree.

The assortativity coefficient is defined as the Pearson correlation coefficient of the degree of the nodes at opposite ends of an edge (Newman, 2003a):

$$r = \frac{\sum_i e_{ij} - \sum_i a_i b_j}{1 - \sum_i a_i b_j} \quad (\text{Equation 3.2})$$

where r is the correlation coefficient ($0 < r < 1.0$), e_{ij} is the fraction of edges that connect a vertex of degree i to a vertex of degree j and $a_i b_i$ are the fraction of each type of edge end that is attached to a vertex with degree i .

3.4.2.3 Cycle basis

Cycle basis are the fundamental set of cycles which make up all cycles found within a network (Paton, 1969; Kavitha *et al.*, 2009). A cycle base can be any length greater than 2 edges as shown in Figure 3.10 (b) which shows the cycle basis in Figure 3.10(a), where a cycle is a path through at least three nodes where all edges and nodes are distinct except the start and finish node (Dolan and Aldous, 1993). The presence of a large number of cycle basis suggests that a network is well connected with nodes having a high number of neighbours (Barthelemy, 2011). This provides a suggestion as to the topological structure of the graph as well as providing an indication to the robustness of the graph, as with more cycles there is likely to be a greater

number of routes between nodes and hence the graph is likely to be topologically more robust to failures (Katifori *et al.*, 2010; Barthelemy, 2011).

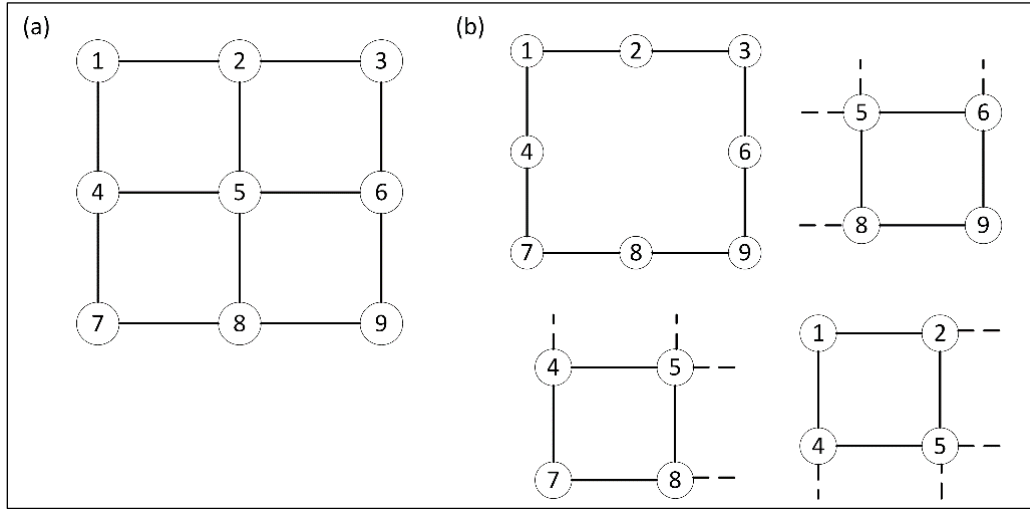


Figure 3.10: Example of cycle basis in a regular grid (a) which has four cycle basis (b).

3.4.3 Multivariate metric analysis

The metric values computed for the suite of synthetic graphs are compared through their multivariate distributions for the three metrics, resulting in three sets of results; assortativity coefficient and maximum betweenness centrality, the assortativity coefficient and the number of cycle basis per node, and finally the maximum betweenness centrality and the number of cycle basis per node. The values for each graph are used for each multivariate combination of metrics. Scatter plots are used to compare the graph models with single standard deviation ellipses used to show the extents of the distribution of the values for the graphs for each model. This method allows the overlap between the different graph models to be shown.

3.4.4 Transformed divergence

As described in Section 3.4.3 the multivariate results from the analysis of the suite of synthetic graphs will be analysed to compare the results from the eight graph models. To statistically test the similarity of the metric values to compare the characteristics of the hierarchical and non-hierarchical graph models, the transformed divergence statistic is used to compute the degree of separability between the graph models for the different metric combinations. The test is a multivariate, bi-directional pairwise test, returning a value which can be used to quantitatively

asses the separability between two sets of values. For each pair of graph types the divergence (D_{ij}) between their structural metrics is defined as (Swain and Davis, 1978) :

$$D_{ij} = \frac{1}{2} \text{tr} \left((C_i - C_j)(C_i^{-1} - C_j^{-1}) \right) + \frac{1}{2} \text{tr} \left((C_i^{-1} - C_j^{-1})(\mu_i - \mu_j)(\mu_i - \mu_j)^T \right) \quad (\text{Equation 3.3})$$

where C_i is the covariance matrix of the metric values for graph type i (C_j for graph type j), C_i^{-1} is the inverse of the covariance matrix of the metric values for graph type i and μ_i is the mean vector of the metric values for graph type i , T is the transposition function and tr is the trace function. In order to get a divergence value in the range 0 – 100 such that it can be interpreted as the probability of separability it is normal to apply a saturation function of the form:

$$TD_{ij} = 100 \left(1 - \exp \left(\frac{-D_{ij}}{8} \right) \right) \quad (\text{Equation 3.4})$$

where $-D_{ij}$ is taken from (Equation 3.3). This results in values between 0 and 100 being returned; 0 indicates two identical sets of values while a value of 100 indicates that they do not overlap at all in their multivariate distributions.

3.5 Topological failures

To explore the characteristics of the synthetic graphs further the robustness of the graphs is explored through a topological failure model. Robustness of graphs is critical for them to withstand perturbations, with infrastructure networks sharing some of the same characteristics exhibited by graph models as discussed in Chapter 2, when exposed to hazards which have the potential to disrupt their functioning (Boccaletti *et al.*, 2006).

A topological failure model has been implemented recognising previous models developed (Albert *et al.*, 2000; Callaway *et al.*, 2000; Crucitti *et al.*, 2004b) for exploring the topological robustness of a network (G), Figure 3.11. Starting with G , with node set N and edge set E , for the first epoch a node, n_i , is selected to be removed based on the method set for the simulation, random, degree or betweenness ((a) in Figure 3.11) (detailed later). When n_i is removed from G , each edge, e , incident to n_i is also removed.

As a result of the removal of a n_i and its edges, $\{e_i^1, e_i^2, \dots, e_i^m\}$, the topology of G changes and it may become disconnected such that there is no longer a path from every node to every other

node (Dolan and Aldous, 1993; Jungnickel, 2004), and/or components form subgraphs which are subsets N and E , but are disconnected (from the largest component) (Dolan and Aldous, 1993) and/or isolated nodes occur which become disconnected from G and have a degree of zero (Dolan and Aldous, 1993). These last two scenarios, the formation of components and isolated nodes, can be handled differently depending on the parameterisation of the failure simulation, either being removed or left in G at the end of each epoch.

For the simulations undertaken, isolated nodes are removed from G as these are disconnected from G , and components are left in G as these still form an ‘active’ part of G . Where either of these features are left in G , all future metric calculations must consider the presence of such features, as metrics such as the average shortest path length of the network, the average of the shortest path between every pair of nodes in the network (Costa *et al.*, 2008), can be affected if G is not fully connected, as no paths will exist between some of the node pairs.

The implemented failure model as described above and in Figure 3.11 allows the removal of nodes for examining the robustness of networks to topological failures. Three methods of node selection are used to explore this, detailed in 3.5.1, falling broadly into two categories, random and targeted, where targeted methods are those which focus on the most critical nodes in a network, identified using metrics such as node degree (Albert *et al.*, 2000; Callaway *et al.*, 2000), and the random method selects nodes at random from those in the network. For each simulation one method is used throughout.

At the end of each epoch (Figure 3.11), metrics are calculated over G to record the changing topological structure of G as a result of the perturbation. A set of metrics have been chosen to record the behaviour of the networks analysed and are detailed in Section 3.5.2.

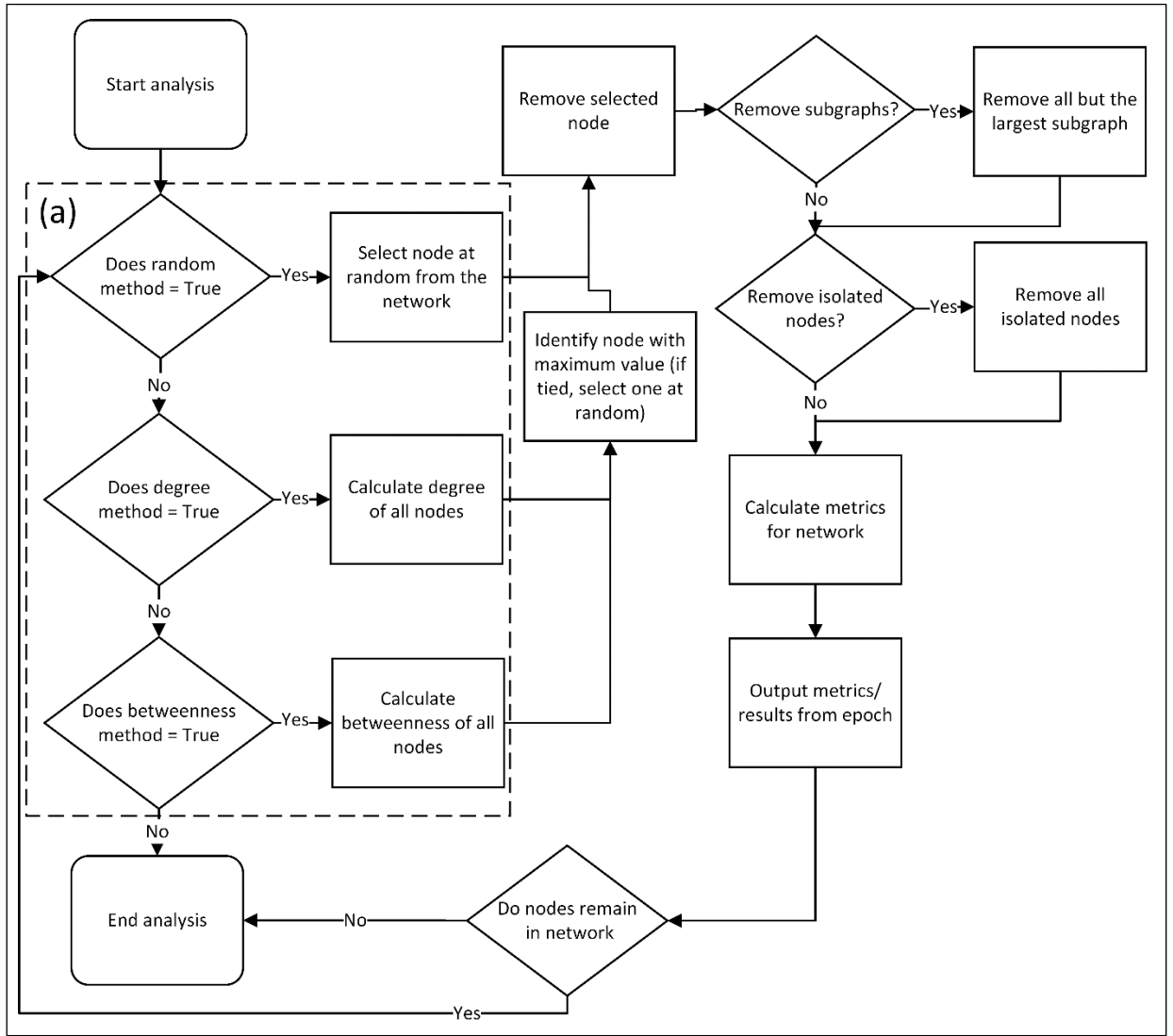


Figure 3.11: Process diagram of the developed topological failure model.

3.5.1 Methods of node selection

Three failure methods are employed in the failure model; (i) random, (ii) node degree and (iii) node betweenness, detailed below. Such methods have previously been employed in the assessment of the robustness of graphs (Albert *et al.*, 2000; Holme *et al.*, 2002; Tanizawa *et al.*, 2005) and infrastructure networks (Bagler, 2008a; Bompard *et al.*, 2011; Lordan *et al.*, 2014) previously, as also detailed in Chapter 2. The difference in the way the three methods target nodes in the graphs and the response from the graphs (Figure 3.12), helps to explore and understand the characteristics of the eight models better than would be possible if only one or two of the methods were employed. This will also help in the recognition of the characteristic differences between the hierarchical and non-hierarchical models.

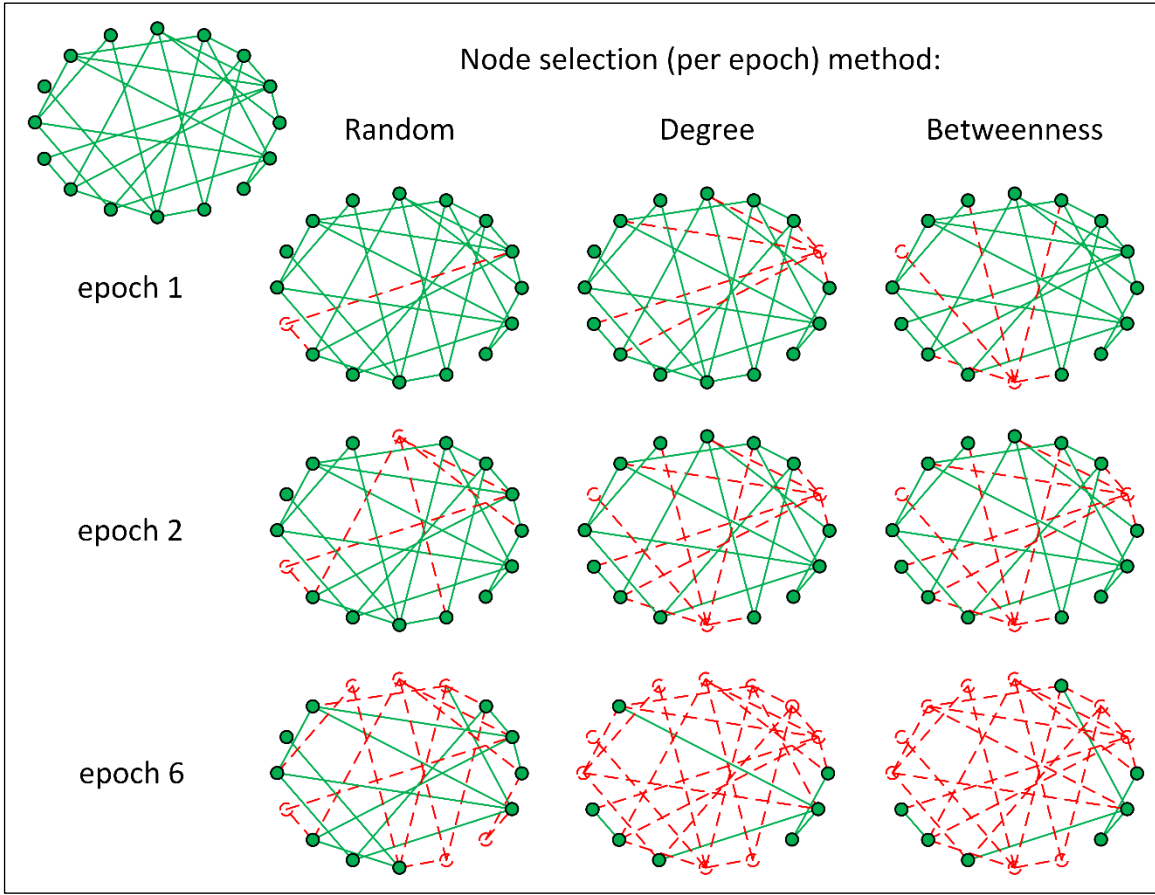


Figure 3.12: Exemplifying the three different node selection methods during a topological failure model using the same example network at three selected epochs, 1, 2 and 6. Failed nodes and edges are shown in red with dashed lines.

The random node selection method simulates the random failure of network assets, such as breakdowns and maintenance periods (Albert *et al.*, 2000). These failures tend to have a lesser impact when compared to the degree based method (Crucitti *et al.*, 2004b). This is demonstrated in Figure 3.12 where the random method only removed six nodes having a smaller impact on the network topology than the other two methods.

The node degree targeted failure method removes the node with the greatest number of incident edges at each epoch. Node degree has been used as a measure of node importance in a graph, as the node with the most edges has been viewed as critical to the connectivity of the network and hence the ease that a network can be traversed (Holme *et al.*, 2002). Using similar methods to those employed previously (Albert *et al.*, 2000; Holme *et al.*, 2002), at each epoch the node with the most edges is removed causing the number of edges in the network to be reduced. Within the failure model, the node degrees are re-calculated at each epoch, ensuring that at each

epoch it is the most connected node that is removed. As expected this method has a tendency to cause greater disruption to the network than the random selection method, as shown in Figure 3.12, which shows the degree based method has a greater impact than the random method by removing two more nodes and six more edges by epoch 6.

The betweenness centrality node selection method removes the nodes which are most critical to shortest paths through the network (Dueñas-Osorio and Vemuru, 2009; Mishkovski *et al.*, 2011), with betweenness centrality defined in Section 3.4.2.1 (Equation 2.2). The nodes with the greatest values are seen as the most critical in the network as these have the greatest number of shortest paths passing through them, thus those removed first in the analysis are the nodes which are on the most shortest paths through the graph and therefore critical to the topological connectivity. The impact of removing these is expected to have a significant impact on the network, equal or greater to that of the degree based method (Holme *et al.*, 2002), as this just selects those nodes which have the most connections, whether are not they are critical to the connectivity of the graph. As with the node degree method, the betweenness values are recalculated in the failure model at each epoch so the node with the greatest value at each epoch is removed.

3.5.2 Recording failure behaviour

At the end of each epoch (T) in the failure model (Figure 3.11), a suite of metrics are computed to characterise the state of the network following the removal of a node. Previous studies have used metrics such as the average shortest path length (Albert and Barabasi, 2002; Holme *et al.*, 2002) and the size of the giant component (Holme *et al.*, 2002; Bassett and Bullmore, 2006; Lordan *et al.*, 2014). However, the average shortest path length as a measure of network robustness becomes poor when a network starts to fragment, when groups of nodes (or single nodes) become separated from the largest connected part of the network (the giant component). The average path length as a metric cannot report the fragmentation of the network as not all nodes in the network are connected to each other, and thus have no path length between them.

The first metric employed is the average number of nodes removed before a graph becomes null. This allows for an assessment of the rate of failure in a network while being perturbed through knowing the fraction of nodes affected/removed (e.g. Barabási *et al.* (2001) and Beygelzimer *et al.* (2005)).

The second method is the number of components and the average size of these as a function of the number of epochs of the simulation(s). This is employed as it shows the behaviour of the

network with regard to its fragmentation and the formation of components (connected nodes which become disconnected from the rest of the network (Albert and Barabasi, 2002)). The fragmentation of a network is associated with a vulnerability to perturbations, with more components expected to form if a network is vulnerable to failures (Newman, 2003b; Costa *et al.*, 2007).

3.6 Real-world spatial infrastructure data

In addition to the suite of synthetic graphs models, spatial infrastructure networks have been generated from a series of spatial datasets. These are split across six infrastructure sectors; air (Section 3.6.1), communications (Section 3.6.2), energy (Section 3.6.3), rail (Section 3.6.4), rivers (Section 3.6.5) and roads (Section 3.6.6). A total of 42 networks were created including some variants of the same infrastructure networks to account for different levels of granularity. For each sector a summary is given in the appropriate sub-sections, with more details available in Appendix B, including a summary of how the networks were generated.

Topologically valid networks are created using a suite of tools developed and provided by the Infrastructure Transitions Research Consortium (ITRC) (ITRC, 2013), and through the use of GIS software. Tools have been developed to solve errors in data such as nodes/edges being disconnected as in Figure 3.13, which shows part of the electricity network provided by National Grid, and for errors such as over and under shoots where edges don't meet exactly.

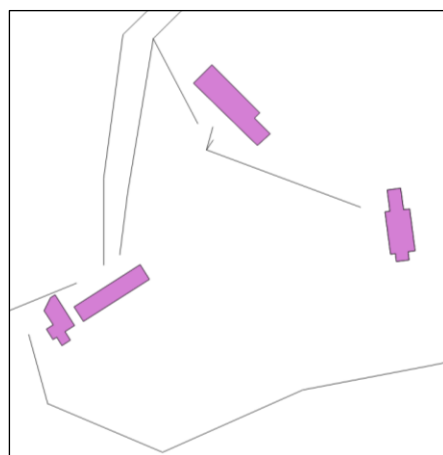


Figure 3.13: Highlighting the topological errors contained within some datasets which required correcting to form topologically valid networks.

3.6.1 Air networks

Six air networks have been created using 2012 data from OpenFlights, a freely available online resource which contains all known airport locations along with the routes which serve these. From this dataset air networks have been generated for four regions; the UK, Europe, United States of America and North America (Figure 3.14(b)), along with the networks of two providers, British Airways and EasyJet (Figure 3.14(a)). These networks have been processed using developed tools to ensure topological integrity.

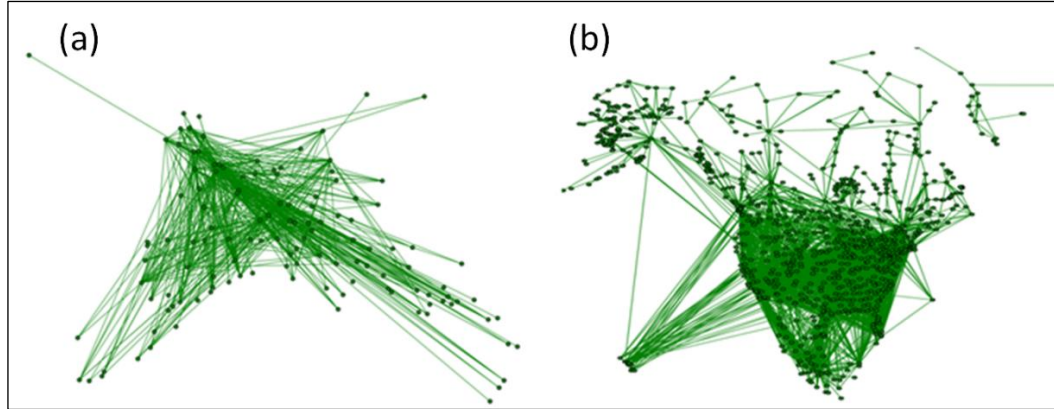


Figure 3.14: EasyJet flight network ($N = 125$, $E = 498$) (a) and the network for North America ($N = 889$, $E = 3760$) (b).

3.6.2 Communication networks

A single network falls within this category, the core JANET network for the provision of high speed internet connections for academic institutions in the UK (Jisc, 2015). This has been digitised using GIS software from schematic network diagrams which are freely available. The network consists of the locations of the main connections within the system resulting in a network with 38 nodes and 58 edges (Figure 3.15).

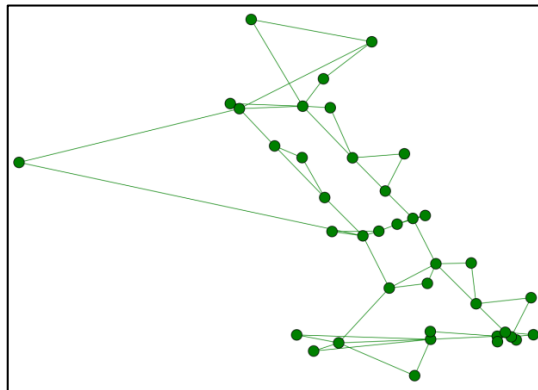


Figure 3.15: The JANET network ($N = 38$, $E = 58$).

3.6.3 Energy networks

Two primary energy networks, electricity and gas, are included in the suite of networks with data provided through the ITRC project, but originally from National Grid, the owners and operators of the main transmission systems. This data has been edited to form topologically valid networks using the suite of tools developed as part of ITRC project as previously mentioned. In total five networks form this category, including three variants of the electricity transmission network which include varying levels of detail with regard to transmission pylons, with the largest, the full network (Figure 3.16(a)), having 23,787 nodes and 24,185 edges and the smallest having 2,218 nodes and 2,520 edges (Figure 3.16(b)). As well as these, a further generated/simulated network for electricity transmission and distribution as provided by the ITRC project has been included which provides a network for England and Wales to the 11Kv level, with 170,667 nodes and 172,019 edges. The full suite of energy networks is detailed in Appendix B with the networks shown along with the size of the node and edge sets.

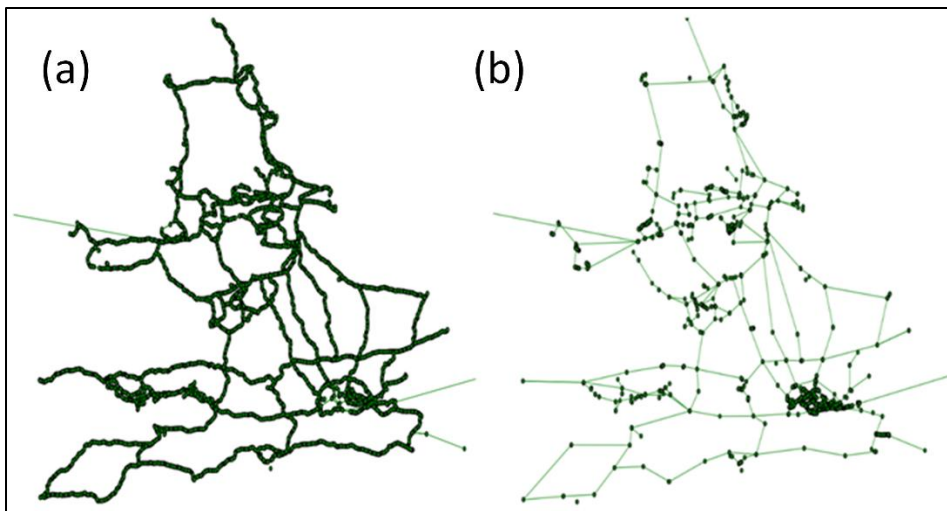


Figure 3.16: The full national grid transmission network ($N = 23787$, $E = 24185$) (a) and the NT (no towers) version ($N = 2218$, $E = 2520$) (b).

3.6.4 Rail networks

A large suite of rail networks, totalling 19 examples (including variants), has been created covering the UK, Ireland, Paris (France) and Boston (United States of America). The UK based examples have been generated from the Ordnance Survey Meridian 2 data using GIS systems and developed tools from the ITRC project, to build topologically valid models. These cover a range of scales, from the national network to local scale such as the London Tube network (Figure 3.17(a)) and the Manchester Metrolink (tram) system. Open Street Map (OSM) data,

along with system maps for verification purposes, have been used to create those networks outside of the UK, such as the network for Ireland (Figure 3.17(b)). The full suite of rail networks can be found in Appendix B with details of the size of the node and edge sets included.

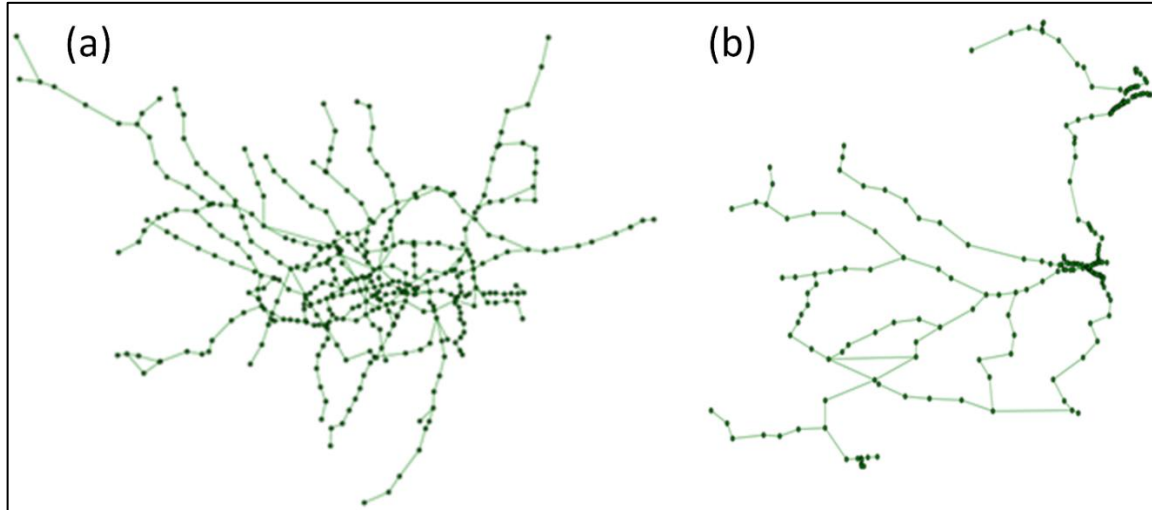


Figure 3.17: Networks for (a) London Tube ($N = 436$, $E = 466$) and (b) Ireland ($N = 201$, $E = 203$).

3.6.5 River networks

Four river networks have been created using the Ordnance Survey Meridian 2 data, including the River Dee, the River Eden, the River Severn (England) and the River Tyne (Northern England), (Figure 3.18). As with the other infrastructure networks these have then been processed to ensure topologically valid models are produced for analysis using the ITRC tools.

3.6.6 Road networks

Ten road networks were generated. Networks for the UK have been built using Ordnance Survey Meridian 2 data with a range of detail with some editing to create topologically valid networks. This includes regional scale networks for three areas, Tyne and Wear, Leeds and Milton Keynes. For these regions/areas networks with different levels of detail through the inclusion of varying road classes were created, from motorways to minor roads, resulting in seven networks. For example, for Tyne and Wear three networks are created, each more detailed than the previous, starting with motorways and A roads (Figure 3.19(b)) culminating in a network with all road classes with 15,249 nodes and 21,817 edges (Figure 3.19(a)). At the national scale, motorways, and A and B roads are used for the UK, generating a network with

24,071 nodes and 50,292 edges. One other network has been created using Open Street Map (OSM) data for Ireland (including Northern Ireland) and contains the motorways, primary and trunk roads. Details for the full suite of road networks is given in Appendix B including the size of the node and edge sets for all networks.

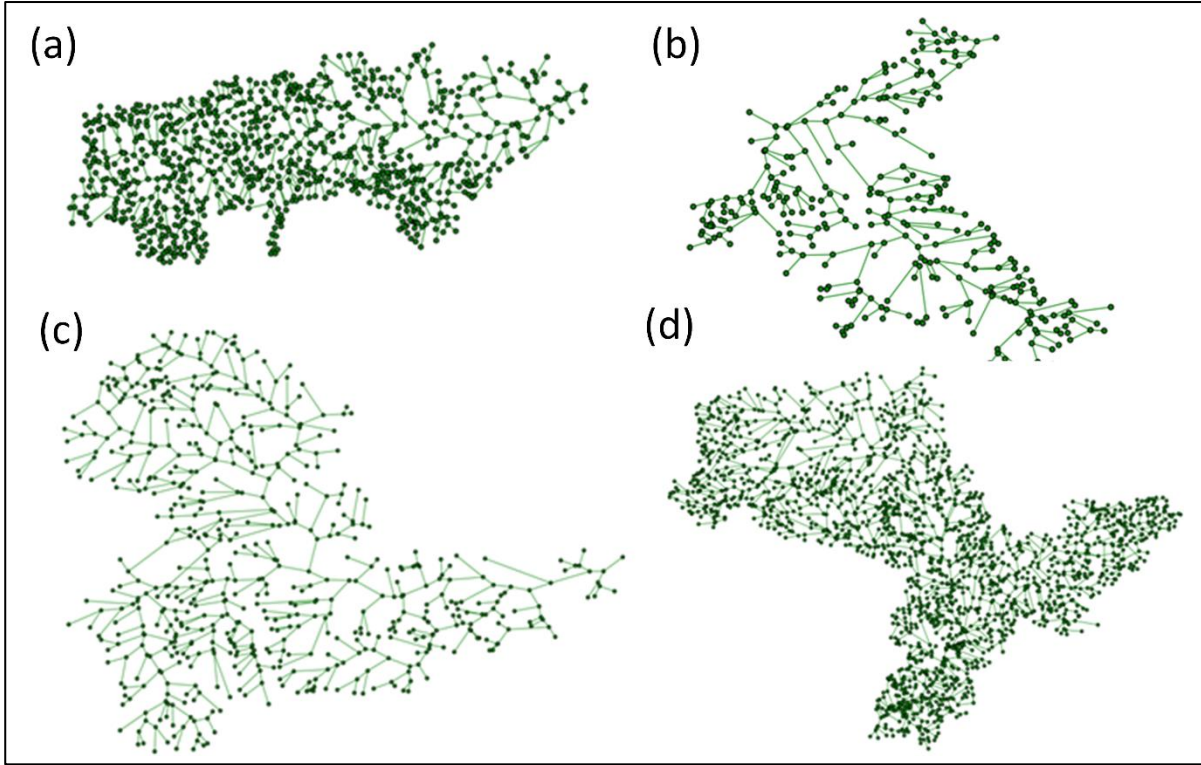


Figure 3.18: The four river networks; (a) Dee ($N = 896$, $E = 900$), (b) Eden ($N = 302$, $E = 301$), (c) Tyne ($N = 616$, $E = 615$) and (d) Severn ($N = 1944$, $E = 2005$).

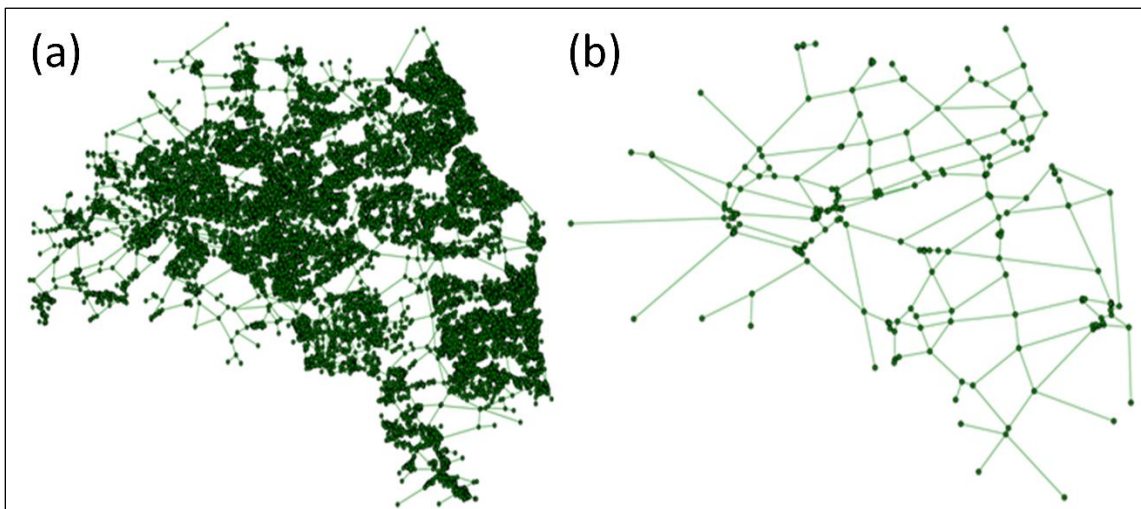


Figure 3.19: Road network for Tyne and Wear with motorways, A, B and minor class roads ($N = 15249$, $E = 21817$) (a) and with only motorways and A roads ($N = 212$, $E = 311$) (b).

3.7 Identifying hierarchical infrastructure networks

Employing the characteristics of hierarchical graphs learned from the analysis of the suite of synthetic graphs (Section 3.4), the suite of critical infrastructure networks (Section 3.6) will be analysed using the same methodological processes to recognise those with hierarchical traits. The degree distributions of the networks will be computed and analysed to identify the topological structure of the infrastructure networks, including a direct comparison to the distributions of the synthetic suite of graphs. A metric analysis will then be performed, calculating the same three metrics as used for the synthetic graphs (Section 3.4.2) to characterise the networks, helping to recognise those which share similar characteristics to hierarchical graphs. Finally the topological robustness of the infrastructure networks will be explored, again employing the same methods as used for the analysis of the suite of synthetic graphs in Section 3.5 (page 59). This analysis is employed in order to recognise any critical spatial infrastructure networks that are seemingly hierarchical and to understand their robustness to different forms of perturbation.

3.8 Enhanced network representation

Infrastructure networks are designed to deliver a commodity/information, both involving a flow over the network (Little, 2003; Ash and Newth, 2007; Dueñas-Osorio and Vemuru, 2009). Therefore, modelling the flows over the network gives a better representation of how network behaves. The developed representation model allows four node and four edge attributes to be modelled (Figure 3.20), and are detailed in Table 3.2. The attributes are stored for each node, $n \in N$, and each edge, $e \in E$. The attributes include the flow (F), n_F and e_F , and flow capacity (FC), n_{FC} and e_{FC} . Included also is resistance (weight) (W) of nodes n_W and edges e_W , buffering (B) n_B and buffer capacity (BC), n_{BC} , latency (time lag between event and response) (L), n_L , the length (D) of edges, e_D , and the stacking (queuing) (S) on edges, e_S . As well as these each node (n) can have a role (R) assigned to it. This is handled differently to the other attributes as this is a description of the function of a node in the network, and is not a quantitative value which can be used to model the behaviour of network assets.

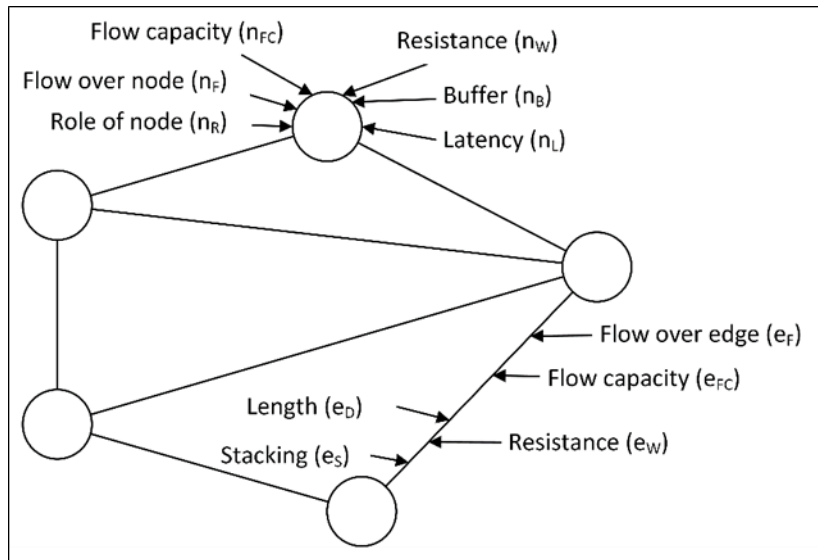


Figure 3.20: Network diagram showing node and edge attribution for modelling of flows over a network.

Attribute	Node (n), Edge (e)	Description
Flow (F)	n_F, e_F	The flow attribute (F) facilitates the modelling of flows over a network by allowing the flow level to be explicitly stored. The flow capacity (n_{FC}, e_{FC}) is stored as well, with the ability for a defined function to be used to adjust the capacity of the node/edge based on other attributes/properties.
Resistance (W)	n_W, e_W	The resistance (or weight) attribute (W) allows for the modelling of characteristics such as travel time over a node/edge which can then be used in the routing of flows for example. As with the flow attribute, functions can be used to alter the resistance value for each distinct node and edge based on other values, such as flow itself.
Latency (L)	n_L	The latency of a node (L), the time it takes for the node to react to an event, allows for the modelling systems which may be operating at a site represented by a node. Again, functions can be used to alter this value based on other values in the network, such as the flow.
Buffer (B)	n_B	Buffering at a node (B), the stock at the location, allows a node to continue to function once a disruption event has occurred. The amount as 'stock' as the node is stored as well as the capacity at the node, n_{BC} .
Role (R)	n_R	Each node is assigned a role (R), normally similar to supply, demand and intermediate, though these can be customised as well as added to.
Length (D)	e_D	The length of an edge (D) can be stored explicitly and used for weighting for the identification of the shortest paths through a network with respect to distance.
Stacking (S)	e_S	The stacking (queuing) attribute (S) allows for values on queues to pass over an edge to be modelled where appropriate.

Table 3.2: Node and edge attributes explicitly stored within the nx_pgnet_atts schema.

The attributes are modelled as part of the network, with each node and edge having its own sets of attributes, as well as function if required. However, traditional graph theory algorithms, such as those for finding the shortest path between nodes, do not consider capacities or weights for traversing nodes, only considering those on edges (Dolan and Aldous, 1993). Networks can possess node attributes which need to be considered when modelling flows over the networks, such as the capacity or the time to traverse a road junction. To facilitate this nodes can be converted into two nodes (Figure 3.21), an in and an out node, linking these with a single new edge and assigning it the node attributes, including the *id* of the node, while also making the network directed (Dolan and Aldous, 1993; Chen, 2003). This process is shown in Figure 3.21 for an example node v which has a capacity of six. Through this method existing algorithms can be utilised for the computation of flows over networks where node and edge attributes are both considered.

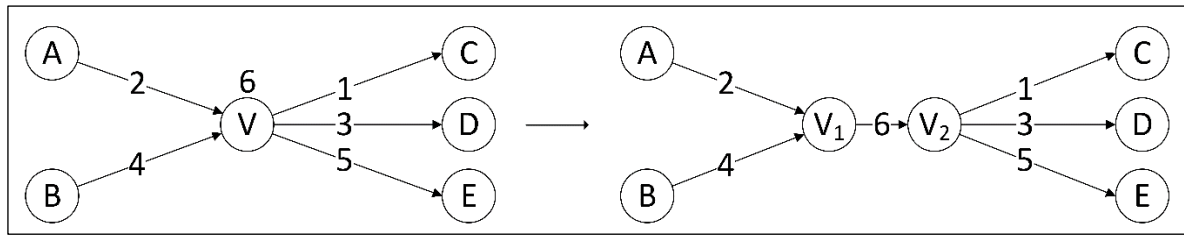


Figure 3.21: Method for modelling node capacities (Dolan and Aldous, 1993), with node and edge capacities shown.

3.9 Capacity constrained failure modelling

Cascading failures, occur where the failure of a small number of nodes or edges is propagated through the redistribution of flows leading to further nodes/edges failing as a result of being over capacity (Crucitti *et al.*, 2004a; Ash and Newth, 2007; Bao *et al.*, 2009a). Such failures have been observed within critical spatial infrastructures, causing disruptions to the service which they supply (Andersson *et al.*, 2005; Havlin *et al.*, 2010), such as in electricity distribution systems where such failures have been well documented (Ash and Newth, 2007; Rosas-Casals and Sole, 2011). A cascading failure is triggered by an initial single, or set of, failures (Crucitti *et al.*, 2004a), where these then trigger the re-distribution of flows (loads) caused by the re-routing of flows resulting from the initial failures. Following the redistribution of flows, some nodes or edges may be over capacity resulting in these failing and thus again a redistribution of flows, resulting in a cascading failure (Crucitti *et al.*, 2004a; Bao *et al.*, 2009a). The trigger for the cascading failure can be a range of causes, from the breakdown of a

component to a natural event which leads to the failure of a component, such as the flooding of a substation (Little, 2003), or as in Europe in 2003, the failure of transmission line due to contact with overhanging vegetation (Andersson *et al.*, 2005).

It has been shown through modelling that cascading failures from a single point of failure can result in complete breakdown of network function (Motter and Lai, 2002; Crucitti *et al.*, 2004a), or at least cause significant disruption to the network (Ash and Newth, 2007; Dueñas-Osorio and Vemuru, 2009; Xia *et al.*, 2010). This potential vulnerability to cascading failures requires networks to have properties which make them robust to such failures to avoid service disruption to users.

To investigate the robustness of different graph topologies and in particular hierarchical graphs, a flow-based capacity constrained cascading failure model is developed (Section 3.9.1), which allows the modelling of flows through synthetic and real-world critical spatial infrastructures networks with supply and demand nodes. The methods of triggering cascading failures are discussed in Section 3.9.2 and the methods employed for recording the results from the simulations using the developed capacity constrained failure model are presented in Section 3.9.3. The analysis and the scenarios investigated are presented in Section 3.9.4.

3.9.1 Developed failure model

A capacity constrained cascading failure model has been developed for modelling cascading failures over graphs (Figure 3.22), with the general approach to the model similar to previously developed models (Crucitti *et al.*, 2004a; Bao *et al.*, 2009a). The failure model allows for the modelling of flows through a graph between supply and demand nodes allowing the effect of a single failure on the graph to be examined.

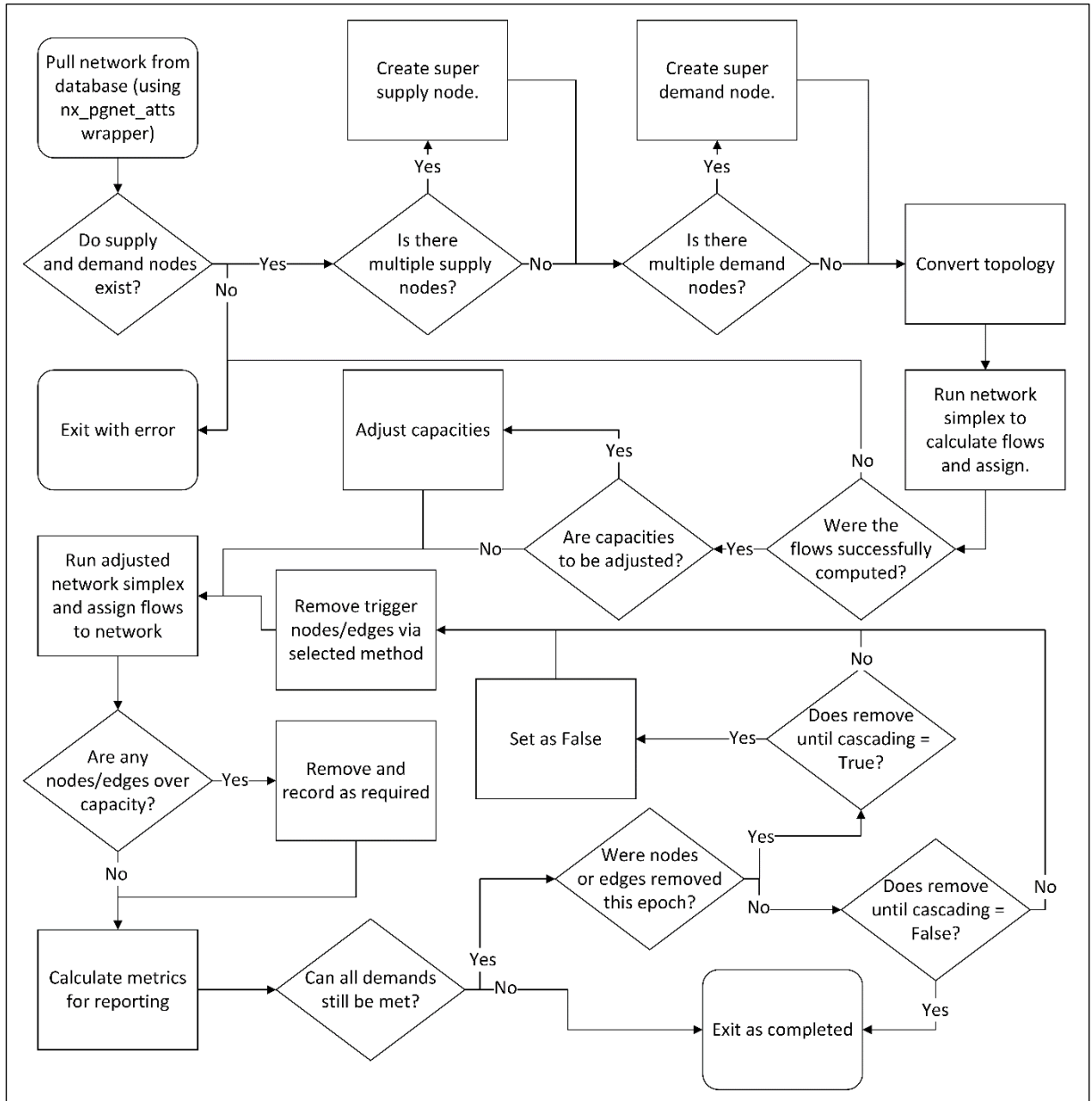


Figure 3.22: Developed capacity constrained cascading failure model.

For a graph, G , supply and demand nodes are defined with values assigned, where the sum of the demand is equal to the sum of the supply. The shortest path is then found between these nodes where the demand is met through the routing of flow(s) from the supply node(s). The supply and demand nodes are defined in an instance of the `nx_pgnat_atts` database schema (Section 3.11.2), where their role is set as appropriate (either as supply or demand). Alternatively, this can be specified when loading the graph from the database, as can the supply and demand values. To simplify the algorithmic approach to calculating flows over a network where multiple supply/demand nodes are used, $\{s_1, s_2, \dots, s_m\}$ and $\{d_1, d_2, \dots, d_m\}$ (Figure 3.23), super supply/demand nodes are created, s^* and d^* . Each has the accumulated sum

(supply/demand value) across the respective node sets, with the edges between these and the original supply/demand nodes ($s^* = \sum\{s_1, s_2, \dots, s_m\}$ and $d^* = \sum\{d_1, d_2, \dots, d_m\}$) having the capacity of the node they link to the super node. This approach reduces the computational complexity of computing flows from multiple supply nodes to multiple demand nodes to just between two nodes with capacities used to constrain the amount which can be supplied by each supply node and the level of demand at each demand node. This allows for existing algorithms for the modelling of flows over networks, such as the network_simplex algorithm (NetworkX, 2015), to be used without the need to develop the algorithms further.

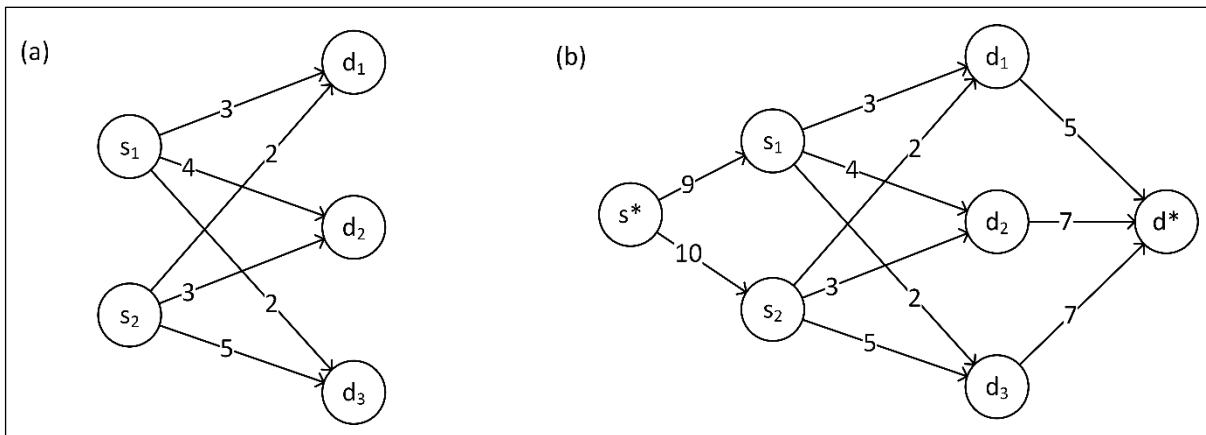


Figure 3.23: Super supply (s^*) and super demand nodes (d^*) (b) added to a network with multiple supply (s_x) and demand nodes (d_x) (a) (Dolan and Aldous, 1993).

The topology of the graph is then converted as detailed in Section 3.8 (page 69) with each node converted into an edge with the node attributes, allowing pre-existing algorithms which only considered edge attributes to use those of the nodes as well. The network simplex algorithm is then used, available in the NetworkX python library, which allows for solutions to flow problems to be found where the supply in the graph is equal to the demand, and where edge weights (the cost of traversing an edge) and capacities are considered in finding the minimum cost solution (NetworkX, 2015). However, where no solution can be found due to a lack of capacity an error is normally returned. Therefore the algorithm has been further developed for this work to return the graph with the flows which have been assigned (Figure 3.24(b)), along with the flows which could not be assigned and the nodes where the lack of capacity is between (edge (1, 4) in Figure 3.24(b)). This then allows the flows which could not be accommodated on the graph without nodes and/or edge capacities being exceeded to be assigned (Figure 3.24(c)).

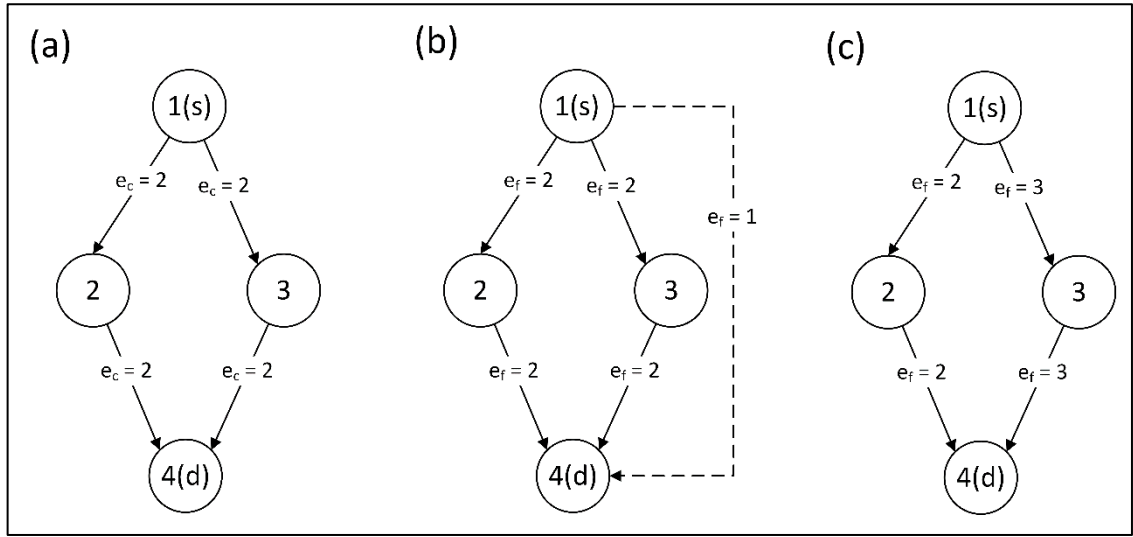


Figure 3.24: The routing of flows (b) to (c) where insufficient capacity is available between the supply node, 1, and the demand node, 4, where a flow of five is required from node 1 to meet the demand at node 4, where all edges have a capacity of 2 (a).

Where the initial calculation of flows is successful over the graph (G), the first step, $T(0)$ (Figure 3.25(a)), the effect of removing nodes/edges from the G can then be explored with the expectation of cascading failures being triggered. Following the removal of a trigger node/edge (Section 3.9.2), the flows are recalculated over G , $T(1)$ (Figure 3.25(b)), using the adapted network simplex algorithm to find routes for the flows through G from the supply to the demand node(s). From the solution for the routing of flows for $T(1)$, the first epoch of the simulation, any nodes over capacity ($n_F^{T(1)} > n_{FC}$) or edges over capacity ($e_F^{T(1)} > e_{FC}$), are deemed as failed and thus are removed from G , $T(2)$ (Figure 3.25(c)). The process of repeating the calculation of flows and the identification and removal of nodes/edges over capacity continues until no more node or edges are over capacity, G becomes completely disconnected, or there is no path from the supply to demand node at all (Figure 3.25(d)).

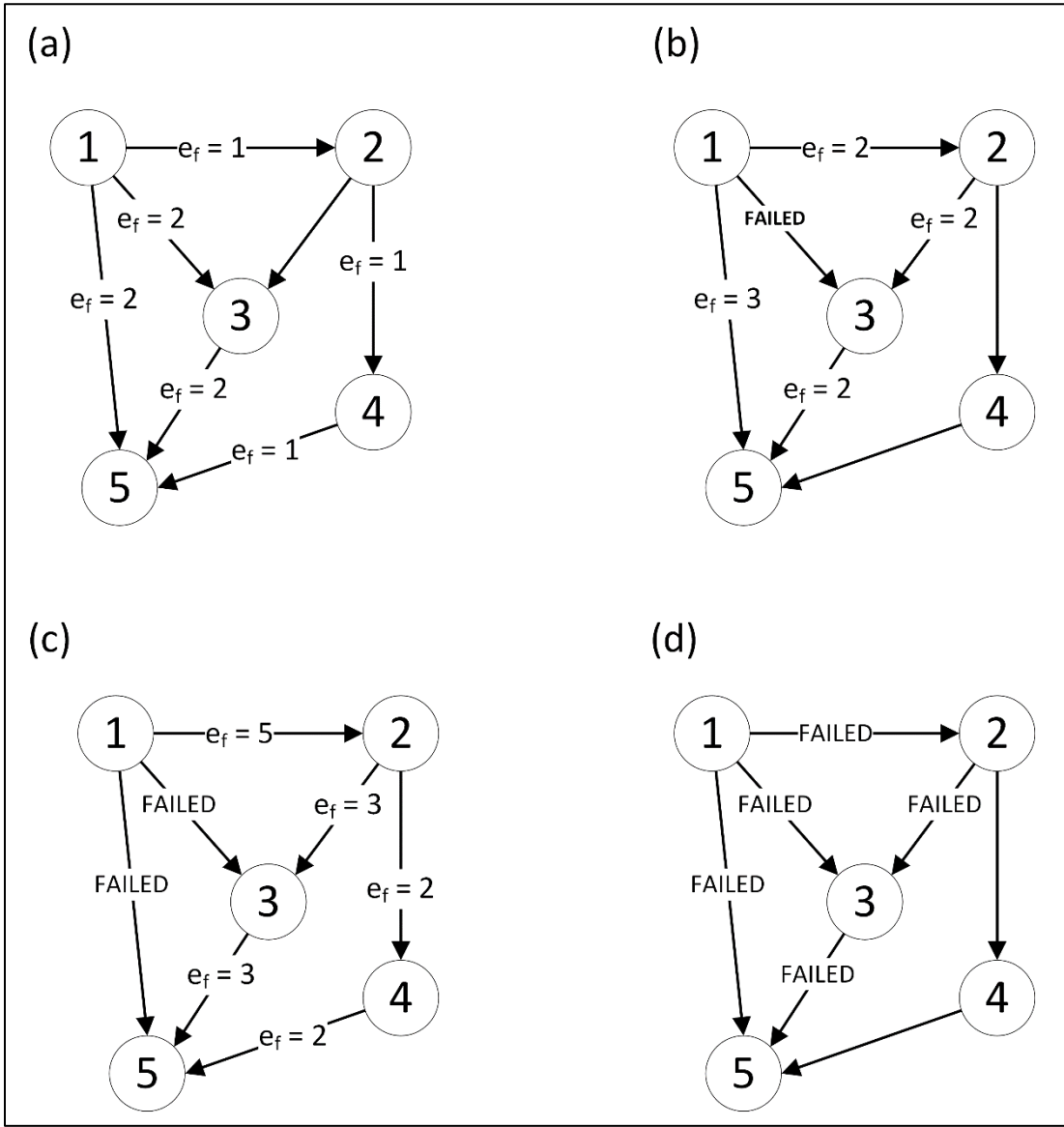


Figure 3.25: Capacity constrained cascading failure example where node 5 has a demand of five and node 1 has a supply of five, with each node and edge having a capacity of two. (a) shows the network at $T(0)$, (b) after the rerouting of flows after the removal of edge (1,3) as a trigger, (c) shows the failure of edge (1,5) as a result of being over capacity and the flow on each edge after the re-routing of flows again, and (d) shows the results of the failure of those edges over capacity, with no route from the supply node, node 1.

3.9.2 Triggering cascading failures

Following the initial calculation of flows $T(0)$ of a graph G , cascading failures are then simulated through the removal of trigger edges, Z (Bao *et al.*, 2009a). Trigger edges are removed from a network, with flows then re-routed giving with the potential for other nodes and/or edges to fail by then being over capacity ($n_F > n_{FC}$, $e_F > e_{FC}$) in an attempt to continue to route all flows from the supply to the demand nodes (Ash and Newth, 2007). This process of flow redistribution as described in Section 3.9.1, continues leading to failures to propagate

through G with the potential for the entire network to be affected. Therefore, the selection of the trigger edges affects the effectiveness of the resulting simulation and the size of the cascading failure which is triggered.

A number of methods can be used to select the trigger edges, from random selection (Crucitti *et al.*, 2004a), to topological methods such as node degree (Dueñas-Osorio and Vemuru, 2009) or betweenness (Crucitti *et al.*, 2004b; Dueñas-Osorio and Vemuru, 2009), as well as the initial load following the flow calculations ($F_{T(0)}^*$) (Bao *et al.*, 2009a; Wang and Rong, 2011) and those that fall within a defined spatial area. The random selection method, along with the degree and betweenness node selection methods have been previously explained in Section 3.5.1.

The success of the removal of trigger edges in causing cascading failures depends on those selected to be removed, the topological structure of the graph and the node and edge characteristics. Of the targeted approaches, betweenness (as defined in Section 3.4.2.1) or the initial flow load as methods are more successful in triggering cascading failures than the node degree selection method (Crucitti *et al.*, 2004a; Dueñas-Osorio and Vemuru, 2009). These targeted approaches remove the nodes/edges with the greatest flow and hence appear more effective at causing disruptions in the network as more flows have to be re-routed than compared to a degree based node removal. Despite the targeted methods, some networks may exhibit a robustness to a single failure of a node/edge with no cascading failure being triggered. However, multiple trigger nodes/edges can be removed to further explore the robustness to failures (Dueñas-Osorio and Vemuru, 2009), with the greater the number of trigger nodes/edges removed, the greater likelihood of a cascading failure being triggered as the capacity within the network will be reduced.

For the simulation of hazard events, such as flooding, trigger nodes/edges can also be those which lie within the effected geographic area. This method maybe less disruptive with regard to the cascading event triggered by the removal of nodes/edges based on flows (Dueñas-Osorio and Vemuru, 2009), but attempts to simulate events such as natural hazards where geographic areas are effected. Events such as floods and wind storms can hence be simulated and the robustness of networks to such failures examined.

3.9.3 Recording failure behaviour

The effect of the removal of trigger edges from a network are recorded through the impact these have on the network including the length of the cascading failure if one is triggered. For each simulation the reason for the simulation ending is recorded. If capacity is still available between

the supply to the demand nodes for the flow the network is said to be in equilibrium as the failure of the edge has had no effect on the ability of the flow to reach the demand node. The graph is regarded as failed where the failure of the trigger edge, or the subsequent failures caused by a cascading failure, leave the graph with no route between the supply and demand nodes, with the graph consisting of a number of components. The length of the cascading failure observed in each simulation is also recorded as is the number of epochs (T) the cascading failure lasted for until the network failed.

3.9.4 Analysis scenarios

A set of eight graphs, one from each of the eight synthetic graph models (Section 3.3) where $250 \leq N \leq 260$, is used to explore the robustness of the hierarchical and non-hierarchical graph models to the capacity constrained cascading failures. This helps to characterise the robustness of the graph models to facilitate the flow of a commodity/information/traffic over them (Motter and Lai, 2002; Ash and Newth, 2007) while being perturbed and their ability to withstand cascading failures, a reason for many failures experienced in real-world infrastructure networks (Andersson *et al.*, 2005; Havlin *et al.*, 2010). To explore the robustness of hierarchical networks to cascading failures six scenarios have been developed which explore the characteristics of the graph models to flow based failures. In each scenario five simulations are run with a single supply and a single demand node, both of which are randomly assigned to nodes in the graphs.

The first two scenarios, (i) and (ii) explore the ability for the graphs to accommodate flows and their susceptibility to cascading failures. Scenario (i) explores the robustness of the graphs through using a node and edge capacity equal to the supply/demand in the graph. A single trigger edge is removed following the initial computation of flows over the graph, as long as a single path between the supply and demand nodes exists the demand will always be met. Scenario (ii) uses the same parameterisation as scenario (i), but instead of a uniform node and edge capacity being used the capacities are assigned based on the graph structure (Table 3.3). This allows for the graph type and its structure to be explicitly considered in the assessment of the robustness of the graphs to cascading failures, providing a greater insight into the characteristics of each model.

Graph	Node/edge capacity
ER	Capacities (C) are assigned randomly to the nodes and edges due to the random topological structure of the graph model, where $1 \leq C \leq 2 \times \text{scenario demand}$. This allows the potential for sufficient capacity to exist between supply and demand nodes, while still matching the nature of the graph with the values assigned randomly.
GNM	Same as above.
WS	Capacities are assigned using the betweenness centrality (0-1) of the nodes and edges, and are thus correlated to the importance of the nodes given with those with the greatest betweenness having the greatest capacity, with $C = (2 \times \text{scenario demand}) \times \text{betweenness centrality}$, while $C \geq 1$. By assigning those edges with the greatest number of shortest paths passing through them with the greatest capacity, the hub nodes which are a feature of the WS model are assigned the greatest values, replicating the nature of the organisation of the graph.
BA	Same as above.
HR	Assigned based in the hierarchy of nodes and edges, where $1 \leq C \leq 2 \times \text{scenario demand}$. This results in those nodes and edges at the top of the hierarchy having the greatest capacity values, allowing for flows to pass through them, with the capacities decreasing approaching 1 as nodes and edges get further away from the top of the tree hierarchy.
HR+	Same as above (HR).
HC	Same as above (HR).
TREE	Same as above (HR).

Table 3.3: Details of the graph based assignment of node and edge capacities.

Scenarios (iii) and (iv) are used to explore the extent of the robustness of the hierarchical and non-hierarchical graph models to cascading failures. Trigger edges are removed until either a cascading failure is triggered or the graph fails with the supply and demand nodes no longer connected. Scenario (iii) is similarly parametrised to scenario (i) though multiple trigger edges are removed rather than just one. This allows the robustness of the graphs to be assessed to their ability to continue to supply the demand while being perturbed, with the more robust graphs expected to require a much large proportion of edges to be removed before the supply can no longer be routed to the demand node. Scenario (iv) uses the same parameterisation as (iii) with the exception that the capacities assigned to the nodes and edges are based on the graph type (Table 3.3), rather than using a uniform capacity. Through the removal of multiple trigger edges each graph will be perturbed until it fails or a cascading failure starts, and thus the extent of their robustness, measured through the fraction of edges removed as trigger edges, can be analysed.

The final two scenarios, (v) and (vi), explore the robustness of the graphs where the supply/demand is greater than the node and edge capacities. A uniform capacity for the nodes and edges is used with scenario (v) exploring the robustness to a single trigger edge and (vi) the

extent of the robustness through multiple trigger edges being removed until the graph fails or a cascading failure starts. These scenarios explore the ability of the graphs to accommodate a greater flow than in the previous scenarios, with the potential to provide greater insights into the robustness of the hierarchical and non-hierarchical graphs to cascading failures.

For two of the scenarios, (ii) and (iv), capacities have been assigned based on the structural organisation of the graphs. This assumes that the graphs behave and have characteristics which match their topological structure rather than the capacities being un-related to the topological structure. This assumption has a significant effect on how the graphs will respond to the perturbations during the flow modelling and simulation as the assigned capacities directly affect the assignment of flows onto the networks, thus affecting their ability to continue to function following perturbations.

3.10 Hierarchical flow robustness modelling

The hierarchical structure of graphs/networks, as discussed in Chapter 2 Section 2.3.3, results in a number of levels within the network topology (Gagneur *et al.*, 2003; Clauset *et al.*, 2008). Such levels may not all be equally important to the functioning of the network. Indeed, it can be hypothesised that one may expect the top-level of a hierarchical network to be the most important as it connected the network together (Barabasi *et al.*, 2003) and as such disruption at this level may have significant potential to decrease robustness. However, relatively little analysis has been undertaken to quantify the sensitivity of different forms of hierarchal networks to disruptions at different levels of their topological organisation. The robustness of a hierarchical network to failures at different levels of its hierarchy can be explored through examining the potential for flows to traverse the hierarchy, from the top level to the lowest level, such as might be required for the delivery of data in the internet or the transmission of electricity in the electricity transmission/distribution network, both of which are hierarchical (Pastor-Satorras *et al.*, 2004; Sanchez-Garcia *et al.*, 2014).

A failure model has been developed to explore the effects of failures within a hierarchical graph/network where flows can be modelled. This employs a generic approach analysing all levels of the hierarchy as part of the same network, rather than separating these into different networks and using a staged analysis approach or focusing on the transmission aspect only (Bompard *et al.*, 2009; Chang and Wu, 2011), therefore allowing the method to be applied across different infrastructure sectors. The model uses a list of supply/source nodes, and a second set of demand/sink nodes (Jungnickel, 2008; Chang and Wu, 2011). The failure analyses

for a given perturbation, be it a single node or edge failure, or a larger set of failures, checks if any paths remain available to all the demand nodes from the supply nodes. Where no path exists to a demand node, the node is regarded as failed (Figure 3.26).

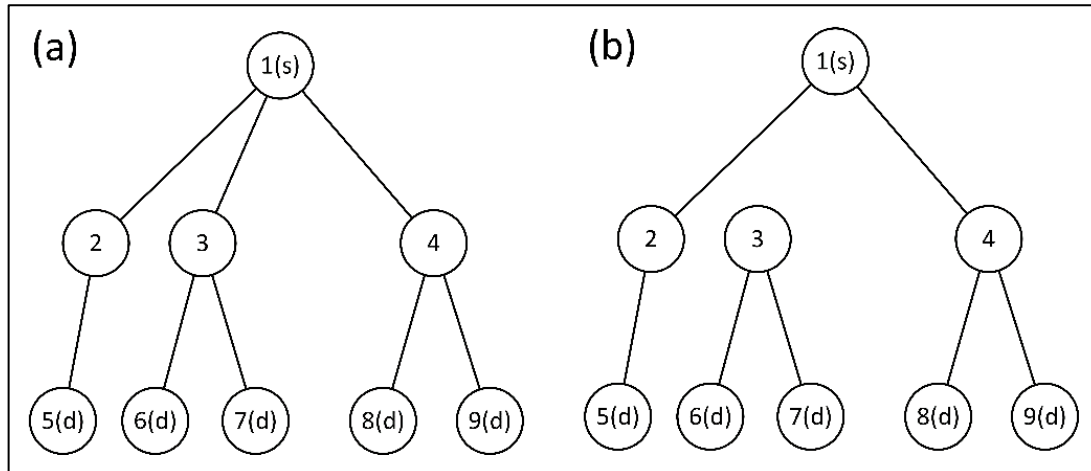


Figure 3.26: Hierarchical failure analysis example, where node 1 is the supply node and nodes 5,6,7,8 and 9 are demand nodes. When edge (1,3) has been removed (b), the failed nodes and edges are shown with dashed lines. Demand nodes 6 and 7 fail as they are no longer connected to the supply node, while node 3 also fails as this is also no longer connected to the network.

For the two sets of analysis undertaken using the developed failure model described above, Sections 3.10.1 and 3.10.2, the electricity transmission and distribution network for England and Wales is used (Figure 3.27), as provided by the Infrastructure Transitions Research Consortium (ITRC) (ITRC, 2013). Based on Ordnance Survey (OS) points of interest data, the network has been generated using this base data and mathematical modelling approaches to fill in the gaps in the network. This has created an attributed electricity transmission and distribution network down to 11Kv substations. The full network comprises of 170,669 substation nodes and 173,039 transmission or distribution cable edges.

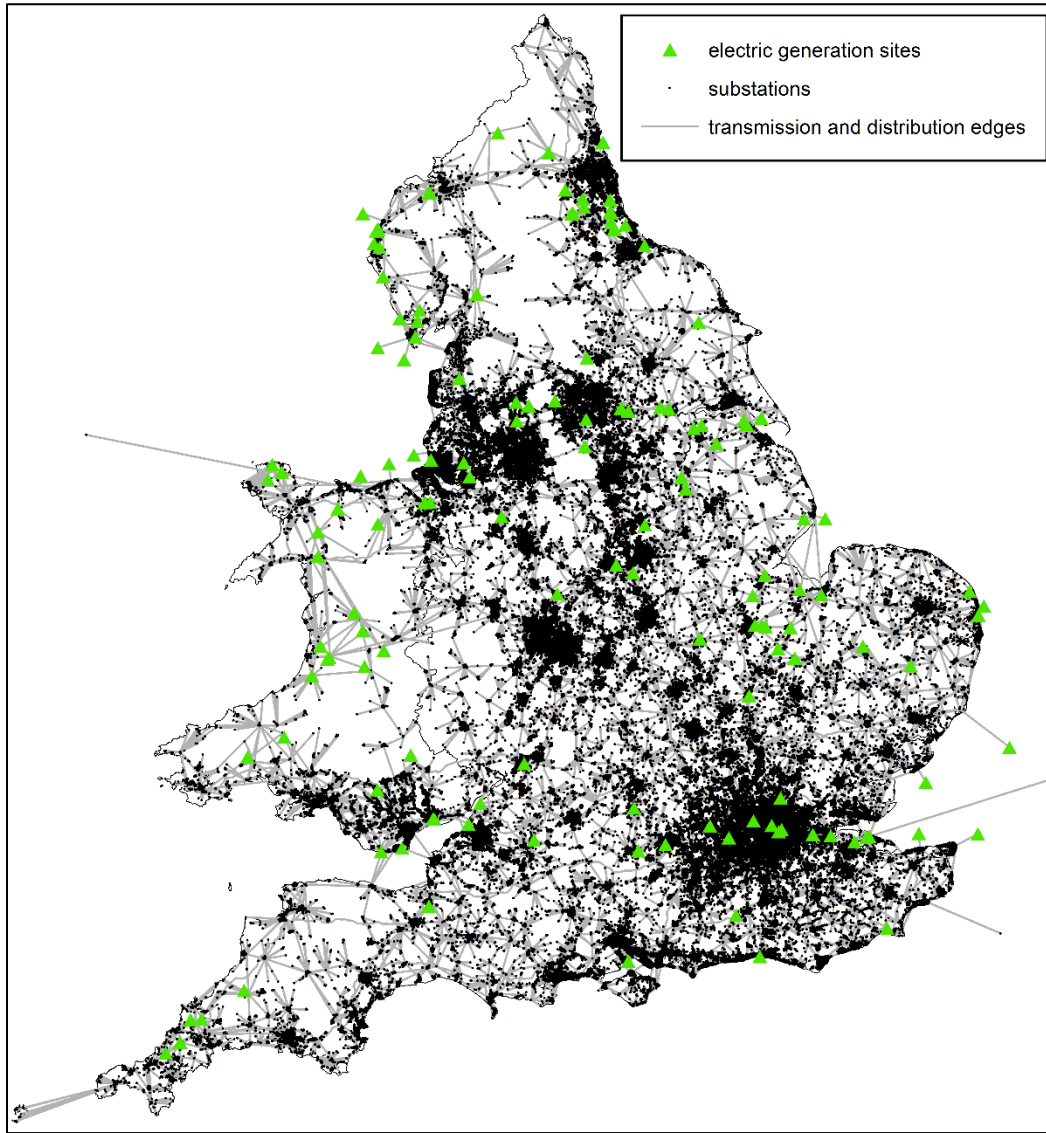


Figure 3.27: Electricity transmission and distribution network for England and Wales (ITRC, 2013).

3.10.1 Hierarchical connectivity modelling

The robustness of the electricity transmission and distribution network for England and Wales is used to examine the effect of removing the critical edges at the highest level of the hierarchy, those edges with a voltage of at least 400Kv, of which there are 323. This allows the redundancy in the network to be assessed in terms of whether paths still exist from the highest level of the transmission network the demand nodes, the 11Kv distribution substations. To achieve this each 400Kv edge, each possible pair of 400Kv edges and each possible combination of three 400Kv edges are removed to explore the effects on the ability for each of the 164,090 11Kv substations to connect to the 400Kv part of the transmission network. Due to the complexity of the analysis, and the large number of simulations and the checking of routes for over 164,000 nodes for each failure a maximum of three edges were removed at once.

3.10.2 *Spatial hazard modelling*

As identified in Chapter 2 infrastructure networks are exposed to a range of hazards including those which are explicitly spatial, including ice storms and flooding events (Little, 2003). It was highlighted that there has been little attention on the explicit geography of infrastructure networks when considering their robustness to perturbations (Barthelemy, 2011). This analysis will explore the robustness of the hierarchical electricity transmission and distribution network to different spatial hazard configurations, with each at first affecting a similar proportion of the network, 2% of node assets. It is presumed that any node asset falling within a hazard area fails, a first-order failure. The effect of these first-order failures on the network is then explored through the ability of the other substations in the network being able to connect through hierarchy to the 400Kv transmission part of the electricity network. Those which cannot are regarded as second-order failures.

To explore the robustness to different hazard footprint sizes and spatial distributions, three scenarios have been derived within which five different realisations are randomly generated and their robustness explored. Scenario (i) simulates single large events such as an isolated storm, as in the ice storm which hit Canada in 1998 (Chang *et al.*, 2007). Scenario set (ii) simulates four hazard areas, and as with (i) removes approximately 2% of nodes assets from the network, with these randomly distributed over the network. Scenario (iii) extends (ii) to use eight hazard areas simulating much smaller hazards which again are distributed randomly over the network.

The effect of the failures caused by the hazards in each simulation are recorded. For the first-order failures, the network assets which have failed as a direct result of the hazard area(s), are recorded in detail, with the number of substations, grouped by operating voltage, and the number of transmission and distribution lines, again grouped by voltage, which have failed recorded. The second-order failures are also recorded, including the counts of the network assets which have failed, again broken down by the operating voltage. However, with these the geographic distance of the failures from the hazard boundary are also recorded, with the average and maximum distances reported to measure the spread of the second-order failures from the hazard area(s). This can be used as an indicator of the robustness of the network, with the further failures are occurring away from the hazard area, the poorer the robustness of the hierarchical network to perturbations.

3.11 Software stack

3.11.1 Framework

A software framework (Figure 3.28) has been developed and employed for the analysis and modelling undertaken in this thesis to provide a consistent platform from which all work can be based. The framework is an extension of that developed and employed successfully by the ITRC (ITRC, 2013) for the analysis and simulation of the UK's national infrastructure networks, with the components from this highlighted in Figure 3.28. All other components shown have been developed for this work, extending the functionality of the ITRC framework.

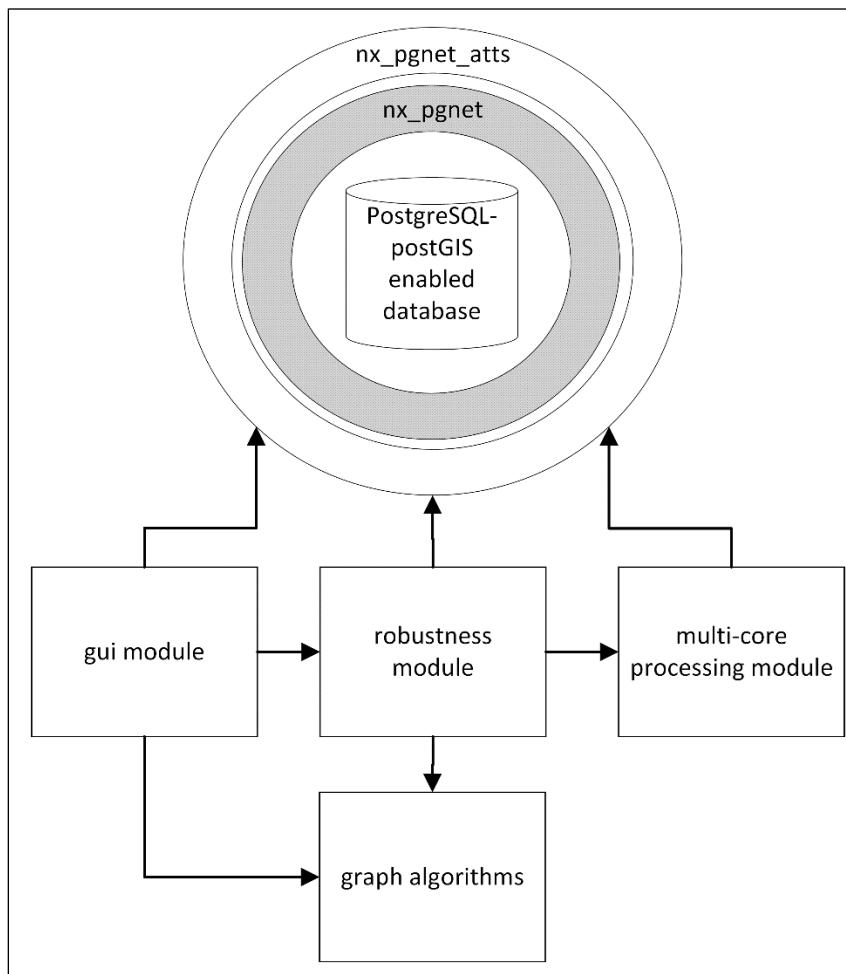


Figure 3.28: Developed software framework. Shaded features indicate those developed by the ITRC (ITRC, 2013).

Underpinning the framework is a central data repository, a PostgreSQL relational database with the spatial extension postGIS (Section 3.11.3). As well as facilitating the storage of data, the database allows for the easier management of data and results within a formal framework. This has been implemented with the ITRC developed 'nx_pgnet' database schema for the storage of networks, along with an extension of this 'nx_pgnet_atts' developed as part of this research for

the explicit handling of network attributes to support the modelling of flows on networks (Section 3.11.2). A further schema has been developed to manage the data and result sets from all the simulations undertaken in the work which is presented in this thesis (Section 3.11.3). PostgreSQL was chosen as the relational database management system (RDMS) due to the existing set of tools available and the extent of experiences with managing large network datasets with the software. Alternatives RDMS such as MySQL and Oracle both offer the required base functionality including their own spatial extensions, however the extent of experiences with these and the existing network handling tools makes PostgreSQL the preferred option.

Many components of the developed framework (Figure 3.28) have been written using the python programming language (Python Software Foundation, 2015). Python is a flexible language with a powerful and ever growing set of libraries being created by an active community (Aruoba and Fernández-Villaverde, 2015), providing users with the capability to perform a vast range of analysis. The language has full interoperability with PostgreSQL with the python psycopg2 PostgreSQL database connector, the most popular PostgreSQL adapter for python (psycopg, 2015), being employed to secure direct reading and writing to and from the database. The nx_pgnet database schema includes a wrapper, written in python, which manages the reading and writing of networks to and from the database. This was also extended through the developed nx_pgnet_atts wrapper. With the existing nx_pgnet wrapper already written in python and the extensive set of libraries available along with an existing familiarity with the language, python was viewed as the preferable language for code/software development for this research.

The analysis and processing undertaken in this research has been achieved by developing a set of modules (Figure 3.28). These include a module for robustness modelling (Section 3.11.4.2), a module for the developed graph models (Section 3.11.4.3), as well as a module for multi-core processing (Section 3.11.4.4). To compliment these a GUI, Graphical User Interface (Section 3.11.5), has also been developed with access to much of the functionality available including database access. This allows users to generate and save networks, calculate graph metrics over the networks, run failure simulations and visualise them.

A common denominator amongst all modules is the use of the NetworkX python library (NetworkX, 2014). The NetworkX package allows for the creation, analysis and visualisation of complex networks/systems through an extensive range of functions/algorithms. Alternative packages were considered, such as Network Workbench (NWB) (NWB Team, 2006), graph-tool (Peixoto, 2015) and igraph (igraph, 2016), all designed for the analysis of complex

networks. However, Network Workbench is itself a standalone application and thus is difficult to integrate into a framework and to develop custom methods. Both the graph-tool library and igraph libraries are similar to NetworkX, however offer a lower level of built-in functionality with regard to graph generating algorithms and analysis options. Despite NetworkX being slower than both of these libraries, the built in functionality and the ease of extending this functionality further has resulted on the implementation of this package for the research.

3.11.2 The network database schema

A network schema, nx_pgnet (Barr *et al.*, 2013) (Figure 3.29), has been used for the storage of networks in a PostgreSQL database. As described in the previous Section (3.11.3), the ‘Graphs’ relation (Figure 3.35) stores the key metadata for a network, including the ID which it can be referred to throughout the database, along with the network name (‘GraphName’) and the names of the tables which store its Node and Edge set (‘Nodes’ and ‘Edges’). Attributes for the nodes and edges are also stored in the respective relation, including the geometry for the nodes. For edges the geometry is stored in the ‘Edge_Geometry’ relation allowing edges with identical geometries to be stored while being part of the same network. The schema also has the capacity for storing edges which map dependencies as well as interdependencies between multiple network instances. A python wrapper allows networks to be read to and from the python programming environment facilitating the analysis of networks stored within the database. Networks can be imported through a selection of options including from shapefiles and from relations already within the database.

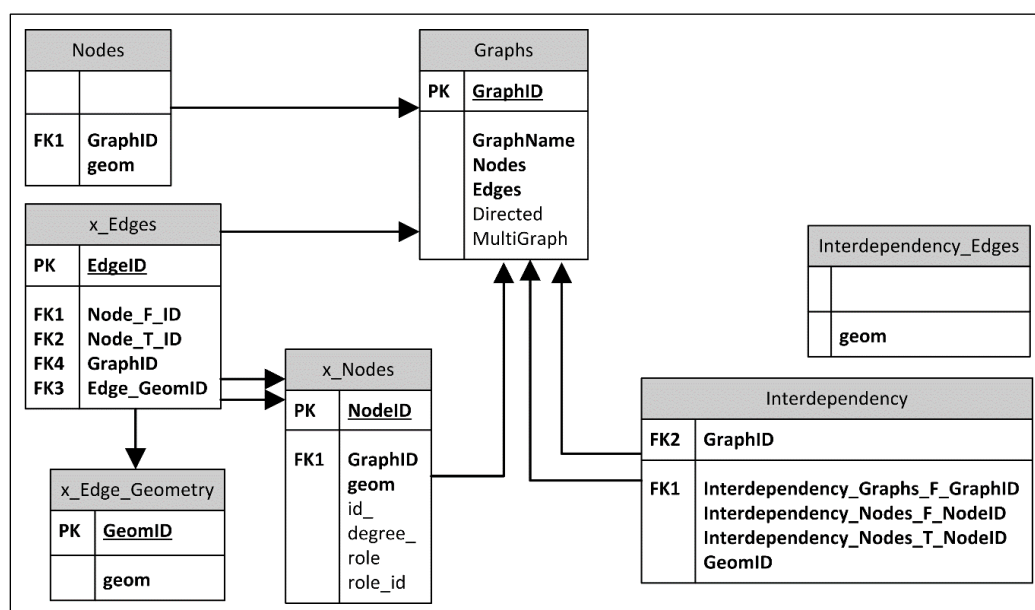


Figure 3.29: nx_pgnet schema example, where ‘x’ is the name of the network.

The nx_pgnet schema has in this research been extended to create nx_pgnet_atts, to allow for the flow based failure modelling as detailed in Section 3.9. The developments allow for the explicit representation of different node and edge types required for the modelling and analysis of flows and capacities within a network. The new schema results in the addition of 11 relations (Figure 3.30), including four new node type relations (flow, buffer, latency and resistance) and four edge attribute relations (flow, length, stacking and resistance) and generic relations for node function, node/edge units of measurement, along with node/edge analytical functions.

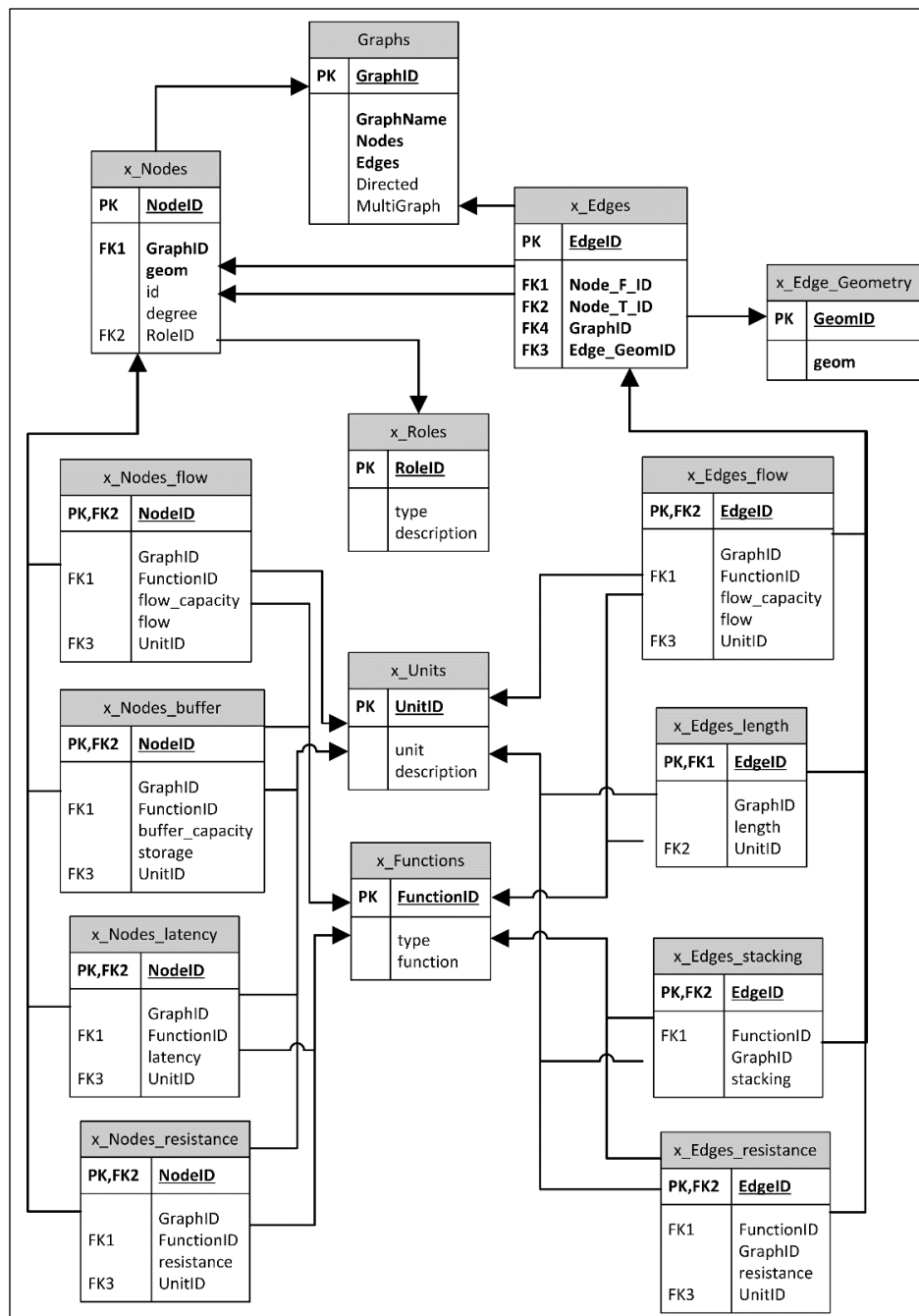


Figure 3.30: Enhanced nx_pgnet schema, where 'x' is the name of the network/graph.

The role of a node is assigned through a ‘RoleID’, stored in the nodes relation (Figure 3.31), which references, through a foreign key, the ‘Roles’ relation (Figure 3.32), which stores all roles of nodes and edges in a network. For flow based failure modelling (Section 3.9) these roles can be used to identify the supply and demand nodes in the network.

GraphID integer	geom geometry	NodeID bigint	id_ integer	degree integer	role_id integer
2123	01010000000000000000	1	0	9	1
2123	01010000000000000000	2	1	24	1
2123	01010000000000000000	3	2	27	1

Figure 3.31: Three rows from an example ‘Nodes’ relation.

RoleID [PK] integer	type character varying(255)	description character varying(255)
1	transfer	
2	supply	
3	demand	

Figure 3.32: Example of the ‘Roles’ relation.

For each attribute the ‘GraphID’ and the id’s of nodes/edges is stored with the attribute data, including the ‘FunctionID’, the attribute value, and capacity in the case of flow and buffer attributes, along with the ‘UnitID’ (Figure 3.33). The ‘FunctionID’ references the ‘Functions’ relation, as a foreign key, allowing functions describing the behaviour of the metric to be used in any analysis. The ‘UnitID’ references the ‘Units’ relation using a foreign key, which stores a list of all attribute units which can then be assigned to any of the attributes being used in the network.

GraphID integer	EdgeID [PK] bigint	FunctionID integer	flow_capacity double precision	flow double precision	UnitID integer
2123	1	0	9		1
2123	2	0	9		1
2123	3	0	6		1

Figure 3.33: Three rows from an example ‘Edges_flow’ relation.

The extended/enhanced nx_pgnnet network schema is accessed through the developed nx_pgnnet_atts wrapper, which allows the attribute data to be returned as part of the network, while still utilising the underlying nx_pgnnet wrapper functions. The wrapper allows networks to be loaded into python as well as written back to the database populating the schema while all constraints are adhered to. This is either done manually or by specifying the name of the network attributes which contain this information. Through a set of primary and foreign keys, as well as designed constraints, the schema is enforced ensuring data integrity and reducing the need for users to check the network manually for issues.

The schema is limited to only handling the defined attributes explicitly, with all others having to remain as part of the attributes for nodes and edges and therefore losing the ability to quickly assess and manipulate where required attribute values and the functions related to these. For those attributes handled explicitly, not all have to be used, and instead these can be chosen for each network with only those attribute tables required for the analysis being built when writing a network to the database through the nx_pgnnet_atts wrapper. Similarly not all attributes, and thus associated data, have to be read into the python environment, with these again specified when initiating the read process.

3.11.3 The analysis database schema

The developed software stack utilises a database for the storage of the complete suite of synthetic and real-world networks as well as for the results from the analysis and failure modelling. A schema, Figure 3.34, has been developed to allow the results to be stored while being intrinsically linked to the networks stored using the network database schema (Section 3.11.2). Results from the analysis of networks are linked directly through the ‘Graphs’ relation (Figure 3.35), which stores records for each network in the database with each having a unique id, the ‘GraphID’, allowing results from the analysis of a network to be linked to the network through a foreign key relationship to this. Further metadata on each network/graph is obtained from the ‘network_type_reference’ relation which records for each graph_id the ‘type_id’ of the model used to generate it, Figure 3.36 (left), of which details can be found in the ‘network_types’ relation (Figure 3.36, right) using the foreign key relationship between ‘type_id’ and ‘type’ in the ‘network_types’ relation.

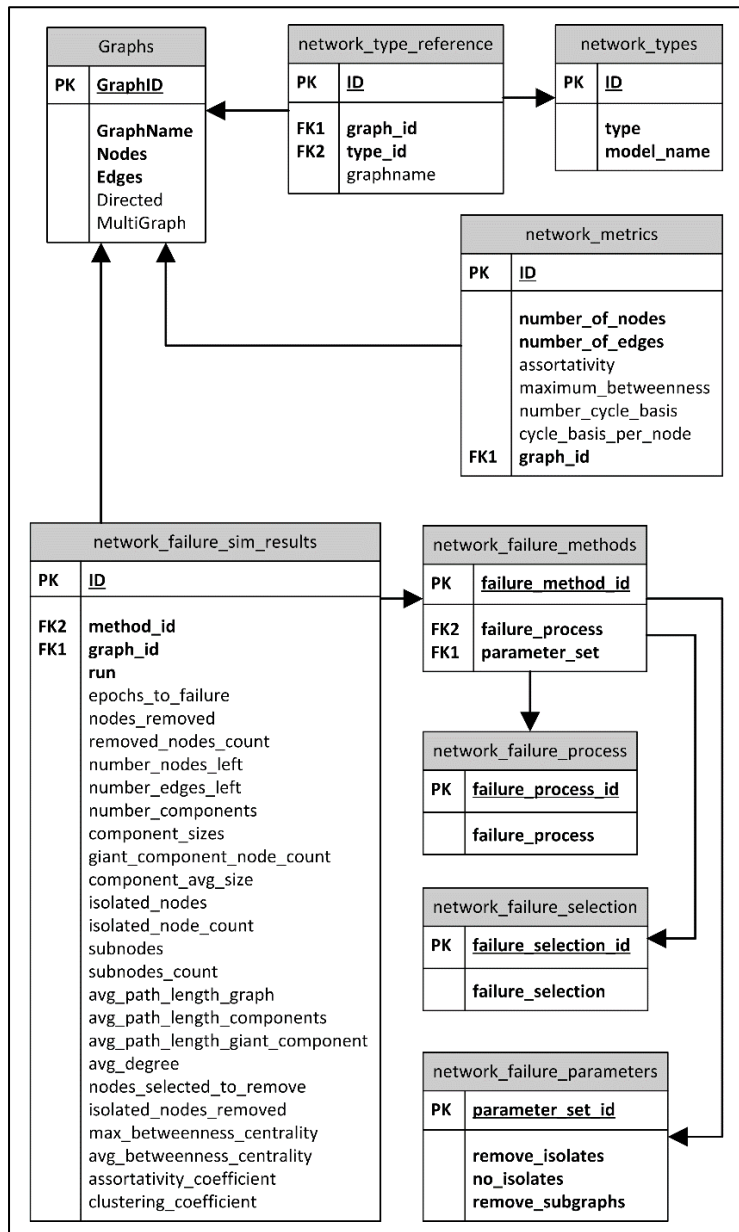


Figure 3.34: Database schema as employed for the research.

GraphID bigint	GraphName character varying	Nodes character varying	Edges character varying	Directed boolean	MultiGraph boolean
1	er0	er0 Nodes	er0 Edges	f	f
2	er1	er1 Nodes	er1 Edges	f	f
3	er2	er2 Nodes	er2 Edges	f	f

Figure 3.35: The top three rows from the ‘Graph’ table.

ID [PK] integer	graph_id integer	type_id integer	graphname text
1	1	1	er0
2	2	1	er1
3	3	1	er2

	ID [PK] integer	type text	model_name text
1	1	ER	Erdos-Renyi
2	2	GNM	GNM
3	3	BA	Barabasi-Albert

Figure 3.36: The top three rows from the ‘network_type_reference’ relation (left) and from the ‘network_types’ relation (right).

Two relations are used to store results, one for those from the analysis of the networks (Figure 3.37), and a second for storing the failure modelling results (Figure 3.38). These both use the ‘graph_id’ as a foreign key to reference the network used for the analysis. The ‘network_metrics’ table which stores results from the analysis of networks stores the calculated metric values allowing these to quickly be returned rather than being re-calculated when needed for subsequent analysis or reporting.

ID integer	graph_id integer	number_of_nodes integer	number_of_edges integer	assortativity double precision	maximum_betweenness double precision
1	1	950	5034	-0.0126778973	0.0095959821851
2	2	1244	13279	-0.00240265068	0.00416735588087
3	3	1669	19509	-0.0030038773	0.00285295818817

Figure 3.37: Top three rows from the ‘network_metrics’ table.

ID integer	method_id integer	graph_id integer	run integer	epochs_to_failure integer	nodes_removed text	removed_nodes_count text	number_nodes_left text	number_edges_left text
1	2	716	0	1059	[], [660], [1060]	0, 1, 2, 3, 4, 5, 6,	1151, 1150, 1149,	7266, 7255, 7240
2	2	716	1	1060	[], [416], [563]	0, 1, 2, 3, 4, 5, 6,	1151, 1150, 1149,	7266, 7257, 7243
3	2	716	2	1067	[], [1121], [330]	0, 1, 2, 3, 4, 5, 6,	1151, 1150, 1149,	7266, 7260, 7247

Figure 3.38: Top three rows of the ‘network_failure_sim_results’ relation with a sample set of result metrics shown.

The second results relation, ‘network_failure_sim_results’ (Figure 3.36), is used to record results from the failure simulations over networks. This is done through storing the results at the end of a simulation, with the details of the failure simulation, the parameterisation, stored in another relation (Figure 3.39), and referenced through a foreign key relationship between ‘method_id’ which references the ‘failure_method_id’ column. As a network can be analysed multiple times with the same failure method the ‘run’ field is used to track each run, with this forming part of a unique constraint, with the ‘method_id’ and ‘graph_id’, to ensure each row is unique.

failure_process integer	selection_process integer	failure_method_id [PK] serial	parameter_set integer
2	1	2	4
2	2	3	4
2	3	4	4

Figure 3.39: First three rows of the ‘network_failure_methods’ relation.

Three parameters are used to define a failure method as shown in Figure 3.39 and thus each unique set of these values (detailed in a further three relations – ‘network_failure_process’, ‘network_failure_selection’ and ‘network_failure_parameters’, Figure 3.40) has a unique failure_method_id which allows results to be searched for based on the failure method used.

failure_process text	failure_process_id integer	failure_selection text	failure_selection_id integer
single	1	random	1
sequential	2	degree	2

parameter_set_id integer	remove_isolates boolean	no_isolates boolean	remove_subgraphs boolean
4	t	t	f
5	t	f	f

Figure 3.40: Two rows from the three failure methods tables – Top left: ‘network_failure_process’. Top right: ‘network_failure_selection’. Bottom: ‘network_failure_parameters’.

3.11.4 Developed modules

As shown in Figure 3.28 and briefly discussed in Section 3.11.1, there are three key modules which form part of this research as well as the nx_pgnet_atts wrapper for the developed database schema (Section 3.11.2). Each of these modules are described in the following subsections, with details provided on their use and purpose with the developed integrated software framework.

3.11.4.1 nx_pgnet_atts wrapper

The developed and employed nx_pgnet_atts wrapper, facilitates reading and writing of networks to and from a PostgreSQL database using the nx_pgnet_atts database schema. The wrapper adds functionality to the existing nx_pgnet wrapper so that the extended schema, Section 3.11.2, can be used. The nx_pgnet_atts wrapper has two main functions, Table 3.4. Example uses and further details are given in Appendix C, with the developed documentation in Appendix F.

Function	Input parameters	Description
read_from_db	<i>database connection:</i> An open OGR connection to the database.	The <i>read_from_db</i> function allows a network to be read from a database using the nx_pgnet_atts schema given an open database connection (<i>database connection</i> variable), a network exists with the given name (the <i>network name</i> variable) and the <i>attributes</i> variable (a dictionary of attributes) is correctly set.
	<i>network name:</i> The name of the network as stored in the database.	
	<i>attributes:</i> The attributes to be returned which have been stored using the schema representation.	
write_to_db	<i>database connection:</i> An open OGR connection to the database.	To write a network to a database using the nx_pgnet_atts schema, the <i>write_to_db</i> function can be used given the correct set of parameters/variables. An open database connection must be provided (<i>database connection</i> variable) along with a name for the network (<i>network name</i> variable) and the network itself (<i>network</i> variable). Other details must be provided such as the attributes which are to be stored explicitly (<i>attributes</i>) and if the network contains the attribute values and functions as attributes.
	<i>network name:</i> A name for the network for the database.	
	<i>the network:</i> The actual network as a NetworkX graph.	
	<i>attributes:</i> The attributes to be stored in schema.	
	<i>contains atts:</i> Does the attributes of the nodes and edges contain the attribute values to be stored in the database as defined by <i>attributes</i> parameter.	
	<i>contains functions:</i> Are the functions for the attributes stored as attributes in the network.	
	<i>overwrite:</i> If to overwrite a network with the same name in the database	
	<i>srid:</i> The spatial reference (coordinate) id, e.g. 27700 for British National Grid. -1 if a spatial.	
	<i>directed:</i> Is the network directed.	
	<i>multigraph:</i> Is the network a multigraph (multiple edges between the same pair of nodes).	

Table 3.4: Key functions within the nx_pgnet_atts wrapper.

3.11.4.2 Robustness

The robustness module contains the methods developed for both the topological and flow based failure modelling as set out in earlier Sections (3.5 and 3.6 respectively), as well as some extra functionality, for which details can be found in Appendix C. Some of the key functions, the two ways in which failure simulations can be run, are described in Table 3.5.

Function	Input parameters	Description
main	<i>network A:</i> Name of network A.	This allows for a full failure analysis to be run, with the steps automated. The failure analysis to be run is defined within the <i>failure parameters</i> variable (dictionary). The <i>when to calculate metrics</i> allows for metrics to be calculated at equal intervals rather than at every epoch. The <i>view failure</i> metric is used when the function is called through the GUI (Section 3.11.5) where failure simulations can be viewed if this is set to True.
	<i>network B:</i> Name of second network for dependency analysis.	
	<i>failure parameters:</i> Parameters for the failure method including node selection method.	
	<i>log file path:</i> Location of log file.	
	<i>view failure:</i> Used in GUI where the failure can be viewed.	
	<i>when to calculate metrics:</i> How often to calculate metrics.	
	<i>failures to occur:</i> List of epochs where failures will occur.	
step	<i>graph parameters:</i> Parameters specific for the network e.g. is it directed.	This allows a user to run the failure analysis step by step, with each call of this function running one failure epoch. This is also called by the <i>main</i> function. The details of the networks are stored in the <i>graph parameters</i> variable, the parameters for the analysis in the <i>parameters</i> variable and the metrics to be calculated are held within the <i>metrics</i> variable
	<i>parameters:</i> General analysis parameters including failure parameter	
	<i>metrics:</i> List of metrics and whether they are to be calculated.	
	<i>iterate:</i> If another epoch of analysis is needed.	
	<i>log file path:</i> Location of log file.	
	<i>when to calculate metrics:</i> How often to calculate metrics.	
	<i>failures to occur:</i> List of epochs where failures will occur.	
	<i>node to fail list:</i> List of dependent nodes to fail during dependency analysis.	

Table 3.5: Key functions in the developed robustness module.

Apart from allowing for failure simulations to be run, a host of other functions also exist for the manipulation of networks for the capacity constrained failure model (Section 3.9). The key functions include those for creating super supply and demand nodes and converting the topology to the enhanced representation, Table 3.6.

Function	Input parameters	Description
create_superdemand_node/ create_supersupply_node	<i>network name:</i> The name of the network as stored in the database	Allow the creation of a super supply node and super demand node as required for developed supply and demand modelling methods. Requires the network, the appropriate nodes, and the list of nodes/edges added if the topology has been converted (below).
	<i>demand nodes/supply nodes:</i> The list of nodes for which a super node will be created	
	<i>added edges:</i> A list of the edges added in the conversion to the enhanced network representation (Section 3.8)	
	<i>added nodes:</i> A list of the nodes added in the conversion to the enhanced network representation (Section 3.8)	
convert_topo	<i>network name:</i> The name of the network.	Converts the topology creating a directed network where a node is replaced with an edge (and the subsequent two nodes). Requires the network only.
revert_topo	<i>network name:</i> The name of the network.	Converts a network back to the original topology, reversing the actions performed in the <i>convert_topo</i> function above.

Table 3.6: Key functions available for network manipulation for failure simulations.

The module can be used as a stand-alone entity, though is dependent on the NetworkX library. However, it can also be used in conjunction with the multi-core processing module (Section 3.11.4.4), allowing multiple simulations to run simultaneously. Further to this, it can also be accessed through the developed GUI (Section 3.11.5) allowing users to access the functionality

of the robustness methods using a graphical interface to set the inputs and parameters for the desired analysis. Further documentation for this module is available in Appendix C.

3.11.4.3 *Graph algorithms*

As described in an earlier section (Section 3.3) three graph generation algorithms have been developed for the generation of the suite of synthetic networks. Algorithms have been developed for three graph types named, hierarchical random, hierarchical random+ and hierarchical communities. These are contained within the same module which has been used throughout the research ensuring consistency in the generation of these networks, with each being easily called/used (Table 3.7). More details of the algorithms themselves can be found in Appendix A. All other graph algorithms/generators used are found within the NetworkX package and thus are not included in this module.

Function	Input parameters	Description
hr	<i>number of levels:</i> The number of levels in the tree graph.	Through calling the <i>hr</i> function, hierarchical networks are generated.
	<i>nodes per branch:</i> The number of branches per parent node in the tree graph.	
	<i>probability:</i> Use to calculate the number of edges to add to the tree graph.	
ahr	<i>number of levels:</i> The number of levels in the tree graph.	The <i>ahr</i> function allows the generation of networks using the HR+ model.
	<i>nodes per branch:</i> The number of branches per parent node in the tree graph.	
	<i>probability:</i> Use to calculate the number of edges to add to the tree graph.	
hc	<i>size of cluster:</i> The number of nodes per community.	The <i>hc</i> function allows the generation of hierarchical community networks with a cluster size of three or four, with up to 4 levels.
	<i>number of levels:</i> The number of levels of communities in the graph.	

Table 3.7: Key functions in the graph algorithms module.

3.11.4.4 *Multi-processing*

The multi-core processing module has been developed to allow modern computing power/technologies to be better utilised and multiple processes performed at once. This was developed to allow multiple failure simulations to be run simultaneously, thus speeding up the processing of this aspect of the research. The module utilises the python multiprocessing library which manages the processes allowing for multiple threads to be used simultaneously.

3.11.5 *Graphical User Interface (GUI)*

The developed graphical user interface (GUI) provides a graphical interface (Figure 3.41), from which nearly all other modules (currently it does not support use of the multi-core processing module) are accessible, allowing users to explore the field of complex networks from the generation through to the visualisation of networks, including metric computation and robustness analysis. A developed short user guide is given in Appendix G.

The interface has been developed using PyQt4 (River Bank Computing, 2013), a python library used for interface design, the interface allows a user to perform a variety of tasks and analysis, summarised below.

- Generation of networks using existing algorithms (8 available), from csv files, from manual entry of node and edge sets, or from a database using the nx_pgnnet or nx_pgnnet_atts schema (detailed in Sections 3.11.3) (Figure 3.42). All networks can then be exported in a number of formats as well.
- Metric analysis of generated networks from a selection of over 20 metrics exploiting the NetworkX python library.
- Robustness simulations using topological failure methods of networks for a single network, where networks have dependencies or interdependencies (Figure 3.43).
- Visualisation of static networks (Figure 3.44), with the ability to render components based on metric values in a number of layouts (Figure 3.45), including geographic if the network has been imported via a database connection.
- Live visualisations of failure simulations with the ability to step through and pause.

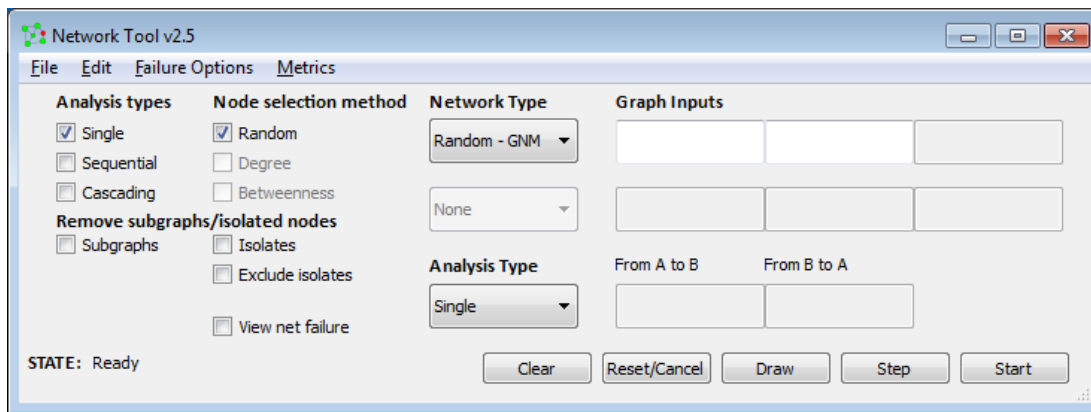


Figure 3.41: GUI interface.

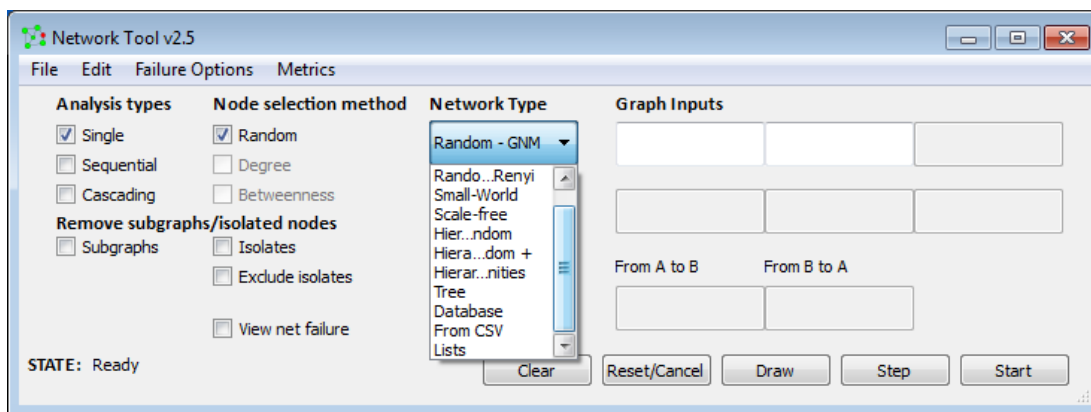


Figure 3.42: The list of available graph generation methods in the GUI.

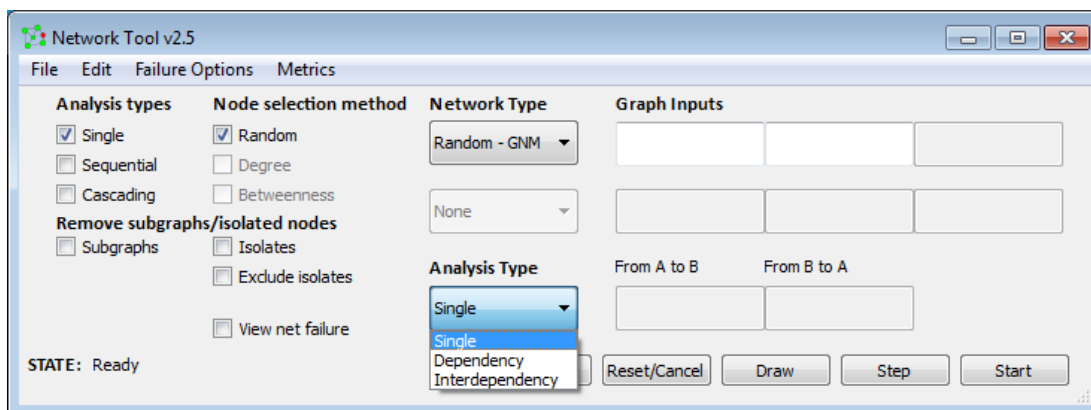


Figure 3.43: Showing the range of failure analysis methods available.

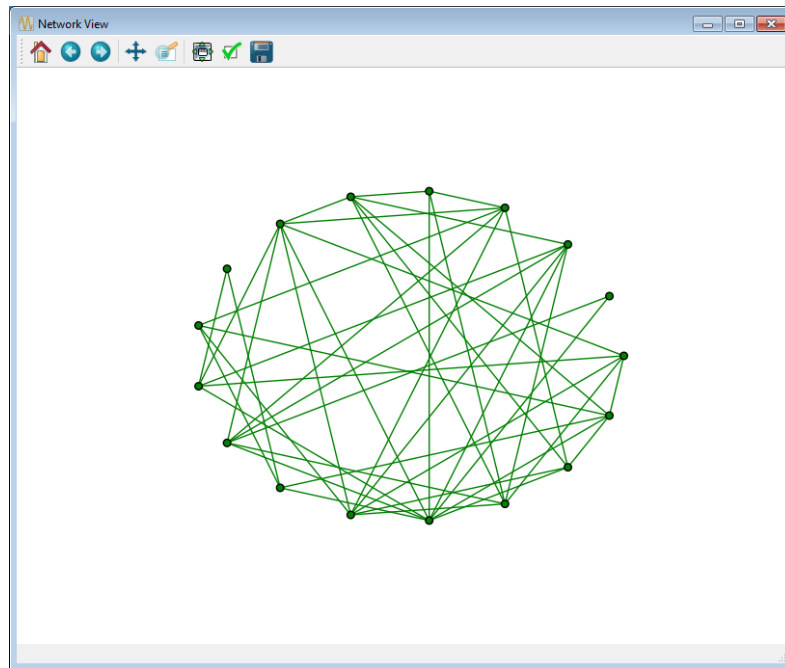


Figure 3.44: Example visualisation of a network as generated through the GUI.

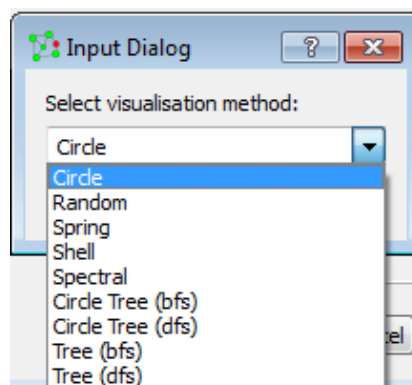


Figure 3.45: Layout options for visualising networks.

3.12 Conclusions

This chapter has presented and detailed the methodology designed to allow the aims and objectives as set out in Chapter 1, Section 1.2 (page 4) to be addressed. The generation of the suite of synthetic graphs has been detailed with the algorithms and the parameterisation of these specified. The suite of critical spatial infrastructure networks has also been presented including details on the generation of these networks from available spatial data.

For the characterisation of the suite of synthetic graphs methods have been presented which will allow the statistical analysis of the eight synthetic graph models and leading to the recognition of the key characteristic differences between hierarchical and non-hierarchical

graphs. The methods which will be used to explore the robustness of the synthetic graph models have then been detailed including the developed topological failure model. The application of this will allow for the comparison between the response of the non-hierarchical and hierarchical graphs. The application of these same methods, the characterisation and topological failure modelling, on the suite of spatial infrastructure networks has also been discussed, with the ability to start to recognise the infrastructure networks which are hierarchical and those which are not an outcome of the methods presented.

To further explore the robustness of hierarchical graphs, a capacity constrained failure model is detailed along with the developed scenarios which will be used to investigate the robustness of the hierarchical graphs to flow based cascading failures. Methods have also been presented which will assess how a hierarchical infrastructure network will respond to hierarchically targeted failures, through the removal of network assets in the highest level of its hierarchy using a number of scenarios. The robustness of a hierarchical infrastructure network to spatial failures is then detailed with a number of scenarios developed which explore the effect of varying configurations of hazards have on the function of the network.

Finally, the employed software stack is detailed including the database framework and the developed software modules which allow the analysis detailed in the methods to be undertaken within a single environment. The integrated database framework ensures all generated graphs and networks are stored in a single easy to access location along with the results from the analysis of these.

Chapter 4: Results

4.1 Introduction

This chapter presents the results from the analysis performed to address the aims of this research outlined in Chapter 1 using the methods described and presented in Chapter 3. The chapter is split into four distinct sections with each presenting results that address the aim and objectives of the research. The first section, Section, 4.1, presents the results aimed at identifying the structural characteristics of hierarchical graphs, including their behaviour and response to perturbations and how this compares to other graph topologies. Section 4.3 compares the structural characteristics of hierarchical graphs and those of a suite of critical spatial infrastructure networks in order to ascertain if any exhibit hierarchical characteristics. Section 4.4 extends the robustness analysis performed on the different network models in Section 4.1 to explore the robustness of the different graph models to capacity constrained cascading failures, thus investigating the difference in response between those which are hierarchical and non-hierarchical. The robustness of hierarchical infrastructure networks is then investigated in more detail using the electricity transmission and distribution network for England and Wales as a case study in Section 4.5. The robustness to simultaneous failures in the highest level of the network hierarchy is explored, as well as the vulnerability of such a network to different spatial hazard scenarios.

4.2 Characteristics of hierarchical graphs

4.2.1 Introduction

Using the suite of graph models developed in Chapter 3, Section 3.3, the characteristics of the eight different graph models have been analysed as described in Chapter 3 Section 3.4 in order to evaluate whether hierarchical graphs exhibit distinctive topological structure and organisation. The results are presented across five Sections, with the first, Section 4.2.2, comparing the degree distributions using a subset of the full suite of synthetic graphs. Using a selected set of graph metrics (Chapter 3, Section 3.4.2, page 53) (assortativity coefficient, maximum betweenness centrality and number of cycle basis per node), a metric based analysis of the full suite of graphs is performed to improve our understanding on the topological characteristics of the graphs. In particular, Section 4.2.4 presents the results of a multivariate statistical analysis of the metric values returned for each graph model using the transformed divergence similarity measure. Following a metric based assessment of the different models a

failure analysis is undertaken in Section 4.2.5 in order to investigate the robustness of the different graph models to perturbations, looking at the number of failures required for a null graph to form. The final section, Section 4.2.6, presents in more detail how each of the models behave when perturbed through an analysis of the how the graphs fragment into components, providing further insights into the robustness of each of the eight synthetic graph models.

4.2.2 Graph degree distributions

The degree distribution of a graph describes its topological structure through the probability of selecting a node at random with a particular degree (Newman, 2003b). This has been calculated for the all graphs within the synthetic suite allowing these to be compared to explore the difference in topological structure between each graph model. However, due to the nature of degree distributions, a single example from each graph model is shown in Figure 4.1, with six example plots for each shown in Appendix D Section D.1. Most strikingly there are clear differences between the random, scale-free and small-world models (ER, GNM, BA and WS) and the suite of hierarchal models generated (HR, HR+, HC and TREE) (Figure 4.1). The random graph models, ER and GNM, exhibit the expected bell-curve normal distribution while the BA model generates a scale-free degree distribution with the proportion of nodes with a high degree decreasing gradually. The WS model exhibits a similar distribution, though with a much smaller maximum degree resulting in the tail being much shorter.

However, the four hierarchal models exhibit strikingly different degree distributions particularly in term of exhibiting a series of discrete peaks in terms of their degree distribution. This is clearly demonstrated by the plot for the TREE model where there are two peaks with no other degrees represented in the plot. Similarly, a number of clear peaks can be seen in the plots for the HR and HR+ plots; however, the distributions also show the existence of other node degrees although at much lower probabilities between these peaks.

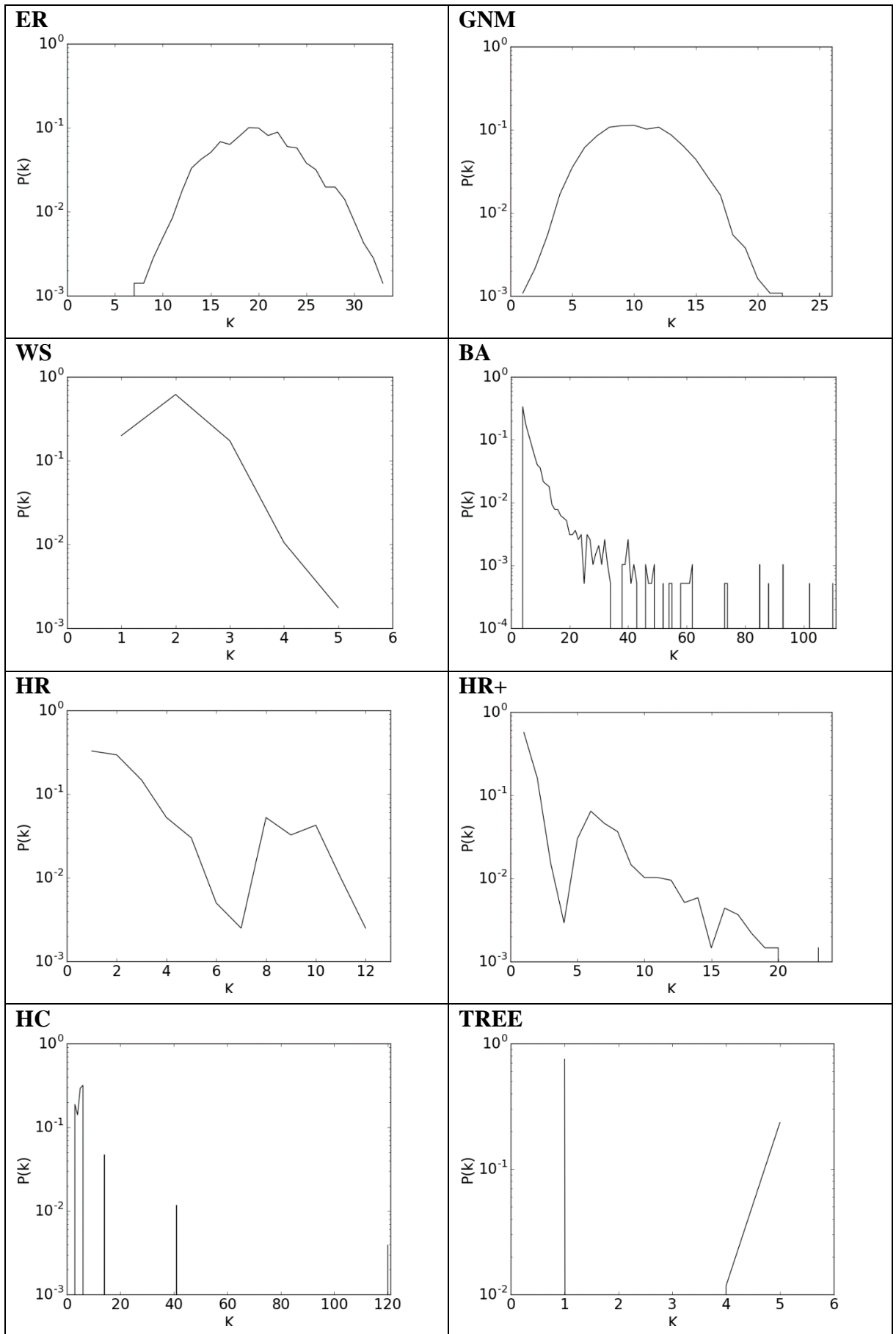


Figure 4.1: Example degree distribution plots for the eight graph models.

The clear differences between the standard graph models (ER, GNM, WS and BA) and the four hierarchical (HR, HR+, HC and TREE), implies that the presence of a hierarchical structure in an infrastructure network may have a significant effect on its topological structure. The presence of a small number of discrete peaks in the degree distributions of the hierarchical models, as seen in the plots for the HR, HR+, HC and TREE graph models (Figure 4.1) means that these models are comprised of a few key nodes with high node degrees (connectivity); this may make them more sensitive to perturbations particularly when focused on these key nodes. This is investigated in detail in Section 4.2.5.

4.2.3 *Assessment of graph metrics*

Figure 4.2 shows the multivariate distribution (mean and standard deviational ellipse) for the calculation of the maximum betweenness centrality (MBC) and the assortativity coefficient (AC) for the full suite of synthetic graph models (Table 4.1), (for more details see Chapter 3, Section 3.3, page 42). Detailed plots for each graph model showing all graphs are given in Appendix D Section D.2. It is evident in Figure 4.2 that the TREE and HC (hierarchical community) graphs exhibit different characteristics compared to the non-hierarchical graph types; which are clustered around the origin of the plot (0.0,0.0). The TREE model has a much higher MBC, 0.76, as expected (Newman, 2002), compared to a mean of 12.86 for the ER, GNM, WS and BA combined (Table 4.2). The HC graph type exhibits AC values much closer to zero, -0.183, rather than the -0.65 for the TREE graphs, which is similar to those associated with the non-hierarchical graph types, though it has MBC values similar to the TREE model, 0.77. This shows that the HC model, although hierarchical, also shares some properties with the non-hierarchical graphs suggesting a greater similarity with the BA and WS graph models which have been strongly associated with infrastructure networks previously (Albert and Barabasi, 2002; Newman, 2003b). The two in-house models, HR and HR+, show a greater similarity with the non-hierarchical graph types than the hierarchical graph types (Figure 4.2), despite their origins being the TREE model. Both models have MBC values, 0.25 and 0.31 (Table 4.2), more closely related to those for the WS and BA graphs, 0.03 and 0.08. This highlights the effects of the extra edges which have been added to create a better connected network as these result in more paths between previously disparate parts of the graphs and thus avoids shortest paths having to pass through the same critical nodes and hence results in a lower MBC.

Graph model	Metric analysis	Failure analysis
ER	1000	500
GNM	1000	500
WS	1000	500
BA	1000	500
HR	1000	500
HR+	1000	500
HC	7	7
TREE	31	31

Table 4.1: Number of exemplars for the eight graph types used in employed analysis.

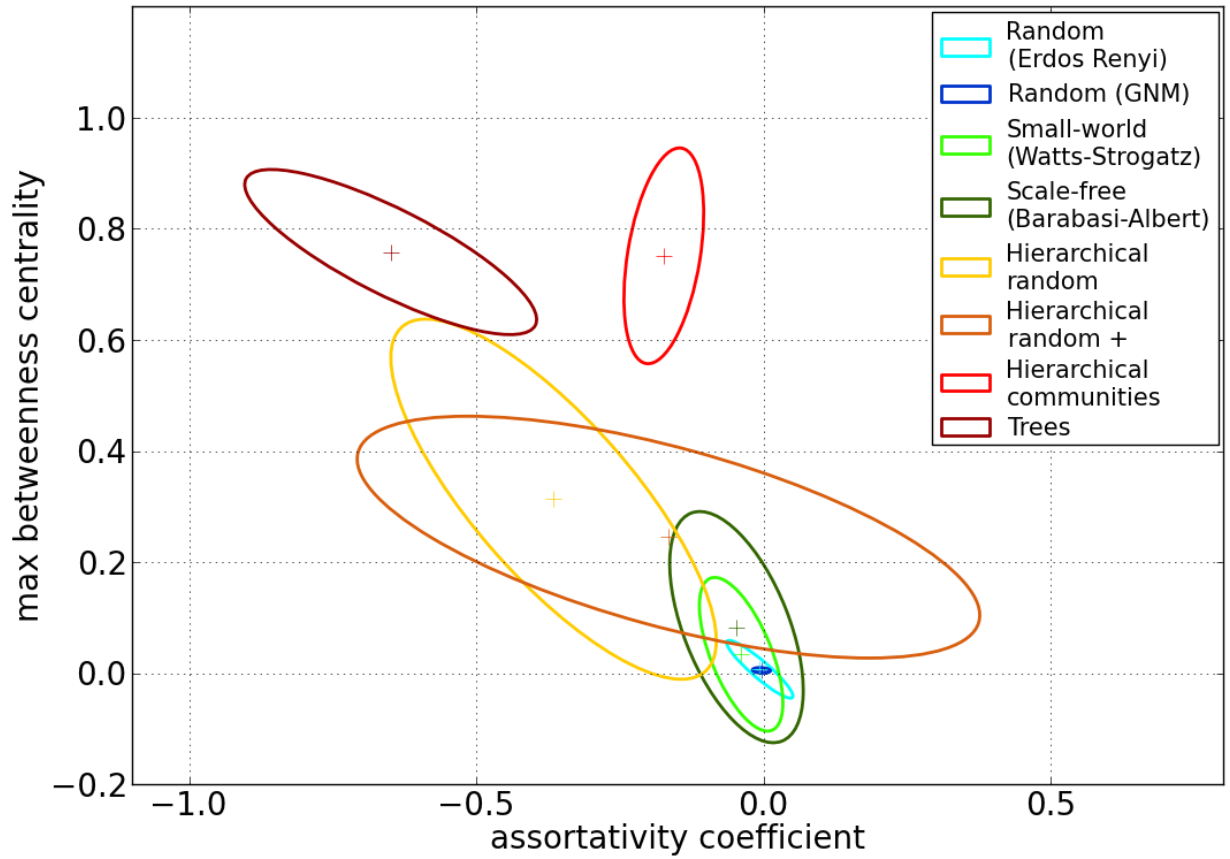


Figure 4.2: Showing the single standard deviation ellipses for the distribution of the assortativity coefficient values and maximum betweenness centrality value for each graph from the eight models in the spectrum (Table 4.1).

Graph model	Graph theme	Assortativity coefficient \bar{x} (σ)	Maximum betweenness centrality \bar{x} (σ)	Number of cycle basis per node \bar{x} (σ)
ER	Random	-0.01 (0.04)	0.01 (0.04)	15.53 (13.75)
GNM	Random	-0.00 (0.01)	0.01 (0.00)	15.09 (14.34)
WS	Small-world	-0.04 (0.05)	0.03 (0.10)	6.86 (4.16)
BA	Scale-free	-0.05 (0.08)	0.08 (0.15)	13.96 (8.38)
HR	Hierarchical	-0.37 (0.20)	0.31 (0.23)	0.51 (0.29)
HR+	Hierarchical	-0.17 (0.38)	0.25 (0.15)	0.55 (0.29)
HC	Hierarchical communities	-0.18 (0.05)	0.77 (0.14)	1.91 (0.38)
TREE	Tree/hierarchical	-0.65 (0.18)	0.76 (0.11)	0.00 (0.00)

Table 4.2: The mean values for each of the eight graph types across the graph metrics computed.

Figure 4.3 compares the multivariate relationship between the AC values and the number of cycle basis (CB) per node, again calculated for all graph models. The standard deviation ellipses highlight the large variation in the CB per node values for the non-hierarchical graphs; a mean of 10.16 compared to 0.24 for the hierarchical models (Table 4.2). In particular, the two random models, ER and GNM, have the highest CB per node values, 15.53 and 15.09 respectively. In contrast the TREE graph and the two in-house developed models based on this, HR and HR+, all have a very low count of CB per node values, with the TREE having zero, and HR and HR+ having an average value of 0.51 and 0.55 respectively (Table 4.2). The final hierarchical graph type, HC, has a greater number of CB per node ($\bar{x} = 1.91$), though this is still much lower than the next highest for the BA (scale-free) graph model ($\bar{x} = 6.86$), making it very clear that there is a difference between the number of CB per node in the hierarchical and non-hierarchical graphs. This indicates that the hierarchical graphs are not as well connected with potential consequences on the level of redundancy in the graphs which may affect their robustness to failures.

It is evident that where the number of CB per node is low, the AC value will be negative, meaning the graph is disassortatively mixed with respect to the degree of the nodes (nodes are connected to nodes with different degrees rather than nodes with the same degrees). The reverse of this can also be inferred, that where the CB per node values are high, the AC values will tend towards one. This is a result of the nodes in the weaker connected graphs, which have a lower

CB per node value, having similar degrees or a low variation as each node tends to have just enough incident edges to allow a connected graph. In the stronger connected graphs, more edges are present and this leads to greater redundancy in the graph connectivity, allowing the number of incident edges to vary more per node leading to a disassortatively mixed graph.

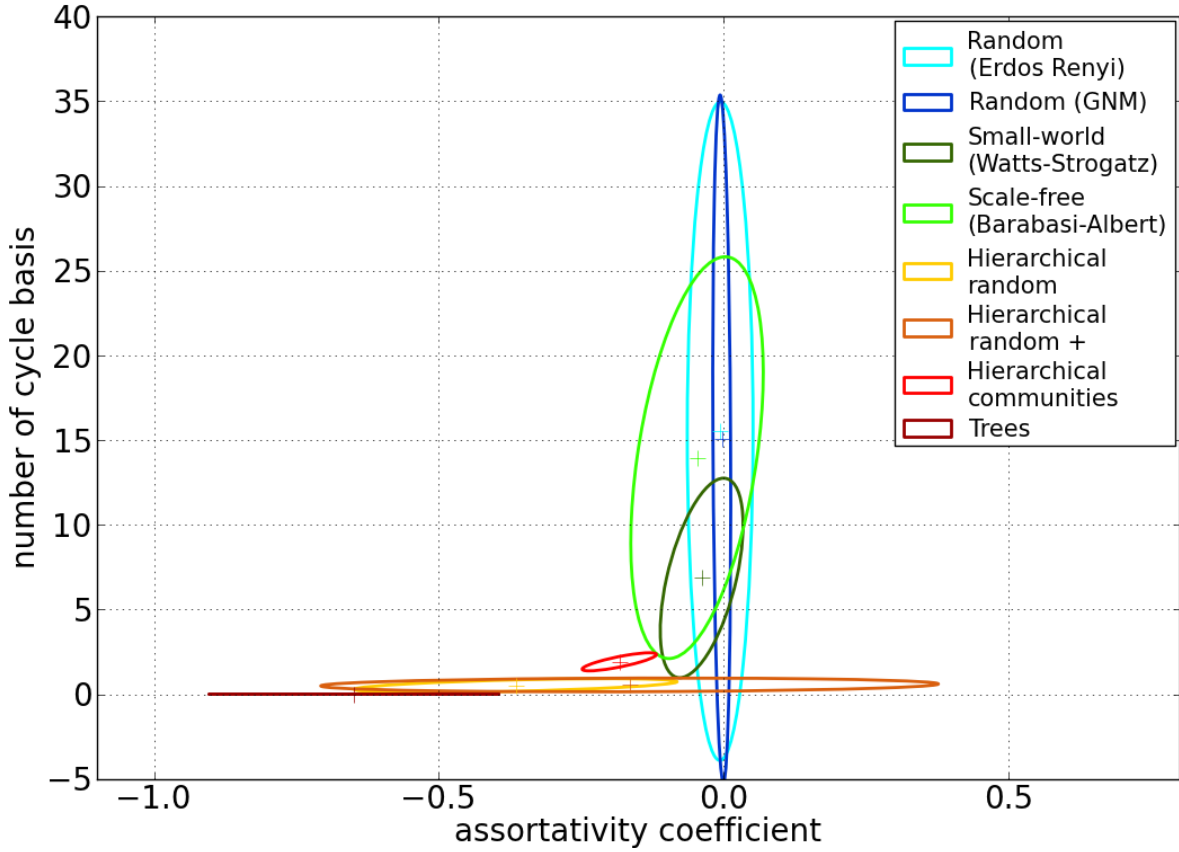


Figure 4.3: The single standard deviation ellipses comparing the relationship between the assortativity coefficient and the number of cycle basis for each graph within the graph suite (Table 4.1).

A comparison of the values of CB per node and the MBC values suggests there is a clear difference between the hierarchical and non-hierarchical models (Figure 4.4). The single standard deviation ellipses show that the hierarchical graph types, while having high MBC values (≥ 0.25) (Table 4.2) also have a low number of CB per node (0-1.91). However, the non-hierarchical graph types exhibit much higher CB values (≥ 6.86), while having low MBC values which tend towards zero (highest being 0.08). This shows that in general the greater the number of CB the lower the MBC value will be, a result of greater connectivity in the graphs providing more paths between nodes. This is an observed trend with one notable exception being the HC graph type which exhibits a higher than expected MBC for the number of CB. Having a modular structure, with communities of well-connected nodes at levels within a hierarchy allows some cycle basis to form, though globally the graphs still have an explicit hierarchical structure

meaning the graph is still topologically reliant on a small number of critical nodes as illustrated with a high MBC value.

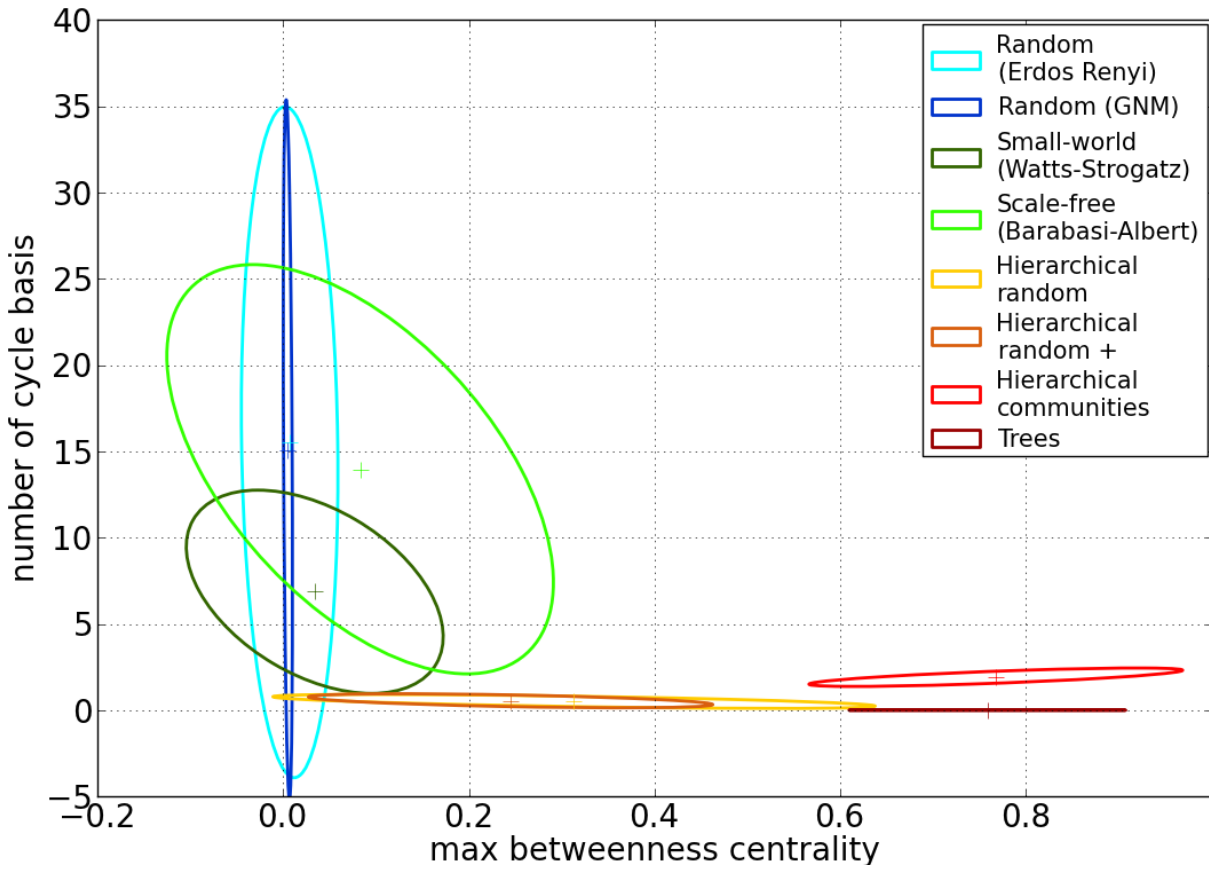


Figure 4.4: The single standard deviation ellipses comparing the distribution of values for all graphs and graph types (Table 4.1) for the maximum betweenness value and the number of cycle basis.

4.2.4 Statistical graph similarity

In order to quantify the statistical difference between the graph models on the basis of the multivariate distribution of the metrics employed, a transformed divergence analysis was performed. This performs a bi-directional pairwise test over two sets of values to quantitatively assess the degree of statistical separability between them (Chapter 3, Section 3.4.4, page 58). Table 4.3 shows the results of this, where a value of 100 indicates that any two graph-models in terms of their multivariate distribution exhibit no overlap, while a value of zero indicates identical distributions; values between 0-100 can be interpreted linearly as the percentage of multivariate separability that exists between any two graph models. Within the range of values returned, a value ≥ 85.00 is considered in many cases to be indicative of very good separability between the distributions being assessed (Swain and Davis, 1978). The separabilities are presented for each pair-wise combination of metrics for each pair of graph models. The key

values of interest are those towards the top right corner of the Table 4.3 where the statistical values of similarity between the hierarchical and non-hierarchical graphs are located.

		ER	GNM	WS	BA	HR	HR+	HC	TREE
ER	AC-MBC	-	99.91	84.78	99.58	100.00	100.00	100.00	100.00
	AC-CB	-	53.22	70.14	27.90	100.00	100.00	100.00	_*
	MBC-CB	-	99.96	75.43	72.62	100.00	100.00	100.00	_*
GNM	AC-MBC	-	-	100.00	100.00	100.00	100.00	100.00	100.00
	AC-CB	-	-	95.81	98.99	100.00	100.00	100.00	_*
	MBC-CB	-	-	100.00	100.00	100.00	100.00	100.00	_*
WS	AC-MBC	-	-	-	14.13	97.79	99.55	99.99	100.00
	AC-CB	-	-	-	43.86	100.00	100.00	100.00	_*
	MBC-CB	-	-	-	43.98	100.00	100.00	100.00	_*
BA	AC-MBC	-	-	-	-	75.24	83.36	99.81	99.90
	AC-CB	-	-	-	-	100.00	100.00	100.00	_*
	MBC-CB	-	-	-	-	100.00	100.00	100.00	_*
HR	AC-MBC	-	-	-	-	-	39.84	98.85	90.23
	AC-CB	-	-	-	-	-	27.34	99.37	_*
	MBC-CB	-	-	-	-	-	6.53	99.76	_*
HR+	AC-MBC	-	-	-	-	-	-	99.99	94.71
	AC-CB	-	-	-	-	-	-	100.00	_*
	MBC-CB	-	-	-	-	-	-	99.72	_*
HC	AC-MBC	-	-	-	-	-	-	-	100.00
	AC-CB	-	-	-	-	-	-	-	_*
	MBC-CB	-	-	-	-	-	-	-	_*

Table 4.3: The transformed divergence analysis of the distribution of the metric values for each graph type. * Results cannot be computed as the TREE model has no cycles.

The values presented in Table 4.3 suggest only in a relatively small number of cases similarities exist between any two graph models/types. The first is the HR and HR+ graph types, where values of 27.34 and 6.53 were returned for the AC-CB and MBC-CB metrics. This similarity was expected as both of these graph types originate from the same underlying TREE model with the only difference being the method employed for the addition of edges to the topological structure (Chapter 3, Section 3.3.5 and 3.3.6, pages 48 and 49). The second similarity identified is between the BA and WS graph types for the AC-MBC relationship, where a value of 14.13 is returned, indicating that the two models may at least share some characteristics. The other two metric relationships, AC-CB and MBC-CB, do show less similarity with both returning values of 43.86 and 43.98 respectively, but low enough values to indicate that there is a degree of similarity between the two graph models. This similarity can be seen in the standard deviation ellipses for the metrics (Figure 4.2, Figure 4.3 and Figure 4.4) where there is always a degree of overlap between them.

Other values of note include those which suggest a similarity between the ER graph model and the GNM and BA models, where values for the AC-CB distribution of 53.22 and 27.90 have been returned respectively. The low value for ER-BA relationship, 27.90, is unexpected, especially where values of 99.58 and 72.62 have been returned for the other two metric distributions, AC-MBC and MBC-CB respectively. The structure of both graphs, and these values, suggest the similarity between these models as a result of the number of cycle basis (CB), the common metric between the two distributions with the lowest statistical values. Within both the ER and BA graph models there is no limit or constraints on the generation of cycles, with these generated as nodes are connected at random, unlike in the other non-hierarchical models. The BA model allows cycle basis to be generated, with these likely involving the few nodes with high degrees which form in this model, whereas within the random models the cycles are likely to be more evenly distributed through the nodes of the network. The other relationship, between ER and GNM models, where a value of 53.22 was returned for the AC-CB distribution, both use methods where edges are added to connected nodes at random. This value is much lower than those returned for the other two metric distributions, 99.91 and 99.96 respectively, showing that the graphs are not similar.

The top right corner of Table 4.3 compares the hierarchical graph types (HR, HR+, HC and TREE) against non-hierarchical graph types (ER, GNM, BA and WS). In these cases, there are only two values below 97.00; 75.24 for BA and HR and 83.36 for BA and HR+ both for the combination of AC and MBC metrics. This indicates that there is a significant difference

between those graphs with a non-hierarchical organisation and those with a hierarchical organisation.

As expected the results in Table 4.3 show that the TREE network type has very few similarities to the other network types analysed, with only a slight similarity being suggested to the HR and HR+ graph types with values of 90.23 and 94.71 respectively for the AC–MBC metrics. Uniquely, the HC graph type has a high separability to all the other graph types, even more so than the TREE model, with no value below 98.85. This suggests that despite being a hierarchical graph, it shares little in common in terms of its topological structure with the other hierarchical models (and the non-hierarchical models).

The results presented statistically show that there are two sets of graphs within the suite generated from the eight models, with the characteristics of the four hierarchical models showing no similarities with the four non-hierarchical graph models through three metrics, AC, MBC and number of CB per node. These characteristics explicitly show that topology of the hierarchical graphs is different from the non-hierarchical graphs, with the differences also suggesting that the hierarchical graphs may exhibit a poorer robustness to perturbations given their high MBC and low CB values.

4.2.5 Topological hierarchical graph robustness

Using the developed topological failure model presented in Chapter 3, Section 3.5 (page 59), the robustness of the suite of synthetic graphs is examined. The developed failure model iteratively removes a single node and those edges which are incident to it, repeating this until no edges are left in the graph (i.e., a null graph is formed). Three methods are used in the selection of the node to remove at each epoch, random selection, the node with the greatest degree and the node with the greatest betweenness centrality. For the second two approaches, after each epoch the degree and betweenness centrality are recalculated ensuring the most critical node, that with the greatest value, for these metrics is removed at each epoch. All three methods are used to explore the robustness of the suite of synthetic graphs with the results presented using the mean and standard deviations of the percentage of nodes required to be removed before null graphs were obtained. A total of 3038 graphs are analysed for each failure method, with 500 randomly selected from all graph models except for the HC and TREE models where the full suite, 7 and 31, are used (Table 4.1, page 105). Five simulations are performed for each graph for each failure method.

For the random node removal approach (Figure 4.5(a)), the random graph types show a greater robustness to the perturbations, with on average over 90% of the nodes needing to be removed for a null graph to form. The hierarchical graphs, with the exception of the HC graph type, exhibit a greater vulnerability to the random removal of nodes, with only 70% of nodes required to be removed (73.46, 74.68 and 68.20 for the HR, HR+ and TREE models respectively). The HC graph type displays a response similar to the non-hierarchical graphs, with 88.06% of the nodes needing to be removed on average to produce a null graph. In a manner similar to the transformed divergence analysis, while a hierarchical model, it displays a very different response compared to the other hierarchical graphs.

The first targeted node removal strategy employed, based on removing the node with the greatest degree at each epoch, results in a similar pattern of response (Figure 4.5(b)), with the hierarchical graph models remaining the most sensitive to perturbation. Again the non-hierarchical graphs exhibit a greater robustness to the failures than the hierarchical graph types (with the exception of the HC model), with between 70 and 85% of nodes needing to be removed before a null graph is formed (85.35% (ER), 82.77% (GNM), 72.57% (BA) and 77.89% (WS)). Figure 4.5(a) and (b) show that the difference between the GNM and ER models and the BA and WS models is greater in the degree based failure, with a difference of approximately 10% between compared to approximately 2% for the random failure simulations. This highlights the greater sensitivity of the BA and WS models to targeted failure methods compared to the random failure method, likely as a result of the presence of hub nodes within these graphs which makes them more vulnerable to the targeted approaches.

The HC graph type shows an average response to the node degree failure method (Figure 4.5(b)), which is much closer to the random graph models (83.70%), while the other hierarchical graph models exhibit a much greater vulnerability to the degree-based failure, with less than 50% of the nodes required on average to be removed to form null graphs (45.36% (HR), 47.16% (HR+) and 36.61% (TREE)). This combined with the metric analysis results presented in Section 4.2.3 and 4.2.4, suggest that the topological structure of the HC model may lead to a greater robustness to random or targeted perturbations compared to the other hierarchical structures, and indeed a behaviour more typical of non-hierarchical models.

Finally Figure 4.5(c) presents the results for the betweenness based method, showing a similar pattern of behaviour as the degree-based mechanism (Figure 4.5(b)). Again the non-hierarchical graphs exhibit a greater robustness to the failures compared to the hierarchical graphs, with approximately 80% of nodes needing to be removed compared to approximately 40% for the

hierarchical graphs. Again, the HC graph type exhibits a behaviour similar to the random, non-hierarchical models with 84% of nodes having to be removed.

The patterns of robustness observed through the three failure mechanisms suggests that the hierarchical graphs, with the exception of the HC graph type, are much more vulnerable to failures than the non-hierarchical graphs (i.e., ER, GNM, BA and WS graph models). For the random failure strategy, the hierarchical graphs fail 19% quicker, where for the targeted methods the hierarchical graphs fail approximately 34% quicker (Table 4.4). The HC graph type has a behaviour that is more like that of a random graph than any of the other hierarchical models suggesting that aspects of its topological structure allow it to be robust to perturbations more effectively than the other hierarchical models.

Node removal strategy	Non-hierarchical graphs \bar{x} (%)	Hierarchical graphs \bar{x} (%)	% difference/change
Random	94	76	19
Degree	80	53	34
Betweenness	81	54	33

Table 4.4: Mean percentage of nodes removed for the node removal options for the hierarchical and non-hierarchical network groups and the percentage difference between these.

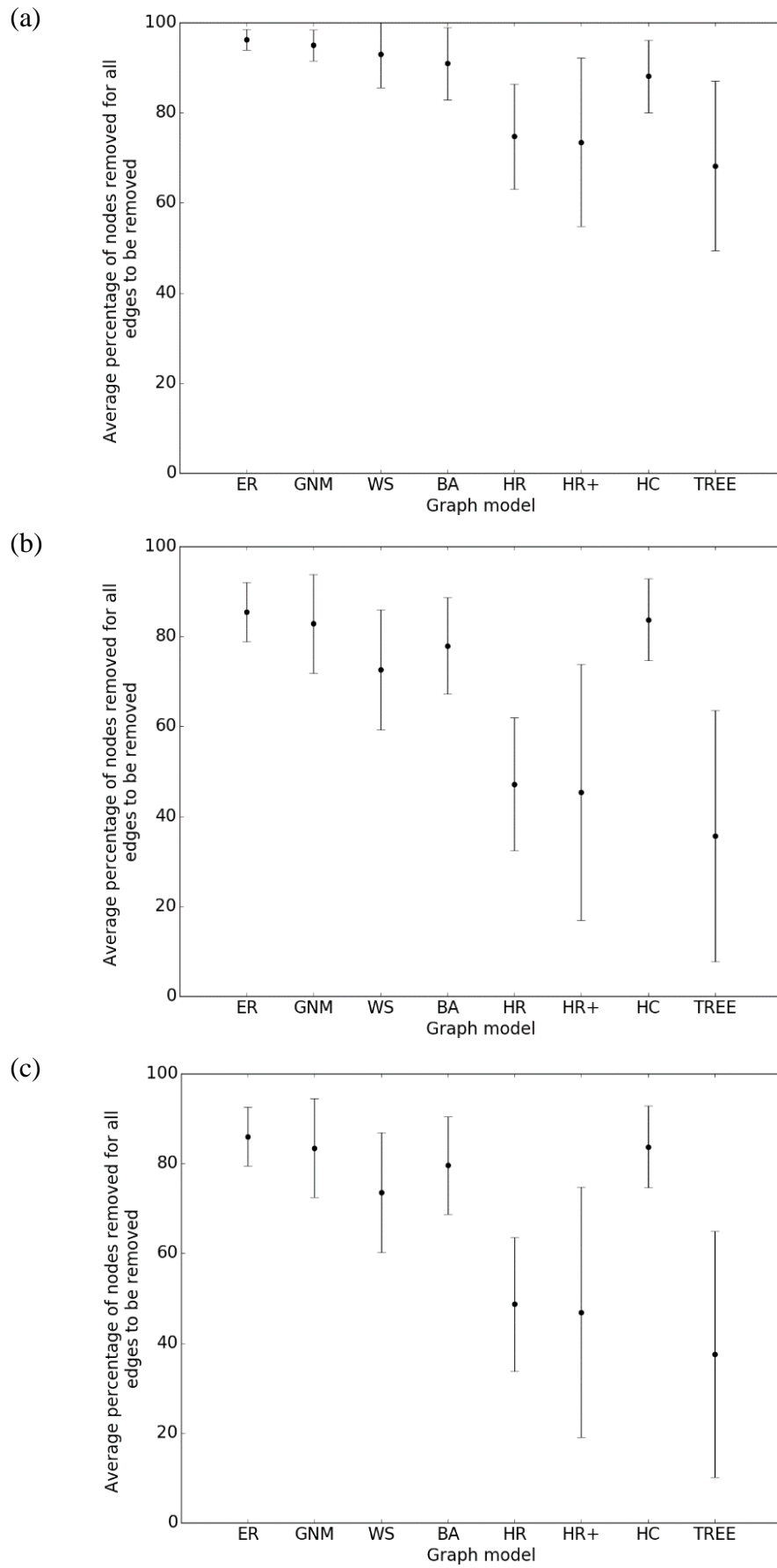


Figure 4.5: Average percentage of nodes removed for each graph to become empty for (a) random node selection, (b) degree based node selection and (c) betweenness centrality based node selection.

4.2.6 Failure characteristics of hierarchical graphs

The results from the failure analysis over the suite of synthetic graph models have shown a consistent set of observed behaviour, highlighting the lack of robustness of the hierarchical graphs other than the HC model (Section 4.2.5). This sub-section presents a more detailed analysis of the graph failure model results, presenting how they transition from the original state to null graphs. Due to the large number of graphs analysed, the results from a single graph which represents the collective behaviour of each model is presented in Figure 4.6 and Figure 4.7, with a greater set of results presented in Appendix D Section D.3.

Figure 4.6 and Figure 4.7, show the number of components as well as the average size of these for each graph model for the three different failure approaches. The non-hierarchical graphs, Figure 4.6 (a - d) only start to fragment once at least 20% of their nodes have been removed (20% for the BA and WS graphs and 50% for ER and GNM graphs). Once these models start to fragment, the number of components rises quickly. This indicates that the non-hierarchical models are initially robust to the failures but once between 20-50% of the nodes have been removed they rapidly start to form a large number of relatively small (in terms of the number of nodes) disconnected components. This is likely a result of the greater connectivity in these graphs as indicated by the greater number of cycle basis and lower maximum betweenness centrality presented in Section 4.2.3.

In contrast the hierarchical graphs (Figure 4.7 (a - d)), start to fragment after <5% of the nodes have been removed, with components quickly forming showing a much weaker robustness. The TREE graph exhibits characteristics suggesting this is the least robust graph with a null graph forming for the three failure methods after as little as 25% of nodes have been removed. The HR graph appears the next least robust graph with all edges having been removed by the time 30-35% of nodes have been removed. The peak in the number of components in the network also comes after approximately 15% of nodes have been removed whereas for the TREE model this is much closer to 10% of nodes. The HR+ model exhibits a more robust behaviour with the peak in the number of components approximately after 35% of nodes have been removed from the graph and a null graph not forming until 50-60% of nodes have been removed.

The HC graph, Figure 4.7 (c) shows a very different failure pattern to the other graph models, both hierarchical and non-hierarchical. The difference in behaviour is caused by the structure of the network, with disparate modules/communities embedded within the topology (Chapter 3, Section 3.3.7, page 51) meaning that the graph relies on a small number of highly connected hub nodes to connect communities. Once these are removed the graph rapidly fragments, as

shown in the results (Figure 4.7(c)), where the removal of $<5\%$ of nodes results in >100 components forming for both the targeted failure mechanisms. However, thereafter, the number and size of the components is sustained for many epochs due to the communities staying ‘internally’ connected. This therefore means these communities can still function as graphs, and it is not until at least another 20-40% of nodes have been removed that the topology of the network begins to change significantly again, with complete failure occurring after $> 80\%$ of nodes have been removed, compared to 20-60% of the nodes in the other hierarchical graphs. The betweenness response differs from the degree response slightly as this targets those nodes within each community which have the greatest number of flows passing through them, causing the communities to fragment quicker with the key connecting nodes failing first. In many cases these nodes may have the same degree of the other nodes in the communities, but their location and the nodes they are connected too makes them more central to the ability to traverse the graph and for the community to stay connected. Thus, while the HC graph fragments after a small number of perturbations, the community structure within its modules allow it to remain robust for much longer.

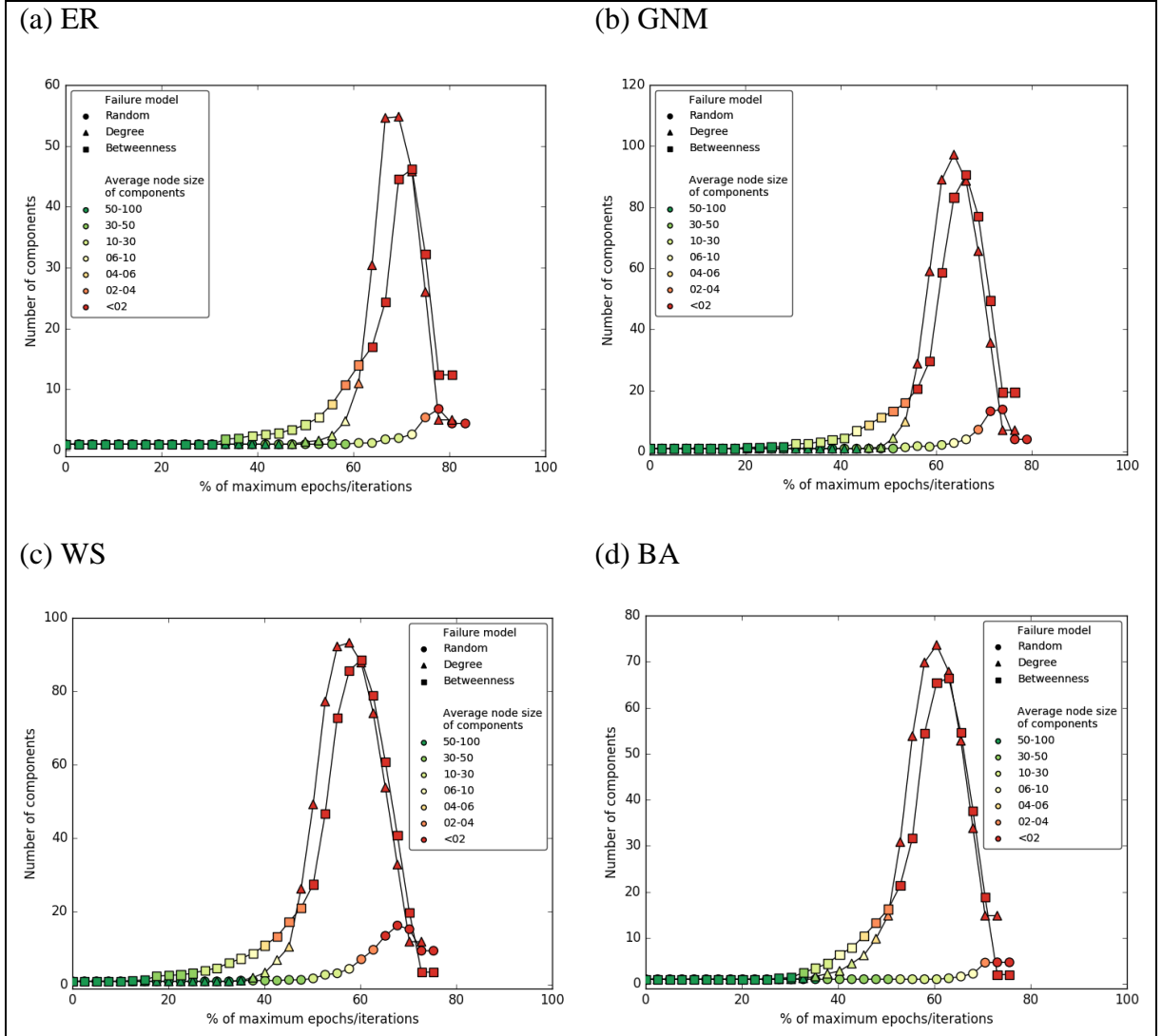


Figure 4.6: The typical response observed for each of the non-hierarchical graph models, ER (a), GNM (b), WS (c) and BA (d) to the three failure models. The plots show the rate of which each graph fragments via the proportion of subgraphs relative to the original number of nodes in the graph (y-axis) as nodes are removed from the graph, shown as a percentage of the original count (x-axis). The coloured symbols show the average size of the subgraphs as a percentage of the number of nodes in the original graph.

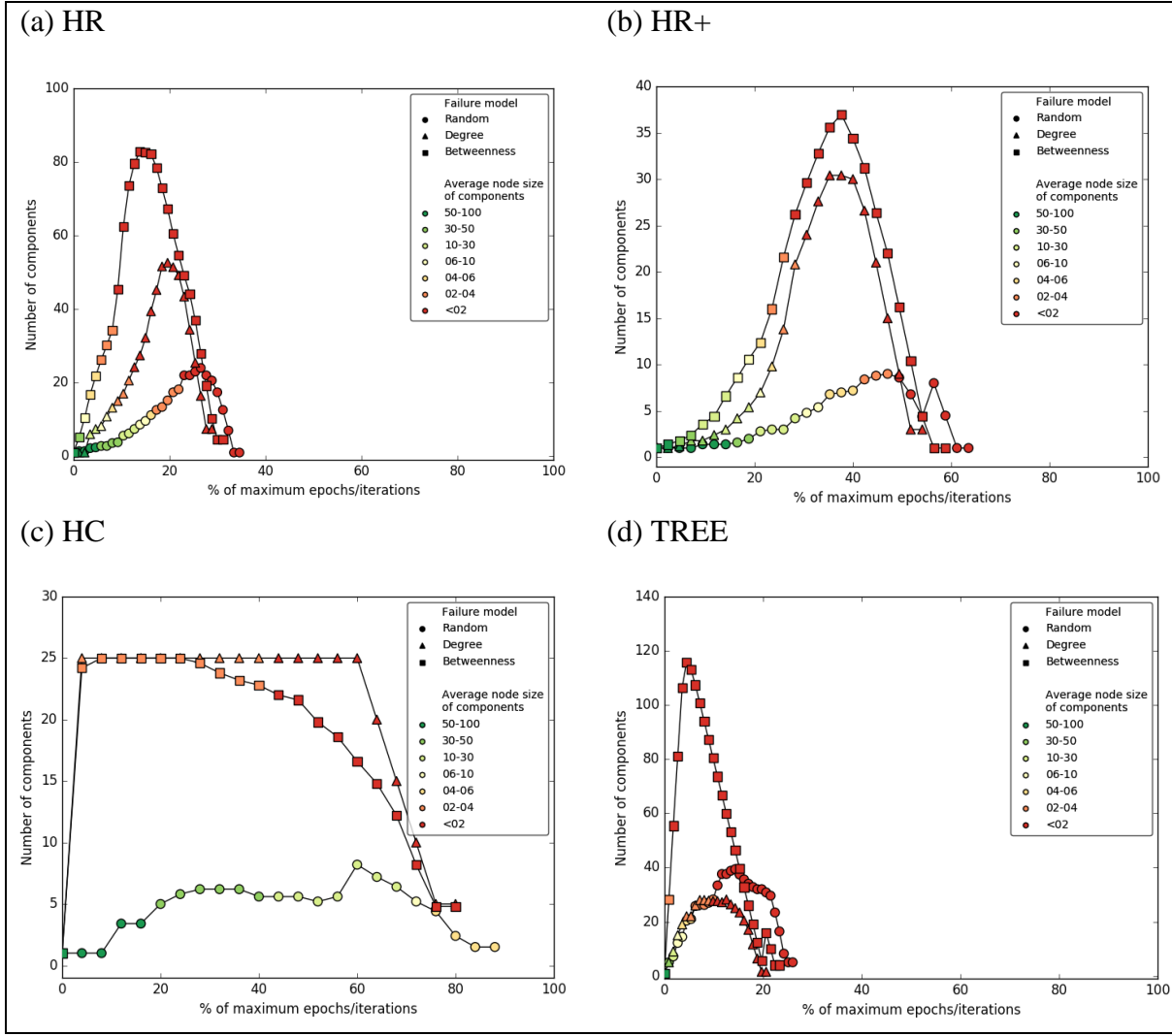


Figure 4.7: The response of the hierarchical graph types, HR (a), HR+ (b), HC (d) and TREE (d), to the three failure mechanisms. These show the rate of which each graph fragments via the proportion of subgraphs relative to the original number of nodes in the graph (y-axis) as nodes are removed from the graph, shown as a percentage of the original count (x-axis). The coloured symbols show the average size of the subgraphs as a percentage of the number of nodes in the original graph.

4.3 The hierarchical characteristics of infrastructure networks

4.3.1 Introduction

Section 4.2 has presented an analysis of the synthetic graph models exploring the characteristics of each of the eight models including those which are hierarchical and non-hierarchical. Chapter 2 Section 2.3.3 also reviewed the emerging literature which suggests that some infrastructure networks may be hierarchically organised. Given this, the characteristics of the analysed hierarchical graphs are used to explore the existence of a hierarchical organisation in critical infrastructure networks.

A suite of critical infrastructure networks has been generated, Table 4.5, as detailed in Chapter 3, Section 3.6 (page 64), totalling 42 infrastructure networks. The characteristics of the groups and the individual infrastructure networks will be analysed allowing the characteristics of these to be compared to the synthetic graph models presented in Section 4.2. This includes identifying within the suite of infrastructures networks any which exhibit characteristics of hierarchical networks. The suite of infrastructure networks is analysed using the same methods applied to the suite of synthetic graphs in Section 4.2.

Infrastructure group	Infrastructure networks
Air (flights)	British airways, EasyJet, European, Northern America, UK, USA
Communications	Janet
Energy	National electricity transmission, National electricity transmission NT (no towers), National electricity transmission MT (minimal towers), National gas transmission, Synthesised electricity transmission
Rail – national	GB rail, Ireland rail, Ireland rail with shortcuts
Rail – regional	Boston (MA, USA) subway, Boston subway with TAPAN, London DLR, London light rail, London Overground, London tube, Manchester Metrolink, RATP (Paris public transports) rail, RATP metro, RATP RER, RATP tram, Tyne and Wear metro, Tyne and Wear metro with shortcuts
Rivers	Dee, Eden, Severn, Tyne
Roads – national	Great Britain motorways, A and B roads, Ireland motorways and trunk roads, Ireland motorways, trunk and primary roads
Roads – regional	Leeds motorways, A, B and minor roads, Leeds motorways, A and B roads, Milton Keynes motorways, A, B and minor roads, Milton Keynes motorways, A and B roads, Tyne and Wear motorways, A, B and minor roads, Tyne and Wear motorways, A and B roads, Tyne and Wear motorways and A roads

Table 4.5: For each infrastructure group the infrastructure networks included in the suite.

4.3.2 Degree distributions of critical infrastructure networks

With 42 infrastructure networks analysed in total, only a sub-set of the degree distributions from each infrastructure group are presented. The complete set of degree distributions for all networks that belong to each group is presented in Appendix E Section E.1.

The degree distributions of the rail infrastructure networks, both national and regional, show a similar set of distributions (Figure 4.8), with all having a single peak where the node degree is two ($k = 2$). Although the tail of the four distributions vary in shape, they all indicate a similar trend with a decreasing proportion of nodes having higher node degree. The shape of the distributions for the national and regional rail networks all suggest that these infrastructures

have a similar distribution to those of small-world (the WS model) graphs (Section 4.2.2, page 102).

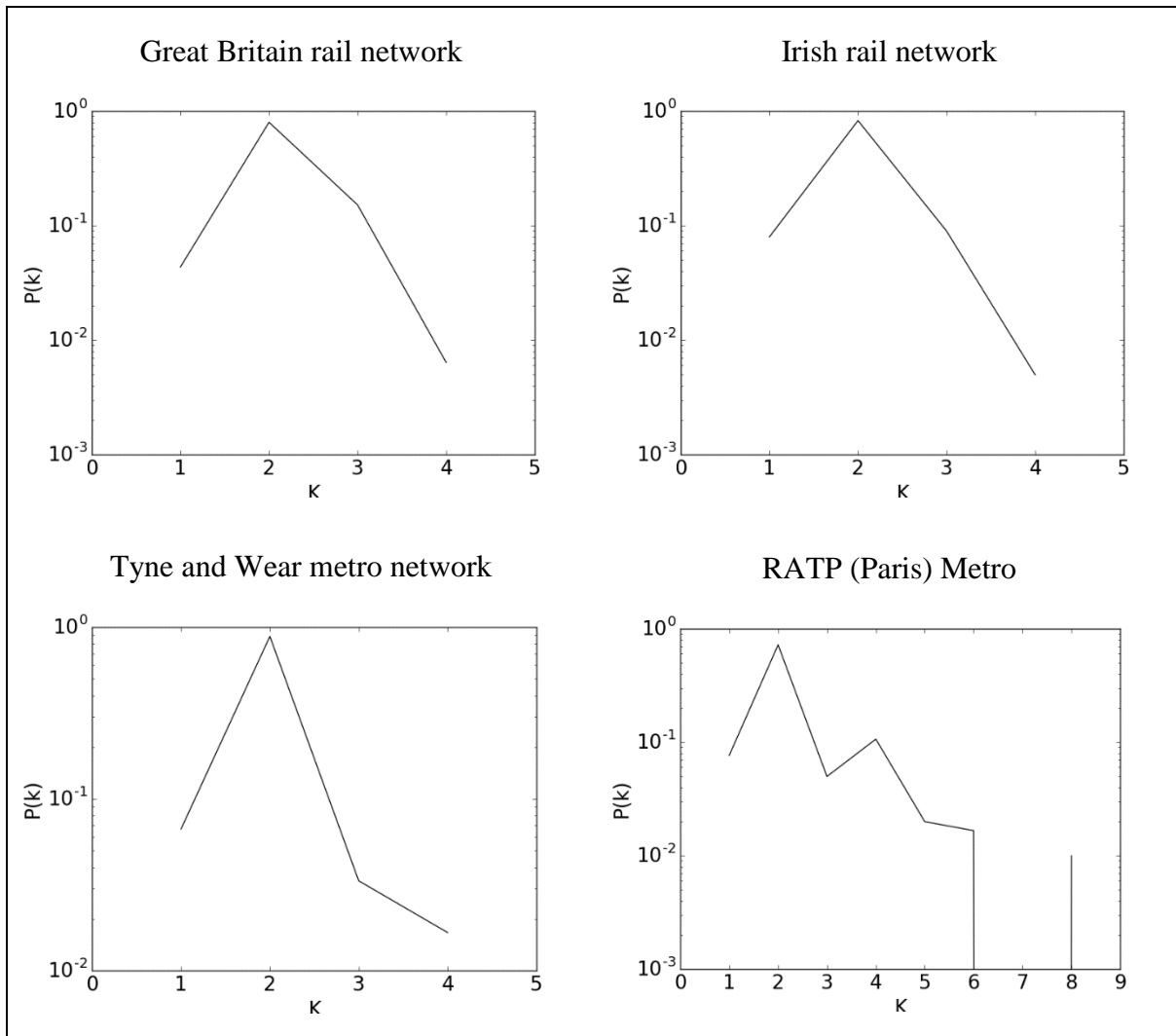


Figure 4.8: Degree distribution plots for selected rail critical infrastructure networks.

Degree distributions are shown for two of the air networks (Figure 4.9). The two presented and indeed the other air networks (Appendix E Section E.1) exhibit similar degree distributions which is similar to graphs generated using the scale-free BA model; the distributions quickly tail off from the peak proportion of nodes having low degrees to a high proportion of nodes with low degree.

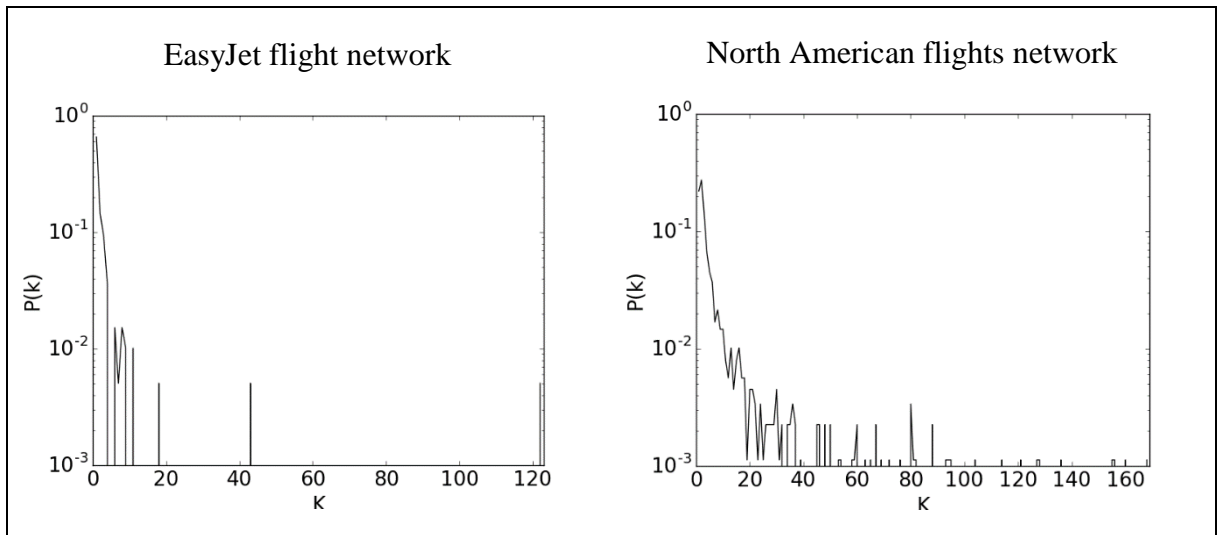


Figure 4.9: Degree distributions for selected air networks.

For the four river networks in the suite of infrastructure networks the degree distribution of two are shown in Figure 4.10 due to all four displaying similar characteristics. The distributions bear the greatest resemblance to the TREE network, with two clear distinct peaks indicating that the greatest proportion of nodes have a degree of one or three. This result is expected as river networks have previously been discussed as having a tree like structure (Barthelemy, 2011). However, as can be seen in the plot for the River Severn in Figure 4.10, as well as nodes with a degree of one and three as expected, there are also a small proportion of nodes with a degree of four. This is caused by rivers naturally braiding and altering the pure tree like topology. Further to this man-management of rivers, with courses diverted for canals and water wheels/turbines, results in the networks also being altered from the expected tree like topology.

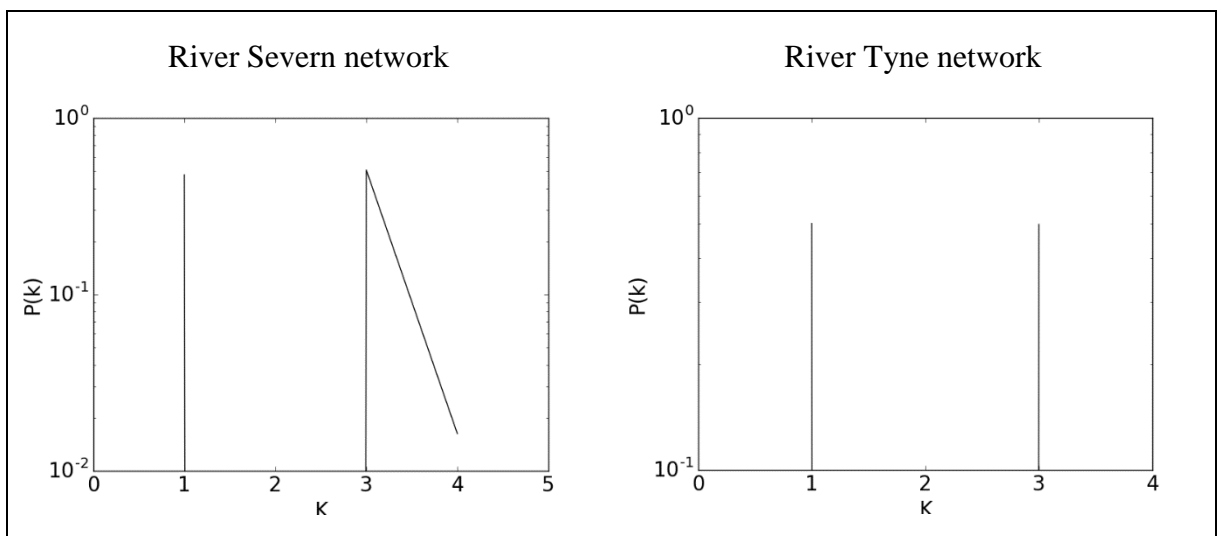


Figure 4.10: Degree distributions for selected river networks.

Two of the three distributions for the national road networks are presented in Figure 4.11. The plots show that there is a large number of nodes with a degree of one ($k = 1$), and a degree of three ($k = 3$). The proportion of nodes with degrees greater than three then decreases giving a tail to the distribution. The distributions don't match any of the non-hierarchical graph models or the HR and HR+ hierarchical models due to having no nodes with a degree of two, and with the distributions showing a small number of nodes with increasing degrees from the peak where $K=3$, these are not like the TREE or HC models either. However, road networks have previously been found to have a scale-free degree distribution (Kalapala *et al.*, 2006), a distribution which has some (albeit limited) similarities with those shown the Figure 4.11, including the tail, from the peak for those nodes with grater degrees.

Figure 4.12 shows the degree distributions for two of the seven regional road networks. These distributions are similar to the national road networks and as such again do not match any of the distributions generated by the eight graph models.

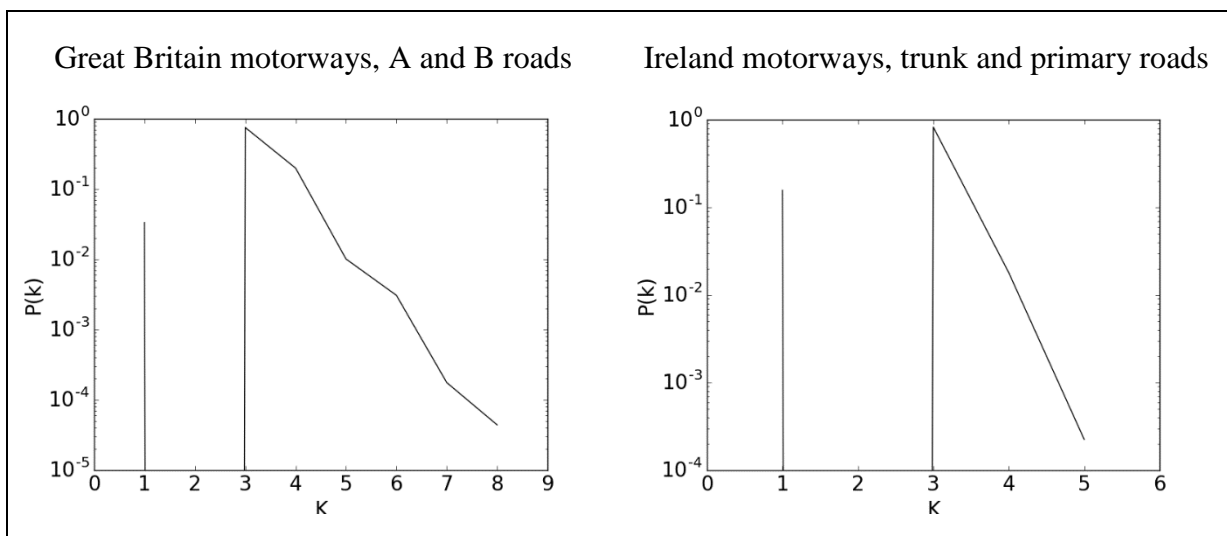


Figure 4.11: Degree distribution plots for selected national scale road networks.

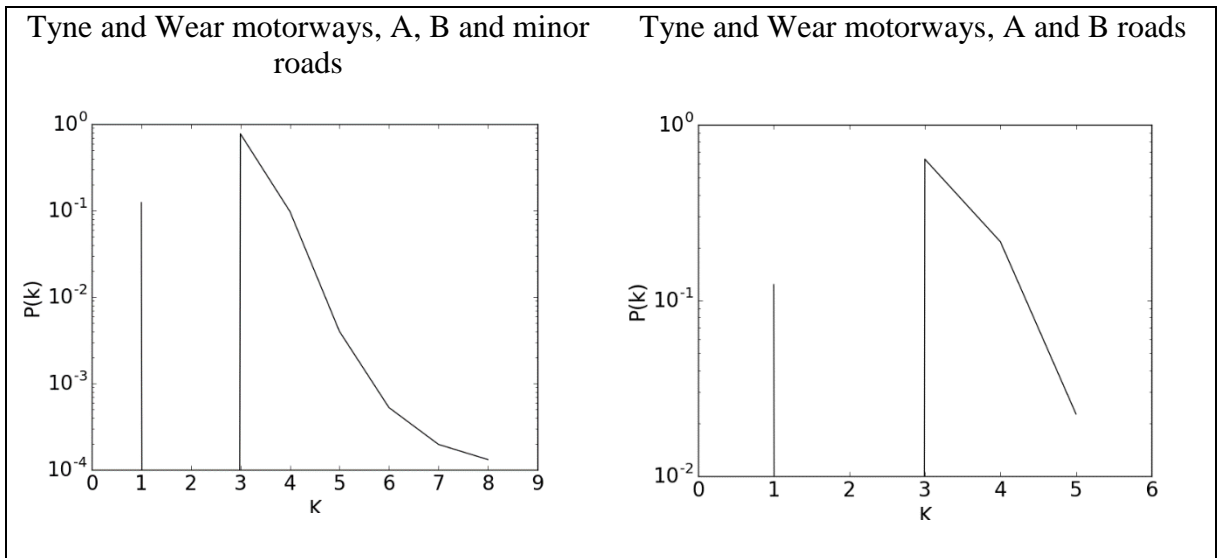


Figure 4.12: Degree distributions for selected regional road networks.

Degree distribution plots for two of the five energy networks are presented in Figure 4.12. The distributions have long right tails, a relatively large number of nodes that exhibit a high degree, indicative of scale-free distributions (Section 4.2.2, page 102).

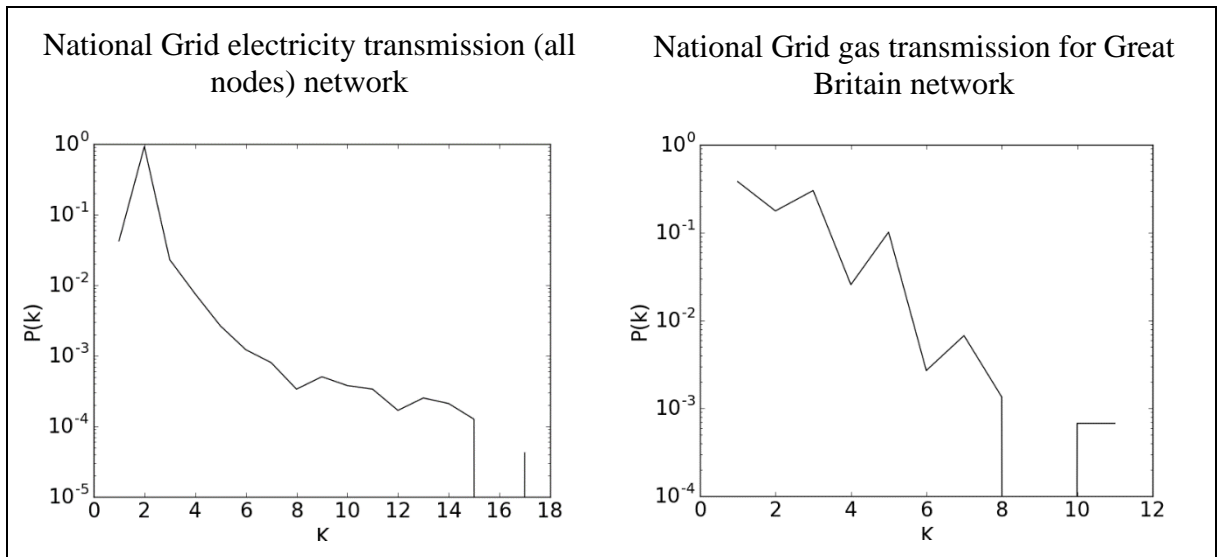


Figure 4.13: Selected degree distribution plots for energy networks.

4.3.3 Assessment of critical infrastructure network metrics

In order to better characterise the infrastructure networks compared to the graph models the same multivariate graph-metric analysis has been undertaken as in Section 4.2.3. For each group of infrastructure networks the multivariate results are plotted with the single standard deviation ellipses of each of the eight graph models (more detailed plots are shown in Appendix E Section E.2 where plots are presented showing the individual infrastructure networks).

For the assortativity coefficient (AC) and maximum betweenness centrality (MBC) (Figure 4.14) 24 of the 42 infrastructure networks lie within the standard deviation ellipses of the synthetic graph models, with 19 of these within the ellipses for the HR and HR+ graph models, and five within the ellipses for the WS (small-world) and BA (scale-free) models (noting that three of these lie closer to the mean of the HR+ model) (Table 4.6). For all the infrastructure networks not within an ellipse, there are all closer to hierarchical graphs. These results clearly suggest the infrastructure networks are most similar to the hierarchical graphs, with all but five having MBC values more similar to those. The AC metric appears to be less of a differentiator between hierarchical and non-hierarchical, though together the two metrics suggest the infrastructure networks are more similar to the hierarchical models. The AC and MBC values show that the infrastructure networks don't share characteristics with the non-hierarchical graph models, despite a number of them, including the rail, air and energy networks, all exhibiting degree distributions most similar to either the WS or BA model. This suggests that the metrics used return different characteristic properties to those shown by the degree distributions, a result of the metrics proving a higher level set of graph characteristics not captured by the degree distribution which captures the topological structure only.

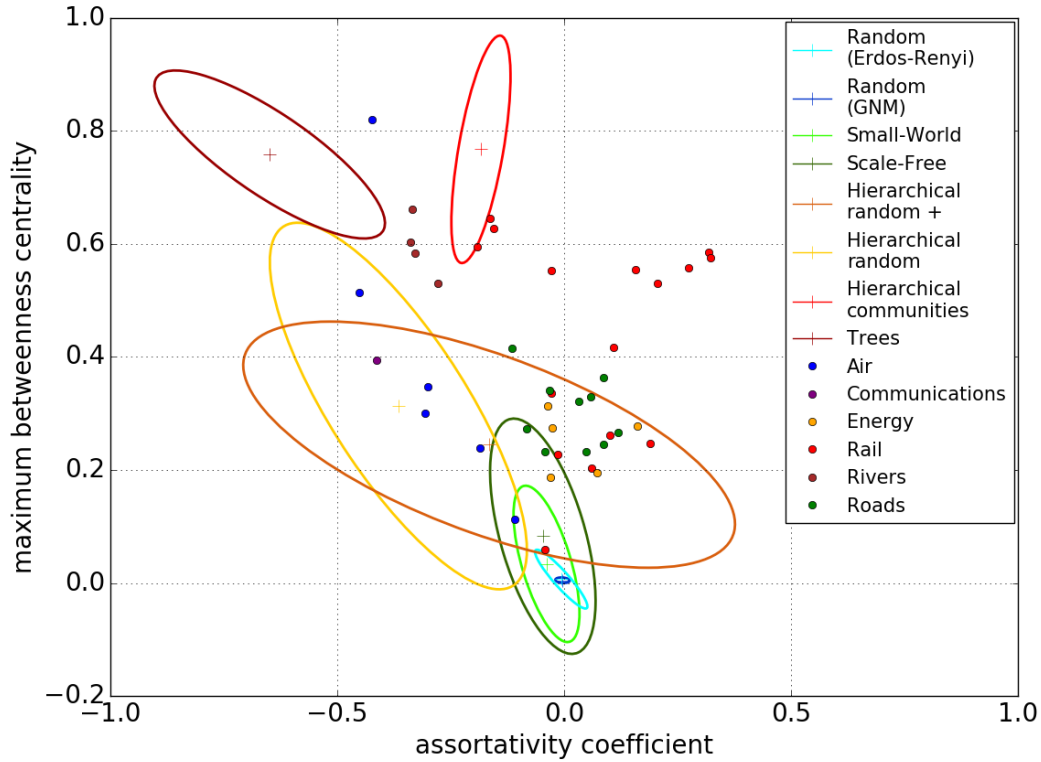


Figure 4.14: Showing for every infrastructure network (presented by group only) the maximum betweenness centrality value against its assortativity coefficient with the single standard deviation ellipses for the synthetic graph models.

Infrastructure	AC value	MBC value	Euclidean distance to WS	Euclidean distance to BA	Euclidean distance to HR+
RATP tram	-0.04	0.06	0.01*	0.03	0.16
European flights	-0.11	0.11	0.09	0.06*	0.11
Electricity transmission MT	0.07	0.20	0.19	0.16	0.16*
Leeds motorways, A and B roads	-0.04	0.23	0.19	0.15	0.04*
Ireland motorways and trunk roads	-0.03	0.34	0.30	0.25	0.14*

Table 4.6: The AC and MBC values for the five infrastructure networks within the non-hierarchical ellipses and the Euclidean distance to the nearest mean values for the synthetic models. * denotes lowest value.

There is a clear grouping of infrastructure networks within the HR single standard deviation ellipse where $-0.1 < AC < 0.3$ and $0.2 < MBC < 0.4$ (Figure 4.14). This is seen in the results for the energy networks (Figure 4.15), where all six are clustered in this area, though one of the national electricity transmission networks does fall within the scale-free ellipse. The MBC values returned for these energy networks indicates that they have at least a single node which is critical, though with values less than those observed in the TREE/HC models there is the potential for multiple nodes within these networks to have high values, with a subset of hub nodes rather than just one as in the TREE and HC graph models. The AC value is found to be similar to that found for many of the non-hierarchical graph models, but also is within the range for the HR and HR+ models.

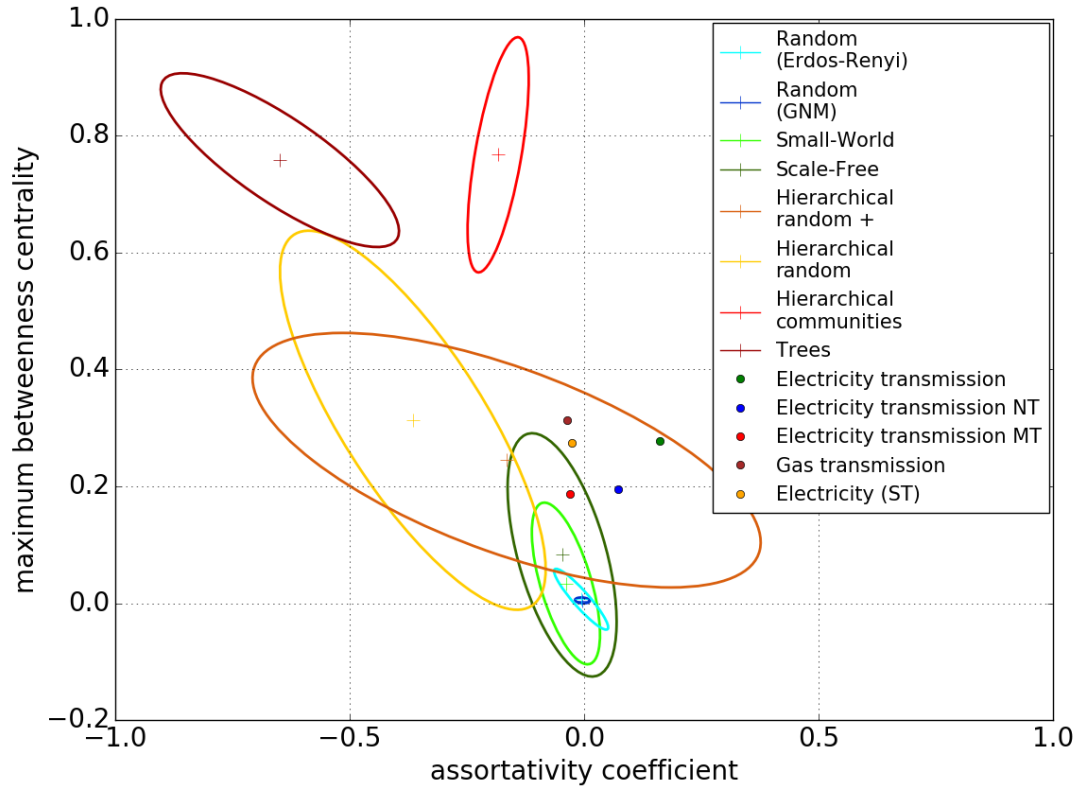


Figure 4.15: Comparing the values of the energy networks compared to the single standard deviations for the graph models for the assortativity coefficient and maximum betweenness centrality metrics.

A similar set of results have been returned for the national and regional road networks (Figure 4.16 and Figure 4.17). As with the energy networks the road networks have returned values for the maximum betweenness which suggests a greater likeness to the hierarchical models and the presence of critical nodes within the network which are relied upon for connectivity of the network. However, for the AC metric the networks have returned values more familiar to the non-hierarchical models, with only the HR+ hierarchical model graphs returning similar AC values. As a result 8 of the 10 networks lie within the HR+ ellipse, with the rest lying above these due to having higher MBC values.

The degree distributions for the road networks were not easily identifiable with any of the synthetic graphs, exhibiting some limited similarities with the hierarchical models as well as the WS and BA models (Section 4.3.2).

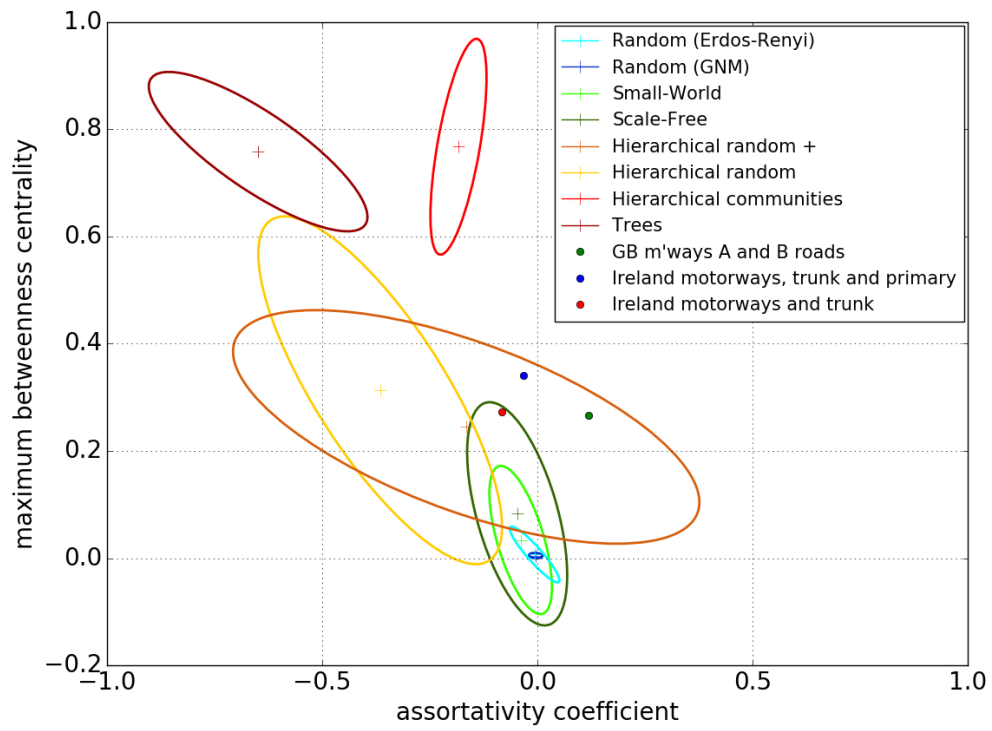


Figure 4.16: Assortativity coefficient and maximum betweenness centrality results for the national road networks compared to the standard deviation ellipses for the eight graph models.

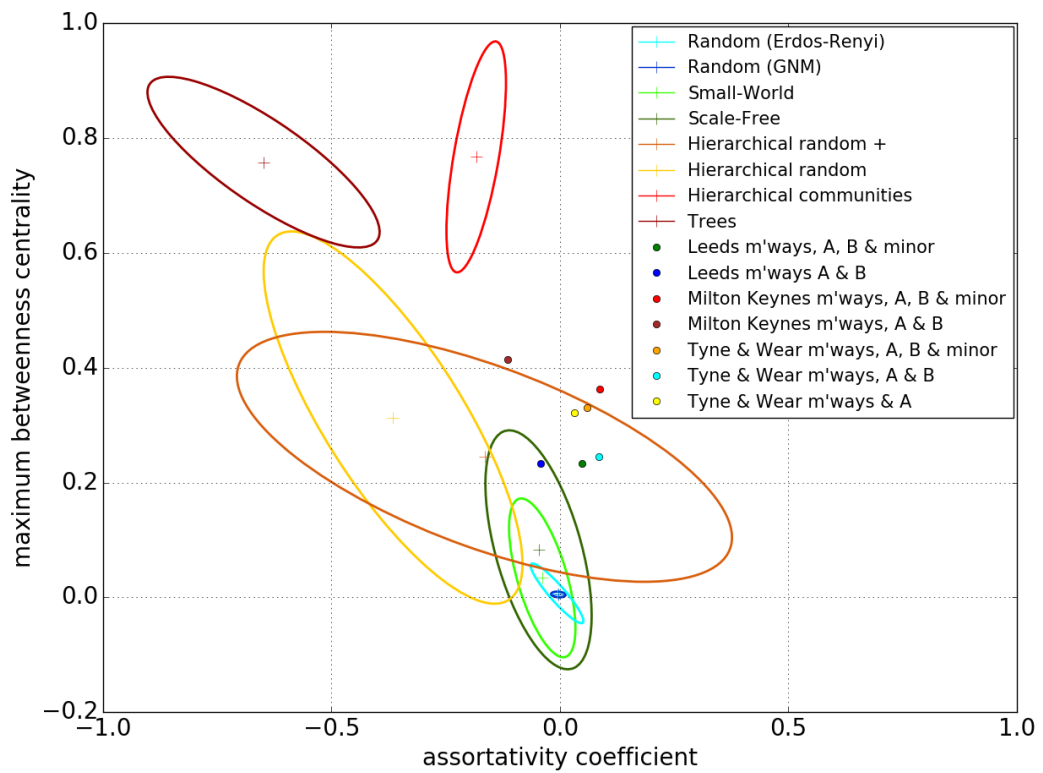


Figure 4.17: Comparing the results for the regional road networks and the distribution of the eight graph models for the assortativity coefficient and maximum betweenness centrality.

Both national and regional rail networks predominantly return similar values for the AC metric (Figure 4.18 and Figure 4.19) to those for the road networks, $-0.3 < AC < 0.3$, similar to the values for the non-hierarchical graphs and as well as the HR+ model. However, again they also have MBC values greater than 0.2, higher than those for the non-hierarchical graphs rendering the network closer in similarity to the hierarchical graphs. Three of the regional networks, Tyne and Wear Metro, London Overground and the RATP RER, lie very close to the HC model ellipse, the closest of the all the infrastructure networks to either the HC or TREE ellipse. For these networks in particular, these have high MBC values, greater than 0.6, suggesting that most shortest paths go through a single node in the networks.

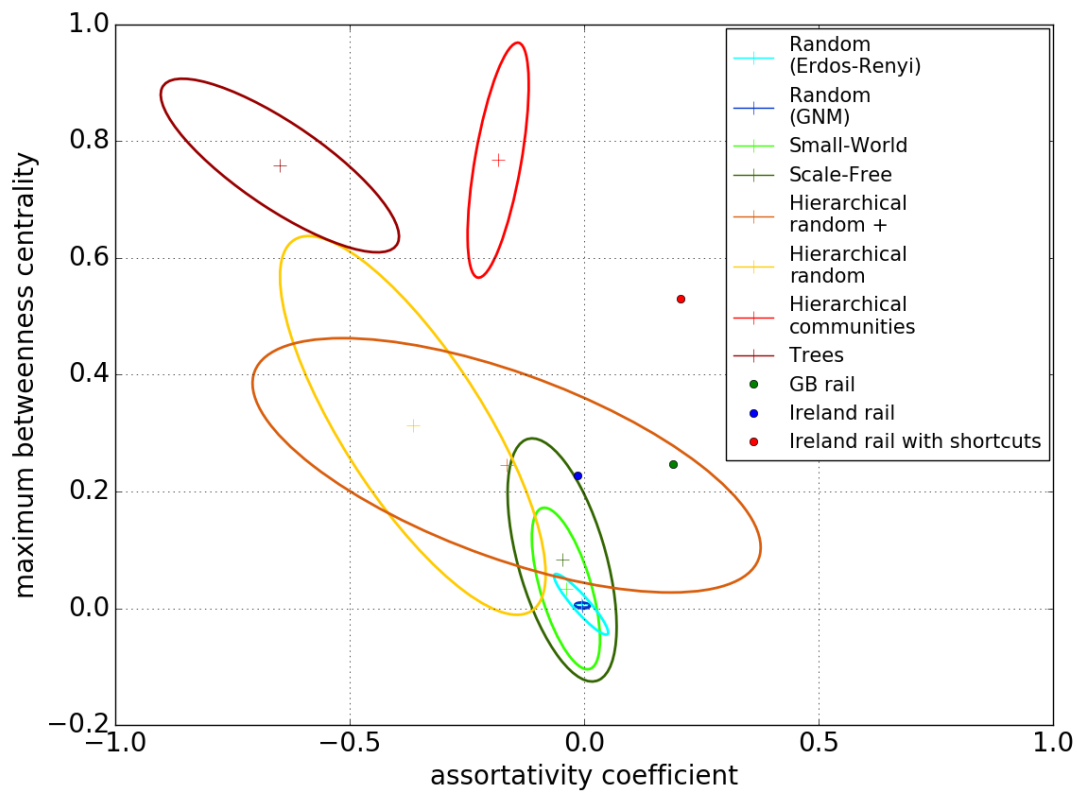


Figure 4.18: Metric distribution of the assortativity coefficient and maximum betweenness centrality metrics for the national rail networks.

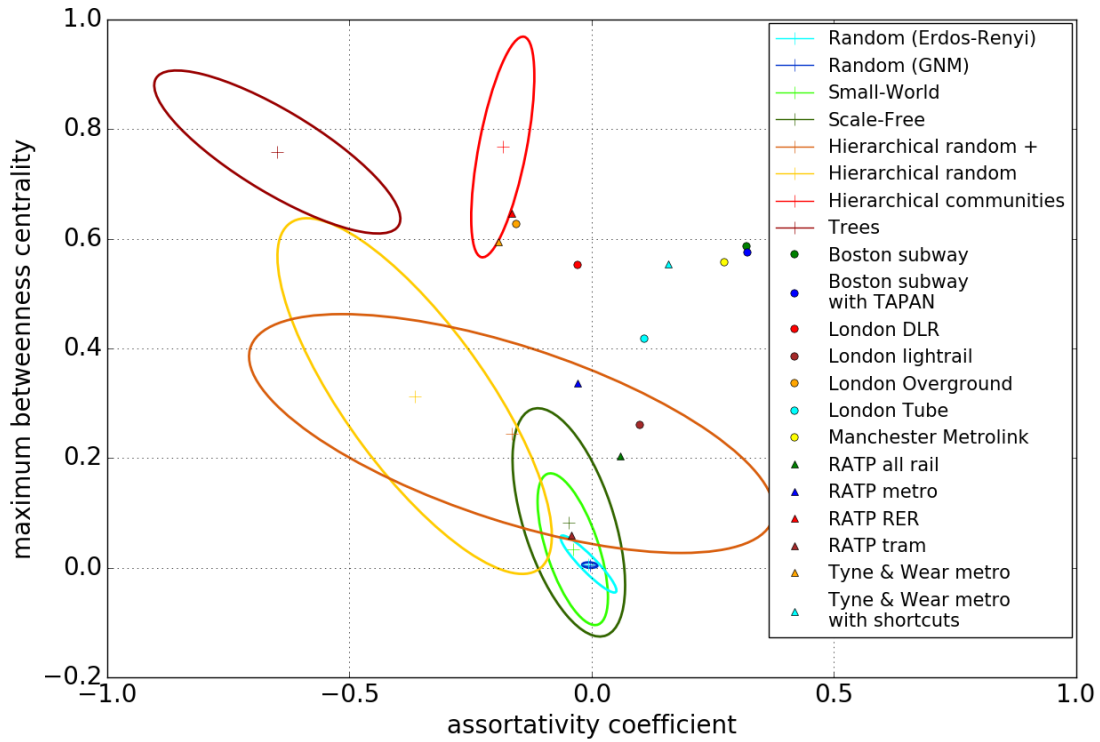


Figure 4.19: Metric distributions for the regional/metropolitan rail networks for the assortativity coefficient and maximum betweenness centrality metrics.

Some of the networks with the lowest AC values are those for air networks (Figure 4.20), where $-0.5 < AC < -0.1$. These low values suggest that nodes with different node degrees are connected to each other, rather than nodes with similar degrees being connected. This structure suggests that these networks, especially those where the values are closer to -1, may have significant hub nodes where the majority of the nodes linked to this have much lower degrees. This is also suggested by the MBC values returned for some of these networks, where the higher the value, the more likely there is a single hub node in the network. Specifically the network for British Airways has an AC value of -0.42 and MBC value of 0.82, the highest for any of the infrastructure networks, and indicative of having a hub node. This potentially makes these networks vulnerable to perturbations with a strong dependence on a single node for the connectivity of the network, as shown by the results in the failure modelling of the hierarchical graphs, and specifically the tree graph model, which the British Airways network is most like with regards to the metric values returned.

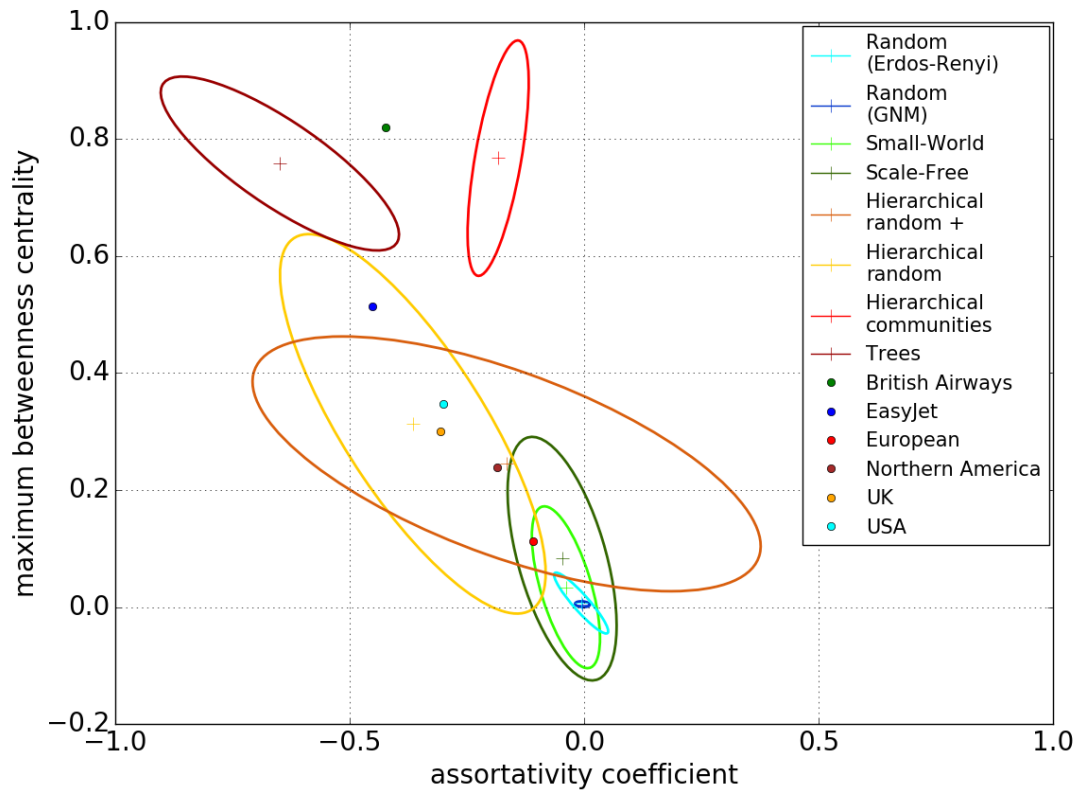


Figure 4.20: Location of the six air networks with regard to the single standard deviation ellipses for the eight graph models using the maximum betweenness centrality and assortativity coefficient values.

The suite of River networks (Figure 4.21) also exhibit AC values below 0, again suggesting a structure for these networks where nodes are not connected to nodes with similar degrees. However, in the case of the river networks this is caused by nodes having either a degree of one, those representing sources as well as the sink/river mouth, or a degree of three, those nodes where rivers/tributaries merge. Due to the tree like structure of river networks, they also have high MBC values, $0.5 < \text{MBC} < 0.7$, though not as high as those seen in the TREE and HC model graphs. This suggests that the River networks have a structure which may be vulnerable to failures as was exhibited by the tree model graphs, given the exhibited metric similarities.

The final network, the JANET communication network, exhibits similar values to the HR and HR+ graphs lying within the ellipses for both of these (Figure 4.22). The network has an AC value of -0.41 and a MBC value of 0.39, suggesting that no single node acts as a hub in the network, but that the network has a mixed set of node degrees leading to nodes being connected to others with different degrees. Without a node with a high MBC, the network may be more robust to failures than some of the other hierarchical graphs such as the HC and TREE model graphs.

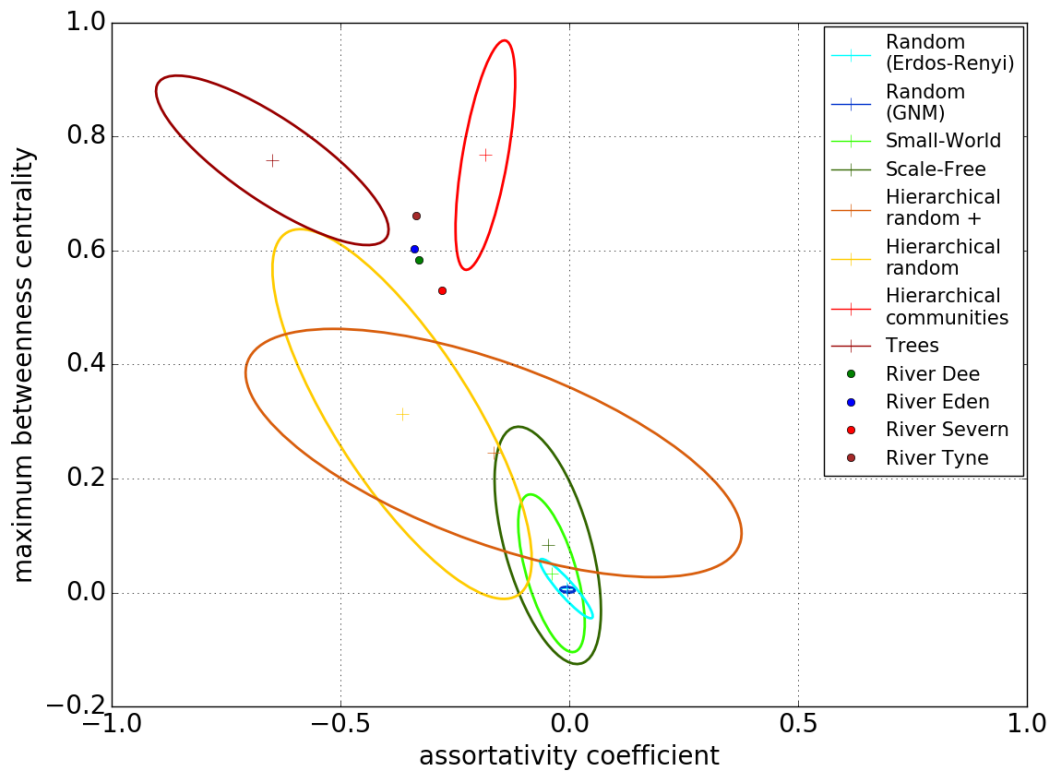


Figure 4.21: Distribution of the river networks using the assortativity coefficient and maximum betweenness centrality metrics with reference to the eight graph models.

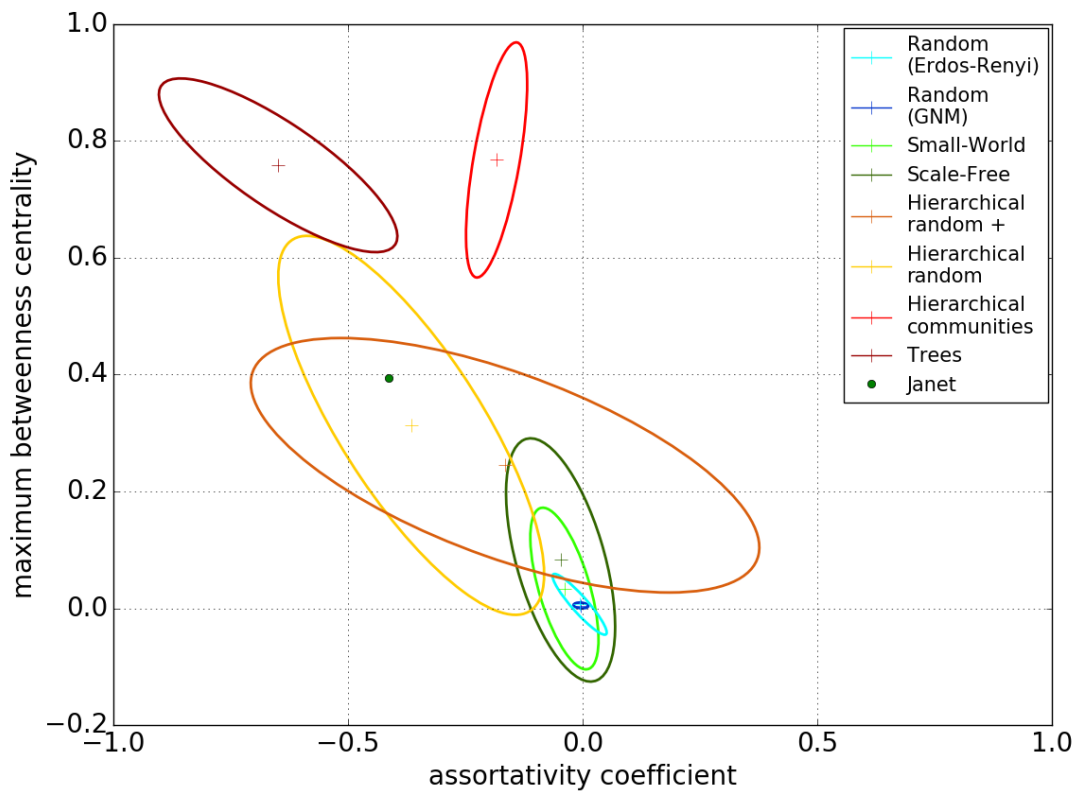


Figure 4.22: Metric result for the communication network for the assortativity coefficient and maximum betweenness centrality metric values with reference to the eight graph models.

A number of the infrastructure networks, all four river networks along with a two air networks (British Airways and EasyJet) and three rail networks (London Overground, Tyne and Wear Metro and Paris RER), return values for the AC and MBC metrics which suggest a similarity with the HC and TREE models. However, none of these actually lie within the ellipses, though the mean for the HC model is closest for six of the nine networks (Table 4.7), while the other three are closer to the TREE or HR models. This indicates that the aforementioned infrastructure networks analysed are hierarchical with a greater likeness to the hierarchical models than the non-hierarchical models.

Infrastructure	AC value	MBC value	Euclidean distance to HR	Euclidean distance to HC	Euclidean distance to TREE
Dee	-0.33	0.58	0.28	0.26*	0.37
Eden	-0.34	0.60	0.30	0.26*	0.35
Severn	-0.28	0.53	0.24*	0.29	0.44
Tyne	-0.34	0.66	0.36	0.21*	0.34
British Airways	-0.42	0.82	0.52	0.25*	0.25*
EasyJet	-0.45	0.51	0.22*	0.40	0.31
London Overground	-0.16	0.63	0.39	0.17*	0.52
Tyne and Wear Metro	-0.19	0.59	0.34	0.21*	0.49
Paris RER	-0.16	0.65	0.40	0.16*	0.51

Table 4.7: The AC and MBC values for the nine infrastructure networks closest to the HC and TREE mean AC and MBC values, with the Euclidean distance calculated. * denotes shortest distance.

The relationship between the AC and CB is shown in Figure 4.23 where all but the air networks show a greater similarity to the hierarchical graphs than the non-hierarchical graphs. The air networks in all but one case (British Airways), have returned values greater than the HR, HR+ and TREE with one of those, flights for Northern America, having similar values to the HC model. Table 4.8 shows the Euclidean distance to the mean of the WS, HR and HC graph models for the six air networks, indicating that many of the networks are in fact more similar to the HC model, with one similar to the HR and one similar, the British Airways network, to the WS model. Although five of the air networks have a greater number of CB, the AC values result in them being closer to the hierarchical networks, and in particular the HC model. This suggests that not only are these five networks hierarchical, they are similar to the most robust hierarchical network (Section 4.2.5, page 111).

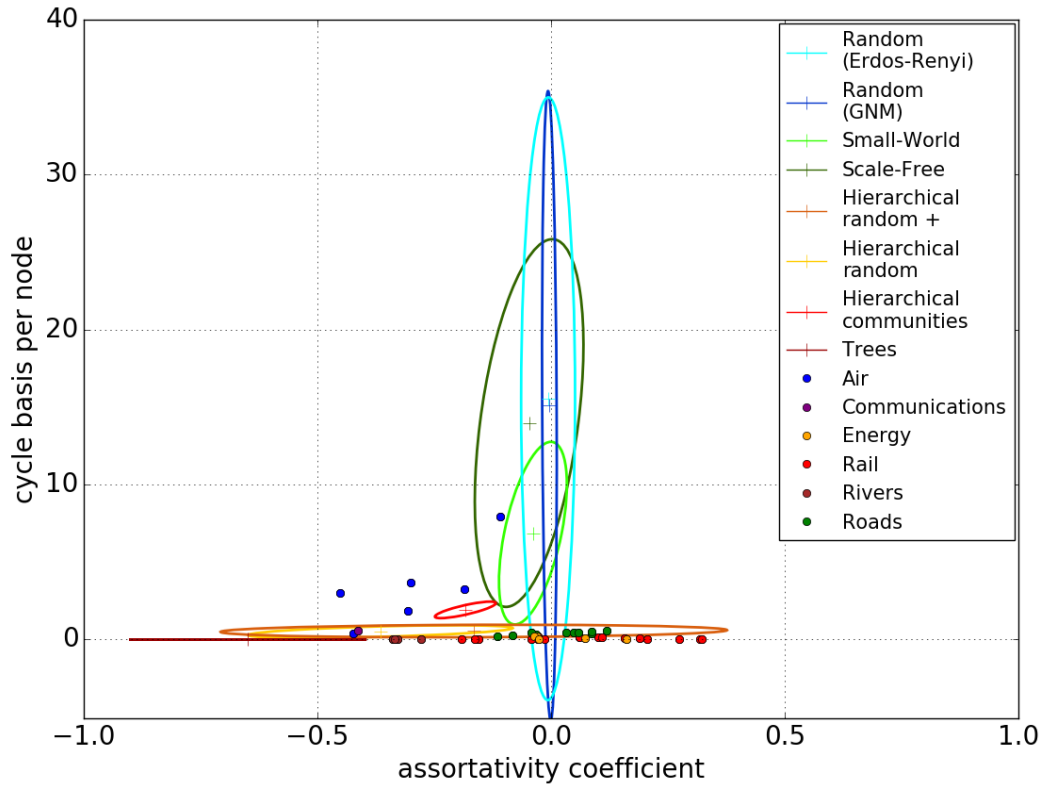


Figure 4.23: Single standard deviation ellipses for assortativity coefficient and number of cycle basis for each synthetic network type and each infrastructure.

Infrastructure	AC value	CB value	Euclidean distance to WS	Euclidean distance to HR	Euclidean distance to HC
British Airways	-0.42	0.38	6.50	0.14*	1.55
EasyJet	-0.45	2.99	3.89	2.48	1.11*
European	-0.11	7.93	1.07*	7.42	6.01
North American	-0.19	3.24	3.63	2.73	1.32*
USA	-0.30	3.68	3.19	3.17	1.77*
UK	-0.31	1.83	5.04	1.32	0.15*

Table 4.8: The AC and CB values for the six air networks and the Euclidean distance to the mean centres of the three graph models which are the shortest distance from each of the air networks, denoted by ‘*’.

All other in infrastructure networks returned values for the CB metric similar or lower than the HR, HR+ and TREE models (Figure 4.23). The previous analysis of the synthetic graphs shows that such low values reduce the robustness of the networks to perturbations (Section 4.2.5), suggesting the infrastructure networks with similar values may also share this characteristics of poor robustness.

Figure 4.24 shows the multivariate plot between the MBC and CB metrics. Nearly all of the infrastructure networks, with the exception of five of the six air networks, lie much closer to the ellipses of the hierarchical graphs than the non-hierarchical graphs. The five air networks all exhibit a greater number of cycle basis per node than the hierarchical graphs, with the average for the suite of air networks, 3.34, being greater than that for all of the other infrastructure networks, <0.55 (Table 4.9).

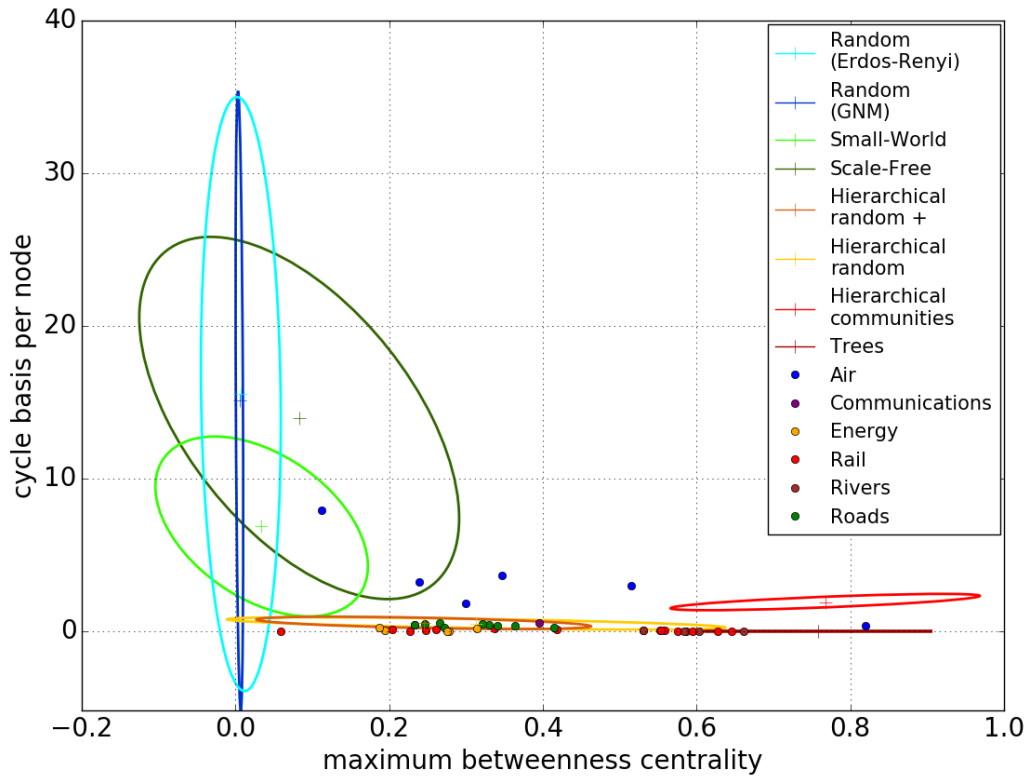


Figure 4.24: Single standard deviation ellipses comparing the relationship between maximum betweenness centrality and the number of cycle basis for the eight synthetic network types.

Infrastructure group	Assortativity coefficient \bar{x} (σ)	Maximum betweenness centrality \bar{x} (σ)	Number of cycle basis per node \bar{x} (σ)
Air	-0.30 (0.12)	0.39 (0.23)	3.34 (2.32)
Communications	-0.41 (0.00)	0.39 (0.00)	0.55 (0.00)
Energy	0.03 (0.08)	0.25 (0.05)	0.12 (0.08)
Rail – national	0.13 (0.10)	0.34 (0.14)	0.06 (0.01)
Rail - regional	0.06 (0.17)	0.45 (0.18)	0.06 (0.06)
Rivers	-0.30 (0.03)	0.57 (0.06)	0.01 (0.01)
Roads – national	-0.06 (0.02)	0.30 (0.03)	0.31 (0.04)
Roads – regional	0.02 (0.07)	0.31 (0.07)	0.42 (0.08)

Table 4.9: Average metric values for the infrastructure network groups.

4.3.4 Topological robustness of infrastructure networks

A comparison of how the infrastructure networks respond to the same failure perturbations as the synthetic graph models explored in Section 4.2.5 was undertaken in order to evaluate whether any infrastructure networks exhibited a similar response to any of the models. As with the synthetic graph models, three failure models have been applied to the infrastructure networks, with the results averaged across all networks of each particular group.

Figure 4.25(a) shows for the random node removal method that all infrastructure network types respond similarly, exhibiting a response broadly the same as the hierarchical graphs (Figure 4.5). The least robust infrastructure group is the river networks with only on average 62.99% of nodes needing to be removed to generate a null state; similar to the TREE graph model (Chapter 4, Section 4.2.5, page 111). Both the regional and national versions of the rail and road networks exhibit a similar robustness, with there being only a 1.5% difference between all road networks and 0.4% between all rail networks. With an average of 67.0% of nodes removed for the rail networks these behave most similarly to the TREE (68.2%). The road networks required on average 70.9% of nodes to be removed before failing, most similar to the HR model (73.4%). It must be noted that although the behaviour of the road networks is most similar to that of the HR graphs, this is only marginal, with the difference to the results for the TREE model only 0.2% greater.

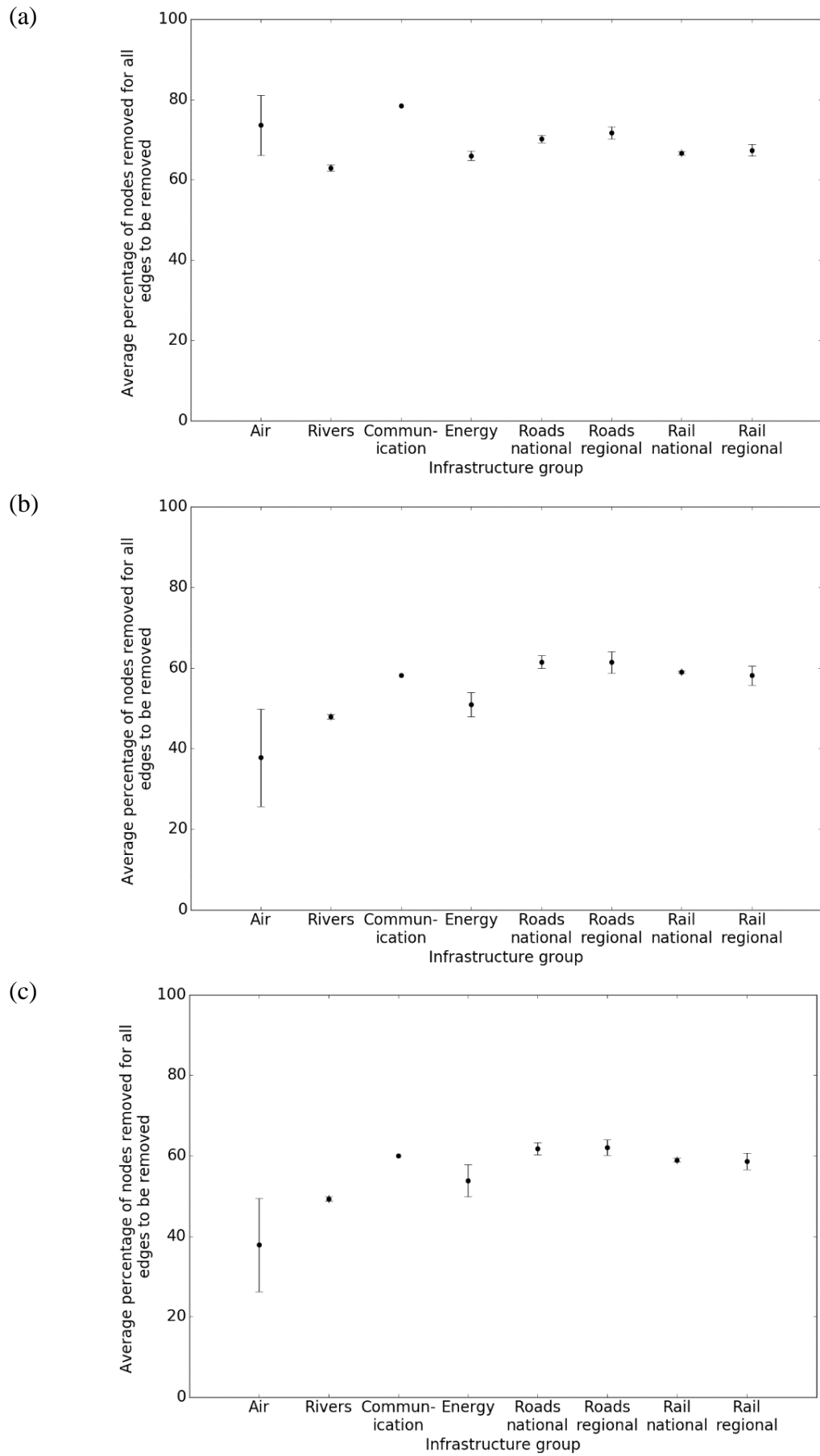


Figure 4.25: Plots showing the average response across the infrastructure groups to the three failure models. Failure plots for each infrastructure network can be found in Appendix E Section E.3.

In the case of the degree (Figure 4.25(b)) and betweenness centrality based failure methods (Figure 4.25(c)), the responses of the infrastructure networks show a similar pattern with all networks exhibited a decreased robustness to these node selection methods than for the random method. This was also the case for the synthetic models (Figure 4.5 Section 4.2.5), with both the hierarchical and non-hierarchical graphs all being less robust to the targeted methods.

It is clear that the air networks are the least robust to targeted attacks with only 37.78% and 37.85% of nodes needing to be removed for both approaches to become null. In the case of the river networks 47.98% and 49.28% of nodes needed to be removed before a null state was reached. The most robust networks are the roads, with the national and regional variations failing after 61.50% and 61.41% of nodes have been removed. A number of other infrastructure networks show similar values such as communications (58.16% and 60.00%), the national rail networks (58.98% and 58.94%) and the regional rail networks (58.16% and 58.63%). The infrastructure networks all fail after 50% - 60% of nodes have been removed (with the exception of the air networks noted above); a similar level as the majority of hierarchical networks (with the exception of the HC model) (Figure 4.5).

In general, the air networks exhibit a greater vulnerability to the failure models than all the other infrastructures, with the road networks appearing to have the greatest levels of robustness to failures. The river networks also exhibited a poor robustness to failures; an expected result given their similarity to tree networks, as discussed in Chapter 2 and highlighted in the degree distribution plots (Section 4.3.2). As expected none of the infrastructure networks behave like the random graph models (ER and GNM models) (Chapter 2, Section 2.3, page 11), or have a behaviour like that observed for the scale-free (BA) and small-world (WS) models, especially for the targeted failure models. Instead the majority exhibit a behaviour more similar to the HR+ model which has values of 74.71%, 52.42% and 52.37% respectively for the three failure models. This suggests that many infrastructure networks may have a hierarchical organisation given the results for the random, degree and betweenness failure methods.

The results from each infrastructure sector for both targeted methods are similar suggesting they have a similar effect on the networks with percentage of nodes needing to be removed similar. Both failure methods however identify and target nodes in a different order so a difference in response is expected, with the networks expected to fragment differently as the betweenness method removes the most critical nodes to the connectivity of the network first. The difference in response although is not clearly visible when the results are shown as above using the average percentage of nodes which need to be removed. However, the response is

more visible in the following section where the results are shown per network and the behaviour of the networks and how they fragment while being perturbed is plotted.

4.3.5 Failure characteristics of infrastructure networks

Given the results presented in the previous section, Section 4.3.4, more detail on how the individual infrastructure networks fail is presented in this section. This allows the behaviour of the infrastructures to be analysed in detail to examine how they fail rather than just looking at how long it takes for null networks to emerge. As in the case of the synthetic graph models (Section 4.2.6), only the results from a selected number of infrastructures are presented in this section. However, the full set of results can be found in Appendix E, Section E.3. The selected results include at least one infrastructure network from each of the infrastructure groups. Where this is considered not to be indicative of all networks within a group, multiple examples are presented.

Figure 4.26 shows how one of the six air networks responded to the three failure methods. The results presented in Figure 4.25 (Section 4.3.4) showed that the air networks were robust to random failures, but also the least robust of the infrastructure networks to the targeted methods. Again the air networks become fragmented very quickly, with $< 5\%$ of nodes needing to be removed for all three failure methods. The response to random failure is much worse than for the TREE example, with the peak in the number of components being much closer to that seen for the targeted methods, though the mean size of these components is much less at between 6-10% of the total size of the network, indicating that although the network fragments, the largest connected component remains large, with much smaller subgraphs forming around this. This results in the network being more robust with the greater proportion of the nodes in the network remaining connected despite the perturbations, a feature not seen in the tree network, or any other hierarchical network. The results for air networks again clearly indicate the network has a small number of highly connected nodes which make it vulnerable to the targeted failure methods, but robust to the random method. The small number of hub nodes in comparison to the number with much lower degrees mean the chances of these being removed is low resulting in the network staying connected for longer.

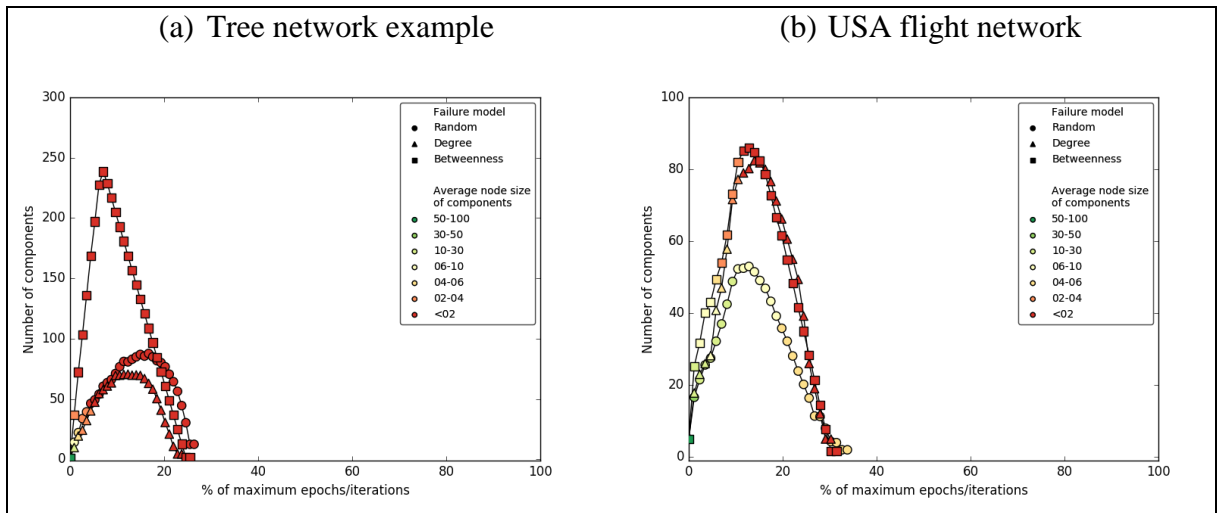


Figure 4.26: Two plots showing the behaviour of networks to the three failure models. The first shows the behaviour of a tree network (left) and (right) shows the behaviour of the flight network for the USA.

The river networks from the previous analysis suggest a hierarchical structure with degree distributions most like those for the TREE graphs (Section 4.3.2), and metric values similar to the HC/TREE graphs (Section 4.3.3). The response of the river networks is exemplified through the result of the River Dee (Figure 4.27), where the response appears to be similar to that observed for the HR graphs. Both networks fragment quickly from the outset, <5% of nodes removed, and fail completely when only 40-50% of the nodes have been removed. The River Dee does however not fail as quickly producing a more gradual failure, with the mean size of the subgraphs decreasing more slowly, to approximately 5% of the size of the network when 15-20% of nodes have been removed. In the case of the HR graph this occurs within 10% of the nodes being removed. This behaviour of the river network makes the response different to the TREE graphs (Figure 4.26).

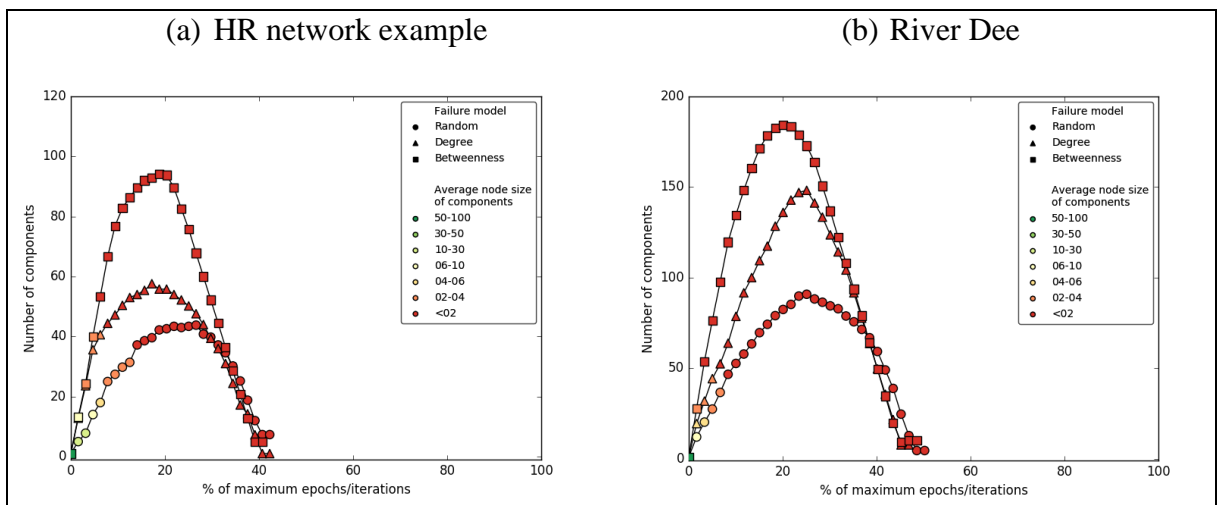


Figure 4.27: Comparing the behaviour to all three failure models the River Dee network (right) and that of a randomly selected HR network (left), highlighting the comparable similarities between the responses for both networks.

The behaviour of the energy networks suggests a better robustness to the three failure models compared to many of the other critical infrastructures (Figure 4.25). The behaviour associated with this poor robustness is exemplified in Figure 4.28 which shows the response of the full transmission network to the failure models. The plot shows how the network fragments at a slower rate than the results presented for many of the other infrastructure networks (Figure 4.29, Figure 4.26 and Figure 4.27), with the peak in the number of components in the network occurring after 30% of nodes have been removed (compared to 15-20% of nodes for other infrastructure examples presented). The profile shown in Figure 4.28 is different to those observed in the response of the non-hierarchical networks (Section 4.2.6, page 115) and instead is most similar to that for HR+ graphs. In contrast to this, the degree distribution of the energy networks was most similar to the BA model (Section 4.3.2), and had metric values most similar to the BA/HR graph models (Section 4.3.3). This set of results suggests that the topological structure of the energy networks are most like that observed for the HR graphs given both the similar behaviour observed to when perturbed and the metric results returned.

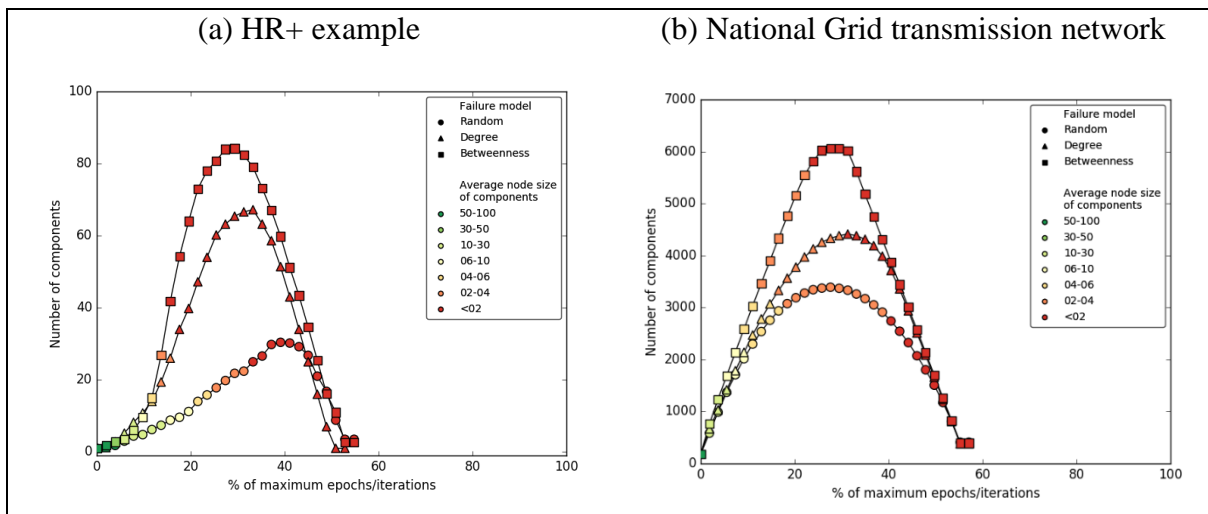


Figure 4.28: Showing the plot for the behaviour of the electricity transmission network for England and Wales (b) and the plot for a HR+ graph.

The rail networks all returned metric values which suggested they had a hierarchical topology, driven by high MBC values (Section 4.3.3), though exhibited degree distributions most similar to the WS model (Section 4.3.2). The failure response to the three perturbation methods, exemplified by the response of the London Tube network (Figure 4.29(b)) shows a response different from that exhibited by the WS (small-world) model (Figure 4.29(a)). Instead the response is more similar to that exhibited by the HR+ model example (Figure 4.28(a)). Both the Tube network and the HR+ example have a peak number of components after approximately 30% of nodes have been removed from the network, though the HR+ model appears to be more robust initially, with $> 10\%$ of nodes removed before the number of components starts to increase rapidly. However, the rail network does appear to be more vulnerable to the random failure method than the HR+ graph, with the greater peak in the number of components compared to those for the targeted methods. This suggests that although the degree distribution may suggest a non-hierarchical organisation, that the rail networks are actually hierarchical, with both the failure and metric results suggesting a similarity to the HR/HR+ models.

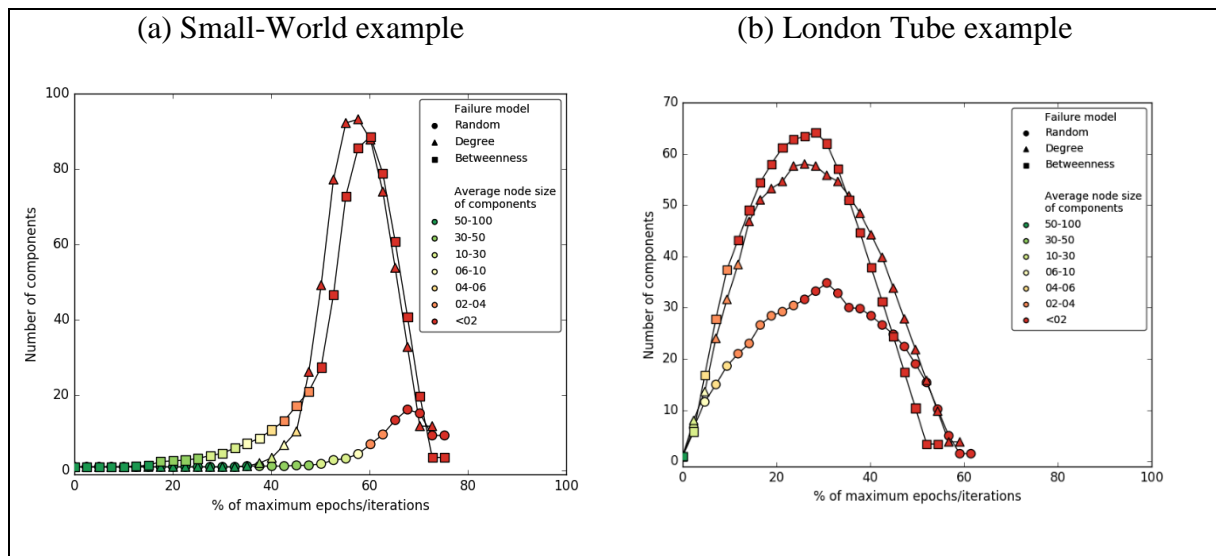


Figure 4.29: Failure plots for a Small-World graph (a) and the London Tube (d).

The road networks, exemplified by the response of the Irish road network (down to the trunk road level) (Figure 4.30(a)), exhibits a similar response to that of the Tube network (Figure 4.29(b)), and thus the HR+ response is the most similar from the synthetic graphs (Figure 4.28(a)). The shape of the response from the road network is similar, though the network appears to form components when it starts to fragment which are smaller than those formed in the failure of the HR+ graph. This could suggest that the network does not break in to large subgraphs, but instead the giant component of the network remains large but with many

components each with only a small number of nodes in forming. The metric values for the road networks also suggest that these networks were hierarchical (HR model), though the degree distributions of the network were unclear as to the topological structure of the networks.

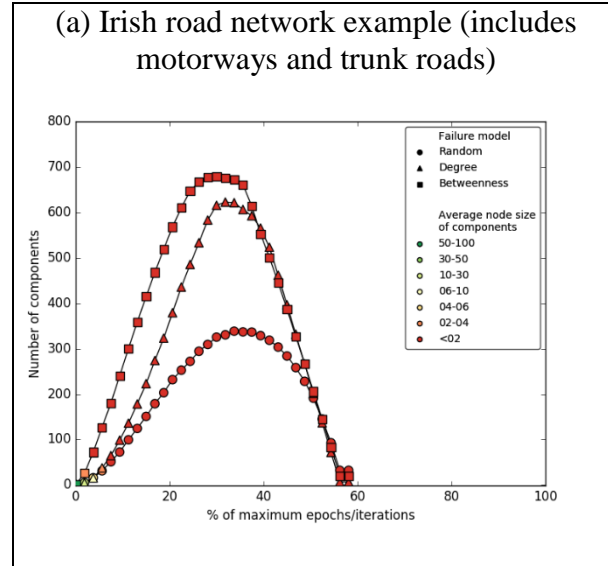


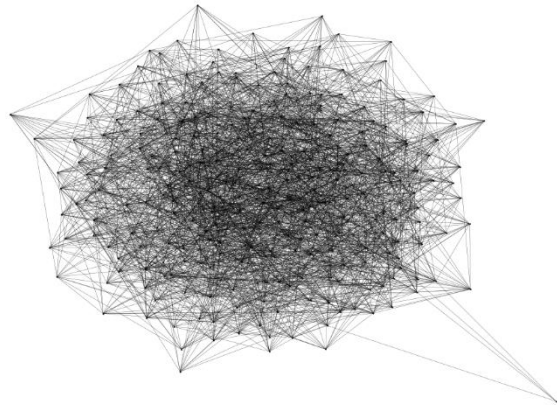
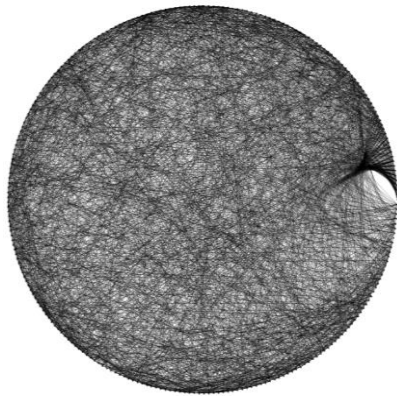
Figure 4.30: Failure plot the Irish road network including motorways and trunk roads.

4.4 Capacity constrained cascading failures on hierarchical graphs

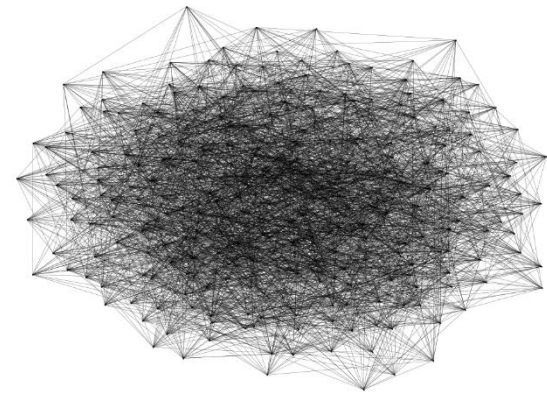
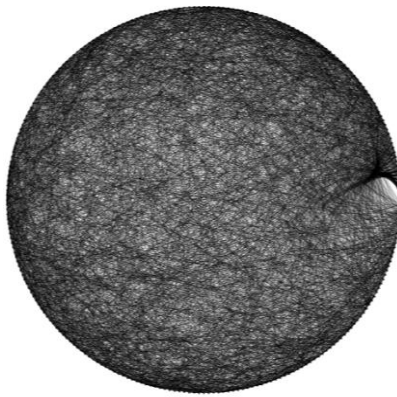
4.4.1 Introduction

The results in this section explore the effect of structure on the flows through a network and consequentially the robustness to perturbations. Previous sections have addressed the robustness of graphs/networks to topological based failures. However, the function of many infrastructure networks involves the flow of commodities and information or the delivery of a service between supply and demand points (Little, 2002), examples including the supply of electricity and gas (Bao *et al.*, 2009b), and water distribution networks (Shuang *et al.*, 2014). As such the attributes of a graph/network, including the capacity of the nodes/edges, are critical to how the flows can move over the network and as such effect the robustness of the graphs/networks to perturbations. A failure model has been developed (Chapter 3, Section 3.9) to explore how failures affect flows over a network and can lead to cascading failures through a graph. This failure model has been used to investigate the robustness of hierarchically structured graphs to cascading failures, with a set of eight synthetic graphs (Figure 4.31) employed corresponding to one graph from each of the eight graph models types.

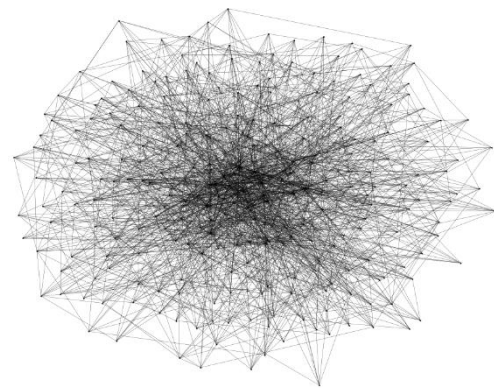
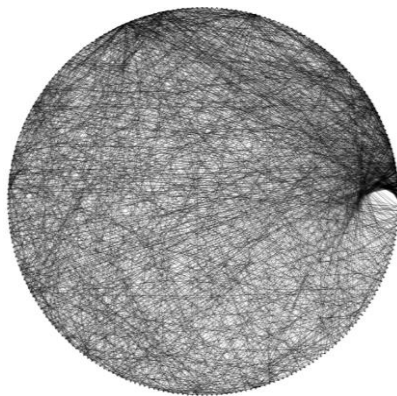
Erdos-Renyi (ER):



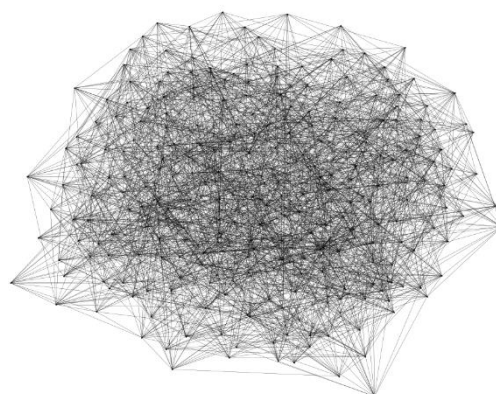
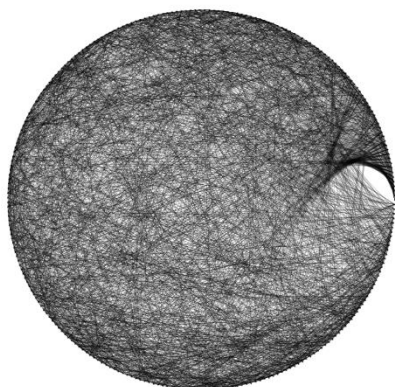
GNM:



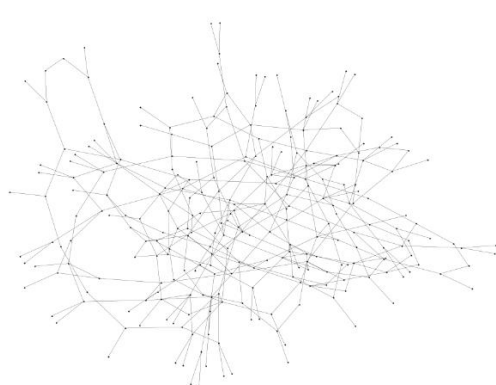
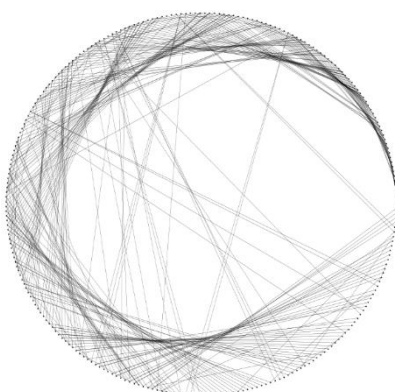
Barabasi-Albert (BA):



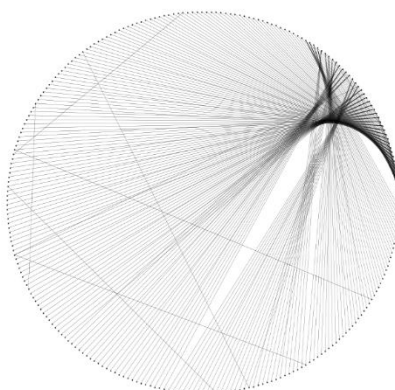
Watts-Strogatz (WS):



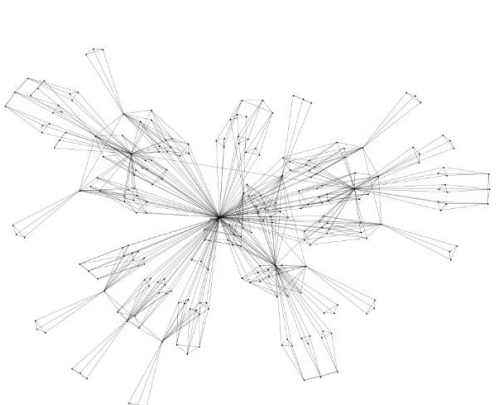
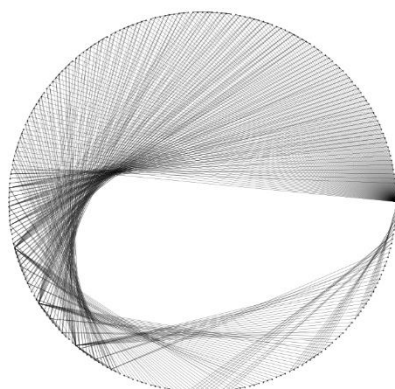
Hierarchical-Random (HR):



Hierarchical Random + (HR+):



Hierarchical Communities (HC):



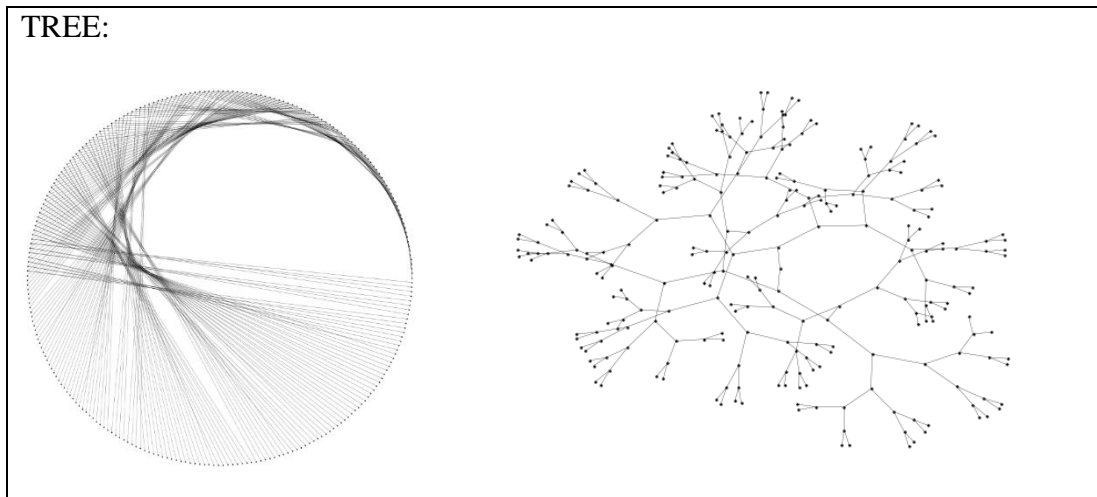


Figure 4.31: Synthetic network exemplars used in the analysis of robustness to cascading failure represented using a circular layout (left) and spring layout (right) which puts the most connected nodes at the centre of the plot.

4.4.2 Scenarios

The analysis is run using a capacity constrained cascading failure model, Chapter 3, Section 3.9 (page 71). For a graph a node is nominated at random as a supply node and a second node is nominated at random as a demand node with each assigned equal values so the supply in the graph always equals the demand. Each node and edge in the graph is then assigned a capacity, the maximum amount of the flow which can pass over the node/edge. The value assigned to the supply node is then routed, if possible, through the network to the demand node where the assigned capacity to the nodes and edges is used to constrain the number of flows which can use each node and edge. The simulation then attempts to trigger a cascading failure by removing the edge with the greatest flow, known as the trigger edge. Where multiple edges share the greatest flow value, one of these is picked at random.

Following the removal of the trigger edge, the routing of flows is performed again. As long as a route is available, irrespective of the capacity of the nodes and edges, the simulation continues. Where the route or routes found have sufficient capacity, the nodes and edges used are not over capacity, the network is said to be in a state of equilibrium. This describes the network as still being able to function following the perturbation.

Where at least a single route exists between the supply and demand nodes, the flows are routed along this, though when there is insufficient capacity along a single route the flow will be distributed across multiple routes if possible. If the flow to be routed cannot be accommodated on all possible routes without node and/or edge capacities being exceeded, the flows are assigned to the shortest route resulting in some nodes and/or edges being over capacity. Nodes

and edges over their capacity are regarded as failed in the next epoch, simulating a cascading failure. In the following epoch, the route searching process is repeated with the nodes and edges over capacity identified and tagged as failed. Only once no nodes or edges are found to be over capacity or no route exists between the supply and demand node does the simulation stop.

As only one supply node and one demand node are assigned in each graph, the analysis is sensitive to the location of the selected nodes, and therefore each node selection is at random for each of the 5 simulations run for each scenario for each graph. This attempts to negate the potential bias of any specific node locations may have on the results, either positively or negatively for the robustness of the graphs being analysed. It should also be noted that the only flows on the graphs are those assigned between the supply and demand nodes, with no other flows on the graphs, therefore simplifying the analysis undertaken.

The eight graphs, Figure 4.31, were analysed using six developed scenarios (Table 4.10) to investigate the robustness of hierarchical graphs to cascading failures. Scenario (i) investigates the robustness of the hierarchical graphs to simple cascading failures through assigning node and edges capacities equal to that of the supply/demand in the graph. A value of four has been used for this, but the value itself is immaterial as long as the capacity of the nodes and edges is equal the supply/demand. A single trigger is removed to analyse the robustness of the graph to a single failure. With each node and edge having the capacity to accommodate all the flow the inability of the flow to reach the demand from the supply node following the single perturbation highlights a poor robustness in the graph, indicating the network has fragmented into at least two components. Scenario (ii) is parameterised similarly to scenario (i), though the capacities of the features are assigned based on the graph model as described in Chapter 3, Section 3.9.4 (page 78), rather than with a uniform value. Nodes and edges within the hierarchical graphs are assigned capacities based on the level of the hierarchy they are in, whereas the random graphs are assigned capacities randomly. Scale-free and small-world graphs are assigned values based on the betweenness centrality of the nodes and edges. The assigned capacities all lie between 1 and 8. This allows for more graph structure specific simulations to be undertaken, and is an attempt at assigning capacities which are more closely related to the role the nodes and edges play in the topology of the graph.

Scenario	Method of node and edge capacity assignment	Number of trigger edges removed
(i)	Capacities equal to supply value	Single trigger edge removed
(ii)	Capacities assigned based on graph model	Single trigger edge removed
(iii)	Capacities equal to supply value	Trigger edges removed until cascading failure triggered
(iv)	Capacities assigned based on graph model	Trigger edges removed until cascading failure triggered
(v)	Capacities lower than supply value	Single trigger edge removed
(vi)	Capacities lower than supply value	Trigger edges removed until cascading failure triggered

Table 4.10: Scenarios employed in the modelling of cascading failures.

Scenarios (iii) and (iv) investigate the robustness of the graphs to cascading failures by removing trigger edges until a cascading failure is triggered or until no path exists between the supply and demand nodes (Table 4.10). These two scenarios are otherwise the same as scenarios (i) and (ii) respectively. This analysis allows for those graphs which are robust to the removal of a single trigger edge to be analysed further to identify the strength of the robustness. This is reported using the proportion of edges removed as trigger edges until a cascading failure is triggered or the supply and demand nodes become disconnected. Through this the strength of the robustness across the hierarchical and non-hierarchical graphs can be compared. This analysis allows for the identification of those graphs which exhibit the greatest robustness to cascading failures and the characteristics which facilitate this.

The final two scenarios, (v) and (vi) explore the robustness of the graph models where the supply/demand value is greater than the capacity of the nodes and edges in the graph (Table 4.10). For each graph this means multiple routes must exist between the supply and demand nodes, unlike in the previous scenarios where only a single route was required. This requires the graphs to be better connected with greater redundancy if they are to be robust to cascading failures. For both scenarios a capacity of four is assigned to the nodes and edges, whereas a supply of six is assigned to the single supply node and a demand of six to the single demand node. In scenario (v) only a single trigger is used, but in scenario (vi) multiple triggers are removed to identify the most robust graphs to the cascading failures where each node and edge lacks sufficient capacity to accommodate all the flow between the supply and demand nodes. Five simulations have been run over each graph for each scenario.

The performance of graphs to cascading failures is measured in three ways; the reason for each simulation terminating, the average length of cascading failure and for those scenarios where

multiple trigger edges are removed, the average number of trigger edges removed over the simulations. Simulations can terminate for three separate reasons, the first being due to a lack of a route of sufficient capacity between the supply and demand node before a trigger edge has been removed, referred to as ‘Not computable’. Where the flow has been routed from the supply to a demand node without any nodes or edges being over capacity after the removal of trigger edge, the graph is said to be in equilibrium. The third reason for a simulation terminating arises when the supply and demand nodes become disconnected from each other, such that they are in different component/subgraphs of the original graph, which can be a result of trigger edges being removed or from the failure of nodes or/and edges being over capacity. The second method of reporting the robustness from this analysis is the average length of the cascading failures, measured by the average number of epochs across the five simulations for each graph. In those scenarios where multiple trigger edges are removed, the number of these is recorded, and then averaged for each simulation for each graph model, providing a third metric on the robustness of hierarchical graphs to cascading failures.

4.4.3 Cascading failures results

The results for scenario (i) (Figure 4.32) shows the reason for the each simulation stopping as a percentage across the five simulations run for each graph and the average length of the cascading failure for each graph analysed. Of the four hierarchical graphs the TREE graph was the least robust, failing in 100% of simulations after the removal of a trigger edge, with no route between the supply and demand nodes (Figure 4.32) and the network becoming disconnected. The HR and HR+ graphs however exhibited a greater robustness to the removal of the trigger edge with 80% and 40% of simulations reaching equilibrium, with a path with sufficient capacity still existing between the supply and demand nodes. In 20% and 60% of the simulations however the network failed with no route available between the supply and demand nodes for the HR and HR+ graphs. This is a result of a single route existing between the locations of the supply node and demand node, which is then broken when the trigger edge is removed. These results indicate that given the random allocation of supply and demand nodes, that the HR and HR+ graphs are more robust than the TREE graph. This is likely a result of the better connectivity within the HR and HR+ graphs than in the TREE graph as a result of the addition of new edges in the graph models for these graphs (Chapter 3, Section 3.3.5 and 3.3.6). The HC graph exhibited a mixed response, with 60% of simulations reaching equilibrium and 40% resulting in no path from the supply to the demand node. This shows a similar behaviour to the HR and HR+ graphs, indicating that the network is reasonably well connected with in

some cases multiple routes which can accommodate the flow. However, in contrast, the non-hierarchical graphs exhibit a much greater robustness with 100% of simulations resulting in the graphs being in equilibrium. This is a result of the greater connectivity of these graphs providing an increased level of redundancy and thus robustness to perturbations, with in this case this resulting in multiple paths between the supply node and demand node.

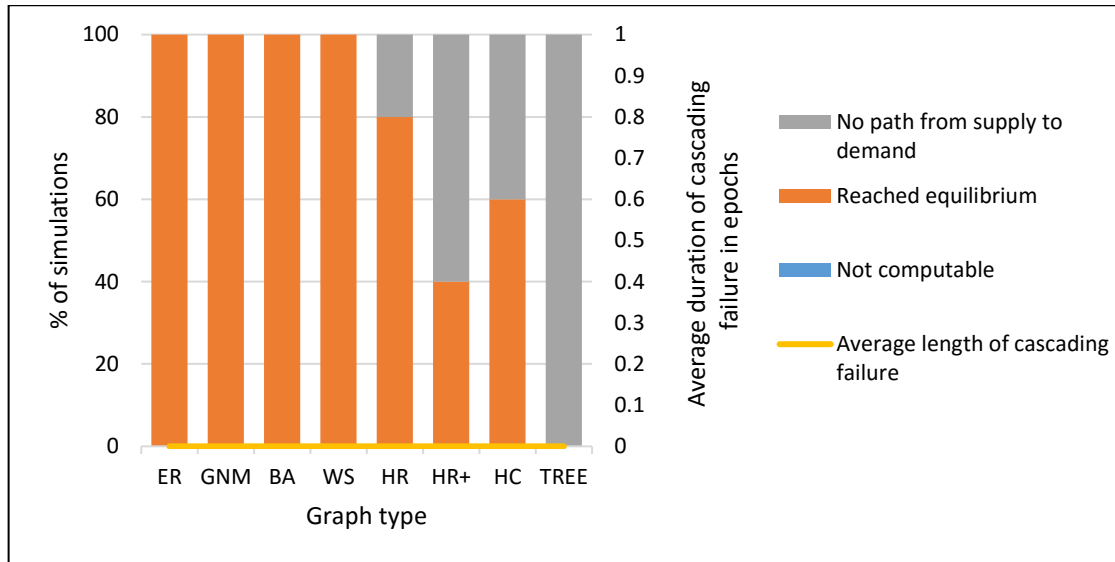


Figure 4.32: Results from cascading failure simulation (scenario (i)) over the synthetic network exemplars with a single trigger edge, a demand of four and capacities of four.

Scenario (ii) was parameterised similarly to scenario (i), but with node and edge capacities assigned based on the graph structure. The results (Figure 4.33), show that three of the hierarchical graphs, the HR, HR+ and TREE, could no longer be solved with 100% of the simulations finding no path with sufficient capacity between the supply and demand nodes even before a trigger edge was removed. It is strikingly clear that these graphs obviously lack the capacity to accommodate the flow from the supply node to the demand node. However, the fourth hierarchical graph model, the HC model, exhibited a greater robustness with the network not being affected by the removal of a single trigger edge in 80% of simulations. In the one simulation where the HC graph failed following the removal of the trigger edge, it is likely the location of the supply/demand node was such that they were poorly connected and the trigger edge selected on this occasion resulted in the only path between the two nodes being broken. However, due to its modular structure flows within communities of nodes that are well connected are supported. Thus, this model seems to be sensitive to failure on the edges that connect the different communities within the graph.

The four non-hierarchical graphs in scenario (ii) (Figure 4.33), exhibited a similar response to that observed in scenario (i). Three of the four graphs, those generated by the ER, GNM, and BA models all exhibited a strong robustness to cascading failures with all ending in a state of equilibrium. However, the WS model in one simulation failed following the removal of the trigger edge but after a cascading failure of 18 epochs (5×3.6 average epoch length of cascading failures), while the other four simulations were not affected by the removal of the trigger edge. This indicates that the graph was connected for much of the simulation with redundancy allowing solutions for the demand to be met. However, the capacities of the nodes and edges on the shortest path were not great enough to accommodate the flow, resulting in the cascading failure until no more routes were available (as demonstrated by the example in Figure 3.25, page 76). The only differentiator between the five WS simulations is the location of the supply and demand nodes, which highlights the sensitivity of the analysis to the location of these, and therefore the structure of the graph as well.

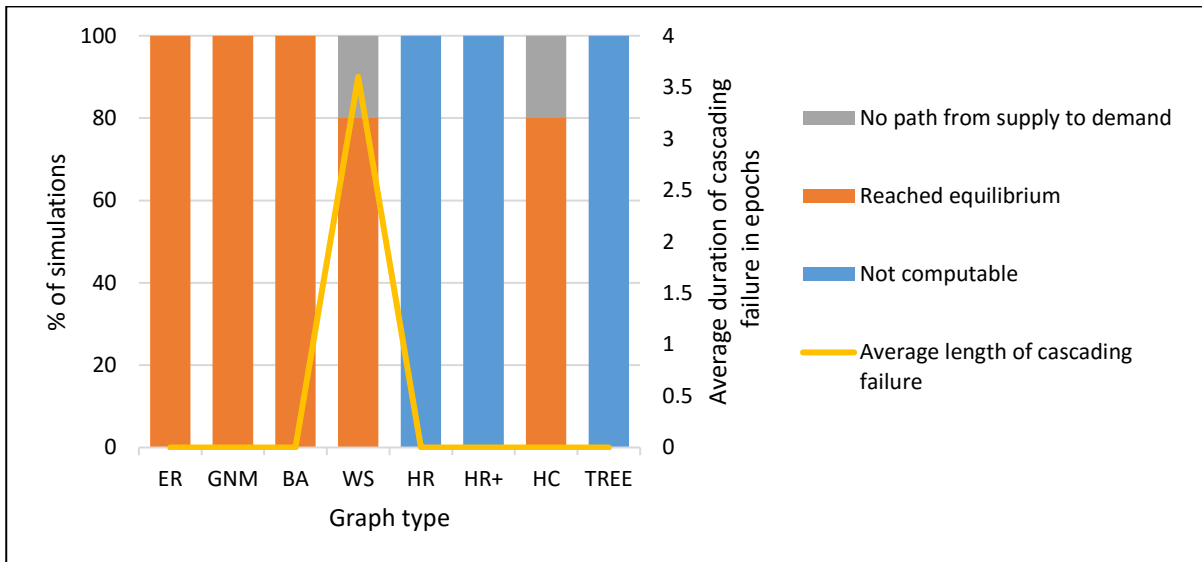


Figure 4.33: Results of the cascading failure simulation (scenario (ii)) over the selected synthetic networks where the demand was four, the capacities were based on structure and a single trigger edge was removed.

The results for scenario (iii) are presented in Figure 4.34 and Figure 4.35 which shows the proportion of edges removed before the failed graphs. For each graph, as with scenario (i), the capacity of the nodes and edges was set to four, with the supply and demand both set to four. The supply and demand nodes were also assigned randomly. Unlike scenario (i), trigger edges were removed until a cascading failure started or there was no longer a route between the supply and demand nodes. This allows the most robust graphs to cascading failures to be identified, with the least robust failing after a smaller proportion of edges have been removed and the

robust graphs not suffering from a cascading failure and only failing after a greater proportion of edges removed. For all graph models, 100% of the simulation runs failed with no path from the supply and demand node existing and no cascading failures occurring, suggesting a robustness to cascading failures (Figure 4.34). Figure 4.35, which shows the percentage of edges removed until the networks failed, shows that the hierarchical graphs, with the exception of the HR model, were the least robust to the removal of trigger edges. The HR+, HC and TREE model all failed with <0.8% of edges removed, whereas the non-hierarchical networks, as well as the HR model, failed after 2.0-4.9% of edges had been removed. Through the shortcuts added to the TREE graph to form the HR graph being unconstrained in terms of the nodes they could connect, these have added a much greater redundancy into the network. Whereas for the HR+ model the edges were only added to nodes in the same level and those in adjacent levels, thus the impact of these on the robustness of the network seems to be much less. The BA model behaves differently to the other graph models, failing after 2.0% of edges have been removed, rather than 4.3-4.9% as for the ER, GNM and WS models, likely due to a greater dependency on hub nodes for flows passing over the network, as indicated by the higher maximum betweenness centrality compared to the other three non-hierarchical models.

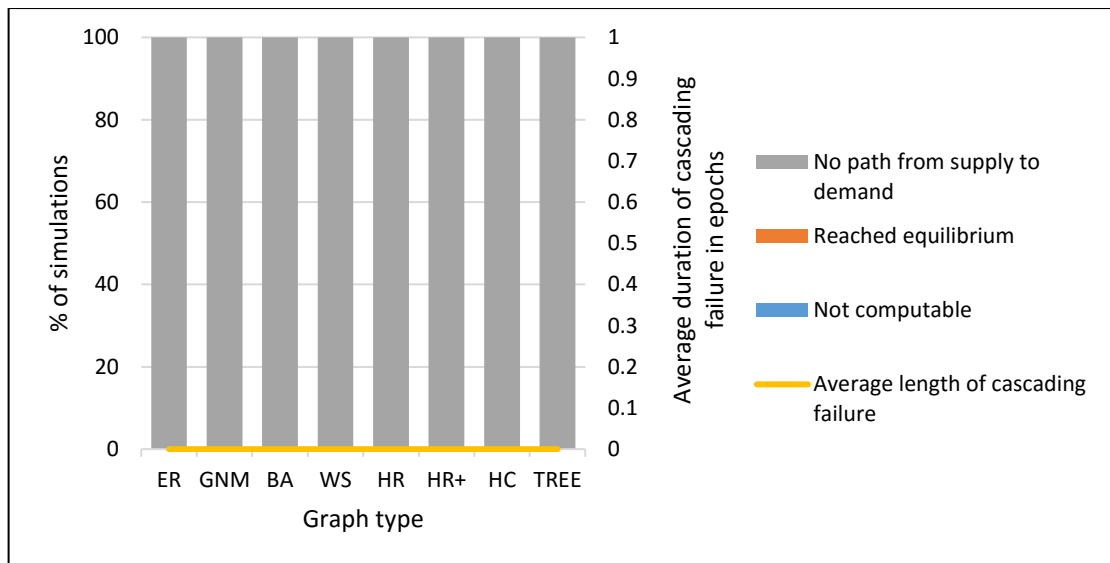


Figure 4.34: Result of the cascading failure simulations (scenario (iii)) on the selected synthetic networks where a supply of four has been used, with capacities set to four and multiple trigger edges removed.

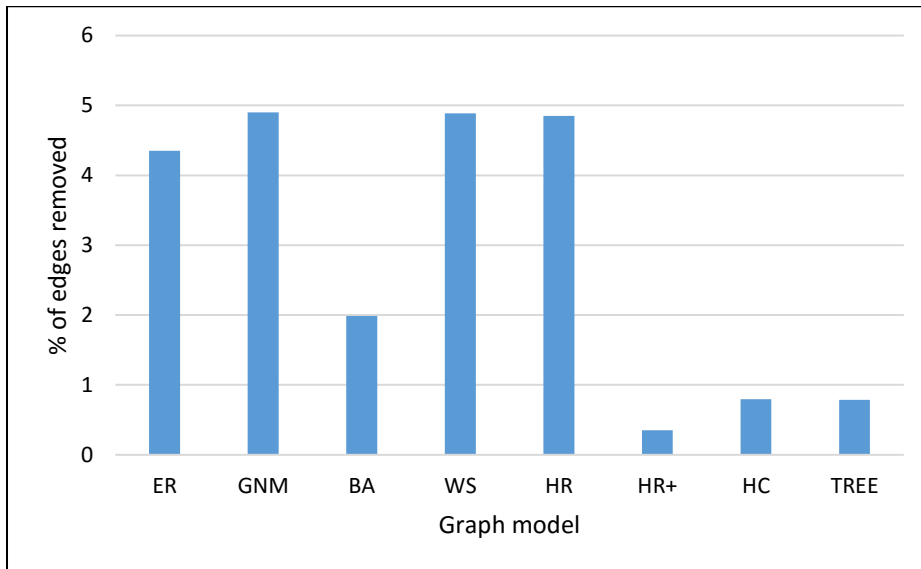


Figure 4.35: Percentage of edges removed as trigger edges for the analysed synthetic graphs for scenario (iii).

Figure 4.36 and Figure 4.37 show the results from scenario (iv). This has been parameterised similarly to scenario (iii), but with node and edge capacities based on the graph structure. The results presented in Figure 4.36 show that for 100% of the simulations with the HR, HR+ and TREE models no path between the supply and demand nodes existed with sufficient capacity to accommodate the flow of four from the supply node to the demand node. This is a result of the parameterisation and the redundancy in the graphs which is non-existent in the TREE graph, and is clearly limited in the HR and HR+ graphs. The HC model fails with an average cascading failure value of 0.6 epochs over the five simulations, showing a cascading failure of either 6 epochs occurred in one simulation, or 1 epoch in three. This short cascading failure suggests that the edges or nodes removed were part of the only path between the supply and demand. These results suggest that the HC model is more robust than the other three hierarchical models which didn't have the capacity to facilitate the flow from the supply to the demand node before any edges were removed. However, the HC graph failed after an average of 2.2 edges (0.3%) were removed Figure 4.37, which suggests the graph is vulnerable.

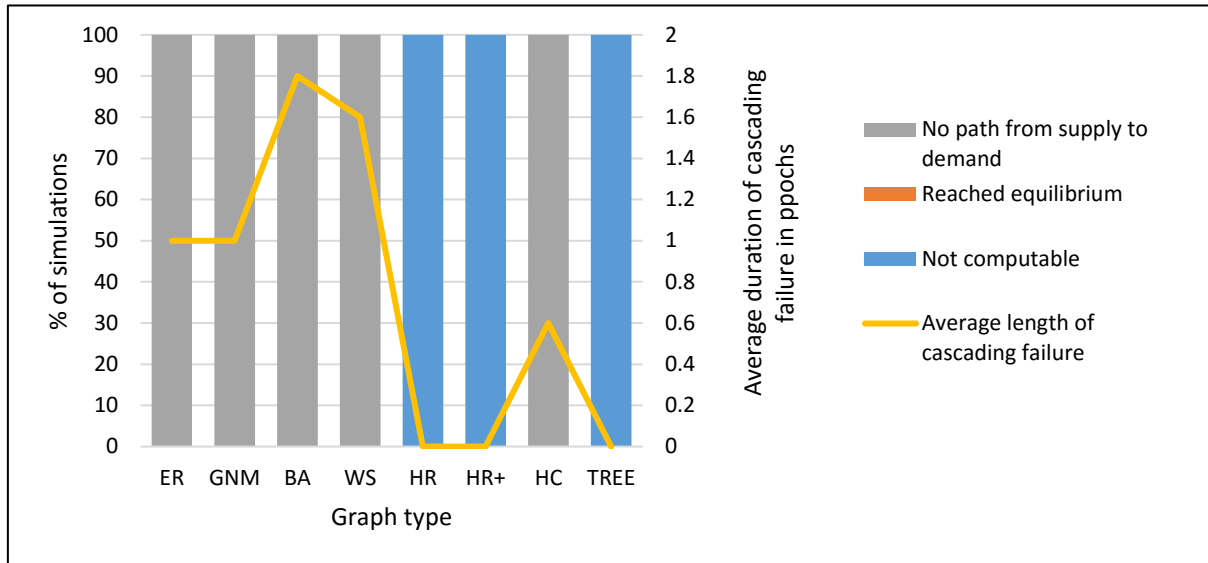


Figure 4.36: The results from the cascading failure simulations (scenario iv) on the selected synthetic networks where a single demand of four was used, capacities were based on the graph structure and multiple trigger edges were removed.

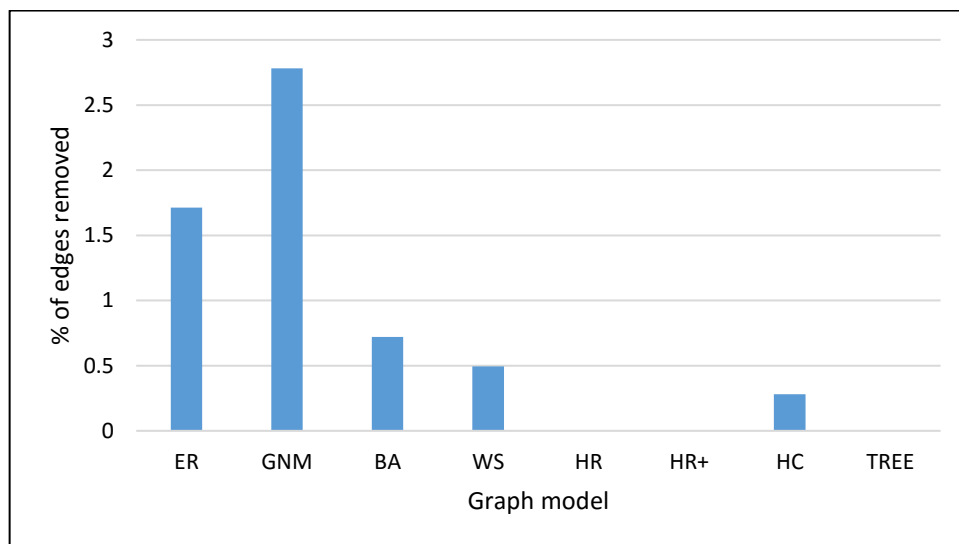


Figure 4.37: Percentage of edges removed as trigger edges for the cascading failure scenario (iv).

In contrast the non-hierarchical graphs have demonstrated a greater robustness with a greater proportion of edges needing to be removed before the graphs failed; for the ER and GNM graphs this was 1.7% and 2.8%, though for the BA and WS models this was only 0.7% and 0.5%. However, all four graphs exhibited cascading failures, with those for the BA and WS models averaging at 1.8 and 1.6 epochs respectively. This indicates that although only a small proportion of edges were removed as triggers, many more may have been removed as a result of being over capacity during the cascading failure. More generally, these results clearly

demonstrate the robustness of the two random models to these type of failures, with a large percentage of edges removed in comparison to the other graph models, likely a result of greater connectivity, and hence redundancy. This is linked to the previous metric analysis of the graph models which indicated that the random models both have a greater number of cycle basis and a lower maximum betweenness centrality (Section 4.2.3).

The result of the simulations for scenario (v) are shown in Figure 4.38, where the graphs have been parameterised so the supply and demand in the graphs is six, though the capacity of the nodes and edges is four. The supply and demand nodes have been selected at random and only a single trigger edge is removed from the graphs. For both the HR+ and TREE graphs there was insufficient capacity on the available routes between the supply and demand nodes for the required flow from the outset, with 100% of the simulations not being computable. This makes it strikingly clear that these graphs lack connectivity and the redundancy to continue to function when perturbed. For the HR graph it was also the case that 60% of the simulations could not be computed, but due to the random location of the supply and demand nodes, solutions could be computed for 40% of the simulations, with 20% of these failing after the removal of a trigger edge and 20% remaining computable with enough capacity on paths between the supply and demand nodes. This reveals that the HR graph is more robust than the HR+ and TREE models.

However, the HC model is more robust than the HR+ and TREE models, only not being computable in 40% of simulations due to insufficient capacity between the supply and demand node, compared to 100% for the HR+ and TREE models. In 60% of simulations the graph reached equilibrium meaning there was sufficient capacity between the two nodes as well as some redundancy with this still being the case following the removal of a trigger edge. In contrast to the hierarchical graphs, the non-hierarchical graphs all were robust to the single failure, with 100% of simulations reaching equilibrium for each of the four models, further highlighting the clear differences between the two sets of graphs. The non-hierarchical graphs clearly have a greater connectivity and redundancy and thus are robust to the greater flow on the network, with multiple routes between the supply and demand nodes.

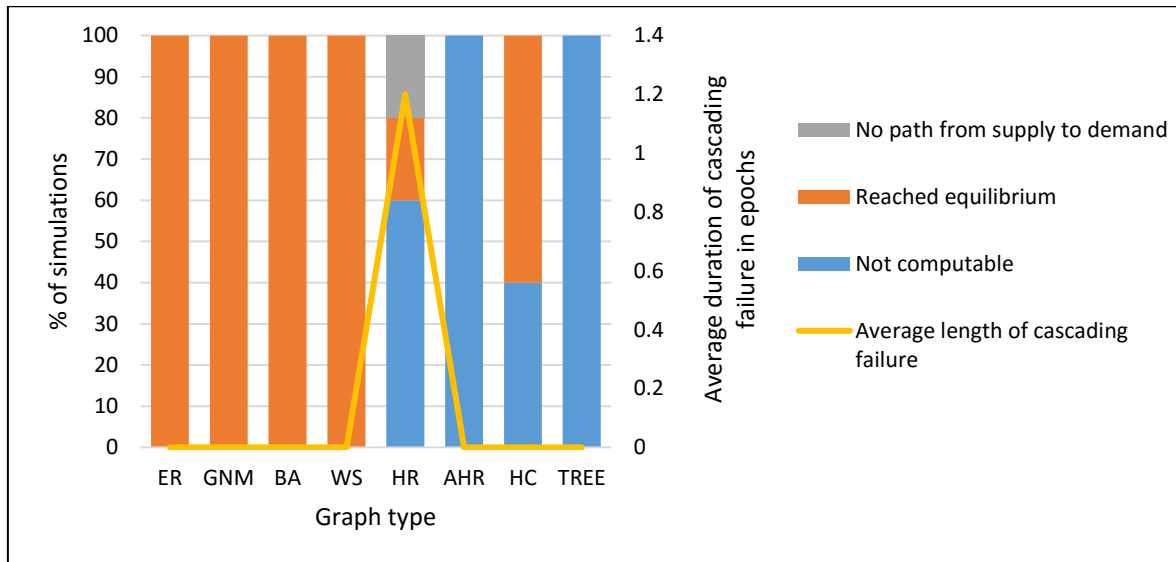


Figure 4.38: The results from the cascading failure simulations (scenario (v)) over the selected synthetic networks where the demand is six, the capacities are four, and a single trigger edge is removed.

Scenario (vi) runs a similar set of simulations as done in scenario (v), but with trigger edges removed until a cascading failure is triggered or there is no longer a path between the supply node and demand node, to identify the most robust graph model to cascading failures. The results are presented in Figure 4.39 and Figure 4.40 which respectively show the performance of the graphs to the scenario and the proportion of edges removed for the graphs to fail or a cascading failure to start. 100% of the simulation for the HR+, HC and TREE graphs show that insufficient capacity was available between the supply and demand nodes to accommodate the required flow. Although the node and edge capacities and the supply/demand in the graph are parameterised the same as in scenario (v), the random location of the nodes has had a significant effect with no simulations being computable. The HR graphs however were computable in 60% of the simulations, highlighting again the greater availability of routes between the supply and demand nodes. In contrast to the hierarchical graphs 100% of the simulations for the four non-hierarchical graphs were computable. As well as being computable, a cascading failure of one epoch was seen in each simulation for the non-hierarchical graphs. This is a result of when two routes exist between the demand and supply node; one of these is broken by the removal of trigger edge resulting in the remaining single route being overloaded and thus failing in the next epoch and resulting in no routes between the supply and demand nodes. Figure 4.40 shows that the random graph models, ER and GNM, are the most robust with 3.8% and 3.7% of edges being removed before there was insufficient capacity between the supply and demand nodes. In the BA and WS models, 2.0% and 3.3% of edges were removed showing a lower robustness

to cascading failures than the two random graphs. The HR model failed with a mean of 2.3% of edges removed indicating that it is more robust than the non-hierarchical graph with the exception of the BA graphs, which failed with 2% of edges removed, though in absolute terms 7.6 edges were on average removed from the HR graph and 38.6 on average from the BA graph.

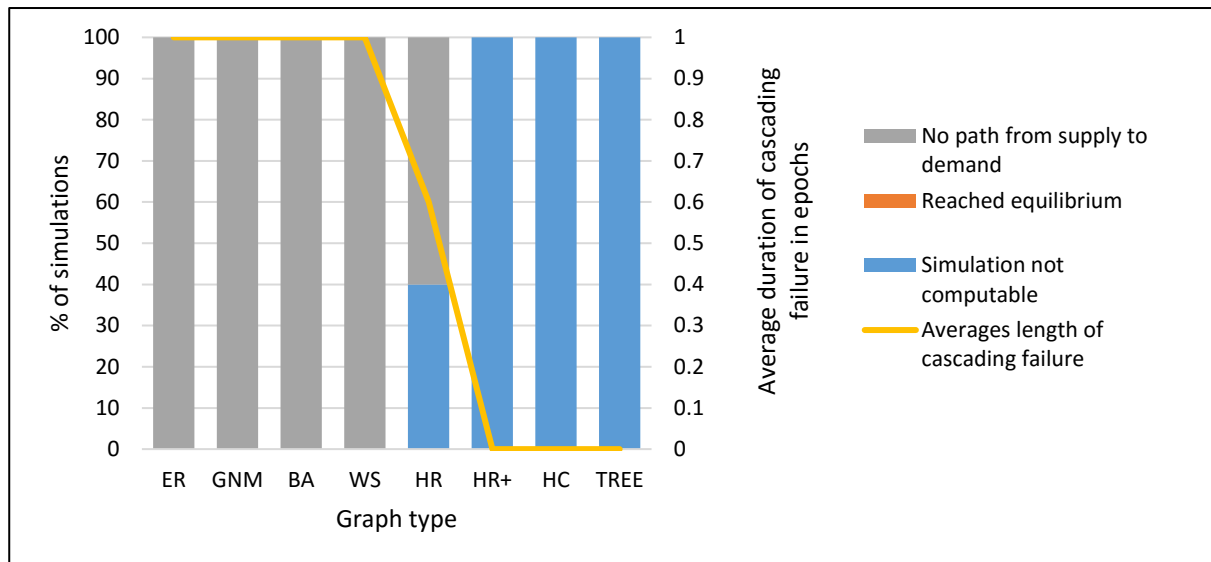


Figure 4.39: The results from the cascading failure simulation (scenario (vi)) over the selected synthetic networks where the edge capacities were set to four, the demand six and multiple trigger edges were removed.

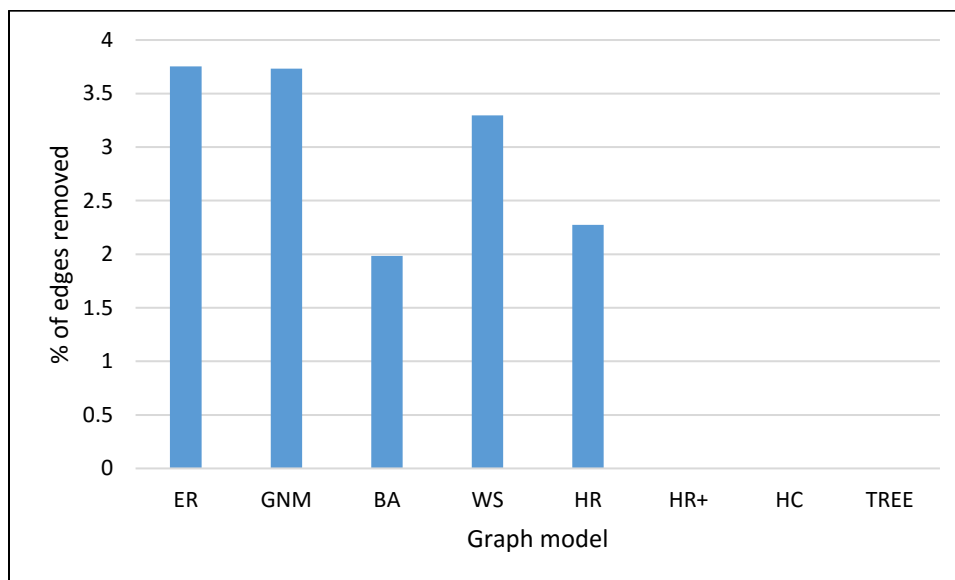


Figure 4.40: Percentage of edges for cascading failure scenario (vi) removed from the graphs.

4.5 Robustness of a hierarchical critical infrastructure network

4.5.1 Introduction

There is building evidence of critical spatial infrastructures having a hierarchical structure, from the electricity network (Chang and Wu, 2011) to the road network (Yerra and Levinson, 2005). Such infrastructure networks use a hierarchy to transmit/support the delivery of flows over the network between two locations. At the centre of this is highest level of the hierarchy, be that the high voltage transmission network for electricity (Chang and Wu, 2011) or the motorways/highways in the road network (Yerra and Levinson, 2005).

The electricity transmission and distribution network for England and Wales is used as an example of an infrastructure network which has a hierarchical organisation, with energy flowing from the power stations down to the substations (Chang and Wu, 2011). Using the electricity transmission network two forms of failure analysis are undertaken to explore the robustness of this hierarchical network. The first section, Section 4.5.2 explores the robustness of the hierarchical electricity network to multiple failures at the highest level of its hierarchy, the 400Kv transmission edges, exploring the robustness of the network to failures at the highest level of its structure. Three scenarios are investigated including the failure of one, two and three 400Kv transmission line simultaneously. The second analysis undertaken, Section 4.5.3, investigates the robustness of the network to different configurations of spatial hazards, from a single large hazard, to eight smaller hazards but which affect the same proportion of network assets. Hazard areas are spatially generated randomly to affect 2% of the network in total in each scenario. This allows for an analysis of how the number of hazards on a hierarchical network affect its ability to function and how the spatial configuration of these also affects the infrastructure network.

4.5.2 Hierarchical dependency of the electricity network

The electricity transmission network, shown in Figure 4.41 and described in Chapter 3, Section 3.9, has 323 400Kv or higher transmission lines (Table 4.11), forming the highest level in the hierarchy of transmission and distribution network.

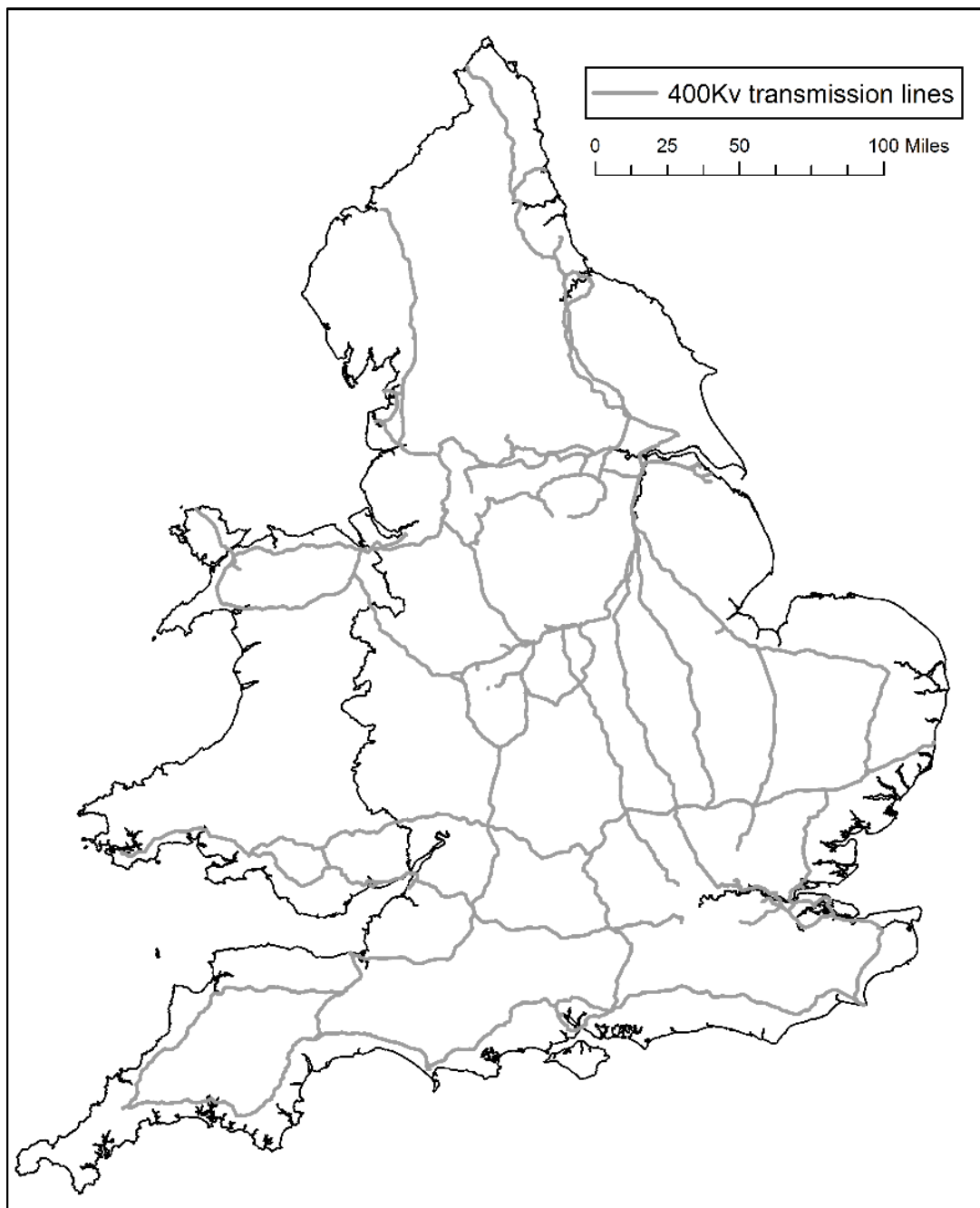


Figure 4.41: 400Kv transmission line network for England and Wales.

Role/voltage (Kv)	Number of nodes (substations)	Number of edges
Generators	156	NA
400	161	323
275	123	201
132	1146	1361
66	19	19
33	4808	5781
25	2	7
22	2	1
11	164090	164090
0 (retired)	162	32
Generator edges (connect generators to the network)	NA	204
Logical (connect co-located substations)	NA	1020

Table 4.11: A breakdown of the nodes and edges in the electricity transmission and distribution network for England and Wales.

Failures in the full electricity transmission and distribution network (Figure 3.27, page 82) are simulated using three scenarios, with the number of the edges removed in the highest level of the hierarchy in the network increasing with each scenario. In scenario (i) each 400Kv edge is removed, of which there are 323, in turn and the connectivity, through the hierarchy of the network to the 11Kv substations is checked to ensure each is still connected to the 400Kv transmission network. Scenario (ii) removes every possible pair of 400Kv transmission edges again checking that each 11Kv substation is still connected to the 400Kv network through the hierarchy of substations. This results in 52,003 different configurations of edge failures. Finally, scenario (iii) investigates the robustness of the electricity network to the failure of three 400Kv transmission edges, resulting in 5,564,321 failure simulations. The connectivity of the 164,090 11Kv substations is then checked in each simulation to find those which are no longer connected through the hierarchy of the electricity network to the 400Kv substations.

For all three scenarios the results returned that none of the 164,090 11Kv substations were disconnected from the 400Kv level of the electricity transmission and distribution network. This observed robustness to failures of multiple edges in the highest level of the electricity transmission network shows that together the transmission and distribution aspects of the network are robust to these failures, even when they are co-located spatially (Figure 4.42). Along with this suggesting that the 400Kv aspect of the transmission network is robust to

failures with sufficient redundancy to remain connected despite 3 three failures, it could also be interpreted that it is the hierarchical nature of the network which is robust. The second level of the hierarchy, 275Kv, and the third level, 132Kv, both have a large number of assets, 201 and 1361 edges respectively (Table 4.11) along with 123 and 1146 substations. These form an extensive set of connections below the top level 400Kv assets, with the potential to provide alternative routes to the 400Kv level of the network. However, it should be noted that these results are based on the topology of the network alone, ignoring the capacities of the network assets, substations and transmission lines, along with the actual supply from the power stations and demand from the substations. It is only with greater data which allows a better attribution of the network assets that these factors could be considered in the analysis.

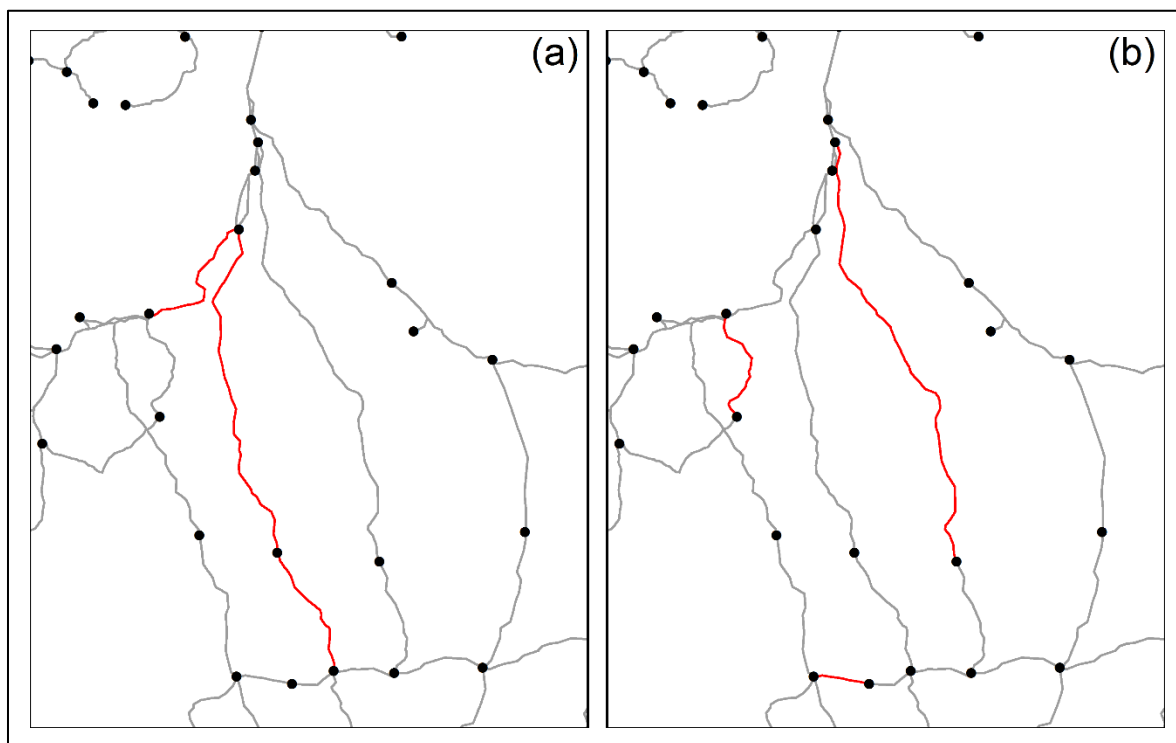


Figure 4.42: (a) showing three co-located transmission edge failures (red) and (b) showing three spatially distributed transmission edge failures.

4.5.3 Robustness to geographic hazards

The robustness of infrastructure networks to geographic hazards is critical given that many of the threats they face are inherently spatial such as flooding and strong winds (Little, 2003). These events directly affect the specific parts of a network that fall within their spatial footprint, but can also lead to failures outside of the hazard, second-order impacts, through the failures cascading to those assets which lie outside of the footprint of the hazard (Little, 2002). When infrastructure networks are exposed to such hazards they have the potential to affect the ability

of the entire network to function (Little, 2003; Sterbenz *et al.*, 2011). A model has been developed to measure the robustness of networks when exposed to spatial hazards, where a hazard is applied as an area (Figure 4.43(a)). All nodes and edges (assets) which fall within the spatial hazard are presumed to fail and referred to as first-order failures (Figure 4.43(b)). Following these failures, subsequent failures caused by nodes and edges becoming disconnected from the network through the first-order failures, are recorded and referred to as second-order failures (Figure 4.43(c)).

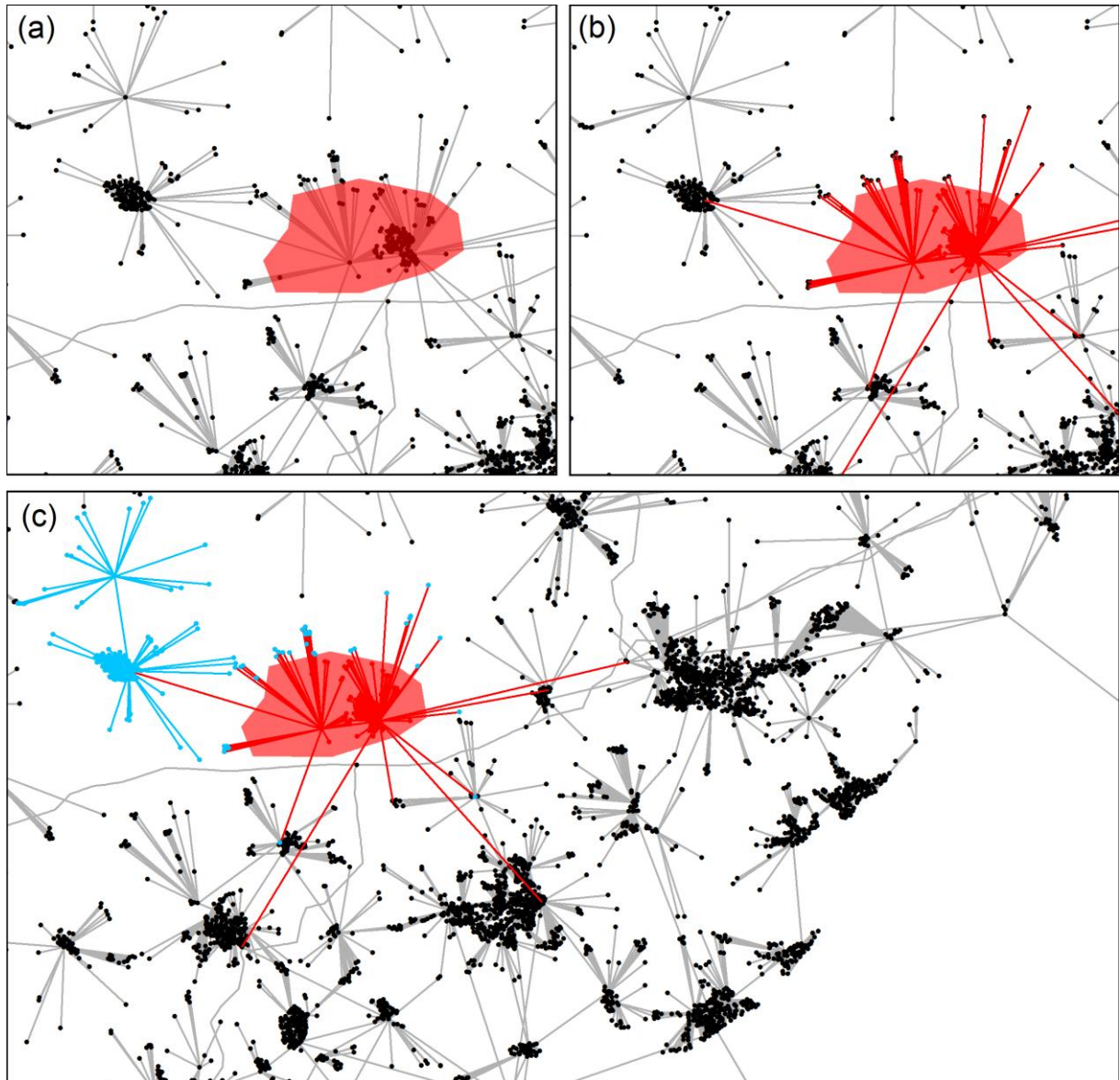


Figure 4.43: Exemplifying the cascade of failures through the electricity transmission and distribution network from an initial hazard. (a) an example hazard area with the electricity network, (b) the first-order failures shown in red and (c) with second order failures shown in blue.

In order to investigate the impact different spatial configurations of hazards may have on a hierarchical infrastructure network, an analysis of the robustness of the electricity transmission and distribution network has been performed using three sets of scenarios. The first (A) uses a single spatial large hazard, the second (B) uses four spatial hazards and the third (C) employs eight spatial hazards. In each case five randomly generated realisations were used to test the robustness of the network. For each realisation the size of the spatial hazard(s) has been set to affect approximately 2% of the total number of node assets in the network. Both the first-order failures, those nodes and edges which lie within the hazard areas are recorded, as well as the second-order failures, those nodes and edges which lie outside the hazards but are no longer connected to the network as a result of the first-order failures. With each hazard realisation set to effect 2% of the substations in the network, the relationship between the size of the hazard area(s) as well as their location(s) and the second-order failures can be examined across the three different scenarios.

Scenario set A is shown in Figure 4.44, with counts of the first-order failures as a result of the hazards shown in Table 4.12. With each hazard area set to approximately 2% of nodes in the electricity network a similar number of nodes are seen to fail in each hazard realisation. These first-order failures show that there is a significant difference between many of the realisations, with (iii) for example having a larger impact on the transmission lines with twelve 400Kv lines affected compared to six and seven for realisation (i) and (ii) respectively (Table 4.12). Realisation (i), around the Lake District (Figure 4.44), affects 19 generators, many more than either of the other four realisations which affect 1-7. For all realisations a large number of the failed nodes are the 11Kv substations (Table 4.12), the lowest level of distribution assets within the electricity network.



Figure 4.44: The single hazard areas for the five simulations in scenario set A.

Realisation		Total affected	Breakdown of affected network assets (nodes and edges)										
			generators	400Kv	275Kv	132Kv	66Kv	33Kv	25Kv	11Kv	0	logical	gen
i	nodes	3403	19	5	1	24	0	118	1	3231	4	NA	NA
	edges	3468	NA	7	2	37	0	148	2	3231	0	20	21
ii	nodes	3383	1	2	0	19	0	84	0	3275	2	NA	NA
	edges	3534	NA	6	0	28	0	97	2	3383	0	14	6
iii	nodes	3259	7	3	0	21	0	89	1	3134	4	NA	NA
	edges	3472	NA	12	0	34	0	126	2	3270	0	19	10
iv	nodes	3292	1	1	0	15	0	66	0	3207	2	NA	NA
	edges	3414	NA	4	1	22	0	79	0	3292	0	13	3
v	nodes	3193	2	1	3	24	0	109	0	3051	3	NA	NA
	edges	3344	NA	9	3	32	0	142	0	3136	0	20	2

Table 4.12: First-order failure counts for the realisations in scenario set A. Figure 4.44 shows the location of the each hazard areas.

The second-order failures, are recorded in Table 4.13 and shown in Figure 4.45. In a number instances the second-order failures are small in number, such as for realisation (i) where only 23 substations are affected where the hazard area covers an area of 12086.7Km² of land. In contrast, 1087 substation failures have been identified in realisation (iii) which has a land area of 5043.8Km². This suggests little relationship between the size of the hazard area and the number of second-order failures, and although realisation (i) has a large coastal boundary unlike realisation (iii), they both have perimeters on land which are less than 20Km's different, 248.2Km and 167.4Km respectively. However, the results (Table 4.12) show that for realisation (iii) over 30 33Kv substations were affected, all of which will have supplied the subsequent 11Kv substations which failed. This is not the case in realisation (i) with only 3 such substations failing. This further suggests a lack of robustness in the distribution aspect of the electricity network and not within the transmission part of the hierarchy.

The spread of the effect of the second-order failures is greatest for realisation (iii) where the average distance of a second-order substation failure is 7.32Km from the hazard area, Table 4.15 and Figure 4.44, with the furthest individual failure being over 20.45Km away. In contrast the lowest spread is seen in realisation (i) where the average distance for a second-order substation failure is just 0.76Km and the maximum just 2.65Km (Table 4.15). Realisation (i) is very much an extreme case, with the other realisations, (ii), (iv) and (v), having average values closer to that of realisation (iii), 6.29Km, 3.18Km and 4.09Km respectively, and maximum distance values of 12.99Km, 15.819Km and 6.669Km. The comparatively shorter distances in realisation (i) are likely a result of the predominantly rural area around the hazard footprint resulting in a low density of network assets.

Realisation		Total affected	Second-order failures of network assets (nodes and edges)										
			generators	400Kv	275Kv	132Kv	66Kv	33Kv	25Kv	11Kv	0	Logical	gen
i	nodes	23	0	0	0	0	0	3	0	20	0	NA	NA
	edges	18	NA	0	0	0	0	0	0	18	0	0	0
ii	nodes	359	4	0	0	0	0	5	0	350	0	NA	NA
	edges	242	NA	0	0	0	0	2	0	240	0	0	0
iii	nodes	1087	0	0	0	5	0	31	0	1051	0	NA	NA
	edges	940	NA	0	0	0	0	22	0	913	0	5	0
iv	nodes	166	0	0	0	0	0	2	0	164	0	NA	NA
	edges	79	NA	0	0	0	0	0	0	79	0	0	0
v	nodes	171	0	0	0	0	0	2	0	169	0	NA	NA
	edges	82	NA	0	0	0	0	0	0	82	0	0	0

Table 4.13: Second-order asset failure counts for the realisations in scenario set A. Mapped in Figure 4.45.

Realisation	Node distance from hazard (Km):	
	\bar{x}	Maximum
i	0.76	2.65
ii	6.29	12.99
iii	7.32	20.45
iv	3.18	15.82
v	4.09	6.67

Table 4.14: Average and maximum distance of the second-order substation failures from the hazard of the realisations in scenario set A.

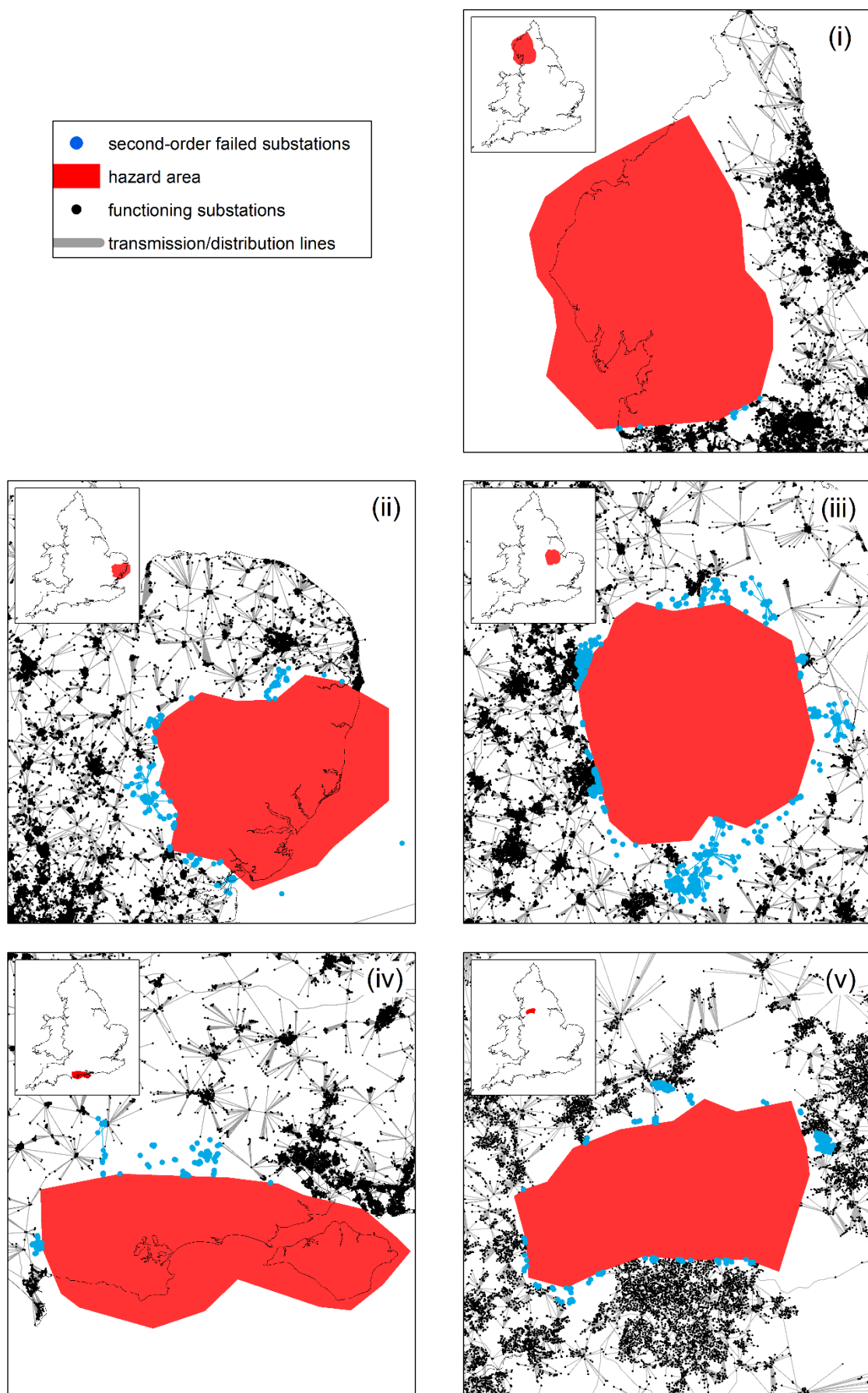


Figure 4.45: Second-order failures for realisations in scenario set A.

The hierarchical electricity network was exposed to four smaller hazards, scenario set B (Figure 4.46), created randomly, with again approximately 2% of the nodes in the network affected for each realisation. For each of the five realisations the first-order failures vary with regard to the type of features which are affected (Table 4.15). Across all realisations very few generators are effected, though a small number of assets towards the highest level of the hierarchy, the transmission substations and lines, are affected due to the random location of the hazard areas. Realisation (iv) has the largest effect on the transmission network with the greatest number of high voltage substations, two 400Kv and four 275Kv, affected as well as eight and nine 400Kv and 275Kv transmission lines respectively. The location of the hazard areas for this realisation, three of which are within or very close to dense network areas (Figure 4.46), suggesting that those lines affected may be those serving the urban areas, potentially having less impact on the wider network. With a hazard area of 1756.0Km², realisation (v) has less of an impact on the transmission network than realisation (iv), which affects an area of 995.5Km², indicating that realisation (v) hazard areas are located in slightly more rural areas where the network is less dense requiring greater area for 2% of the network to be affected.

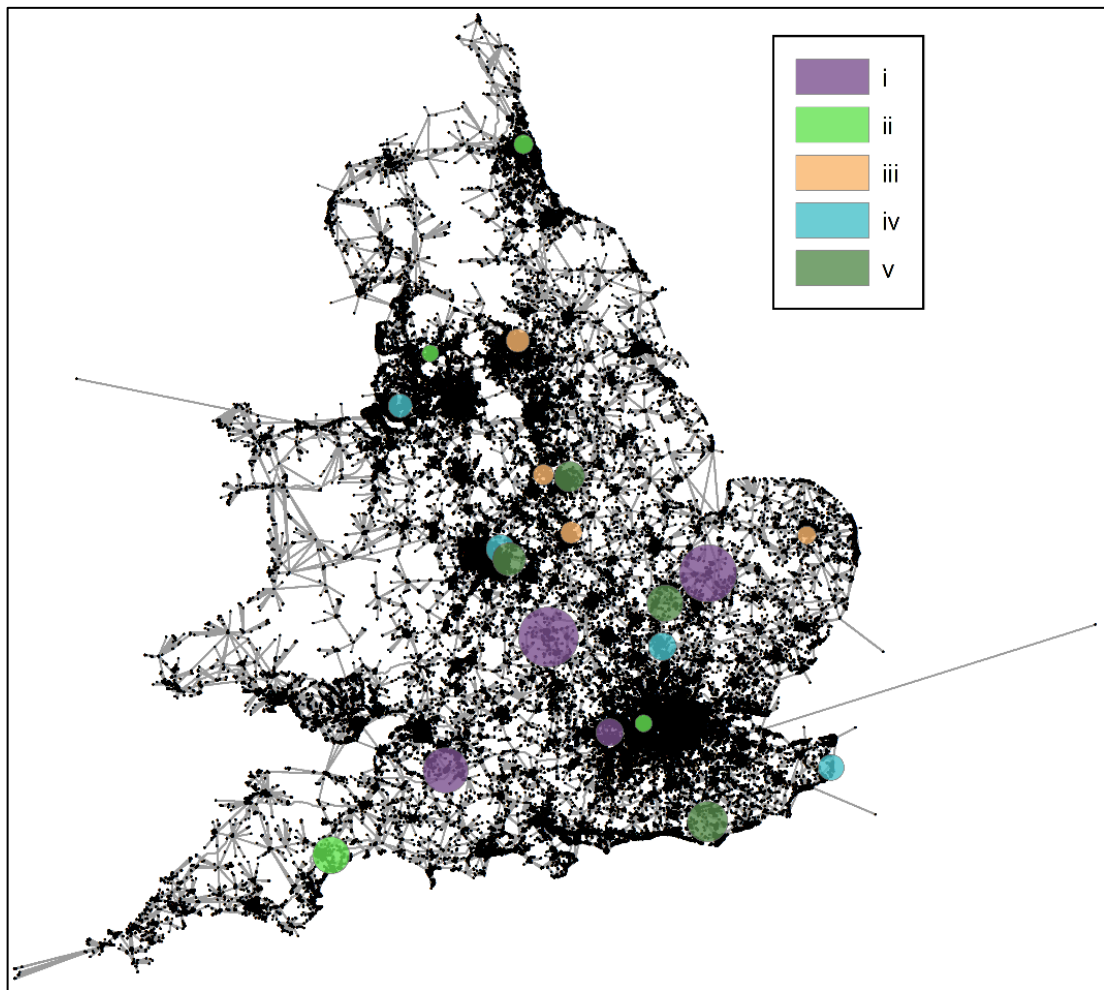


Figure 4.46: Map of the four hazard areas for each of the five realisations in scenario set B.

Realisation		Total affected	Breakdown of affected network assets (nodes and edges)										
			generators	400Kv	275Kv	132Kv	66Kv	33Kv	25Kv	11Kv	0	logical	Gen
i	nodes	3059	2	1	0	13	0	61	0	2982	0	NA	NA
	edges	3513	NA	3	0	24	0	96	0	3377	0	11	2
ii	nodes	3207	0	1	2	20	1	97	0	3085	1	NA	NA
	edges	3506	NA	6	5	34	1	142	0	3295	2	21	0
iii	nodes	3209	0	1	1	23	0	93	0	3091	0	NA	NA
	edges	3642	NA	3	1	36	0	135	0	3444	1	22	0
iv	nodes	3211	1	2	4	26	0	104	0	3071	3	NA	NA
	edges	3605	NA	8	9	43	0	132	1	3387	1	23	1
v	nodes	3160	2	2	2	24	0	105	0	3023	2	NA	NA
	edges	3467	NA	8	5	41	0	140	0	3249	0	22	2

Table 4.15: First-order asset failure counts for the five realisations in scenario set B. Figure 4.46 shows the hazard areas for the five realisations.

The second-order effects of these hazards show a greater impact (Table 4.16), than for the single hazard realisation in scenario set A (Table 4.13), with many more substations becoming disconnected from the network. Realisation (v) had the greatest impact with 6580 substations affected as a result of the first-order network asset failures. These hazards affected an area of 1755.9Km^2 , whereas realisation (i) which affects the largest area, 3429.6Km^2 , resulted in 5452 substation failures. The realisation with the hazards which affected the smallest land area, 633.6Km^2 in realisation (iii), resulted in 6069 substation failures in total, clearly demonstrating that as with the single hazard realisations in scenario set A, that the areas affected has little relationship to the number of second-order failures. Instead, the results from Table 4.15 and Table 4.16 suggest that the number of second-order failures, rather than being dictated by the size of the hazard areas or the number of failures in the transmission level of the network, are instead more related to the number of first-order failures in the lower level distribution network. The trends shown in Figure 4.47 shows that the total second-order node and edge failures follow a similar trend, with the closest match shown in the trend of the first-order failures for the 33Kv edges. The trend line for the high level transmission edges, the 400Kv, 275Kv and 132Kv edges, follow the general trend for the second-order failures apart from in realisation (iv). This suggest a correlation between the number of second-order edge failures and second-order node failures, a relationship which was expected given the failure of an edge will lead to the failure of at least one node. It is also noticeable that across all five scenarios the number of second-order node and edge failures was similar, despite each scenario having four hazard areas which were

randomly distributed over the network. This suggests that the impact of the hazards in each scenario, although they may have varied individually, when aggregated results in a similar number of failures.

Realisation		Total affected	Second-order failures of network assets (nodes and edges)										
			generators	400Kv	275Kv	132Kv	66Kv	33Kv	25Kv	11Kv	0	Logical	gen
i	nodes	5452	4	4	1	28	0	121	0	5294	0	NA	NA
	edges	3183	NA	0	0	2	0	43	0	3086	0	5	2
ii	nodes	6463	4	5	7	40	2	164	0	6239	2	NA	NA
	edges	4226	NA	0	0	4	0	95	0	4113	0	11	3
iii	nodes	6059	4	4	1	32	0	147	0	5869	2	NA	NA
	edges	3552	NA	0	0	4	0	72	0	3465		9	2
iv	nodes	4953	4	3	4	28	0	131	0	4778	5	NA	NA
	edges	2648	NA	0	0	0	0	56	0	2589	0	3	0
v	nodes	6580	2	6	2	39	0	179	1	6343	8	NA	NA
	edges	4158	NA	0	0	4	0	74	0	4071	0	8	1

Table 4.16: Second-order asset failures for the five realisations in scenario set B, also mapped in Figure 4.48.

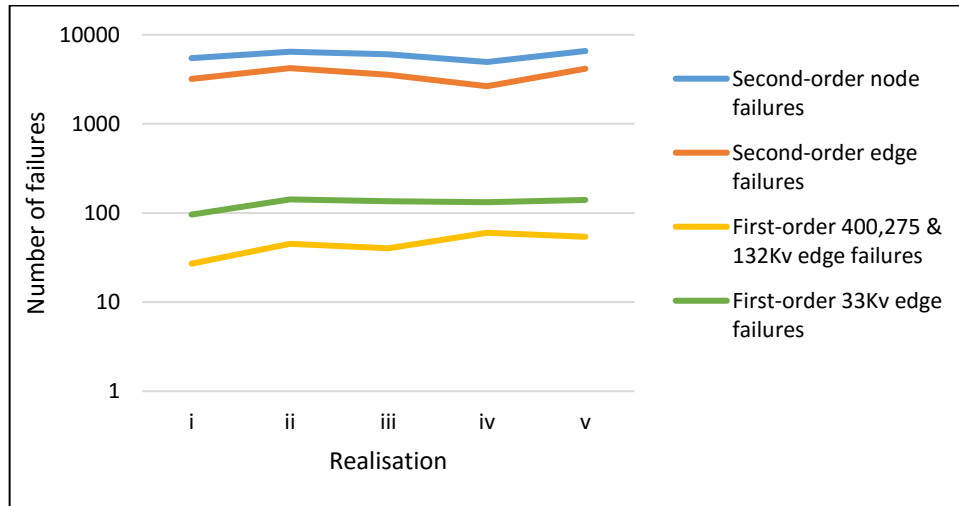


Figure 4.47: Trends between the number of first and second-order asset failures for selected nodes and edges for each realisation in scenario set B, shown using a log scale.

The geographic pattern of second-order failures are shown in Figure 4.48. It is clear that in both realisations (ii) and (iii) there is wide dispersal of second-order failures, where (ii) shows a large spread of failures in the North East of England and (iii) shows this in East Anglia also. This large spread of failures can also be seen in Table 4.17 where the maximum distance for the second-order failures for realisations (ii) and (iii) are 85.7Km's and 60.1Km's, over 10Km's greater than for any other of the realisations. This suggest a poor robustness to failures in these areas, though the North East of England would also normally be connected to the power network in Southern Scotland which may improve the robustness in this region. The average distance from the hazard areas for the second-order failures for all five realisations is 9Km - 19Km (Table 4.17). This is greater than the observed values for scenario set A, 0.7Km – 7.3Km (Table 4.14), a result of the smaller hazard areas being more likely to lie within urban areas whereas the urban areas tend to fall completely within the single hazard areas. This causes fewer second-order failures as there is a lower density of assets around the boundaries of the single hazard areas compared to those smaller areas which lie within urban areas.

Realisation	Node distance from hazard (Km):	
	\bar{x}	Maximum
I	10.93	44.58
Ii	13.16	85.74
Iii	18.24	60.08
Iv	9.63	32.51
V	11.23	49.31

Table 4.17: Average and maximum distance of the second-order substation failures from the hazard areas for each realisation in scenario set B.

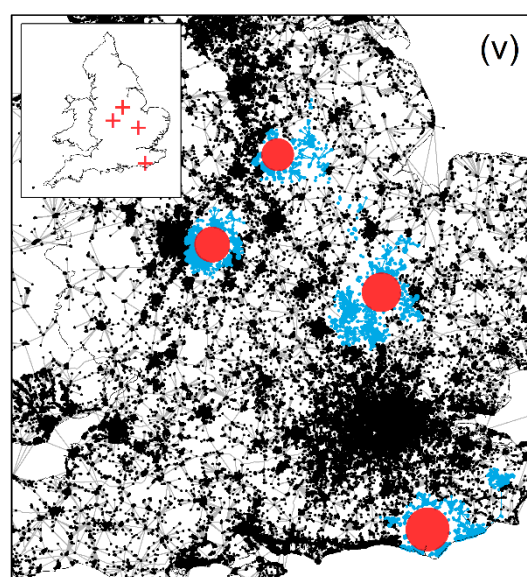
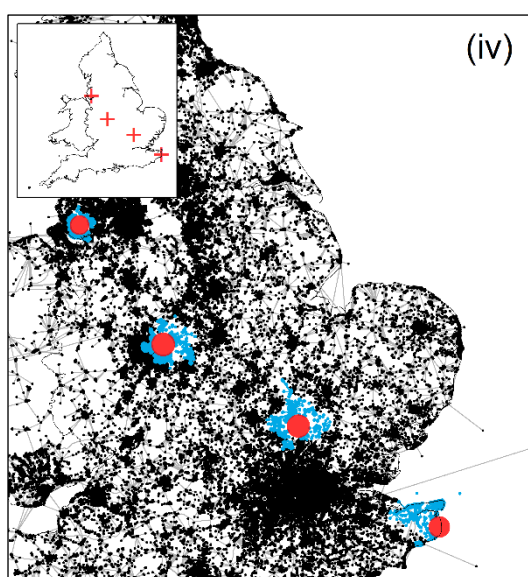
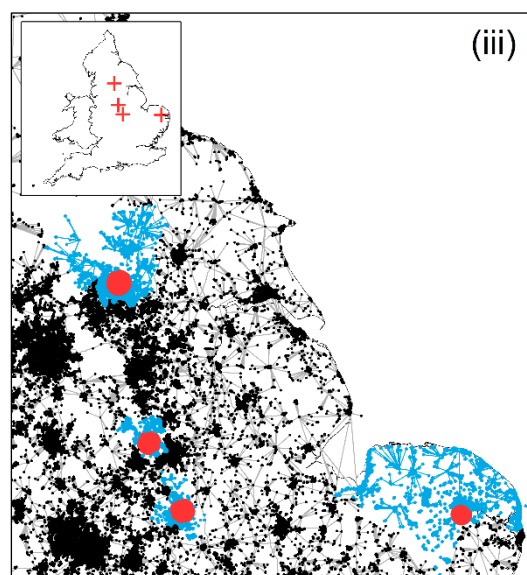
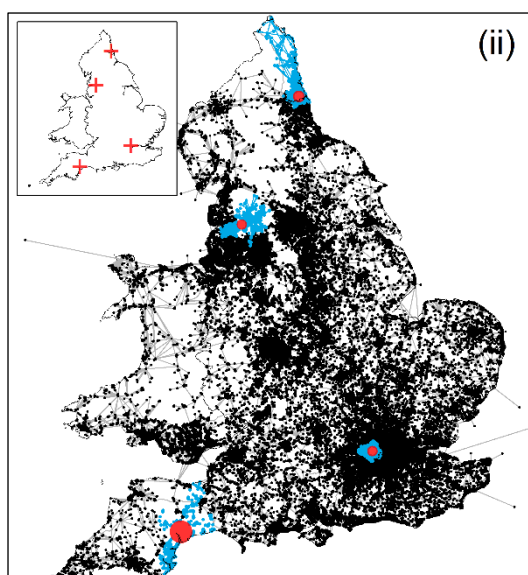
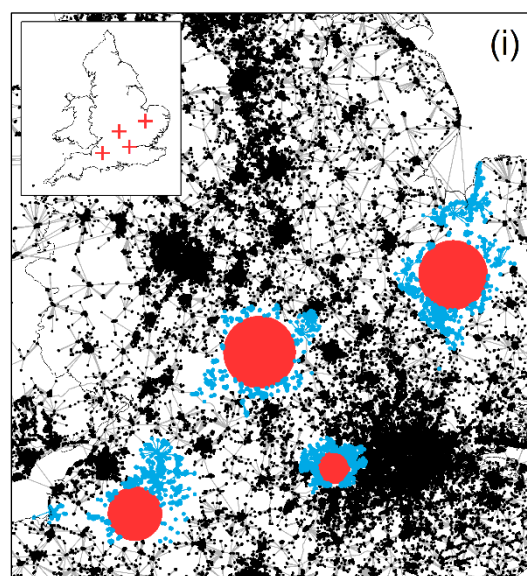
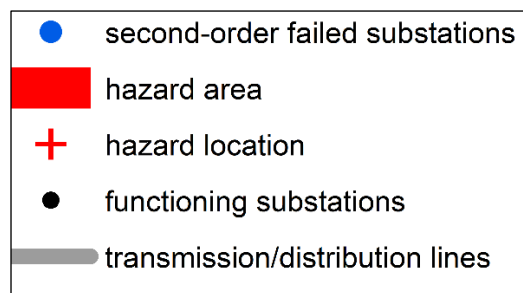


Figure 4.48: Second-order substation and edge failures for the five realisations in scenario set B.

The final set of scenarios, set C, uses eight small hazard areas to perturb the hierarchical electricity network, with hazard sets for each of the five realisations shown in Figure 4.49. Again the areas are selected at random and with the total number of nodes directly affected approximately 2% of the total in the network. All five realisations effect the transmission features, the highest level of the hierarchy, in the network similarly with regard to the failure of the 400Kv and 275Kv substations (3, 6, 3, 4 and 3) and edges (16, 16, 10, 9 and 12) (Table 4.18), with little discernible difference between the different realisations. There are no direct failures of generators in realisation (iii), but all others result in at least a single failure, Table 4.18, which has the potential to remove critical power supply stations from the network.

Realisation		Total affected	Breakdown of affected network assets (nodes and edges)										
			generators	400Kv	275Kv	132Kv	66Kv	33Kv	25Kv	11Kv	0	logical	gen
i	nodes	3323	3	2	1	31	0	128	0	3156	2	NA	NA
	Edges	3805	NA	9	7	56	0	198	0	3501	2	27	5
ii	nodes	3244	2	3	3	29	0	91	0	3115	1	NA	NA
	Edges	3742	NA	8	8	58	0	134	0	3508	3	21	2
iii	nodes	3371	0	1	2	21	1	102	0	3244	0	NA	NA
	Edges	3893	NA	4	6	35	1	144	0	3681	0	22	0
iv	nodes	3065	1	2	2	18	2	86	0	2953	1	NA	NA
	Edges	3518	NA	6	3	32	2	130	0	3325	1	18	1
v	nodes	3145	1	1	2	24	0	94	0	3020	3	NA	NA
	Edges	3690	NA	5	7	40	0	156	0	3457	1	23	1

Table 4.18: First-order network asset failures for each realisation in scenario set C. The location of the hazards for these realisations are shown in Figure 4.49.

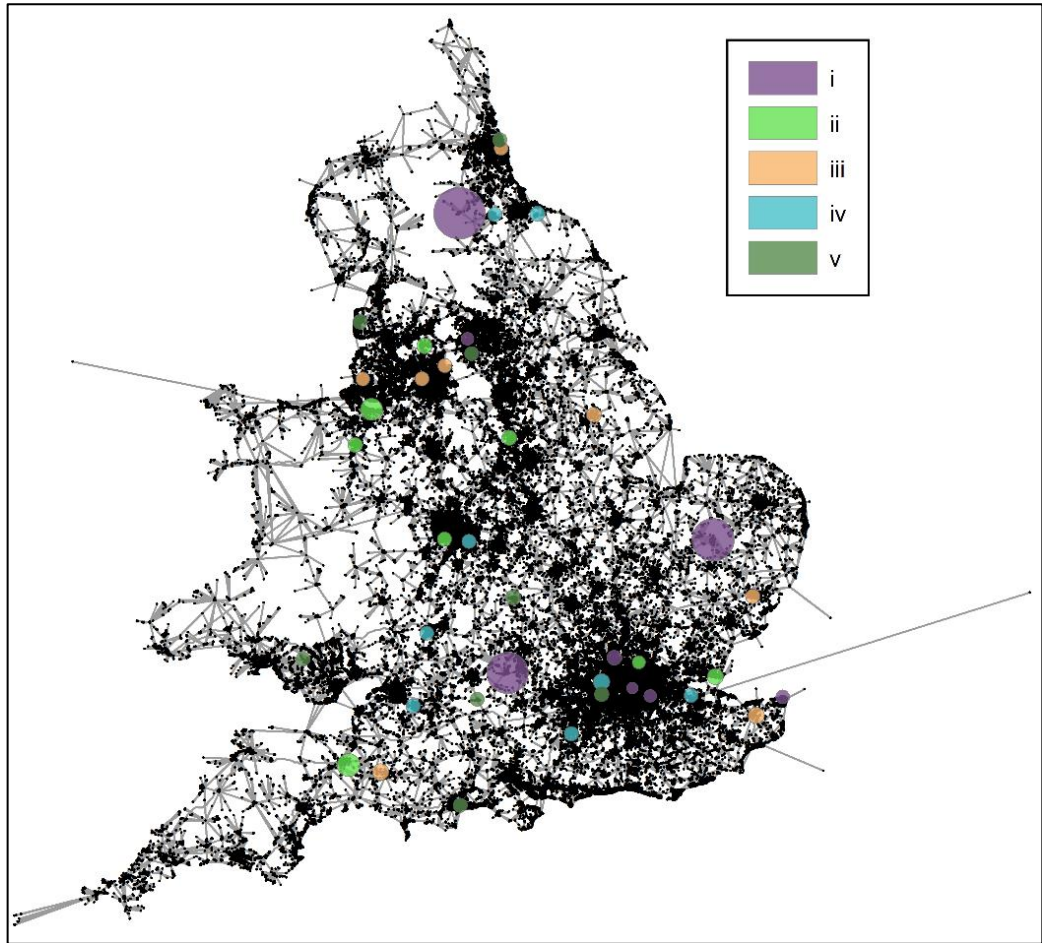


Figure 4.49: Showing the location of the eight hazard areas for each of the five realisations in scenario set C.

The second-order failures for the five hazard realisations in scenario set C show a significant variation in effect on the electricity network with realisation (iii) only resulting in the failure of a further 5187 substations/generators (Table 4.19), whereas realisation (i) results in the failures of 9173 substations. With both realisations (i) and (iii) having four hazard areas in urban areas as well as four non-urban locations, the distribution between dense and less dense areas of network assets is similar. However, realisation (i) has three hazards in close proximity to each other in London, potentially causing a greater proportion of failures as the failure caused by each hazard may have affected the robustness of the substations around the other hazards, increasing the number of second-order failures observed.

The total for second-order substation failures in realisation (i) is greater by 2153 than the total for any other of the realisations (Table 4.19). This same difference is not observed for the first-order failures (Table 4.18) with the number of substations failures for realisation (i) being 3323, less than observed for realisation (iii) which has a total of 3371. However, there is a noticeable difference between the five realisations with the number of 33Kv edge failures, with 198 first-

order failures for realisation (i), with the next highest being 156 for realisation (v). The trend for the first-order 33Kv edge failures seems to provide an explanation for the varying number of second-order substation failures, as was the case for scenario set B (Figure 4.47). Although the hazard areas in realisation (i) cover the largest area, 2811.587Km², this is a result of the three of the hazard areas lying in rural areas. This is further exemplified by realisation (ii) which covers the second largest area of land at 940.0Km², yet has the second fewest second-order substation failures. These results strongly suggest a relationship between the number of second-order failures and the location of the hazards, rather than the size, with those in urban areas affecting a greater number of distribution lines and substations, such as the 33Kv edges, leading to a greater number of second-order failures. The trend lines showing the relationship between the first and second-order node and edge failures across the five scenarios for scenario set C (Figure 4.50) indicate that the strength of the relationship between first-order failures and second-order failures is weaker than for scenario set B (Figure 4.47). This is likely a result of the greater number of hazard areas in each scenario and the local network structure in and round these areas creating a greater variation in robustness than when only four hazard areas were used in scenario set B. As with the results from the previous scenario the trend lines appear relatively flat, a result of the similar robustness to the perturbations exhibited in each of the five simulations. This suggests that the electricity network may have a uniform robustness, with the same response to failures no matter the location of the hazards.

Realisation		Total affected	Second-order failures of network assets (nodes and edges)										
			generators	400Kv+	275Kv	132Kv	66Kv	33Kv	25Kv	11Kv	0	logical	Gen
i	Nodes	9173	5	10	7	59	1	276	0	8813	2	NA	NA
	Edges	5306	NA	0	0	6	0	120	0	5162	0	17	1
ii	Nodes	6325	7	11	7	47	0	191	0	6055	7	NA	NA
	Edges	3635	NA	0	1	1	0	79	0	3549	0	4	1
iii	Nodes	5187	4	6	9	45	0	139	0	4977	7	NA	NA
	Edges	2596	NA	0	0	1	0	48	0	2541	1	5	0
iv	Nodes	7020	1	4	4	41	2	183	0	6779	6	NA	NA
	Edges	4144	NA	0	0	6	0	78	0	4047	0	12	1
v	Nodes	6770	4	5	8	41	2	173	0	6532	5	NA	NA
	Edges	3788	NA	0	0	2	0	80	0	3693	0	11	2

Table 4.19: Second-order network asset failure counts for each realisation in scenario set C.

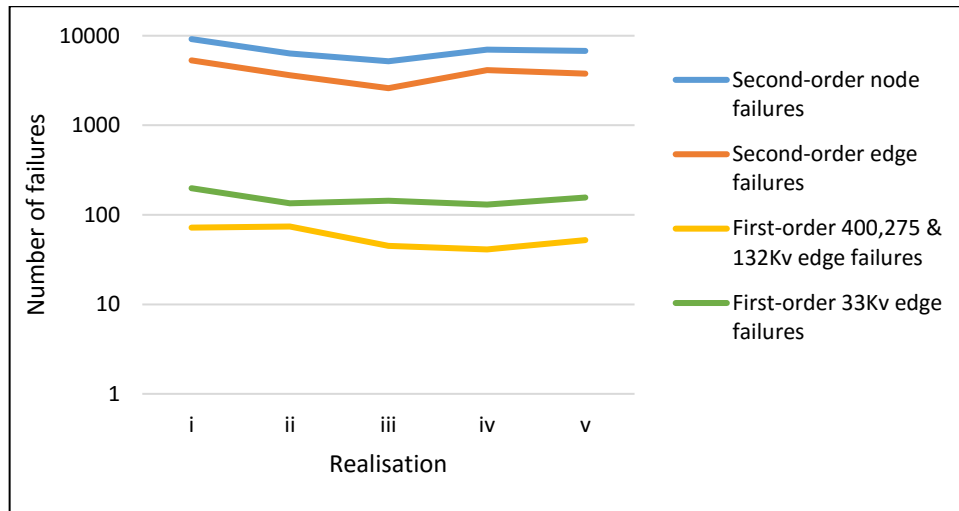


Figure 4.50: Trends between the number of first and second-order asset failures for selected nodes and edges for each realisation in scenario set C, shown using a log scale.

The average distance of the second-order substation failures for all realisations is very similar, between 7.7Km and 11.5Km suggesting that these failures occur relatively close to the hazard areas with a similar behaviour across all areas. For the maximum distance of second-order failures from the hazard areas there is much greater range (55.2Km), with the maximum distance found in realisation (v) being 87.8Km compared to the shortest at 32.6Km for realisation (ii). The second greatest maximum distance (56.4Km) is 30Km less than realisation (v), showing that realisation (v) is an extreme case. This distance is observed in the North East England (Figure 4.51), where the failures spread to the border with Scotland, where no network has been included in this analysis. It is suggested that if a network for Scotland were to be included, the results would differ as a second high level supply line may connect the affected area reducing the number of second-order failures. However, again there is no relationship evident between the distance of second-order failures in this realisation set and the size of the areas, with realisation (v) affecting the smallest areas (657.6Km^2), while realisation (ii) affects a greater area (940.0Km^2), with realisation (i) affecting the greatest area (2811.6Km^2). This suggests the effect of the hazard areas is again dependent entirely in the location of the hazard areas.

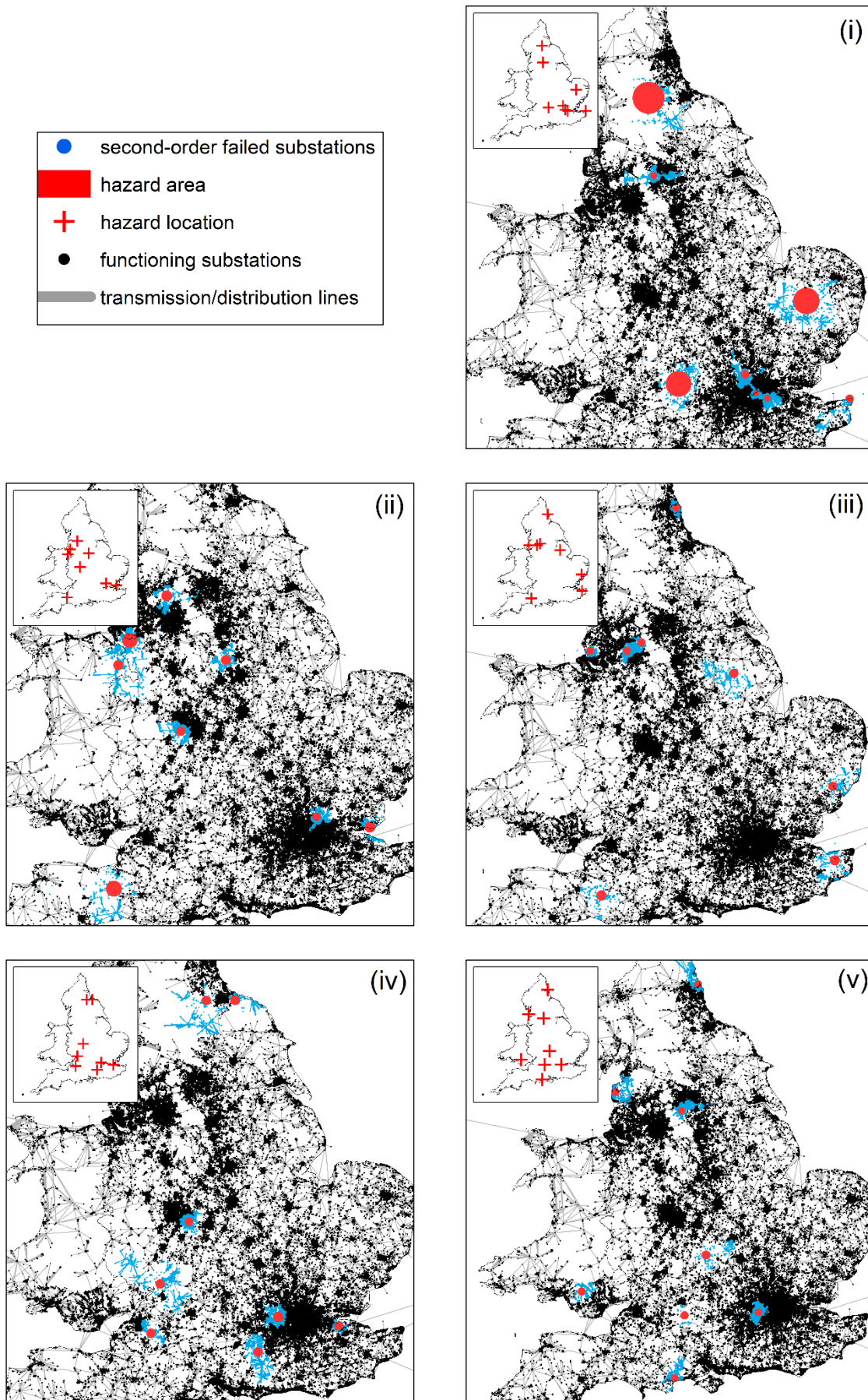


Figure 4.51: Second-order asset failures for the five hazard realisations in scenario set C.

Realisation	Node distance from hazard (Km):	
	\bar{x}	Maximum
i	10.76	47.85
ii	7.75	32.56
iii	8.44	56.43
iv	11.50	48.82
v	11.36	87.79

Table 4.20: Average and maximum distance from hazard areas of second-order substation failures for the five realisations in scenario set C.

4.6 Summary

This chapter has presented the results from the analysis of a suite of synthetic graph models and critical spatial infrastructure networks in order to understand whether the structure of hierarchical infrastructure networks results in a different level of robustness than non-hierarchical infrastructure networks.

The results have shown that hierarchically organised graphs can be recognised from non-hierarchically organised graphs. The most effective method employed was through the use of graph metrics such as the number of CB and the MBC where the hierarchical graphs were found to be statistically different from the non-hierarchical graphs using these metrics. Typically, the hierarchical graphs had < 2 CB per node and MBC values > 0.3 . The response exhibited by hierarchical and non-hierarchical graphs to perturbations also shows significant differences. The hierarchical models have been found to exhibit a much weaker robustness to perturbations, with the hierarchical graphs on average failing 28.37% quicker than the non-hierarchical graphs. The hierarchical HC model did however fail more similarly to the non-hierarchical graphs than the other hierarchical graphs, exhibiting a greater robustness to perturbations. An analysis using a capacity constrained cascading failure model over the hierarchical and non-hierarchical graphs has shown that the hierarchical graphs were much less robust, failing nearly twice as quick in all six scenarios, or not being computable due to poor connectivity. However, the HR graph was an exception to this exhibiting a robustness more similar to the non-hierarchical graphs.

It has been found that critical spatial infrastructure networks including air networks, river networks, and rail networks exhibit characteristics, such as values for the MBC and number of CB which are the same, to a greater extent, as those found the already analysed hierarchical graphs. Some of these networks, such as the river networks, have hierarchical degree

distributions, though some infrastructure networks including some rail and energy networks do also share similarities with the non-hierarchical graph models such as the scale-free and small-world models. For a number of infrastructure networks, including the road and rail networks, the metric values have returned a different similarity to graph models than the degree distributions, highlighting the ability of the higher level metrics to provide deeper insights into the organisation of the networks.

The robustness of the infrastructure networks also showed similarities to the hierarchical graphs, with all infrastructure networks analysed behaving like the hierarchical TREE, HR and HR+ graphs. None of the infrastructure networks returned a response similar to that exhibited by the most robust hierarchical model, the HC model however.

Finally, a case study using the electricity transmission and distribution network for England and Wales has shown that the network is robust to a range of failure scenarios. This included the targeted removal of multiple 400Kv transmission lines, the highest level in the network, finding that this had no effect on the connection of the 11Kv substations to the transmission network. The robustness of the network to a range of spatial hazards was also explored using three scenarios showing as the number of hazards increases with the total number of affected node assets remaining similar (2% of substations), the network becomes more vulnerable to second-order failures. For single hazards a mean of 358.8 second-order substation failures was recorded, with a mean of 6366.8 for the eight hazard realisations. The land area covered by the hazards was found not to be a factor in the resulting number of second-order substations failures, with instead the first-order failure of 33Kv edges, those which feed the 11Kv substations, found to be a possible factor. However, the location of the hazard areas seems to be the biggest factor in the scale of the second-order failures, with those in urban areas having the greatest effect.

Chapter 5: Discussion

5.1 Introduction

The dependence on critical infrastructure networks within societies around the world, and the vulnerability of these to a range of hazards and subsequent failures, has been highlighted in Chapter 2. The literature review also showed that graph models, such as the scale-free and small-world models, have been used for the analysis of critical infrastructure networks to improve our understanding of their characteristics, including their robustness to failures (Albert *et al.*, 2000; Bassett and Bullmore, 2006). This review also highlighted the existence of a hierarchical structure in some real networks such as those in social sciences and biological networks (Ravasz *et al.*, 2002; Clauset *et al.*, 2008), with graph models being developed for such graph topologies (Ravasz *et al.*, 2002; Ravasz and Barabasi, 2003). However, there exists relatively little understanding of whether certain critical infrastructure networks exhibit a hierarchical structure either topologically or in terms of the flows and movements taking place on them. To address this, Chapter 3 presented a methodological framework and suite of software to investigate the characteristics of hierarchical networks and if critical infrastructure networks exhibit these characteristics. Chapter 4 then presented the result of this analysis. In this chapter a discussion is presented which synthesises the findings of Chapter 4 in order to consider the robustness of critical infrastructure networks and the subsequent potential impacts on the broader systems they operate within.

5.2 Characteristics of hierarchical graphs and networks

5.2.1 Graph models and characteristics

The analysis of the suite of eight graph models has shown that hierarchical graphs can be distinguished from non-hierarchical graphs (Chapter 4, Section 4.2, page 101) using three metrics, the maximum betweenness centrality (MBC), the assortativity coefficient (AC) and the number of cycle basis per node (CB). Via a pair-wise analysis of the three metrics (Figure 4.2, Figure 4.3 and Figure 4.4), it is clear that the graphs generated by the four hierarchical graph models return different characteristics to those generated by the four non-hierarchical models. In general the hierarchical models exhibit lower AC values, ≤ -0.17 with greater MBC values, ≥ 0.25 and lower CB values, ≤ 1.91 compared to AC values ≥ -0.05 , MBC values ≤ 0.08 and CB values ≥ 6.86 for the non-hierarchical models (Table 4.2, page 106).

The MBC values, an indication of the level of connectivity within a graph (Girvan and Newman, 2002), are higher for the hierarchical graphs, with an average of 0.52, whereas the non-hierarchical models have an average of 0.03. This indicates a reliance on one or possibly several critical nodes in the hierarchical graphs in order to remain traversable. The HC and TREE models exhibit particularly high values for the MBC metric, an average of 0.77, indicating that these two models in particular are more dependent on a single node than all other models. Those graphs with high MBC values are more likely to be relying on a hub node, with a large proportion of shortest paths passing through this, and thus is more likely to be vulnerable to the failure of these nodes, especially in targeted failure models.

Inversely related to the MBC metric is the CB metric, with lower values, showing fewer loops/cycles in the graph and hence potentially a smaller number of unique paths between any pairwise set of nodes. This is shown by the hierarchical models, where they have an average CB of 0.74, whereas the non-hierarchical graphs have an average of 12.86. A low value also indicates a lack of redundancy in the graphs, with few cycles indicating the lack of alternative paths between nodes (Katifori *et al.*, 2010). Therefore the analysed hierarchical graphs may be more vulnerable to failures compared to the non-hierarchical graphs. The AC metric shows a difference in the structure of the hierarchical and non-hierarchical graphs with the hierarchical models having a mean of -0.34, indicating nodes with different degrees are connected to each other (Newman, 2002), as might be found in a network relying on hub nodes. The non-hierarchical models on the other hand have a mean of 0.02, similar to Newman (2002) who found random and scale-free models have a value of zero. These values suggest the degree of connected nodes are not correlated, indicating a more mixed structure which may be more robust to perturbations (Newman, 2002). This indicates no reliance, or a limited reliance, on hub nodes within the non-hierarchical graphs and thus less reliance on a small subset of hub/critical nodes.

With the distribution of metrics being statistically different between the four suites of hierarchical and non-hierarchical graphs, it is clear the metrics can be employed to distinguish between the two sets of models (Table 4.3). Only two of the metric distributions between the hierarchical and non-hierarchical graphs fall below the critical value of 85 which indicates statistically different distributions (Swain and Davis, 1978), with these being the HR and HR+ values against the BA model for the AC-MBC metric combinations (75.24 and 83.36 respectively). The identified similarity between these models is only across one combination of the metric distributions, and the values are still high indicating that the similarity between them is relatively similar. There is thus a clear difference between the two sets of graph models, with

each having disparate characteristics as identified through the three graph metrics, implying that hierarchical graphs are distinct from non-hierarchical graphs.

The results have highlighted that the use of the three metrics, MBC, AC and CB, has allowed the characteristics of the hierarchical graphs to be identified, with these being unique from those found for the non-hierarchical graphs. Previous research has focused on identifying the characteristics of graphs using metrics such as the degree distribution, the average path length and the clustering coefficient (Newman, 2003b; Amaral and Ottino, 2004; Boccaletti *et al.*, 2006). These all focus on the topological structure of the graphs, though as discussed previously (Chapter 3, Section 3.4.2), these also have a number of limitations when measuring the characteristics of graphs (Ouyang *et al.*, 2009). The application instead of the MBC, AC and CB metrics enabled high-level insights into the graphs analysed to be ascertained compared to metrics such as a degree distribution. The betweenness centrality metric for example relates to the connectivity of the graph through reporting on the importance of each node (Girvan and Newman, 2002). The metric calculates a value for each node on its importance for the shortest paths through the network with regard to all other nodes, and thus the role the node plays in keeping the network well connected, rather than just reporting on its topological properties. The successful application of these higher-level metrics in this research further indicates that research can look beyond the traditional metrics, degree distribution, average path length and clustering coefficient, for the characterisation of graph structures. A number of other studies have examined the use of these metrics including research by Caldarelli *et al.* (2004) who examined using alternatives to the clustering coefficient when characterising the structure of graphs and Foster *et al.* (2010) who employed assortativity measures when comparing the structure of directed graphs and networks. However, none of those studies, or others found, have performed a comprehensive ensemble analysis either over an extensive range of graph models as employed in this work, or with as many exemplars with regards to the models employed.

The comparison of the degree distributions over the graphs generated by the eight models are less conclusive at being able to differentiate between hierarchical and non-hierarchical graphs. There is a tendency for the plots for the hierarchical graphs to have a number of peaks and troughs when plotted (Figure 4.1, page 103). These features though also appear in the plots for the non-hierarchical graphs such as in the scale-free plots, especially towards the tail of the distribution. Further to this, there is little differentiation possible between the degree distribution plots for the HR and HR+ graph models, indicating that these models may produce graphs which are near identical. However, these have been shown not to be identical in the

metric results discussed in the above paragraphs highlighting the ability of the higher-level metrics to identify characteristics otherwise not seen in metrics such as the degree distribution.

The results have also shown that within the two groups of models, the non-hierarchical and hierarchical graph models, there is little homogeneity within them, highlighting the breadth of the spectrum of graph models employed in the research. Of the 18 statistical values from the transformed divergence analysis measuring the overlap between the non-hierarchical graphs five values were below 70, with the greatest similarity being between the BA and WS model where values of 14.13, 43.86 and 43.98 were returned (Table 4.3, page 109). Both models share some similar characteristics including similar average path lengths (Albert and Barabasi, 2002), giving rise to the observed similarity. For the four hierarchical graph models, the similarities observed were between the HR and HR+ models, with values less than 40 for the three metric combinations. The similarity of these two models is to some degree expected given that both models share the same base graph (the TREE model) and have similar generation methods (Chapter 3, Section 3.3). Other studies examining the properties of graphs have focused on a smaller number of models, or have been specific to a single model. Work by Barrat and Weigt (2000) for example examined the properties of small-world graphs with the only the one graph type used while Newman (2002) only used two graph models along with a number of real-world networks to examine assortative mixing in networks. A greater a number of models were used by Costa and Silva (2006) with three employed to investigate the effect of a new set of metrics. However, this still falls short of the eight employed in this work. Further to this, none of these previously mentioned studies have employed large ensembles of the graphs employed, with these limited to single figures in some cases, with Costa and Silva (2006) only employing three versions of each model, significantly less than 1000 employed for six of the eight models in this research (with the other two have 7 and 31 exemplars).

5.2.2 Critical spatial infrastructure networks

Comparison of the graph model results with real critical infrastructure networks using the AC, MBC and CB metrics (Chapter 4, Section 4.3.3, page 123) revealed that 37 of the 42 networks were more similar to the hierarchical models than the non-hierarchical graph models (Figure 4.14, Figure 4.23 and Figure 4.24). This indicates that many of the spatial infrastructure networks are better represented through one of the four hierarchical graph models than the four non-hierarchical graph models, two of which, the small-world and scale-free models, have long been described as being similar to infrastructure networks (Newman, 2003b; Boccaletti *et al.*, 2006).

Of the infrastructure networks investigated the air networks showed a greater similarity to the hierarchical HR/HR+ graph models, with the river networks also showing the greatest similarity to the hierarchical TREE model. These both exhibited high MBC values, especially the river networks, along with AC values <-0.2 , resulting in these being similar to the TREE and HR/HR+ models. These characteristics also suggest the presence of hub nodes within the networks, potentially making them vulnerable to failures. As well as these, a number of the rail networks, including local system such as the Tyne and Wear Metro, showed the greatest similarity to the HC graph model, whereas the rest, including national networks for Ireland and Great Britain, were most similar to the HR+ graph model. A spread of values were observed for the MBC metric across the rail networks. The local rail networks which were similar to the HC graph model, exhibiting high values, >0.5 , making these appear vulnerable to failures. However, the other rail networks exhibited values <0.5 , and with AC values >-0.1 , suggesting a degree of robustness. The energy networks were most similar to the HR+ graph model along with the road networks, with values of $0.18 < \text{MBC} < 0.5$, and $-0.2 < \text{AC} < 0.2$. Both of these sets of networks are hierarchical, but show a tendency for a more mixed graph structure with greater redundancy than in the other infrastructure networks, potentially making these more robust to failures.

These results are in contrast to findings from other research, with the structure of infrastructure networks varying from those found through the employed metrics. For example air networks have previously found to be scale-free (Verma *et al.*, 2014) or to have small-world properties (Bagler, 2008a), and electricity networks found to possess scale-free properties (Albert *et al.*, 1999; Rosas-Casals *et al.*, 2007; Hines and Blumsack, 2008). Sen *et al.* (2003) has also shown the Indian rail network shares the characteristics of the small-world graph model, although Latora and Marchiori (2002) found for Boston the local rail network was not similar to the small-world model, but the transport system as a whole was. There continues to appear to be a discontinuity between the results reported in these studies and the characteristics identified in this work, suggesting the employed metrics are identifying different characteristics which provide greater insights into the structure of the networks, and in many cases, the hierarchical form they take. This implies that many infrastructure networks may have a different structure to those which have been previously reported using *traditional* metrics, such as the degree distribution, the average path length and the clustering coefficient.

The higher-level metrics employed show that the characteristics of most infrastructure networks imply a greater similarity with the hierarchical models suggesting these are a better representation of spatial infrastructure networks. Hierarchical models, such as the HR and HR+,

can be used to represent infrastructure networks due to the similarities that are present between the real-world networks and the graph models. The use of hierarchical models has been suggested previously but with little reference to the infrastructure sector, for example simulating networks from biology (Ravasz *et al.*, 2002). However, it is now clear, through the results presented in this work, that infrastructure networks can be hierarchical.

It was previously suggested by Barabasi *et al.* (2003) that geographical constraints prevent hierarchies forming in spatial networks, with the cost of building and adding long links a cause for this. This research has found otherwise, with a number of infrastructure networks, including some rail and energy networks, where costs can be associated to the development of links, appearing to be more similar to hierarchical models. The air networks analysed also returned hierarchical characteristics, though there are no (or little) costs associated with the construction of physical links between airports unlike in all other sectors analysed. None of the aforementioned networks were analysed by Barabasi *et al.* (2003), which instead was limited to the analysis of an actor network, a language network, a metabolic network, a protein interaction network and sample networks of the world wide web and the internet (at the AS (autonomous system) level). The suite of networks analysed was limited in comparison to those analysed in this work, which covered a breadth of spatial infrastructure sectors with over 40 networks analysed. It is therefore clear despite the suggestion by Barabasi *et al.* (2003) that spatial networks can indeed be hierarchical.

It is important to note that none of the infrastructure networks fell within the ellipses of the TREE or HC graph models (Figure 4.14, Figure 4.23 and Figure 4.24), while at least 20 infrastructure networks for each metric distribution fell within those for the HR and HR+ graph models. This indicates that such infrastructure networks do not have the repeating deterministic hierarchical structure found in the TREE model, but have some degree of ‘positive’ stochastic redundancy in their hierarchical structure such as the intra and inter level connectivity of the HR and HR+ models (Chapter 3, Section 3.3). This is likely a result of the evolution of the real-world networks, where the hierarchical structure, although not necessarily designed, has emerged through a drive for operational efficiency and economic viability, e.g. short links which provide the service to more users (Gastner and Newman, 2004), along with the evolution of the network over time to meet changing user requirements. Infrastructure networks, while being hierarchical, are closer to those models which do not represent the extremes of hierarchical structure, suggesting any models such as the HR and HR+ as developed in this work could be adopted as better representations of infrastructure networks. These offer the

ability to retain an underlying hierarchical tree structure, though with the flexibility to add new links which break the strict hierarchical rules found in the TREE model.

While the results presented for the infrastructure networks using the metrics of MBC, AC and CB have found them to be hierarchical, the results from the degree distribution comparisons between the infrastructure networks and graph models are less conclusive (Chapter 4, Section 4.3.2). For example, the air networks have degree distributions closest to the scale-free model, but from the analysis of MBC, AC and CB these are shown to be hierarchical (Figure 4.20), and the regional rail networks have distributions most similar to the small-world model, but return metric values most similar to the hierarchical graphs. The observed differences are brought about through the deeper insights the higher-level metrics, such as the betweenness centrality and assortativity coefficient, can reveal about a graph, as these reveal characteristics that are not just topologically based. The degree distribution has also been suggested as being redundant in the characterisation of spatial planar networks (Barthelemy, 2011), with it being stated that for road networks “it is of little interest” due to string spatial constraints resulting in a cut-off in the distribution. Other studies have used alternative methods to the degree distribution to examine the characteristics of graphs, such as the diameter (Albert *et al.*, 1999; Bollobás and Riordan, 2004), centrality measures (Freeman, 1978; Barthélemy, 2004; Newman, 2005; Crucitti *et al.*, 2006) and the cyclic nature of the graphs (Caldarelli *et al.*, 2004; Ginestra and Matteo, 2005; Rozenfeld *et al.*, 2005; Klemm and Stadler, 2006), all allowing greater insight into the graph structure. As such, there is growing evidence that these higher-level metrics should be used in the characterisation of infrastructure networks.

Although the results have found that 37 of the 42 infrastructure networks investigated are hierarchical, at least 17 of the infrastructure networks fell outside of the single standard ellipses of all graph models across the three metrics (Figure 4.14, Figure 4.23 and Figure 4.24). This suggests that the suite of graph models employed could be extended to try and find models which better match some of those infrastructure networks. Recently models have been developed (Barrat *et al.*, 2005; Barthelemy, 2011) which focus on representing specific infrastructure systems where spatial constraints are considered. Specific examples include models for air networks where the distance between nodes is considered in the growth of new links (Wilkinson *et al.*, 2012), with a similar model proposed by Gastner and Newman (2006) for both road and air networks. Such models generate networks which reportedly represent spatial infrastructure networks more realistically, though this field is still emerging with work focused on a select number of infrastructure sectors. The adoption of these models in future research would allow further insights to be gained on the infrastructure networks analysed, with

the role geographic constraints play in the evolution of real-world spatial infrastructure networks being considered in the network models. The extent geographic constraints also play on the structure of and thus characteristics of networks, and the similarity of these to hierarchical models should also be explored.

5.2.3 Robustness analysis

The robustness analysis shows that hierarchical graphs are, on average, less robust to all three failure methods than the non-hierarchical models, failing 27.9% quicker (Chapter 4, Section 4.2.5, page 111). As mentioned in the previous section, 37 of the 42 analysed infrastructure networks appear to be hierarchical, and were found fail in a similar manner to the hierarchical graphs with the infrastructure networks failing on average after 59.8% of nodes have been removed (Chapter 4, Section 4.3.4, page 135), whereas the hierarchical graph models failed after an average of 61.1%. This highlights a critical lack of robustness across the spatial infrastructure networks analysed, making them vulnerable to perturbations. The results have also shown that the infrastructure networks fail 21.3% (Chapter 4, Section 4.3.4) quicker when exposed to the targeted failure methods compared to the random method, indicating a greater vulnerability to the failure of the most critical nodes in the networks. This same susceptibility to the failure of critical nodes in hierarchical graphs/networks was also suggested by Wuellner *et al.* (2010) following the analysis of the networks for a range of air passenger carriers in the USA. This highlights a major vulnerability of the hierarchical infrastructure networks, with poor robustness both in terms of random failures or the failures of critical components as suggested through the targeted failure analysis. With modern societies depending on the services these infrastructure networks provide (Boin and McConnell, 2007; Sterbenz *et al.*, 2011), this result clearly indicates much work is required to strengthen and improve the robustness of these hierarchically structured networks to make them much more robust and dependable.

The HC model exhibited very different failure characteristics to the other hierarchical graph models. The HC model was much more robust to the failure methods, failing 21.7% quicker than the other hierarchical models for the random failure method and an average 48.0% quicker for the targeted failure methods. The structure of the HC model, a hierarchical set of communities, results in the network becoming fragmented into disparate communities (Section 4.2.6 and Figure 4.7(c)), but these themselves remain connected and robust due to a strong level of connectivity within them (Ravasz *et al.*, 2002). This makes these graphs much more robust, with disparate communities still connected and individually robust, giving a model that behaves

more like the non-hierarchical models when perturbed. Employing the HC structure in infrastructure networks could potentially make them more robust to perturbations. The structure would allow for regional parts of the network to continue to function when the network is perturbed where these areas are not affected directly, improving the robustness of those infrastructures which have a different hierarchical organisation.

However, despite the HC model being more robust to failures than the other three hierarchical models, the metric analysis reveals that there are no infrastructure networks which appear to consistently share similar characteristics with this model (Figure 4.14, Figure 4.23 and Figure 4.24, Chapter 4, Section 4.3.3)). As well as this, none of networks exhibit similar failure characteristics to the HC model (Figure 4.25, page 136), again suggesting that this structure is not found in any of the analysed critical infrastructure networks. It is clear the adoption of such a structure for infrastructure networks would improve their robustness to failures, at least where communities of nodes, normally part of the wider network, can continue to function and deliver the service they are intended for. This includes networks such as road networks, where this would enable road users to still travel locally when the network is perturbed, but not to other parts of the network (communities) which have become disconnected. The electricity network is an example of an infrastructure network where the adoption of the HC model is less applicable, with this relying on centralised electricity generation (Bouffard and Kirschen, 2008; Bayod-Rújula, 2009). For this to be robust while having a HC structure, each community would have to have sufficient capacity to meet the electricity demands within it and therefore local generation sites would be imperative. Although there is an increasing trend towards local generation, fuelled by the development of the renewable energy sector (Alanne and Saari, 2006; Bouffard and Kirschen, 2008; Bayod-Rújula, 2009), careful planning would be required to ensure this was achievable.

The most robust infrastructure networks on average across all three failure models were the road networks, failing after an average of 64.8% of nodes had been removed, compared to the least robust, the air networks, which failed on average after only 49.7% of nodes had been removed. The air networks, with such a low average, suggests an inherent weakness to perturbations, with the network being vulnerable any failures, and nearly as vulnerable as the hierarchical TREE model, which on average failed after the removal of 47.1% of nodes. The inherent weakness exhibited by the network, also found by Lordan *et al.* (2014) in an analysis of the global air network, exposes the potential for the significant disruption to these networks, such as seen following the ash cloud generated by a volcanic eruption in 2010 in Iceland (from the Eyjafjallajökull volcano), which caused widespread disruption to the European air network

(Wilkinson *et al.*, 2012). However, unlike the hierarchical TREE model, the air networks do exhibit some redundancy, with them having a large number of CB (3.34) compared to the TREE model which has none. The poor robustness is instead caused by the presence of hub nodes, as indicated by the low AC value and a high MBC, which when removed from the network resulted in the loss of the cycles which would have made the networks more robust to failures. This is clearly exhibited by the targeted failure results, where the air networks failed 48.6% quicker when compared to random failures. This highlights the criticality of hub nodes in some networks, a feature also identified by Lordan *et al.* (2014) and Wuellner *et al.* (2010) in the analysis of air traffic networks, and the need to ensure such structures have redundancy within them to avoid the potential for large-scale disruption through a small number of perturbations.

The road networks, despite exhibiting a similar set of values for the MBC to the air networks and having a lower CB average, were found to be much more robust. These networks aren't reliant on hub nodes (Dorogovtsev and Mendes, 2002), as illustrated by the higher AC value and lower MBC value, and thus there is also a greater redundancy within the network. This allows them to withstand perturbations, whether random or targeted, failing only 13.0% quicker for the targeted method compared to the random failure method. The difference in response highlights the importance and role hub nodes can play in a network, and the effect avoiding such components can have on the ability of a network to respond to failures. It also highlights the need for networks to have redundancy in order to be robust to perturbations (Jenelius, 2010).

The failure results (Chapter 4, Section 4.3.4, page 135) also demonstrated that while the three metrics used in the characterisation of hierarchical networks can help to differentiate between hierarchical and non-hierarchical graphs/networks, they can also help us to understand how a graph, or infrastructure network, may behave when perturbed. The results above suggest that no one metric can be used as a guide to a networks robustness, with instead a combination of these required to understand how a network may respond to perturbations. It is clear that that networks which have MBC values tending to 1 and AC values approaching -1 rather than 1 are the least robust to failures, such as the river and the air networks. Networks with such features begin to rely on hub nodes and also are more likely to lack a high level of redundancy. Combined with a low CB value, the networks can be considered particularly weak when perturbed, as exemplified by the TREE model which has no loops (Table 4.2, page 106) and is the least robust graph model (Figure 4.5, page 114). Although networks with such characteristics may be unavoidable in the real-world through their natural development and the role of hubs in some infrastructure systems, such as air networks, measures are required to improve the robustness of such networks to ensure a more robust structure. Through the

introduction of greater redundancy, the addition of a greater number of alternative links and thus routes through a network, the robustness can be improved. From the analysis it is clear that the two metrics, the number of cycle basis and the maximum betweenness centrality, are intrinsically linked with regard to how they can help to characterise the redundancy and robustness of a network.

Adopting more robust structures for infrastructure networks, such as the earlier mentioned HC model, or altering networks to move away from the characteristics which make networks vulnerable as discussed in the previous paragraphs, will allow networks to continue to function when perturbed more often, lessening the disruption to users. However, the way to achieve this for already developed infrastructure networks, such as those analysed in this research, remains an open question (Little, 2003). With many of the analysed infrastructure networks being closest to the HR and HR+ graph models, a transition away from these models which have shown a lack of robustness, to one which is more robust like the HC model, will allow the networks to withstand certain types of perturbations better. The way in which the transition might occur is however beyond the scope of this work, and instead remains a question for future research. Implementing a transition to more robust network structures through the addition of new assets, links/nodes, naturally incurs costs (Barthelemy, 2003; Royal Academy of Engineering, 2011), potentially inhibiting adaptations being made, and thus restricting the ease and speed at which such changes are made.

5.3 Cascading failure analysis

The cascading failure analysis (Chapter 4, Section 4.4.3, page 148) showed that as with the topological robustness analysis the hierarchical graphs are less robust than the non-hierarchical graphs. Across all six scenarios investigated the hierarchical graphs, HR, HR+, HC and TREE, on average performed worse than the non-hierarchical graphs, with these failing quicker in all six scenarios analysed, reaching equilibrium (not failing) in 14% of simulations compared to 49% for the non-hierarchical graphs. This highlights the lack of robustness present in the hierarchical networks when flows are considered, with them consistently failing quicker. Further to this, in a number of cases, the analysis could not be performed on the hierarchical graphs as the initial flow from the supply to the demand node could not be supported within the capacity constraints of the nodes and edges. This further highlights when comparing the hierarchical and non-hierarchical graphs that the hierarchical graphs have less redundancy within their structure.

As with the topological analysis, the presence of hierarchical redundancy (redundancy within the levels of the hierarchy) within the networks has a strong influence on the ability of the networks to withstand the perturbations. This is highlighted by the response of the TREE, HR and HR+ models to the failure scenarios, where the HR and HR+ which both have hierarchical redundancy exhibited a much greater robustness in three of the six scenarios. Across the three scenarios the HR and HR+ model on average reached equilibrium in 23% for the simulations compared to 0%. Equally, the HR and HR+ models failed on the removal of the trigger edge in 27% of the simulations compared to 33% for the TREE. The ability to support flows over the networks is critical as already mentioned, and thus the results from this analysis clearly indicate that the TREE model is especially weak at supporting flows, especially when perturbed. The HR and HR+ models exhibited a greater robustness despite being based upon the TREE model (Chapter 3, Section 3.3, page 42), with the addition of links within levels of the TREE model and links (shortcuts) between different levels of the TREE hierarchy. These extra links increase the capacity for flows as well as redundancy and hence improve the robustness of models, a feature also shown to work by Helbing *et al.* (2006a) in the analysis of hierarchical management systems.

Of the HR and HR+ models, the HR model exhibited a greater robustness than the HR+ model, with this being more robust across four of the six scenarios. The difference between the two models, the type of links which are added to the base TREE model, has allowed the HR model to be more robust. The HR model has links which form shortcuts between different levels of the hierarchy, whereas the HR+ model only has links between adjacent levels and within a level. The addition of these longer *shortcuts* enables flows to move much more freely through the network, increasing the capacity between levels while also improving redundancy (Helbing *et al.*, 2006a). These longer level links thus appear more effective at improving the redundancy while also improving the capacity of the network, suggesting the HR model as a better model to move towards than the HR+ for infrastructure networks.

The flow based results show that not only does redundancy allow graphs/networks to be more robust to topological failures, but also improves their ability to handle increased volumes of flow and thus also improve the robustness to cascading failures. This has been shown by the inability of the hierarchical models to support some of the scenarios being employed (Chapter 4, Section 4.4.3, page 148). As such, it is even clearer that hierarchical models, such as the TREE model, should be avoided in infrastructure networks as they are particularly weak in their ability to withstand both topological and cascading failures. The HR and HR+ models, the models found to share the most similar characteristics with the infrastructure networks analysed,

were more robust than the TREE model to all failures. There is therefore an indication that with the HC model being most robust, that infrastructure networks should migrate away from the HR+ model towards the HR or HC.

5.4 Analysis of a hierarchical electricity network

The initial analysis of the combined transmission/distribution electricity network has shown that the network is robust for up to three failures in the transmission network (Chapter 4, Section 4.5.2, page 157), with the analysis not considering more failures than this due to the computational overhead of the analysis. The ability of the network to continue to function, with all substations remaining connected to the transmission and distribution hierarchy despite any three transmission line failures, shows that the network has a good degree of robustness within the transmission aspect of the network. This allows electricity to flow from the major sources of generation to all points of the network ensuring demand can be met. This robustness ensures that multiple failures cannot disrupt the functionality of the network.

The results from this spatial failure analysis (Chapter 4, Section 4.5.3, page 160) revealed that as the number of hazard areas increased and the number of initially affected infrastructure node assets remained similar, the number of second-order failures rose, suggesting a vulnerability to multiple hazards on the network. This was shown by an increase of second-order node failures on average from 361 for a single hazard (Table 4.13, page 165), to 5901 for four hazards (Table 4.16, page 169) and 6895 where there was eight hazards (Table 4.19, page 174). This clearly indicates a susceptibility to multiple hazards, and suggest that the network is more vulnerable to many smaller hazards rather than single large hazards. This makes the network vulnerable to events such as flooding, where multiple disparate locations can be affected at the same time (Suarez *et al.*, 2005), or during wind storm events where individual network assets within a wider area can be affected. These sort of events are also expected to increase in frequency due to climate change (Royal Academy of Engineering, 2011), and thus having infrastructure networks which are more robust to such events is becoming more critical. This vulnerability is caused by the larger hazards causing the failure of the higher-level substations and those which depend on them, unlike the smaller hazards where each only result in first-order failures for a few substations, thus leaving dependent substations to fail as second-order failures, rather than first. The smaller hazard areas can cause large failures by affecting the key substations which many nodes are dependent upon, and are just as likely to do this as the larger hazard areas, with in each scenario the hazards affecting the same number of nodes.

Previous analysis of the vulnerability to spatial hazards has shown that as the size of the hazard increases, the greater the impact on the network (Sterbenz *et al.*, 2010; Wilkinson *et al.*, 2012; Ouyang, 2016). Although not directly comparable, these results differ slightly from the results found in this work. Although Wilkinson *et al.* (2012) did not consider multiple hazards, they did analyse how the size and location of the hazard affects the robustness of the European air network. This work showed that location of the hazards was an important factor in the effect of a hazard, with a small hazard in the right location able to create the same disruption as a larger hazard. This, to some extent, has also been found in the analysis of the electricity network, with those hazards in rural locations having a larger effect due to the lack of redundancy in these areas (Chapter 4, Section 4.5.3, page 160). The spatial distribution of the hazards within the research was random, though exploring the effects of the clustering of hazards will provide different insights into how robust the network is to such events, providing more detail on how the network responds to perturbations. This extra insight can help in future decisions on how to improve the robustness of the network, ensuring potential adaption strategies consider a wide spectrum of results.

As the number of hazard footprints increased, the average distance of the second-order failures from these areas also increased; from an average of 4.3km for the single hazard scenarios (Table 4.14) to 10.0km for the eight hazard scenarios (Table 4.20). This again indicates a susceptibility to multiple hazards, with the smaller the hazards, the greater the propagation of failures following the initial failures within the hazard area. The propagation of failures is caused by a lack of robustness, with lower level (voltage) substations only being supplied by a single, or a small number, of high-level substations, resulting in a large number of second-order failures when these supply options are all removed following a spatial hazard. This is highlighted by the results, where on average 93% of the second-order failures in the simulations for the eight hazard scenarios were 11Kv substations, the lowest level in the network. Therefore, only 7% of the failed substations were those responsible for the further distribution of electricity to other demand points, limiting the total impact on the network and the further propagation from the initial hazard areas.

The identified susceptibility to multiple hazards raises many issues regarding the robustness of the electricity transmission and distribution network. Improving the robustness of this is critical, with the ability to change how the network responds to perturbations and the service the network is still able to deliver to customers. The hierarchical structure of the network, most similar to the HR+ graph model (Figure 4.15, page 126), suggests that the network structure is one of the more robust hierarchical structures (Figure 4.5, page 114), though improvements can

still be made. This could include adopting a structure more like the HC model which would allow communities within the network to continue to function given the adequate provision of electricity generation facilities within the community. The robustness of the HC model to spatial hazards has not been explored, though the results from the topological failure analysis suggest this is a more robust structure (Figure 4.5, page 114), so may improve the response to multiple hazards. The community structure may reduce the second-order failures by providing greater redundancy to the substations within the communities, resulting in the lowest level, and second-lowest level substations being dependent on more than just one substation, improving robustness. A transition to towards the HC model is also supported by a trend towards a more decentralised electricity generation systems (Alanne and Saari, 2006; Bouffard and Kirschen, 2008; Bayod-Rújula, 2009), with more electricity being generated within local communities. This offers the ability for electricity systems to become less dependent on centralised generation facilities, and hence local communities maybe able to provide their own energy, improving the potential for these to be robust to failures which result in their disconnection from the central transmission network.

5.5 Future analysis opportunities

The previous sections have discussed the outcomes from the research undertaken, identifying a hierarchical organisation within many spatial critical infrastructure networks, a structure that also makes the networks vulnerable to random, targeted and cascading failures. Within the analysis performed further opportunities to learn more about the behaviour of the hierarchical graphs and infrastructure networks are identifiable, with some clear opportunities arising following this work.

The failure based analysis undertaken in this research can be extended in multiple ways, building on the methods already developed. The flow based cascading failure analysis undertaken can be extended to include more detail in the modelling of the flows, with the ability to model features such as buffering and latency (Evans, 2010; Eusgeld *et al.*, 2011; Filippini and Silva, 2014) already within the developed model. These allow a greater analysis of how the network assets, nodes and edges, respond to failures, and can provide more a in-depth understanding of how the system as a whole may behave. There is also the capacity for more physically based modelling to be undertaken, with the functionality for modelling the behaviour of flows over the network, providing much more detail in the behaviour of the system when it is perturbed. Linked to this is also ability to use more than one supply node and demand node, with many networks featuring multiple of these, including networks such as that for electricity

(Carreras *et al.*, 2002; Dueñas-Osorio and Vemuru, 2009) and transport, functionality which again is already possible with the developed framework and model. Performing this more in-depth analysis over the networks will allow for a greater understanding of the characteristics of the infrastructure networks, and in particular of how the hierarchical structured networks behave when a more detailed flow based analysis is undertaken to test their robustness to cascading failures.

The failure analysis performed in this work is based on the assumption that each asset shared the same level of vulnerability when exposed to a hazard, namely that it would fail. This is however not always the case, with some network assets being more robust than others, with the point of failure varying from asset to asset (Dueñas-Osorio and Vemuru, 2009). Assessing the robustness of a network to hazards with this extra consideration adds a further dimension to the analysis, allowing more in-depth analysis of the robustness of individual infrastructure networks. Previous research has identified and explored methods through which this can be done, especially when analysing the robustness to cascading failures, through assigning each asset a value with the probability of failure when exposed to a hazard/overloaded (Motter, 2004; Bao *et al.*, 2009b; Dueñas-Osorio and Vemuru, 2009). Within the framework developed for this work (Chapter 3, Section 3.11, page 84) this functionality can be easily added to work within the existing failure models. Adopting this method would improve the analysis of hierarchical networks to spatial failures, allowing for new understanding to be learned on the ability of hierarchically organised networks to withstand perturbations, with individual assets modelled with greater realism.

As mentioned in the previous sections there is a need to improve the robustness of critical infrastructure networks which share characteristics with graph models such as the TREE, HR and HR+. Achieving this has been discussed by adopting a structure similar to the HC model to allow a move away from the characteristics which make networks vulnerable to failures. How this can be achieved has not been assessed in this research and is in itself a major research undertaking. However, such an analysis may offer potential insights on how improving network structure can assist in the overall robustness of different infrastructure systems. Further options, including improving the redundancy of the network or improving the ability of individual assets to withstand hazards (Little, 2003; Beygelzimer *et al.*, 2005) should also be explored, with these also being areas that have been recognised as requiring further work within infrastructure systems research (Little, 2003).

All of the failure analysis undertaken in this research on the graphs and spatial infrastructure networks has focused on the robustness of these to a range of failures scenarios (Bruneau and

Reinhorn, 2007; McDaniels *et al.*, 2008). This can be extended though to consider the resilience of the networks, how they recover to full functionality following the failures (Reed *et al.*, 2009; Ouyang *et al.*, 2012; Hosseini *et al.*, 2016), providing greater insights into the ability of the networks to continue to function when perturbed. This will help to assess the ability of hierarchical networks to respond to failures with a greater realism, with the ability of nodes/edges affected by a hazard to be repaired and become available for use again within a network critical to this. The functionality to perform this analysis is not currently within the developed framework, though extending it to enable this analysis would be relatively straight forward.

5.6 Conclusion

The results from the analysis have made it clear that hierarchical graphs and networks are distinct from non-hierarchical graphs, with three key higher-level metrics, the assortativity coefficient, the maximum betweenness centrality and number of cycle basis per node, allowing these to be distinguished from one another. These high-level metrics provide a deeper insight than traditional topological methods, and have identified many infrastructure networks as possessing a hierarchical structure, in many cases contrary to previous research. The implications of infrastructure networks being hierarchically structured has been examined with regard to the robustness of hierarchical graphs/networks to perturbations, showing these are less robust than non-hierarchical graphs, to both topological failures and cascading failures. The HC hierarchical model was an exception, with this appearing to be as robust as the non-hierarchical models, with the community structure making it robust to failures through the ability of the communities, which the graph fragmented into, to stay connected within. With many infrastructure networks appearing to be hierarchical, this indicates the need to improve the robustness of hierarchical graphs, and thus infrastructure networks, to perturbations. Methods have been discussed, from increasing the redundancy in the networks, to adopting the HC structure, with the applicability for different infrastructure networks mentioned. Further to this, through a series of scenarios with spatial hazards, the hierarchical electricity transmission and distribution network for England and Wales was shown to be susceptible to multiple hazards, while the initial number of affected assets remained constant, highlighting the importance of considering such failure scenarios. The network was robust at the transmission level, with the lack of robustness identified in the lower-levels of the distribution network, making this the area of the network which needs improving.

Chapter 6: Conclusion

6.1 Introduction

This chapter presents the main findings of the research presented in the previous chapters. The research aimed to identify the hierarchical organisation of critical spatial infrastructure networks and the robustness of these to a range of failure scenarios. The following objectives were set out in order to address the aim:

1. Review the research field pertaining to hierarchical networks and graph models and their application in the analysis of critical spatial infrastructure networks.
2. Investigate the properties of hierarchical graphs to identify the characteristics which makes them recognisable from non-hierarchical graphs.
3. Identify examples of hierarchically organised critical spatial infrastructure networks using the outcomes from objective 2.
4. Explore the robustness of hierarchical infrastructure networks to perturbations and the reasons why such networks behave differently to those of other topological structures.

The main findings are presented in the following section, Section 6.2, with areas of future work given in Section 6.3.

6.2 Main findings

6.2.1 *Hierarchical networks and critical infrastructures*

The research began with a review of literature around the structure and robustness of critical infrastructure networks (objective 1). This found that there was a growing body of research into the hierarchical organisation in networks, with a clear understanding of the organisation found in biological and social networks. To this end, models had been developed to model such networks allowing improved understanding of their characteristics and behaviour (Ravasz *et al.*, 2002; Ravasz and Barabasi, 2003). However, there is only a small body of literature which has addressed the possible hierarchical organisation of critical spatial infrastructure networks. The review found that the structure of hierarchical networks was such that the levels of redundancy in networks in some cases was not as strong as non-hierarchical models, leading to potentially less robust networks (Helbing *et al.*, 2006a; Helbing *et al.*, 2006b). With all critical infrastructure networks embedded in space, a review of the literature where the robustness to

spatial hazards has been analysed showed only a small selection of studies have addressed this important issue. The findings of this review have highlighted a significant amount of research in the field of infrastructure robustness to perturbations, though it has also found that the presence and impact of a hierarchical organisation of infrastructure networks has not been examined in detail despite an emerging literature on the subject of hierarchically organised networks.

6.2.2 Hierarchical graphs

Objective 2 was addressed through the analysis of a suite of non-hierarchical and hierarchical graphs, allowing for the structure and characteristics to be compared. A range of methods were employed initially, with the use of three graph metrics (assortativity coefficient, maximum betweenness centrality and number of cycle basis per node), appearing to provide a method of recognising those graphs with a hierarchical organisation. These metrics were selected based upon the known structure of the hierarchical and non-hierarchical graph models as identified in Chapter 2. A statistical difference was shown to exist between the hierarchical and non-hierarchical graphs using the metrics, with two, the maximum betweenness centrality and the number of cycle basis per node, shown to be the clearest indicators of a possible hierarchical organisation.

Using a developed topological based failure models, similar to those reviewed in Chapter 2, the robustness of hierarchical and non-hierarchical networks was explored showing that three of the four models, the tree models and those based on it, exhibited a poor robustness to perturbations. The fourth model on the other hand, based on a hierarchy of communities, was much more robust to failures locally due to a modular community structure, though globally became disconnected very quickly, similar to the failure behaviour exhibited by the other hierarchical models.

6.2.3 Hierarchical infrastructure networks

A suite of spatial critical infrastructure networks was analysed using the results from objective 2 with the same methods employed again, in order to address objective 3. Over 40 infrastructure networks were analysed and compared to the suite of synthetic graphs, with the majority of infrastructure networks found to share the same characteristics exhibited by the hierarchical graph models including road networks and air networks. A further comparison was made through the results of the topological robustness analysis undertaken over the networks where

it was found that most exhibited a behaviour when perturbed similar to the hierarchical graphs; the least robust graph models analysed.

6.2.4 Robustness of hierarchical infrastructure networks

Infrastructure networks are designed to deliver a service which results in the movement of a commodity/information over the network. An analysis of the hierarchical and non-hierarchical models with a developed capacity constrained cascading failure model was used to investigate the robustness of the different models with flows on them and their susceptibility to cascading failures, Objective 4. This analysis found that the hierarchical models were less robust to the cascading failures than the non-hierarchical models, with the hierarchical communities model shown to be the most robust hierarchical model. The other three hierarchical models, were as in the case of the topological failures, the least robust to the cascading failures.

The robustness of a hierarchical spatial critical infrastructure network, the electricity transmission and distribution network for England and Wales, was explored to both hierarchical failures and spatial hazards. The network was first subjected to a failure model that targeted network assets at the highest level of the hierarchy, exploring the ability to the network to still function when such assets were unavailable. This analysis shows that the network was robust for all combinations of three 400Kv transmission lines failing, indicating a high level of tolerance to such failure events. The robustness to spatial hazards was then explored by applying a series of different spatial configurations of spatial hazards to the network, where those assets within the hazard footprints failed (restricted to 2% in all scenarios), and the failures caused by these assessed. The analysis showed that as the number of areas increases the number of second-order failures, substations outside of the hazard areas that become disconnected, increased. This has highlighted a sensitivity to an increasing number of hazards, with a greater number of failures occurring as the number of hazard areas increased despite the number of initially affected assets remaining the same. However, the network was found to be robust to the hazards with regard to how far failures propagated, with second-order failures tending to occur close to the hazards except in some rural areas where the density of network assets is low.

6.3 Future work

6.3.1 Geographic and spatial graph models

The graph suite used to explore characteristics of hierarchical networks consisted of eight graph models, all of which generated networks in graph space with no consideration for geographic

patterns or the cost of building links between nodes over spatial distances, all of which have been shown to affect the development of real-world infrastructure networks (Barrat *et al.*, 2005; Barthelemy, 2011). Some studies have suggested that such constraints effect the characteristics of the real-world networks including their degree distributions (Herrmann *et al.*, 2003) and to thus replicate this graph models need to consider the same geographic/spatial constraints (Barthelemy, 2003; Gastner and Newman, 2004). To this end, there is an emerging suite of models (Gastner and Newman, 2004; Masuda *et al.*, 2005; Gastner and Newman, 2006; Wilkinson *et al.*, 2012; Fu *et al.*, 2015) which attempt to consider some the geographic/spatial constraints such as the length of links.

The inclusion of graphs generated by models which explicitly consider geographic/spatial constraints may present characteristics which better match some of the infrastructures analysed in the research presented in this thesis. For example, the model developed by Wilkinson *et al.* (2012) has been designed for air networks which one may expect to better match the characteristics exhibited by the suite of air networks employed in this thesis. This may help in a better characterisation of the hierarchical organisation in critical infrastructure networks, allowing for a greater understating of their characteristics as well as robustness to perturbations and hence how such networks may be made less vulnerable to failures.

6.3.2 Cascading failures on hierarchical graphs

The cascading failure model developed (Chapter 3, Section 3.9 (page 71)) to examine the robustness of hierarchical networks has been used over a suite of synthetic graphs. The analysis undertaken (Chapter 4, Section 4.4 (page 142)), has provided some insights into the robustness of the eight different graph models, showing that the random graphs are much more robust than any other models and the four hierarchical graph models are extremely vulnerable to cascading failures. The designed analysis only employed six scenarios and thus further insights into the behaviour and the robustness of the hierarchical models to cascading failures can be learned through a more detailed analysis driven by a greater range of scenarios and graphs being utilised. Example future scenarios should consider the dynamics of the hierarchical graphs in more detail, including the use of multiple supply and/or demand nodes in the graphs, such as in energy networks where generators supply electricity demand (Chang and Wu, 2011). Through the application of these methods to both hierarchical and non-hierarchical graphs, the characteristics which make graphs robust to cascading failures can be further understood, with the potential to develop an understanding of how to improve the least robust graphs.

6.3.3 Modelling of infrastructure flows

Critical infrastructure networks provide a service to users (Little, 2002; Little, 2003), be that the delivery of commodities, data or methods of transport (Murray, 2013; Ouyang, 2014). These travel through infrastructure networks as flows, the movement of commodities/data over edges and through nodes (Murray, 2013), with such assets designed to handle a certain capacity (Murray *et al.*, 2008; Ouyang, 2014). Therefore, the movement of flows through a network is dictated by the properties of the network assets such as their capacity. Through not only using the assets in a network, but also the attributes of these, the flows through the network can begin to be simulated in greater detail.

A developed capacity constrained cascading flow model has been developed (Chapter 3, Section 3.9) which allows for the explicit modelling of flows over networks. This supports physically based models, where the behaviour of the flows are modelled following perturbations to the network. However, throughout the analysis undertaken in this thesis the potential of this model has not been fully utilised due to insufficient data being available for the parameterisation of the infrastructure networks investigated. Future work should thus look to utilise this developed model further through compiling more complete datasets which will facilitate physically based flow modelling to be undertaken on critical infrastructure networks. This will allow for the further analysis of hierarchical critical spatial infrastructure networks to be undertaken improving our knowledge of their robustness to failures. From those analysed lessons can be learned as to how the different characteristics of the networks affects their robustness.

6.3.4 Spatial hazard modelling

Critical spatial infrastructure networks are exposed to a large number of natural hazards, most of which affect a spatial area and can be modelled as such (Li *et al.*, 2016). However, a review of the literature (Chapter 2) highlighted that only a small number of studies have examined explicitly the robustness of spatial critical infrastructure networks to spatial hazards, with Barthélemy (2011) noting that many studies neglect the spatial aspects of networks. This thesis has presented a case study whereby the robustness of a hierarchical critical infrastructure network was analysed with a range of spatial hazard scenarios used to perturb the network, exploring the robustness of the infrastructure network to different configurations of spatial

hazards. However, flows through the network were not modelled due to insufficient data for the parametrisation for this.

Spatial hazards can come in different sizes, from a local flood in a street to extreme windstorms. With critical infrastructure networks often covering large spatial domains, the network can be exposed to many hazards of different size simultaneously. Although this study used three different scenarios which varied in the number and size of spatial hazards much more detailed analysis can be undertaken with an extended scenario set to further explore the robustness of networks to multiple hazards over their spatial domain. This could include a deeper exploration of the effect the spatial configuration of hazards has on the robustness of networks as well as scenario data from past events or data from projections for future events.

A greater appreciation of the characteristics of the network assets could also be incorporated, with each treated individually with regard to its exposure to a hazard, with each asset potentially having a different critical failure threshold resulting from additional engineered protection for example. Such in-depth analysis would be specific to an infrastructure network, but would provide a greater depth of understanding on the network's robustness and how this could be potentially improved through the identification of less robust assets. Given the developed capacity constrained cascading failure model, the spatial hazard modelling could also be coupled with this if the correct data for the parameterisation of the networks was available. This would allow for a much more detailed analysis of the robustness of hierarchical critical spatial infrastructure networks with the loss of supply and demand assets in the network being handled explicitly. The most vulnerable areas of the infrastructure networks with regard to flows through them are more likely to be identified through this modelling approach, allowing those critical assets to be either hardened to failures or the redundancy improved to provide alternative routes for flows near the most vulnerable areas.

6.3.5 Improving the robustness of hierarchical networks

The results presented in this thesis have highlighted the existence of a hierarchical organisation in some critical infrastructure networks, such as rivers, railways, energy and air flight networks. The robustness analysis also undertaken has shown that the hierarchical models, especially those based around the tree model, are vulnerable to both targeted and random failures. Together this suggests an inherent lack of robustness in some of the critical infrastructure networks to perturbations, which given the results, could have catastrophic consequences on the services they deliver.

Given these findings, there is clearly a need to improve the robustness of hierarchical infrastructure networks to give them a greater robustness to failures. This can be achieved through increasing the redundancy within the networks (Doyle *et al.*, 2005; Haines, 2009; Jenelius, 2010), such as through the addition of new edges (Helbing *et al.*, 2006a; Wuellner *et al.*, 2010). However, in spatial critical infrastructures there is a financial cost associated with the construction of each new edge (or node) asset in the network (except for networks of air flights) (Barrat *et al.*, 2005; Barthelemy, 2011), so the addition of new assets must to be optimised against the cost and the improvement in robustness that they offer. The rewiring of networks has also been analysed as a method of improving robustness in networks (Beygelzimer *et al.*, 2005), though in many critical spatial infrastructure networks this is again limited by the costs associated with the development of new connections between locations. Therefore, the development of an optimisation approach/software for improving redundancy while minimising the economic expenditure in hierarchical critical infrastructure networks could provide potential solutions to improving the robustness of networks. Through integrating such software with tools similar to those developed for this research the improvement in robustness could also be quantified.

6.3.6 Robustness to dependencies in hierarchical infrastructures

As noted in Chapter 1, infrastructure networks do not operate in isolation, but instead form an interconnected web of networks providing services for users (Rinaldi *et al.*, 2001; Vespignani, 2010), and other infrastructure networks in some cases (Rinaldi *et al.*, 2001). Dependencies, where one infrastructure network relies upon another (Rinaldi *et al.*, 2001; Rinaldi, 2004), can affect the functioning of infrastructure systems and networks, and thus it is important to consider these when analysing the robustness/resilience of critical spatial infrastructure networks (Rinaldi *et al.*, 2001). Therefore, by including and modelling dependencies between infrastructure networks a better understanding of their robustness to failures can be developed (Zimmerman, 2001; Rinaldi, 2004; Dueñas-Osorio, 2005). Models have previously been developed to perform such modelling (Dueñas-Osorio *et al.*, 2007b; Havlin *et al.*, 2010; Johansson and Hassel, 2010), although the focus has not been on hierarchical infrastructure networks and instead focused on specific infrastructure networks. However, such modelling requires knowledge of the dependencies existing between infrastructure networks and therefore the success of such an approach would be heavily dependent on such knowledge and data being available.

6.3.7 Software framework

A developed framework (Chapter 3, Section 3.11.1, page 84)) has been used to undertake the research presented in this thesis. Included in this is a suite of tools and modules which have been developed as detailed in Chapter 3 Section 3.11.4, page 92), and have enabled the analysis and simulations which have been undertaken. The data for the analysis, including the graphs and networks, have been stored in a developed database (Chapter 3, Section 3.11.2 and 3.11.3), with these being easily accessible to the developed modules and tools through developed database wrappers. As well as storing the input data, the database has also stored the results from the analysis, with these written in the database following each simulation. The database schemas include over 40 developed procedural SQL database functions which allow the wrappers to effectively read and write networks and data to and from the database.

The developed software framework has underpinned the research undertaken, and it is suggested that future work should look to develop the framework extending its capability and performance further. With the size and complexity of critical infrastructure networks ever increasing (Klein *et al.*, 2008; Agarwal *et al.*, 2014), the ability for a framework to be able to effectively manage and support the analysis of large networks is critical. The largest network analysed in the research presented in this thesis had less than 200,000 nodes. However it is known that much larger infrastructure networks exist with node counts greater than 1million (Newman, 2003b). Relational databases become slow as the size of the networks increase (due to the need to join data from different parts of the database), though new and emerging technologies such as graph databases offer the potential for the more efficient storage, management and analysis of large networks (Álvarez *et al.*, 2010; Dominguez-Sal *et al.*, 2010). High performance computing methods such as cloud computing offer greater computing power (Armbrust *et al.*, 2010) and thus the potential to better support the analysis of the increasingly large critical infrastructure networks. These new, emerging and still developing technologies should be adopted where faster solutions are required.

6.4 Key findings and implications

The importance of national critical infrastructure is evident through the dependence on the services they provide including energy, water and communications. However, the growing complexity of modern infrastructure networks makes understanding their robustness to perturbations increasingly difficult in a world where the number of hazards which they are exposed to are increasing. As an aid graph models have been used to represent complex infrastructure networks to improve our understanding of the behaviour of the networks when

perturbed by different events, from the breakdown of components to natural hazard events. More recently new models have been developed with an explicit hierarchical structure and employed for the analysis of networks from social and biological sciences, where it has been shown that such networks do have a hierarchical organisation. Some studies have also suggested that spatial critical infrastructure networks exhibit a hierarchical organisation, though this analysis has been limited.

This thesis has presented research which has examined the robustness of hierarchical critical infrastructure networks. The characteristics of hierarchically organised graphs was first identified prior to an analysis of their robustness to perturbations. Using the identified characteristics critical infrastructure networks were then analysed to identify the presence of a hierarchical organisation within their topological structures. The robustness of these to perturbations, using the same methods as employed on the graphs, was also explored as a measure of the impact of critical infrastructure networks being hierarchically organised. An example hierarchical network, the electricity transmission and distribution network for England and Wales, was used to explore how the hierarchical organisation and the hierarchical nature of the flows through the network, are effected by hierarchically targeted perturbations and spatial hazards.

The results presented in Chapter 4 and discussed in Chapter 5 have shown that a hierarchical organisation exists in some critical infrastructure networks, most notably the road and air infrastructure networks. The hierarchical organisation of networks has also been shown to make them less robust to failures compared to non-hierarchical networks, with a decreased level of redundancy in the networks and a greater reliance on a single critical node, a hub, or a small subset of hub nodes. However, the hierarchical electricity network for England and Wales has been shown to be robust topologically to targeted hierarchical failures as well as spatial hazards, with the effects of spatial hazards and in particular the single hazards, at least topologically to the second-order failure level, being limited with regard to the size of the network affected.

The outcomes of this research highlight the need for greater redundancy within critical infrastructure systems, with the addition of new edges between nodes which are not considered to be hubs suggested, lessening the impact of the removal of existing hub nodes. This also offers the potential to decrease the pressure/reliance on the existing hub nodes, making them less critical to the functioning of the network. This single strategy would likely improve the robustness of infrastructure networks across multiple sectors, including rail and energy networks where employed strategically. Although the addition of new assets is inhibited by the economic cost of construction, the benefits possible through the greater redundancy which

could be achieved go beyond the networks being more robust, but could also allow the potential for less inconvenience when assets are closed for maintenance/repair/replacement amongst other things.

Policies otherwise should look to encourage a move away from the reliance on hub nodes and a strict hierarchical structure, with greater flexibility required to achieve a greater robustness to perturbations. A move towards a community/modular structure, for example as seen in the HC model, would allow infrastructure networks to retain a hierarchical structure while having a greater robustness to failures. However, this is limited to those systems which are not dependent on a system wide network, such as electricity and gas, where supplies are transmitted at a national scale, and instead is better suited to networks such as those for air, road and rail where the systems can still function at a local scale when the network is perturbed.

Understanding the dependency of network assets within a network is critical, with this highlighted by the extent at which some failures could propagate through the electricity network for England and Wales given an exposure to spatial hazards. Failures were exacerbated by a lack of redundancy in locations, especially in more rural areas, allowing the impact of an initial hazard to spread beyond the boundaries of the hazard. Identifying locations where redundancy is weak enables investment to be directed at these locations strengthening the network and its ability to withstand perturbations. This work would help reduce the identified susceptibility to multiple hazards identified in the electricity network, a characteristic making the network vulnerable, especially to climate based events, such as flooding and extreme temperatures, which are forecasted to increase in frequency and inherently affect multiple areas rather than a single point location.

This research has provided new knowledge on the characteristics of hierarchically organised graphs as well as identifying those infrastructure networks which present a hierarchical organisation. The properties which cause non-hierarchical networks to exhibit a greater robustness than the hierarchical networks have also been identified. These findings can be used to identify further infrastructure networks with a hierarchical organisation which given the robustness of the topology is critical. Further to this, from the findings reported in this thesis, the robustness of hierarchical networks is now much better understood allowing for the potential to improve this using the identified weaknesses in the hierarchical organisation of networks when compared to the more robust non-hierarchical networks. To this end, some suggestions have been made as to how infrastructure networks might be made more robust to perturbations, including around improving the redundancy within the networks.

Bibliography

- 107th Congress (2001) 'USA Patriot Act (Uniting and Stengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001)' *Public Law*. pp. 1-132.
- Agarwal, J., Liu, M. and Galvan, G. (2014) 'Vulnerability and Resilience of Networked Infrastructures', in *Vulnerability, Uncertainty, and Risk*. American Society of Civil Engineers, pp. 2811-2820.
- Ahmed, A., Dywer, T., Hong, S.-H., Murray, C., Song, L. and Wu, Y.X. (2005) 'Visualisation and Analysis of Large and Complex Scale-free Networks', *IEEE VGTC Symposium on Visualization*. 1-8.
- Alanne, K. and Saari, A. (2006) 'Distributed energy generation and sustainable development', *Renewable and Sustainable Energy Reviews*, 10(6), pp. 539-558.
- Albert, R., Albert, I. and Nakarado, G.L. (2004) 'Structural Vulnerability of the North American Power Grid', *The American Physical Society*, 69, pp. 1-4.
- Albert, R. and Barabasi, A.-L. (2002) 'Statistical mechanics of complex networks', *Reveiws of Modern Physics*, 74, pp. 47-97.
- Albert, R., Jeong, H. and Barabasi, A.-L. (1999) 'The diameter of the world wide web', *Nature*, 401.
- Albert, R., Jeong, H. and Barabasi, A.-L. (2000) 'Error and attack tolerance of complex networks', *Nature*, 406, pp. 378-382.
- Álvarez, S., Brisaboa, N.R., Ladra, S. and Pedreira, Ó. (2010) *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*. ACM.
- Amaral, L.A.N. and Ottino, J.M. (2004) 'Complex networks', *The European Physical Journal B - Condensed Matter*, 38, pp. 147-162.
- Amaral, L.A.N., Scala, A., Barthelemy, M. and Stanley, H.E. (2000) 'Classes of small-world networks.', *Proceedings of the National Academy of Sciences of the United States of America*, 97, pp. 11149-11152.

- Andersson, G., Donalek, P., Farmer, R., Hatziaargyriou, N., Kamwa, I., Kundur, P., Martins, N., Paserba, J., Pourbeik, P., Sanchez-Gasca, J., Schulz, R., Stankovic, A., Taylor, C. and Vittal, V. (2005) 'Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance', *Power Systems, IEEE Transactions on*, 20, pp. 1922-1928.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Petterson, D., Rabkin, A., Stoica, I. and Zaharia, M. (2010) 'A View of Cloud Computing', *Communications of the ACM*, 53, pp. 50-58.
- Aruoba, S.B. and Fernández-Villaverde, J. (2015) 'A comparison of programming languages in macroeconomics', *Journal of Economic Dynamics and Control*, 58, pp. 265-273.
- Ash, J. and Newth, D. (2007) 'Optimizing complex networks for resilience against cascading failure', *Physica A: Statistical Mechanics and its Applications*, 380, pp. 673-683.
- Bagler, G. (2008a) 'Analysis of the airport network of India as a complex weighted network', *Physica A: Statistical Mechanics and its Applications*, 387, pp. 2972-2980.
- Bagler, G. (2008b) 'Complex Network view of performance and risks on Airport Networks', *Airports: Performance, Risks, and Problems*, pp. 1-7.
- Balijepalli, C. and Oppong, O. (2014) 'Measuring vulnerability of road network considering the extent of serviceability of critical road links in urban areas', *Journal of Transport Geography*, 39(0), pp. 145-155.
- Bao, Z.J., Cao, Y.J., Ding, L.J. and Wang, G.Z. (2009a) 'Comparison of cascading failures in small-world and scale-free networks subject to vertex and edge attacks', *Physica A: Statistical Mechanics and its Applications*, 388, pp. 4491-4498.
- Bao, Z.J., Cao, Y.J., Wang, G.Z. and Ding, L.J. (2009b) 'Analysis of cascading failure in electric grid based on power flow entropy', *Physics Letters A*, 373, pp. 3032-3040.
- Barabasi, A.-L. and Albert, R. (1999) 'Emergence of Scaling in Random Networks', *Science*, 286, pp. 509-512.
- Barabási, A.-L., Ravasz, E. and Vicsek, T. (2001) 'Deterministic scale-free networks', *Physica A: Statistical Mechanics and its Applications*, 299(3-4), pp. 559-564.

- Barabasi, A.-L., Zoltan, D., Erzsebet, R., Soon-Hyung, Y. and Zoltan, O. (2003) 'Scale-Free and Hierarchical Structures in Complex Networks', *AIP Conference Proceedings*, 661, pp. 1-16.
- Barabasi, A., Albert, R. and Jeong, H. (2000) 'Scale-free characteristics of random networks: the topology of the world-wide web', *Physica A: Statistical Mechanics and its Applications*, 281, pp. 69-77.
- Barabasi, A. and Oltvai, Z.N. (2004) 'Network biology: Understanding the cell's functional organization', *Nature Reviews Genetics*, 5(2), pp. 101-113.
- Barr, S.L., Alderson, D., Robson, C., Otto, A., Hall, J., Thacker, S. and Pant, R. (2013) 'A National Scale Infrastructure Database and Modelling Environment for the UK', *International Symposium for Next Generation Infrastructure*. Wollongong, New South Wales, Australia.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R. and Vespignani, A. (2004) 'The architecture of complex weighted networks', *Proceedings of the National Academy of Sciences of the United States of America*, 101, pp. 3747-3752.
- Barrat, A., Barthelemy, M. and Vespignani, A. (2005) 'The effects of spatial constraints on the evolution of weighted complex networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2005, pp. 1-20.
- Barrat, A. and Weigt, M. (2000) 'On the properties of small-world network models', *The European Physical Journal B - Condensed Matter and Complex Systems*, 13(3), pp. 547-560.
- Barthelemy, M. (2003) 'Crossover from scale-free to spatial networks', *Europhysics Letters*, 63(6), pp. 915-921.
- Barthelemy, M. (2011) 'Spatial networks', *Physics Reports*, 499, pp. 1-101.
- Barthélemy, M. (2004) 'Betweenness centrality in large complex networks', *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2), pp. 163-168.
- Barthelemy, M. and Amaral, L.A.N. (1999) 'Small-world networks: Evidence for a crossover picture', *Physical Review Letters*, 82, pp. 3180-3183.
- Bassett, D.S. and Bullmore, E. (2006) 'Small-world brain networks', *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 12, pp. 512-23.

- Bayod-Rújula, A.A. (2009) 'Future development of the electricity systems with distributed generation', *Energy*, 34(3), pp. 377-383.
- Berizzi, A. (2004) *Power Engineering Society General Meeting, 2004. IEEE*. 10-10 June 2004.
- Beygelzimer, A., Grinstein, G., Linsker, R. and Rish, I. (2005) 'Improving network robustness by edge modification', *Physica A: Statistical Mechanics and its Applications*, 357(3-4), pp. 593-612.
- Birmelé, E. (2009) 'A scale-free graph model based on bipartite graphs', *Discrete Applied Mathematics*, 157, pp. 2267-2284.
- Bobbio, A., Bonanni, G., Ciancamerla, E., Clemente, R., Iacomini, A., Minichino, M., Scarlatti, A., Terruggia, R. and Zendri, E. (2010) 'Unavailability of critical SCADA communication links interconnecting a power grid and a Telco network', *Reliability Engineering & System Safety*, 95(12), pp. 1345-1357.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.U. (2006) 'Complex networks: Structure and dynamics', *Physics Reports*, 424, pp. 175-308.
- Boin, A. and McConnell, A. (2007) 'Preparing for critical infrastructure breakdowns: the limits of crisis management and the need for resilience', *Journal of Contingencies and Crisis Management*, 15, pp. 50-59.
- Bollobás, B. and Riordan, O. (2004) 'The Diameter of a Scale-Free Random Graph', *Combinatorica*, 24(1), pp. 5-34.
- Bompard, E., Napoli, R. and Xue, F. (2009) 'Analysis of structural vulnerabilities in power transmission grids', *International Journal of Critical Infrastructure Protection*, 2, pp. 5-12.
- Bompard, E., Wu, D. and Xue, F. (2011) 'Structural vulnerability of power systems: A topological approach', *Electric Power Systems Research*, 81, pp. 1334-1340.
- Bonabeau, E. (2002) 'Agent-based modeling: Methods and techniques for simulating human systems', *Proceedings of the National Academy of Sciences*, 99(suppl 3), pp. 7280-7287.
- Börner, K., Sanyal, S. and Vespignani, A. (2007) 'Network science', *Annual Review of Information Science and Technology*, 41(1), pp. 537-607.

- Bouffard, F. and Kirschen, D.S. (2008) 'Centralised and distributed electricity systems', *Energy Policy*, 36(12), pp. 4504-4508.
- Bronk, C. (2015) 'Two securities: How contemporary cyber geopolitics impacts critical infrastructure protection', *International Journal of Critical Infrastructure Protection*, 8(0), pp. 24-26.
- Bruneau, M., Chang, S.E., Eguchi, R.T., Lee, G.C., O'Rourke, T.D., Reinhorn, A.M., Shinozuka, M., Tierney, K., Wallace, W.A. and von Winterfeldt, D. (2003) 'A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities', *Earthquake Spectra*, 19, pp. 733-752.
- Bruneau, M. and Reinhorn, A.M. (2007) 'Exploring the Concept of Seismic Resilience for Acute Care Facilities', *Earthquake Spectra*, 23(1), pp. 41-62.
- Cabinet Office (2008) *The Pitt Review: Learning lessons from the 2007 floods*. London. [Online]. Available at: http://webarchive.nationalarchives.gov.uk/20100807034701/http://archive.cabinetoffice.gov.uk/pittreview/ /media/assets/www.cabinetoffice.gov.uk/flooding_review/pitt_review_full%20pdf.pdf.
- Cabinet Office (2010) *Strategic Framework and Policy Statement: on Improving the Resilience of Critical Infrastructure to Disruption from Natural Hazards*. London.
- Caldarelli, G., Pastor-Satorras, R. and Vespignani, A. (2004) 'Structure of cycles and local ordering in complex networks', *The European Physical Journal B*, 38, pp. 183-186.
- Callaway, D.S., Newman, M.E.J., Strogatz, S.H. and Watts, D.J. (2000) 'Network robustness and fragility: percolation on random graphs.', *Physical Review Letters*, 85, pp. 5468-5471.
- Carreras, B.A., Lynch, V.E., Dobson, I. and Newman, D.E. (2002) 'Critical points and transitions in an electric power transmission model for cascading failure blackouts.', *Chaos (Woodbury, N.Y.)*, 12, pp. 985-994.
- Chang, L. and Wu, Z. (2011) 'Performance and reliability of electrical power grids under cascading failures', *International Journal of Electrical Power & Energy Systems*, 33, pp. 1410-1419.

- Chang, S.E., McDaniels, T.L., Mikawoz, J. and Peterson, K. (2007) 'Infrastructure failure interdependencies in extreme events: power outage consequences in the 1998 Ice Storm', *Natural Hazards*, 41, pp. 337-358.
- Chen, W.-K. (2003) *Net Theory And Its Applications: Flows In Networks*. London, UK: Imperial College Press.
- Chiaradonna, S., Di Giandomenico, F. and Lollini, P. (2009) 'Interdependency Analysis in Electric Power Systems', *Critical Information Infrastructure Security*, 5508, pp. 60-71.
- Clauset, A., Moore, C. and Newman, M.E.J. (2008) 'Hierarchical structure and the prediction of missing links in networks', *Nature*, 453, pp. 98-101.
- Cohen, R., Erez, K., Ben-Avraham, D. and Havlin, S. (2001) 'Breakdown of the Internet under Intentional Attack', *Physical Review Letters*, 86, pp. 3862-3865.
- Costa, L.D.F., Rodrigues, F.a. and Cristino, A.S. (2008) 'Complex networks: the key to systems biology', *Genetics and Molecular Biology*, 31, pp. 591-601.
- Costa, L.D.F., Rodrigues, F.A., Travieso, G. and Villas Boas, P.R. (2007) 'Characterization of complex networks: A survey of measurements', *Advances in Physics*, 56, pp. 167-242.
- Costa, L.F. and Silva, F. (2006) 'Hierarchical Characterization of Complex Networks', *Journal of Statistical Physics*, 125(4), pp. 841-872.
- Crucitti, P., Latora, V. and Marchiori, M. (2004a) 'Model for cascading failures in complex networks', *Physical Review E*, 69.
- Crucitti, P., Latora, V. and Marchiori, M. (2004b) 'A topological analysis of the Italian power grid', *Physica A: Statistical Mechanics and its Applications*, 388, pp. 92-97.
- Crucitti, P., Latora, V. and Porta, S. (2006) 'Centrality measures in spatial networks of urban streets', *Physical Review E*, 73.
- Danziger, M.M., Bashan, A., Berezin, Y. and Havlin, S. (2014) 'Percolation and cascade dynamics of spatial networks with partial dependency', *Journal of Complex Networks*, pp. 1-15.
- Dennis, N. (2005) 'Industry consolidation and future airline network structures in Europe', *Journal of Air Transport Management*, 11, pp. 175-183.

- Doglioni, A., Primativo, F., Laucelli, D., Monno, V., Khu, S.-T. and Giustolisi, O. (2009) 'An integrated modelling approach for the assessment of land use change effects on wastewater infrastructures', *Environmental Modelling & Software*, 24(12), pp. 1522-1528.
- Dolan, A. and Aldous, J. (1993) *Networks and Algorithms (An introductory approach)*. 1 edn. Chichester, UK: John Wiley and Sons.
- Dominguez-Sal, D., Urbon-Bayers, P., Gimenez_vano, A., Gomez-Villamor, S., Martinez-Bazan, N. and Larriba-Pey, J.L. (2010) 'Survey of graph database performance on the HPC scalable graph analysis benchmark', *Web-Age Information*, pp. 37-48.
- Dorogovtsev, S.N. and Mendes, J.F.F. (2002) 'Evolution of networks', *Advances in physics*, 51(4), pp. 1079-1187.
- Doyle, J.C., Alderson, D.L., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R. and Willinger, W. (2005) 'The “robust yet fragile” nature of the Internet', *Proceedings of the National Academy of Sciences of the United States of America*, 102, pp. 14497-14502.
- Duan, Y. and Lu, F. (2014) 'Robustness of city road networks at different granularities', *Physica A: Statistical Mechanics and its Applications*, pp. 1-14.
- Dueñas-Osorio, L.A. (2005) 'Interdependent response of networked systems to natural hazards and intentional disruptions'.
- Dueñas-Osorio, L.A., Craig, J.I. and Goodno, B.J. (2004) 'Probabilistic response of interdependent infrastructure networks', *2nd annual meeting of the Asian-pacific network of centers for earthquake engineering research (ANCER). Honolulu, Hawaii. July*, pp. 28-30.
- Dueñas-Osorio, L.A., Craig, J.I. and Goodno, B.J. (2007a) 'Seismic response of critical interdependent networks', *Earthquake Engineering & Structural Dynamics*, 36, pp. 285-306.
- Dueñas-Osorio, L.A., Craig, J.I., Goodno, B.J. and Bostrom, A. (2007b) 'Interdependent Responce of Networked Systems', *Journal of Infrastructure Systems*, 13, pp. 185-194.
- Dueñas-Osorio, L.A. and Vemuru, S.M. (2009) 'Cascading failures in complex infrastructure systems', *Structural Safety*, 31, pp. 157-167.
- Dunn, S., Fu, G., Wilkinson, S. and Dawson, R. (2013) 'Network theory for infrastructure systems modelling', *Engineering Sustainability*, 166(5), pp. 281-292.

- Egan, M.J. (2007) 'Anticipating future vulnerability: Defining characteristics of increasingly critical infrastructure-like systems', *Journal of contingencies and crisis management*, 15, pp. 4-17.
- Electricity Consumers Resource Council (2004) *The economic impacts of the August 2003 blackout*. Washington DC: Council, E.C.R.
- Erdos, P. and Renyi, A. (1959) 'On random graphs I.', *Publ. Math. Debrecen*, 6, pp. 290-297.
- Eusgeld, I., Nan, C. and Dietz, S. (2011) "“System-of-systems” approach for interdependent critical infrastructures', *Reliability Engineering & System Safety*, 96(6), pp. 679-686.
- Evans, A.J. (2010) 'Complex Spatial Networks in Application', *Complexity*, 16, pp. 11-19.
- Filippini, R. and Silva, A. (2014) 'A modeling framework for the resilience analysis of networked systems-of-systems based on functional dependencies', *Reliability Engineering & System Safety*, 125(0), pp. 82-91.
- Fortunato, S. (2010) 'Community detection in graphs', *Physics Reports*, 486, pp. 75-174.
- Foster, J.G., Foster, D.V., Grassberger, P. and Paczuski, M. (2010) 'Edge direction and the structure of networks.', *Proceedings of the National Academy of Sciences of the United States of America*, 107, pp. 10815-10820.
- Freeman, L.C. (1978) 'Centrality in Social Networks Conceptual Clarification', *Social Networks*, 1, pp. 215-239.
- Fu, G., Wilkinson, S. and Dawson, R. (2015) 'A Spatial Model for Infrastructure Network Generation and Evolution', in Sanayei, A., E. Rössler, O. and Zelinka, I. (eds.) *ISCS 2014: Interdisciplinary Symposium on Complex Systems*. Springer International Publishing, pp. 365-371.
- Gagneur, J., Jackson, D.B. and Casari, G. (2003) 'Hierarchical analysis of dependency in metabolic networks', *Bioinformatics*, 19, pp. 1027-1034.
- Gao, J., Buldyrev, S.V., Havlin, S. and Stanley, H.E. (2011) 'Robustness of a Network of Networks', *Physical Review Letters*, 107, pp. 1-5.
- Gastner, M.T. and Newman, M.E.J. (2004) 'Shape and efficiency in spatial distribution networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2006.

- Gastner, M.T. and Newman, M.E.J. (2006) 'The spatial structure of networks', *The European Physical Journal B - Condensed Matter and Complex Systems*, 49, pp. 247-252.
- Genge, B., Kiss, I. and Haller, P. (2015) 'A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures', *International Journal of Critical Infrastructure Protection*.
- Ginestra, B. and Matteo, M. (2005) 'Loops of any size and Hamilton cycles in random scale-free networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2005(06).
- Girvan, M. and Newman, M.E.J. (2002) 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences*, 99, pp. 7821-7826.
- Grubestic, T.H., Matisziw, T.C. and Zook, M.A. (2009) 'Spatio-temporal fluctuations in the global airport hierarchies', *Journal of Transport Geography*, 17, pp. 264-275.
- Guimera, R. and Amaral, L.A.N. (2004) 'Modeling the world-wide airport network', *The European Physical Journal B*, 38, pp. 381-385.
- Gursesli, O. and Desrochers, A.A. (2003) *Systems, Man and Cybernetics, 2003. IEEE International Conference on*. 5-8 Oct. 2003.
- Haimes, Y.Y. (2009) 'On the definition of resilience in systems.', *Risk analysis : an official publication of the Society for Risk Analysis*, 29, pp. 498-501.
- Havlin, S., Araujo, N.A.M., Buldyrev, S.V., Dias, C.S., Parshani, R., Paul, G. and Stanley, H.E. (2010) 'Catastrophic Cascade of Failures in Interdependent Networks', *Nature Letters*, 464, pp. 1025-1028.
- Helbing, D. (2012) 'Agent-Based Modeling', in Helbing, D. (ed.) *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 25-70.
- Helbing, D., Ammoser, H. and Kühnert, C. (2006a) 'Information flows in hierarchical networks and the capability of organizations to successfully respond to failures, crises, and disasters', *Physica A: Statistical Mechanics and its Applications*, 363(1), pp. 141-150.

- Helbing, D., Armbruster, D., Mikhailov, A.S. and Lefebvre, E. (2006b) 'Information and material flows in complex networks', *Physica A: Statistical Mechanics and its Applications*, 363(1), pp. 11-16.
- Herrmann, C., Barthélemy, M. and Provero, P. (2003) 'Connectivity distribution of spatial networks', *Physical Review E*, 68(2).
- Hines, P. and Blumsack, S. (2008) *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*. 7-10 Jan. 2008.
- HM Treasury (2010) *Strategy for National Infrastructure*. UK: HM Treasury,.
- Holling, C.S. (1973) 'Resilience and stability of ecological systems', *Annual review of ecology and systematics*, 4.
- Holme, P., Huss, M. and Jeong, H. (2003) 'Subnetwork hierarchies of biochemical pathways', *Bioinformatics*, 19(4), pp. 532-538.
- Holme, P., Kim, B.J., Yoon, C.N. and Han, S.K. (2002) 'Attack vulnerability of complex networks', *Physical Review E*, 65(5).
- Holmgren, Å.J. (2006) 'Using Graph Models to Analyze the Vulnerability of Electric Power Networks', *Risk Analysis*, 26(4), pp. 955-969.
- Homeland Security Advisory Council (2011) *Community Resilience Task Force Recommendations*.
- Hoogendoorn, S.P. and Bovy, P.H.L. (2001) 'State-of-the-art of vehicular traffic flow modelling', *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 215(4), pp. 283-303.
- Hosseini, S., Barker, K. and Ramirez-Marquez, J.E. (2016) 'A review of definitions and measures of system resilience', *Reliability Engineering & System Safety*, 145, pp. 47-61.
- Huang, L., Yang, L. and Yang, K. (2006) 'Geographical effects on cascading breakdowns of scale-free networks', *Physical Review E*, 73(3).
- igraph (2016) *igraph*. Available at: <http://igraph.org/redirect.html> (Accessed: 04/08).

- ITRC (2013) *Infrastructure Transitions Research Consortium*. Available at: <http://www.itrc.org.uk/> (Accessed: 12/03/13).
- Jenelius, E. (2010) 'Redundancy importance: Links as rerouting alternatives during road network disruptions', *Procedia Engineering*, 3(0), pp. 129-137.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabasi, A.L. (2000) 'The large-scale organization of metabolic networks', *Nature*, 407(6804), pp. 651-654.
- Jisc (2015) *Janet network*. Available at: <https://www.jisc.ac.uk/janet> (Accessed: 24/2).
- Johansson, J. and Hassel, H. (2010) 'An approach for modelling interdependent infrastructures in the context of vulnerability analysis', *Reliability Engineering & System Safety*, 95(12), pp. 1335-1344.
- Jungnickel, D. (2004) *Graphs, networks and algorithms*. Third edn. Springer.
- Jungnickel, D. (2008) 'Algorithms and Computation in Mathematics', 5.
- Kalapala, V., Sanwalani, V., Clauset, A. and Moore, C. (2006) 'Scale invariance in road networks', *Physical Review E*, 73(2).
- Katifori, E., Szollosi, G.J. and Magnasco, M.O. (2010) 'Damage and Fluctuations Induce Loops in Optimal Transport Networks', *Physical Review Letters*, 104(4).
- Kavitha, T., Liebchen, C., Mehlhorn, K., Michail, D., Rizzi, R., Ueckerdt, T. and Zweig, K.A. (2009) 'Cycle bases in graphs characterization, algorithms, complexity, and applications', *Computer Science Review*, 3(4), pp. 199-243.
- Kim, H.-J. and Kim, J.M. (2005) 'Cyclic topology in complex networks', *Physical Review E*, 72(3).
- Klein, R., Rome, E., Beyel, C., Linnemann, R., Reinhardt, W. and Usov, A. (2008) 'Information Modelling and Simulation in large interdependent Critical Infrastructures in IRRIS', in *Critical Information Infrastructure Security*. Springer, pp. 36-47.
- Klemm, K. and Stadler, P.F. (2006) 'Statistics of cycles in large networks', *Physical Review E*, 73(2).

- Lancichinetti, A., Fortunato, S. and Kertész, J. (2009) 'Detecting the overlapping and hierarchical community structure in complex networks', *New Journal of Physics*, 11.
- Laprie, J.C. and Kanoun, K. (2007) 'Modelling interdependencies between the electricity and information infrastructures', *Computer Safety, Reliability, and*, pp. 54-67.
- Latora, V. and Marchiori, M. (2002) 'Is the Boston subway a small-world network?', *Physica A: Statistical Mechanics and its Applications*, 314, pp. 109-113.
- Leavitt, W.M. and Kiefer, J.J. (2006) 'Infrastructure Interdependency and the Creation of a Normal Disaster: The Case of Hurricane Katrina and the City of New Orleans', *Public Works Management & Policy*, 10(4), pp. 306-314.
- Leu, G., Abbass, H. and Curtis, N. (2010) 'Resilience of ground transportation networks: a case study on Melbourne', *33rd Australian Transport Research Forum Conference*.
- Li, Y., Zhang, L., Huang, C. and Shen, B. (2016) 'The structural robustness of geographical networks against regional failure and their pre-optimization', *Physica A: Statistical Mechanics and its Applications*, 451, pp. 420-428.
- Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E. and Åberg, Y. (2001) 'The web of human sexual contacts', *Nature*, 411(6840), pp. 907-908.
- Little, R.G. (2002) 'Controlling cascading failure: understanding the vulnerabilities of interconnected infrastructures', *Journal of Urban Technology*, pp. 37-41.
- Little, R.G. (2003) *International Conference on System Sciences*.
- Lordan, O., Sallan, J.M., Simo, P. and Gonzalez-Prieto, D. (2014) 'Robustness of the air transport network', *Transportation Research Part E: Logistics and Transportation Review*, 68(0), pp. 155-163.
- Louf, R., Jensen, P. and Barthelemy, M. (2013) 'Emergence of hierarchy in cost driven growth of spatial networks', *Proceedings of the National Academy of Sciences of the United States of America*, 110(22).
- Luca, D.A., Alain, B., Marc, B. and Alessandro, V. (2006) 'Vulnerability of weighted networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2006(04).

- Macal, C.M. and North, M.J. (2010) 'Tutorial on agent-based modelling and simulation', *Journal of Simulation*, 4(3), pp. 151-162.
- Masuda, N., Miwa, H. and Konno, N. (2005) 'Geographical threshold graphs with small-world and scale-free properties', *Physical Review E*, 71.
- McDaniels, T.L., Chang, S.E., Cole, D., Mikawoz, J. and Longstaff, H. (2008) 'Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaptation', *Global Environmental Change*, 18, pp. 310-318.
- Merabti, M., Kennedy, M. and Hurst, W. (2011) *Communications and Information Technology (ICCIT), 2011 International Conference on*. 29-31 March 2011.
- Mishkovski, I., Biev, M. and Kocarev, L. (2011) 'Vulnerability of complex networks', *Communications in Nonlinear Science and Numerical Simulation*, 16, pp. 341-349.
- Motter, A.E. (2004) 'Cascade Control and Defense in Complex Networks', *Physical Review Letters*, 93(9).
- Motter, A.E. and Lai, Y.-C. (2002) 'Cascade-based attacks on complex networks', *Physical Review E*, 66(6).
- Murata, T. (1989) 'Petri nets: Properties, analysis and applications', *Proceedings of the IEEE*, 77(4), pp. 541-580.
- Murray, A.T. (2013) 'An overview of network vulnerability modeling approaches', *GeoJournal*, 78(2), pp. 209-221.
- Murray, A.T., Matisziw, T.C. and Grubestic, T.H. (2008) 'A Methodological Overview of Network Vulnerability Analysis', *Growth and Change*, 39(4), pp. 573-592.
- Nagel, K. and Rickert, M. (2001) 'Parallel implementation of the TRANSIMS micro-simulation', *Parallel Computing*, 27(12), pp. 1611-1639.
- Nagel, K., Stretz, P., Pieck, M., Donnelly, R. and Barrett, C.L. (1997) 'TRANSIMS traffic flow characteristics', *arXiv preprint adap-org/9710003*.
- NetworkX (2012) *NetworkX - balanced_tree*. Available at: http://networkx.lanl.gov/reference/generated/networkx.generators.classic.balanced_tree.html (Accessed: 16/03).

- NetworkX (2014) *NetworkX: Overview*. Available at: <https://networkx.github.io/>. (Accessed: 24/10).
- NetworkX (2015) *network_simplex*. Available at: https://networkx.github.io/documentation/latest/reference/generated/networkx.algorithms.flow.network_simplex.html (Accessed: 23/07).
- Newman, M.E.J. (2000) 'Models of the small world', *Journal of Statistical Physics*, 101, pp. 819-841.
- Newman, M.E.J. (2002) 'Assortative Mixing in Networks', *Physical Review Letters*, 89, pp. 1-4.
- Newman, M.E.J. (2003a) 'Mixing patterns in networks', *Physical Review E*, 67, pp. 1-13.
- Newman, M.E.J. (2003b) 'The Structure and function of complex networks', *Physics*, pp. 1-58.
- Newman, M.E.J. (2004) 'Analysis of weighted networks', *Physical Review E*, 70.
- Newman, M.E.J. (2005) 'A measure of betweenness centrality based on random walks', *Social Networks*, 27(1), pp. 39-54.
- Newman, M.E.J., Watts, D.J. and Strogatz, S.H. (2002) 'Random graph models of social networks.', *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 1, pp. 2566-2572.
- Ng, K.M., Reaz, M.B.I. and Ali, M.A.M. (2013) 'A Review on the Applications of Petri Nets in Modeling, Analysis, and Control of Urban Traffic', *IEEE Transactions on Intelligent Transportation Systems*, 14(2), pp. 858-870.
- NWB Team (2006) *Network Workbench Tool*. Available at: <http://nwb.cns.iu.edu/> (Accessed: 28/02/2012).
- O'Rourke, T.D. (2007) 'Critical Infrastructure, Interdependencies, and Resilience', *The Bridge*, 37(1), pp. 22-29.
- Oliva, G., Panzieri, S. and Setola, R. (2010) 'Agent-based input–output interdependency model', *International Journal of Critical Infrastructure Protection*, 3(2), pp. 76-82.

- Ouyang, M. (2014) 'Review on modeling and simulation of interdependent critical infrastructure systems', *Reliability Engineering & System Safety*, 121(0), pp. 43-60.
- Ouyang, M. (2016) 'Critical location identification and vulnerability analysis of interdependent infrastructure systems under spatially localized attacks', *Reliability Engineering & System Safety*, 154, pp. 106-116.
- Ouyang, M., Dueñas-Osorio, L. and Min, X. (2012) 'A three-stage resilience analysis framework for urban infrastructure systems', *Structural Safety*, 36–37(0), pp. 23-31.
- Ouyang, M., Hong, L., Mao, Z.-J., Yu, M.-H. and Qi, F. (2009) 'A methodological approach to analyze vulnerability of interdependent infrastructures', *Simulation Modelling Practice and Theory*, 17, pp. 817-828.
- Palla, G., Derényi, I., Farkas, I. and Vicsek, T. (2005) 'Uncovering the overlapping community structure of complex networks in nature and society.', *Nature*, 435, pp. 814-818.
- Pastor-Satorras, R., Alexei, V. and Vespignani, A. 440 (2004) 'Topology , Hierarchy , and Correlations in Internet Graphs' *Lecture Notes in Physics*. pp. 425-440.
- Pastor-Satorras, R., Vázquez, A. and Vespignani, A. (2001) 'Dynamical and Correlation Properties of the Internet', *Physical Review Letters*, 87, pp. 3-6.
- Paton, K. (1969) 'An algorithm for finding a fundamental set of cycles of a graph', *Communications of the ACM*, 12(9), pp. 514-518.
- Peixoto, T.P. (2015) *graph-tool: efficient network analysis*. Available at: <http://graph-tool.skewed.de/> (Accessed: 28/02/2012).
- Petreska, I., Tomovski, I., Gutierrez, E., Kocarev, L., Bono, F. and Poljansek, K. (2010) 'Application of modal analysis in assessing attack vulnerability of complex networks', *Communications in Nonlinear Science and Numerical Simulation*, 15(4), pp. 1008-1018.
- Pimm, S.L. (1984) 'The complexity and stability of ecosystems', *Nature*, 307, pp. 321-326.
- psycopg (2015) *psycopg*. Available at: <http://initd.org/psycopg/> (Accessed: 25/08).
- Purcell, M. and Fyfe, S. (1998) *Queen's University Ice Storm '98 study: emergency preparedness and response issues*. Ottawa: Canada, E.P.

Pye, G. and Warren, M. (2006) *Proceedings of the 7th Australian Information Warfare and Security Conference*. [Australian Information Warfare & Security Conference].

Python Software Foundation (2015) *Python*. Available at: <https://www.python.org/>
(Accessed: 24/08).

Ravasz, E. and Barabasi, A.-L. (2003) 'Hierarchical organization in complex networks', *Physical Review E*, 67.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.-L. (2002) 'Hierarchical organization of modularity in metabolic networks.', *Science (New York, N.Y.)*, 297, pp. 1551-1555.

Reed, D.A., Kapur, K.C. and Christie, R.D. (2009) 'Methodology for assessing the resilience of networked infrastructure', *Systems Journal, IEEE*, 3, pp. 174-180.

Richards, M.G., Hastings, D.E., Rhodes, D.H. and Weigel, A.L. (2007) 'Defining Survivability for Engineering Systems', *Conference on Systems Engineering Research*. pp. 1-12.

Rinaldi, S.M. (2004) 'Modeling and Simulating Critical Infrastructures and Their Interdependencies', *International Conference on System Sciences*. Hawaii.

Rinaldi, S.M., Peerenboom, J. and Kelly, T. (2001) 'Identifying, understanding and analysing Critical Infrastructure Interdependencies', *IEEE Control Systems Magazine*, 21, pp. 11-25.

River Bank Computing (2013) *PyQt4*. Available at:
<https://www.riverbankcomputing.com/software/pyqt/download>.

Rosas-Casals, M. and Sole, R. (2011) 'Analysis of major failures in Europe's power grid', *Electrical Power and Energy Systems*, 33, pp. 805-808.

Rosas-Casals, M., Valverde, S. and Sole, R.V. (2007) 'Topological Vulnerability of the European Power Grid Under Errors and Attacks', *International Journal of Bifurcation and Chaos*, 17, pp. 2465-2475.

Royal Academy of Engineering (2011) *Engineering the Future*. London, UK: Royal Academy of Engineering.

- Royal Academy of Engineering (2014) *Counting the cost: the economic and social costs of electricity shortfalls in the UK*. London, UK: Royal Academy of Engineering.
- Rozenfeld, H.D., Kirk, J.E., Boltt, E.M. and Ben-Avraham, D. (2005) 'Statistics of cycles: how loopy is your network?', *Journal of Physics A: Mathematical and General*, 38, pp. 4589-4595.
- Sales-Pardo, M., Guimerà, R., Moreira, A.A. and Amaral, L.A.N. (2007) 'Extracting the hierarchical organization of complex systems.', *Proceedings of the National Academy of Sciences of the United States of America*, 104, pp. 15224–15229.
- Sanchez-Garcia, R.J., Fennelly, M., Norris, S., Wright, N., Niblo, G., Brodzki, J. and Bialek, J.W. (2014) 'Hierarchical Spectral Clustering of Power Grids', *IEEE Transactions on Power Systems*.
- Scawthorn, C., Porter, K.A. and Risk, S.P.A. (2011) 'Aspects of the 11 March 2011 Eastern Japan Earthquake and Tsunami', *Reconnaissance Report*.
- Schulman, P.R. and Roe, E. (2007) 'Designing infrastructures: Dilemmas of design and the reliability of critical infrastructures', *Journal of Contingencies and Crisis Management*, 15, pp. 42-49.
- Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P.A., Mukherjee, G. and Manna, S.S. (2003) 'Small-world properties of the Indian Railway network', *Physical Review E*, 67.
- Shi, X., Gaoxi, X. and Tee Hiang, C. (2008) 'Tolerance of intentional attacks in complex communication networks', *Communications Magazine, IEEE*, 46, pp. 146-152.
- Shuang, Q., Zhang, M. and Yuan, Y. (2014) 'Node vulnerability of water distribution networks under cascading failures', *Reliability Engineering & System Safety*, 124(0), pp. 132-141.
- Sterbenz, J.P.G., Cetinkaya, E.K., Hameed, M.A., Jabbar, A. and Rohrer, J.P. (2011) 'Modelling and analysis of network resilience', *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on*. IEEE, pp. 1-10.
- Sterbenz, J.P.G., Hutchison, D., Çetinkaya, E.K., Jabbar, A., Rohrer, J.P., Schöller, M. and Smith, P. (2010) 'Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines', *Computer Networks*, 54, pp. 1245-1265.

- Suarez, P., Anderson, W., Mahal, V. and Lakshmanan, T.R. (2005) 'Impacts of flooding and climate change on urban transportation: A systemwide performance assessment of the Boston Metro Area', *Transportation Research Part D: Transport and Environment*, 10(3), pp. 231-244.
- Swain, P.H. and Davis, S.M. (1978) *Remote Sensing: The Quantitative Approach*. 1st edn. McGraw-Hill International Book Company.
- Tanizawa, T., Paul, G., Cohen, R., Havlin, S. and Stanley, H.E. (2005) 'Optimization of network robustness to waves of targeted and random attacks', *Physical Review E*, 71(4), pp. 1-4.
- Ten, C.W., Liu, C.C. and Manimaran, G. (2008) 'Vulnerability Assessment of Cybersecurity for SCADA Systems', *IEEE Transactions on Power Systems*, 23(4), pp. 1836-1846.
- Trusina, A., Maslov, S., Minnhagen, P. and Sneppen, K. (2004) 'Hierarchy Measures in Complex Networks', *Physical Review Letters*, 92(17).
- Ulieru, M. (2007) 'Design for resilience of networked critical infrastructures', *Digital EcoSystems and Technologies Conference, 2007. DEST'07. Inaugural IEEE-IES*. IEEE, pp. 540-545.
- Valeria, C., Emiliano, C. and Emanuele, G. (2007) 'Agent-based modeling of interdependencies in critical infrastructures through UML', *Proceedings of the 2007 spring simulation multiconference - Volume 2*. Norfolk, Virginia. Society for Computer Simulation International, pp. 119-126.
- van Dam, K.H. and Lukszo, Z. (2006) *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*. 8-11 Oct. 2006.
- Velykiene, R. and Jones, C.B. (2011) *A Fast Track Analysis of ICT Constraints on Evolving Physical Infrastructure*.
- Verma, T., Araujo, N.A.M. and Herrmann, H.J. (2014) 'Revealing the structure of the world airline network', p. 12.
- Vespignani, A. (2010) 'The fragility of interdependency', *Nature*, 464, pp. 984-985.

- Walker, B., Holling, C.S., Carpenter, S.R. and Kinzig, A. (2004) 'Resilience , Adaptability and Transformability in Social – ecological Systems', *Ecology And Society*, 9(2).
- Wang, J.-W. and Rong, L.-L. (2011) 'Robustness of the western United States power grid under edge attack strategies due to cascading failures', *Safety Science*, 49, pp. 807-812.
- Wang, J. and Rong, L. (2009) 'Cascade-based attack vulnerability on the US power grid', *Safety Science*, 47, pp. 1332-1336.
- Wang, S., Hong, L. and Chen, X. (2012) 'Vulnerability analysis of interdependent infrastructure systems: A methodological framework', *Physica A: Statistical Mechanics and its Applications*, 391(11), pp. 3323-3335.
- Watts, D.J., Dodds, P.S. and Newman, M.E.J. (2002) 'Identity and Search in Social Networks', *Science*, 296(5571), pp. 1302-1305.
- Watts, D.J. and Strogatz, S.H. (1998) 'Collective dynamics of 'small-world' networks.', *Nature*, 393, pp. 440-442.
- Wilkinson, S., Dunn, S. and Ma, S. (2012) 'The vulnerability of the European air traffic network to spatial hazards', *Natural Hazards*, 60, pp. 1027-1036.
- Wuellner, D.R., Roy, S. and D'Souza, R.M. (2010) 'Resilience and rewiring of the passenger airline networks in the United States', *Physical Review E*, 82(5).
- Xia, Y., Fan, J. and Hill, D. (2010) 'Cascading failure in Watts–Strogatz small-world networks', *Physica A: Statistical Mechanics and its Applications*, 389(6), pp. 1281-1285.
- Yazdani, A. and Jeffrey, P. (2012) 'Applying Network Theory to Quantify the Redundancy and Structural Robustness of Water Distribution Systems', *Journal of Water Resources Planning and Management*, 138(2), pp. 153-161.
- Yerra, B. and Levinson, D. (2005) 'The emergence of hierarchy in transportation networks', *The Annals of Regional Science*, 39, pp. 541-553.
- Zimmerman, R. (2001) 'Social Implications of Infrastructure Network Interactions', *Journal of Urban Technology*, 8, pp. 97-119.
- Zio, E. (2014) 'Vulnerability and Risk Analysis of Critical Infrastructures', in *Vulnerability, Uncertainty, and Risk*. American Society of Civil Engineers, pp. 23-30.

Appendix A: Suite of synthetic networks

More details on the suite of synthetic networks as presented in Chapter 3.

A.1 Summary of synthetic suite

The synthetic suite of networks employed in the analysis has been generated randomly to produce an ensemble of 6043 networks with each being generated through one of eight graph models. Statistics on the networks generated by each of the models are given in Table A.1.

Graph type	Number of exemplars	Average number of nodes	Average number of edges
ER	1000	1013	12726
GNM	1000	1004	11657
BA	1000	996	15501
WS	1000	962	7676
HR	1000	348	548
HR+	1000	386	586
HC	7	140	451
TREE	36	372	371

Table A.1: Statistics covering the set networks for each of the eight graph types within the suite of synthetic networks.

A.2 Generation of synthetic suite

The suite of synthetic networks has been generated using eight models, with five using existing algorithms available in the NetworkX python library (Erdos-Renyi, GNM, Watts-Strogatz, Barabasi-Albert and TREE), and three using developed algorithms (Hierarchical random, Hierarchical random + and Hierarchical communities).

A.2.1 *Erdos-Renyi*

Networks within the suite use the Erdos-Renyi algorithm available in the NetworkX library. This uses the function ‘`erdos_renyi_graph`’ with the two input parameters, the number of nodes of the proportion of the total possible edges for the network to be added to the generated network.

A.2.2 *GNM*

Again the GNM networks are generated using the GNM algorithm within the NetworkX library, ‘`gnm_random_graph`’. This function is used with two input parameters, the number of nodes and the number of edges.

A.2.3 *Watts-Strogatz*

Watts-Strogatz networks are created using the Watts-Strogatz algorithm available in the NetworkX library. The function, ‘`watts_strogatz_graph`’, creates a network using an

implementation of the Watts-Strogatz algorithm requiring three input parameters, the number of nodes, the number of neighbours each node is joined to and the probability of rewiring each edge.

A.2.4 Barabasi-Albert

Scale-free networks are generated using the Barabasi-Albert model which is part of the NetworkX library. Networks can be created using the ‘`barabasi_albert_graph`’ function with two parameters, the number of node and the number of edges to connect each new node with.

A.2.5 Hierarchical random

The hierarchical random graph type is a custom graph type developed for the purposes of this research and based upon the TREE model (A.2.8). The algorithm, Figure A.1, at first builds a tree network (line 7) using the first two specified input parameters. Using the third parameter the number of edges to add to the network is calculated by multiplying this with the number of nodes in the network (line 10). While adding the new edges (line 12) the start and end nodes are selected at random (lines 13 and 14), with the first constraint being that the start and end nodes must be different (line 15) thus avoiding self-loops, and the second constraint (lines 19 – 22) being that the edge must already exist in the network. Once all edges are added, the graph, *G*, is returned.

```

1  def hr(a,b,p):
2      '''
3      a = number of levels,
4      b = nodes per level,
5      p = multiplier value for the number of edges to be added
6      '''
7      G = nx.balanced_tree(a,b) #generate tree graph
8      no_of_nodes = G.number_of_nodes() #get the number of nodes
9
10     yadd = round(p * no_of_nodes)
11
12     while y < yadd:
13         w1=r.randint(0,nodes) #generate two random node values
14         w2=r.randint(0,nodes)
15         while w1==w2: #need to check not the same value
16             w2=r.randint(0,nodes) #if both same node, change node 2
17
18         # check edge not in network already
19         conflict = False
20         for edg in G.edges():
21             if (edg[0] == w1 and edg[0] == w2) or (edg[1] == w2 and edg[0] == w1):
22                 conflict = True
23
24         # if edge not in network add and add to y
25         if conflict == False:
26             G.add_edge(w1,w2)
27             y+=1
28
29     return G

```

Figure A.1: HR graph model code.

A.2.6 Hierarchical random +

The hierarchical random + model is, like the HR model, based on the Tree model (A.2.8). However, due to the constraints in the way in which new edges can connect nodes, restricted to those is adjacent levels or those in the same level, and the way in which the number of edges to add is calculated, the algorithm is much longer and more complex. As with the HR model the network is generated from a balanced tree network (line 5, Figure A.2). The number of nodes in each level, and the id of those nodes in each level, are then identified allowing for the subsequent calculations of the number of new edges to add to each level and the nodes which those edges can connect to.

```

1 def ahr(a,b,p):
2     '''a = number of levels, b = nodes per level, p = probability for the
3       number of edges to add'''
4
5     #generate graph
6     G = nx.balanced_tree(a,b)
7
8     #get the nodes in each level of the network
9     #add them to a list, identify them through there predecessor being 0,
10    #and then get the node number from a list of nodes
11    gnodes=G.nodes()
12    mlevel=[]
13    mlevel.append([0])
14    z=1#z will be the bottom of the range of nodes to add to the level list
15    #level1 will be 0 - a - needs nodes 1-2
16    temp = []
17    while z <= a:
18        temp.append(gnodes[z])
19        z=z+1
20
21    mlevel.append(temp)
22    gno = 1 + len(temp)
23    ai = a*(a+1)
24
25    #get the nodes in each level
26    while gno <> G.number_of_nodes():
27        temp = []
28        while z <= ai:
29            temp.append(gnodes[z])
30            z += 1
31        mlevel.append(temp)
32        ai = ai * a + a #counter used for inner loop
33        gno = gno + (len(temp)) #counter used for primary loop

```

Figure A.2: The initial steps of the HR+ graph model.

A.2.7 Hierarchical community

The hierarchical community (HC) networks have been generated using a developed algorithm which produced networks derived by Ravasz *et al.* (2002) and Ravasz and Barabasi (2003). The networks proposed by the authors vary in the size of community, Figure A.3, thus two different algorithms are used, but a higher level function, Figure A.4, controls which algorithm is used to generate the requested network. Each algorithm has a similar structure, with Figure A.5 exemplifying the initial structure which controls the building of the models depending on the level specified. The ‘createnodes’ function, Figure A.6, is then used to create all the nodes required, connecting them into the applicable community size, in this case five. These are then stitched together as required forming the final network.

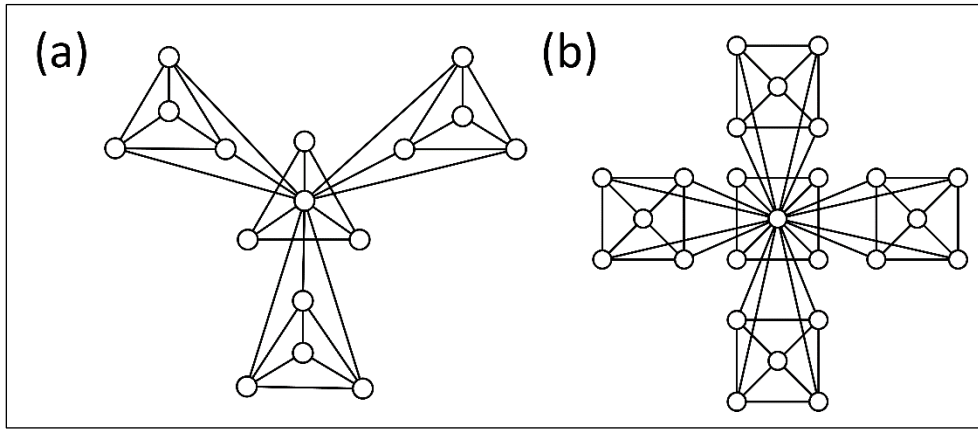


Figure A.3: The two HC graph types. Community size of four (triangles) (a) and (b) a community size of five (squares).

```

1 def hc(type,level):
2     """
3     Top level code for HC model. Allows generating of the two types of HC
4     model.
5     """
6     if type == 0:
7         G = triangle(level)
8     elif type == 1:
9         G = square(level)
10    else:
11        'Incorrect type parameter supplied. Use 0 for triangle based and 1
12        for square based model.'
13    return G

```

Figure A.4: HC graph model code. Code controls the generation of a HC model based network by calling either of the model network types.

```

1 def square(level):
2     #produce a square hierarchical community graph
3     def five_cluster(level):
4         G = nx.Graph()
5         if level==1:
6             limit=1
7             G,a,b,c,d,e=createnodes(G,limit)
8         elif level==2:
9             limit=5
10            G,a,b,c,d,e=createnodes(G,limit)
11            G=level2(G,a,b,c,d,e)
12        elif level==3:
13            limit=25
14            G,a,b,c,d,e=createnodes(G,limit)
15            G=level3(G,a,b,c,d,e)
16        elif level==4:
17            limit=125
18            G,a,b,c,d,e=createnodes(G,limit)
19            G=level4(G,a,b,c,d,e)
20        elif level==5:
21            limit=625
22            G,a,b,c,d,e=createnodes(G,limit)
23            G=level5(G,a,b,c,d,e)
24        else:
25            print 'ERROR! Please enter a suitable level value.'
26        return G

```

Figure A.5: The control function for the HC model where the community size of five is used to generate the network.

```

1 def createnodes(G,limit):
2     i = G.number_of_nodes()
3     p = 0
4     while p < limit: #creates the basic 5 node module and edges
5         i = G.number_of_nodes()
6         a = i
7         b = i + 1
8         c = i + 2
9         d = i + 3
10        e = i + 4
11        nodeaddlist = (a,b,c,d,e)
12        edgeaddlist = ([a,b],[a,c],[a,d],[a,e],[b,c],[c,d],[d,e],[
13                        e,b],[b,d],[c,e])
14        p = p+1
15        G.add_nodes_from(nodeaddlist)
16        G.add_edges_from(edgeaddlist)
17    return G,a,b,c,d,e

```

Figure A.6: Algorithm for creating the base level community of five nodes for the second of the network within the HC model.

A.2.8 *Tree*

All tree networks have been created using the Balanced-Tree algorithm, ‘balanced_tree’, in the NetworkX library. This takes two input parameters, the number of levels/layers (excluding the single source node) and the number of branches per node (the number of nodes connected to a parent node).


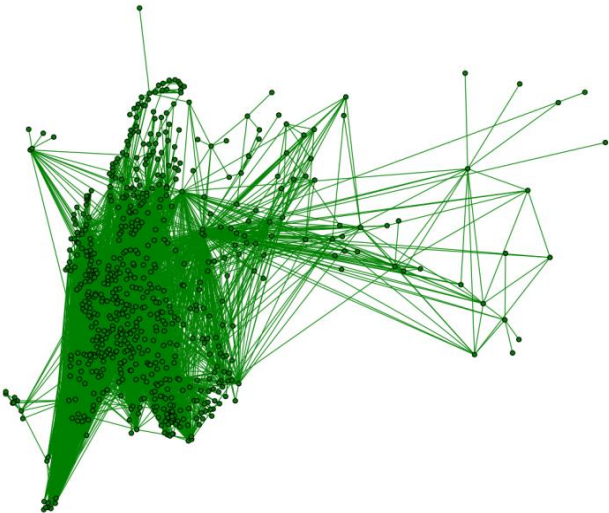
Appendix B: Suite of critical infrastructure networks


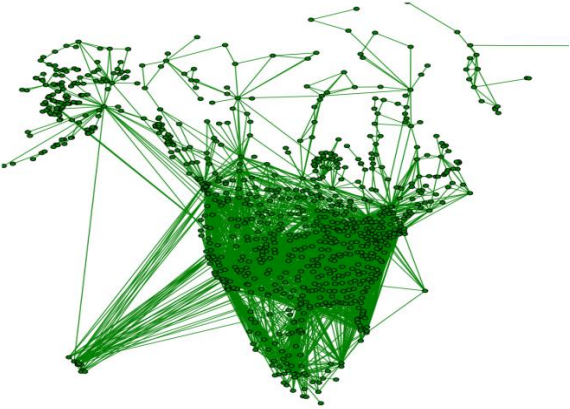
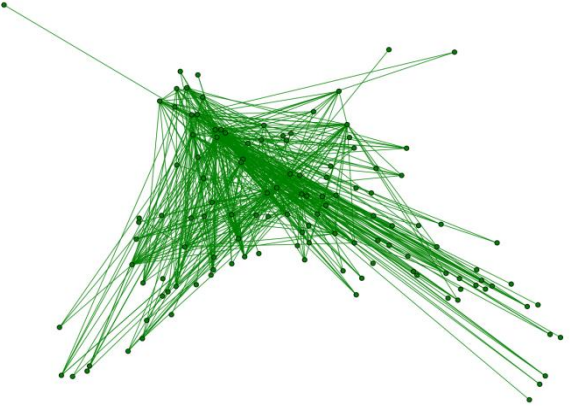
Further details of the suite of critical spatial infrastructure networks.

The second suite of networks generated contains 42 infrastructure networks across six different sectors covering a variety of scales. A mix of data sources has also been used to maximise the number and variety of networks which can be used for analysis. The following sub-sections, B.1-B.8, detail for each subset of networks the networks which fall within these groups, including the data source(s) and the size of the networks.

B.1 Air networks

In total six air networks have been created, all using data from the OpenFlights online resource, Table B.1. These cover flights within four regions as well as for two service providers.

Network	Data source	Size	
UK	OpenFlights	Nodes:48 Edges: 135	
Europe	OpenFlights	Nodes: 643 Edges: 5737	

USA	OpenFlights	Nodes: 601 Edges: 2808	
North America	OpenFlights	Nodes: 889 Edges: 3760	
EasyJet	OpenFlights	Nodes: 125 Edges: 498	


British Airways	OpenFlights	Nodes: 198 Edges: 271	
-----------------	-------------	--------------------------	--

Table B.1: Details of the air networks contained with the infrastructure database.

B.2 Communication networks

Due to the limited availability of data, only one communication network is included within the suite of infrastructure networks, Table B.2. The JANET network which provides high speed network and inter connections between major academic institution in the UK, has been manually digitised using network diagrams to generate the best possible representation

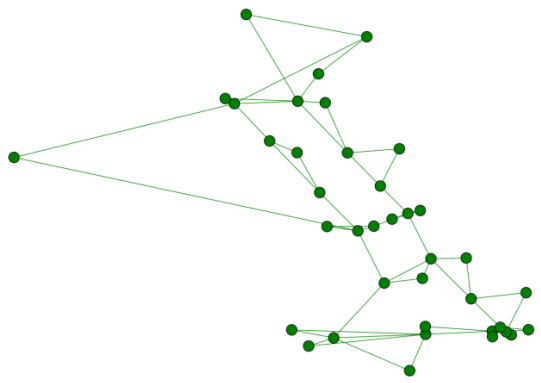
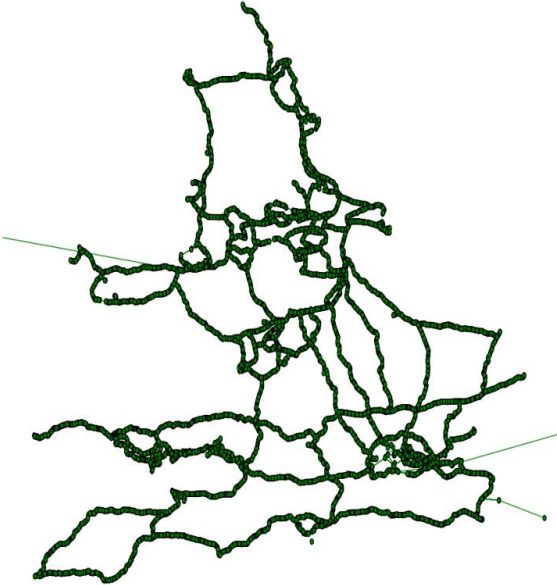

Network	Data source	Size	
JANET	JANET	Nodes: 38 Edges: 58	


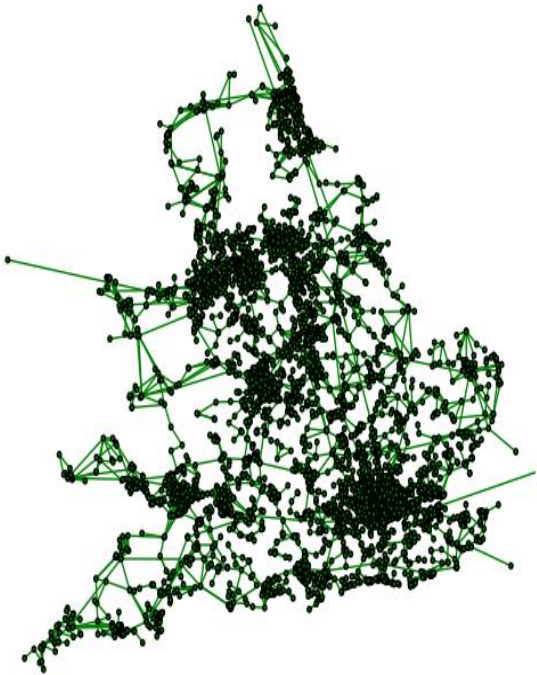
Table B.2: Details of the communication network within the infrastructure suite.

B.3 Energy networks

The suite of energy networks consists of five exemplars covering national gas and electricity transmission networks, as owned by the National Grid, Table B.3. Three of the networks cover the transmission network for electricity, with all based on the same data but variants generated at different scales of simplification. A single gas transmission network is included which is directly based on national grid data again. The fifth network is an electricity transmission and distribution network which has been generated as part of the ITRC project using the national

transmission network, Ordnance Survey Points of Interest dataset, and a sample distribution network for a single area of the UK.

Network	Data source	Size	
Electricity transmission	ITRC/National grid	Nodes: 23787 Edges: 24185	
Electricity transmission (MT)	ITRC/National grid	Nodes: 2218 Edges: 2520	

Electricity transmission (NT)	ITRC/National grid	Nodes: 7980 Edges: 8264	
Electricity transmission and distribution	ITRC (Generated network from National Grid and Ordnance Survey data)	Nodes: 170669 Edges: 173039	

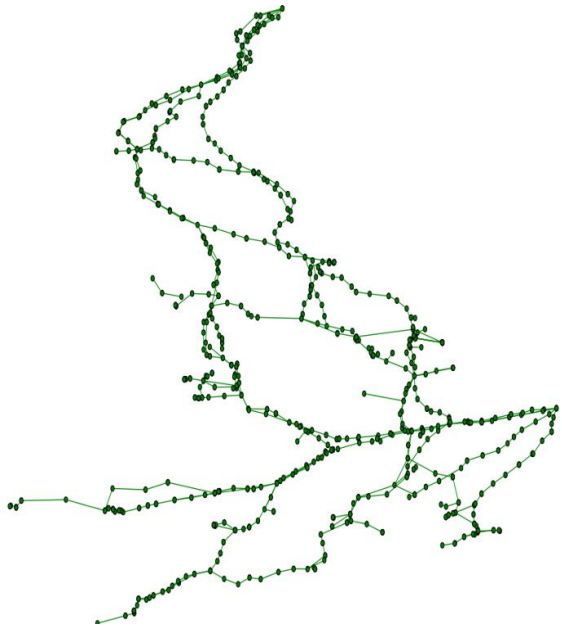

Gas transmission	ITRC/National grid	Nodes: 1486 Edges: 1739	
------------------	--------------------	----------------------------	--

Table B.3: Details of the energy networks within the infrastructure suite.

B.4 Rail – National

Two national rail networks have been generated, one for the UK using Ordnance Survey data and a second for Ireland, for which there are two versions, which have been generated using Open Street Map data, Table B.4. The difference between the two variations for Ireland is the inclusion of apparent rail lines which no routes appear to use.

Network	Data source	Size	
UK rail network	OS Meridian 2	Nodes: 7995 Edges: 8490	


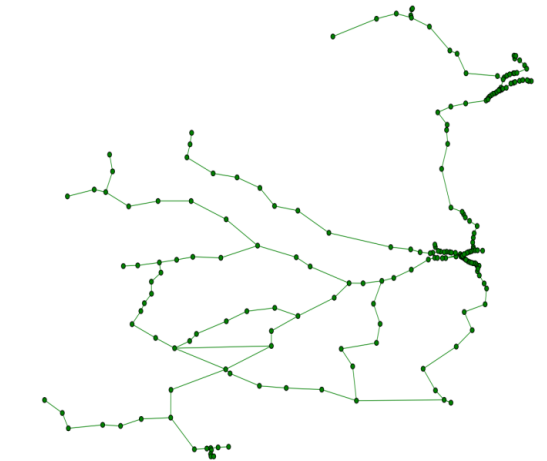
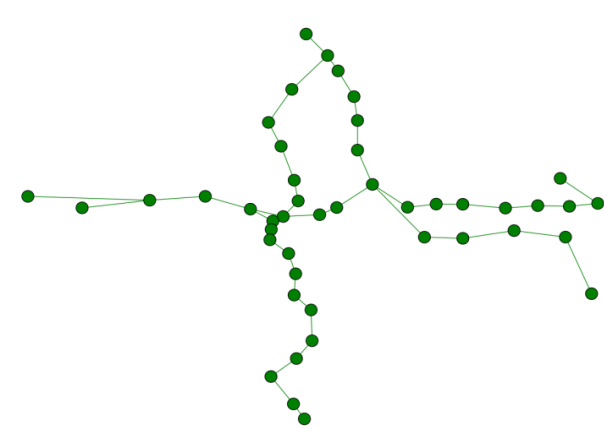
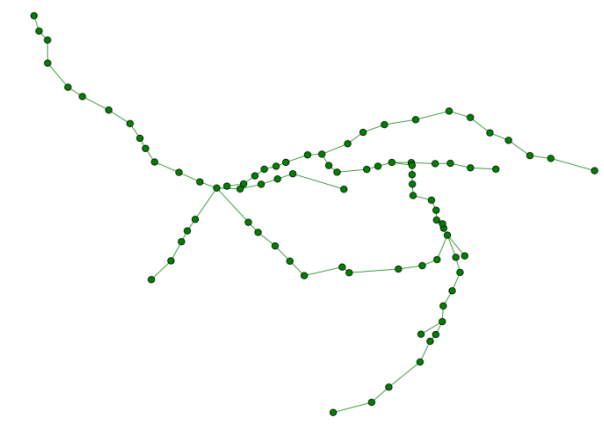
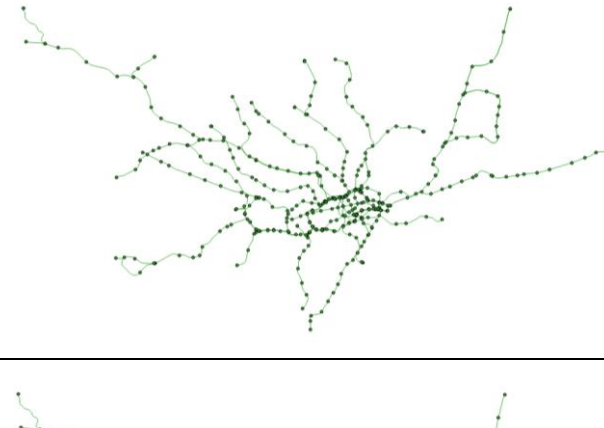
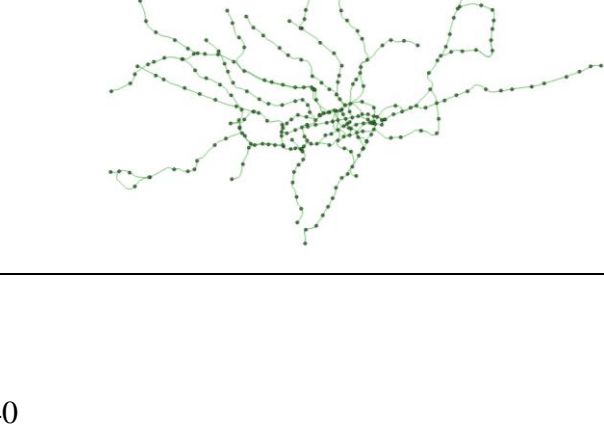
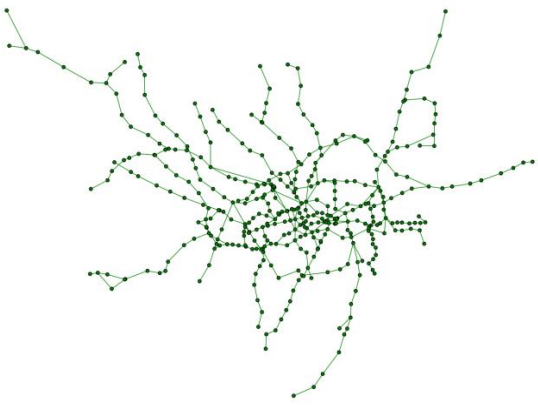
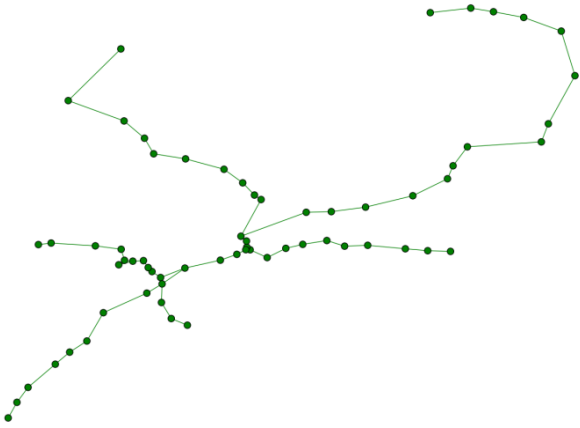
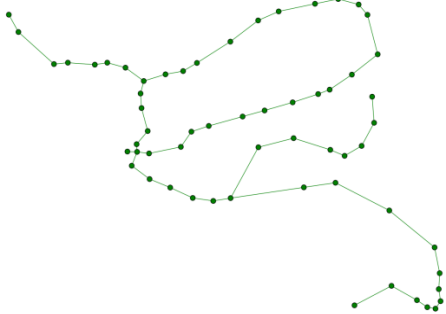
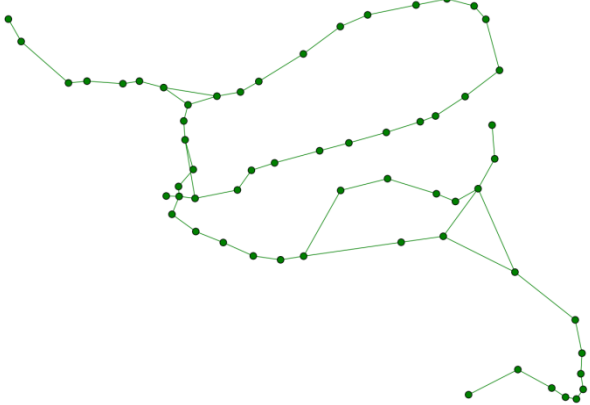
Ireland rail network	Open Street Map	Nodes: 201 Edges: 203	
Ireland rail network (all track)	Open Street Map	Nodes: 201 Edges: 208	

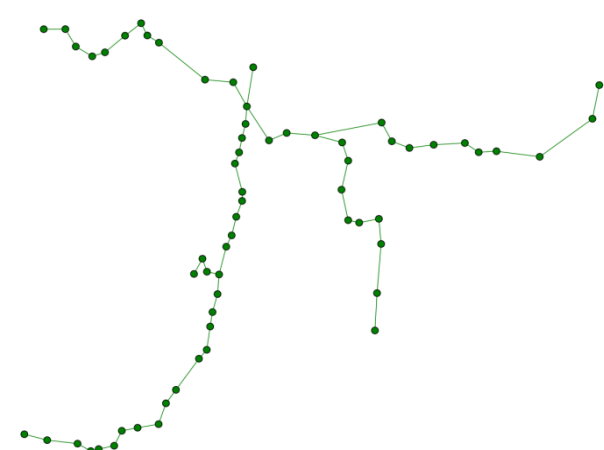
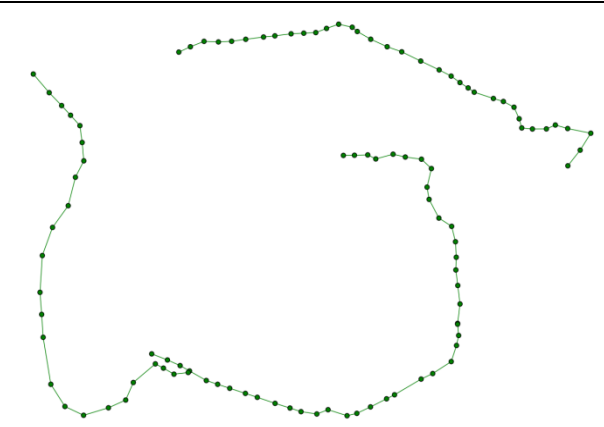
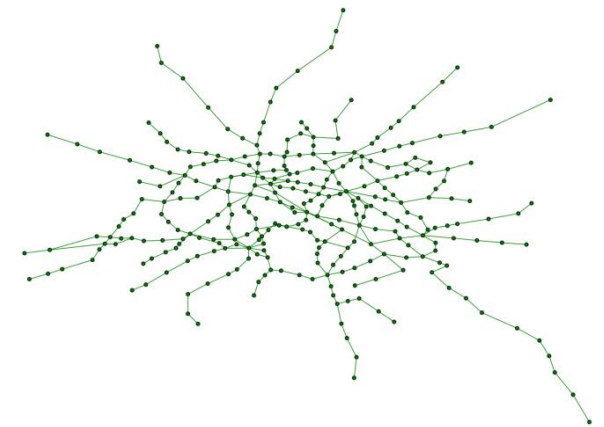
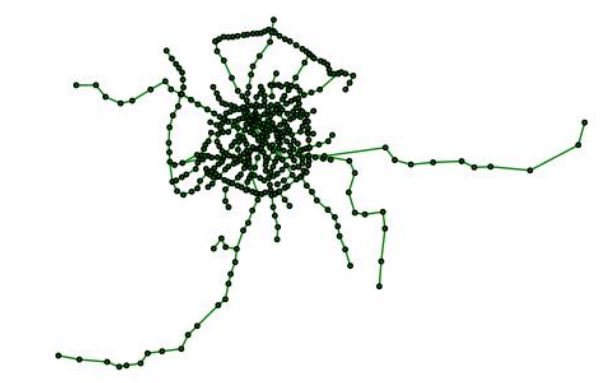
Table B.4: Details of national rail networks within the infrastructure suite.

B.5 Rail – Regional

An array of urban rail systems have been used to build a suite of rail networks of varying degrees of complexity, covering the UK, Paris (France) and Boston (USA), Table B.5. For London the main rail systems, taken from Ordnance Survey data, have been used to generate networks, including the Overground, the Docklands light railway and the tube. All three of which have also been used to generate a combined light rail network for Greater London as well. Networks have also been generated for the Manchester Metrolink and the Tyne and Wear Metro system, both of which have been generated from Ordnance Survey Meridian 2 data sets. As with London, for Paris the main rail systems have been generated as networks, using Open Street Map data. The network sets includes the RER, tram and metro networks, with a fourth network with all systems in as well. Further afield the transport network, the lightrail/subway in Boston, USA, has been used to generate an example network, with two versions, with the second containing the TAPAN extension. These have been created using Open Street Map data with verification using the published system maps.

Network	Data source	Size	
London DLR (Docklands Light Railway)	OS Meridian 2	Nodes: 45 Edges: 46	
London overground	OS Meridian 2	Nodes: 86 Edges: 85	
London tube	Transport for London (TFL)	Nodes: 436 Edges: 466	
London tube (merged stations)	TFL	Nodes: 296 Edges: 332	

London light rail	OS Meridian 2/TFL	Nodes: 399 Edges: 452	
Manchester metrolink	OS Meridian 2	Nodes: 65 Edges: 66	
Tyne and Wear metro	OS Meridian 2	Nodes: 60 Edges: 60	
Tyne and Wear metro (all shortcuts)	OS Meridian 2	Nodes: 60 Edges: 64	

RATP (Paris) RER	Open Street Map	Nodes: 64 Edges: 63	
RATP (Paris) tram	Open Street Map	Nodes: 103 Edges: 99	
RATP (Paris) metro	Open Street Map	Nodes: 301 Edges: 357	
RATP (Paris) integrated rail	Open Street Map	Nodes: 467 Edges: 519	

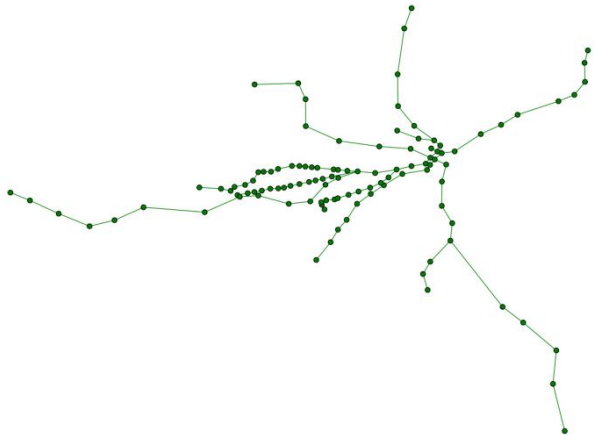

Boston light rail	Open Street Map	Nodes: 113 Edges: 114	
Boston light rail (with TAPAN)	Open Street Map	Nodes: 120 Edges: 121	

Table B.5: Details of the regional rail networks within the infrastructure suite.

B.6 Rivers

Four river networks have been generated as part of the suite of networks. These are all for the UK and have been generated using OS Meridian 2 data. The rivers included are the Eden (North West England), Dee (Scotland), Severn (England) and Tyne (North East England), Table B.6.

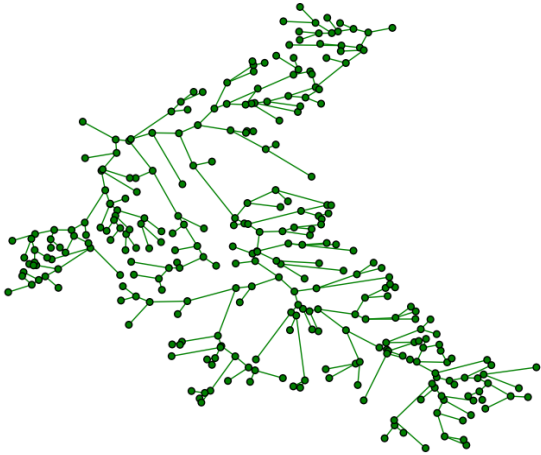


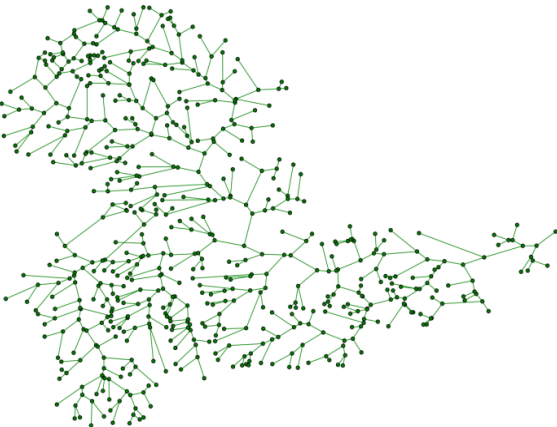
Network	Data source	Size	
River Eden	OS Meridian 2	Nodes: 302 Edges: 301	
River Dee	OS Meridian 2	Nodes: 896 Edges: 900	
River Severn	OS Meridian 2	Nodes: 1905 Edges: 1966	
River Tyne	OS Meridian 2	Nodes: 616 Edges: 615	

Table B.6: Details of the river networks within the infrastructure suite.

B.7 Road – National

Two national scale road networks have been generated, one for the UK and one for Ireland (including Northern Ireland), Table B.7. The UK network, with 24071 nodes, has been created using Ordnance Survey Meridian 2 data, whereas the Irish network has been created from Open Street Map data.


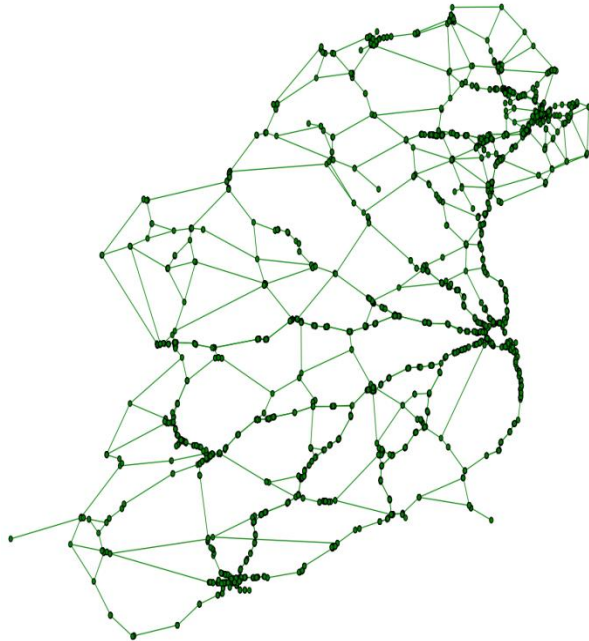


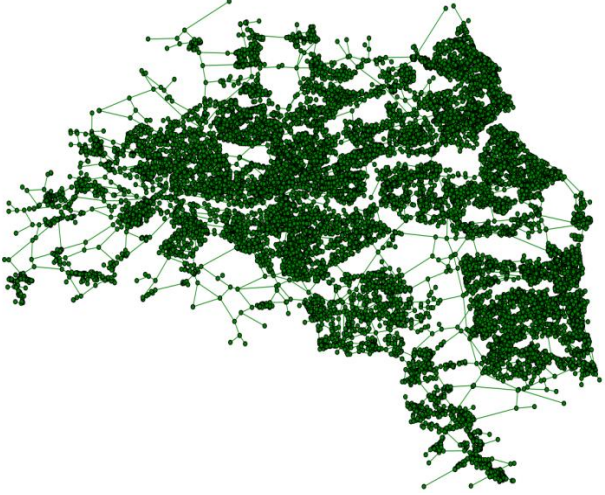

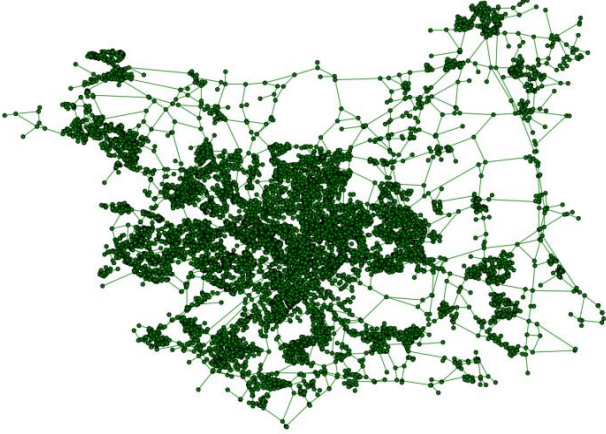
Network	Data source	Size	
UK motorways, A and B roads	OS Meridian 2	Nodes: 24071 Edges: 50292	
Ireland primary and trunk roads	Open Street Map	Nodes:4444 Edges: 6011	

Table B.7: Details of the national road networks within the infrastructure suite.

B.8 Road – Regional

Along with the national road networks, seven regional networks, Table B.8, have also been included in the suite of road networks. Three regions/cities are covered by the suite, Tyne and Wear, Leeds and Milton Keynes. For Tyne and Wear three variations are included, from the first with motorways and A roads, to the third which also has B roads and minor roads. For Leeds and Milton Keynes, each has two versions of the roads networks, with the first having motorways, A and B roads, and the second extended to include minor roads as well.

Network	Data source	Size	
Tyne and Wear Motorways and A roads	OS Meridian 2	Nodes: 212 Edges: 311	
Tyne and Wear Motorways, A and B roads	OS Meridian 2	Nodes: 398 Edges: 600	

Tyne and Wear Motorways, A, B and minor roads	OS Meridian 2	Nodes: 15249 Edges: 21817	
Leeds Motorways, A and B roads	OS Meridian 2	Nodes: 283 Edges: 411	
Leeds Motorways, A and minor roads	OS Meridian 2	Nodes: 9732 Edges: 14015	

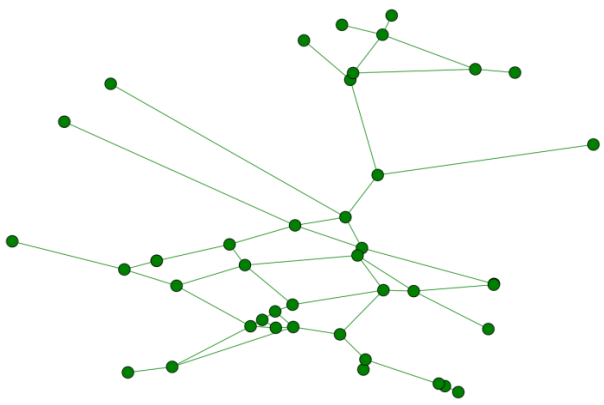

Milton Keynes Motorways, A and B roads	OS Meridian 2	Nodes: 42 Edges: 51	
Milton Keynes Motorways, A, B and minor roads	OS Meridian 2	Nodes: 2587 Edges: 3583	

Table B.8: Details of the regional road networks within the infrastructure suite.

B.9 Summary of edits to infrastructure groups

Table B.9 summarises the edits made to each of the infrastructure network groups using tools developed by ITRC and GIS tools.

Group	Editing
Air	From the open flights data the airports were plotted using the provided coordinates and then the flights plotted by connecting them via their origin and destination airport codes. For the two networks of flight operators, these were generated by using the operator code to select the applicable flights and then selecting the airports based on those required for the flights.
Communication	The JANET network has been replicated at a national level using available diagrams on their website and those of associated partners/groups. This was created using GIS software.
Energy	None – These were provided topologically valid.
Rail - national	The GB rail network was provided via ITRC, though was generated from Ordnance Survey Meridian 2 dataset. The network for Ireland (including Northern Ireland) has been built using Open Street Map data. This data was used as the basis for the network which required editing to make it topologically valid. Public rail maps and timetables were also used to check the integrity of the data and the resulting network.
Rail - regional	For those within England, these were generated using Ordnance Survey Meridian 2 data, with the use of Open Street Map to check the validity of the data and any amendments required to make the network topologically valid. For the Boston rail networks these have been generated using Open Street Map as the main source, with supporting evidence from official documentation and maps. ITRC tools were used in the generation of the valid networks from the datasets employed.
Rivers	These have been created using the Meridian 2 Ordnance Survey product. Editing was required to create networks where only junctions were found at the intersection of three edges (or more where applicable).
Roads - regional	Using Ordnance Survey Meridian 2 these networks have been generated to reproduce the road networks as close as possible. These required some editing, especially those where the full suite of road classification were not used. Networks were edited using GIS and ITRC tools to create nodes only where road junction appeared in the networks.
Roads - national	The networks for Ireland have been generated using Open Street map data only, though required a vast amount of editing due to the way the data has been recorded. Again this was to create valid topological networks where nodes were only present where junctions were found. The national network for GB was provided by ITRC.

Table B.9: Summarising the edits made to the networks for each infrastructure group

Appendix C: Details on utilising developed software

Brief details including code snippets for using the developed network schema and the robustness module.

C.1 nx_pgnet_atts

The module allows for the storage of networks with the explicit handing of a specific set of attributes. Documentation on all the functions developed is available in Appendix F, though examples of use of some of the key functions such as those for writing a network to the database, Section C.1.1, and reading from the database, Section C.1.2, are given here. Further examples are given for other functions to demonstrate how the functions can be used and the breadth of functions available from within python. These examples include how to update the attributes of a node, Section C.1.3, how to add a new function to the database for a network, Section C.1.4, and how to view for all nodes a single attribute, Section C.1.5.

C.1.1 Writing a network (python to database)

Given a NetworkX network instance within the python environment, it can be written to the database using the ‘write_to_db’ function, as exemplified in **Error! Reference source not found.** Line 1 shows the declaration of the attributes which are to be explicitly stored within the database, with those set as False to be ignored, which in this case is none of them. This list is broken into node attributes, then edge attributes. The other variables required include a valid psycopg2 database connection and the name which the network is to be given. Variables such as the presence of the attribute values and functions in the attributes of the nodes/edges in the network, the ‘contains_atts’ and the ‘contains_functions’ variables respectively need to be set to allow the schema to extract these details. Finally, the option to overwrite a network with the same name, the spatial reference id, if the network is directed and if the network is a multigraph also need to be assigned.

```
1 attributes = [{'flow':True,'capacity':True,'storage':True,  
  'resistance':True,'latency':True},{'flow':True,'capacity':True,  
  'length':True,'resistance':True,'stacking':True}]  
2  
3 nx_pgnet_atts.write(dbconn,name).write_to_db(G,attributes,  
  contains_atts=True,contains_functions=True,overwrite=True,srid=  
  27700,directed=False,multigraph=False)
```

Figure C.1: Example code to write a network to a database using the nx_pgnet_atts schema.

C.1.2 Reading a network (database to python)

A network which has been saved in a database using the nx_pgnet_atts schema can be read into a python environment using the ‘read_from_db’ function within the read class. Given a valid database connection, and the name of the network which exists in the database, the function

attempts to read the network into python. Exemplified in **Error! Reference source not found.**, the only other variable required is the dictionary containing the information on the node and edge attributes which are to be read from the database and form attributes of the nodes and edges in the NetworkX python network instance. This is of the same form as explained in Section C.1.1. Not all attributes which are explicitly handled in the database need to be form part of the NetworkX instance.

```

1 attributes = [{'flow':True,'capacity':True,'storage':True,
  'resistance':True,'latency':True},{'flow':True,'capacity':True,
  'length':True,'resistance':True,'stacking':True}]
2
3 G=nx_pgnet_atts.read(dbconn,name).read_from_db(attributes)

```

Figure C.2: Example code to read a network from a database using the nx_pgnet_atts schema.

C.1.3 *Update attributes of a node*

The attributes of a single node can be updated easily without having to read the network into python. This is done using the ‘update_node_attributes’ function within the table_sql class, as exemplified in **Error! Reference source not found.** As with the read and write class’s, a valid database connection is needed and a valid name of a network within the database. Details must then be specified of the attribute to update, the new value, the function id to be assigned, the unit id, and the node id which is to be updated. The overwrite function should be specified as True to ensure the update functions as intended.

```

1 nx_pgnet_atts.table_sql(dbconn,name).update_node_attributes(
  attribute='Flow',att_value=12,functionid=2,unitid=0,nodeid=234,
  overwrite=True)

```

Figure C.3: Example code to update the role of a single node within the database.

C.1.4 *Add a new function to the database*

A function can be added to the functions relation in the database for a specified network using the ‘add_functions’ function from the write class. Given a valid database connection and a network name, the listed functions will be added to the functions table. For each function, the id, the type and the function itself need to be specified in this order. The id is used as the primary key so this needs to be unique from all other functions already in the functions relation.

```

1 functions = [3,'variable','capacity = flow * 1.3'],[4,'variable',
'capacity = flow * 2.6']
2 nx_pgnet_atts.write(dbconn,name).add_functions(functions)

```

Figure C.4: Adding a new function to the functions relation in the network specific functions relation in the database.

C.1.5 View attribute data for all nodes

The attribute data for a specified attribute for all nodes can be viewed in python using the developed ‘get_node_data’ function, part of the table_sql class. The function returns for each node the attribute value(s), the functions and the units. This requires the database connection, the name of the network, the name of the attribute data to be returned and the names of the node table, function table and attribute table. Figure C.5 provides an example of how to use the function.

```

1 result = nx_pgnet_atts.table_sql(dbconn,name).get_node_data(key,
node_table,function_table,att_table)

```

Figure C.5: How to view for all nodes the data for a single attribute as specified.

C.2 Robustness module

The robustness modules allows for the analysis of networks with a large assortment of options available to users, with both the topological (Chapter 3, Section 3.5 (page 59)) and flow modelling (Chapter 3, Section 3.9 (page 71)) supported by this module. The options for analysis can be split into a number of categories including the analysis type (Section C.2.1), the analysis approach (Section C.2.2) and the node selection approach (Section C.2.3).

C.2.1 Analysis type

The analysis type refers to the way in which the networks are to be analysed, with the options including the analysis of a single network (a) and the analysis of a pair of networks where the function of one depends on the function of another (b).

- a. Single network analysis
 - For the analysis of a single network, treating the network as a standalone entity.
- b. Dependency analysis

- For the analysis of two networks, one of which is dependent on the other.
 - Dependencies are stored as the from and to nodes in the respective networks.

C.2.2 Analysis/failure approach

Networks can be analysed in a number of ways, from analysing the robustness of a network to single failures (a), the analysis of multiple failures with one at each epoch (b), and the analysis of cascading failures in networks (c).

- a. Single failure analysis
 - Explores the effect of the removal of a single node in a network by in-turn removing each node from the network, allowing the node with the greatest effect to be identified.
 - Works for all analysis types as specified in Section C.2.1.
- b. Sequential analysis
 - At each epoch a node is removed until no nodes are left in the network (or edges).
 - Works for all analysis types as specified in Section C.2.1.
- c. Cascading analysis
 - Cascading failures are simulated with a single trigger node/edge removed with the nodes/edges which are over capacity checked at each epoch after flows have been re-calculated.
 - Works only on the single network analysis type, Section C.2.1(a).

C.2.3 Node selection approach

For each failure approach the way in which the nodes to fail are selected is critical to how the network may behave as a consequence of the perturbations. Five methods are available, including the random selection (a), node degree based method (b), a node betweenness method (c), a manual/list based method (d) and a geographic option (e) where nodes and edges within the specified spatial area are failed.

- a. Random
 - Selects a node at random from the network.
 - Available for all failure approaches specified in Section C.2.2.

- b. Degree
 - Selects the node with the greatest degree, the number of edges connected to the node.
 - Available for all failure approaches specified in Section C.2.2.
- c. Betweenness
 - Selects the node with the greatest betweenness, the node with the greatest number of shortest paths passing through it when considering those between all node pairs.
 - Available for all failure approaches specified in Section C.2.2.
- d. List
 - Allows a list of nodes to be specified with each entry being removed at an epoch in the order they have been listed. Where a listed node has already been removed, for example it may have become an isolated node and been removed as of this (see Section for details on these options), the node will be skipped and the next node in the list removed at the epoch.
 - Available for the sequential failure approach only, Section C.2.2(b).
- e. Geographic
 - Allows nodes and edges which fall within a geographic area to be removed, as specified by a shapefile.
 - Only available for the cascading failure approach, Section C.2.2(c).

C.2.4 Failure variables

Three variables on how the network, and certain aspects of it, should be handled also exist to customise the analysis being undertaken.

- a. Remove subgraphs
- b. Remove isolated nodes
- c. Exclude isolated nodes

When running failure analysis subgraphs can form, with two options for the handling of these, specified through the ‘remove subgraphs’ option (a). The first is to ignore them and remove them from the network, classing them as failed as they are no longer connected to the largest part of the network. However, depending on the structure of the network, a subgraph could be up to half the size of the original network, and thus removing it is a significant step having consequences on the results of the analysis being performed. Alternatively, the second option

allows subgraphs to remain in the network and instead analysis results consider these as still part of the network.

Along with the formation of subgraphs, isolated nodes can also appear, nodes which have no edges connecting to them to any other node. Variable (b) allows for nodes to either be removed from the network, regarding the nodes as failed, or for them to be left in the network. By removing the nodes from the network all components which have not failed are retained thus retaining the true state of the network topologically. On the other hand, removing the nodes allows for the nodes to be regarded as failed, as may occur in some network when a node is no longer connected. This can be exemplified by a rail station, whereby once no longer connected to the track which train run along, it has to can be regarded as failed as it no longer serves a purpose, presuming the analysis does not consider restoration of failed components (not considered/possible within the developed module).

When selection nodes to fail option (c) can be used to specify if isolated nodes should be ignored when choosing a node to fail. This is obviously not an option if isolated nodes are removed when they appear as specified by option (b).

Appendix D: Synthetic graphs analysis results

Results from the analysis of the suite of synthetic graphs generated through the employed eight graph models.

D.1 Degree distributions for the synthetic graphs

The degree distribution of a graph provides insights into the topological structure of a graph, with the distribution mapping for each node degree in the graph the proportion of nodes which have each degree. For each of the eight graph models employed six example degree distribution plots are shown which represent the spectrum of distributions plotted for the full set of networks generated by each model. The six plots representing the degree distributions of the 1000 graphs generated using the ER model are shown in Figure D.1 and those selected from the suite generated by the GNM model are given in Figure D.2 (page 260). The selected plots for the graphs generated by the WS model and the BA model are presented in Figure D.3 (page 261) and Figure D.4 (page 262) respectively. Figure D.5 (page 263) and Figure D.6 (page 264) present the sample degree distribution plots for the graphs generated by the HR and HR+ models respectively. The example plots for the graphs generated by the HC model are presented in Figure D.7 (page 265) and Figure D.8 (page 266) shows the plots for the graphs generated by the TREE model.

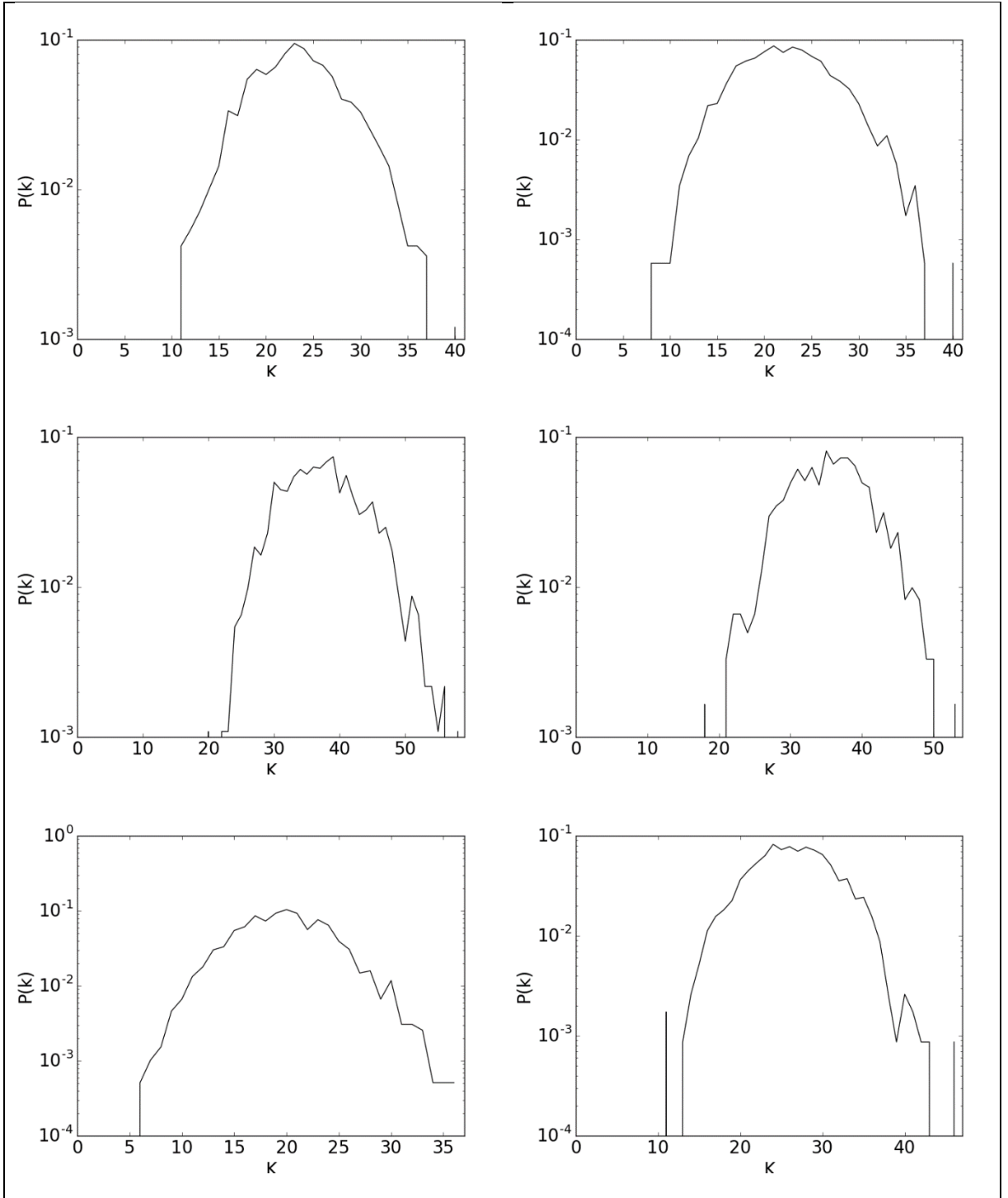


Figure D.1: Degree distribution plots for the six example graphs generated by the ER graph model, Chapter 3, Section 3.3.1.

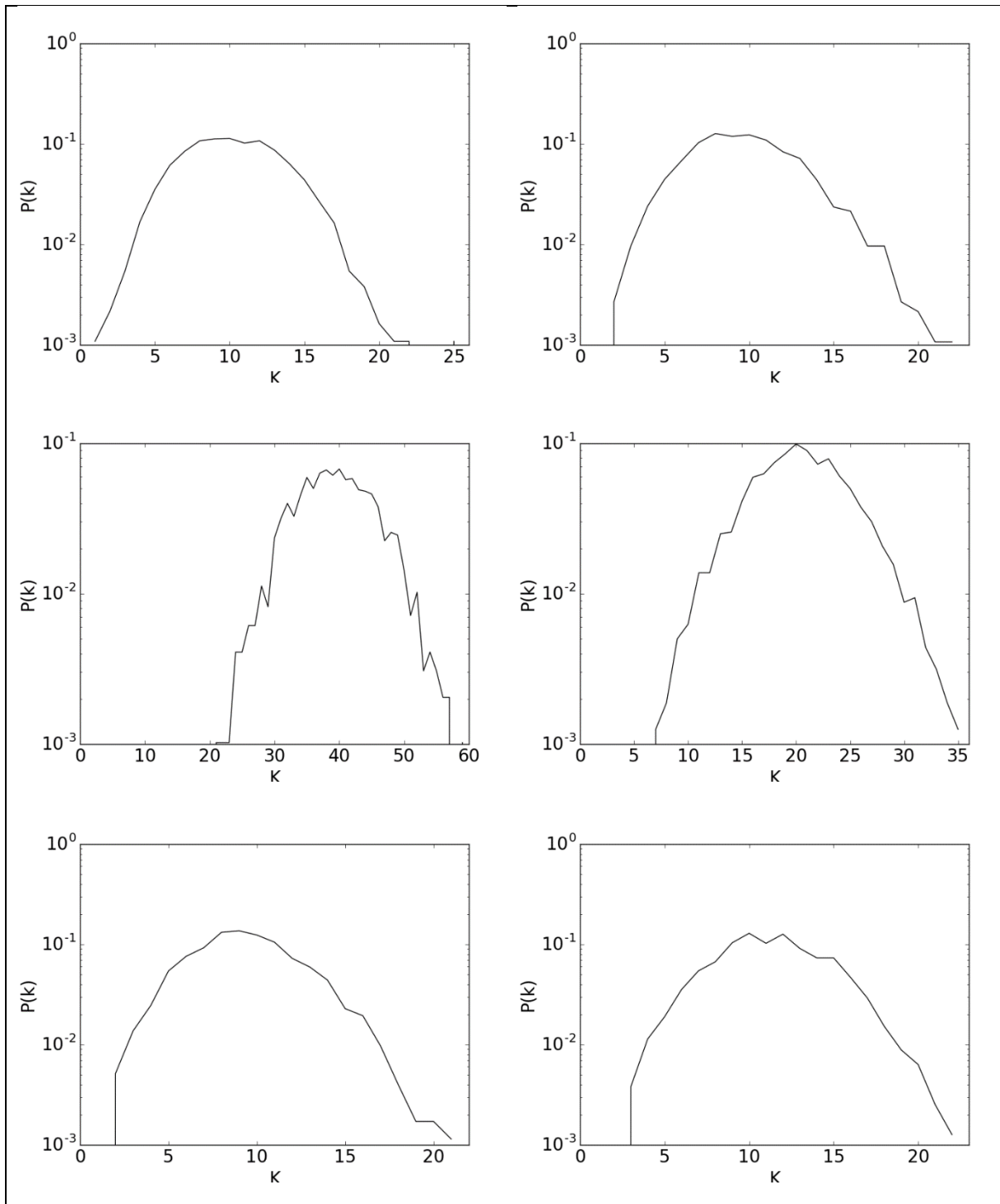


Figure D.2: Degree distribution plots for six example graphs generated by the GNM graph model, Chapter 3, Section 3.3.2.

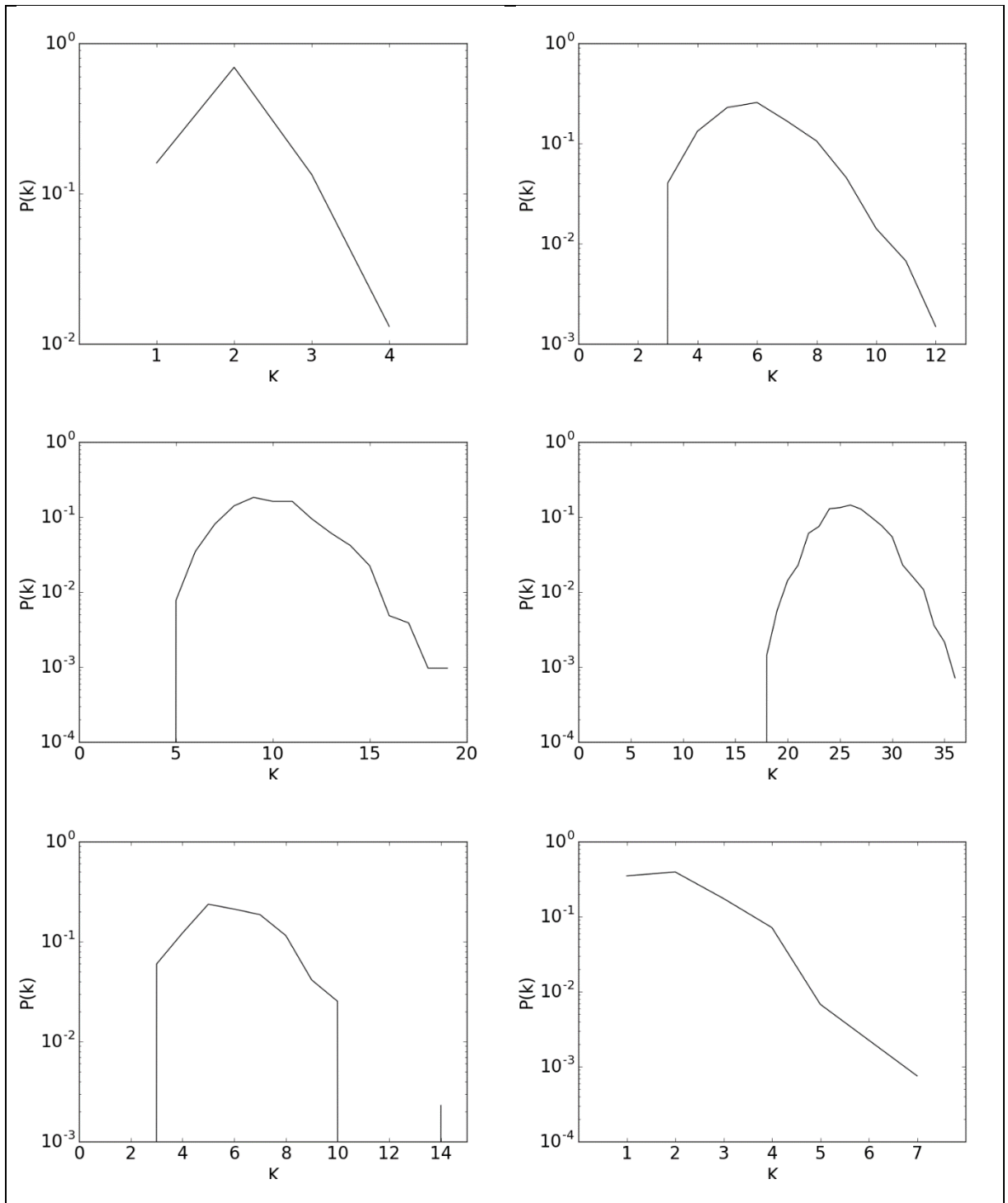


Figure D.3: Degree distribution plots for the six example graphs generated by the WS graph model, Chapter 3, Section 3.3.3.

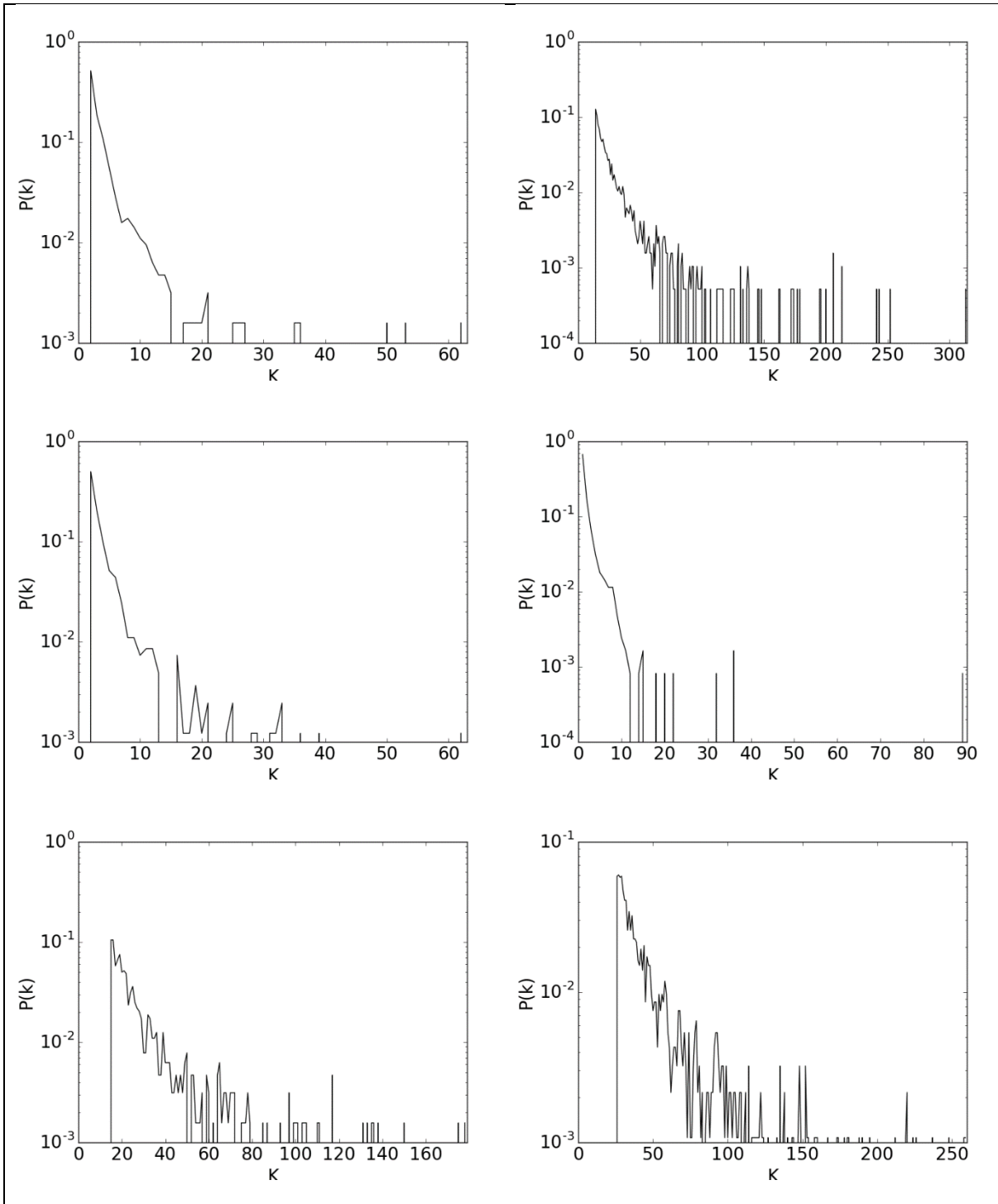


Figure D.4: Six example degree distribution plots for graphs generated by the BA graph model, Chapter 3, Section 3.3.4.

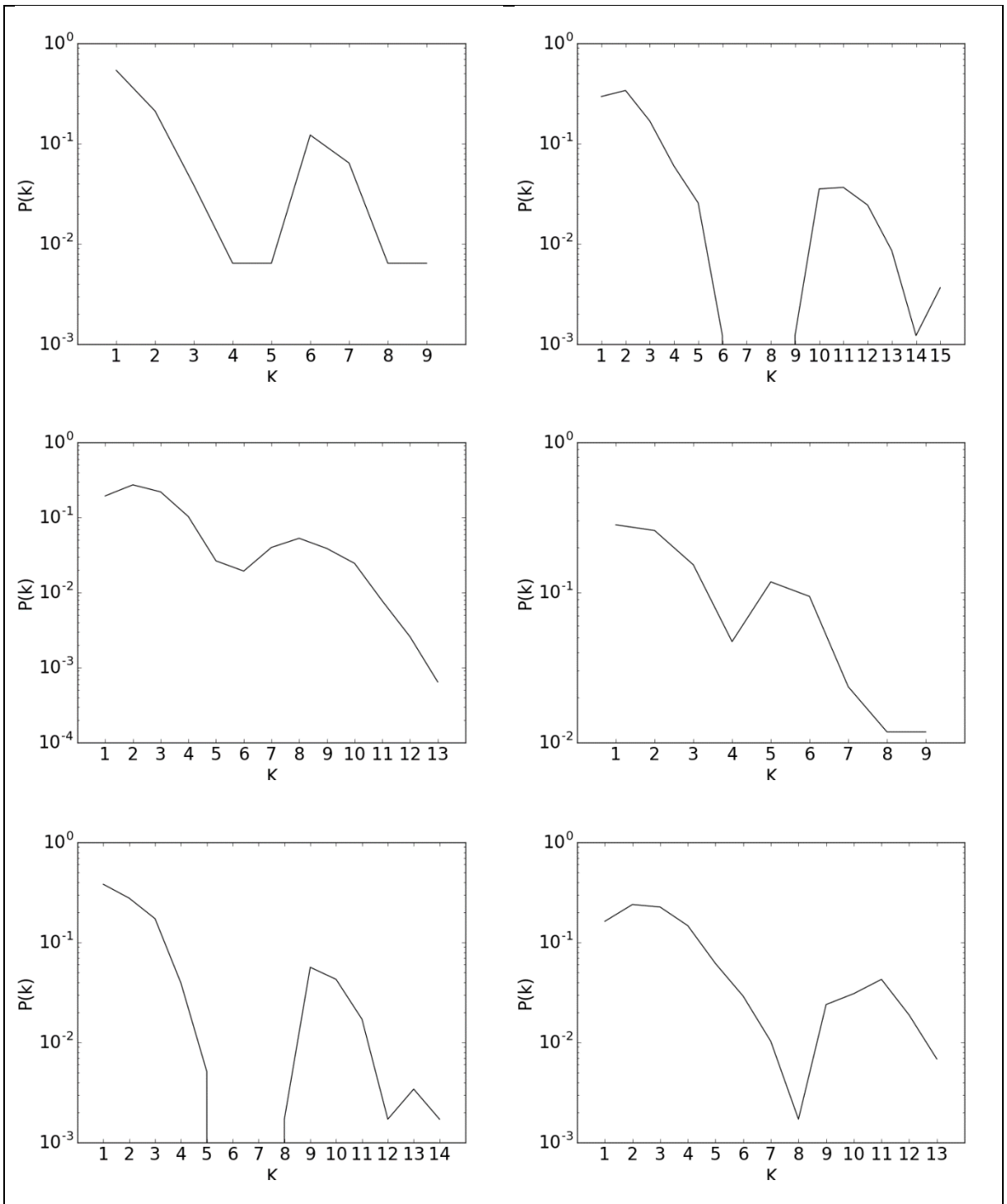


Figure D.5: Six example degree distribution plots for graphs generated by the HR graph model, Chapter 3, Section 3.3.5.

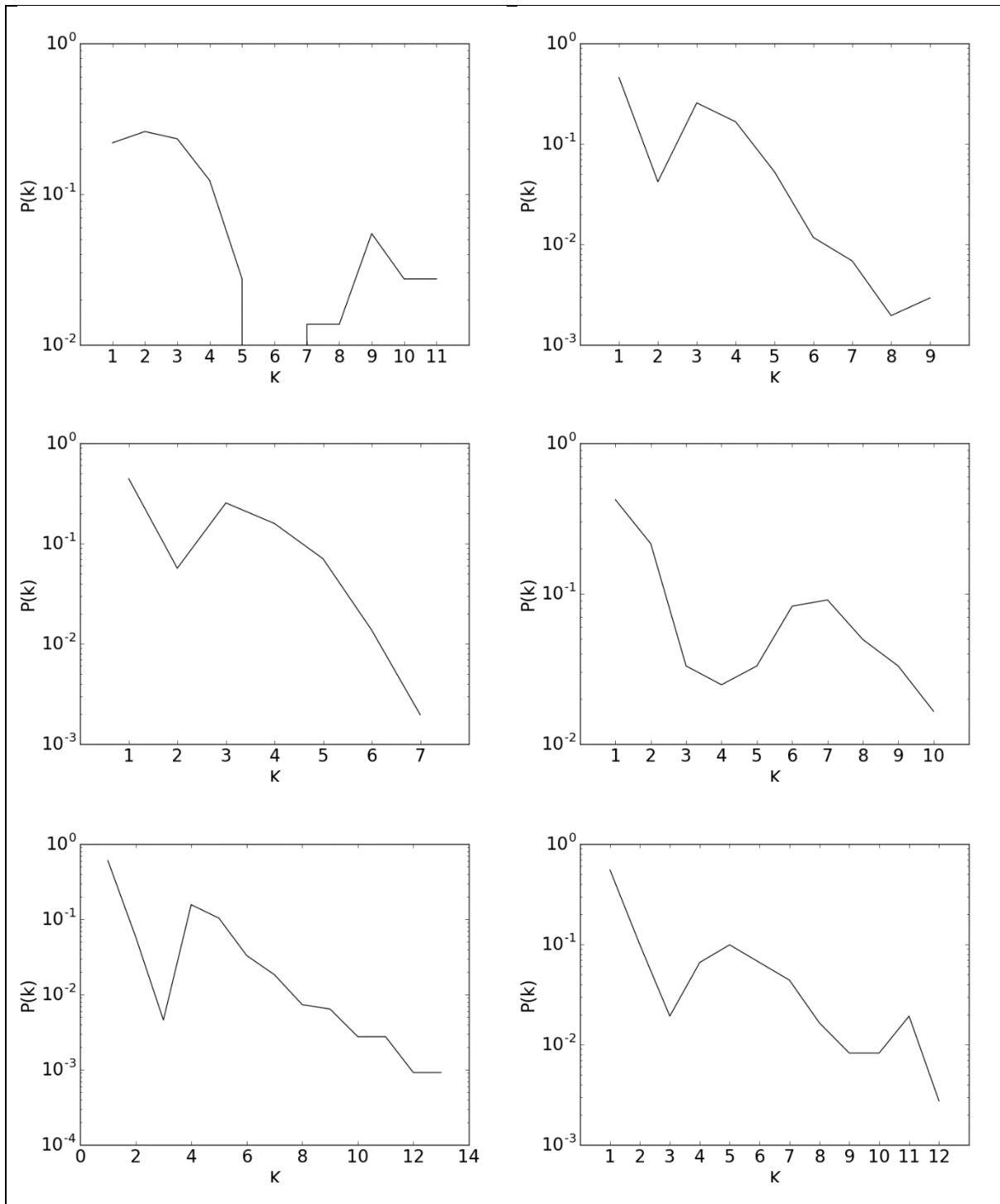


Figure D.6: Example degree distribution plots for six graphs generated using the HR+ graph model, Chapter 3, Section 3.3.6.

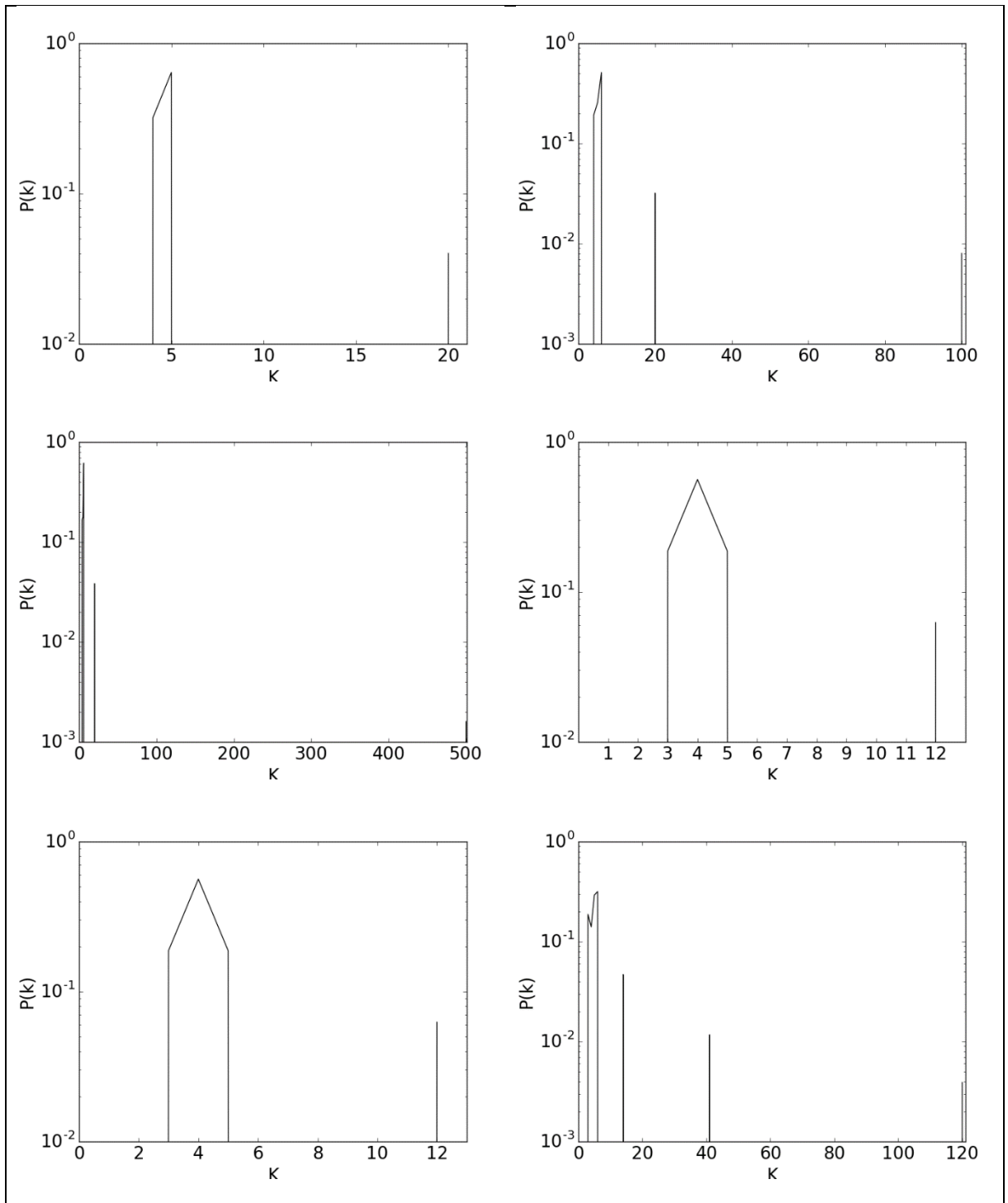


Figure D.7: Six example degree distribution plots for graphs generated by the HC graph model, Chapter 3, Section 3.3.7.

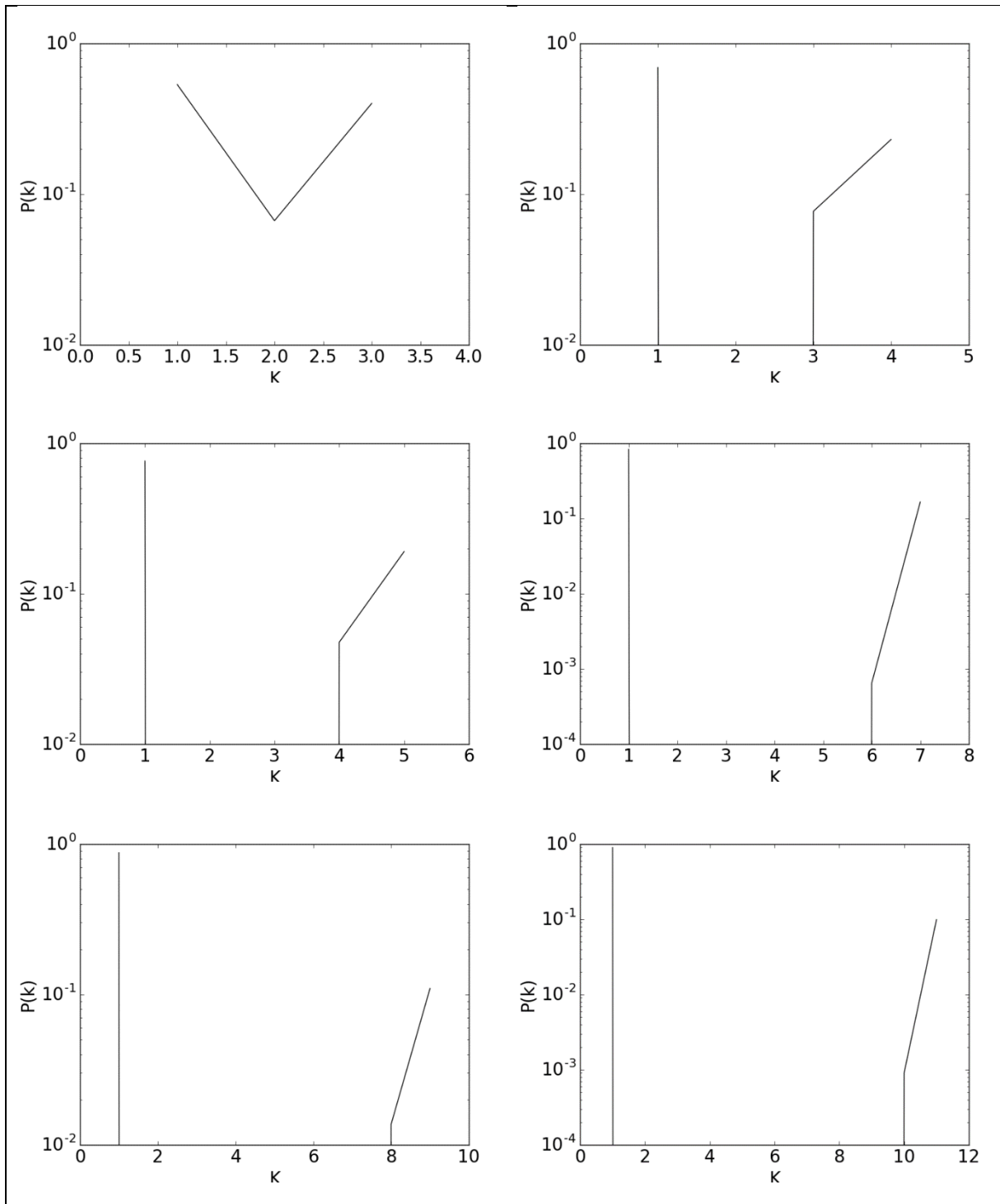


Figure D.8: Six degree distribution plots of graphs generated by the TREE graph model, Chapter 3, Section 3.3.8.

D.2 Metric values for the synthetic graphs

For the characterisation of the synthetic graph models the three selected metric values (Chapter 3, Section 3.4 (page 53)), the assortativity coefficient (AC), the maximum betweenness centrality (MBC) and the number of cycle basis per node (CB), were computed to record key properties of the networks generated by the models. These metric values can be used to characterise the networks as well as for comparing models. For each graph type the full suite of graphs were analysed (Table D.1). The following Sections, D.2.1 - D.2.8, present the metric values for each of the eight graph models, presented with the standard deviation of the metric values for all models to provide context for the results from each suite.

Graph model	Graph theme	Number of graphs in suite	Number of graphs analysed
ER (Erdos-Renyi)	Random	1000	1000
GNM	Random	1000	1000
WS (Watts-Strogatz)	Small-World	1000	1000
BA (Barabasi-Albert)	Scale-free	1000	1000
HR (Hierarchical random)	Hierarchical random	1000	1000
HR+ (Hierarchical random +)	Hierarchical random +	1000	1000
HC (Hierarchical communities)	Hierarchical communities	7	7
TREE	Balanced tree	31	31

Table D.1: The number of graphs for each model in the graph suite and the number used for the metric analysis.

The results for the ER graphs are given in Section D.2.1 (page 268), with the results for the second random graph model, the GNM model presented in Section D.2.2 (page 270). Section D.2.3 (page 272) presents the results for the small-world topological graphs as generated by the WS model, with the scale-free (BA) model results given in Section D.2.4. (page 274) The results for the HR and HR+ model graphs are presented in Sections D.2.5 (page 276) and D.2.6 (page 278) respectively. Section D.2.7 (page 280) gives the results from the graphs generated using the HC model and the results for the graphs generated by the TREE model are presented in Section D.2.8 (page 282).

D.2.1 ER

The three metrics have been calculated for the 1000 graphs generated randomly using the ER graph model. The results are shown with the standard deviations of the values for all of the graph models providing added context with the points for the ER model graphs shown as clear circles. The results are split into three plots, the first showing the AC and MBC values, Figure D.9, the second the AC and the number of CB per node, Figure D.10, and the third the MBC and the number of CB per node, Figure D.11.

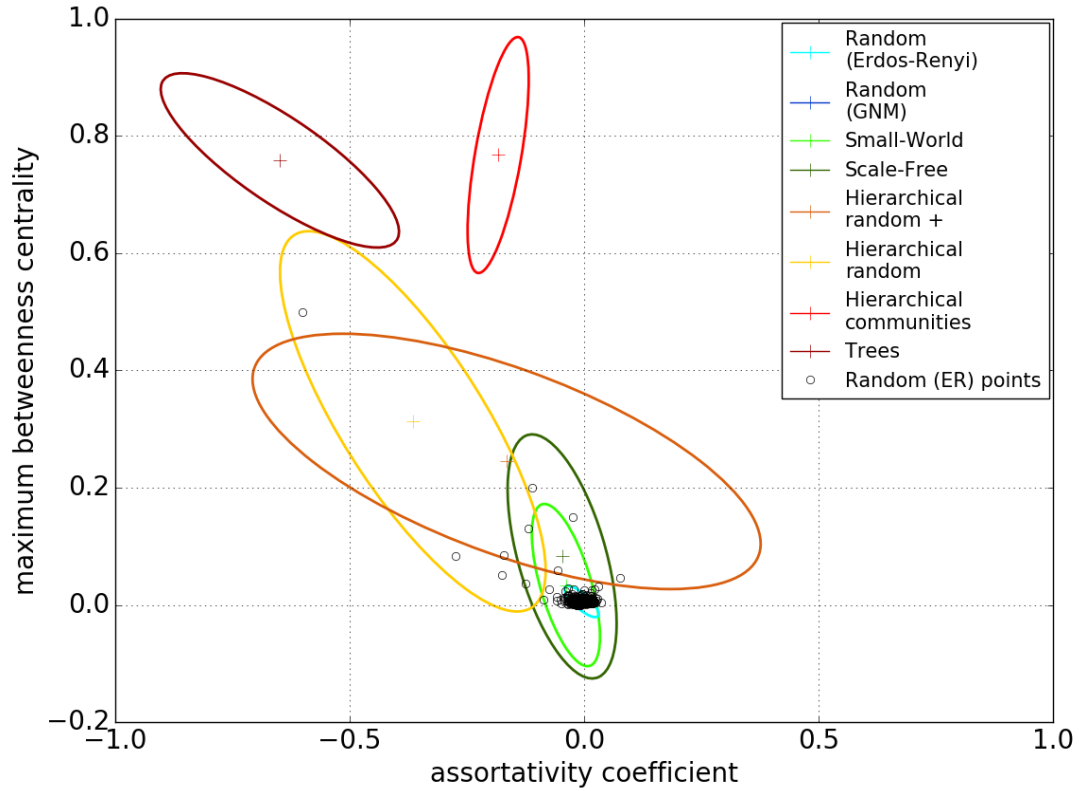


Figure D.9: The metric values for each of the ER model generated graphs for the assortativity coefficient and the maximum betweenness centrality metric.

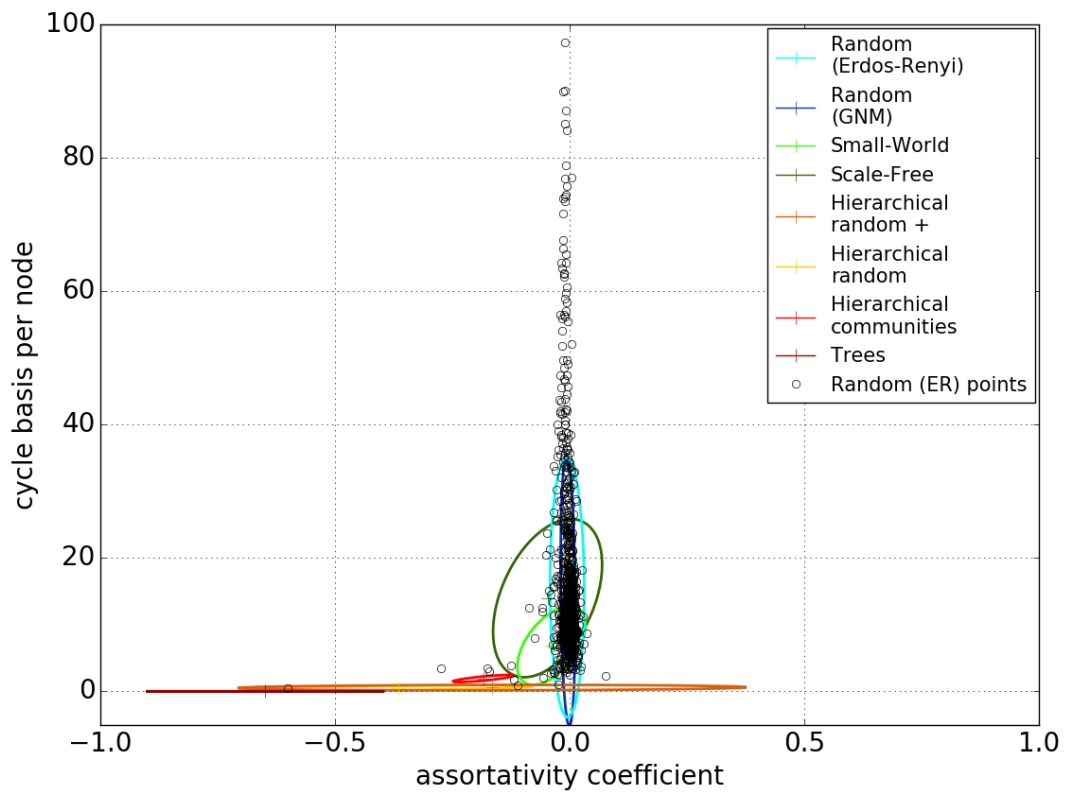


Figure D.10: Computed assortativity coefficient and cycle basis per node values for the graphs generated using the ER graph model.

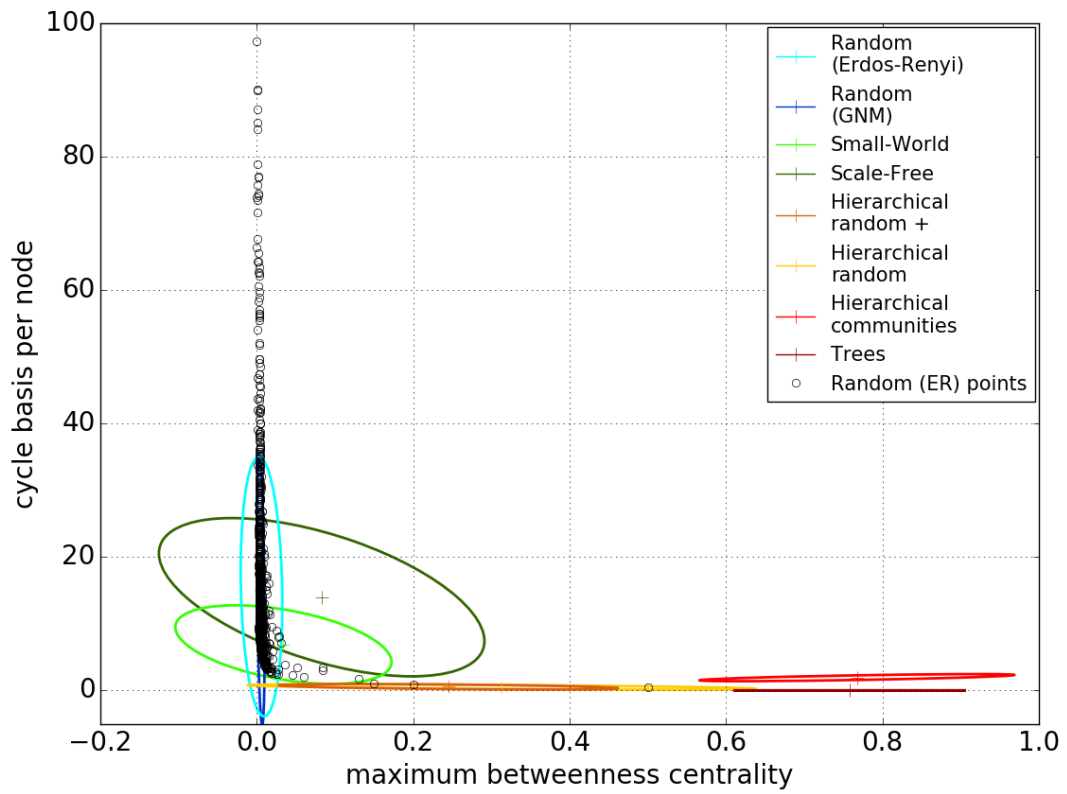


Figure D.11: Results for the graphs generated with the ER graph model for the maximum betweenness centrality and number of cycle basis.

D.2.2 GNM

The GNM model, the second model which generates graphs with a random topological structure, has been used to generate 1000 graphs. The metric values for these have been computed as with the other graph models and are shown alongside the results for the other graph models to enable these to be compared more easily. The results for each graphs are presented, with the first for the AC and MBC, Figure D.12, and the second for the AC and the number of CB per node, Figure D.13. The final set of results, Figure D.14, shows the results for the MBC and the number of CB per node.

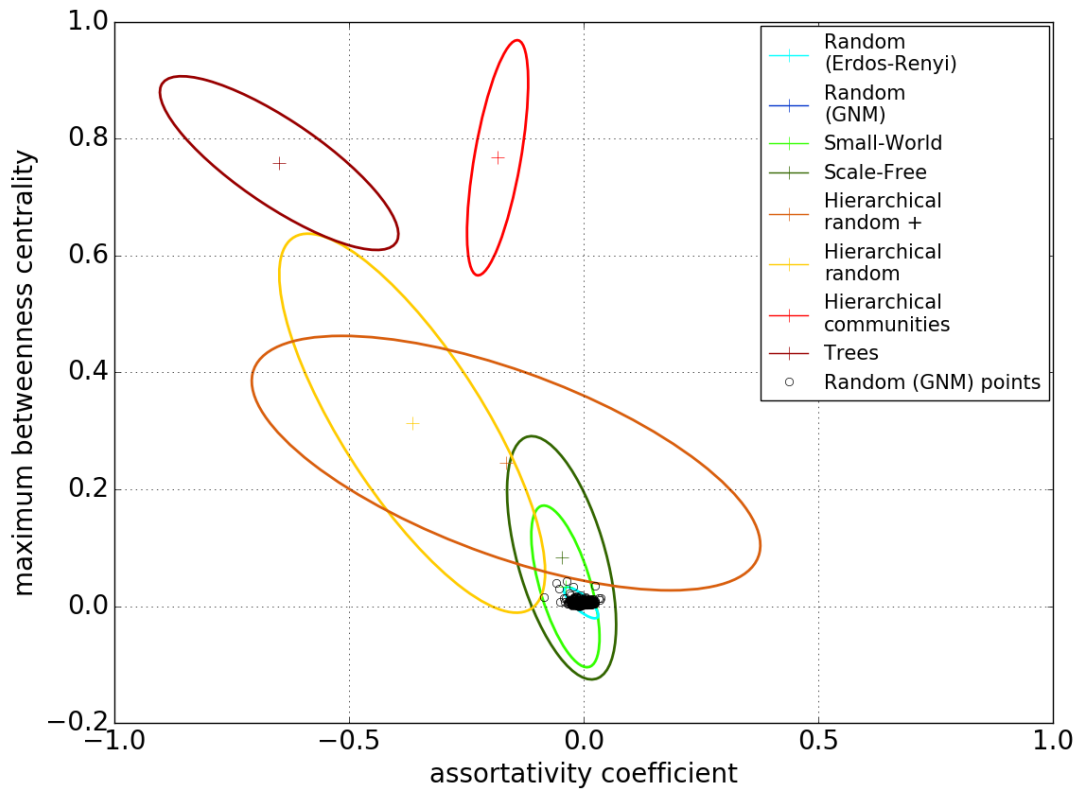


Figure D.12: The assortativity coefficient and maximum betweenness centrality points for the graphs generated by the GNM model.

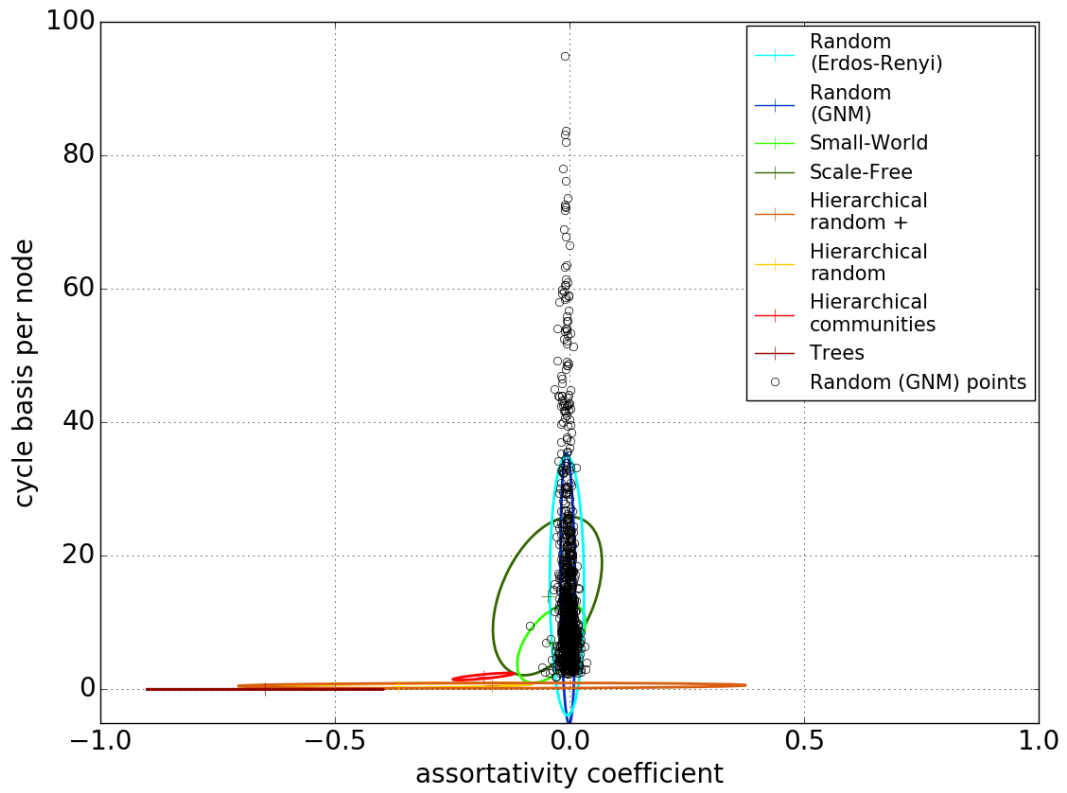


Figure D.13: Points for the assortativity coefficient and number of cycle basis for graphs generated by the GNM model.

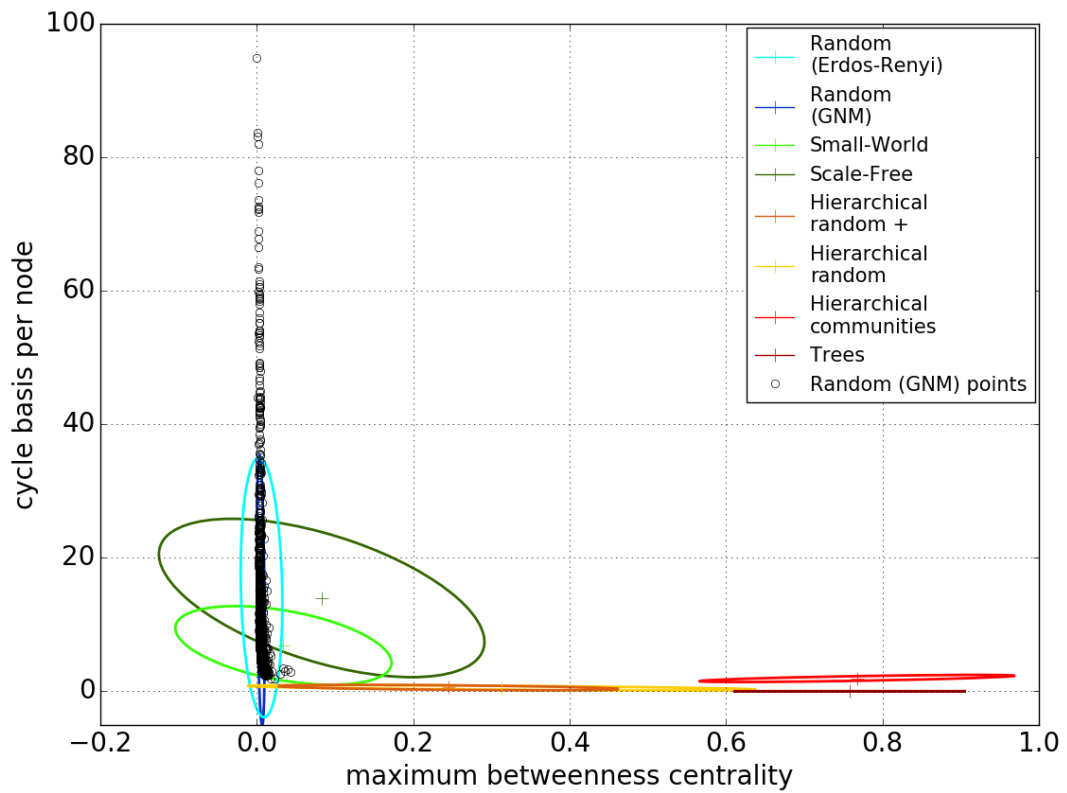


Figure D.14: Points for the maximum betweenness centrality and number of cycle basis for the 1000 graphs generated by the GNM model.

D.2.3 WS

Graphs with a small-world topological structure have been generated using the WS model, with 1000 exemplars generated and added to the suite of synthetic graphs. The characteristics of those networks with regard to the selected metrics, the AC, MBC and number of CB per node, are have been calculated allowing the graphs to be compared to those generated by the other graph models. The values for the AC and the MBC are presented in Figure D.15, with Figure D.16 showing the AC and number of CB per node. Finally Figure D.17 presented the MBC and the number of CB per node.

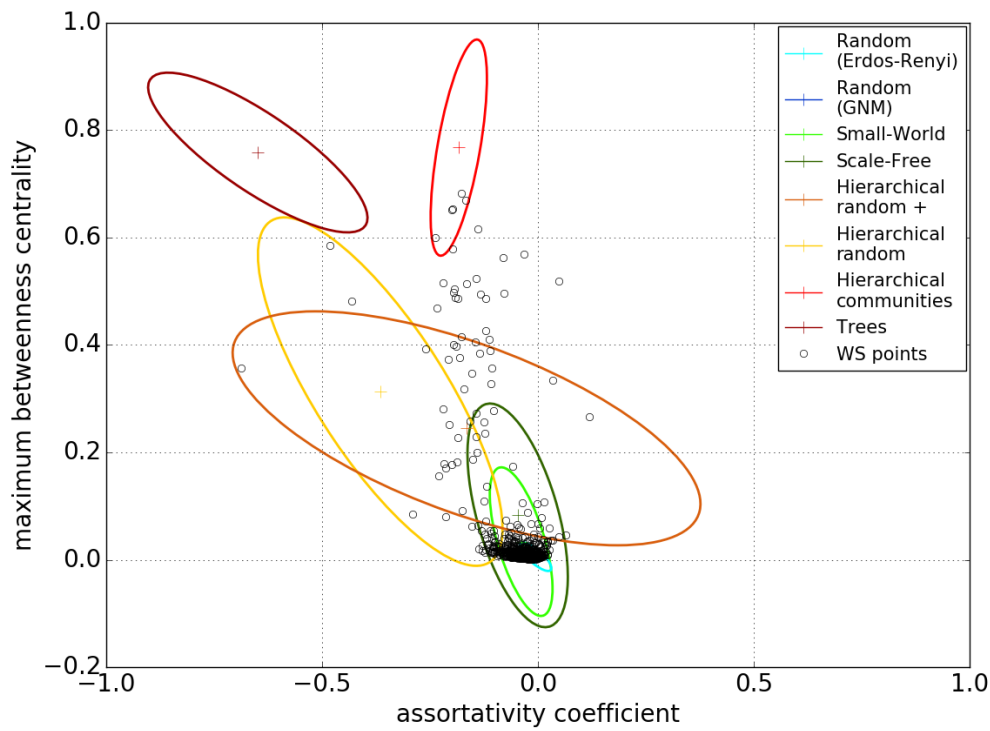


Figure D.15: For the graphs generated by the WS model the values for the assortativity coefficient and maximum betweenness centrality.

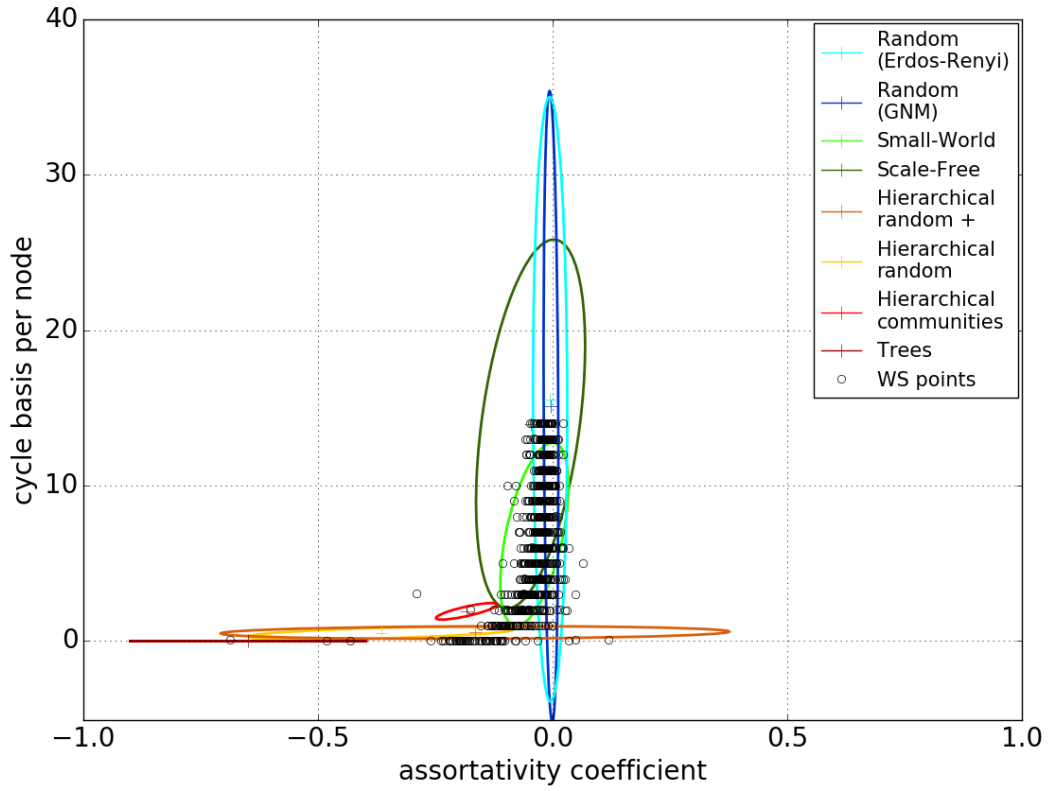


Figure D.16: Points for assortativity coefficient and number of cycle basis for the 1000 graphs generated using the WS model.

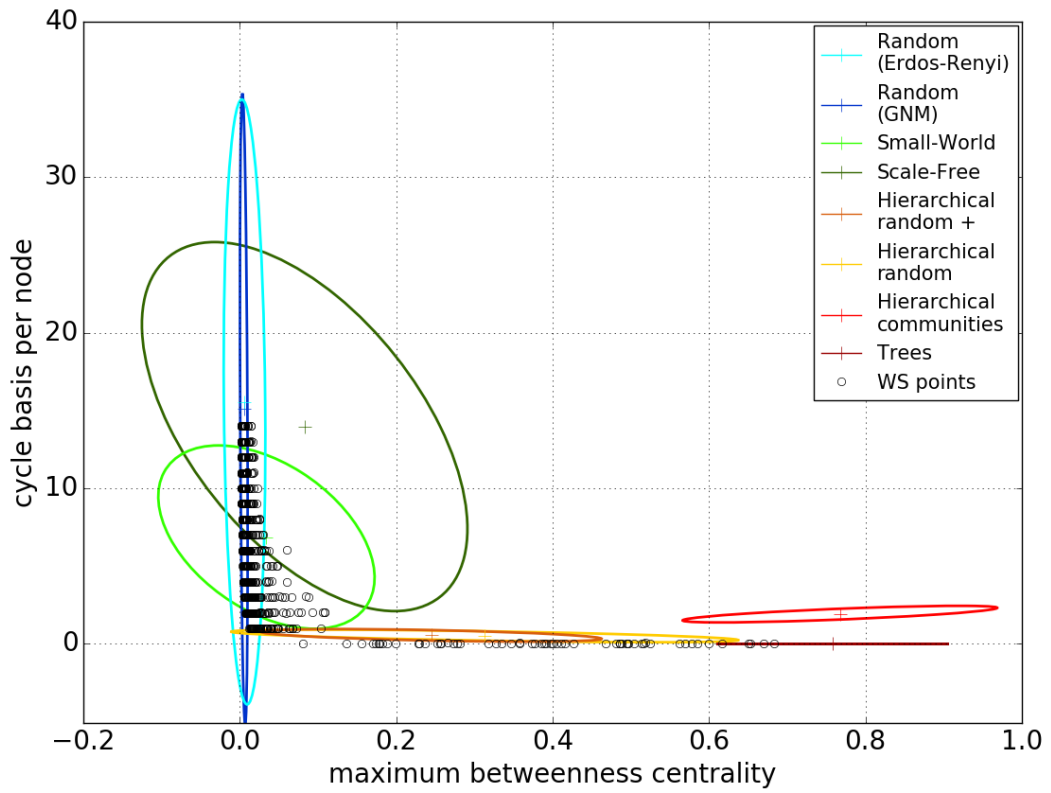


Figure D.17: Values for the maximum betweenness centrality and number of cycle basis for the graphs generated by the WS models.

D.2.4 BA

Using the BA (Barabasi-Albert) model 1000 scale-free networks were generated and added to the suite of synthetic graphs. As with the other graphs, the three selected metrics, AC, MBC and number of CB per node were calculated for each generated graph. The first set of results presented in Figure D.18 shows the AC and the MBC for the networks, with Figure D.19 showing the values for the AC and the number of CB per node. Figure D.20 presents the final set of values, the MBC and the number of CB per node.

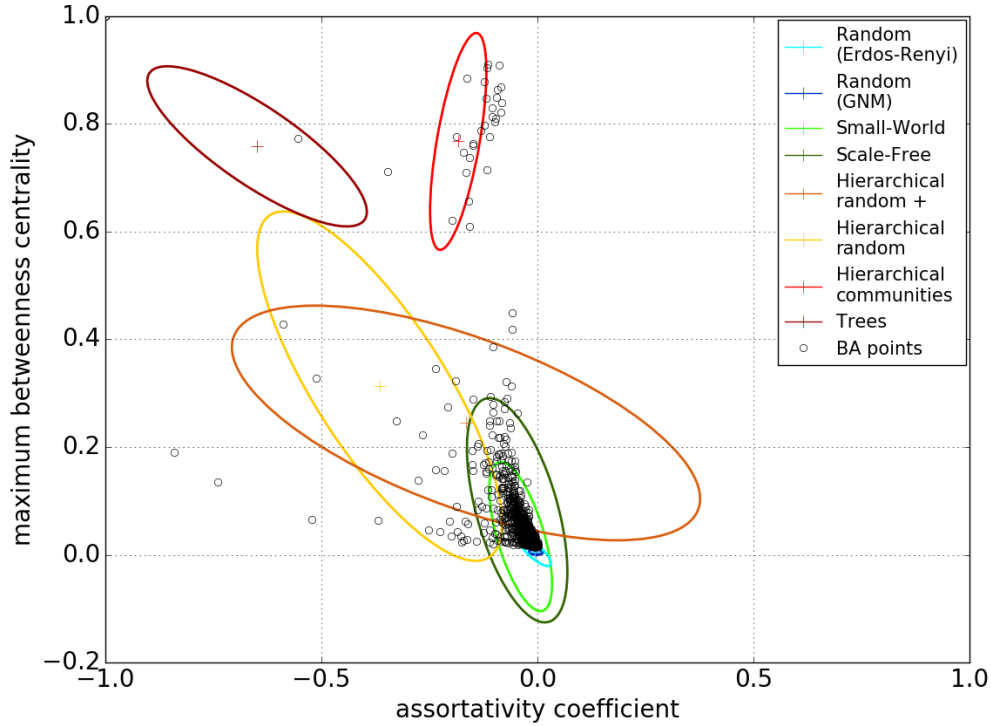


Figure D.18: Assortativity coefficient and maximum betweenness centrality values for the 1000 graphs generated by the BA graph model.

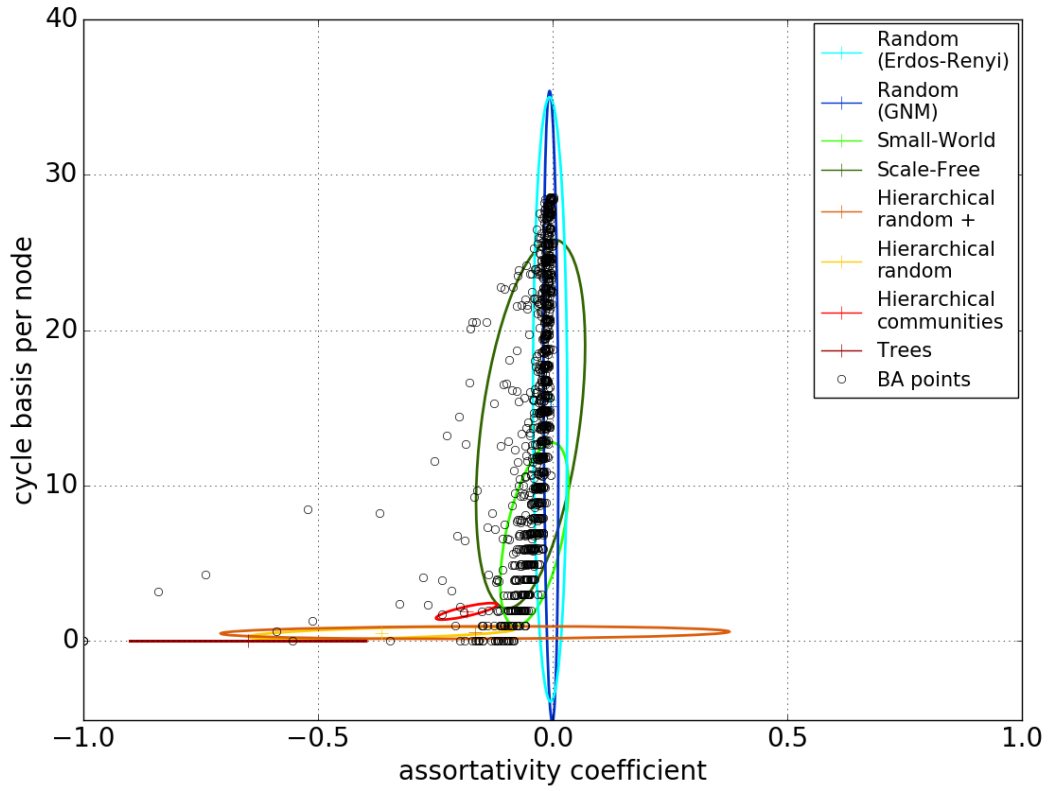


Figure D.19: Assortativity coefficient and number of cycle basis for the graphs generated by the BA graph model.

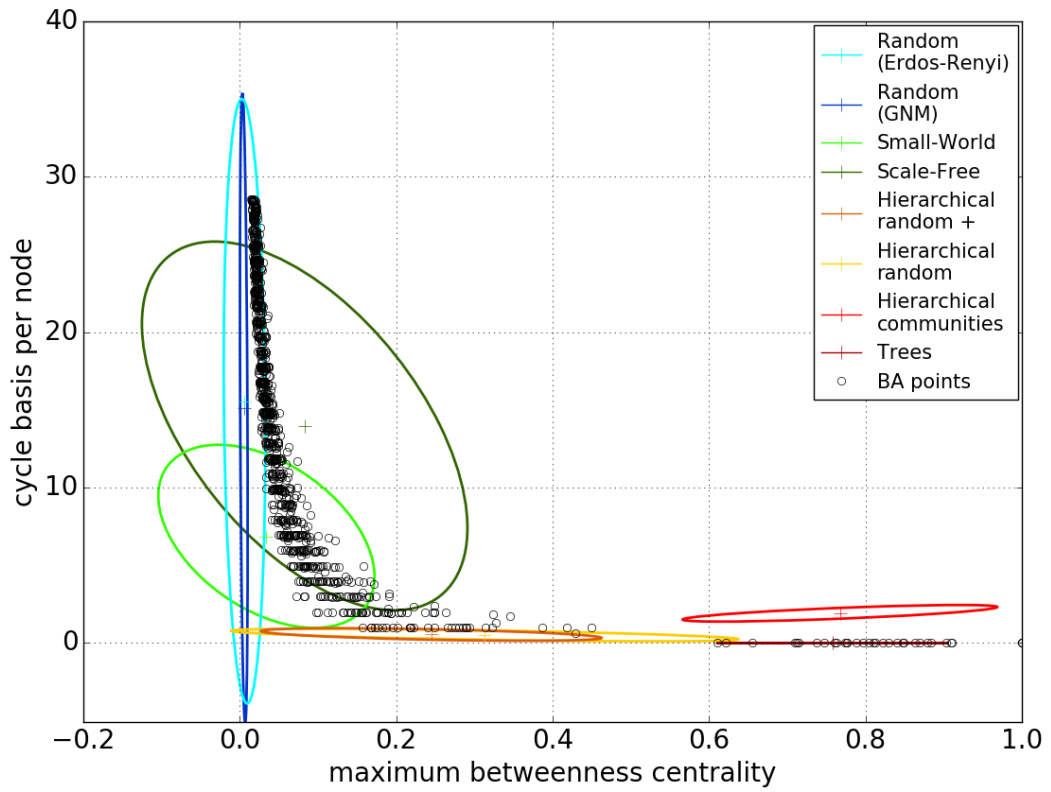


Figure D.20: Points for the maximum betweenness centrality and number of cycle basis as calculated for the set of graphs generated by the BA graph model.

D.2.5 HR

1000 exemplar graphs have been generated using the HR (hierarchical random) graph model to from part of the synthetic suite of graphs. Following the calculation of the selected metrics, AC, MBC and number of CB per node, Figure D.21 shows the values for the AC and the MBC. Points for the AC and number of CB per node for the 1000 networks are presented in Figure D.22 and Figure D.23 shows the values for the MBC and the number of CB per node.

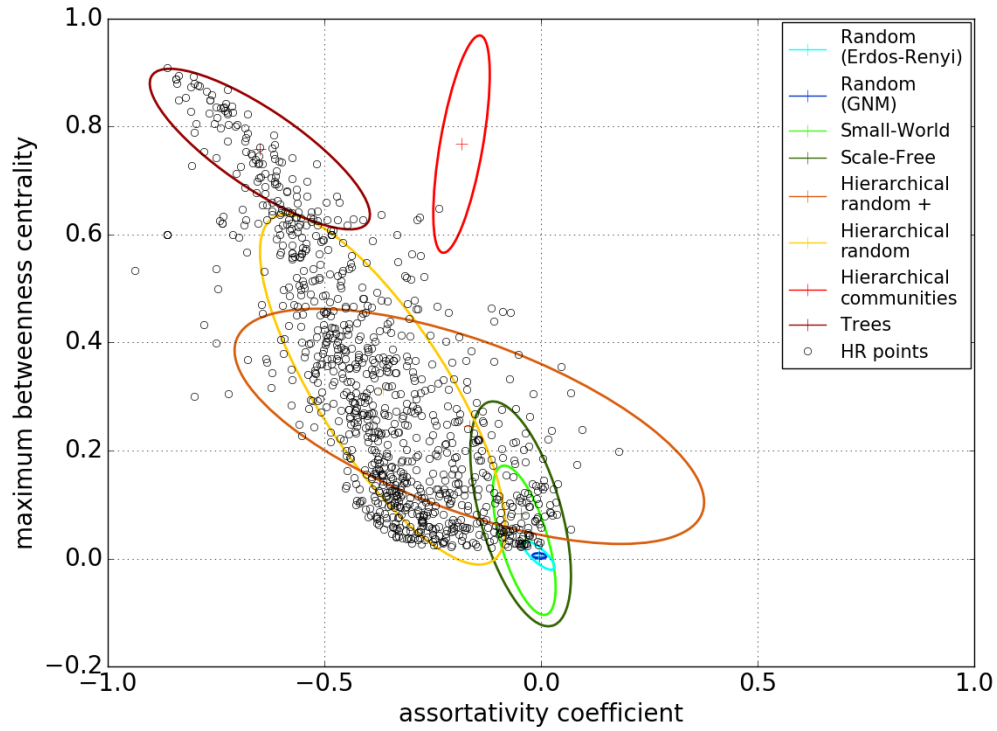


Figure D.21: Points for graphs generated by HR for the assortativity coefficient and maximum betweenness centrality.

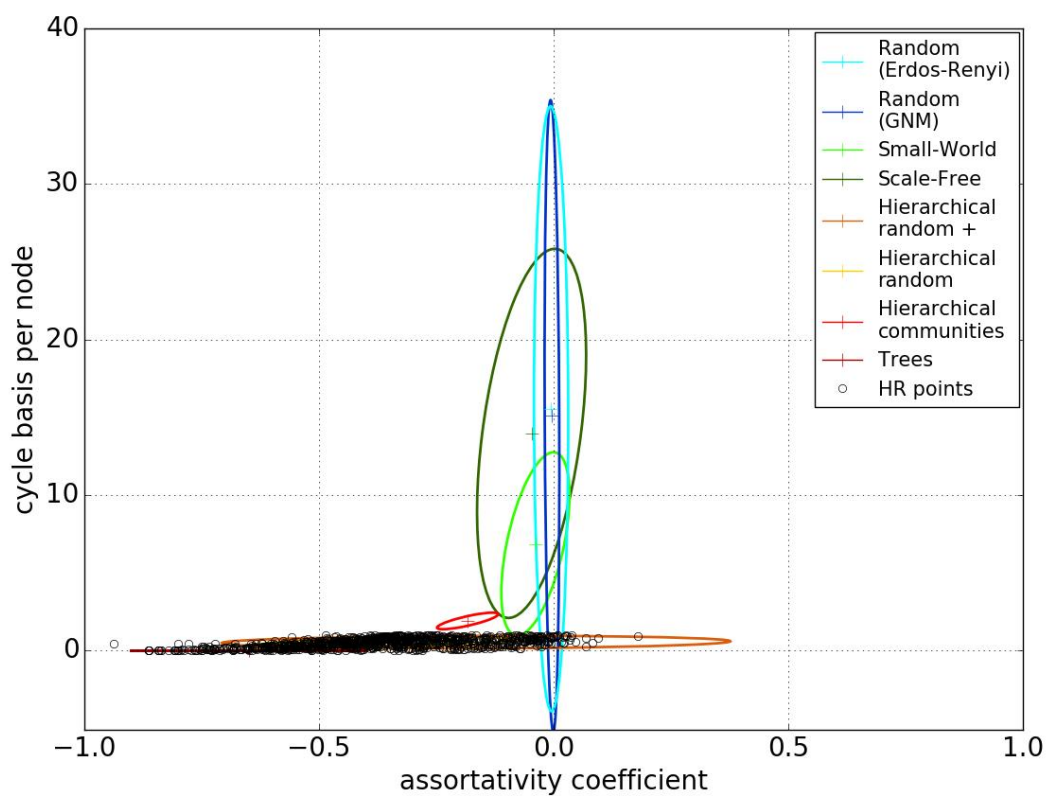


Figure D.22: Values for the assortativity coefficient and number of cycle basis for graphs generated by the HR model.

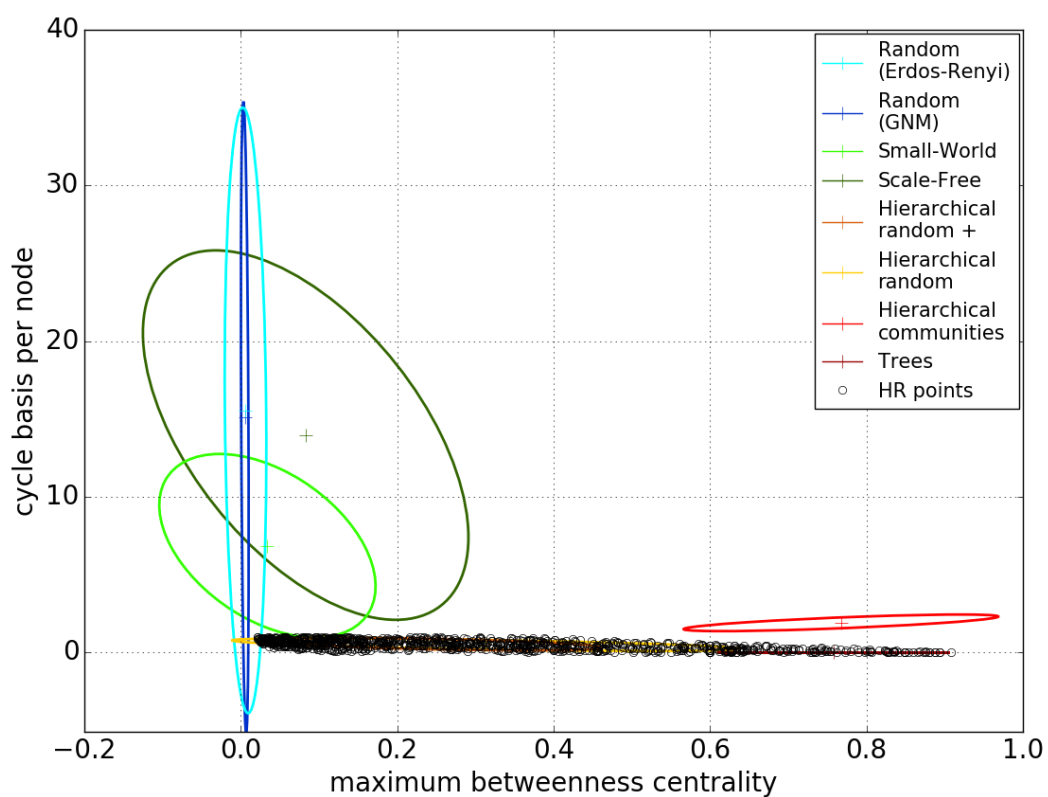


Figure D.23: Maximum betweenness centrality and the number of cycle basis per node for each of the graphs generated by the HR graph model.

D.2.6 HR+

The HR+ graph model has been used to generate an ensemble of 1000 graphs which have subsequently been analysed using the selected graph metrics, the AC, the MBC and the number of CB per node. The results for the AC and the MBC are given in Figure D.24 with the points for the AC metric and number of CB per node in Figure D.25. The final set of points for the MBC and the number of CB per node are given in Figure D.26.

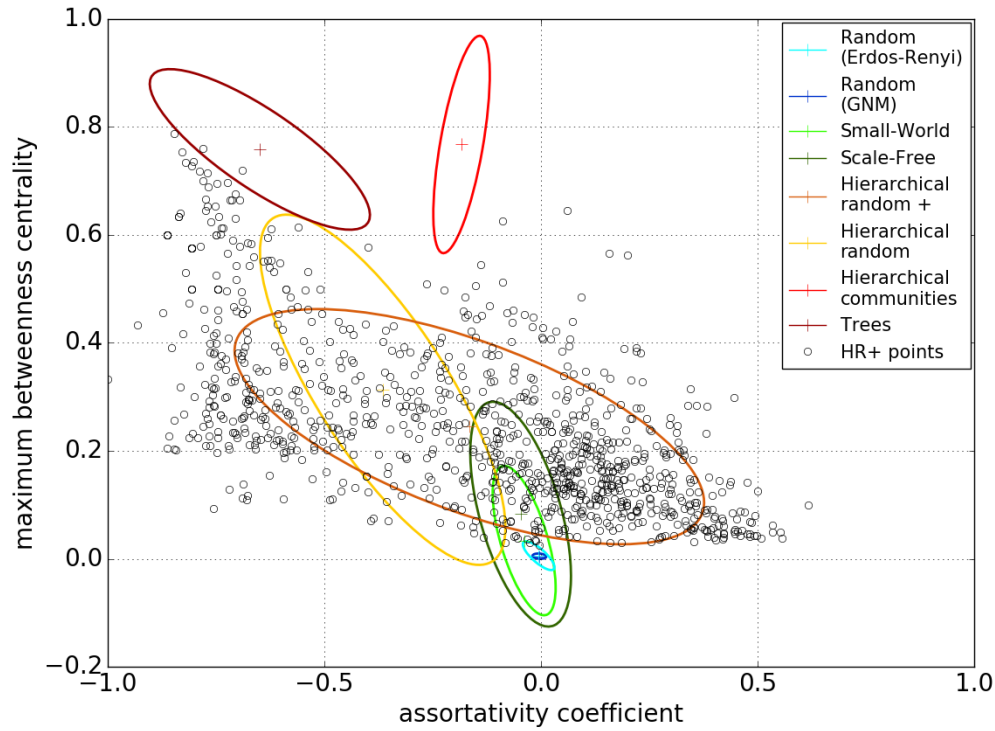


Figure D.24: Values for the graphs generated by the HR+ model for the assortativity coefficient and maximum betweenness centrality metric values..

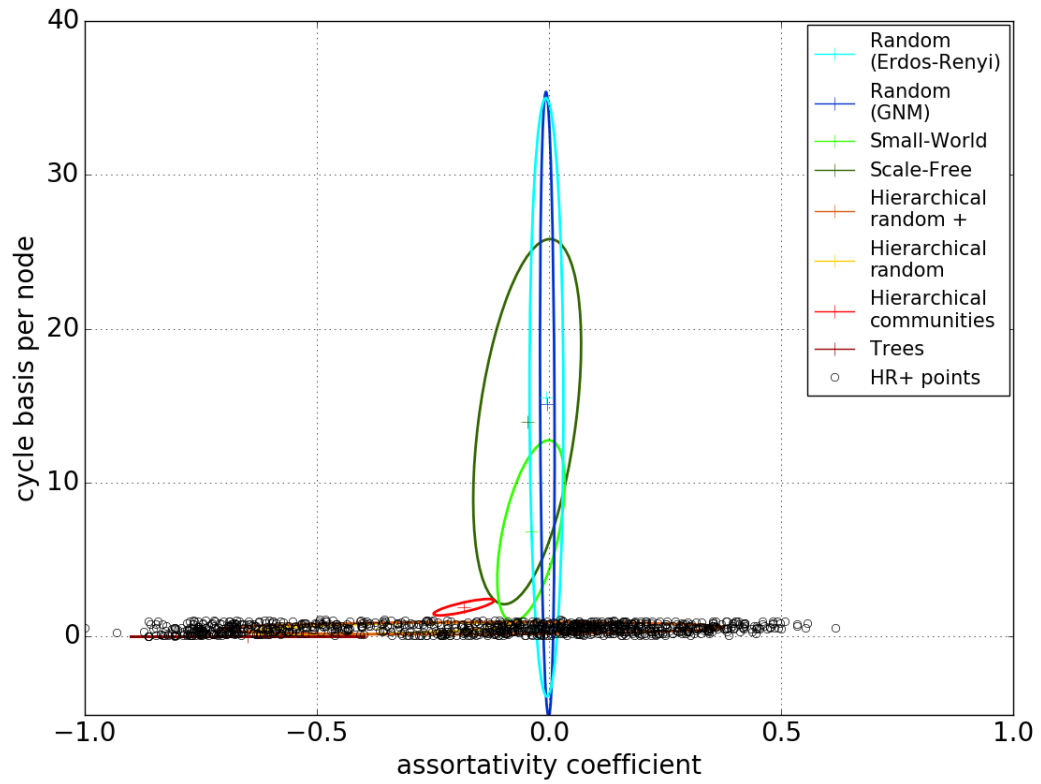


Figure D.25: Values for the assortativity coefficient and the number of cycle basis per node for all graphs generated by the HR+ graph model.

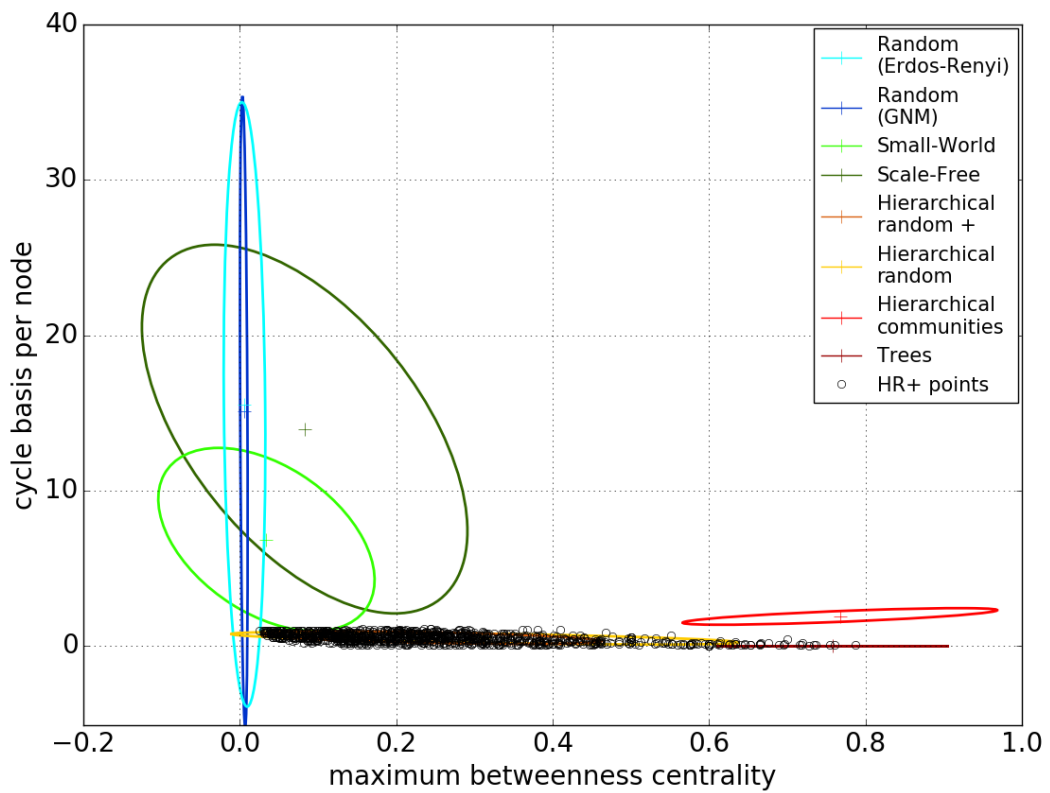


Figure D.26: Maximum betweenness centrality and the number of cycle basis per node values for all graphs generated by the HR+ graph model.

D.2.7 HC

The HC (Hierarchical communities) model has been used in the generation of graphs for the synthetic suite of networks. Through plotting the values for the three selected metrics, the AC, the MBC and the number of CB per node, the characteristics of the graphs generated by the graph model can be compared to those of the other seven models employed in the generation of graphs. Figure D.27 presents the points for the AC and MBC, with the points for the AC and the number of CB per node given in Figure D.28. The values for the final metric pairing, the MBC and the number of CB per node are presented in Figure D.29.

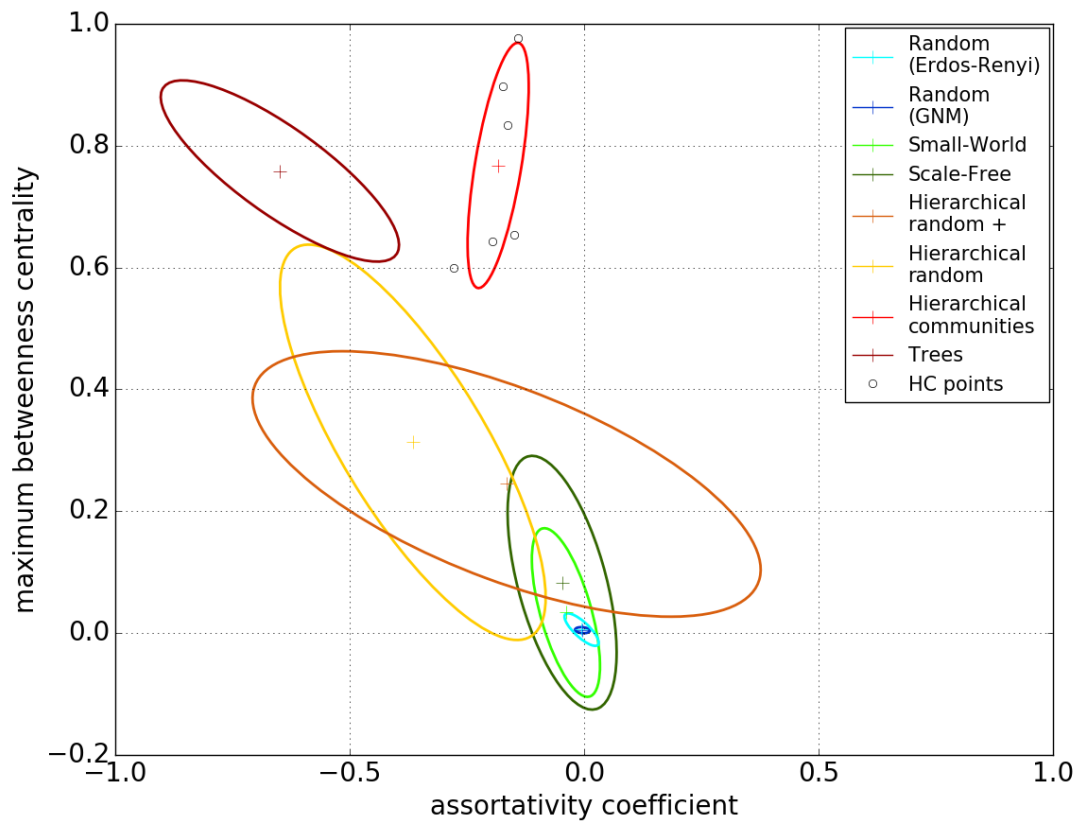


Figure D.27: Points for the HC graph model generated graphs for the assortativity coefficient and the maximum betweenness centrality.

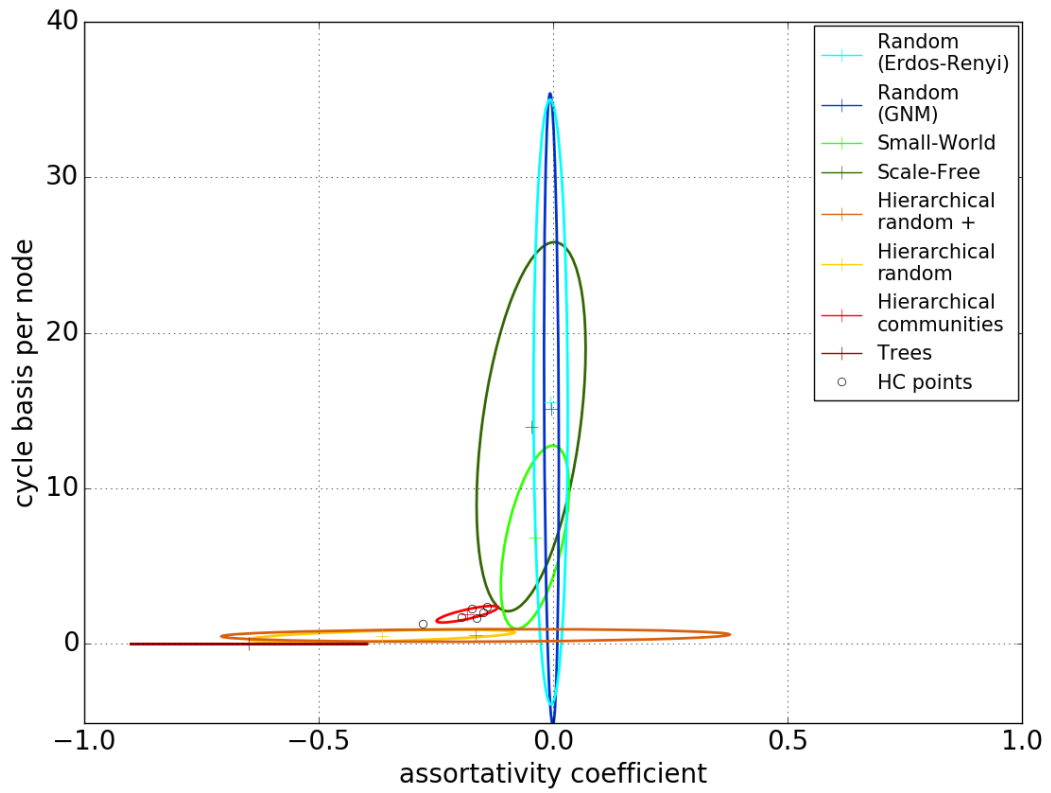


Figure D.28: Points for the assortativity coefficient and number of cycle basis per node for each of the graphs generated using the HC graph model.

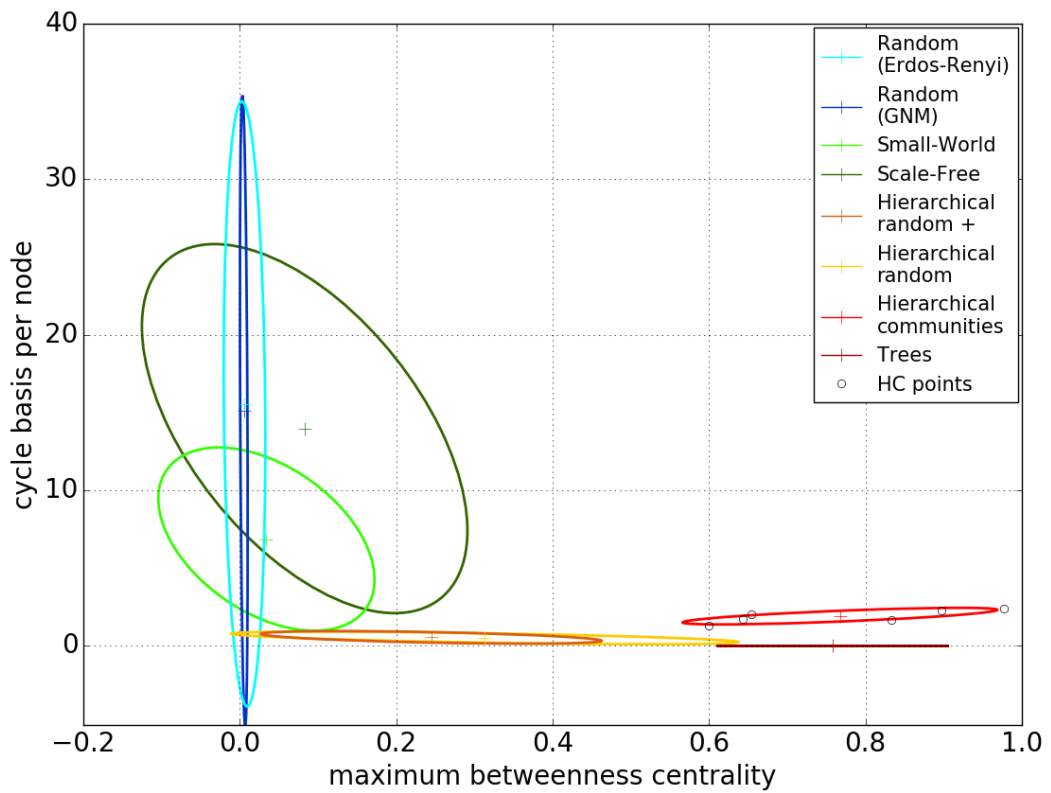


Figure D.29: Calculated values for the graphs generated using the HC graph model for the maximum betweenness centrality and the number of cycle basis per node metrics.

D.2.8 TREE

Using the TREE graph model an ensemble of graphs have been generated for the wider suite of synthetic graphs. Three metrics, the AC, the MBC and the number of CB per node, have been calculated for graphs to help in the characterisation of them. Through plotting each graph against the single standard deviation ellipses for each of the graph models, a comparison between models can be made. Results for the AC and MBC are presented in Figure D.30, with the results for the AC and the number of CB per nodes are presented in Figure D.31. Figure D.32 shows the final set of results for the MBC and number of CB per node.

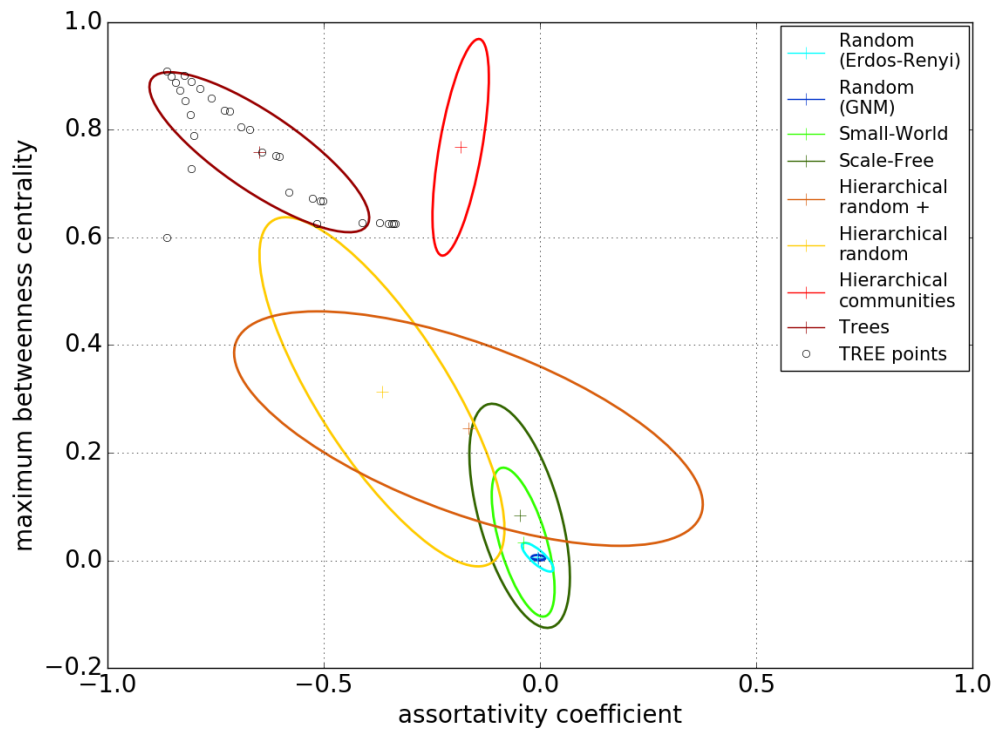


Figure D.30: Metric values for the assortativity coefficient and maximum betweenness centrality values for the TREE graphs.

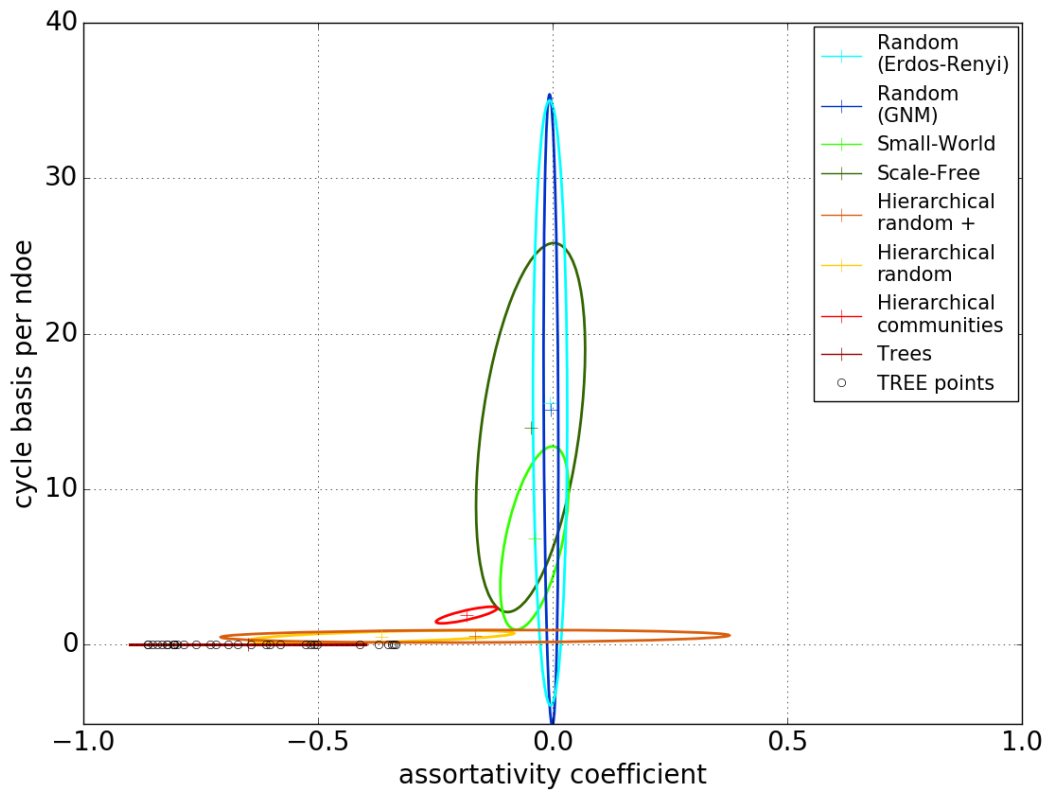


Figure D.31: Assortativity coefficient and number of cycle basis for the graphs generated by the TREE model.

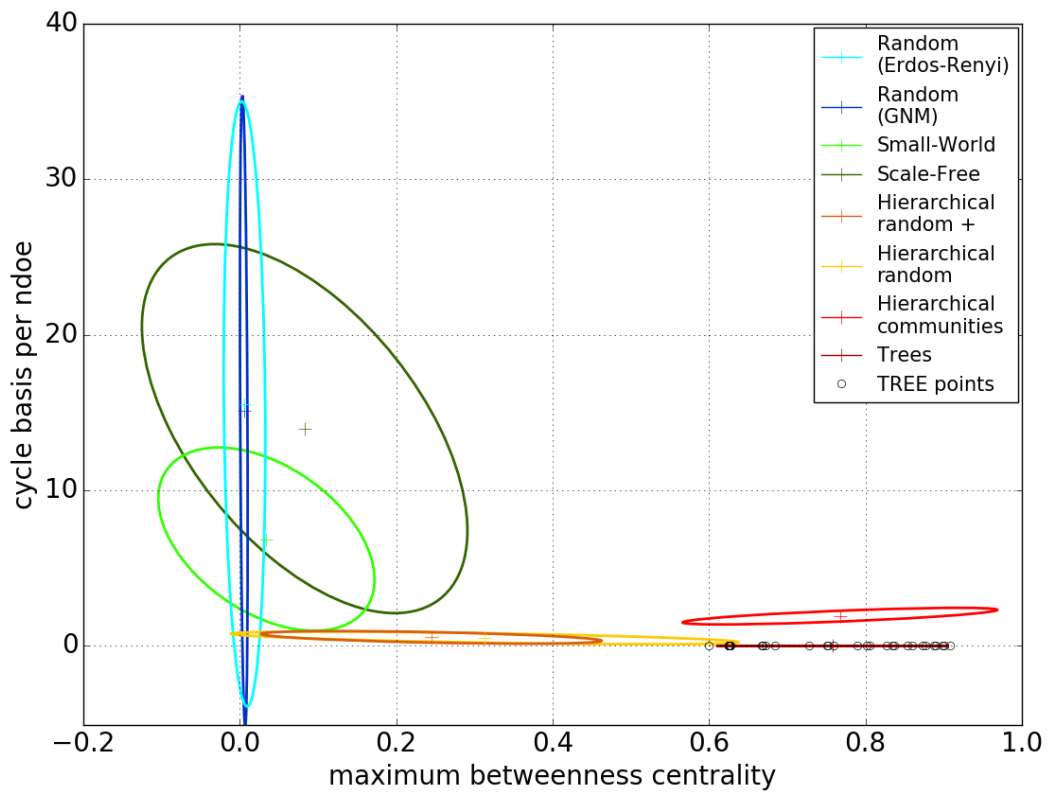


Figure D.32: Points for the maximum betweenness centrality and number of cycle basis per node for each of the graphs generated by the TREE graph model.

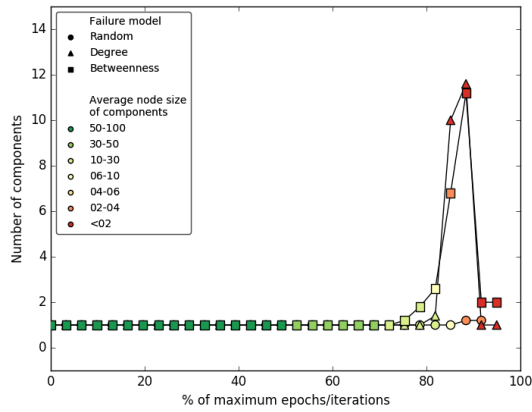
D.3 Robustness of synthetic graphs

Graphs generated using the spectrum of synthetic graph models have been analysed using a developed failure model (Chapter 3, Section 3.5 (page 59)) which employs three methods of assessing the robustness of the graphs to perturbations. The failure simulations were run over 500 graphs from each of the eight graph models with the exception of the HC and TREE models where the full suite was run, 7 and 31 graphs (Table D.2). This allows for the results from each model to be compared against each other as well as for the different cardinality of graphs to be accounted for. A subset of results from the graphs of different types are used to present the results, with those selected presenting the spectrum of behaviours exhibited in each case. The results for the selected six graphs for the ER model are presented in Figure D.33 with those for the second random model presented in Figure D.34 (page 286). Figure D.35 (page 287), shows the results for the WS (small-world) graphs with the results in Figure D.36 (page 288) showing the results for BA graphs. Results for the selected six graphs from the set generated by the HR model are given in Figure D.37 (page 289) and the HR+ graph results are presented in Figure D.38 (page 290). The results for the final two sets of graphs, those generated by the HC and TREE model, are presented in Figure D.39 (page 291) and Figure D.40 (page 292).

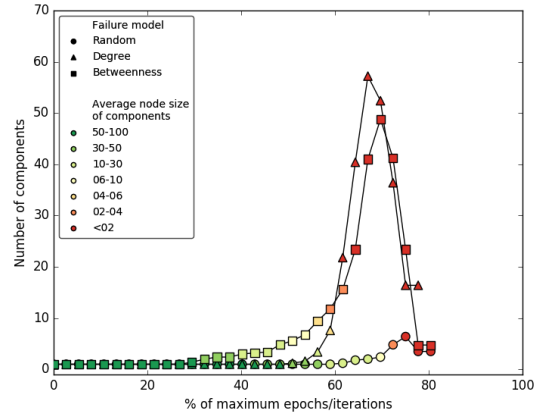
Graph model	Number of graph type in suite	Number of graph employed in failure simulations
ER	1000	500
GNM	1000	500
WS	1000	500
BA	1000	500
HR	1000	500
HR+	1000	500
HC	7	7
TREE	31	31

Table D.2: The number of graphs used for failure simulations from each of the eight graph models.

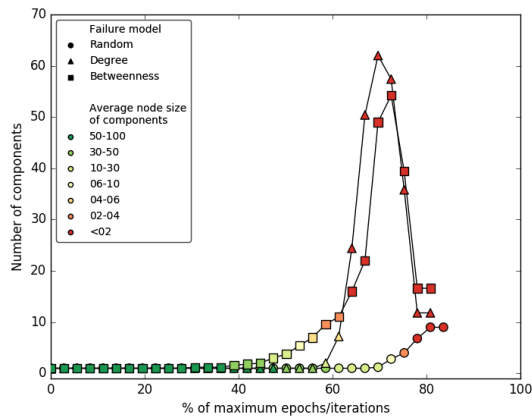
Number of nodes: 397



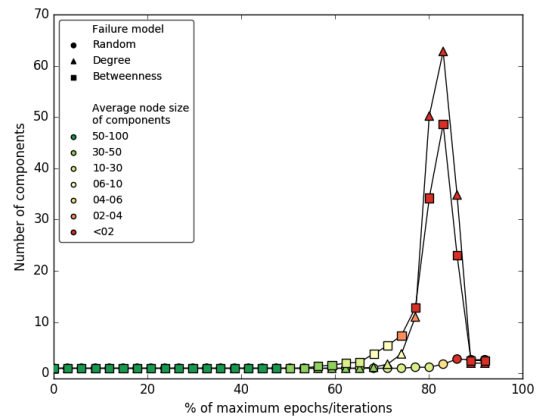
Number of nodes: 746



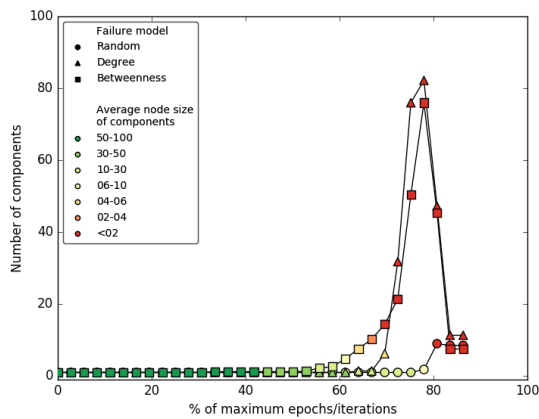
Number of nodes: 861



Number of nodes: 1247



Number of nodes: 1616



Number of nodes: 1965

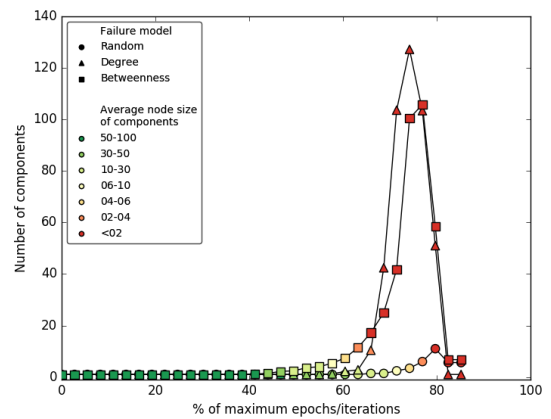
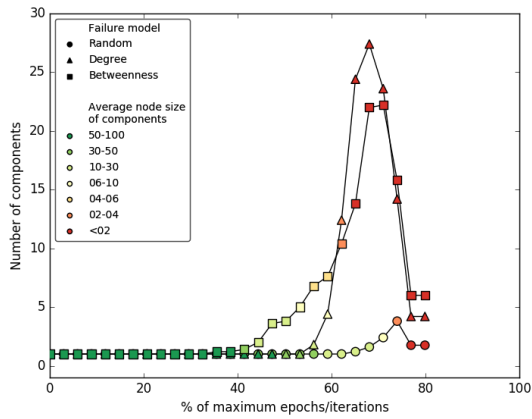
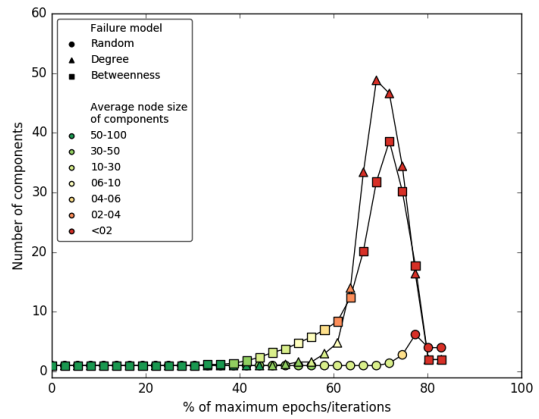


Figure D.33: Behaviour of the selected six ER graphs to the topological failure methods.

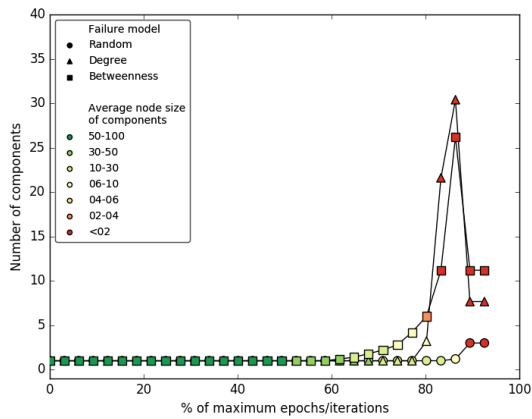
Number of nodes: 338



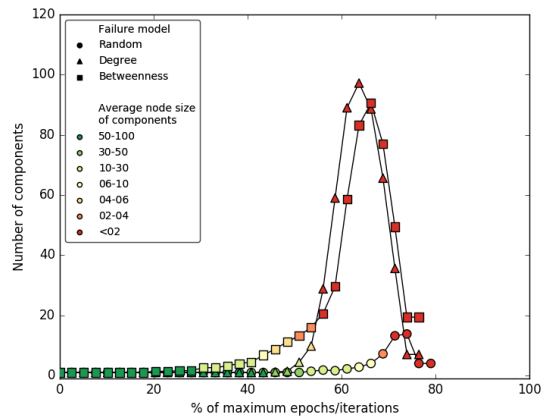
Number of nodes: 651



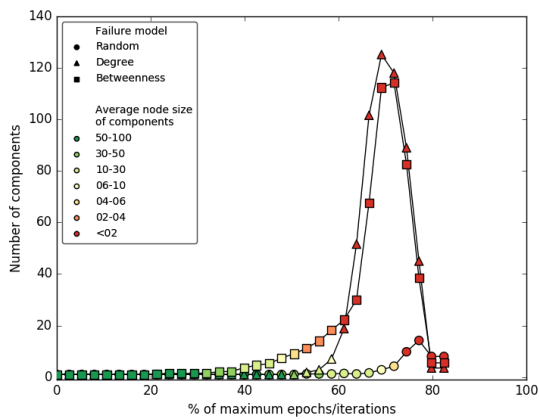
Number of nodes: 940



Number of nodes: 1177



Number of nodes: 1692



Number of nodes: 1982

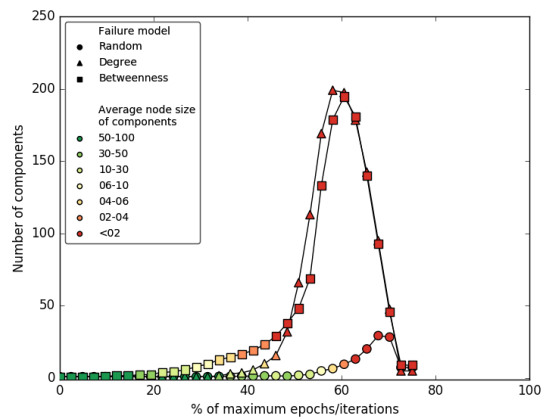
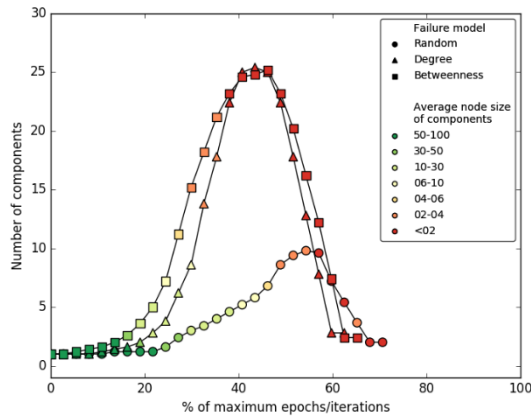
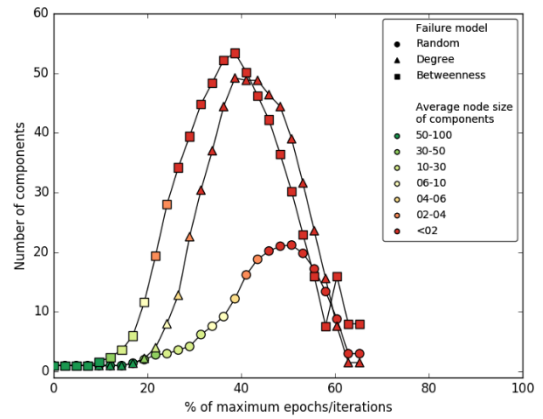


Figure D.34: Behaviour of six selected GNM graphs to the three topological failure methods.

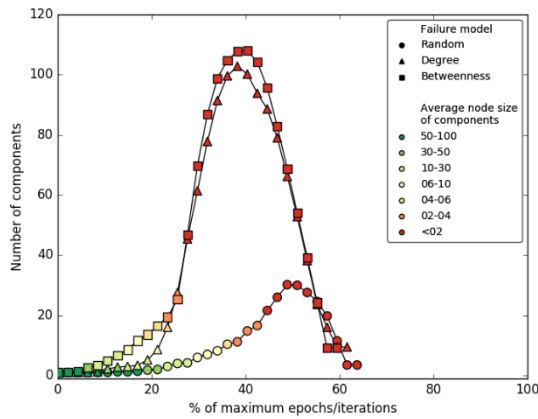
Number of nodes: 184



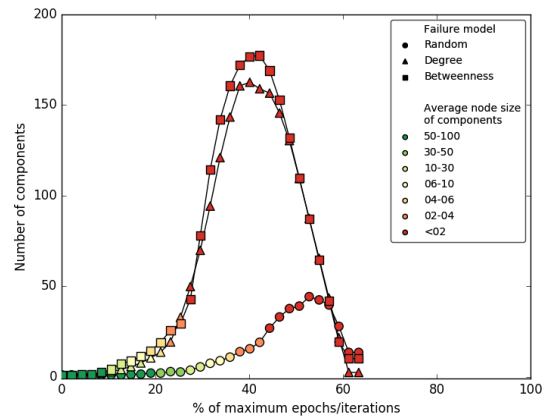
Number of nodes: 331



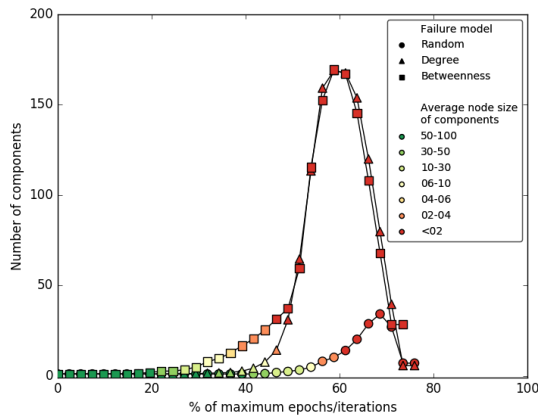
Number of nodes: 706



Number of nodes: 1089



Number of nodes: 1673



Number of nodes: 1966

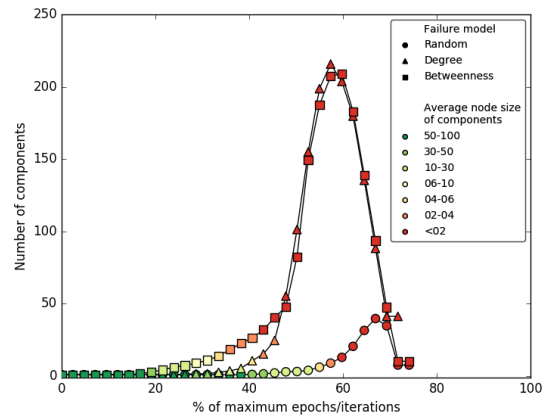


Figure D.35: Behaviour of the six selected WS graphs to the three topological failure methods.

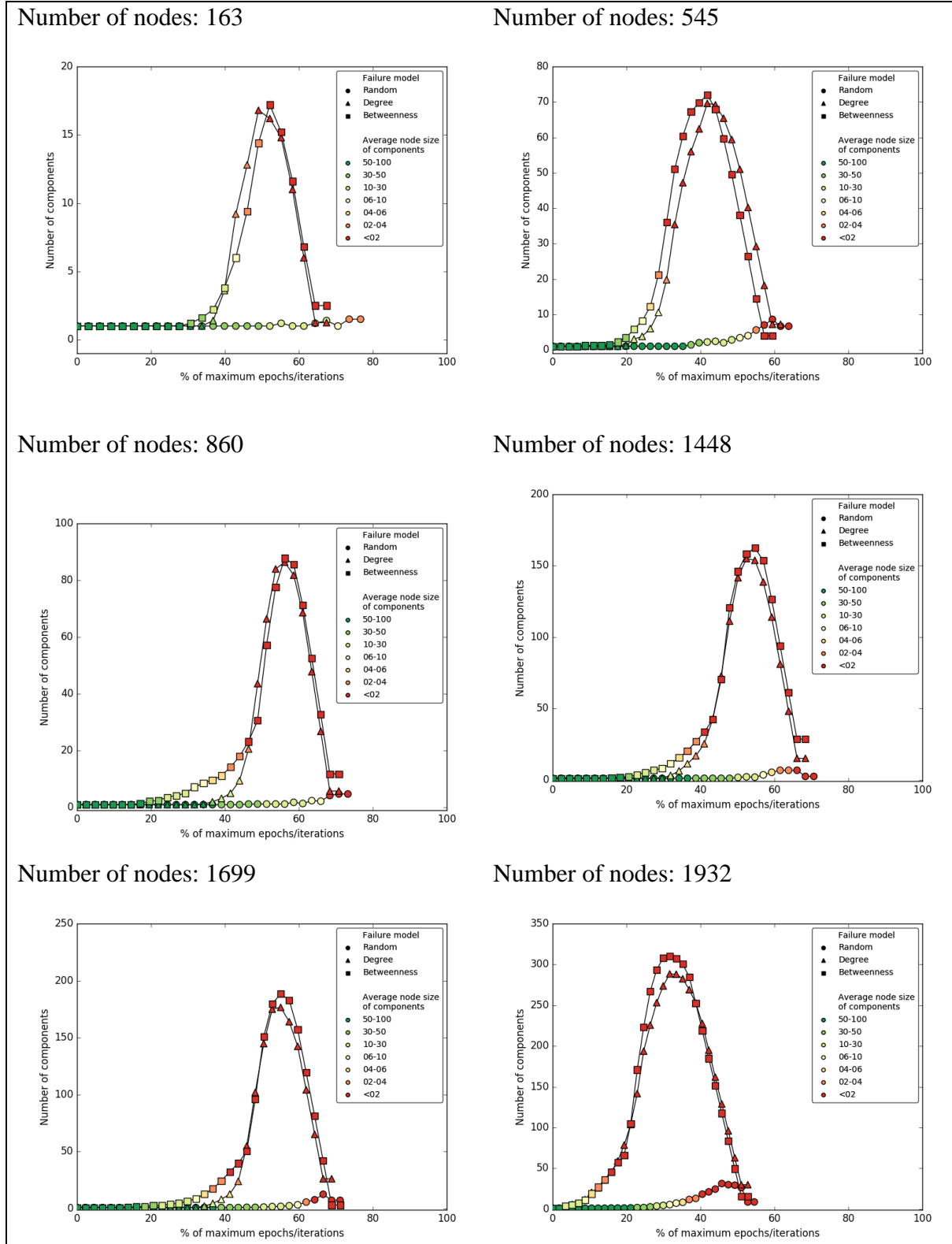
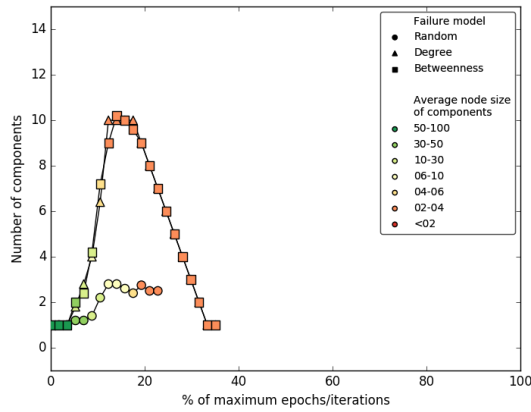
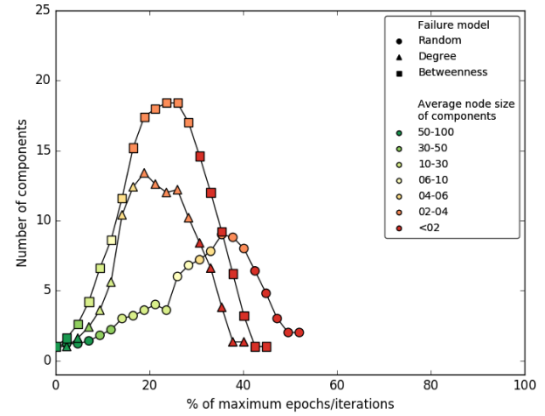


Figure D.36: Behaviour of the six selected BA graphs to the three topological failure methods.

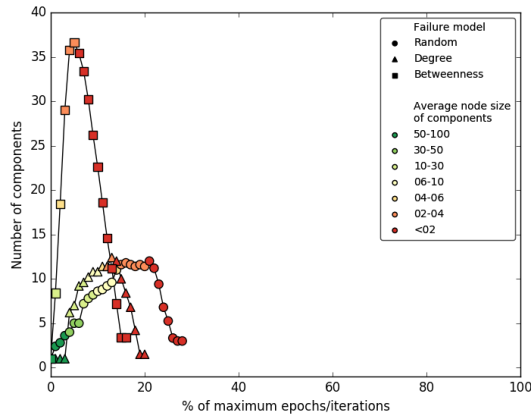
Number of nodes: 57



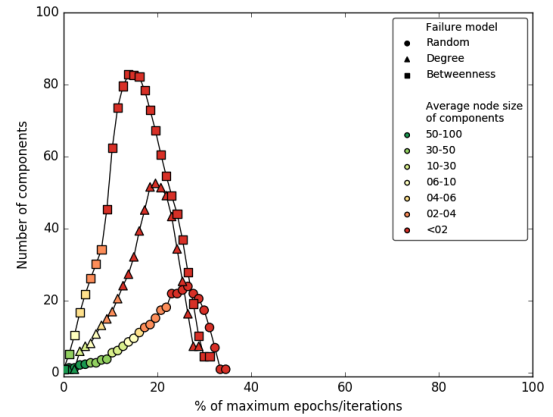
Number of nodes: 127



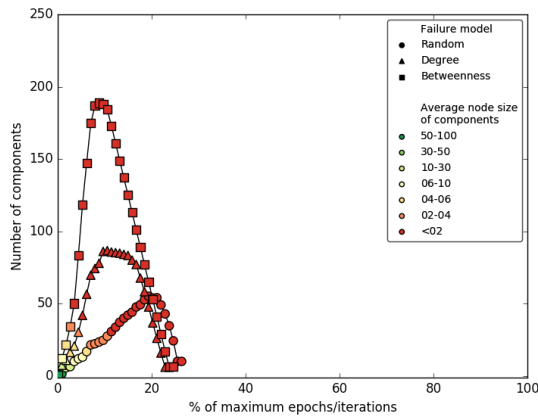
Number of nodes: 400



Number of nodes: 781



Number of nodes: 1365



Number of nodes: 1555

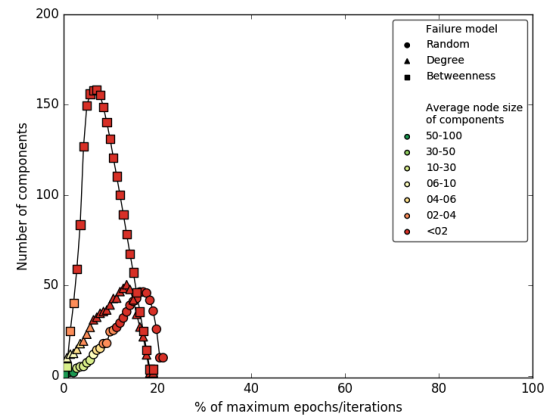


Figure D.37: Behaviour of the six selected HR graphs to the three topological failure methods.

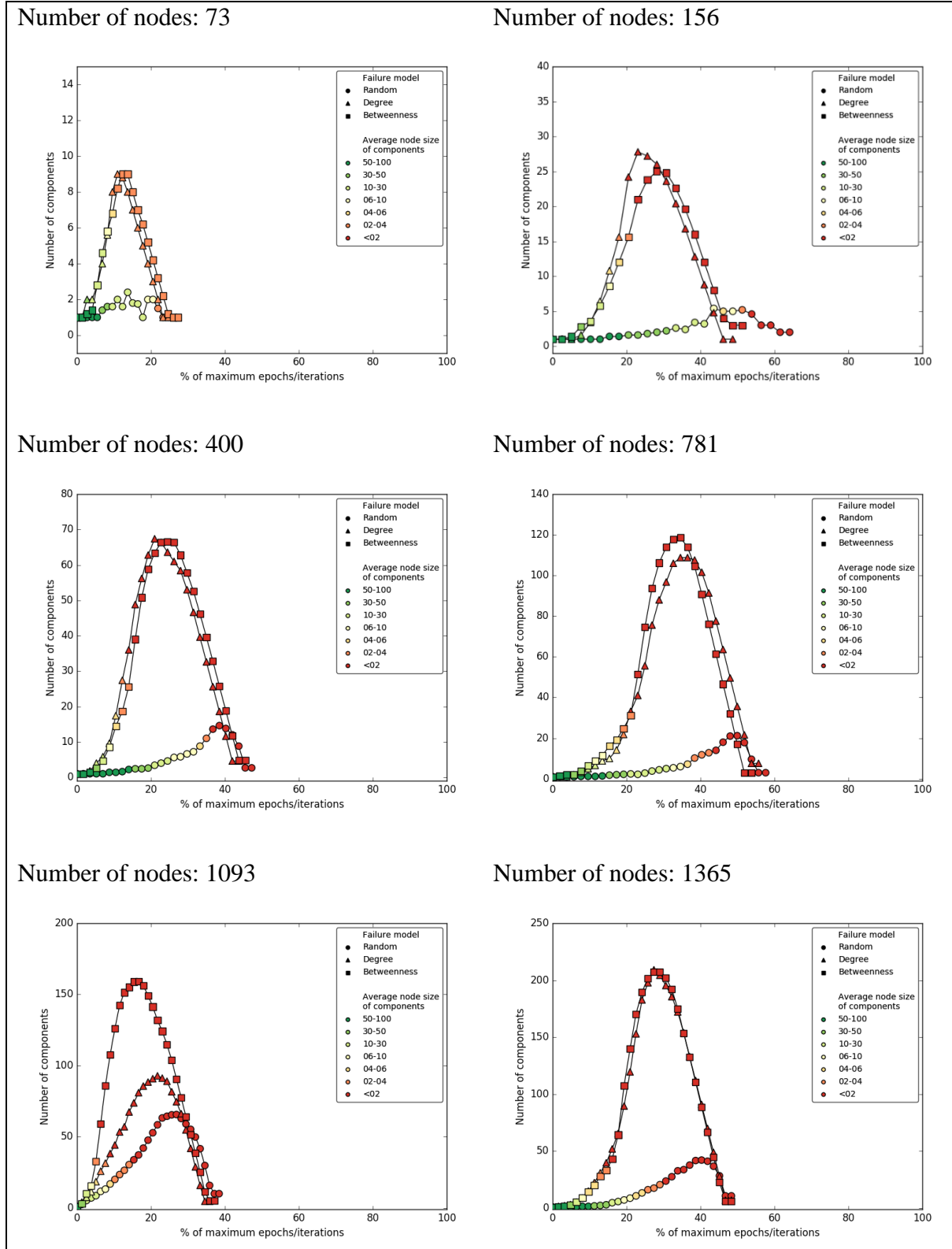
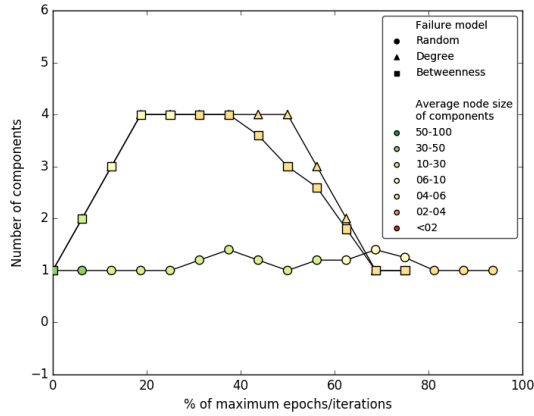
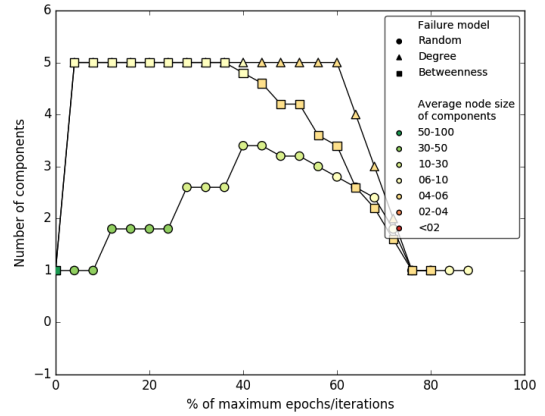


Figure D.38: Behaviour of the six selected HR+ graphs to the three topological failure methods.

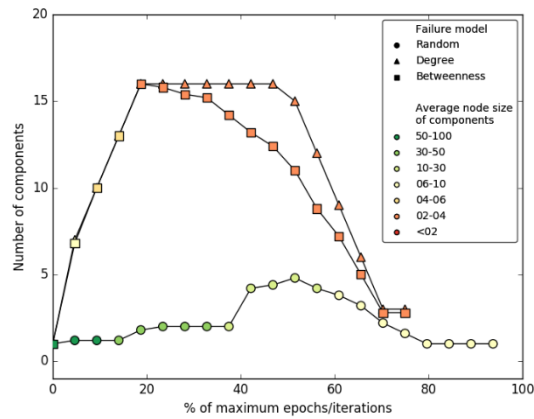
Number of nodes: 16



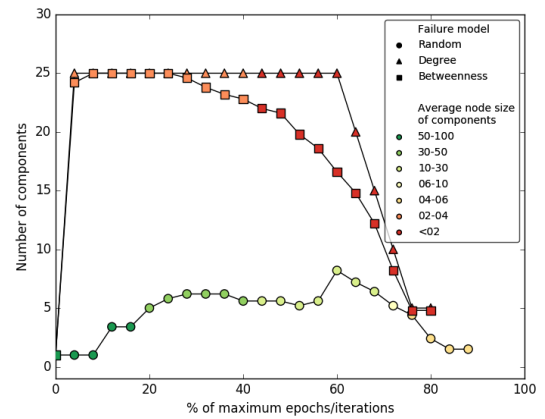
Number of nodes: 25



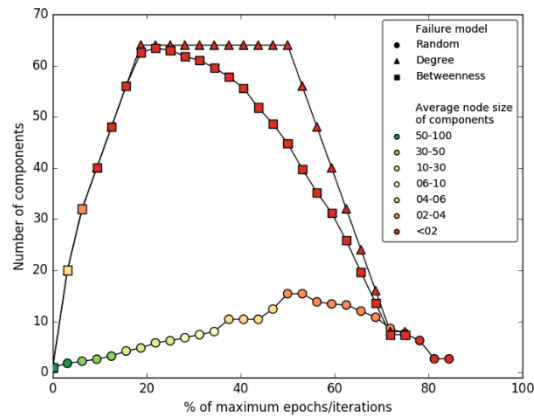
Number of nodes: 64



Number of nodes: 125



Number of nodes: 256



Number of nodes: 625

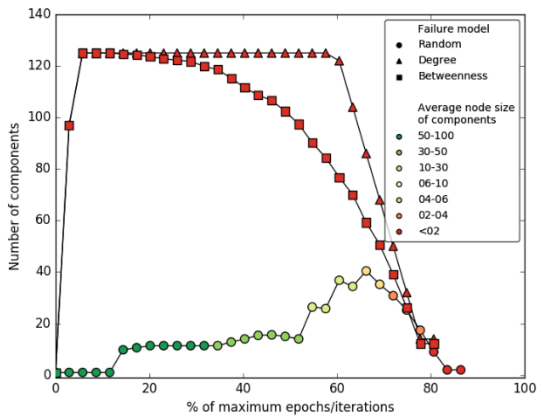
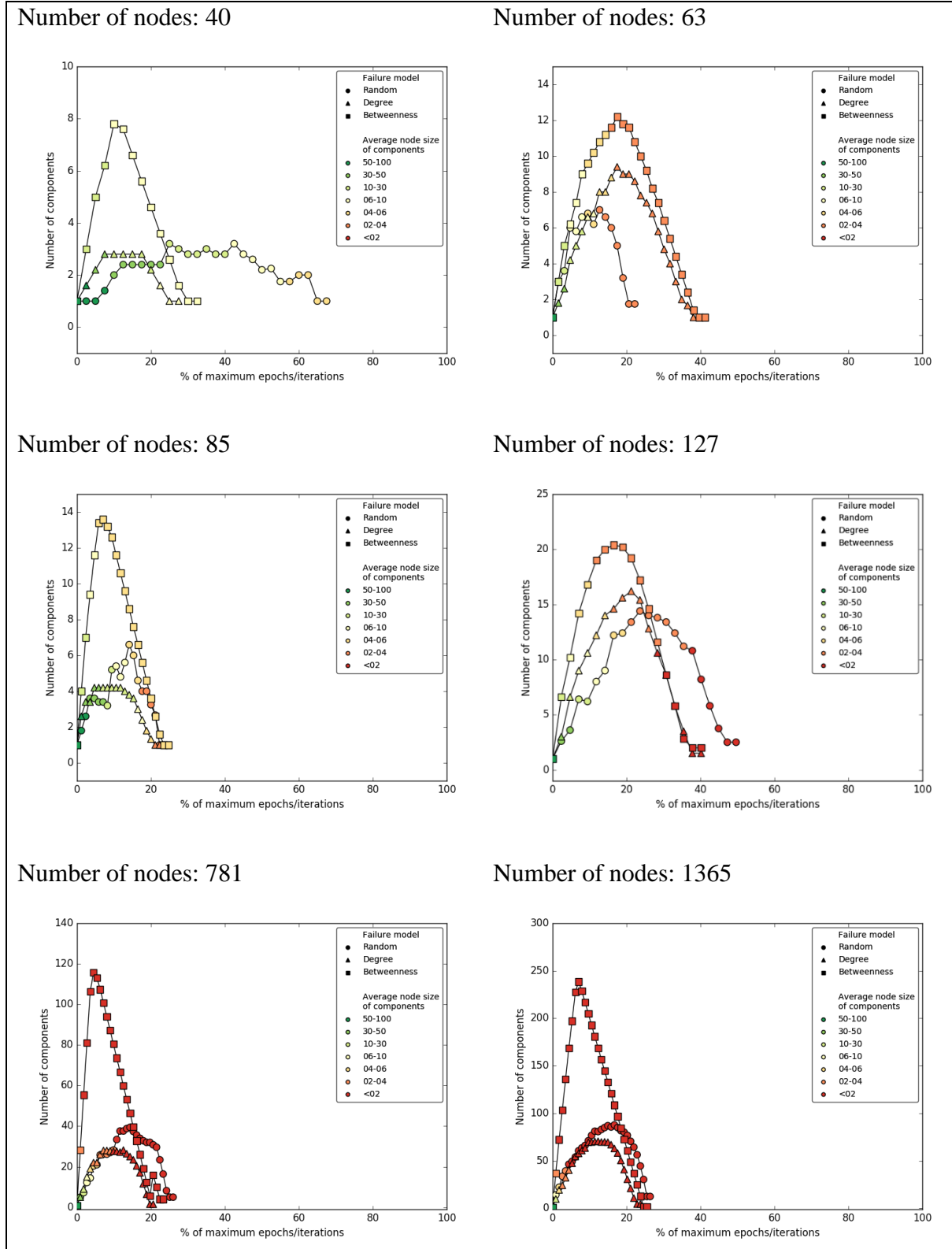


Figure D.39: Behaviour of the six selected HC graphs to the three topological failure methods.



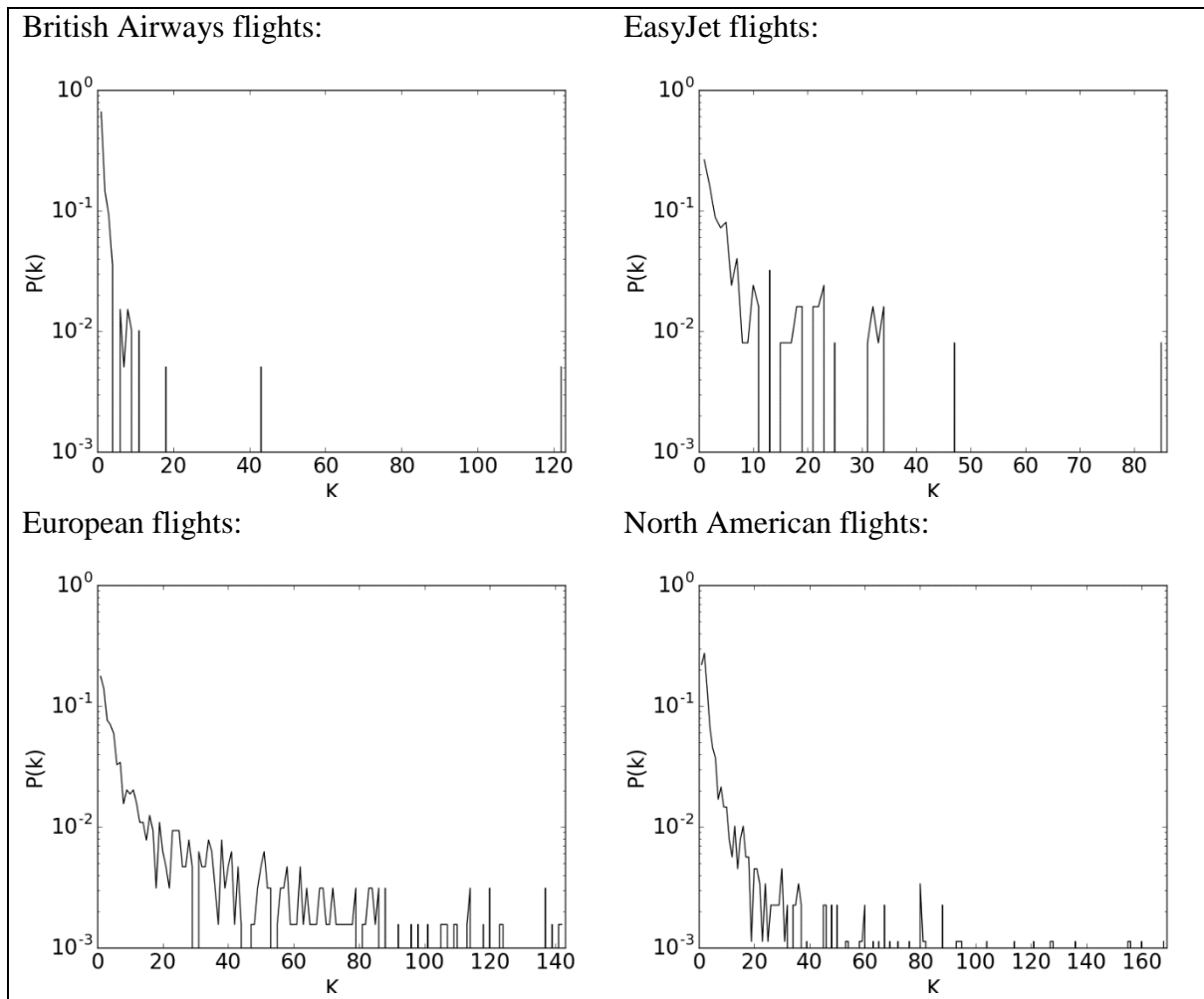
Appendix E: Infrastructure analysis results

Results from the analysis of the suite of critical spatial infrastructure networks with comparisons to the results of the analysis of the synthetic graphs.

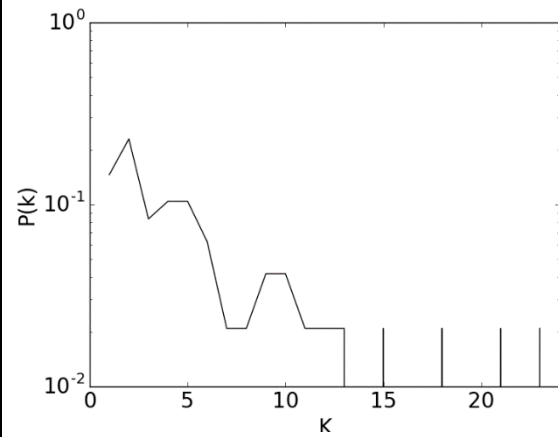
E.1 Degree distributions for critical spatial infrastructure networks

The topology of a network can be characterised in a manner of ways, with one of those being the degree distribution. The degree distribution gives the likelihood of a node being selected with a specific node degree and thus when plotted for a network can be used and compared to those of other networks to characterise a networks topological structure. For the suite of spatial infrastructures developed (Chapter 3, Section 3.3 (page 42)), the plots for each infrastructure network are presented here to provide further insight into each of them.

The plots of the degree distributions for the air networks are presented in Figure E.1 with those for the communication network presented in Figure E.2 (page 295). Figure E.3 (page 296) presents the degree distributions for the energy networks. Figure E.4 (page 297) and Figure E.5 (page 299) present the degree distribution plots for the rail networks, with the national networks presented first followed by the regional light rail networks. The plots for the river networks are then presented, Figure E.6 (page 300), followed by those for the national scale road networks and the regional scale road networks, Figure E.7 (page 300) and Figure E.8 (page 302).



UK flights:



USA flights:

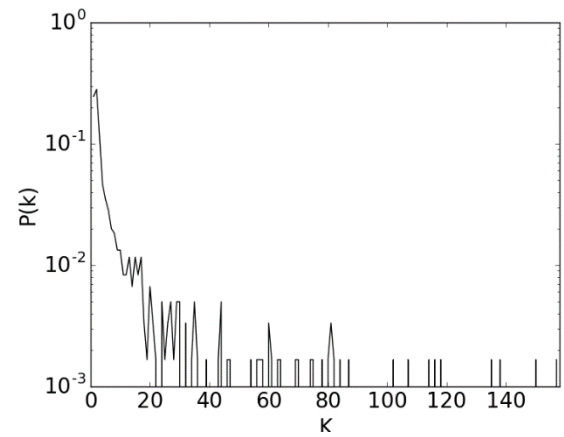


Figure E.1: Degree distribution plots for the suite of air networks, Chapter 3, Section 3.6.1.

Janet:

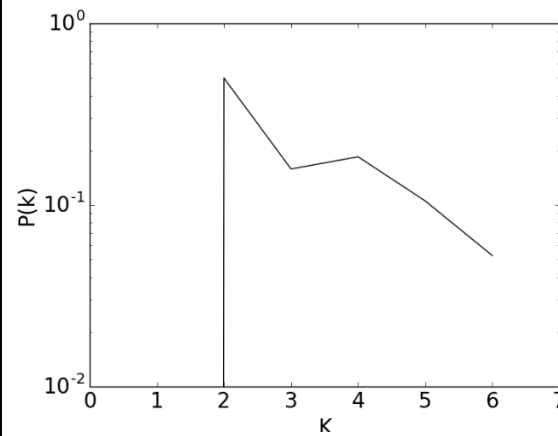
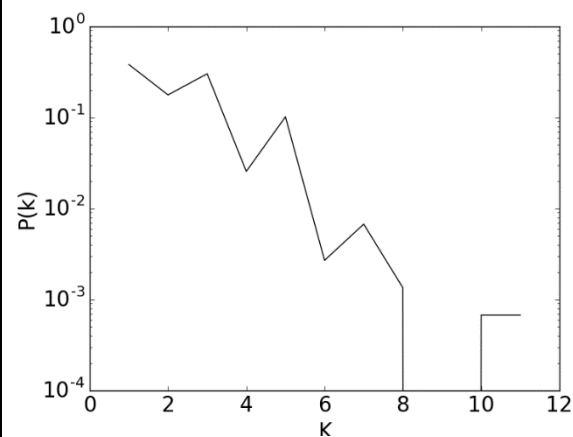
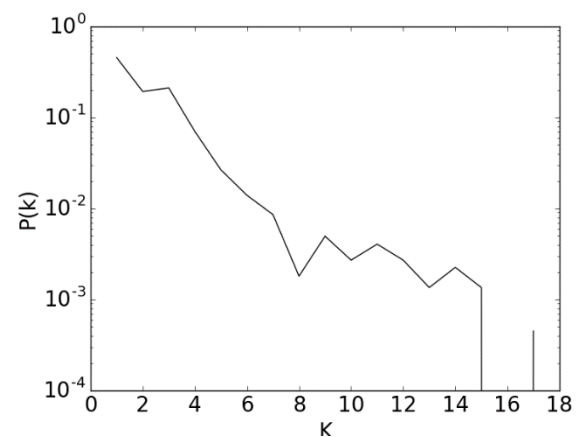


Figure E.2: Degree distribution plot for the communication network, Chapter 3, Section 3.6.2.

National Grid electricity transmission:



National Grid electricity transmission MT:



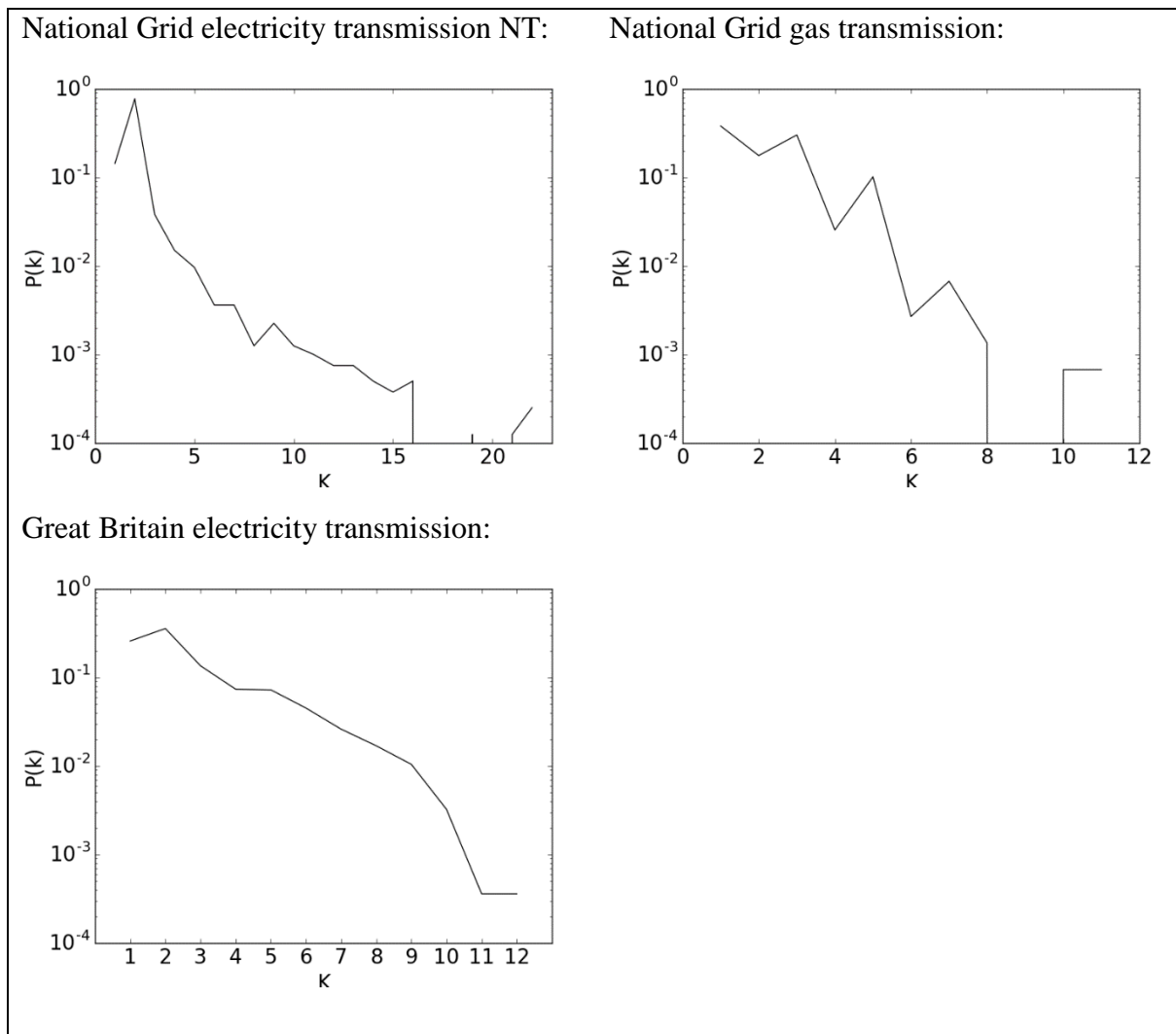
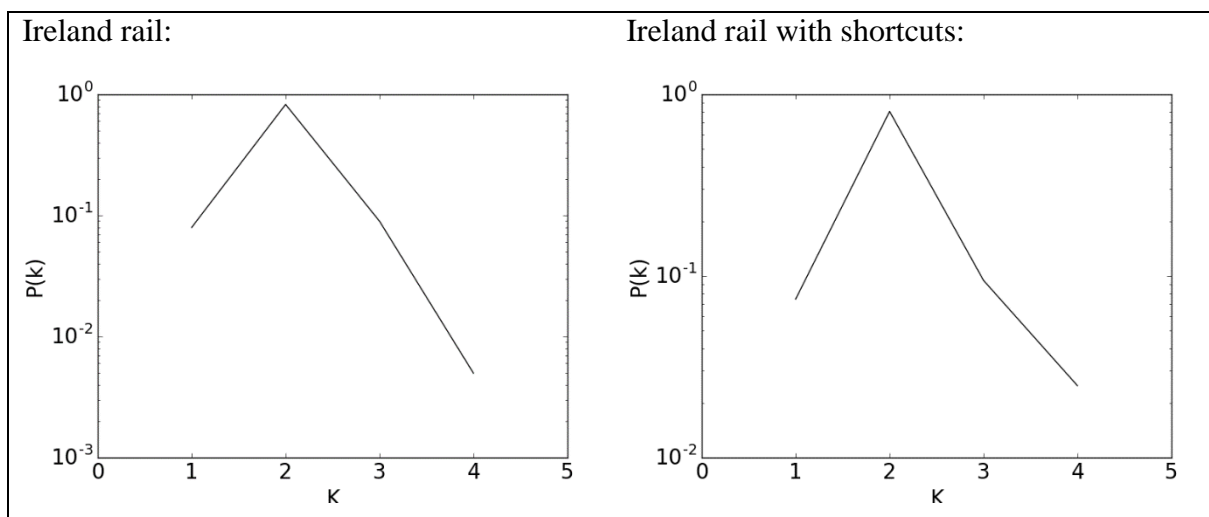


Figure E.3: Degree distribution plots for the suite of energy networks, Chapter 3, Section 3.6.3.



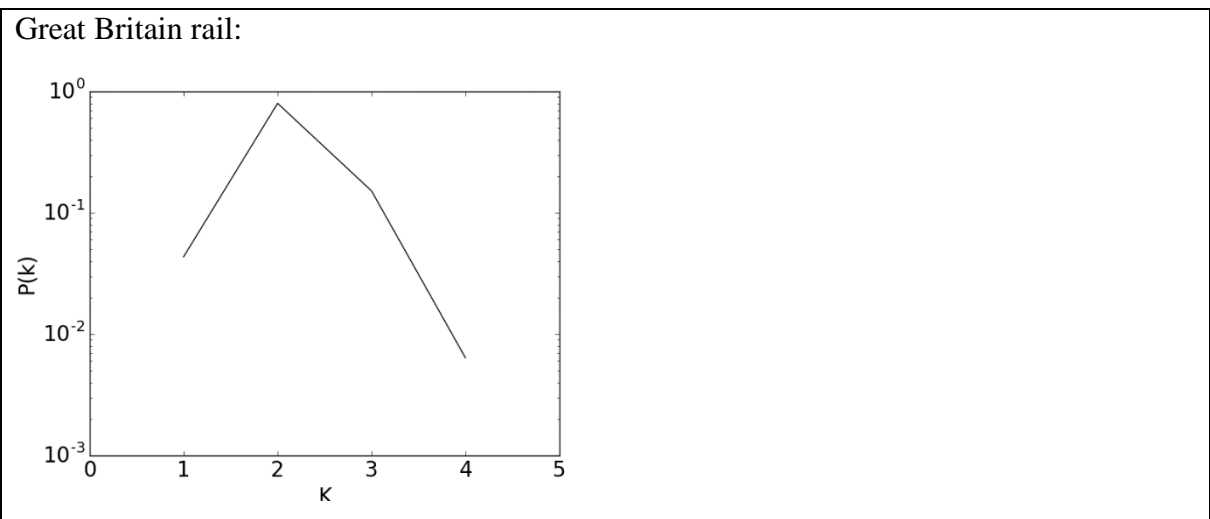
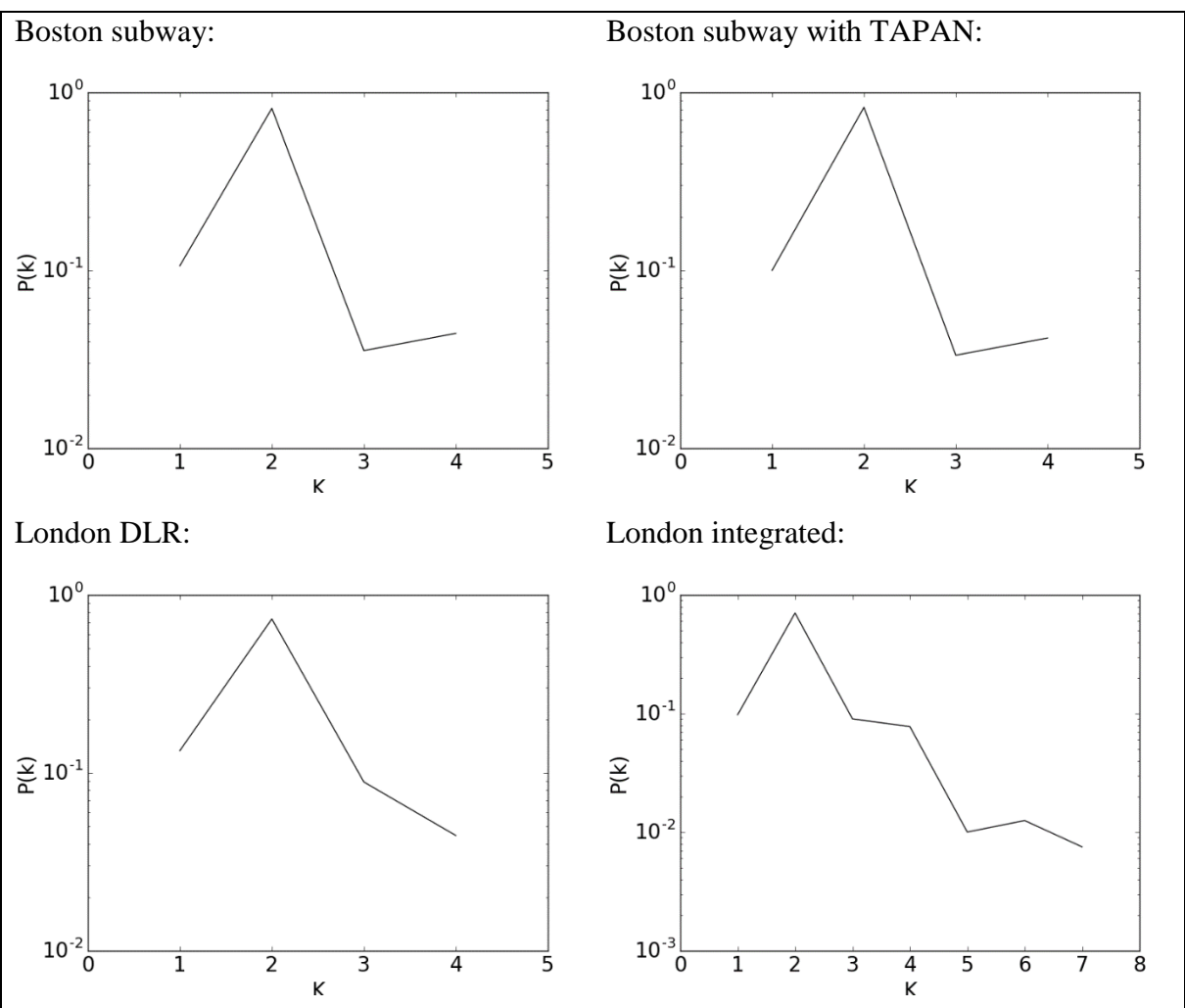
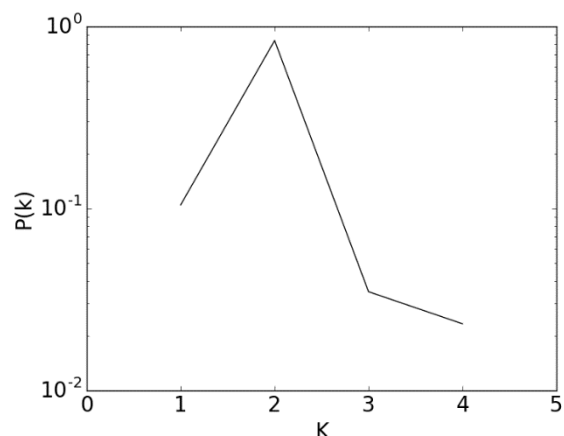


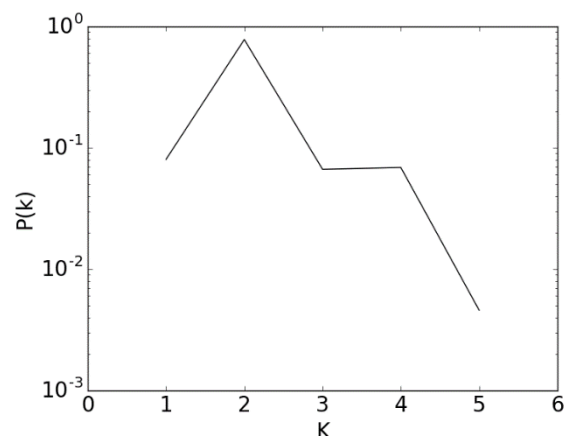
Figure E.4: Degree distribution plots for national rail networks, Chapter 3, Section 3.6.4.



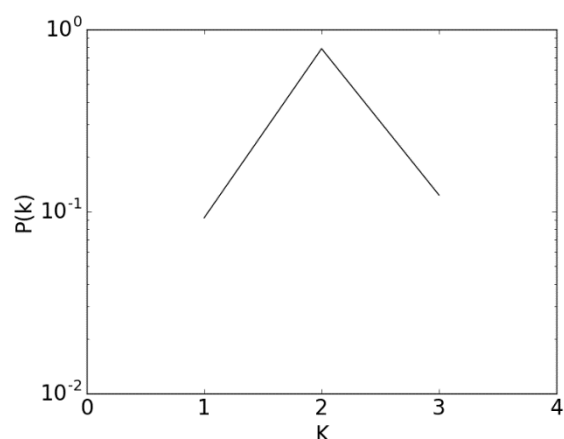
London Overground:



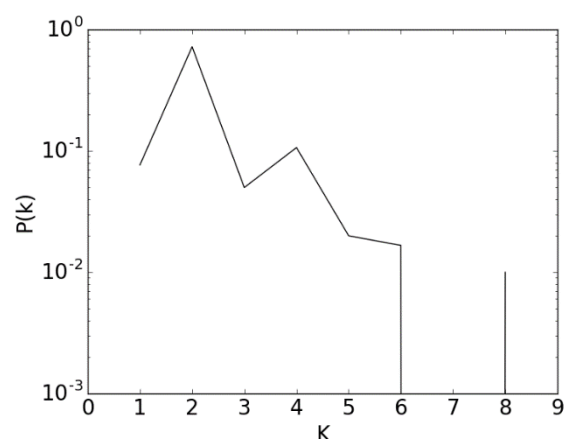
London Tube:



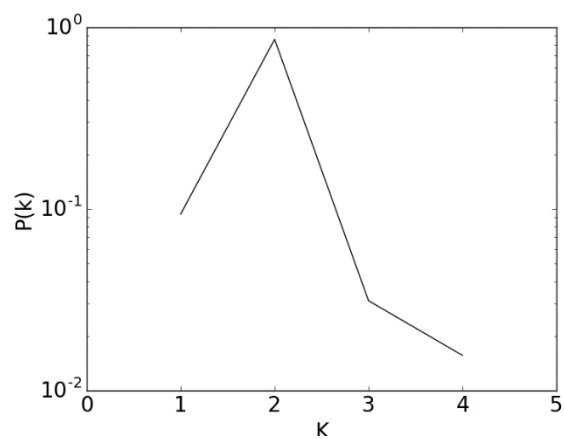
Manchester Metrolink:



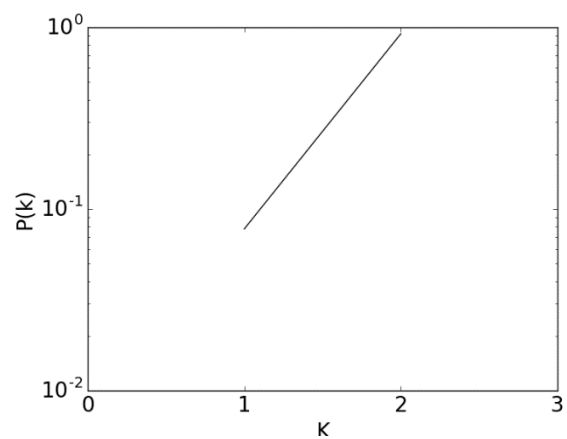
RATP (Paris) metro:



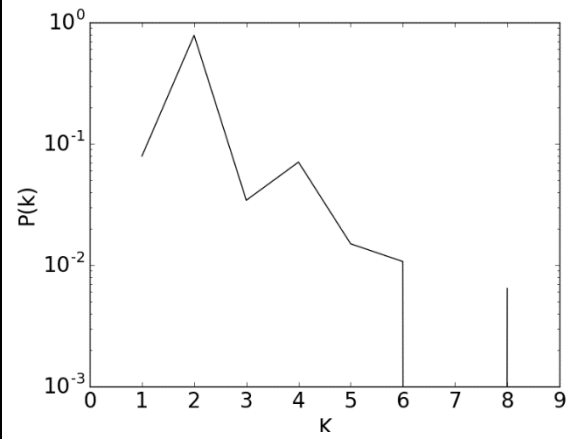
RATP (Paris) RER:



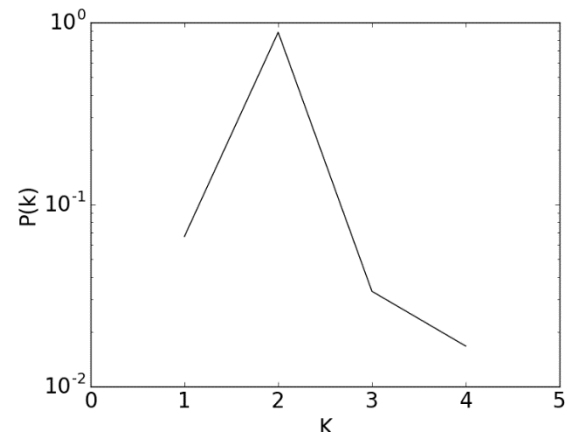
RATP (Paris) tram:



RATP (Paris) integrated:



Tyne and Wear metro:



Tyne and Wear metro with shortcuts:

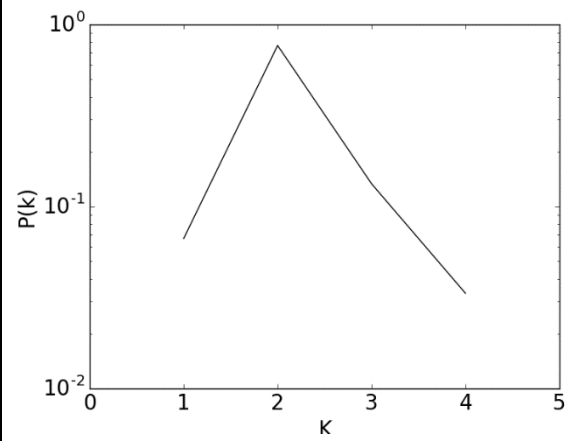
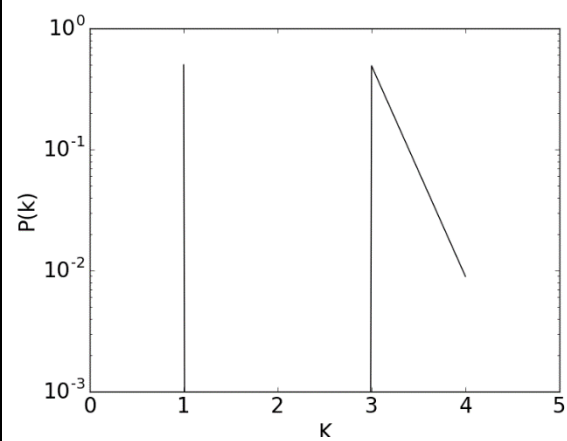
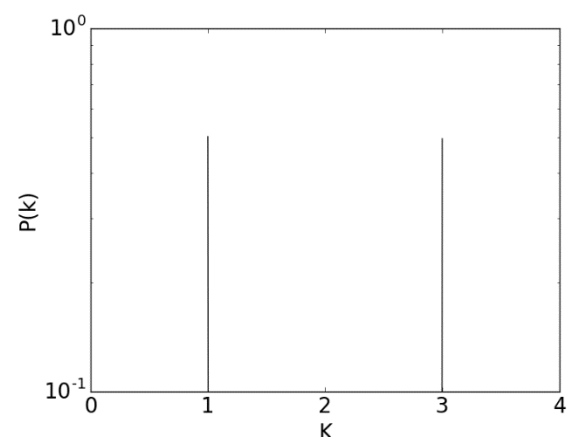


Figure E.5: Degree distribution plots to the light rail networks, Chapter 3, Section 3.6.4.

River Dee:



River Eden:



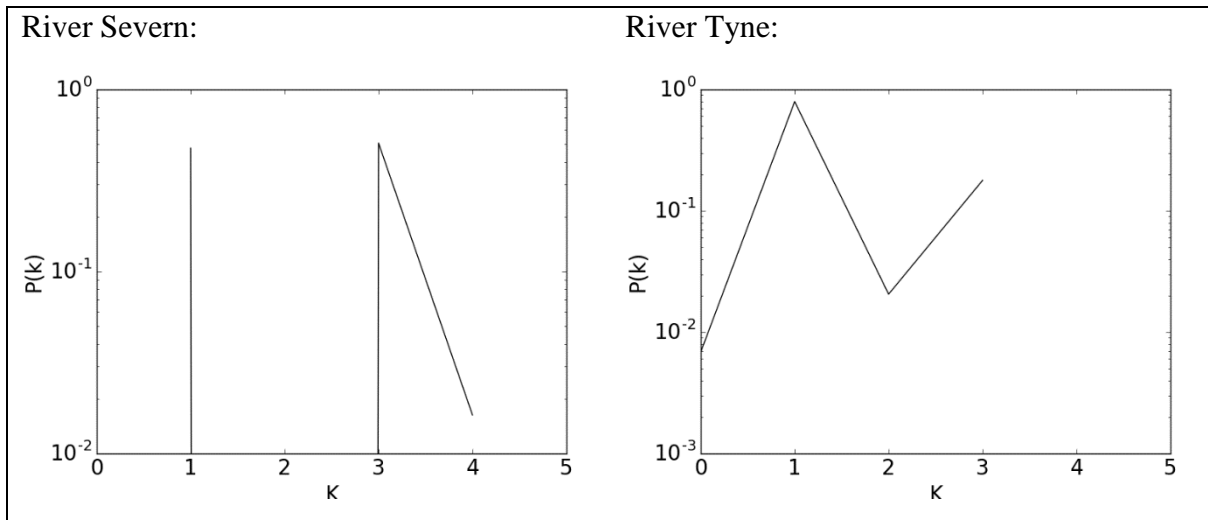


Figure E.6: Degree distribution plots for the suite of four river networks, Chapter 3, Section 3.6.5.

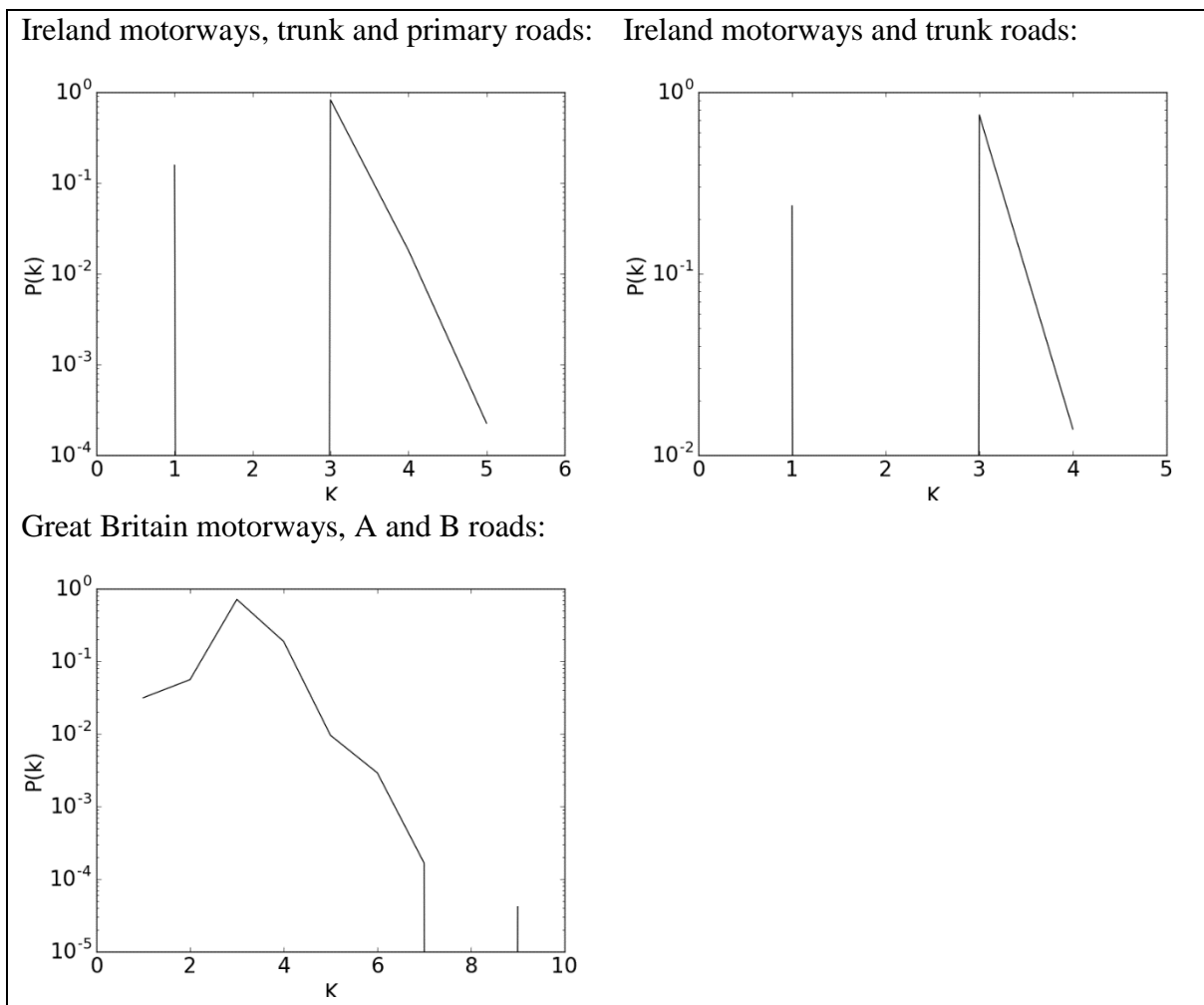
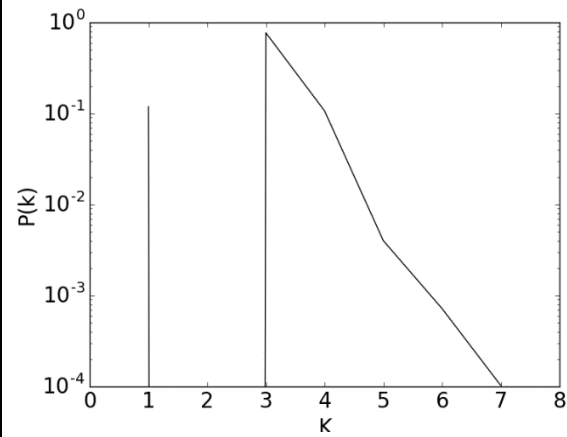
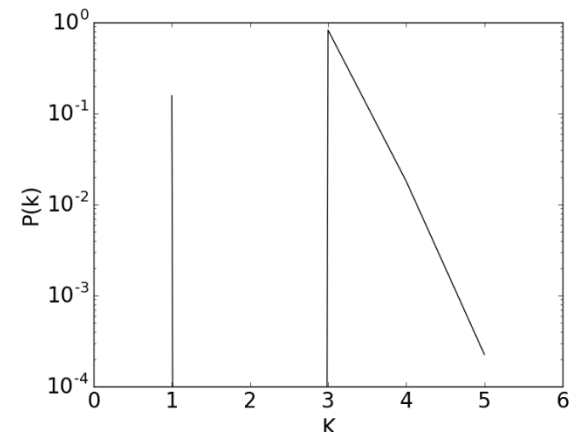


Figure E.7: Degree distribution plots for the suite of national scale road networks, Chapter 3, Section 3.6.6.

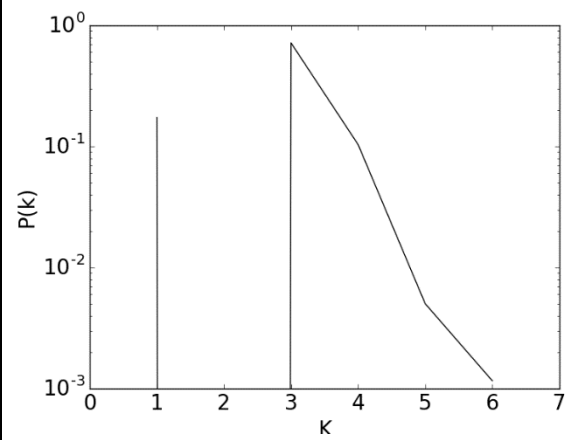
Leeds motorways, A, B and Minor roads:



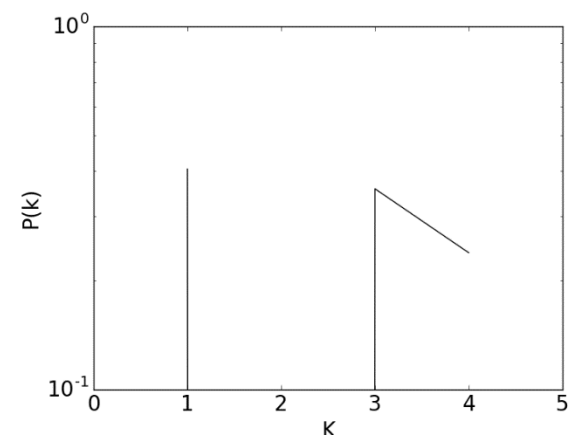
Leeds motorways, A and B roads:



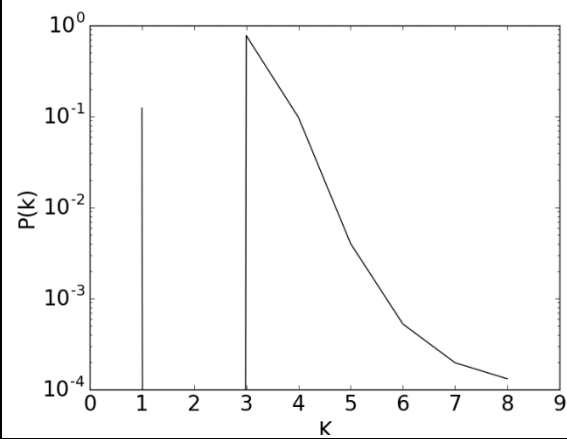
Milton-Keynes motorways, A, B and minor roads:



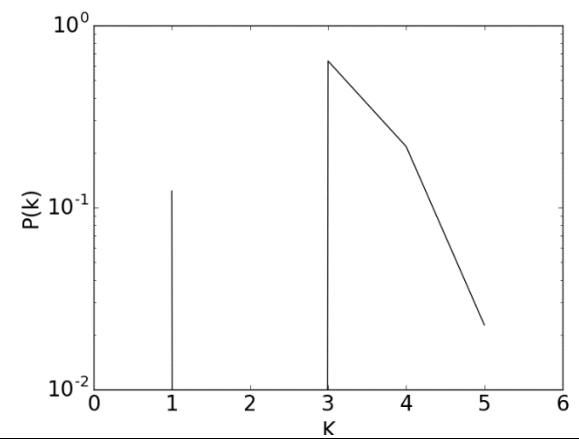
Milton-Keynes motorways, A and B roads:



Tyne and Wear motorways, A, B and minor roads:



Tyne and Wear motorways, A and B roads:



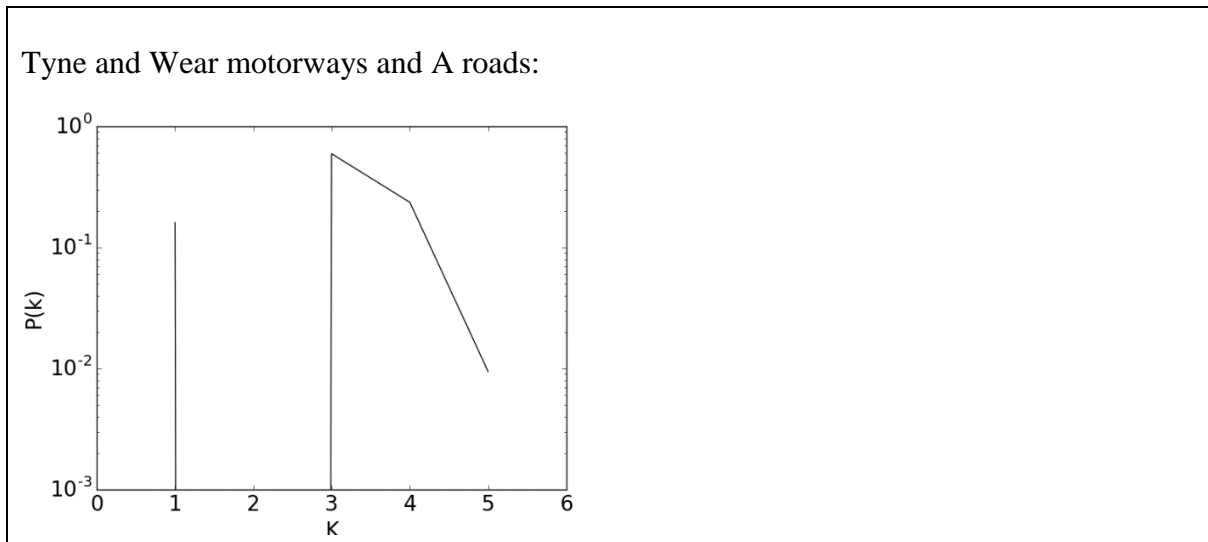


Figure E.8: Degree distribution plots for the regional/city scale road networks, Chapter 3, Section 3.6.6.

E.2 Metric values for the critical spatial infrastructure networks

To characterise the structure of the networks three selected metrics (Chapter 3, Section 3.4 (page 53)), the assortativity coefficient (AC), the maximum betweenness centrality (MBC) and the number of cycle basis per node (CB), have been calculated for each network. The same metrics were calculated for the synthetic graphs allowing the infrastructure networks to be compared to these through the use of single standard deviation ellipses, computed over the full analysed suite (Table D.1, page 267), for each of the eight graph models in the plots for the results. Results are presented for the infrastructure networks in the following sections for each group of infrastructure networks. Section E.2.1 gives the results for the suite of air networks while Section E.2.2 presents the results for the communication network. The results for the energy networks are presented in Section E.2.3. Sections E.2.4 and E.2.5 present the results for the national rail networks and regional networks respectively. The results for the river networks are given in Section E.2.6, with Sections E.2.7 and E.2.8 presenting the results for the national road networks and the regional road networks.

E.2.1 Air networks

Table E.1 shows the metric values for each of the networks, with Figure E.9, Figure E.10 and Figure E.11 showing the metric values for the three different pair-wise combinations with the synthetic graph ellipses shown for reference purposes.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
British Airways	-0.42	0.82	0.38
EasyJet	-0.45	0.51	2.99
European	-0.11	0.11	7.93
North American	-0.19	0.24	3.24
UK	-0.30	0.30	1.83
USA	-0.30	0.35	3.68

Table E.1: The three calculated metric values for each of the six air networks.

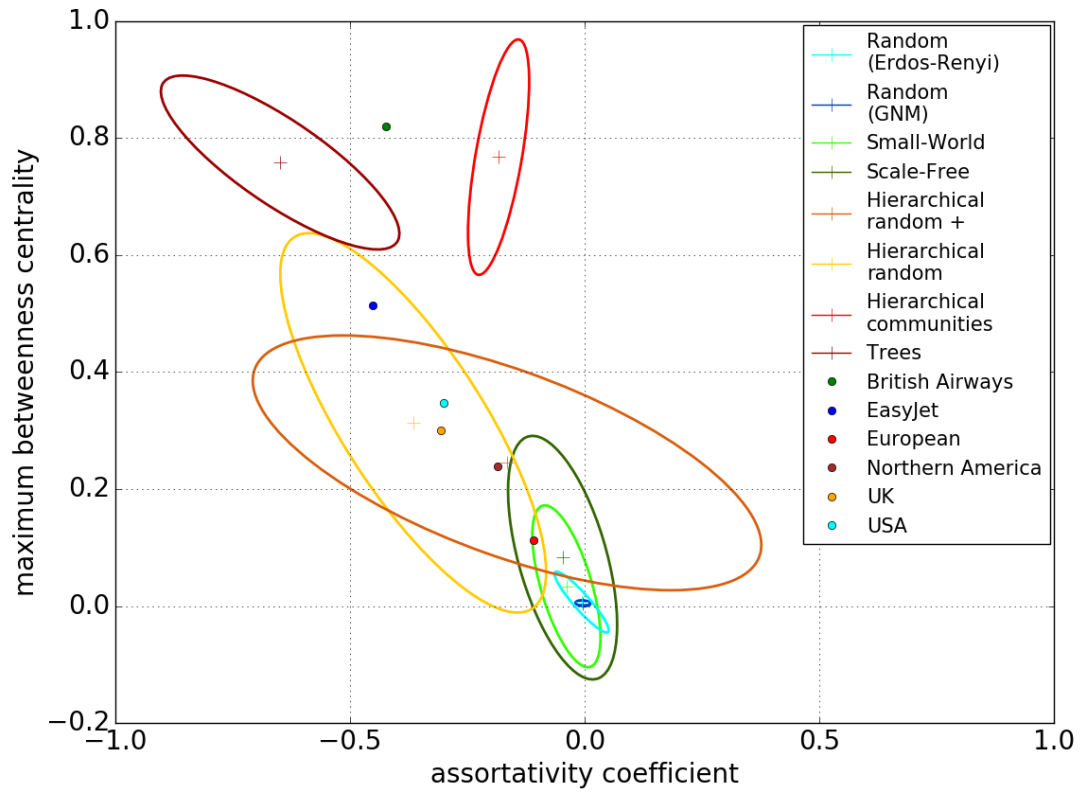


Figure E.9: Points for the air networks when plotted using the assortativity coefficient and the maximum betweenness centrality metric values.

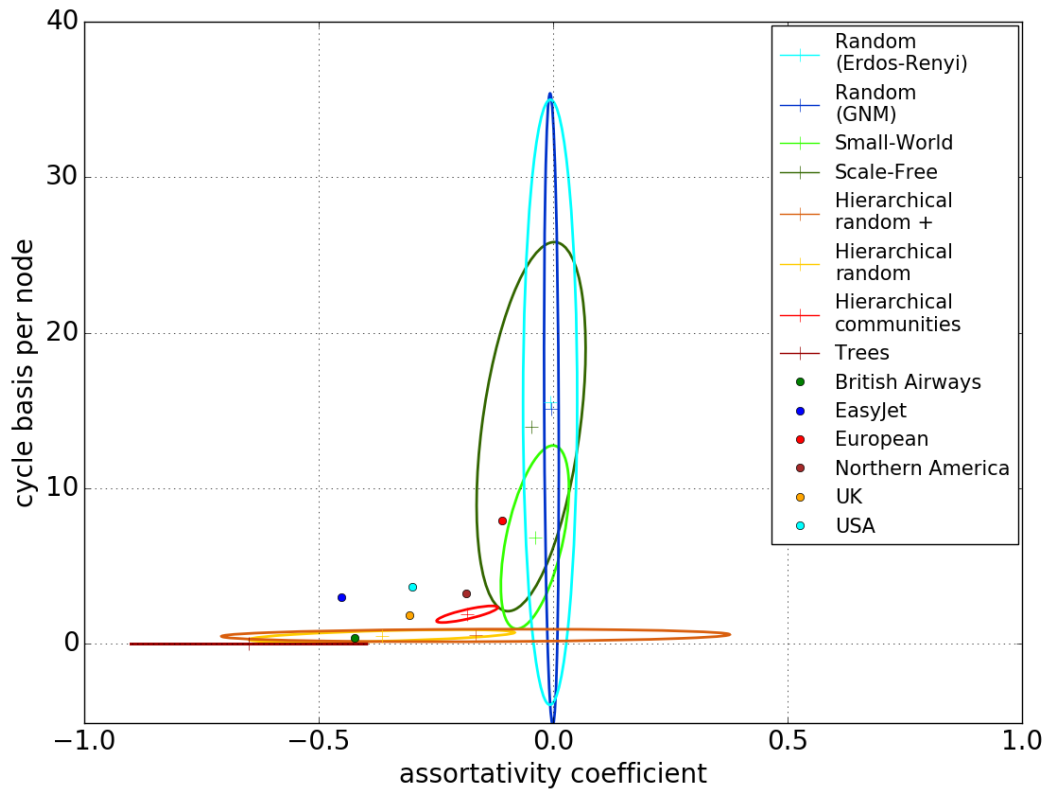


Figure E.10: Values for the air networks with the assortativity coefficient and number of cycle basis per node plotted.

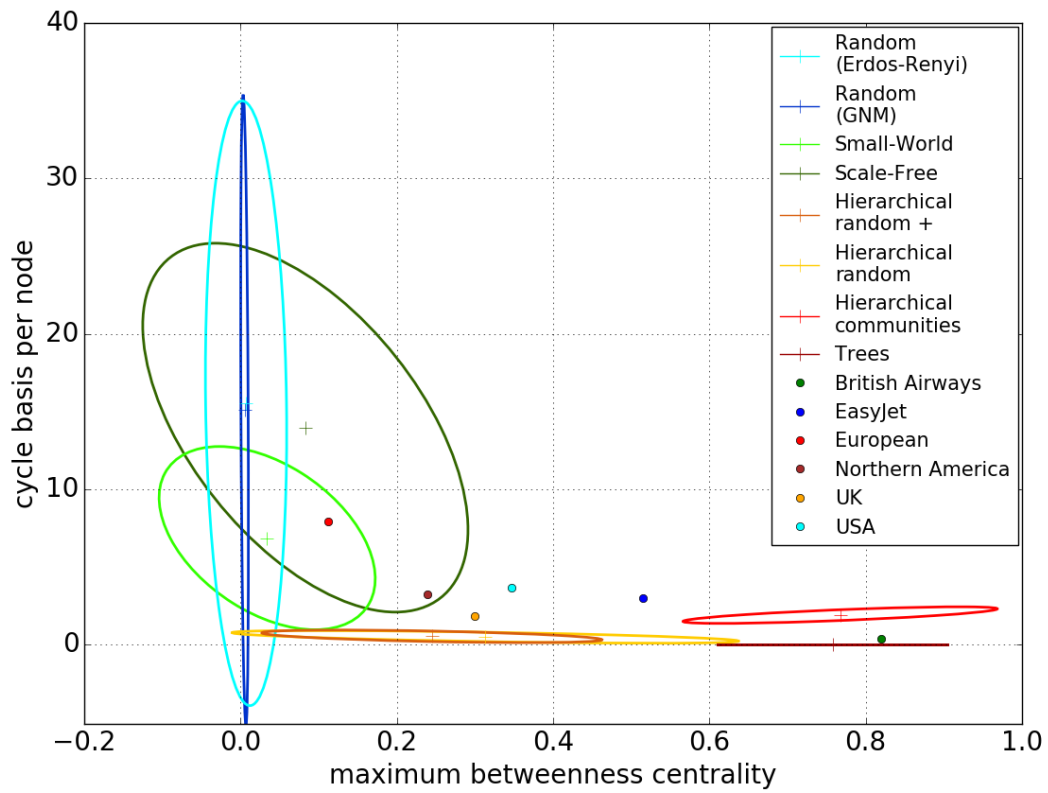


Figure E.11: Points for the air networks for the maximum betweenness centrality and the number of cycle basis per node.

E.2.2 Communications

The three metrics, AC, MBC and number of CB per node, have been calculate for the Janet network, Table E.2, and then compared to the values returned for the same metrics across the suite of synthetic graphs, Figure E.12, Figure E.13 and Figure E.14.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
Janet	-0.41	0.40	0.55

Table E.2: Metric values for the communication network.

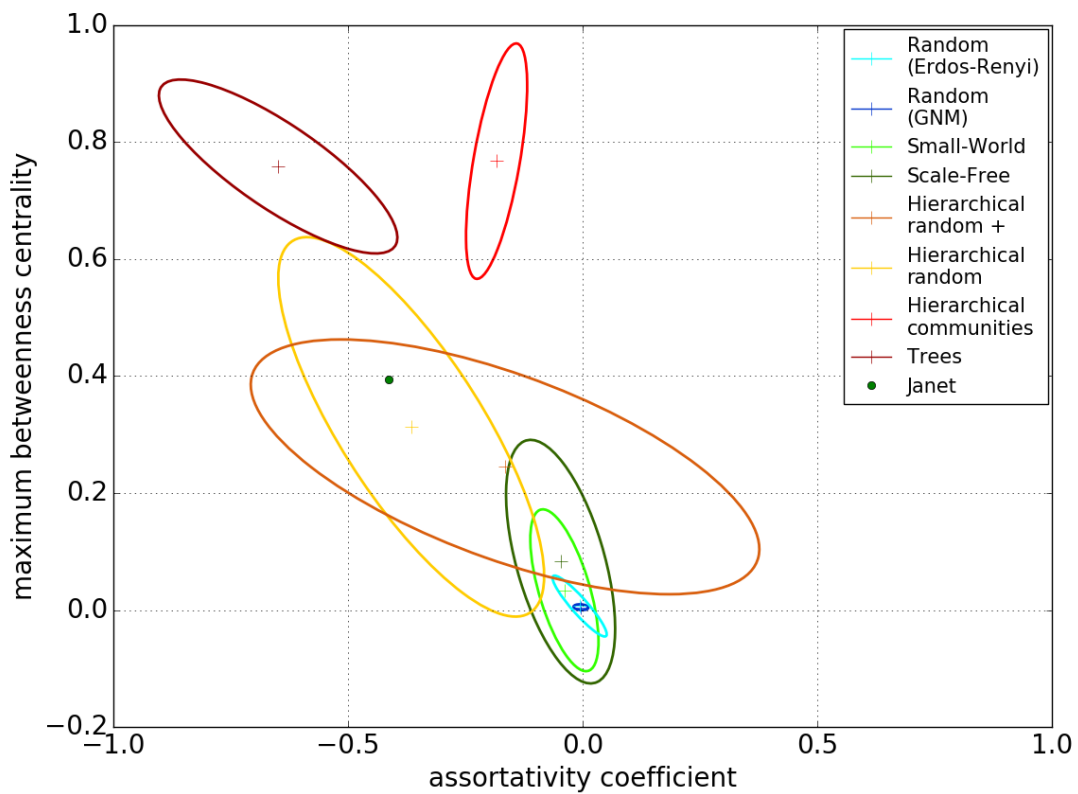


Figure E.12: Assortativity coefficient and maximum betweenness centrality for communications network.

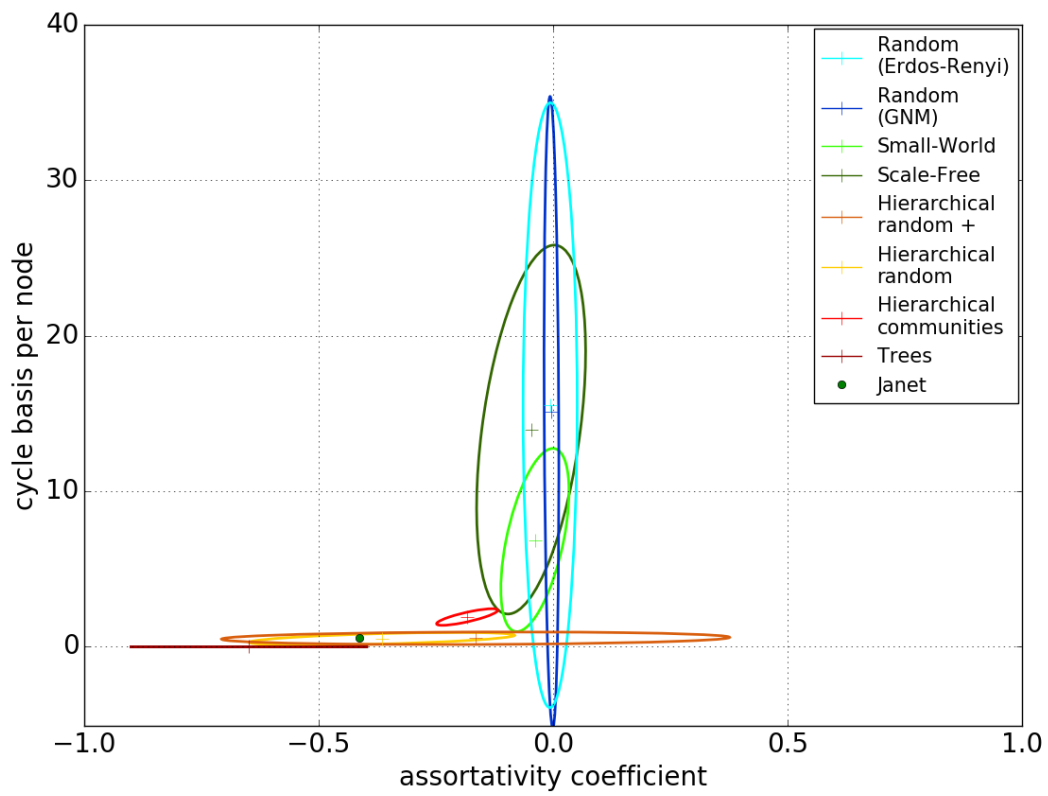


Figure E.13: Communications network result for the assortativity coefficient and number of cycle basis per node.

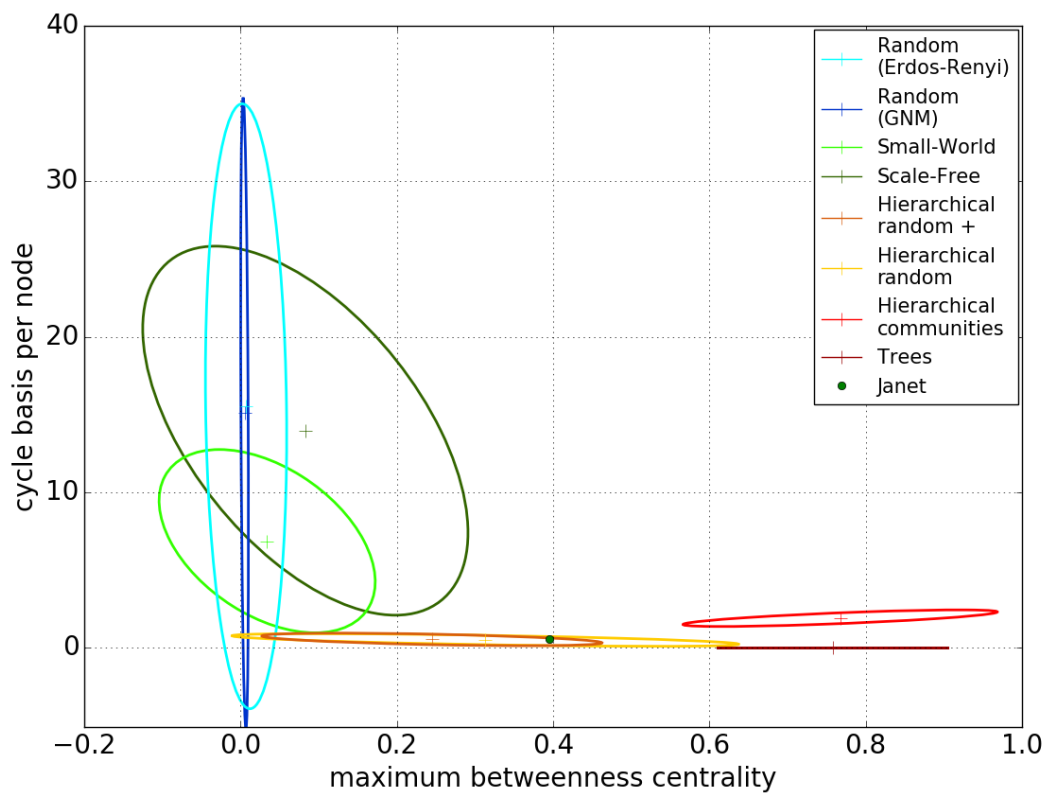


Figure E.14: Results for the maximum betweenness centrality and number of cycle basis per node for the communication network.

E.2.3 Energy

The three metric values have been calculated for the suite of energy networks, Table E.3, as with all other sets of networks allowing them to be compared against each other. Figure E.15, Figure E.16, Figure E.17 compare the metric values of the energy networks to those computed for the suite of eight graph models.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
National Grid electricity transmission	0.16	0.28	0.02
National Grid electricity transmission MT	-0.03	0.19	0.22
National Grid electricity transmission NT	0.07	0.20	0.06
National Grid gas transmission	-0.03	0.31	0.19
Electricity transmission network	-0.03	0.28	0.38

Table E.3: The computed graph metric values for the suite of energy networks.

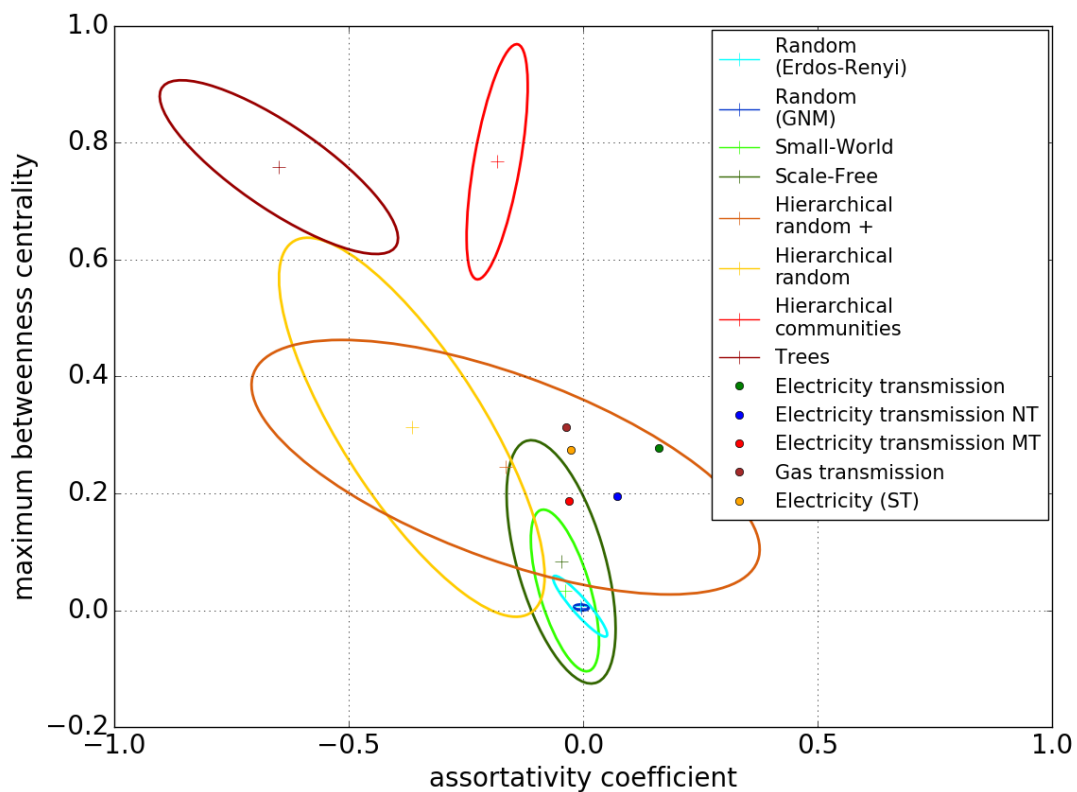


Figure E.15: For the suite of energy networks the results for the assortativity coefficient and the maximum betweenness centrality metrics.

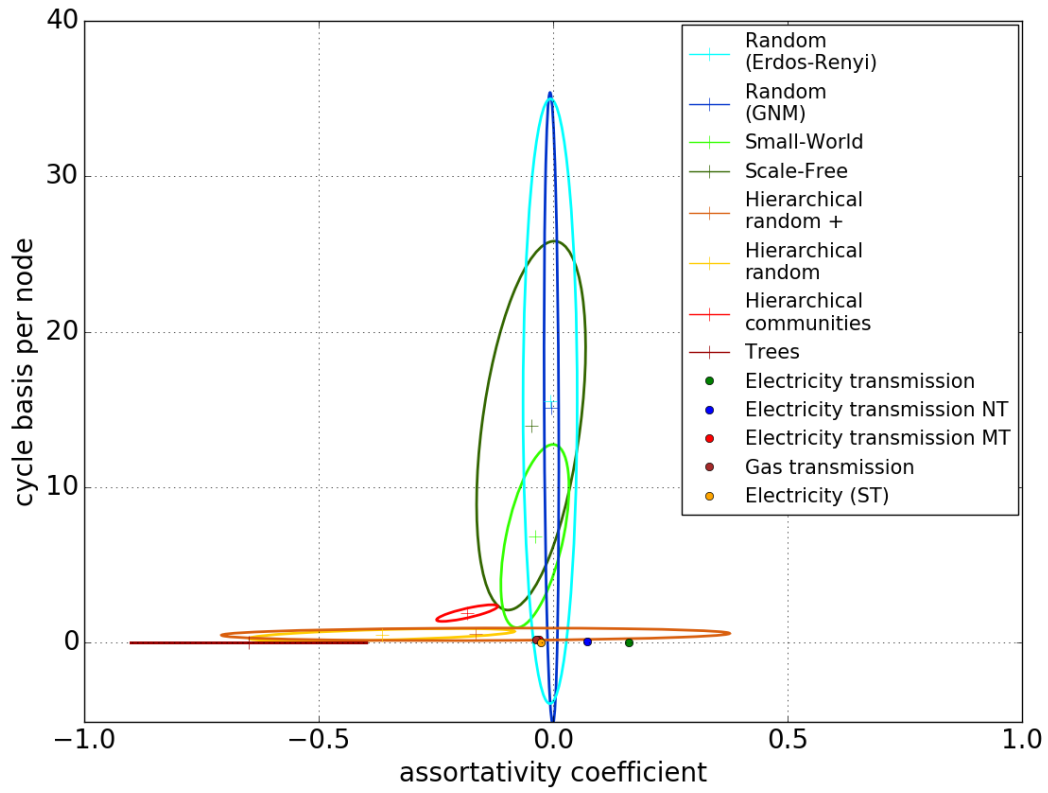


Figure E.16: Results for the assortativity coefficient and number of cycle basis per node metrics for the suite of energy networks.

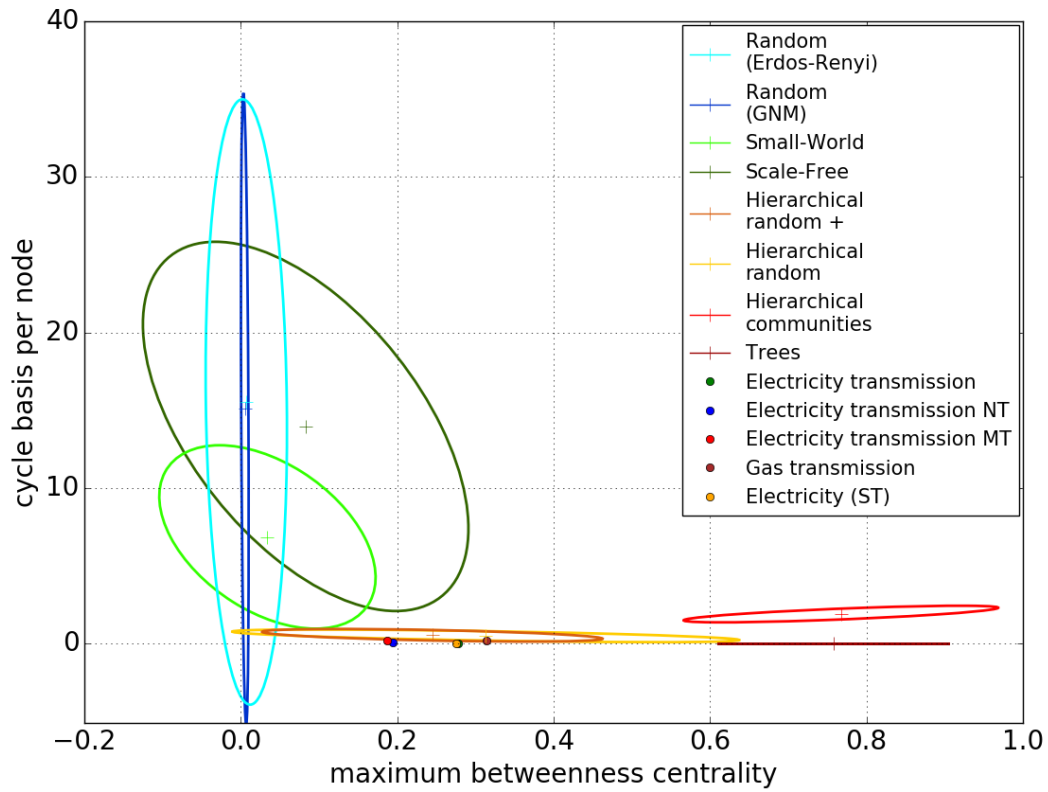


Figure E.17: Maximum betweenness centrality and number of cycle basis per node metric results for the suite of energy networks.

E.2.4 Rail – National

The computed metric values for the national rail networks are presented in Table E.4, with these values then compared to the same values returned for the suite of synthetic graphs, Figure E.18, Figure E.19 and Figure E.20.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
Great Britain rail	0.19	0.25	0.06
Ireland rail	-0.02	0.23	0.02
Ireland rail with shortcuts	0.21	0.53	0.04

Table E.4: The computed metric values for the networks within the suite of national rail networks.

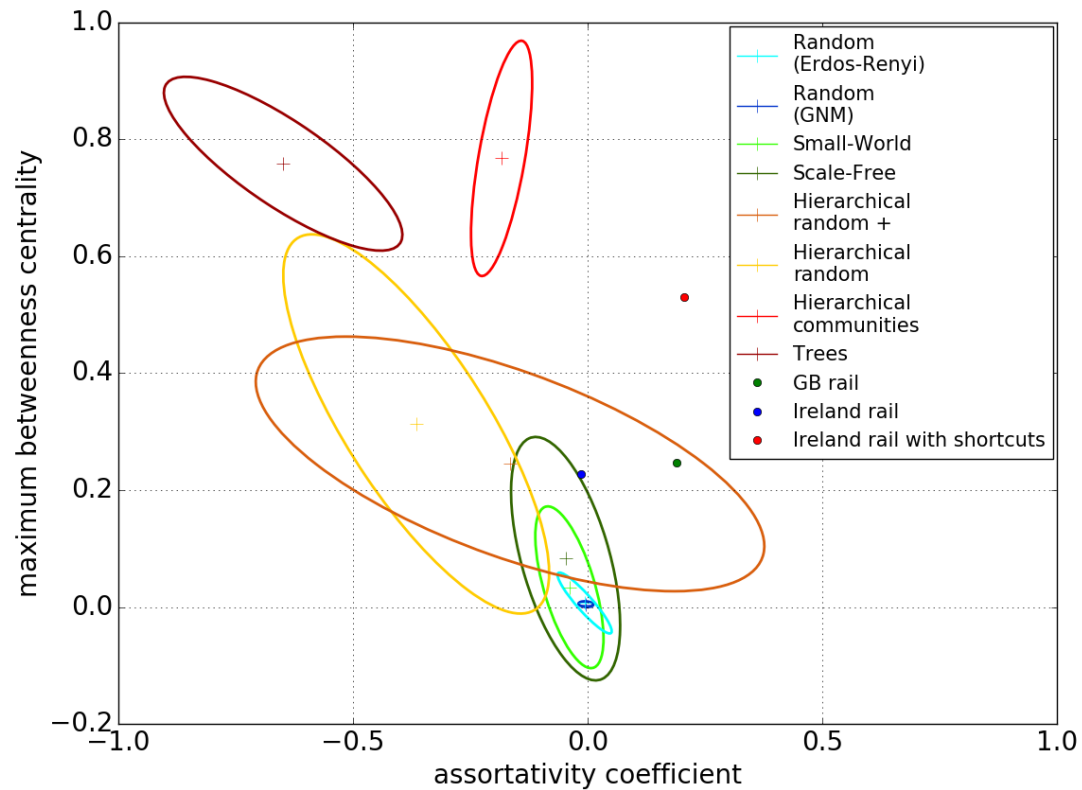


Figure E.18: The metric results for the suite of national rail networks for the assortativity coefficient and maximum betweenness centrality.

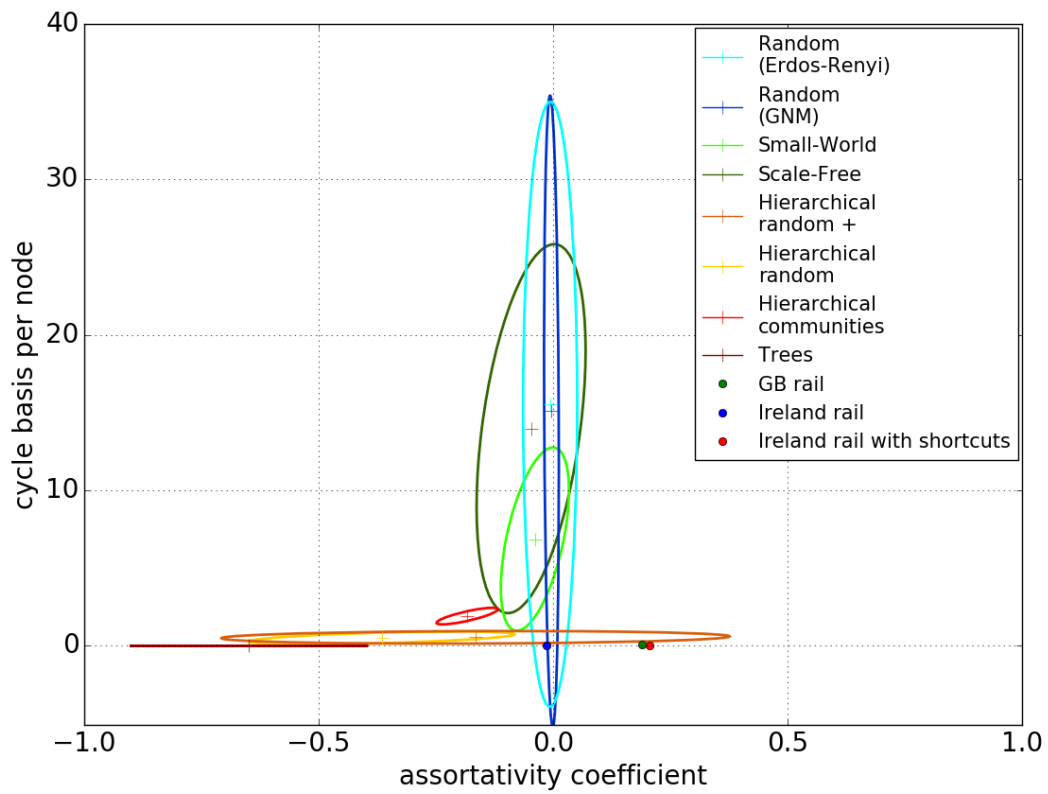


Figure E.19: For the suite of national rail networks the results for the assortativity coefficient and number of cycle basis per node metrics.

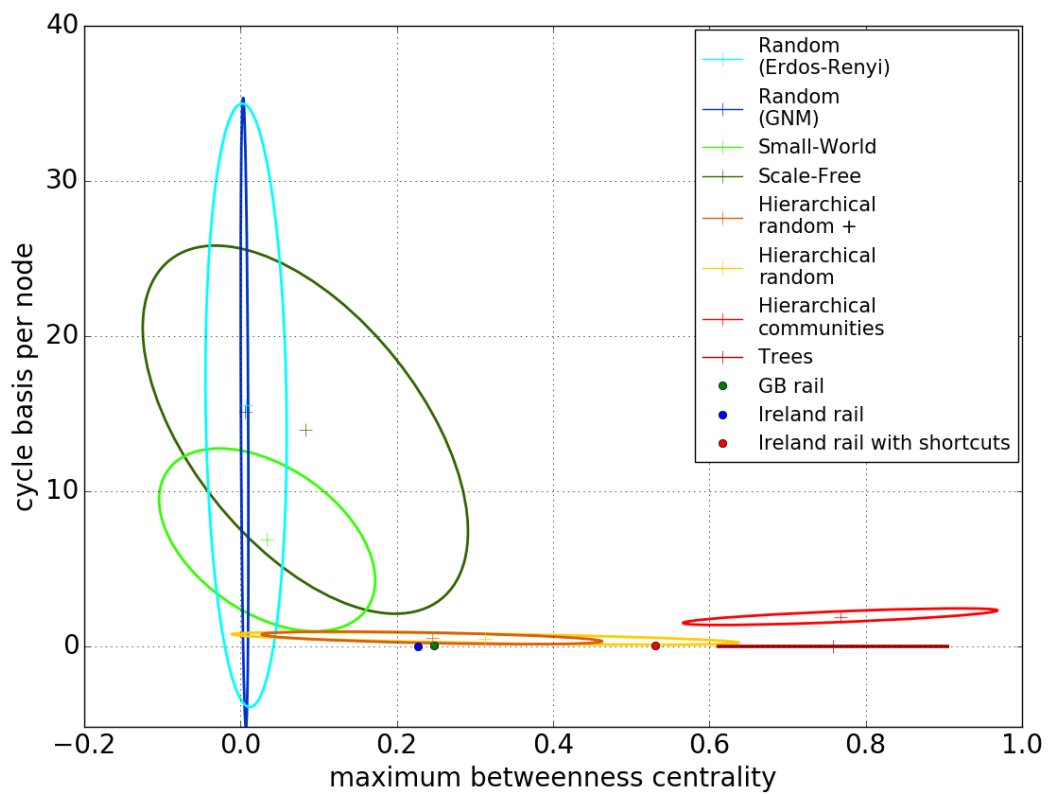


Figure E.20: Results for the suite of national rail networks for the maximum betweenness centrality metric and the number of cycle basis per node.

E.2.5 Rail – Regional (light rail)

The metric values for the three computed metrics, AC, MBC and number of CB per node, for each of the light rail networks have been computed (Table E.5) and also plotted against each other alongside the single standard deviation ellipses for each of the eight synthetic graph models, Figure E.21, Figure E.22 and Figure E.23.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
Boston subway	0.32	0.59	0.02
Boston subway with TAPAN	0.32	0.58	0.02
London DLR	-0.03	0.55	0.04
London light rail	0.10	0.26	0.14
London Overground	-0.16	0.63	0.00
London tube	0.11	0.42	0.13
Manchester Metrolink	0.27	0.56	0.03
RATP rail	0.06	0.20	0.12
RATP metro	-0.03	0.34	0.19
RATP RER	-0.16	0.65	0.00
RATP tram	-0.04	0.06	0.00
Tyne and Wear metro	-0.19	0.59	0.02
Tyne and Wear metro with shortcuts	0.16	0.55	0.08

Table E.5: Values for the three calculated metrics for the regional rail (light rail) networks.

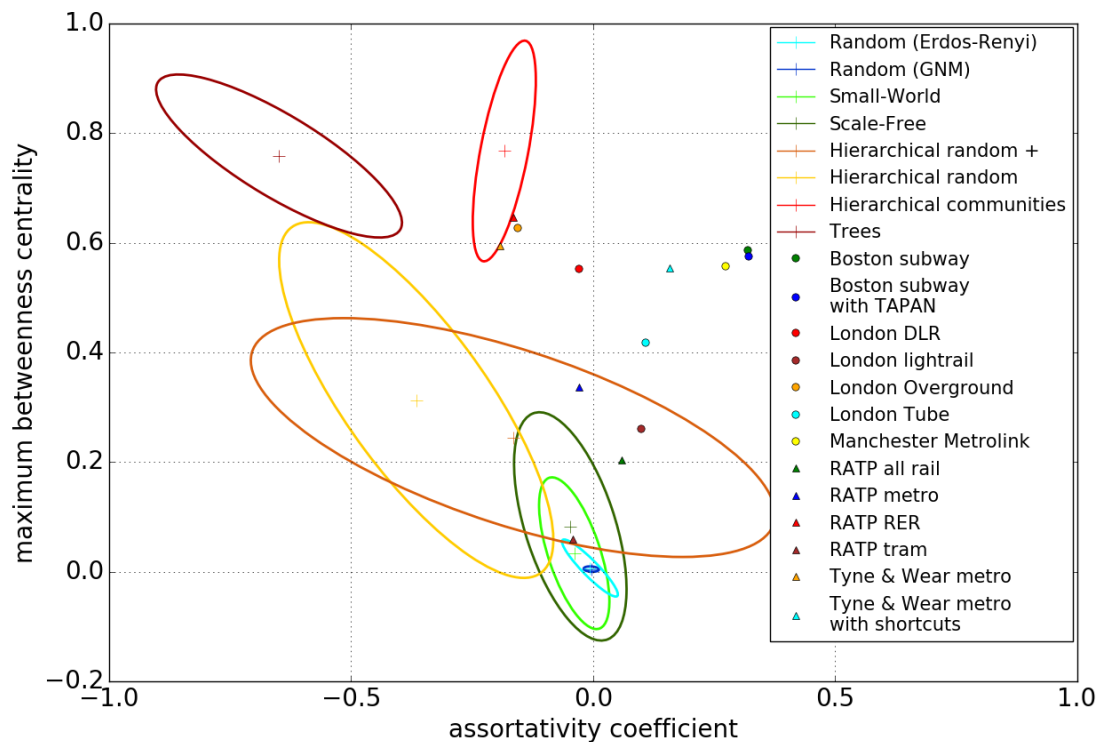


Figure E.21: Results for the suite of regional rail networks for the assortativity coefficient and maximum betweenness centrality metrics.

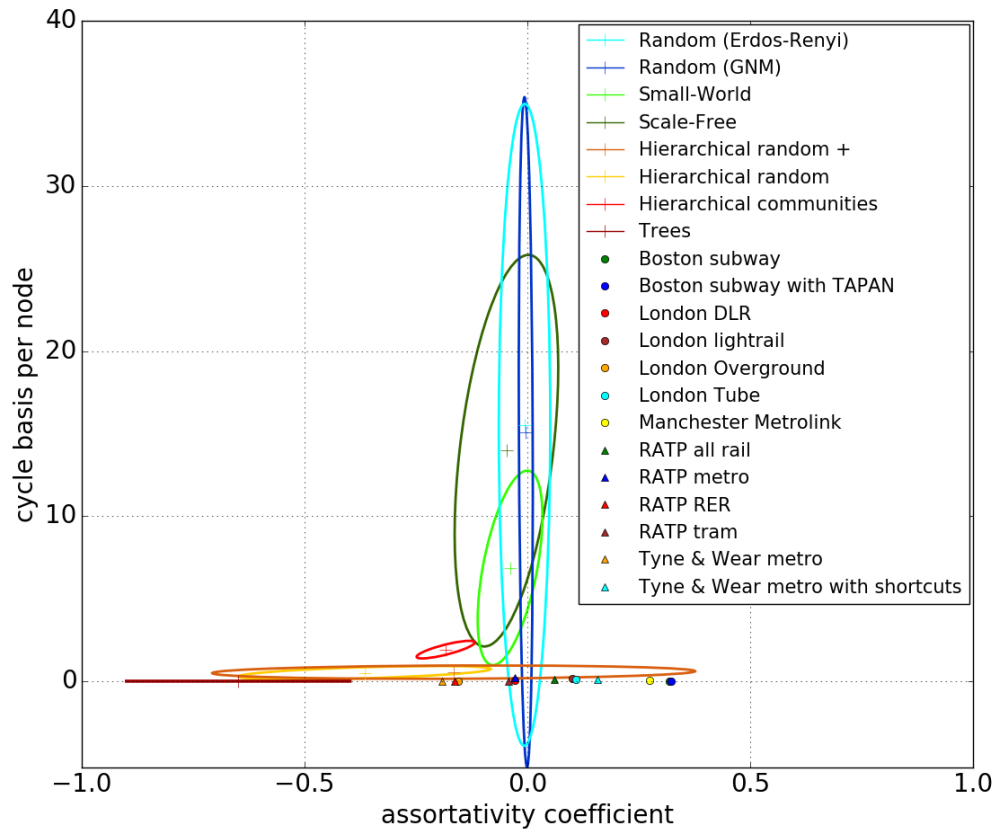


Figure E.22: Assortativity coefficient and number of cycle basis per node results for the suite of regional rail networks.

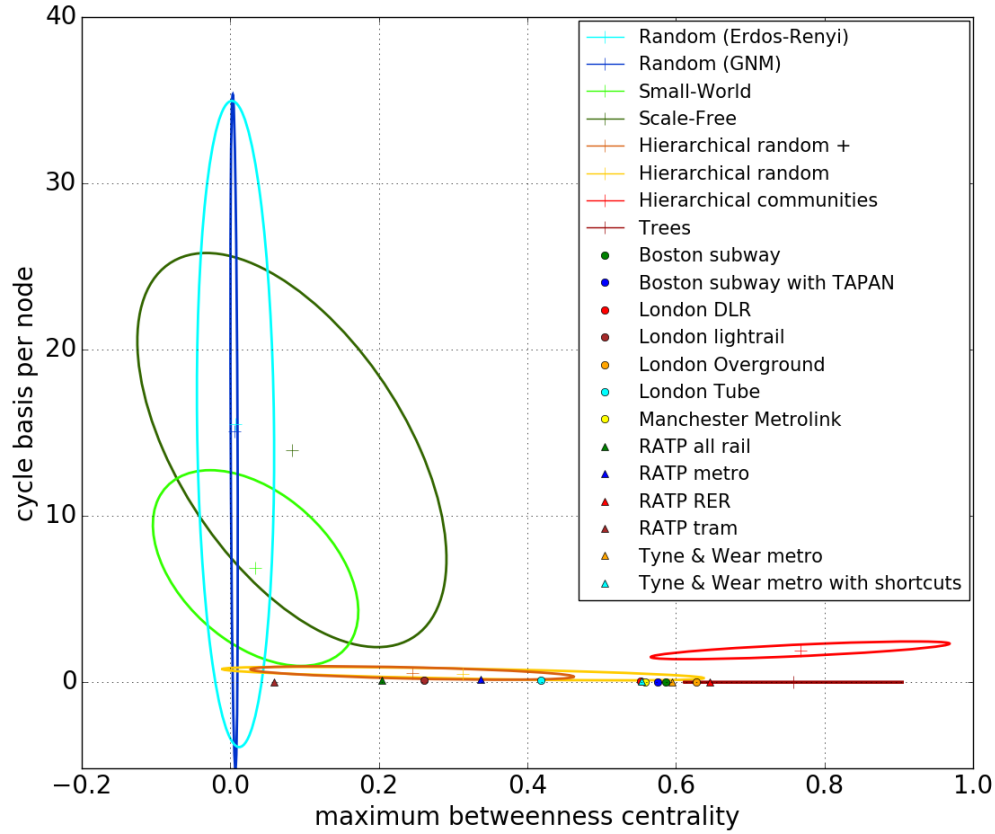


Figure E.23: For the suite of regional rail networks the metric results for the maximum betweenness centrality and number of cycle basis per node.

E.2.6 Rivers

The three selected metrics, the AC, MBC and number of CB per node, have been computed for each of the four river networks, Table E.6, with the values plotted against each other alongside the single standard deviation ellipses for each of the synthetic graph models, Figure E.24, Figure E.25 and Figure E.26, allowing the values for the river networks to be compared to the range of different graph models.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
Dee	0.00	0.00	0.01
Eden	0.00	0.00	0.00
Severn	-0.28	0.53	0.03
Tyne	-0.34	0.66	0.00

Table E.6: Calculated metric values for the suite of river networks.

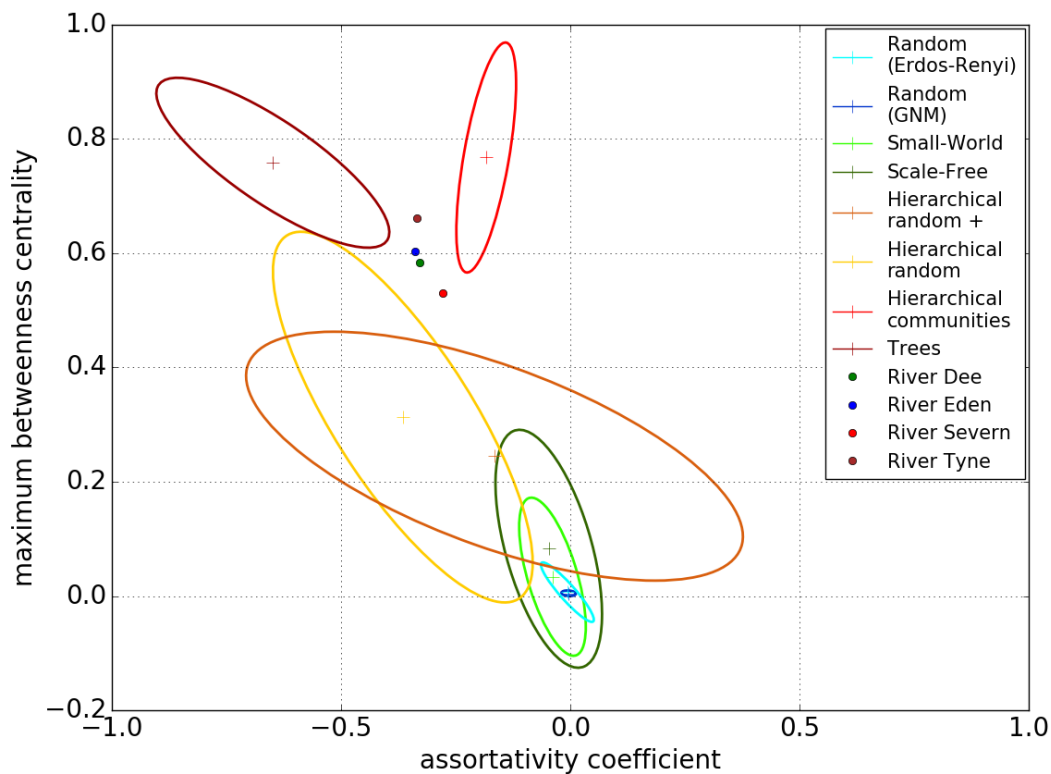


Figure E.24: For the suite of river networks the results for the assortativity coefficient and the maximum betweenness centrality metrics.

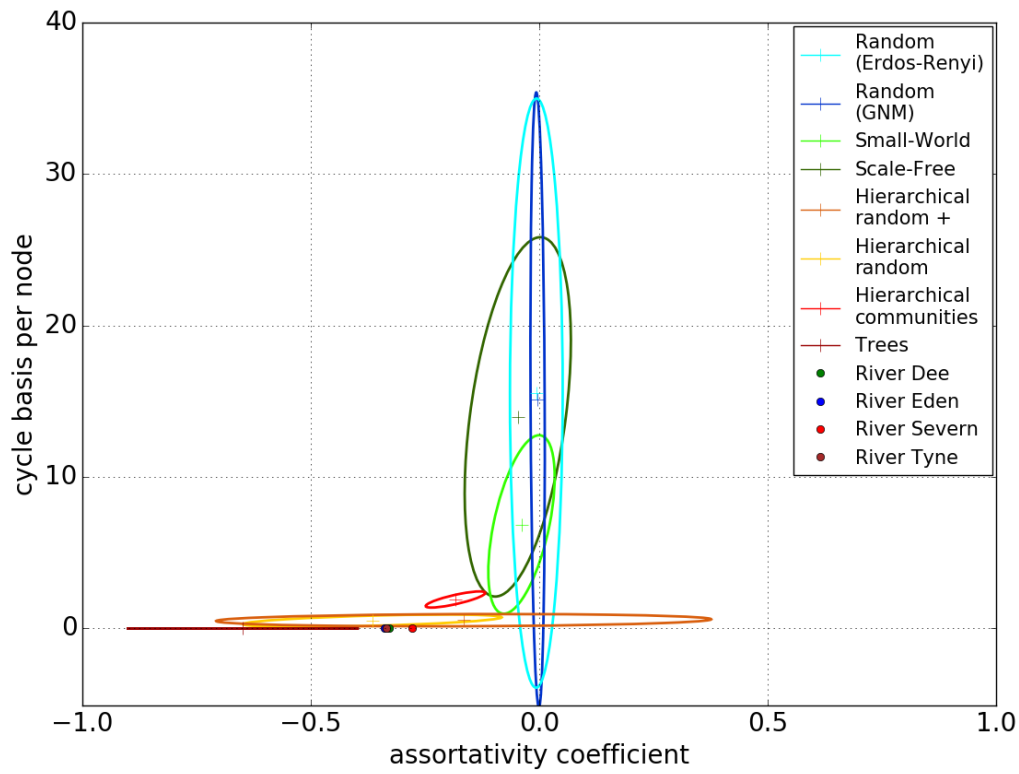


Figure E.25: The values for the assortativity coefficient and the number of cycles basis per node for the suite of river networks.

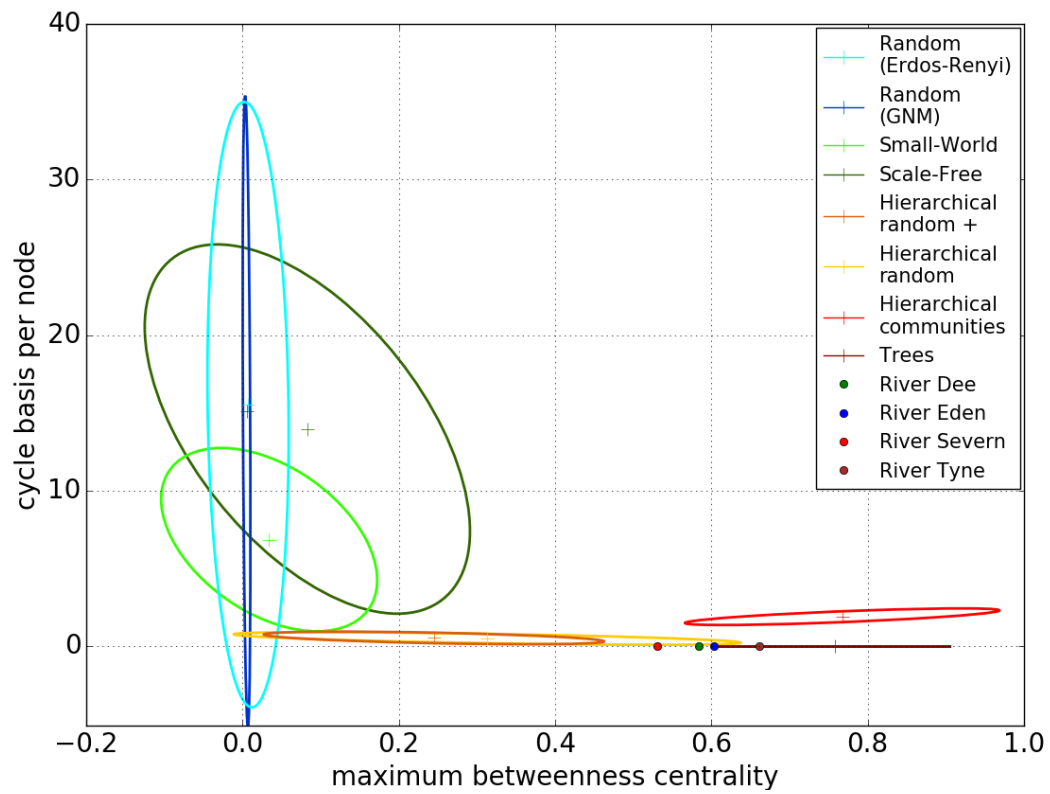


Figure E.26: Results for the maximum betweenness centrality and the number of cycle basis per node for the suite of river networks.

E.2.7 Roads – National

The suite of national road networks consists of three networks, with the values for the selected metrics for the suite are presented in Table E.7. These values for each of the networks are then compared to the eight different graph models in Figure E.27, Figure E.28 and Figure E.29 using the single standard deviation of the metric values for the synthetic models.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
Great Britain motorways, A and B roads	0.12	0.27	0.55
Ireland motorways, trunk and primary roads	-0.03	0.34	0.35
Ireland motorways and trunk roads	-0.08	0.27	0.27

Table E.7: The values for the three calculated metrics for the national road networks.

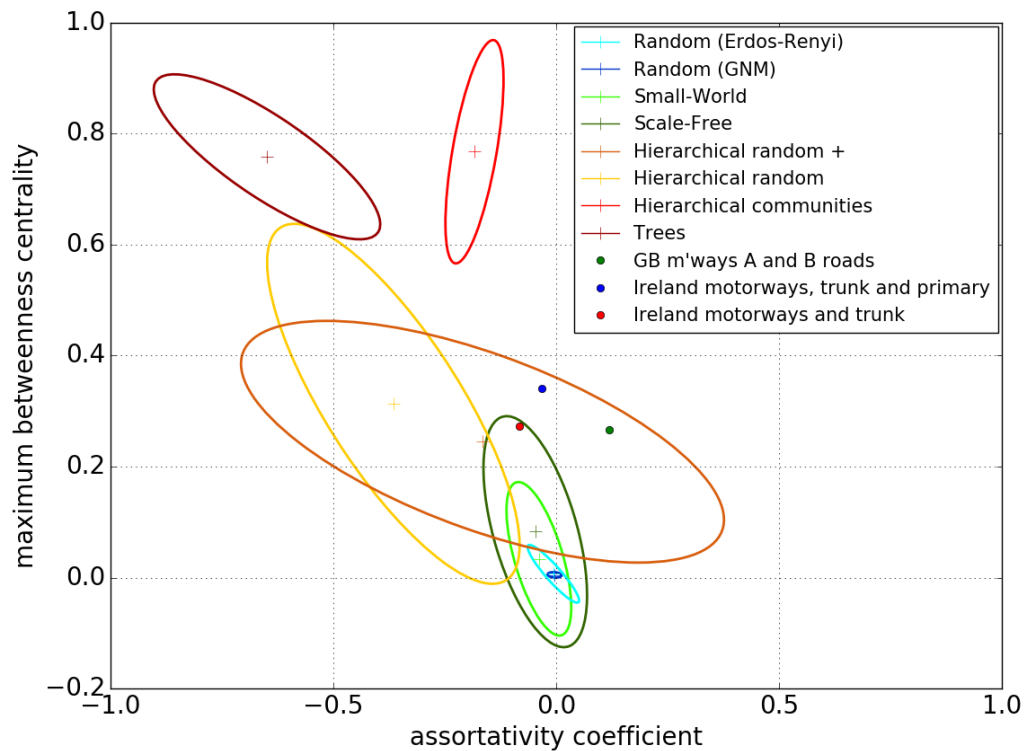


Figure E.27: The calculated metric results for the suite of national road networks for the assortativity coefficient and maximum betweenness centrality metrics.

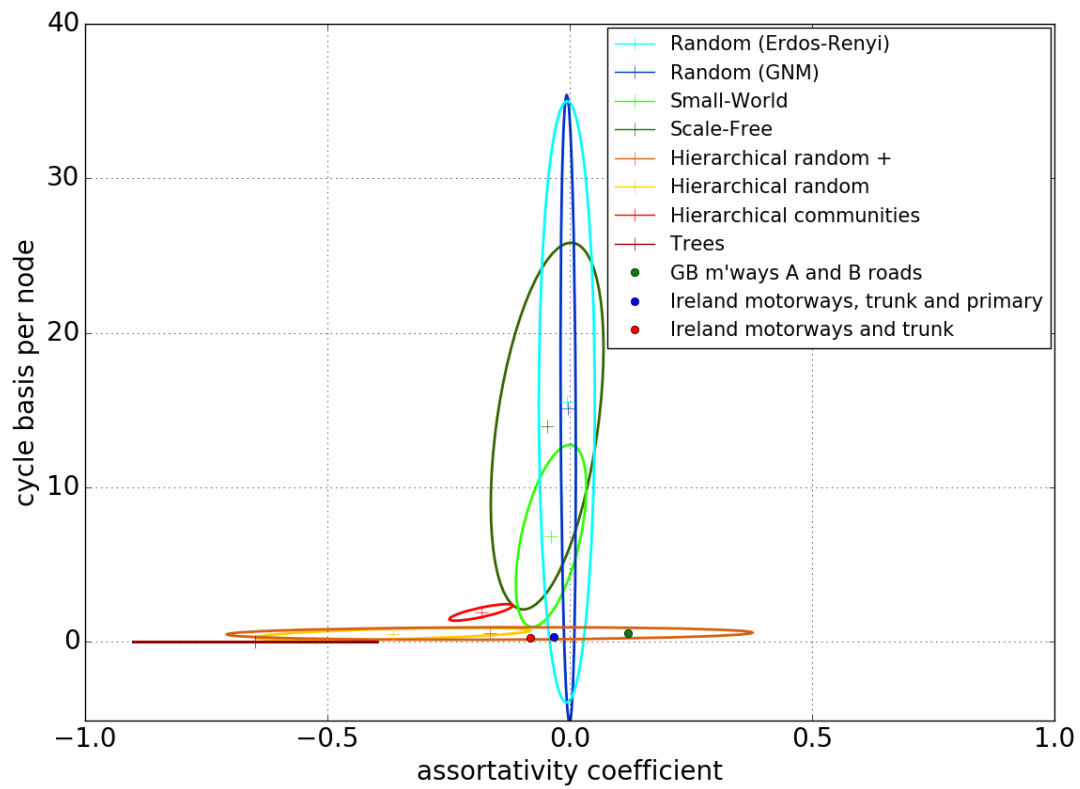


Figure E.28: Assortativity coefficient and number of cycle basis per node metric results for the suite of national road networks.

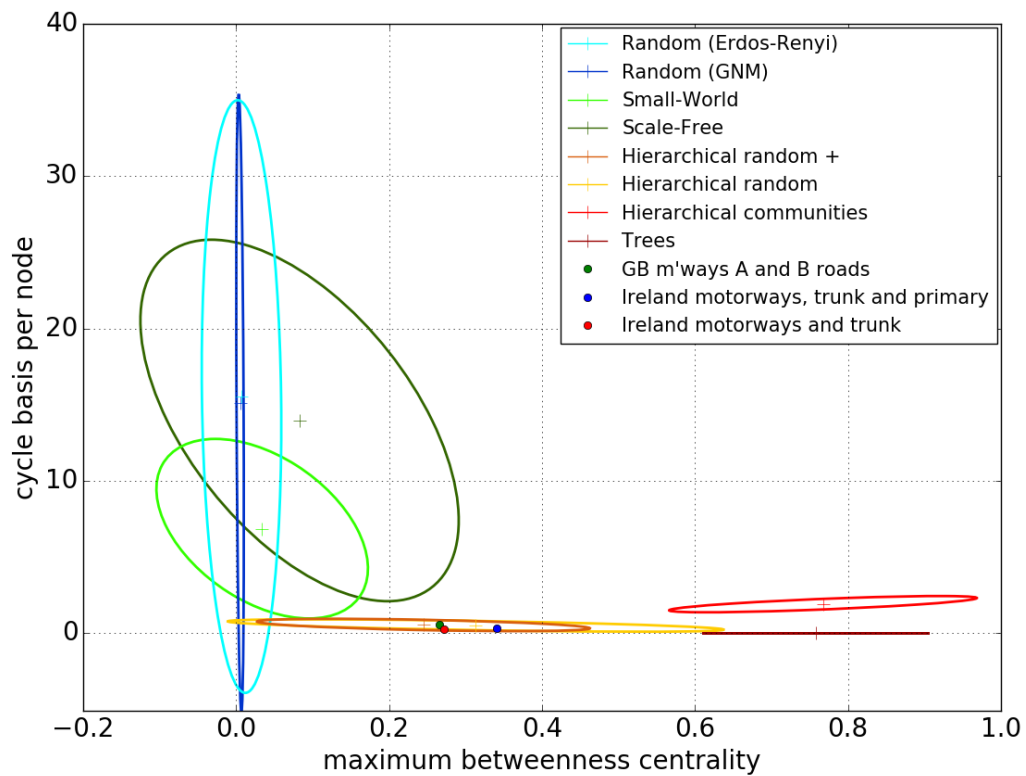


Figure E.29: Results for the maximum betweenness centrality and number of cycle basis per node for the suite of national road networks.

E.2.8 Roads – Regional

Severn networks of regional roads are included in the Table E.8 which shows the values for the computed metrics for each network. These values are then used to compare each network to the eight graph models for the synthetic graphs, Figure E.30, Figure E.31 and Figure E.32.

Infrastructure network	Assortativity coefficient	Maximum betweenness centrality	Cycle basis per node
Leeds motorways, A, B and minor roads	0.05	0.23	0.44
Leeds motorways, A and B roads	-0.04	0.23	0.46
Milton Keynes motorways, A, B and minor roads	0.09	0.36	0.39
Milton Keynes motorways, A and B roads	-0.12	0.42	0.24
Tyne and Wear motorways, A, B and minor roads	0.06	0.33	0.43
Tyne and Wear motorways, A and B roads	0.09	0.25	0.51
Tyne and Wear motorways and A roads	0.03	0.32	0.47

Table E.8: The three calculated metrics for the suite of regional road networks.

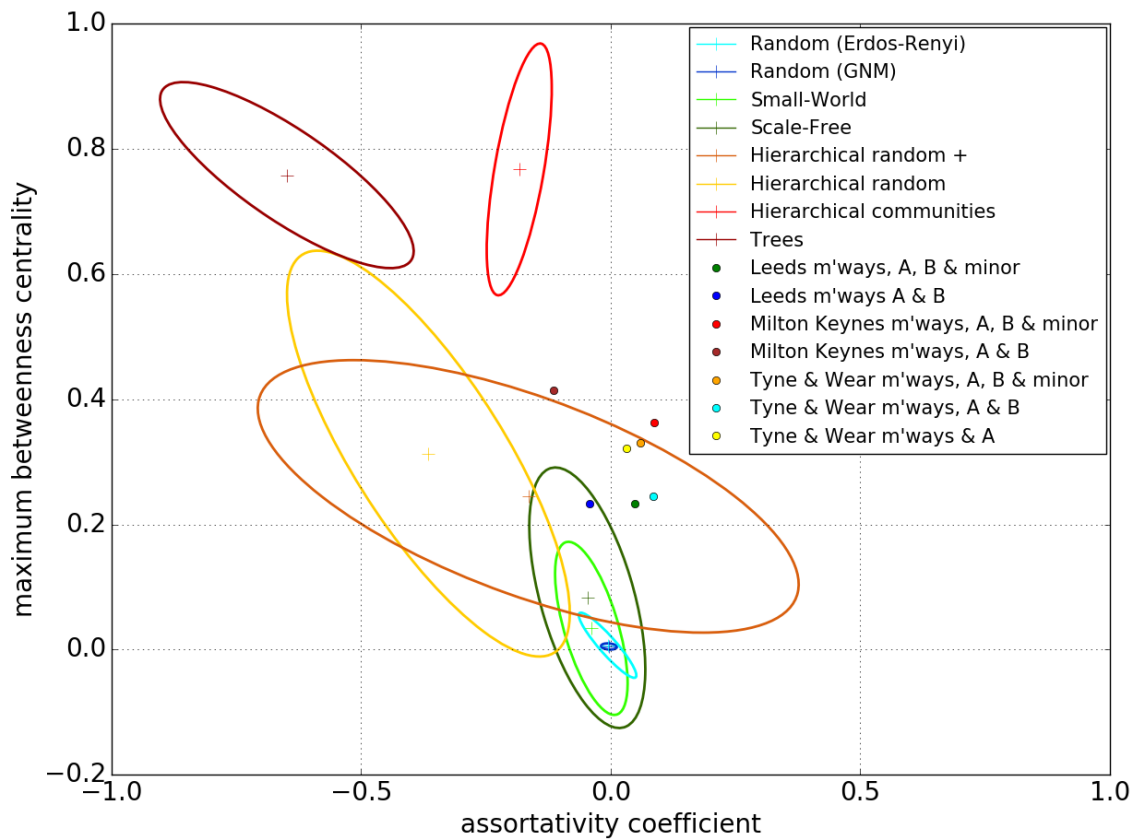


Figure E.30: Values for the suite of regional road networks for the assortativity coefficient and maximum betweenness centrality metrics.

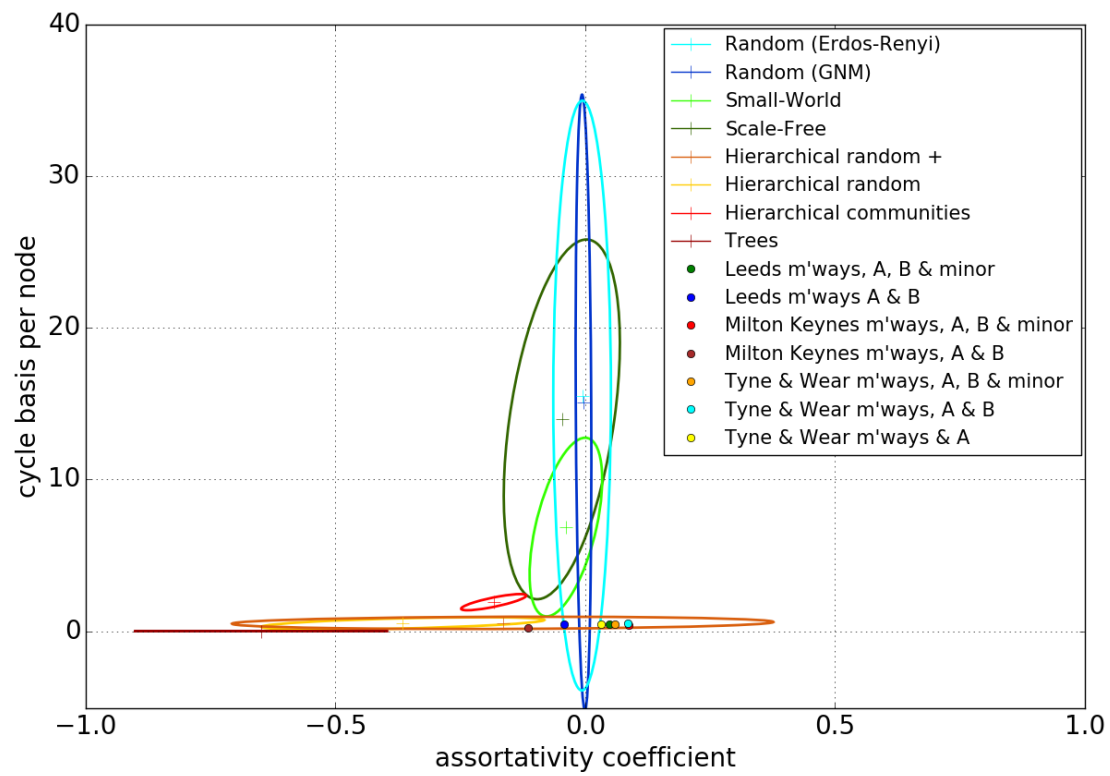


Figure E.31: Results for the assortativity coefficient and number of cycle basis per node for the suite of regional road networks.

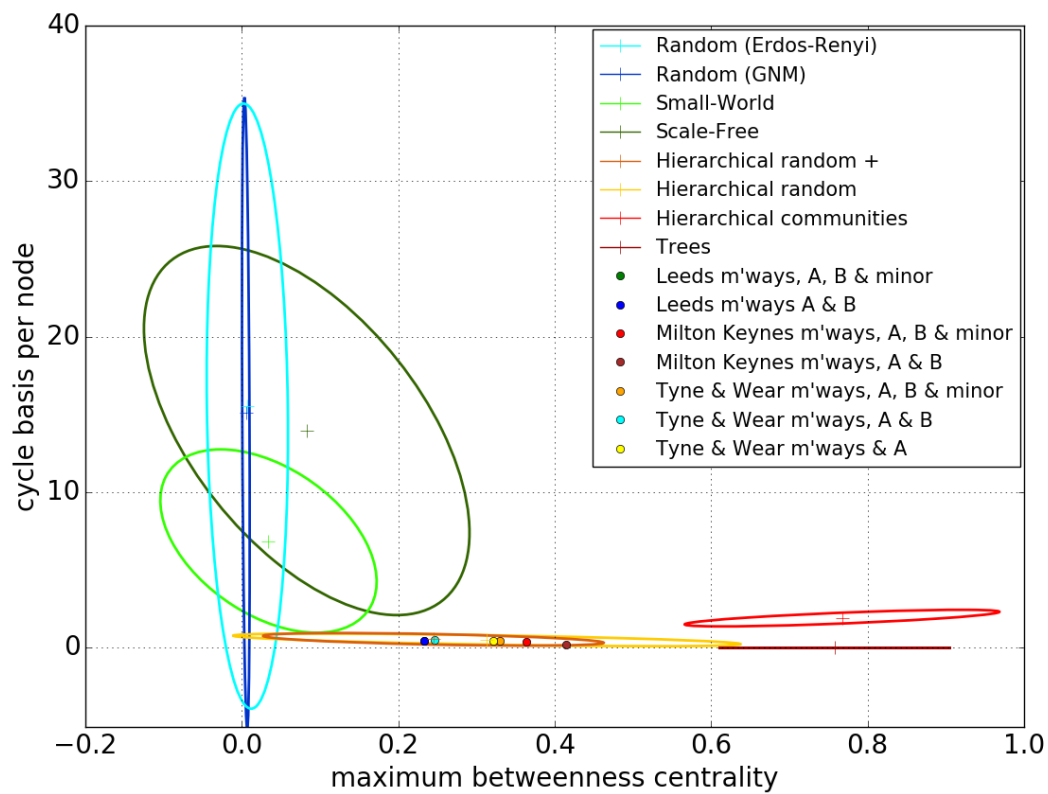
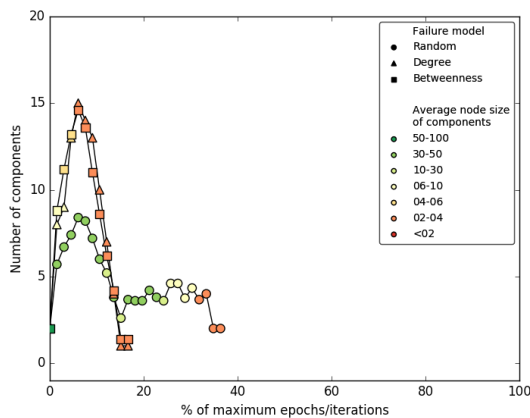


Figure E.32: Values for the suite of regional road networks for the maximum betweenness centrality and number of cycle basis per node values.

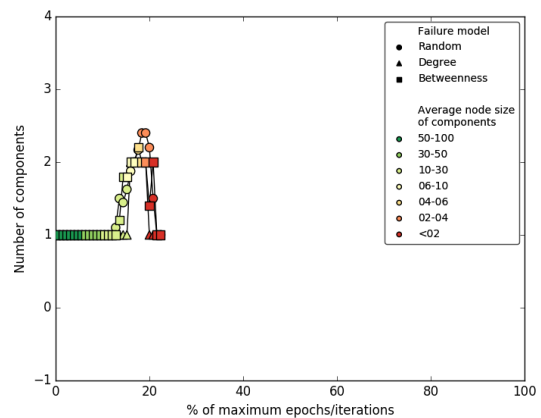
E.3 Robustness of critical spatial infrastructure networks

Each of the spatial infrastructure networks has been analysed for its robustness to three topological based failure methods (Chapter 3, Section 3.5 (page 59)). The results for the suite of infrastructure networks have been split into the respective infrastructure groups, with those for the air networks presented in Figure E.33. Each plot shows, for the three failure methods, random, node degree and maximum betweenness centrality, the number of components in the network on the x-axis as the percentage of nodes removed (y-axis) increases, with the symbols indicating the average size of the components at the selected intervals. Figure E.34 (page 320) shows the results for the communication network, Janet, and Figure E.35 (page 321) presented the failure behaviour results for the suite of energy networks. The behaviour of the rail networks to the failures are presented for the national rail networks in Figure E.36 (page 322) and for the regional rail networks in Figure E.37 (page 324). Results for the river networks are presented in Figure E.38 (page 325), followed by those for the road networks, Figure E.39 (page 325) and Figure E.40 (page 327) for the national and regional networks respectively.

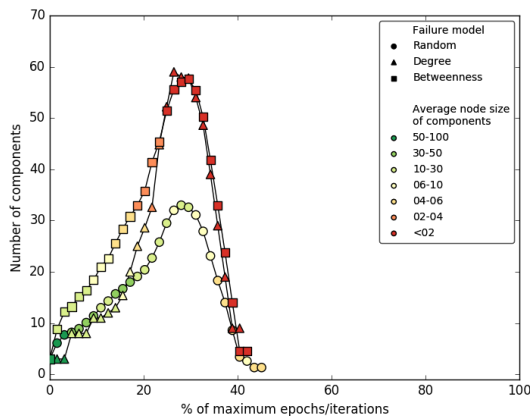
British Airways flights:



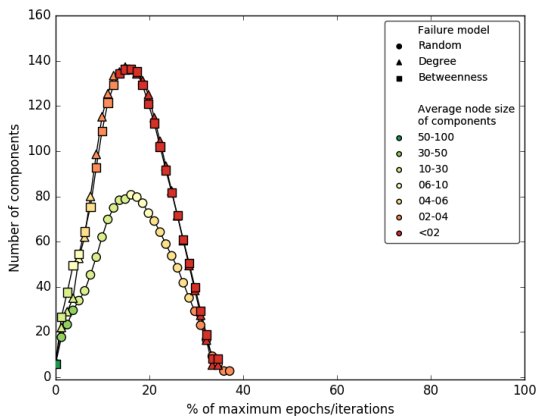
EasyJet flights:



European flights:



North American flights:



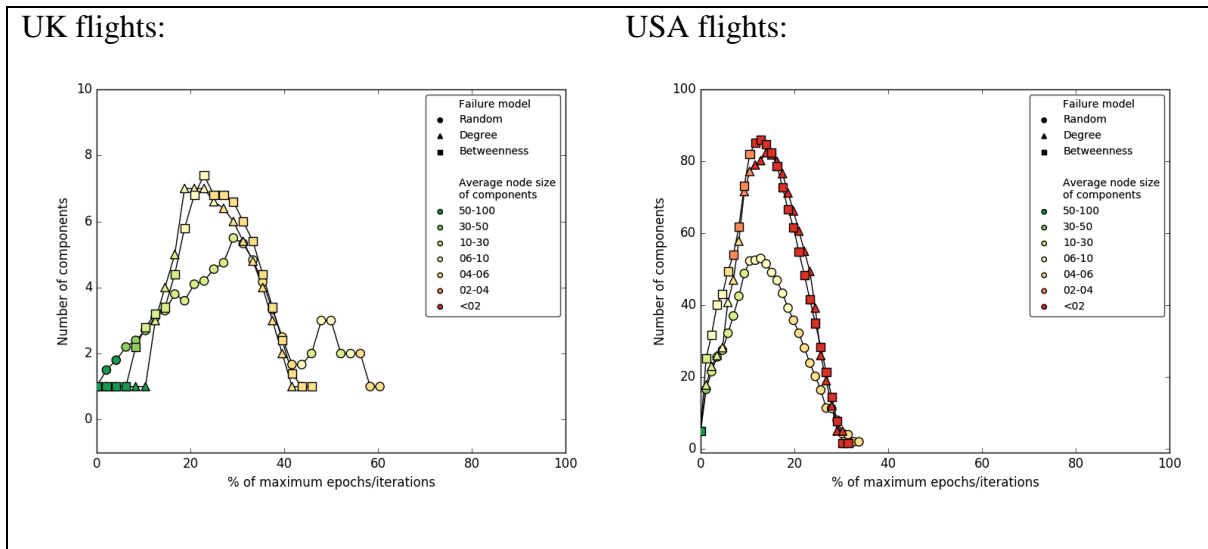


Figure E.33: Failure behaviour plots for the suite of six air networks.

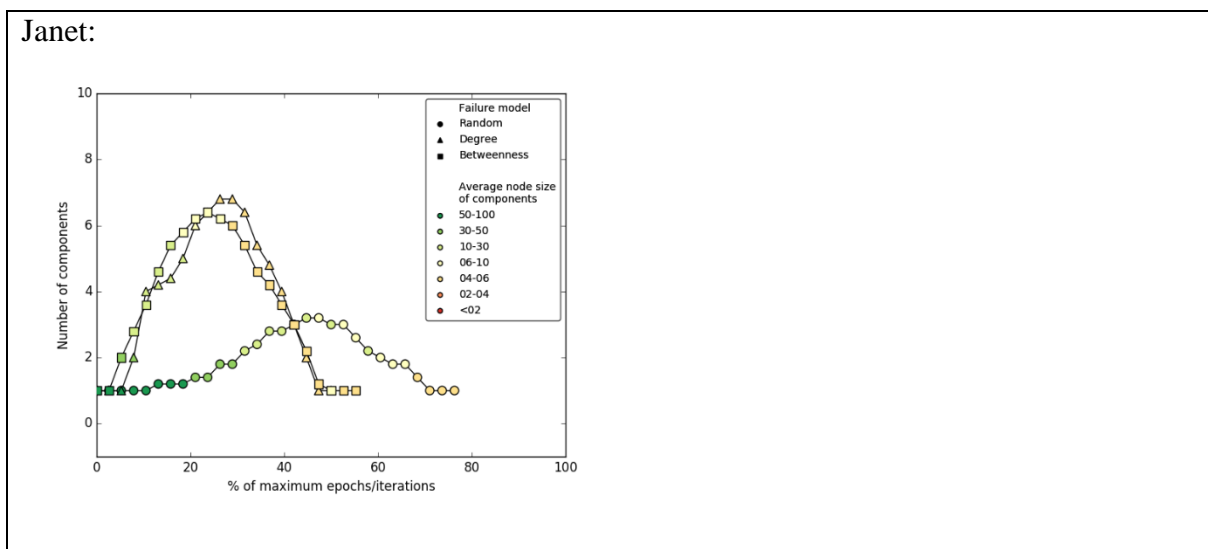
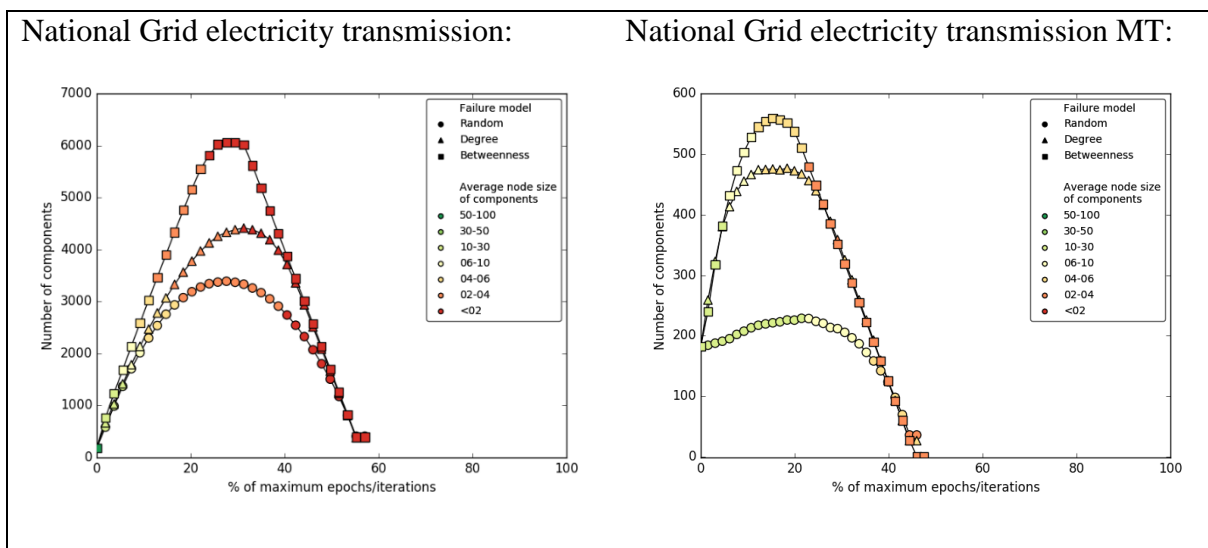
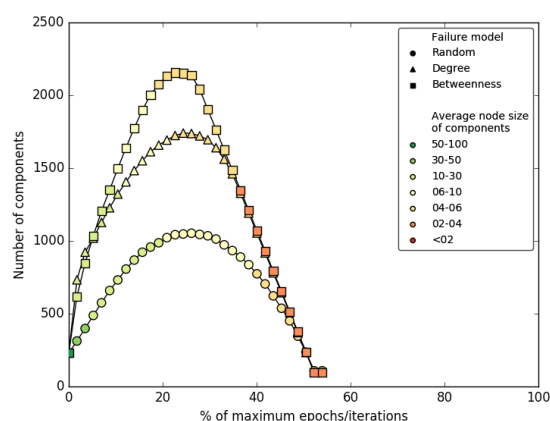


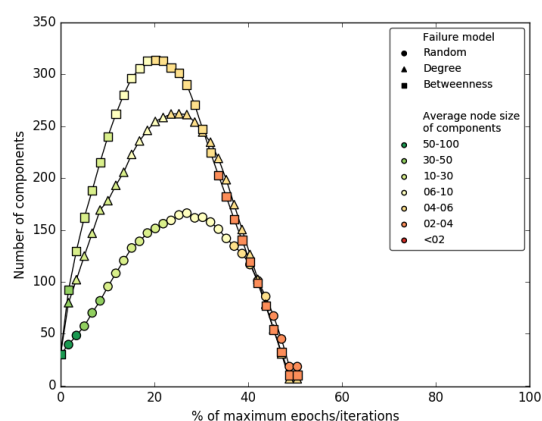
Figure E.34: Behaviour for the Janet communications network to the three topological failure methods.



National Grid electricity transmission NT:



National Grid gas transmission:



Great Britain electricity transmission network:

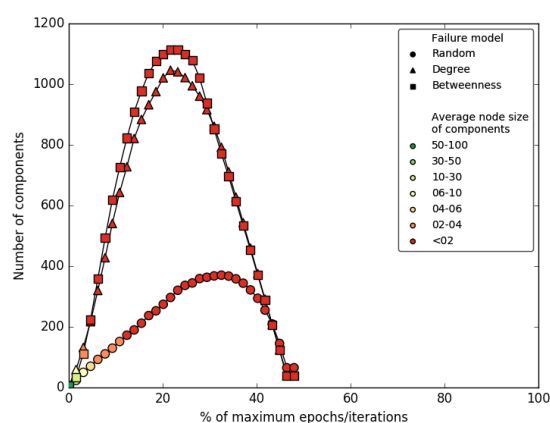
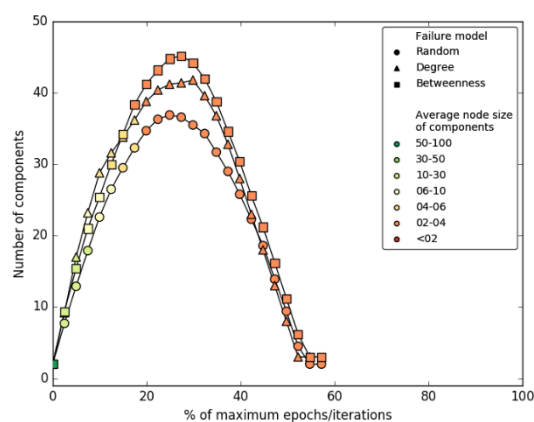
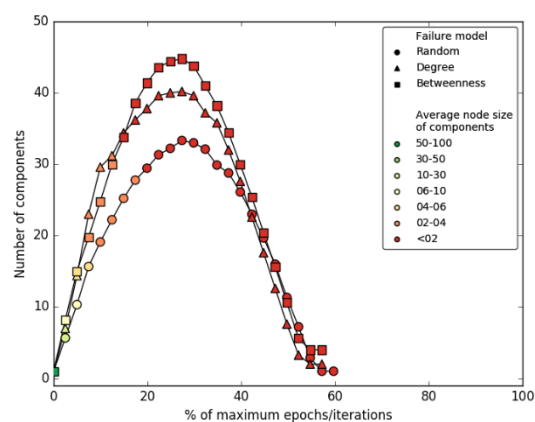


Figure E.35: Response to the three topological failure methods for the energy infrastructure networks.

Ireland rail:



Ireland rail with shortcuts:



Great Britain rail:

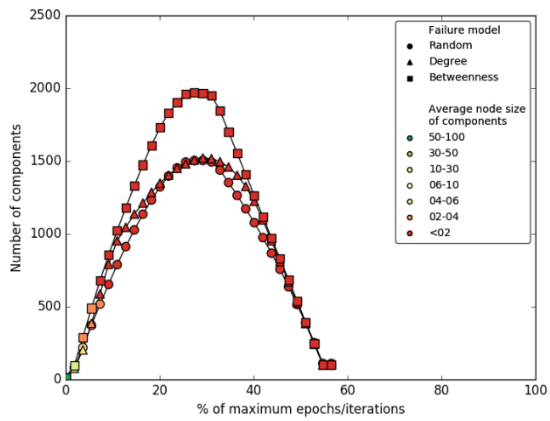
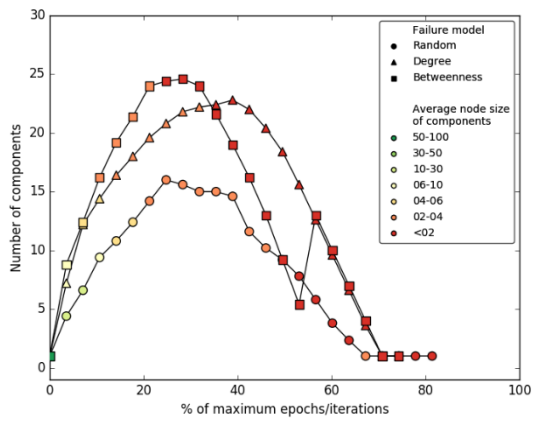
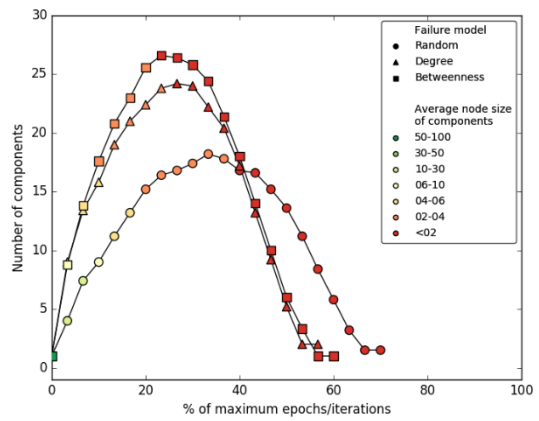


Figure E.36: Response of the national scale rail networks to the three failure methods.

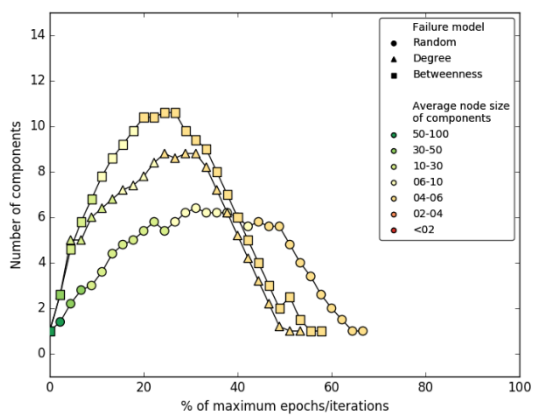
Boston subway:



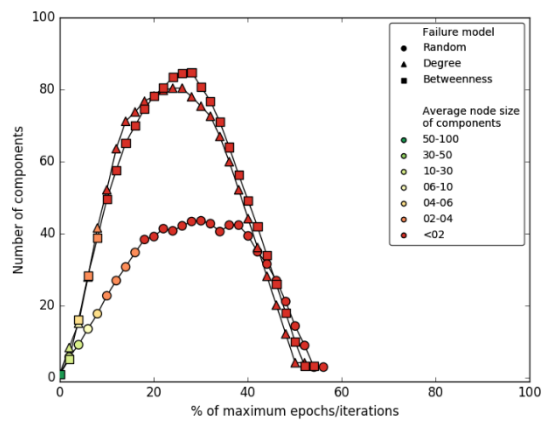
Boston subway with TAPAN:



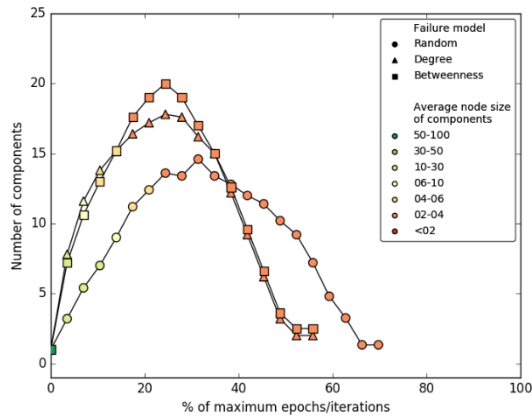
London DLR:



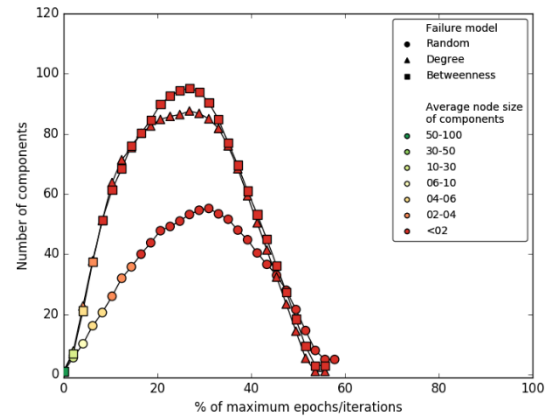
London lightrail:



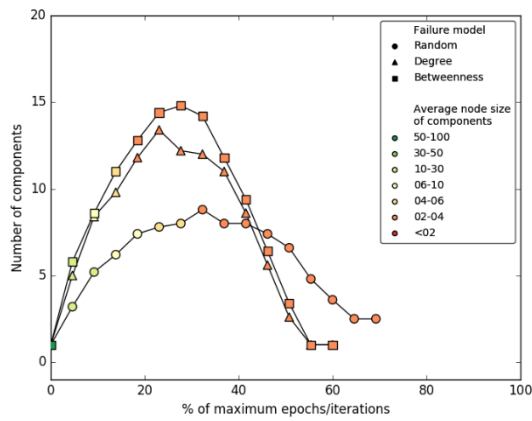
London Underground:



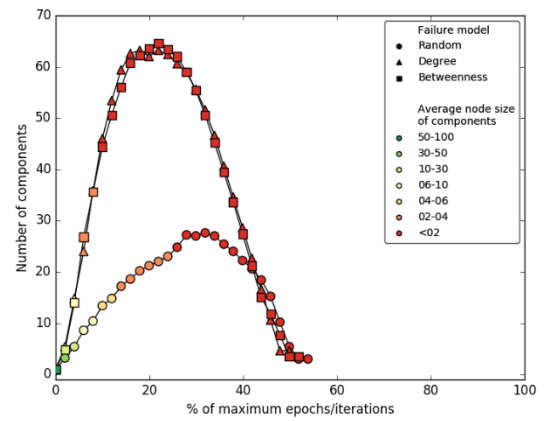
London Tube:



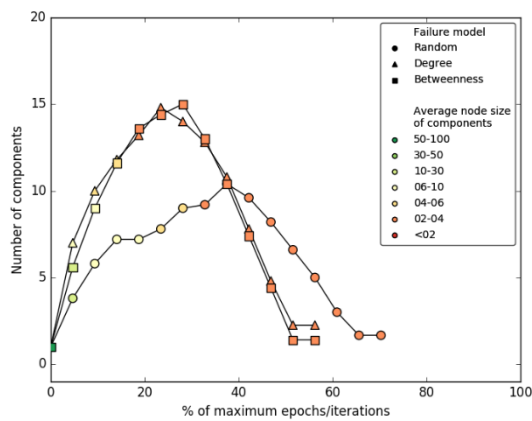
Manchester Metrolink:



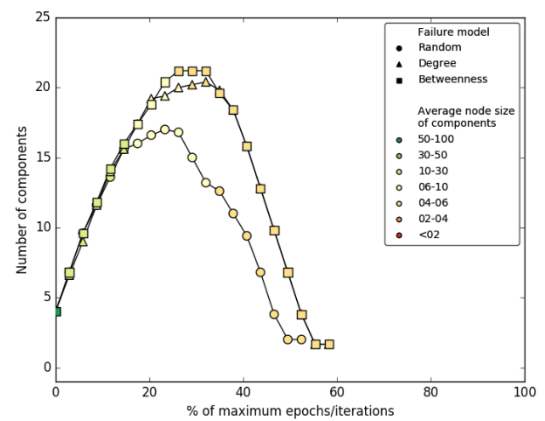
RATP (Paris) metro:



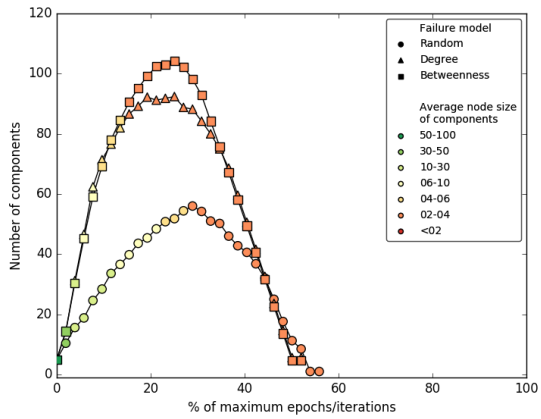
RATP (Paris) RER:



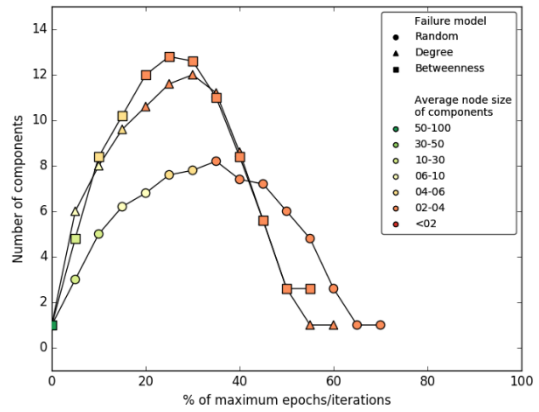
RATP (Paris) tram:



RATP (Paris) integrated:



Tyne and Wear metro:



Tyne and Wear metro with shortcuts:

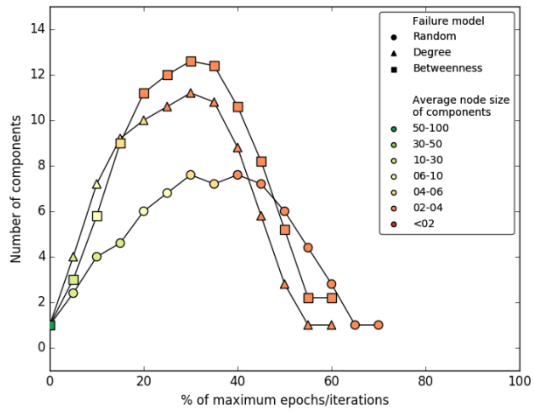
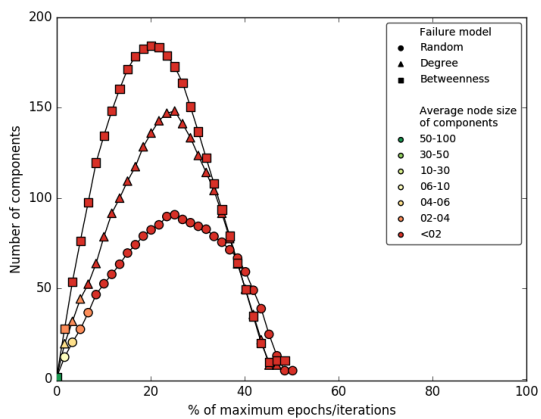
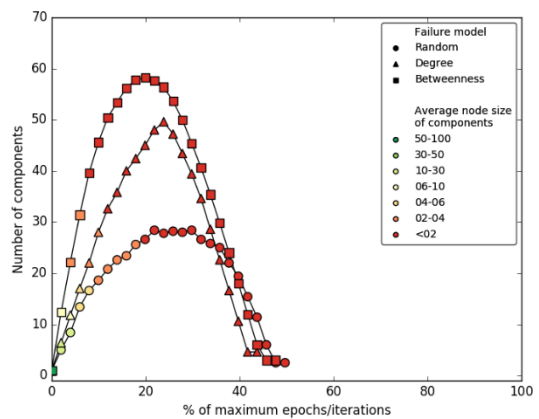


Figure E.37: Behaviour of the regional rail (light rail) networks to the three topological failure methods.

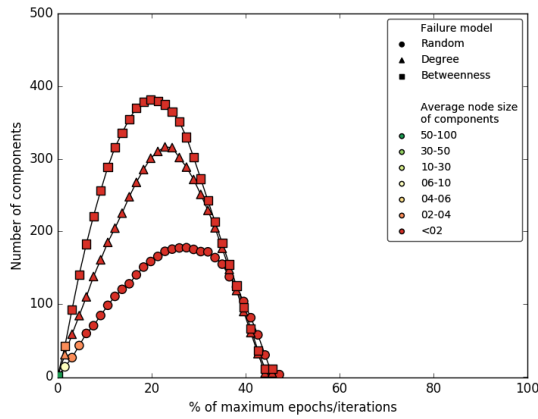
River Dee:



River Eden:



River Severn:



River Tyne:

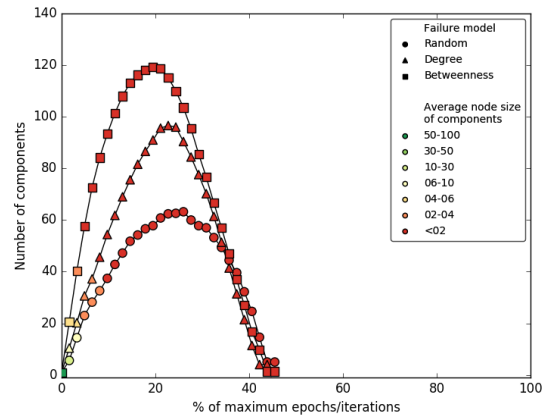
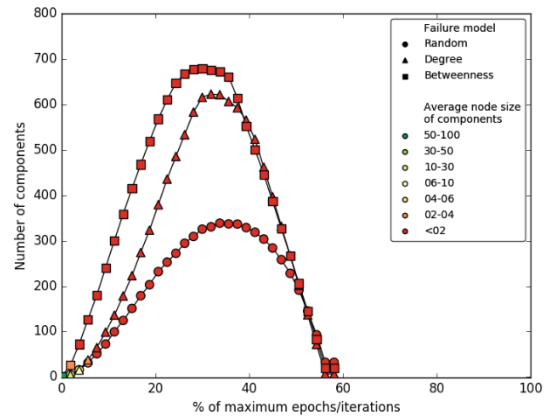
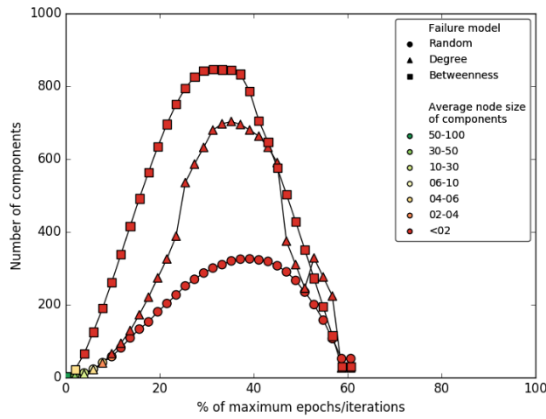


Figure E.38: Behaviour profiles for the four river networks to the three topological based failure methods.

Ireland motorways, trunk and primary roads: Ireland motorways and trunk roads:



Great Britain motorways:

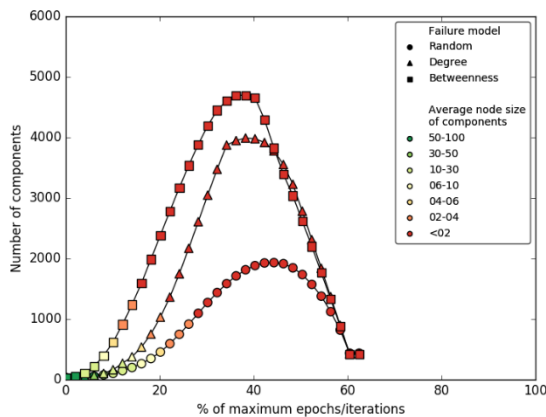
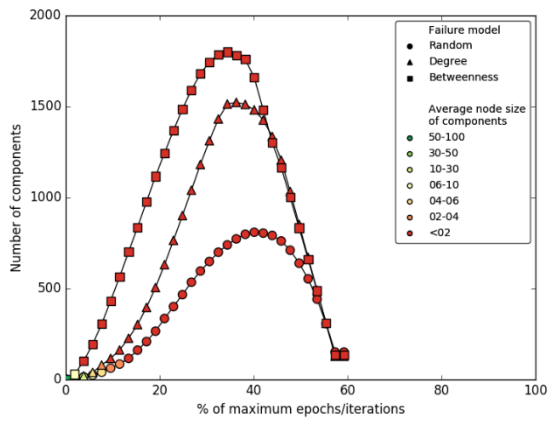
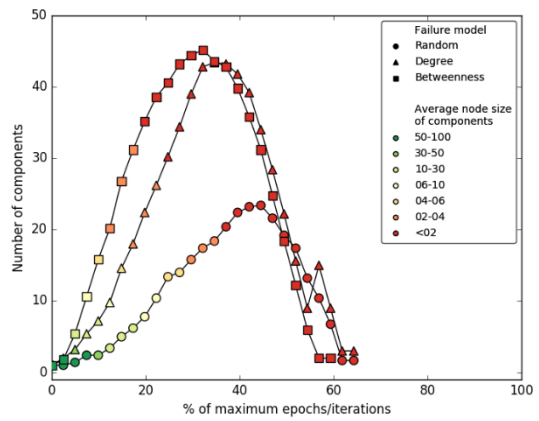


Figure E.39: Response of the national scale road networks to the three topological failure methods.

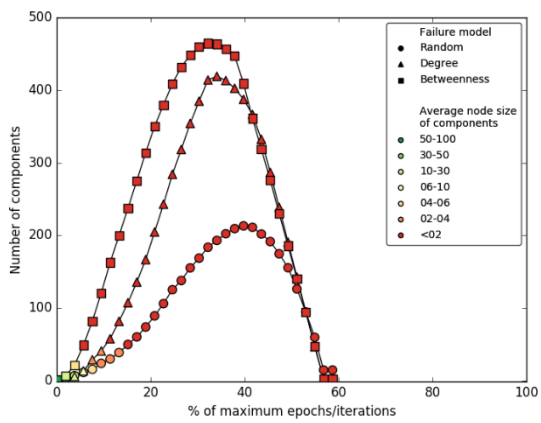
Leeds motorways, A, B and Minor roads:



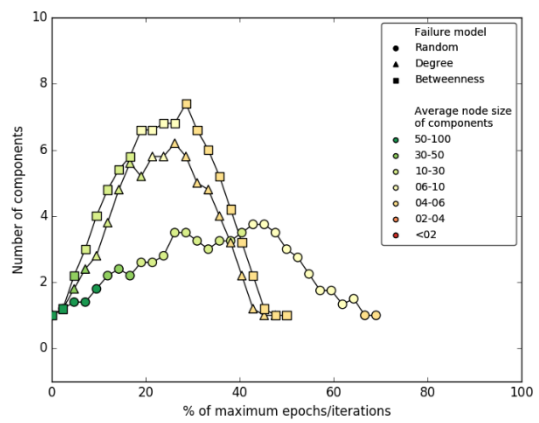
Leeds motorways, A and B roads:



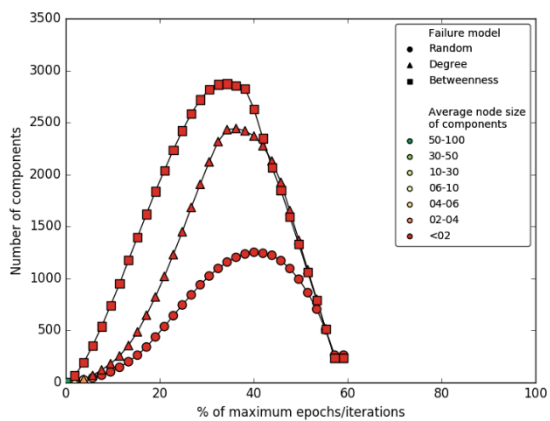
Milton-Keynes motorways, A, B and minor roads:



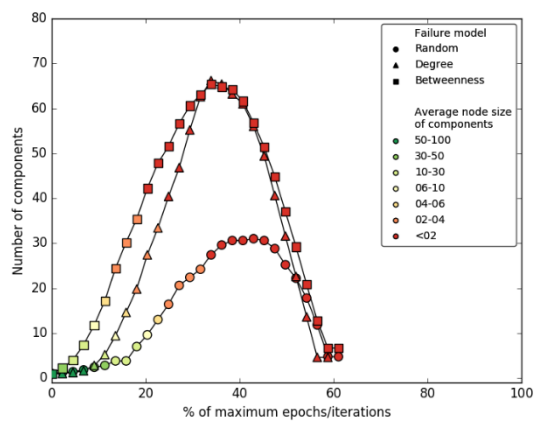
Milton-Keynes motorways, A and B roads:



Tyne and Wear motorways, A, B and minor roads:



Tyne and Wear motorways, A and B roads:



Tyne and Wear motorways and A roads:

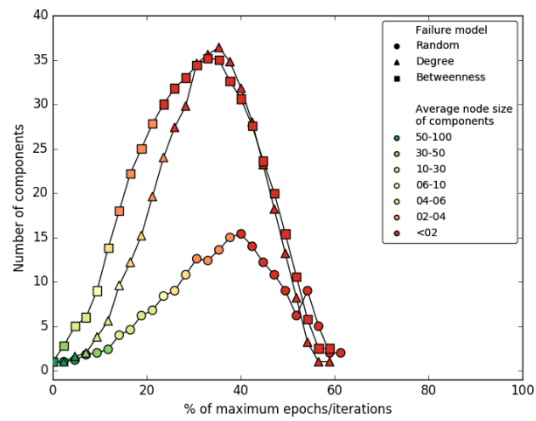


Figure E.40: Behaviour of the suite of regional road networks to the three topological failure methods.

Appendix F: nx_pgnet_atts documentation

Documentation on the developed nx_pgnet_atts python and postgres functions for the nx_pgnets_atts schema.

Description:

An extension to the ITRC interdependency network database model, including database schema, wrappers and functions, to handle explicitly network attributes and their functions for the field of network modelling and simulation.

Developed by: Craig Robson

Newcastle University

January 2016

1 Introduction

The contents of this document sets out the form and structure of an extension to the ITRC database network schema, nx_pgnet, which explicitly handles network attributes and the functions related to these. A schema diagram below, Figure 1, shows the local structure for a single network. The schema can handle multiple networks. See the nx_pgnet library for more details on the global structure of the database.

Section 2 introduces how the extension can be utilised to analyse networks. Section 0 provides some information on the key python functions which have been developed which enable this extension to work seamlessly with the ITRC database schema model. Section 4 introduces the key PostgreSQL functions which have been developed for the handling and manipulation database-side of the network tables.

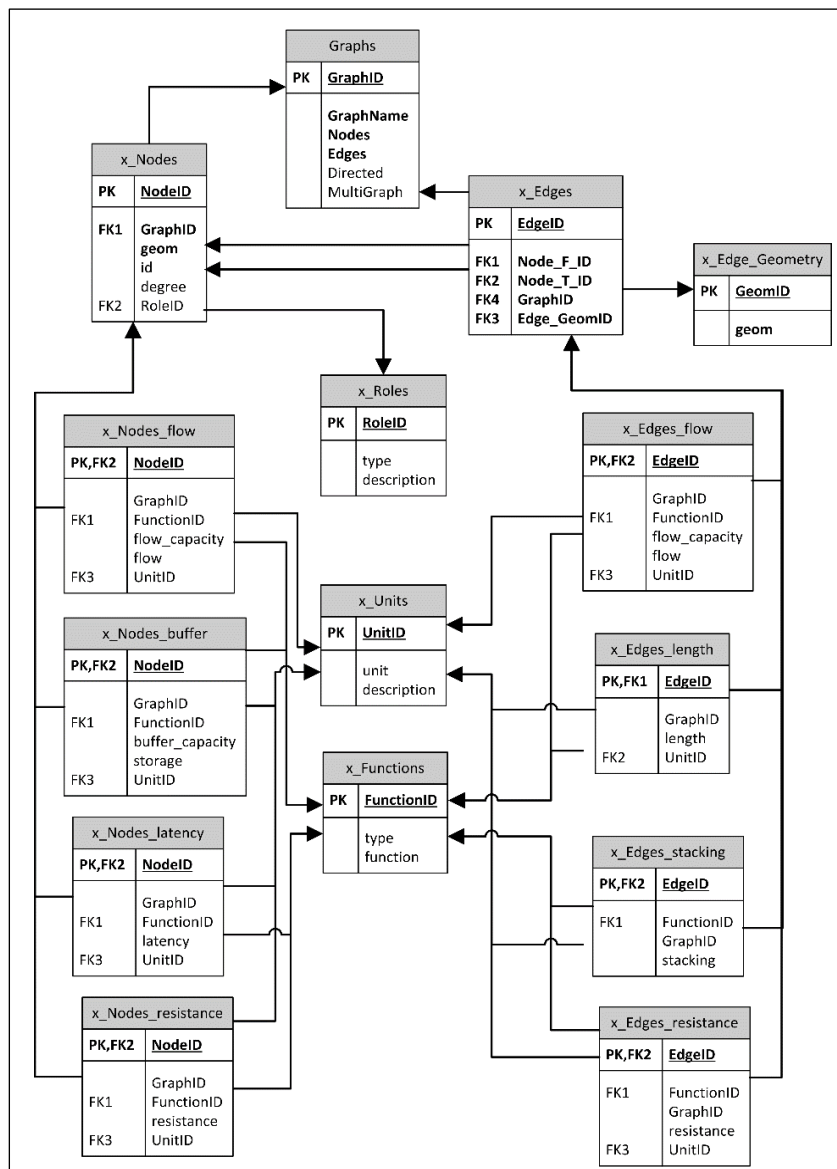


Figure 1: nx_pgnet_atts schema. 'x' refers to the name of the network/graph.

2 Using the extended database schema and functions

Key functions which facilitate the of the database schema for the storage, management and analysis of networks.

2.1 Loading a network from the database

This creates a NetworkX instance of the network from a database. The attributes variable allows the attributes to be added to the network, from a selection of up to five for the nodes and five for the edges) when returned to be specified. Function for each are also retrieved. If attributes set to 'None', a network is returned without any of the attributes are functions attached to nodes or edges.

```
G = nx_pgnet_atts.read(conn, network_name).read_from_db(attributes)
```

```
attributes = [{ 'flow':False, 'capacity':True, 'storage':False, 'resistance':False, 'latency':False},  
{ 'flow':False, 'capacity':True, 'length':False, 'resistance':False, 'stacking':False}]
```

2.2 Adding a network to the database

From a network instance a network can be written to the database schema.

If the network contains the attributes and functions for them, these can be added to the database schema through specifying in the attribute dict those which are present in the network. The contains_atts variable and contains_functions variable must also be set as True. The function will then add the functions to the function table for the network and then add the correct function id for each node and edge.

If the attributes are present in the network instance, but the functions, these can be added in a similar way, though the functioned for the nodes and edges will be left blank.

The overwrite variable allows the function to overwrite a network with the same name in the database with that selected as the input for this function.

```
G=nx_pgnet_atts.write(conn,network_name).write_to_db(G,attributes,contains_atts,contains_  
functions,overwrite)
```

2.3 Add functions to function table

Where the functions for attributes are not present in the network, these can be added separately to the functions table. A list of lists is used so multiple functions can be added in a single process, where the format for each should be 'type,function text, function id'.

```
result = nx_pgnet_atts.write(conn,name).add_functions(functions)
```

2.4 Update a function

A function can be updated directly in the database. This updates the functions table, and thus any network instances will have to be re-created from the database or updated directly as a network instance.

```
result = nx_pgnet_atts.write(conn,name).update_function(functioned,new_function,  
function_type)
```

2.5 Find existing functions

All functions in the network functions table are returned, along with their id's and types.

```
result = nx_pgnet_atts.read(conn,name).return_network_functions()
```

3 Python functions developed

3.1 Write class

write_to_db()

Writes a network to the database, storing the functions and attributes in separate tables for nodes and edges if identified by the user.

populate_roles()

Writes a network to the database, storing the functions and attributes in separate tables for nodes and edges if identified by the user.

add_functions()

Adds a function to the function table with the specified unique function id, the text for the function (written in a pythonic format) and the type of function it is.

update_functions ()

Updates the function text (and the type is required) for a specified function in the function table.

add_atts_randomly()

Writes a network to the database, storing the functions and attributes in separate tables for nodes and edges if identified by the user.

assign_roles_randomly()

Writes a network to the database, storing the functions and attributes in separate tables for nodes and edges if identified by the user.

3.2 Read class

read_from_db()

Loads a network from the database with the attributes and functions if both requested and present in the database, creating a networkx network instance.

return_network_functions()

Returns a list of the functions in the function table.

3.3 Table_sql class

create_units_table()

Calls the developed postgres function 'np_create_units_table'. Creates the relation/table to store the unit information in for the metrics which are handled explicitly.

create_role_table()

Calls the developed postgres function 'np_create_role_table'. Creates the role relation/table which stores each role and it's id (one should be assigned to every node).

create_function_table()

Calls the developed postgres function 'np_create_function_table'. Creates the function relation/table which stores functions and an id which can assigned to each node/edge for each explicitly handled attribute.

create_nodes_attribute_table()

Calls the developed postgres function 'np_create_node_attribute_table'. Creates a node relation/table for the storage of attribute data for the specified attribute including the value, the units and the function for each node.

create_edge_attribute_table()

Calls the developed postgres function 'np_create_edge_attribute_table'. Generated an edge attribute table for the specified attribute for the storage of the attribute values, units and functions for each edge.

rename_node_column()

Calls the developed postgres function 'np_rename_node_column'. Renames a column in the node table if the same as one of the explicitly handled attributes.

rename_edge_column()

Calls the developed postgres function 'np_rename_egde_column'. Renames a column in the edge table if the same as one of the explicitly handled attributes.

check_attribute_table_exists()

Calls the developed postgres function 'np_check_attribute_table_exists'. Checks if the attribute table exists for the network and attribute specified.

update_node_attributes()

Updates for a specified node and attribute the given details in the respective attribute node table.

update_edge_attributes()

Updates for a specified edge and attribute the given details in the respective attribute edge table.

update_fuction()

Updates the details for a function in the function table.

update_node_functionid()

Updates the id of a function for a given node and the given attribute.

update_edge_functionid()

Updates the id of a function for a give node and the given attribute.

get_function_ids()

Return the ids of the functions in the function relation/table.

get_functionid()

Returns the function id of the given function from the function relation/table.

get_node_data()

Returns for all nodes the data (value, function etc.) for them for a specified attribute.

get_edge_data()

Returns for all edges the data (value, function etc.) for them for a specified attribute.

get_units_dict()

Returns a dictionary of the units in the units relation/table.

get_role_dict()

Returns a dictionary of the roles in the roles relation/table.

get_roleid()

Returns the role id for a given role from the role relation/table.

get_unitid()

Returns the unit id for a given unit from the unit relation/table.

check_role_column()

Checks in the node relation/table that the 'role_id' column has been created.

4 PostgreSQL functions developed

A number of PostgreSQL functions have been developed which enable the creation, manipulation and management of the network tables required to full fill the requirements of the developed model. The key functions which may be of interest to a user are specified below. For all others, please see the list of functions in the database itself.

np_add_edge_attribute()

Updates the attribute value (attribute, the value and units user specified) and the functionID (providing that specified exists in the function table).

np_add_edge_attribute_no_units()

Updates the attribute value (attribute and the value user specified) and the functionID (providing that specified exists in the function table).

np_add_function()

Adds a new function to the function table with the user specifying the functionID.

np_add_function_no_checks()

Forces the addition of a new function in the function relation/table without running the required checks. 'np_add_function' (above) should be used rather than this.

np_add_functionid_to_edge_attribute_table()

Updates the functionID, provided by the user, if it exists in the function table, for the specified edge record in the specified attribute table.

np_add_functionid_to_node_attribute_table()

Updates the functionID, provided by the user, if it exists in the function table, for the specified node record in the specified attribute table.

np_add_node_attribute()

Updates the attribute value (attribute, the value and the units user specified) and the functionID (providing that specified exists in the function table).

np_add_node_attribute_no_units()

Updates the attribute value (attribute and value user specified) and the functionID (providing that specified exists in the function table).

np_check_attribute_table_exists()

Given an attribute, checks if it exists that a table exists for the specified network.

np_check_function_exists()

Checks that the specified function exists in the function relation/table.

np_create_edge_attribute_table()

Build an attribute table for the edges in a network, provided with a name by the user.

np_create_edge_view()

Generates the edge view from which the network is built from. Adds attribute columns to the edges and their geometries as requested by the user.

np_create_function_table()

Creates the function table when the network is initially built using the network prefix.

np_create_node_attribute_table()

Build an attribute table for the nodes in a network, provided with a name by the user.

np_create_node_view()

Generates the node view from which the network is built from. Adds attribute columns to the nodes, including their geometry, as requested by the user.

np_create_role_table()

Generates the role table for the specified network.

np_create_units_table()

Generates the units table for the specified network.

np_delete_all_tables()

Deletes all tables related to the network from the prefix provided including those for the network itself and the views.

np_rename_edge_column()

Renames an attribute of the edges in the edge table if the same as one of the specified attributes which are handled explicitly.

np_rename_node_column()

Renames an attribute of the nodes in the nodes table if the same as one of the specified attributes which are handled explicitly.

np_update_edge_attribute()

Updates the attribute of the specified edge record for the supplied attribute. Checks the table exists for attempting to run the update command.

np_update_node_attribute()

Updates the attribute of the specified node record for the supplied attribute. Checks the table exists for attempting to run the update command.

np_update_units_edges()

Updates the units id for an edge for the specified attribute.

np_update_units_nodes()

Updates the units id for an node for the specified attribute.

Appendix G: Graphical User Interface (GUI) documentation

Documentation on using the developed user interface which allows access to many of the models and methods developed for complex network within the research.

resilience_gui

Developed by Craig Robson

Newcastle University

April 2016

1 Introduction

This library provides a user interface for the python library ‘resilience’ along with added visual aspects to improve the user experience and range of analytics available to a user. Through the developed user interface the tool aims to provide the same level of flexibility as available natively through the resilience library thus allowing users to run a multitude of failure simulations. The functionality of the tool also extends from this to offer the ability to compute a range of common graph metrics, a selectin of those available in the NetworkX library. The library also provides the tool, and the user, with the ability to generate networks using some common graph generators. Finally, the tool utilises a range of drawing functions available in NetworkX to offer the user the ability to visualise a network during perturbations or not.

1.1 Dependencies

The developed tool relies on a number of other libraries for which without the tool may not run or will suffer from reduced functionality.

- Python 2.7+
- NetworkX 1.7+
- PyQt4
- robustness 1.0+

1.2 Further reference material

- NetworkX - <https://networkx.github.io/>

2 Using the interface

2.1 User Interface

The user interface has three main portions; (i) input network settings, (ii) failure analysis settings and (iii) the control buttons/menu’s. These are indicated in Figure 1 below.

An error message may appear on the initial loading of the tool about not being able to find/load the ‘nx_pgnet’ module. The location of this, if downloaded and required for the users objectives, can be specified in the ‘View Options’ option in the ‘Edit’ menu.

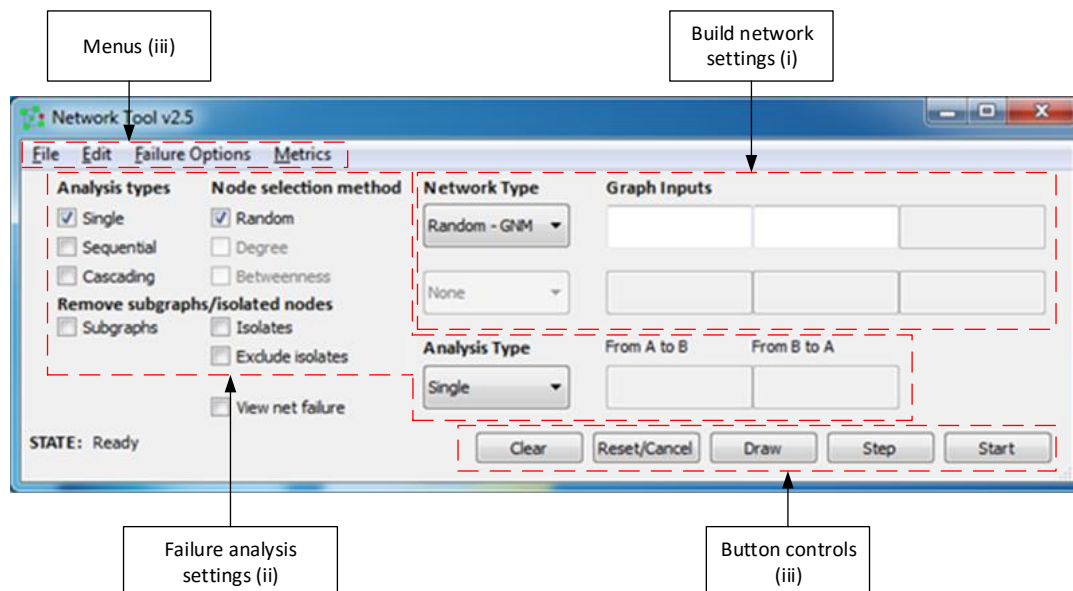


Figure 1: Diagram of interface with main aspects highlighted.

2.2 Building a network

There are number of options available to build a network:

- Graph generator
- CSV
- List
- Database

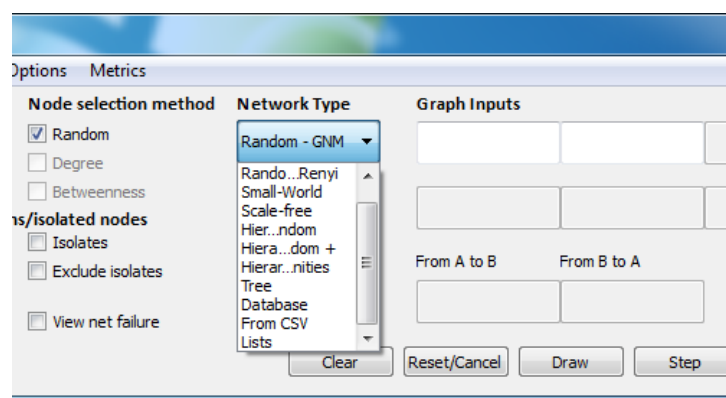


Figure 2: Highlighting the input options to build a network.

Before any analysis can be selected to be run the network build method should be selected and the parameters (file or values) entered to allow the network to be built. For the first option, using a graph generator algorithm, as available in the NetworkX library, help is available by hovering over the input boxes which are not shaded out for the network.

The CSV option requires the user to provide a CSV which contains a list of nodes and edges on consecutive lines in a text file.

The list option requires a list of nodes and a list of edges to be entered manually.

The database option allows users who have access to a database using the nx_pgnnet schema to build a network from this, which opens the option for visualising the network geographically as well. This requires the database connection parameters and the name of the network to be loaded.

2.3 Calculating metrics and simple visualisation

To calculate a metric over a network options are available in the 'Metric' (Figure 3) drop down menu. These allow the computation of a number of metrics using algorithms available in the NetworkX library. Results are returned in a window. There are options available for the computation of multiple metrics simultaneously, again with the results returned in a window.

Visualisations of networks can be quickly obtained through the 'Draw' button, which then gives options specific to the network build method, Figure 4 (when built from the database this allows a geographic visualisation). The visualisation can be customised using the 'Edit' menu and the 'View Options' item, Figure 5, which allows the colouring and size of nodes/edges to be changed, including having the size based on a metric value, for which there a small number of options. The results of any visualisation can be saved in a range of formats using the toolbar available in the visualisation window, Figure 6.

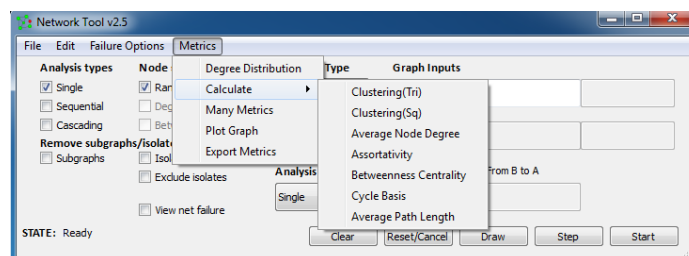


Figure 3: Showing the 'Metrics' menu and the available calculation options.

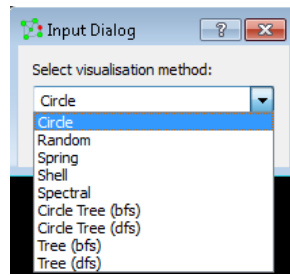


Figure 4: Draw menu showing the drawing algorithms available.

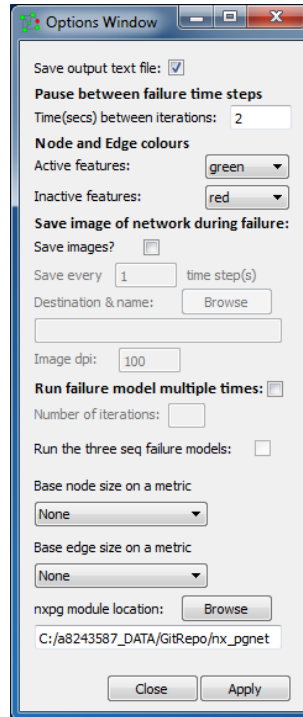


Figure 5: The Options menu.

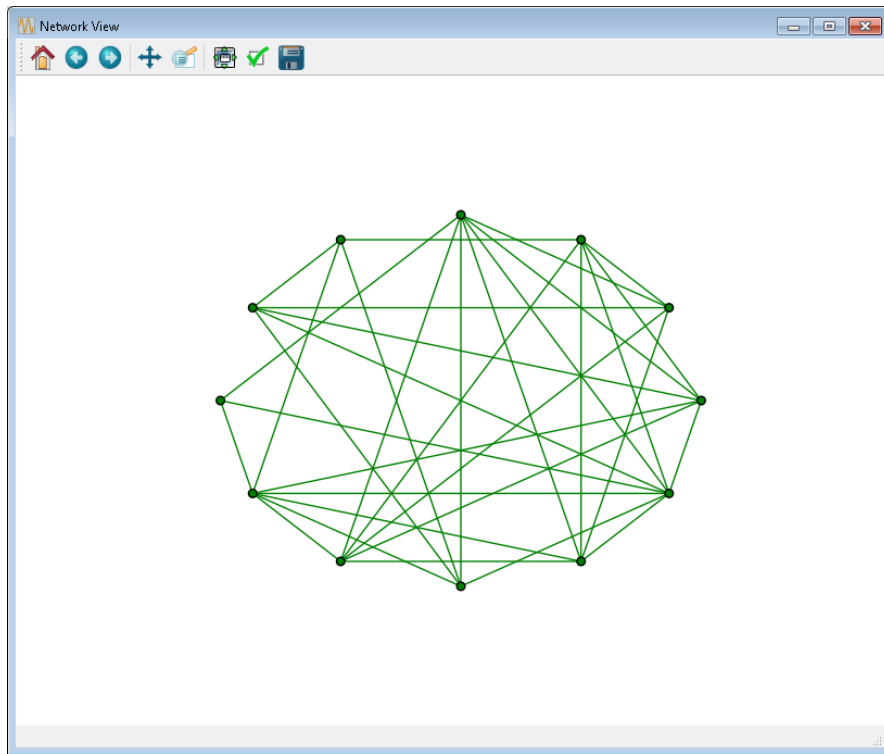


Figure 6: Showing the visualisation window, with the save button in blue on the top toolbar..

2.4 Simple failure simulations

To run a failure simulation first select the ‘network type’ and then enter the parameters required to build the specific network. Following this select one of the ‘analysis types’ and then one of the ‘node selection method’s’. If you click run ‘Start’ now the simulation will run. However if you select the ‘View net failure’ option before clicking ‘Start’ a visualisation window will appear and the failure can be seen. Upon clicking start a window will open asking for a file to write the results to, using a simple text format. A pause time is used to slow the simulation down so the visualisation is readable, this along with many other options are available from the ‘Edit – View Options’ window (Figure 5) (some further advanced options are available through the ‘Options’ option in the ‘Failure Options’ menu). Further options are available on the user interface which allow for the customisation of how the simulation handles subgraphs and isolated nodes (under the ‘Remove subgraphs/isolated nodes’ heading).

While a simulation is running the ‘Pause’ button can be pressed at any time with the simulation stopping at the end of the current iteration. This allows for a more detailed analysis of the graph in the visualisation window. To resume, press the ‘Start’ button again. Rather than the simulation running one go, via the ‘Step’ button, one step can be run at a time. At any time during the simulation the ‘Start’ button can be pressed and the simulation will run automatically until the end.

At the end of a simulation there is the option to view the metrics which have been computed throughout the analysis on two plots (to change the metrics for the next simulation see the following section). Simply select a metric from each of the drop down menus and the plots should be updated automatically, allowing for comparisons to be made between metrics and this the behaviour of the network analysed. At any time the plots can be saved using the save button in the tool bar across the top of the window.

2.5 Complex failure simulations

The tool and the underlying resilience module facilitates the ability to analyse dependencies between networks, using the same failure options as for the single network analysis. Where dependency is concerned, the network which is the ‘parent’ is the one subjected to the failures. The settings/parameters for the second network can be set by changing the ‘Analysis Type’ drop down menu to ‘dependency’ (Figure 7), which should then enable a second network to be created. The dependency links can be created randomly (‘Failure Options’ – ‘Random Dependency Edges’ (Figure 8)) or entered in the input box titled ‘From A to B’. these should be in the form of a list of tuples. As with the analysis of a single network, clicking the ‘Start’ button commences the analysis. It should be noted when doing dependency analysis the simulation cannot be visualised.

It is also possible to run interdependency analysis in a similar fashion to the dependency analysis.

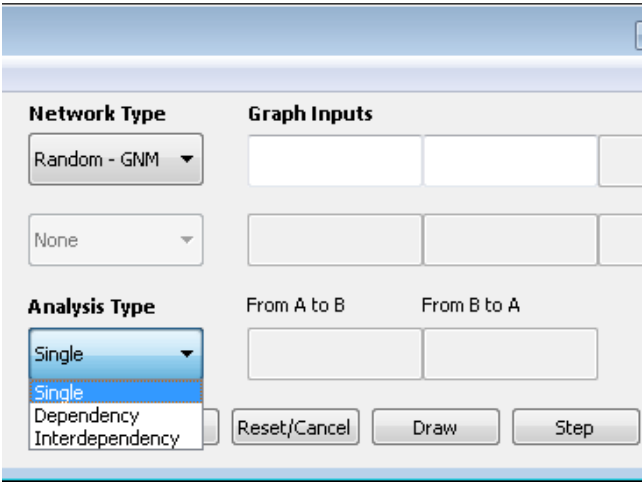


Figure 7: List of the different Analysis Types available, including dependency and interdependency.

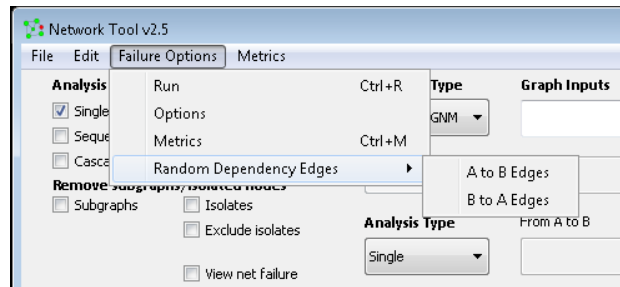


Figure 8: Showing the Failure Options menu and the option to create random dependency links.

3 Further settings

3.1 Metrics

A default set of metrics are calculated during a simulation at each time step, however for each simulation the metrics computed can be changed going to the ‘Failure Options’ – ‘Metrics’ window (Figure 9). Those included, if possible, will also be listed in the visualisation at the end of a simulation.

3.2 Config files

Configuration files can be saved and re-loaded so where a user sets up the gui to run a simulation, the settings can be re-loaded when the gui is next opened, as well as allowing for the settings to be reviewed outside of the gui from the text file directly. This is done through the ‘Edit’ menu.

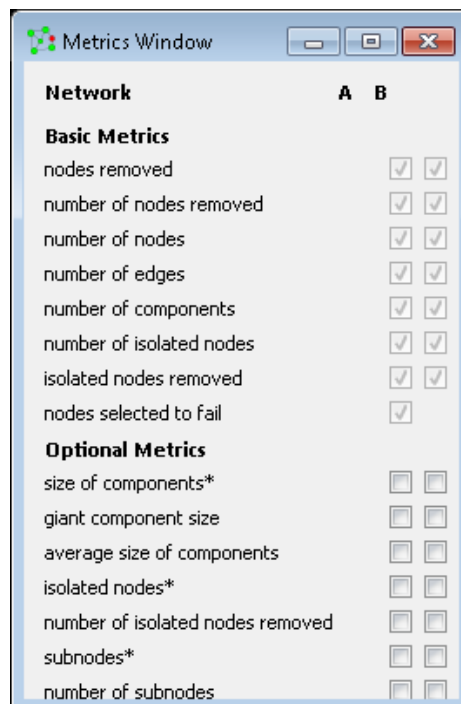


Figure 9: A snapshot of the 'Metrics' window and the possible metrics (note that this shows less than half of the available optional metrics).