Towards the Integration of Complex Systems Theory, Geographic Information Science, and Network Science for Modelling Geospatial Phenomena

by

Taylor Anderson

M.Sc., Simon Fraser University, 2015

B.E.S., University of Waterloo, 2013

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Geography
Faculty of Environment

SIMON FRASER UNIVERSITY

Summer 2019

Approval

Name:	**Taylor Anderson**
Degree:	**Doctor of Philosophy**
Title:	**Towards the Integration of Complex Systems Theory, Geographic Information Science, and Network Science for Modelling Geospatial Phenomena**
Examining Committee:	**Chair:** Margaret Schmidt Associate Professor
	Suzana Dragicevic Senior Supervisor Professor
	Christina Semeniuk Supervisor Assistant Professor Department of Biological Sciences University of Windsor
	Nicholas Coops Supervisor Professor Department of Forest Resources Management University of British Columbia
	Duncan Knowler Internal Examiner Associate Professor Resource and Environmental Management
	Arika Ligmann-Zielinska External Examiner Associate Professor Department of Geography, Environment and Spatial Sciences Michigan State University
Date Defended/Approved:	June 19, 2019

Abstract

A complex systems approach conceptualizes spatial systems from the bottom-up to better understand how local spatial interactions generate emergent system-level behavior and spatial patterns at large spatial extents. This approach can be applied to examine ecological, urban, and social systems within contexts of geographic space and time. Geographic automata systems (GAS) including cellular automata (CA) and agent-based models (ABM) are spatio-temporal modelling frameworks that are rooted in complex systems theory. In a similar manner, network theory uses a complex systems approach to represent and analyze spatial systems as sets of georeferenced nodes and links that form measurable spatial networks. Separately, GAS and network-based approaches offer unique advantages in exploring and analyzing complex systems, however the two approaches are rarely integrated. Therefore, the purpose of this dissertation is to explore the intersection of complex systems theory, geographic information science, and network theory to leverage the advantages of each field for better understanding a variety of complex spatial systems. The main objective is to develop a suite of novel network-based automata modelling approaches that simulate complex dynamic spatial systems as measurable, evolving, spatial networks. Three novel modelling approaches are developed including: a geographic network automata (GNA) model that uses spatial networks, network-based transition rules, and network analysis for the representation of complex spatial systems; a network-based ABM (N-ABM) that integrates networks not as inputs for the ABM, but as a novel way to conceptualize, analyze, and communicate the model and model results; and a network based validation approach for the testing of ABMs. Obtained results demonstrate that the integration of complex systems theory, geographic information science, and network theory offers new means for the representation, analysis, communication, and testing of GAS and the complex systems they represent, thus helping to thus helping to "open the black box". Furthermore, the presentation of modelling results in application to insect infestation and disease transmission contribute to the enhancement of decision-making processes by providing tools that can be used in forecasting and scenario testing. This dissertation contributes new methodological frameworks to the fields of geographic information science, GAS, and network theory.

Keywords: geographic information science, geographic information systems, complex systems science, network science, geographic automata systems, forest insect infestation, epidemiology

This thesis is dedicated to coffee… and maybe also to Gedeon who makes coffee for me every day.

Acknowledgements

I would like to acknowledge the Natural Sciences and Engineering Research Council (NSERC) of Canada for support of this study under the Discovery Grant Program awarded to Dr. Suzana Dragicevic and under the NSERC Canada Graduate Scholarship – Doctoral. In addition, I am very grateful to the Department of Geography and to Simon Fraser University for providing me with the opportunity to obtain several scholarships including graduate fellowships, travel awards, the President's PhD award, as well as the opportunity to obtain several teaching assistantships and sessional instructor positions. I thank the Town of Oakville, the Canadian Food Inspection Agency, and the US Forestry Service for the provision of data to support this research. I would like to express my gratitude to West Grid and Compute Canada for the use of their infrastructure for high performance computing. Lastly, I thank Nick Collier and Eric Tatara, developers of Repast Simphony, for their continued dedication and support to the community that uses this software.

I am very fortunate to have benefitted from the immeasurable investment of time and resources, guidance, and support from several exceptional scholars throughout my academic career. It is largely because of this that I have been inspired to achieve excellence in both research and teaching. I cannot begin to express my gratitude for the dedication, hard work, and mentorship generously offered by my supervisor, Dr. Suzana Dragicevic. Her contributions to this research and to my professional development as an academic and an individual are invaluable. In addition, I would like to thank my supervisory committee members, Dr. Christina Semeniuk and Dr. Nicholas Coops, for their unwavering support and vital guidance and advice. I thank my SAM Lab colleagues as well as other colleagues within and outside of the Department of Geography that I have had the opportunity to meet over the years. Special thanks to the staff in the Department of Geography, most notably Anke Baker for countless administrative support and guidance you have offered.

On a personal level, I am forever grateful to my love, my friends, and my family - for your love, motivation, and patience.

Table of Contents

Approval ii
Abstract iii
Dedication v
Acknowledgements vi
Table of Contents vii
List of Tables xi
List of Figures xiii

Chapter 1. Introduction 1
1.1. Research Problems and Questions 5
1.2. Research Objectives 5
1.3. Dissertation Overview 6
1.4. References 8

Chapter 2. Overview of Network Theory and its Geospatial Applications 11
2.1. Abstract 11
2.2. Introduction 11
2.3. Theory behind Network Science 13
2.3.1. Descriptive Graph Theory Measures 14
Degree 15
Average Degree 16
Degree Distribution 16
Clustering Coefficient 17
Average Clustering Coefficient 17
Path 18
Shortest Path Length 18
Average Path Length 18
Diameter 18
Assortativity 18
Betweenness Centrality 19
Weight 19
Community 20
Transportation indices and Variants 20
Distance Based Metrics 21
2.3.2. Theoretical Network Structures 22
2.4. Graph Theory to Measure the Structure of Complex Spatial Systems 24
2.4.1. Spatial Transportation Networks 25
2.4.2. Spatial Social Networks 26
2.4.3. Spatial Ecological Networks 28
2.4.4. Geophysical Networks 29
2.5. Linking Structure and Dynamics of Complex Spatial Systems using Networks 30
2.6. Conclusion and Future Directions 31

2.7. References 33

Chapter 3. Representing Spatial Systems as Evolving Networks: A Geographic Network Automata Model 43

3.1. Abstract 43
3.2. Introduction 43
3.3. Geographic Network Automata (GNA) 46
3.3.1. GNA Modelling Framework 46
3.3.2. GNA Spatial Network Analysis using Graph Theory 51
3.4. Game of Life GNA Model (GNA_{GOL}) 53
3.4.1. GNA_{GOL} Model Framework 53
3.4.2. GNA_{GOL} Scenarios 56
3.4.3. GNA_{GOL} Model Testing 58
3.5. Results 59
3.5.1. GNA_{GOL} Simulation Results 59
Scenario 1 59
Scenario 2 60
Scenario 3 61
3.5.2. GNA_{GOL} Spatial Network Analysis Results 63
Characterizing the Underlying Network UN 64
General Trends 64
Comparison of Two Different Landscapes 65
Correlation between Graph Theory Measures 66
Degree Distribution 68
3.6. Discussion and Conclusions 69
3.7. References 70

Chapter 4. A Geographic Network Automata Approach for Modelling Dynamic Ecological Systems 73

4.1. Abstract 73
4.2. Introduction 73
4.3. Theoretical Framework of the Geographic Network Automata 75
4.4. GNA_{EAB} Modelling Methodology 81
4.4.1. Characteristics of Emerald Ash Borer Insect Infestation 82
4.4.2. Study Site and Data Sets 83
4.4.3. GNA_{EAB} Model Overview 84
Spatial Network Structure SN_{EAB} 85
Evaluating Habitat Patch Susceptibility 86
Implementing GNA_{EAB} Dynamics 88
GNA_{EAB} Model Output Analysis 91
GNA_{EAB} Model Sensitivity Analysis and Calibration 92
4.5. Results 96
4.5.1. GNA_{EAB} Simulation Results 96
4.5.2. Analysis of GNA_{EAB} Model Outputs 100
4.6. Discussion and Conclusion 104

4.7. References 106

Chapter 5. Network-Agent Based Model for Simulating the Dynamic Spatial Network Structure of Complex Ecological Systems 110
5.1. Abstract 110
5.2. Introduction 110
5.3. Networks and their Representation through Graph Theory 113
5.4. Methods 116
5.4.1. Emerald Ash Borer (EAB) Biological Background 117
5.4.2. N-ABM 118
Simulating Infestation Dynamics using Agents 119
5.4.3. Infestation Dynamics as Networks 123
5.4.4. N-ABM Testing 124
5.4.5. Analysis of Generated Spatial Networks 125
Average Node Degree 125
Degree Distribution 126
Average Path Length 126
Average Clustering Coefficient 127
5.5. Results 127
5.5.1. Simulation Results 129
5.5.2. Analysis of the Spatial Network Structure Results 131
Graph Size 131
Average Node Degree and Degree Distribution 131
Degree Distribution across Time 135
Average Clustering Coefficient 135
Average Path Length 136
5.6. Discussion and Conclusion 139
5.7. References 141

Chapter 6. A Validation Approach for Spatially Explicit Agent-Based Models using Networks 148
6.1. Abstract 148
6.2. Introduction 148
6.3. Overview of Approaches for Model Validation 151
6.4. NEtworks for Agent-based Model Testing (NEAT) Approach 152
6.5. Applying the NEAT Approach to an Epidemiological Agent-Based Model (Epi-N-ABM) 154
6.5.1. Modelling Epidemics 155
6.5.2. Epi-N-ABM Simulating Influenza Dynamics using Agents 155
Model Overview 156
Design Concepts 160
Details 161
6.5.3. Epi-N-ABM: Simulating Influenza Dynamics as Networks 163
6.5.4. Epi-N-ABM Testing Using Networks 163
Sensitivity Analysis and Calibration 163

Sensitivity Analysis and Calibration Results 164
Validation 167
6.6. Results 169
6.6.1. Simulation Results 169
6.6.2. Validation Results 177
6.7. Discussion and Conclusions 180
6.8. References 182

Chapter 7. Conclusions 187
7.1. General Conclusions 187
7.2. Summary of Findings 187
7.3. Future Directions 195
7.4. Research Contributions 198
7.5. References 202

List of Tables

Table 2.1. Transportation indices and variants 20

Table 2.2. Distance based metrics 21

Table 3.1. Examples of spatial, non-spatial, and network properties for both nodes *v* and links *e* 49

Table 3.2. Some graph theory measures for network analysis. 51

Table 3.3. The spatial, non-spatial, and network properties of the nodes and links in the $GNA_{GOL.}$ 54

Table 3.4. Transition rules and their parameterization specific to each scenario. 57

Table 3.5. Sensitivity of the underlying network *UN* to the distance threshold $d_{ij.}$ 58

Table 3.6. Correlation table presenting the correlation between each of the graph theory measures numbered (1) to (5) obtained from the evolving spatial network generated by (a) scenario 1: fluctuating network, (b) scenario 2: network growth, and (c) scenario 3: network shrinkage. 67

Table 4.1. Graph theory measures for network analysis. 92

Table 4.2. Observed real-world rate of spread from 1997 to 2011. 93

Table 4.3. Variation of susceptibility weights. 94

Table 4.4. Defining measures for each network type. 100

Table 4.5. Resulting graph theory measures from the spatial network SN_{EAB} generated by the GNA_{EAB} and the random spatial network $SN_{RAND.}$ 102

Table 5.1. Definitions of important graph theory measures. 113

Table 5.2. A description of (a) agent state variables and (b) agent parameters with references. 120

Table 5.3. Main processes and schedules of each agent. 122

Table 5.4. Summary of network measure results derived from (a) the single N-ABM infestation network chosen for visual presentation of results, and (b) the average value across all 50 generated infestation networks for all network measures with the associated standard error. 127

Table 5.5. Confusion matrix detailing the spatial agreement of the levels of severity of infestation observed in the real world and simulated node degree. 134

Table 5.6. Power law exponent $-\alpha$ over time. 135

Table 5.7. Summary of network measure results derived from the $N\text{-}ABM_{RAND}$ infestation network. 137

Table 6.1. The Epi-N-ABM agent classes, sub-classes, and associated variables. 156

Table 6.2. Epi-N-ABM design concepts as adapted from the ODD protocol. 160

Table 6.3. The Epi-N-ABM subroutines and their description. 162

Table 6.4. The average degree of contact $\langle k_{contact} \rangle$ for the social network obtained using different combinations of interaction likelihoods β_I for each location. 165

Table 6.5. The average degree of infection, average degree of infection if k>0, and the average number of agent's that are infected for the infection network obtained using different combinations of transmission likelihoods β_T. ... 165
Table 6.6. Observed empirical regularities of (a) social networks and (b) infection networks ... 167
Table 6.7. Spatio-temporal network measures obtained from SN_{social} ... 172
Table 6.8. Spatio-temporal network measures obtained from $SN_{infected}$. ... 177
Table 6.9. Table 9. Network measures obtained and averaged over 30 runs for a) the spatial social network and b) its random equivalent ... 178
Table 6.10. Network measures obtained and averaged for the infected network ... 180

List of Figures

Figure 2.1. A simple spatial graph *G* to illustrate graph notations and definitions. 14

Figure 2.2. Adjacency matrix A for an undirected graph G. 15

Figure 2.3. The resulting degree distribution for graph *G,* illustrated in Figure 2.1.....17

Figure 3.1. Different spatial network *SN* dynamics: (a) dynamics *of* a primary network of interest *SN* that evolves over time and (b) dynamics *on* a network *SN* where the network of interest evolves over time as a function of an underlying network *UN*. 47

Figure 3.2. Parameterization of R1-R4 for (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3, where orange values indicate the number of neighbours that results in node death, yellow values indicate the number of neighbours that result in node survival, and blue values indicate the number of neighbours that result in node reproduction.....57

Figure 3.3. Based on (a) the initial state, the evolving spatial networks SN_{GOL} generated from the GNA_{GOL} are presented for scenario 1 (b) t_5, (c) t_{10}, (d) t_{15}, (e) $t_{20,}$ scenario 2 (f) t_5, (g) t_{10}, (h) t_{15}, (i) $t_{20,}$ and scenario 3 (j) t_5, (k) t_{10}, (l) t_{15}, (m) t_{20}. 62

Figure 3.4. Simulation outcomes for scenario 1 fluctuating networks with (a, b) water barrier (blue) and (c, d) street barrier (orange) landscapes for time (a, c) t_5 and (b, d) t_{10} with (e) the values for the calculated number of nodes, number of links, average clustering coefficient, average degree, and average path length for each landscape and all model iterations. 63

Figure 3.5. Calculated number of nodes, number of links, average clustering coefficient, average degree, and average path length for the obtained spatial network *SN* as it evolves over space and time generated from (a) scenario 1: fluctuating network, (b) scenario 2: network growth, and (c) scenario 3: network shrinkage.....65

Figure 3.6. Calculated degree distribution for each evolving spatial network *SN* as it evolves over time generated from (a) scenario 1: fluctuating network, (b) scenario 2: network growth, and (c) scenario 3: network shrinkage.....68

Figure 4.1. Overview of GNA model structure. 76

Figure 4.2. Example of simple spatial networks: (a) undirected (b) directed and the corresponding adjacency matrices (c) undirected and (d) directed.....78

Figure 4.3. Spatial network evolution via (a) new connections, (b) node disconnection, and (c) node rewiring. 80

Figure 4.4. Workflow of the GNA_{EAB} model of forest insect infestation.....85

Figure 4.5. Network evolution as a result of the application of the transition rule *R1* in the case where (a) node *j* has a degree *k=0*, (b) node *j* has a degree *k>=1*, and (c) the cost connection *C1* with barriers of lakes and rivers. ..90

Figure 4.6. Sensitivity of GNA_{EAB} simulated rate of spread to (a) variations in likelihood threshold and (b) weighted combinations for calculating *LDS.* The observed real-world rate of spread (distance per year) is presented in orange and the simulated rate of spread using parameters that best match the observed rate of spread is presented in green. 95

Figure 4.7. The GNA_{EAB} model outcomes: the simulated spatial network SN_{EAB} of EAB dispersal as it evolves from year (a) 1997-2003, (b) 1997-2005, (c) 1997-2009 and (d) 1997-2011 without dead trees, and (e) the simulated forest stands with dead trees from 1997-2011. 98
Figure 4.8. Comparison of the GNA_{EAB} simulated spatial networks of the (a) EAB dispersal SN_{EAB} from year 1997 to 2011 without forest stand death and (b) presented per each county, with (c) the observed real-world spatio-temporal patterns of EAB dispersal. 99
Figure 4.9. Degree distribution *P(k)* as measured for the SN_{EAB} structure for each year. 102
Figure 4.10. The change in (a) the average clustering coefficient <*C*> and (b) average path length <*l*> from 1997-2011. 103
Figure 5.1. Network-agent based model (N-ABM) approach applied for emerald ash borer infestation case study 116
Figure 5.2. Overview of the geospatial N-ABM agents' interactions, processes, and scheduling for the simulation of dynamic spatial networks with the map of the study area in Oakville, Canada. 117
Figure 5.3. Programming logic and pseudocode developed for the N-ABM approach that couple agent interactions and the generation of the dynamic spatial network structures. 124
Figure 5.4. Variation of the degree distribution *P(k)* generated across all 50 model runs. The degree distribution presents the number of trees with degree *k.* 129
Figure 5.5. The N-ABM simulation results depicting spatial EAB infestation extent as a function of EAB and ash tree agents' interactions in the Town of Oakville between 2008 and 2009 for: (a and b) July 2008 at t61; (c an d) August 2008 at t92; (e and f), July 2009 at t426; and (g and h) August 2009 at t457. 130
Figure 5.6. Degree distribution *P(k)* for (a) k_{out} and (b) k_{in} and degree distribution *P(k)* plotted on a log-log scale for (c) k_{out} and (d) k_{in}. 132
Figure 5.7. Spatial distribution of tree node degree for (a) k_{out} and (b) k_{in} demonstrating highly connected trees and less connected trees for EAB insect infestation. The simulated high degree trees correspond to the (c) regions of real-world high severity infestation. 133
Figure 5.8. Spatial distribution of varying clustering coefficients of tree nodes. The inset map corresponds with the red dot on the main map and presents a node exhibiting the star like pattern commonly generated during the EAB host selection process. 136
Figure 6.1. The Epi-N-ABM model structure 158
Figure 6.2. Map of (a) locations of homes, businesses, and schools that agents travel between and the simulated social spatial network SN_{social} as it evolves over time for time steps (b)t_1, (c)t_5, (d)t_8, (e)t_9, (f) t_{10}, and (g) t_{11}. 170
Figure 6.3. The social network SN_{social} in a non-spatial representation for time steps t_1 (a), t_5 (b), t_8 (c), t_9 (d), t_{10} (e), and t_{11} (f) 171

Figure 6.4. Map of (a) spatial location of homes, businesses, and schools that the agents travel between and the simulated infection network $SN_{infected}$ as it evolves over time for time steps (b) t_{168}, (c) t_{1176}, and (d) t_{6216}. 175

Figure 6.5. The infection network $SN_{infected}$ in a non-spatial representation for time steps (a) t_{168}, (b) t_{1176}, and (c) t_{6216}. 176

Figure 7.1. Geographic network automata in relation to other geographic automata systems models. This situates N-ABM and other integrated approaches into the larger context. 200

Chapter 1.

Introduction

Real world geographic phenomena from ecological to social are increasingly conceptualized from the bottom-up and analyzed using a complex systems approach (Thrift, 1999). A complex systems approach aims to explore how small-scale interactions between individual system components across space and time generate system-level behavior and spatial patterns at much larger scales, referred to as emergence (Cilliers, 1992). For example, small scale interactions between insect pests and their hosts generate large scale spatial patterns of infestation. In another example, social interactions between individuals infected with influenza form large scale patterns of virus propagation. A complex systems approach acknowledges that real systems are typically non-linear and thus cannot only be understood by the sum of their parts (Manson, 2001; Rietsma, 2002). The non-linearity in real systems is a function of the behavior of the system components, shaped by their unique characteristics and objectives as well as feedback from other components within the system, their environment, or the system in its entirety. Variations in behavior and dynamics at the local level is often amplified at the system level, making these systems challenging to understand (Manson, 2001).

Researchers model real-world systems in order to better understand them (Cilliers, 1999). Conventionally under the scientific framework of reductionism, real-world systems have been simplified and distilled down to their individual components in order to better understand their behavior. This traditional bottom-up view influenced the design of models that represent real-world systems as closed, balanced, and homogenous (Batty & Torrens, 2005). Complex systems theory offers an alternative to reductionism by focusing on interaction between system components, spurring the development of new types of models to better understand real phenomena by addressing that complexity emerges from system openness, non-linearity, and heterogeneity (Thrift, 1999; Wilson 1994; Batty & Longley, 1995).

More recently, complex geographic systems are represented using geographic automata systems (Torrens & Benenson, 2005). In the same way that complex systems theory conceptualizes geographic phenomena from the bottom-up, geographic automata

provide an explicitly spatial framework to represent and analyze systems from the bottom-up by discretely representing the driving mechanisms observed at the local level in the system from which system-level behavior emerges. The modelling approach cellular automata (CA) represents local dynamics between system components (White & Engelen, 2000) while another approach, agent-based modelling (ABM), represents interactions between individuals or "agents" (Torrens, 2010). Geographic automata offer an alternative to top-down statistical or mathematical models that use equations and mathematical functions to capture system level processes and as such cannot capture the complexity resulting from individual-level heterogeneity and adaptation, local spatial and temporal variations, and subsequent non-linearity inherent to real-world complex systems (Parunak, 1998; Railsback, 2001; Torrens, 2010). These modelling approaches are coupled with geographic information systems (GIS) and geospatial data sets and have been used in application to a variety of real geographic phenomena with early studies in urban (White & Engelen, 1993), social, and ecological systems (DeAngelis & Mooij, 2005). Flexible in their model design, geographic automata can allow for easy manipulation of parameters and scenario building and provide useful contributions to decision-making processes.

Network theory also seeks to represent and analyze systems using a bottom-up complex systems approach. In these network representations, nodes represent system components and links represent the discrete spatial interactions between them. Network representations are particularly advantageous because networks can be measured using graph theory, a branch of mathematics, that mathematically formalises network conceptualizations of the real world to graphs (G), which are made up of a collection of nodes (N) connected links (L) and defined as $G = \{N, L\}$ (Borner et al., 2008). Graphs have a structure that can be solved mathematically to aid in the understanding of the system it represents (Barabasi, 2014). Graphs are traditionally arranged based on their topological relationships. For example, in a graph representing a social network, each node represents an individual and each link represents a friendship between two nodes. Nodes in an aspatial representation of this social network will be arranged topologically based on the strength of their relationships. However, more recently, there is interest in arranging graphs based on the spatial location of their nodes. For example, by embedding that same social network in geographic space, nodes represent the spatial location of individuals and the links represent the friendship between the two individuals.

Graph theory can then be used to characterize the spatial structure of the social network.

The characterization of the spatial structures of urban, social, ecological, or geophysical networks is vital as this information can be used to understand the dynamics and processes occurring on or within that network. Based on this idea, networks are used as inputs in geographic automata where simulated system dynamics operate as a function of an underlying network structure. For example, airline networks, street networks, and origin-destination networks are incorporated as inputs in ABMs to better simulate dynamics of human mobility which consequently has implications for the spread of disease, flow of information, and traffic and congestion (Torrens, 2010; Dibble and Feldman, 2004; Balcan et al. 2009; Burger et al., 2017, Pires & Crooks, 2017).

The relationship between the spatial network structure and the spatial dynamics and processes occurring on or within that network is tightly coupled and can be considered non-linear. More specifically, in the same way that the spatial structure of a system influences the system's dynamics and processes, these dynamics and processes over time change the system's spatial structure. For example, the spatial structure of a transportation network plays a significant role in the patterns of human mobility and in turn, patterns of human mobility may alter the structure of the transportation network over time. The change in a system's spatial structure over time as a function of its network dynamics is typically referred to as network evolution or as Barthelemy (2018) recently termed “network morphogenesis”. Network evolution has been operationalized using an aspatial modelling approach referred to as network automata (Sayama and Laramee, 2009; Smith et al., 2011); however studies exploring this modelling approach do not account for the spatial properties of networks, do not examine real-world networks, nor do they use geospatial data.

The fields of geographic information science (GISc) and geographic automata systems as well as network science have a common interest in both exploring and better understanding the linkage between the spatial structures of geographic systems and their tightly coupled dynamics and processes. Batty (2005, pg. 3) summarizes this from a GISc perspective, stating:

> “Traditional GIS is largely a-temporal, representing locational structures at a single point in time. In so far as time has been involved,

> such systems simply represent a series of cross-sections in comparative static manner, with little functionality or science being developed to deal with processes that link these cross-sections through their temporal evolution".

Furthermore, he suggests that part of the problem may be that GISc models are challenged with the task of embracing spatial relations, interactions, and connectivity and this poses a significant barrier to its continued development. Although network science inherently represents interactions and connectivity between system components within larger complex spatial systems and offers the theoretical basis and tools for other fields like GISc to follow suit, the field of network science also struggles with linking static network structures observed at various points in time with the spatial processes that drive their changes.

Exploring the tight-coupling between network structures and the processes occurring on and within such structures requires a disaggregated approach that acknowledges the complexity inherent to these systems. Specifically, the local dynamics and processes that operate as a function of the underlying spatial structures subsequently spurring the evolution of these structures would need to be recorded and measured. In most cases, field surveys are not extensive enough, nor is data at a fine enough resolution to capture and explore this tight-coupling between spatial structure and dynamics. Instead, computational modelling can be leveraged. Barthelemy (2018) calls for a new modelling framework for studying and testing the interplay between the evolution of network structures as a function of their dynamics and processes and vice versa. However, the development of an entirely new modelling framework is not necessary since the integration of networks with well-developed complex systems modelling approaches would inherently meet those modelling needs, providing the potential to analysis of complex geographical systems while leveraging tools offered by the fields of complex systems theory, GISc, and network science.

This integration of networks and geographic automata also has the potential to overcome traditional challenges faced in complex systems modelling. Geographic automata models are typically heavily parameterized, contain large numbers of heterogeneous and adaptive agents, and incorporate stochastic elements and processes. As a result, it can be challenging to fully understand the link between local processes and emergent behaviour, even in well-developed geographic automata (Topping et al., 2010). Approaches such as sensitivity analysis and robustness analysis

aim to shed light on this link by exploring how changes in model input correspond to changes in model output; however, the complexity of geographic automata limits the ability to obtain a complete understanding of the model (Broeke et al., 2016). As such geographic automata models are referred to as ‘black boxes' and thus modellers face challenges in model analysis, testing, replicability, and communication. Networks and geographic automata can be integrated to form a new framework to characterize and better understand geographic model internal behavior and emergent simulated patterns. In this case, networks are not implemented as a model input for which to inform system dynamics in the model, but rather, the complex system dynamics simulated by the geographic automata are abstracted as a network representation that can then be measured and quantified. This information can provide an opportunity to open the ‘black box' and form additional means for understanding, analyzing, and communicating geographic automata models and the complex spatial systems they represent.

1.1. Research Problems and Questions

This dissertation examines the integration of complex systems theory, geographic information science (GISc), and network theory and seeks to leverage the advantages of all three scientific fields of research for the conceptualization, representation, and analysis of both geographic automata and the complex spatial systems they represent. Specifically, this dissertation addresses the following research questions:

1. To what extent can the integration of network theory and geographic automata systems be leveraged for the development of theoretical frameworks for network-based automata modelling approaches?

2. How can the theoretical frameworks for network-based automata models be implemented for the representation and analysis of real spatial-temporal complex phenomena at a variety of spatial scales?

3. To what extent can network theory be used to enhance the testing of geographic automata models?

1.2. Research Objectives

In order to answer the proposed questions, this dissertation integrates concepts of GISc and complex systems approaches with network theory for the development of

novel network-based automata modelling approaches that can be used to represent and understand a variety of complex dynamic spatial systems. The main objectives of this dissertation are:

1. Development of a theoretical framework for a novel network-based automata modelling approach called Geographic Network Automata (GNA) for the representation and analysis of complex spatial systems as evolving networks.

2. Implementation of the proposed GNA modeling approach to simulate and analyze ecological phenomena at a large spatial scale using the case study of the emerald ash borer (EAB) forest insect infestation.

3. Development and implementation of a network-based ABM (N-ABM) modeling approach applied to the case study of the EAB forest insect infestation at a fine spatial scale.

4. Leverage spatial network analysis for the development of a robust NEtworks for Agent-based model Testing (NEAT) approach and implement the NEAT approach on an epidemiological network-based ABM (Epi-N-ABM) as a case study.

This dissertation meets these objectives using a variety of contexts, ranging from ecological to social.

1.3. Dissertation Overview

The five chapters following the Introduction address the objectives of this dissertation. Chapter 2 offers a comprehensive review of the literature, exploring in detail the intersection of GISc and network science. This synergy between the two fields paves the way for the use of spatially explicit networks to represent and understand complex spatial systems. This chapter also offers an introduction to graph theory, the mathematical foundation for the development of the programming functions implemented in all subsequent chapters that are used to measure and characterize spatial network structures.

Geographic network automata (GNA), a novel modelling approach developed for the simulation of complex systems as evolving spatial networks is presented in Chapter 3. In the same way that CA simulates dynamics between neighbouring cells and ABMs simulate interactions between discrete autonomous agents, GNA simulates local spatial interactions between georeferenced nodes using links from which a spatial network

structure emerges and evolves. In this modelling approach, the tight coupling between spatial network structure and the dynamics and processes occurring on and within these networks can be examined. The GNA modelling approach is well situated under the broader modelling framework of geographic automata systems by using a network automata modelling approach that is explicitly geospatial. The framework is implemented using a spatial network representation of Conway's Game of Life (Conway, 1970), where the implemented transition rules simulate dynamics between nodes at the very local level, altering the structure of the spatial network, which in turn influences the dynamics between nodes. Graph theory measures are used to characterize the structure and behavior of simulated networks.

Chapter 4 presents a study of the application of the GNA modelling framework to the real-world case study of the emerald ash borer (EAB) forest insect infestation at a large spatial extent. The GNA implements real geospatial datasets representing forest stands that have been surveyed to contain ash trees in the state of Michigan, US. These forest stands serve as habitat patches that sustain EAB populations. In the GNA model, these forest stands are represented as nodes and dispersal between nodes are represented using links, forming an evolving network of EAB spread across Michigan from 1997 to 2011. As the infestation worsens, the network structure evolves as new forest stand nodes are added to the network and as long infested forest stands die and are removed from the network. The evolving network's spatial structure can be measured using graph theory to better understand how in combination with landscape connectivity influences the dynamics of the spread of EAB at a large regional scale.

Given that the work so far simulates and analyses patterns of EAB infestation at a coarse spatial resolution (Chapter 4), there is a need to examine network structures representing EAB dynamics at finer spatial scales. Chapter 5 focuses on the integration of networks and ABM modelling approaches, providing new means to analyze ABMs and the complex phenomena they represent. Specifically, Chapter 5 uses a network-based ABM (N-ABM) to examine EAB behavior and the resulting patterns of spread as represented by spatial networks at very fine spatial and temporal scales in the Town of Oakville, Canada from 2008 to 2009. The proposed N-ABM modelling approach differs from the GNA modelling approach in that dynamics governing the emergent patterns of spread are a function of programming complex behaviors and decision making of individual EAB agents. The network component of the N-ABM abstracts the dynamics,

processes, and emergent spatial patterns simulated by the ABM as networks, allowing for the application of graph theory to measure this behavior. This N-ABM differs from the traditional use of networks as inputs in ABMs from which agent behavior and dynamics are influenced. Selected graph theory measures have been used to analyze the simulation outcomes.

The inclusion of networks in geographic automata modelling approaches both requires and offers potential new means for model testing. Chapter 6 presents a NEtworks for ABM testing (NEAT) that focuses on the testing of internal model processes and simulated spatial networks patterns. The approach is applied to an epidemiological network-based ABM (Epi-N-ABM) that simulates the transmission of the influenza virus through a population in Vancouver, Canada. Dynamics simulated in the ABM are abstracted as both spatio-temporal contact and infection networks that can be measured using graph theory and compared with observed properties of real contact and infection networks.

All of the models as mentioned above were developed and implemented using the Java programming language in the Eclipse integrated development environment with a free and open source simulation platform called Repast Simphony (Repast Simphony, 2016, 2017, 2018), capable of working with geospatial data. The ABMs were executed on a Compute Canada, West Grid facility for parallel computing. Visualization of the model outputs was facilitated by a variety of proprietary and open source software including ArcGIS, Idrisi, and Gephi. The concluding chapter of this dissertation summarizes the general findings, offers insight into potential future research, and discusses the contributions to the intended fields of study.

1.4. References

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases*. Proceedings of the National Academy of Sciences,* 106(51), 21484-21489.

Barabâsi, A.-L. (2014). *Linked: The Science of Networks and How Everything is Connected to Everything Else and What it Means.* New York, NY: Plume.

Barthélemy, M. (2011). Spatial networks. *Physics Reports,* 499(1), 1-101.

Barthélemy, M. (2018). *Morphogenesis of Spatial Networks.* Cham, Switzerland: Springer International Publishing.

Batty, M. & Longley, P. (1995). *Fractal Cities.* London, UK: Academic Press.

Batty, M. (2005). Network geography: Relations, interactions, scaling and spatial processes in GIS. *Re-presenting GIS*, 149-170.

Batty, M., & Torrens, P. M. (2005). Modelling and prediction in a complex world. *Futures,* 37(7), 745–766.

Borner, K., Sanyal, S., & Vespignani, A. (2008). Network Science. *Annual Review of Information Science and Technology,* 41(1), 537–607.

Broeke, G.T., Van Voorn, G., & Ligtenberg, A. (2016). Which sensitivity analysis method should I use for my agent-based model?.*Journal of Artificial Societies & Social Simulation*. 19(1).

Burger, A., Oz, T., Crooks, A.T. and Kennedy, W.G. (2017), Generation of Realistic Mega-City Populations and Social Networks for Agent-Based Modeling. *The Computational Social Science Society of Americas Conference*, Santa Fe, NM: ACM.

Cilliers, P., & Spurrett, D. (1999). Complexity and post-modernism: Understanding complex systems. *South African Journal of Philosophy,* 18(2), 258-274.

Conway, J. (1970). The game of life. *Scientific American,* 223(4), 4.

DeAngelis, D. L., & Mooij, W. M. (2005). Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution, and Systematics, 36,* 147–168.

Dibble, C., & Feldman, P. G. (2004). The GeoGraph 3D computational laboratory: Network and terrain landscapes for RePast. *Journal of Artificial Societies and Social Simulation*, *7*(1).

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269-271.

Manson, S. M. (2001). Simplifying complexity: a review of complexity theory. *Geoforum,* 32(3), 405-414.

Parunak, H. V. D., Savit, R., & Riolo, R. L. (1998). Agent-based modeling vs equation-based modeling : a case study and users' guide. *Multi-Agent Systems and Agent-Based Simulation*, Berlin, Heidelberg: Springer.

Pires, B., & Crooks, A. T. (2017). Modeling the emergence of riots: A geosimulation approach. *Computers, Environment and Urban Systems*, *61*, 66-80.

Railsback, S. F. (2001). Concepts from complex adaptive systems as a framework for individual-based modelling. *Ecological Modelling,* 139(1), 47–62.

Reitsma, F. (2003). A response to simplifying complexity. *Geoforum,* 34(1), 13-16.

Repast Simphony. (2016). Version 2.4. [Computer Software]. Chicago, IL: University of Chicago.

Repast Simphony. (2017). Version 2.5. [Computer Software]. Chicago, IL: University of Chicago.

Repast Simphony. (2018). Version 2.6. [Computer Software]. Chicago, IL: University of Chicago.

Sayama, H., and Laramee, C. (2009). Generative network automata: A generalized framework for modeling adaptive network dynamics using graph rewritings. *Adaptive Networks,* 311–332.

Smith, D. M., Onnela, J. P., Lee, C. F., Fricker, M. D., & Johnson, N. F. (2011). Network automata: Coupling structure and function in dynamic networks. *Advances in Complex Systems*, 14(03), 317-339.

Thrift, N. (1999). The Place of Complexity. *Theory, Culture & Society, 1*6(3), 31–69.

Topping, C. J., Høye, T. T., & Olesen, C. R. (2010). Opening the black box—Development, testing and documentation of a mechanistically rich agent-based model. *Ecological Modelling,* 221(2), 245-255.

Torrens, P. M. (2010). Agent-based models and the spatial sciences. *Geography Compass,* 4(5), 428–448.

White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and planning A,* 25(8), 1175-1199.

Wilson, A.G. (1994). *Modelling the City*. London, UK: Routledge.

Chapter 2.

Overview of Network Theory and its Geospatial Applications

2.1. Abstract

Geographic phenomena can be conceptualized and modelled as a complex system where spatial patterns and processes emerge from local interactions between individual components of the system. Complex systems from ecological, urban, social, to geophysical typically endure as non-linear processes that make these systems difficult to model, analyze, and understand, particularly using top-down traditional statistical methods. Both geographic information science (GISc) and network science acknowledge the value in using a decentralized approach to not only represent, but also characterize and analyze complex spatial systems. Despite their complementary nature, approaches that integrate GISc and network science are relatively limited, especially in the representation and analysis of spatial systems in contexts that include both geographic space and time. The objective of this study is to provide a comprehensive review of the intersection of GISc and network science application to geographic phenomena. This study argues that their intersection naturally pushes both fields forward in their combined interest to better represent and understand spatio-temporal phenomena and the link between spatial process and structure.

2.2. Introduction

Geographic Information Science (GISc) is concerned with the fundamental issues surrounding the use of digital technology to collect, store, manage, represent, model, and analyze geographic information (Wright, 2010). Network science is defined as the application of graph theory to problems in a variety of fields (Barabasi, 2016). Although strikingly different by definition, these two fields of science share largely unrevealed commonalities in the way in which they conceptualize and represent the real world. For one, both fields acknowledge the value in organizing real systems under the lens of complex systems and thus, often conceptualize real world phenomena from the bottom-up, in contrast to the top-down way such systems are traditionally conceptualized

(Holland, 1996). This complex systems approach seeks to understand how heterogeneous components i.e. people, institutions, animal species, or financial transactions “interact” or “connect” and self-organize to form emergent systems including societies, cities, ecosystems, and economies. Second, GISc and network science have an aligned interest in shifting from describing the static structure of systems as a whole to understanding and representing the process of change occurring on and within such structures (Batty, 2005). This requires the inclusion of time, not as a set of snapshots of the static system from several time periods, but more specifically to link the changes observed in the system over time with the system’s underlying processes. Finally, and potentially most unifying, is the common interest in the explicit representation of geographic space as a driver of these structures and the tightly-coupled processes that shape them over time.

Early studies that sought to characterize the structure of networks were not necessarily concerned with their geographic properties and thus studies that leveraged both GISc and network science were limited (Barthélemy, 2011). It was not until the 1970s when quantitative geographers began using spatial networks to characterize the spatial structure and evolution of transportation networks that the interest in spatial properties of networks moved to the forefront (Barthélemy, 2018). Researchers observed that rooting networks in geographic space provides additional and important information that may help to explain the behavior of spatial systems. For example, in a topological social network, individuals are represented as nodes and the strength of relationships between two nodes are represented by their length. In this aspatial representation, close relationships will have shorter links than acquaintances. Since this type of network does not capture physical interactions between individuals in geographic space, this representation is not useful for understanding the spread of disease. Instead, a spatial representation of a social network, where links represent physical contact between individuals can be used (Rahmandad and Sterman, 2008). The use of spatial network analysis in regional geography in the 1960-70s ultimately paved the way for first formal intersection of GISc and network science, sparking research in spatial networks that would later make contributions to transportation geography, landscape ecology, epidemiology, biology, and the development of methods for complex spatial systems modelling.

The objective of this study is to critically analyze the way in which GISc and network science thread together as the underpinnings for the use of spatial networks to represent and understand complex spatial systems. This study argues that the development of approaches that integrate GISc and network science meets the common interests of both fields and in turn allows for the quantification of complex spatial structures, provides evidence that spatial networks structures have implications for system processes, seeks to better understand dynamic system properties, facilitates new methods for spatial analysis, and most challenging of all, explores the evolution of network structure as a function of network dynamics (Phillips et al. 2015). The latter has been facilitated by increased data availability, computational power, and sophisticated spatially explicit modelling. This research study first provides an overview of traditional network science theory, next, comprehensively reviews a selection of literature regarding network science in a spatial context, and lastly, finishes by proposing future directions for network science in GISc.

2.3. Theory behind Network Science

Although the first formal intersection of the spatial sciences and network science is marked by the use of spatial networks by quantitative geographers in the 1960-70s, the first known application of graph theory (Euler, 1741), 200 years earlier, can also be considered as an inherently spatial problem. Famously referred to as the “Seven Bridges of Konisberg”, mathematician Leonard Euler represents the islands of Konisberg as nodes and the bridges between them as links, forming a network. Using graph theory, he proves that it is impossible to walk across all seven bridges connecting the islands and never cross the same one twice, what was later more commonly referred to in network science and GISc as a routing problem. He concludes that finding (or not finding) the path is not a matter of intelligence; it is simply an inherent property of the graph. Euler’s findings formed the foundation of network science and gave way to the notion that graphs have properties and structure and graph theory can be used to discover and catalogue these properties. Following in the footsteps of Euler’s findings, early spatial network analysis of real complex spatial systems focused on describing and cataloguing the properties of spatial networks using graph theory.

2.3.1. Descriptive Graph Theory Measures

Spatial networks *SN* are mathematically represented as graphs *G* composed of *spatially embedded* nodes *N* that are connected by links *L*, meaning the node's location can be identified by a pair of coordinates. The structure of a graph can be described and cataloged using a variety of developed graph theory measures. These measures can be found in varying degrees of detail in just about every network science review paper (Newman, 2003, Barthélemy, 2011, Lewis, 2011, Barabasi et al, 2016). Based on the scope of this review, this section *briefly* defines a few important graph measures that are useful for characterizing networks. A simple spatial graph *G* (Figure 2.1) is used to illustrate these measures.

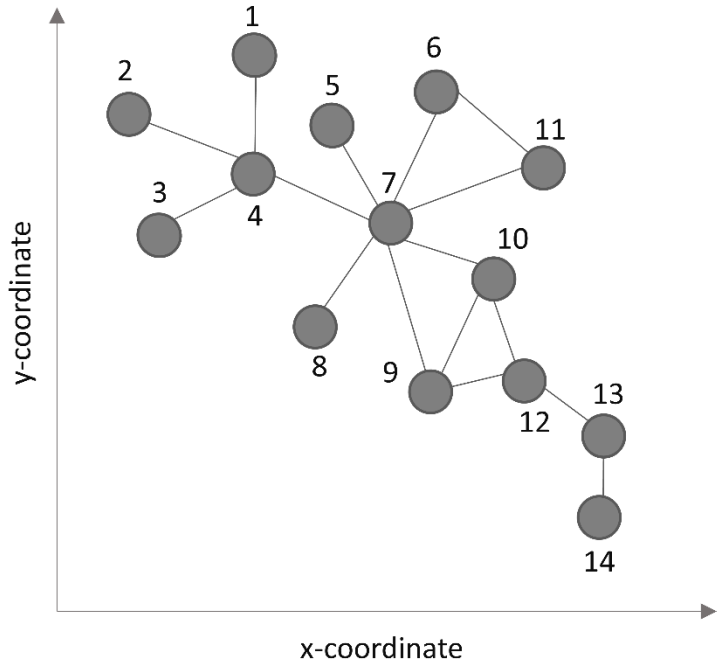

Figure 2.1. A simple spatial graph *G* to illustrate graph notations and definitions.

Graphs are defined as *G = {N, L}* where *N* are the set of nodes and *L* are the set of links that make up the graph. A node *v* is referred to by its order *i* in the set *N*. A link *e* joins nodes v_i and v_j. In Figure 2.1, for graph *G,* the number of nodes n = 14 and the number of links m = 16. Graphs can be undirected *G* or directed *G'*, unweighted or weighted. Social networks are most often abstracted as undirected graphs since each relationship between two people is typically reciprocated. In contrast, flows in a power line network may be unidirectional. In a weighted graph, weight *w* typically represents

the volume of people, objects, or information that “flows” between nodes (Barrat et al., 2004). Graphs can also be bipartite, meaning nodes can be divided into two classes U and V where each link connects a U-node to a V-node and vice versa (Barabasi, 2016).

The topology of a graph can be represented using an adjacency matrix (Figure 2.2). This is an *N x N* matrix *A* where $A_{ij} = 1$ if nodes v_i and v_j connect, and 0 otherwise. If the graph is directed, the adjacency matrix will be asymmetrical. If the graph is undirected, the adjacency matrix will be symmetrical. For example, for the undirected graph in Figure 2.1, the adjacency matrix would appear as follows:

A =

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4	1	1	1	0	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	1	0	0	0	1	0	0	0
7	0	0	0	1	1	1	0	1	1	1	1	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	0	1	0	0	1	0	1	0	0
10	0	0	0	0	0	0	1	0	1	0	0	1	0	0
11	0	0	0	0	0	1	1	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	1	1	0	0	1	0
13	0	0	0	0	0	0	0	0	0	0	0	1	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Figure 2.2. Adjacency matrix A for an undirected graph G.

Common graph measures were developed to characterize the topological structure of a graph and thus are a function of the information stored in the adjacency matrix. Common graph measures (Barthélemy, 2011) include the following:

Degree

The degree of a node *k* is a local measure specific to an individual node and is defined as the sum of links e_j connected to node v_i:

$$k = \sum_j A_{ij} \qquad (1)$$

where the higher the degree, the more connected a node is. For example, for the simple spatial graph *G* (Figure 2.1), the degree *k* of node $v_1 = 1$ and the degree *k* of node $v_{10} = 3$.

Average Degree

The average degree <*k*> is a global measure that calculates the average number of links of nodes across a non-directed graph, defined as:

$$< k > = \frac{1}{N \sum_i k_i} = 2E/N \qquad (2)$$

where twice the number of links (directed both ways) is divided by the number of nodes. In the example of spatial graph *G* (Figure 2.1), the average degree <k> is two, meaning that on average, each node is connected to two other nodes. The degree distribution is useful for providing an overall snapshot of the structure of the network.

Degree Distribution

The degree distribution *P(k)* is the fraction of nodes in a graph with degree *k.* For example using the example graph *G* in Figure 2.1, the fraction of nodes in a graph with degree *k* can be calculated by the number of the nodes with the same degree, divided by the total number of nodes in the graph *G*. The degree distribution *P(k)* can be plotted on a histogram to present the degree distribution of the graph (Figure 2.3). The degree distribution is useful for understanding whether all nodes in the graph have the same number of connections, or whether some have many connections and some only have a few connections.

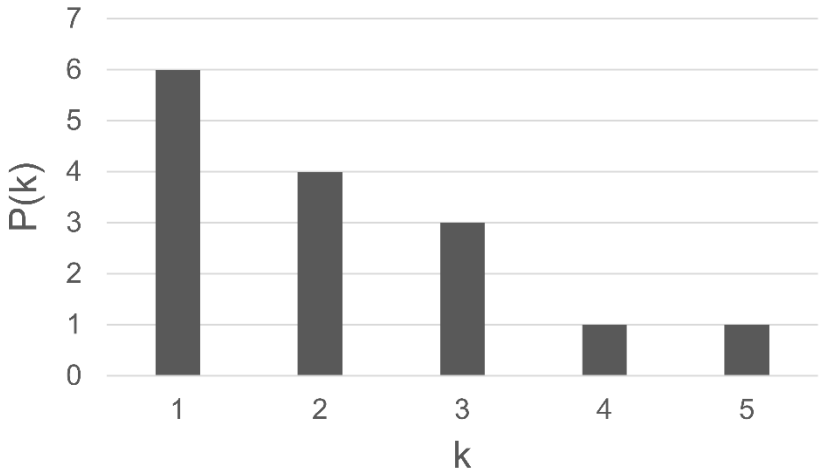

Figure 2.3. The resulting degree distribution for graph *G,* illustrated in Figure 2.1.

Clustering Coefficient

The clustering coefficient *C* measures "cliquishness" of an individual node v_i, commonly described as the probability that "friends" of *i* (i.e. nodes connected to v_i) are also friends of each other (Watts & Strogatz, 1998). When *C = 1,* the node v_i and the nodes connected to node v_i, also known as *neighbouring nodes*, are perfectly connected. In contrast, when *C = 0,* the node's neighbours are not connected at all. For node v_i of degree k_i, the clustering coefficient *C(i)* is defined as:

$$C(i) = \frac{2E_i}{k_i(k_i-1)} \qquad (3)$$

where E_i is the number of links among the neighbors of v_i. The clustering coefficient *C(12)* from the example graph *G* (Figure 2.1), is calculated by counting the number of neighboring nodes of v_{12}, that are connected to each other. In this example, node v_{12} has 3 neighboring nodes or $k_{12} = 3,$ and the number of connections between those neighboring nodes $E_{12} = 1$. Therefore, the clustering coefficient of *C(12)* is 0.166.

Average Clustering Coefficient

The average clustering coefficient <*C*> is a global measure that determines the cliquishness of all nodes across a graph and is calculated as the average *C* over all individual nodes. When <*C*>= *1,* the graph is perfectly connected and thus maintains a lattice like structure. In contrast, when <*C*> = *0,* the graph is not connected at all and

thus maintains a star like structure. The average clustering coefficient <*C*> for graph *G* is 0.26 (Figure 2.1).

Path

A path P_{i_i,i_n} is the collection of nodes and/or links between two nodes v_i and v_j. The collection of nodes and links in a path is defined as:

$$L_P = \{v_0, v_1, v_2, \dots\ v_n\} \qquad (4)$$

$$N_P = \{e_0, e_1, e_2, \dots\ e_m\} \qquad (5)$$

Shortest Path Length

There may be many paths of varying lengths l between two nodes. The shortest path length is calculated by finding all paths between two nodes and finding the shortest path. The shortest path length is calculated by counting the total number of intermediate nodes or links along the shortest path between two nodes and is denoted by l_s. The l_s measure is defined as:

$$l_s(i,j) = \min_{paths\ (i \to j)} \qquad (6)$$

Average Path Length

The size $< l >$ of a graph is the average length between all possible pairs of nodes in the network.

Diameter

The diameter of a graph d_G is the maximum shortest path length l_s.

Assortativity

Assortativity *ρD* measures whether the extent to which nodes that are connected are similar with respect to their graph measures (Newman, 2002). This measure most commonly determines the similarity of degree *k* between nodes. A pair of nodes are considered assortative if they have similar degrees. For example, in graph *G,* connected nodes v_6 and v_{11} both have a degree of 2 and thus are considered to be assortative. A pair of nodes are considered disassortative if the degrees are dissimilar. For example, in

graph *G,* node v_7 has a degree of 7 and v_5 has a degree of 1 and thus this connected pair is considered disassortative. This idea can also be applied to other characteristics such as node weight (i.e. high weight nodes connect to other high weight nodes). This measure can be calculated by looking at the correlation between all possible pairs of nodes in a graph with degrees' *k* and comparing this to the correlation between all possible pairs of nodes in a random network with degrees' *k.*

Betweenness Centrality

The importance of a node in a graph is measured by its centrality. Although there are many centrality measures, one of the most common measures is betweenness centrality g that calculates the total number of shortest paths between any two nodes in the graph that pass through node v_i and is defined as (Freeman, 1977):

$$g(i) = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \qquad (7)$$

where σ_{st} is the number of shortest paths going from node v_s to node v_t and $\sigma_{st}(i)$ is the number of shortest paths going from node v_s to node v_t through node v_i.

Similarly, the importance of a link e_j is measured by betweenness centrality g that calculates the total number of shortest paths between any two nodes in a graph that include link *j* and is defined as (Newman, 2003):

$$g(j) = \sum_{s \neq t} \frac{\sigma_{st}(j)}{\sigma_{st}} \qquad (8)$$

where σ_{st} is the number of shortest paths going from node v_s to node v_t and $\sigma_{st}(j)$ is the number of shortest paths going from node v_s to node v_t through link *j.* Betweenness centrality and degree of a node often correlate, where the most central node has a high number of links.

Weight

The consideration of weight of each node in a graph is important. For example, in a street network, vehicles may more frequently traverse some links which are thus

weighted more heavily. The weight w_i of node v_i is calculated by summing the link weights between two nodes (Barrat et al., 2004):

$$w_i = \sum_{j \in \Gamma(i)} w_{ij} \qquad (9)$$

Community

A community is defined as a set of nodes that have more connections among themselves than other nodes in the graph.

Transportation indices and Variants

Although geographic space may play a role in the topologies characterized by the above graph theory measures, these traditional measures were not designed specifically to measure spatial structure of real networks. As a result, spatial indices such as the gamma index *γ* and the alpha index *α* and their variants were defined by quantitative geographers in the 1960s initially for the characterization of the structure of spatial planar transportation networks (Garrison, 1960; Kansky, 1969). Some of these important indices are presented in Table 2.1.

Table 2.1. Transportation indices and variants

Gamma index γ_p	measures the density of a graph and can be defined by: $\gamma_p = \frac{E}{3N-6}$ where *E* is the number of edges and *N* is the number of nodes.
Alpha index α	a measure of meshedness (Buhl et al., 2006) as an index from 0 to 1 where 0 is a tree-like graph and 1 is a maximal planar graph (no "loose" nodes jutting out) and can be defined as: $\alpha = \frac{E-N+1}{2N-5}$
Completeness Υ_N variant	an index from 0 to 1 where 0 indicates a high number of dead ends or "unfinished" crossings and 1 indicates a more organized and complete city and can be defined as (Barthelemy, 2011): $\Upsilon_N = \frac{N(1)+N(3)}{\sum_{k\neq 2} N(k)}$ where $N(k)$ is the number of nodes of degree k. For example, an unfinished road or dead end will typically have a node degree of 3 or 1, where a finished crossing will have a node degree of 4.

Compactness ψ variant	measures the degree to which a city is filled with roads from 0 to 1 where 0 represents the extreme in a square city with one road encircling it and 1 represents a tight grid and can be defined as (Courtat et al., 2010): $$\psi = 1 - \frac{4A}{(l_T - 2\sqrt{A})^2}$$ where A is the area of the city and l_T is the total length of roads.

Distance Based Metrics

The distance between linked node pairs is an important measure for real networks embedded in geographic space. Distance can be measured in a variety of ways, the most common being Euclidean distance $d_E(i,j)$ or "as the crow flies" distance. In contrast, the route distance $d_R(i,j)$ is computed by summing the length of links which make up the shortest path between node v_i and v_j. There are several distance related functions which are summarized in Table 2.2 as defined in Barthelemy (2011).

Table 2.2. Distance based metrics

Route factor $Q(i,j)$	measures the "directness" between two linked node pairs can be computed. The route factor is defined as the ratio between route distance and Euclidean distance and the closer to 1, the more efficient the graph: $$Q(i,j) = \frac{d_R\,(i,j)}{d_E\,(i,j)}$$
Accessibility $< Q(i) >$	measures the accessibility for a node v_i where the smaller the value, the more accessible the node and can be defined as: $$< Q(i) > = \frac{1}{N-1}\sum Q(i,j)$$ where $Q(i,j)$ is the sum of route factors.
Straightness centrality $C^s(i)$	measures the ease of communication between two nodes when two points are connected through a straight path, defined as (Crucitti et al., 2006): $$C^s(i) = \frac{1}{N-1}\sum_{j \neq i} \frac{d_E\,(i,j)}{d_R(i,j)}$$
Length total l_T	calculated by summing the total lengths of all links in the graph and is defined as: $$l_T = \sum_{e \in E} d_E\,(e)$$ where $d_E(e)$ is the length of link e.

Cost *C*	important in some contexts, particularly in better understanding transportation networks. For example, adding links and nodes to a network an improve transport performance *P* and fault tolerance, but increase cost. Cost *C* can be defined as: $$C = \frac{l_T}{l_T^{MST}}$$ where l_T^{MST} is the length of a minimum spanning tree with the same number of edges (not efficient).
Transport performance *P*	measured as the minimum distance between all pairs of nodes, normalized to the same quantity, but computed for the minimum spanning tree and is defined as: $$P = \frac{< l >}{< l_{MST} >}$$
Efficiency *E*	measures how efficiently a real network can exchange information (Latora & Marchiori, 2001). The quantity is zero when there are no paths between the nodes and is equal to one for a complete graph and can be defined as: $$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{l(i,j)}$$ Efficiency can be used to measure the fault tolerance of a graph, for example, how the *E* changes when a node is removed.

2.3.2. Theoretical Network Structures

Real world spatial phenomena exhibit properties pertaining to one or more mathematical graph types: regular, random, small world, and scale free. Each graph type differs in structure, thus yielding a distinctly different combination of graph properties as measured by graph theory measures. In a regular graph, every node is connected to the exact same number of their most adjacent nodes. Since it is likely that the neighbouring nodes of node v_i are also neighbours, this graph type is characterized by a high clustering coefficient (Watts & Strogatz, 1998). Because of their uniform and highly connected structure, regular graphs have a long average shortest path length, meaning that the minimum number of links separating any two nodes in the graph is long. Real world spatial systems rarely take on the structure of a regular graph, although they offer a useful baseline for comparison to other types of graphs (Lewis, 2011).

Some spatial phenomena may instead exhibit properties of random graphs (Erdos & Renyi, 1959; 1960). Random graphs are formed by taking a set of nodes and randomly generating links between them based on the same probability *p*, resulting in a

Poisson degree distribution characterized by a peak and rapidly diminishing sides (Boccaletti et al., 2006). Some nodes may have more connections than others, but the difference is marginal. In contrast to regular graphs, random graphs have a very short average path length </> and a low average clustering coefficient <C>.

Most real-world spatial networks, however, tend to produce properties that fall between regular and random graph structure, referred to as small world graphs. Small world graphs are modelled by randomly re-wiring regular graphs i.e. removing a connection between two neighboring nodes and forming connections between non-neighboring nodes or long-range connections (Watts & Strogatz, 1998). A regular graph with just a few long-range connections reduces the average path length of the graph dramatically, but maintains a high clustering coefficient (Watts & Strogatz, 1998). Small world graphs contain nodes with varying probabilities of forming connections to other nodes (Sole & Valverde, 2004). For example, a few nodes may characteristically have a denser connectivity than the rest of the nodes, forming the center of local clusters and maintaining connection to distant nodes and generating a very low average path length. Social networks exhibit properties of small world graphs, consisting of several tightly knit clusters, representing close ties such as family and friends, loosely connected to one another by a few weak ties such as acquaintances (Granovetter, 1973). The short average path length of social networks was depicted in Stanley Milgram's work in the late 1960s when he discovered that any two people are separated by just less than six intermediaries, commonly referred to today as the *six degrees of separation* phenomenon (Milgram, 1967).

In fact, six degrees is probably an overestimate, since Milgram's subjects cannot possibly know the *shortest path* between the starter and the target. If anything, social networks are even smaller. The internet, social media, and globalization removes physical distance as a barrier to social networking, potentially challenging Tobler's First Law of Geography, where "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). In addition to social networks, networks of food webs (Williams et al., 2002), molecules in a cell (Fell & Wagner, 2000), scientist co-authorships (Barabasi et al., 2002), and the neurons in the brain of a monkey and cat (Bassett & Bullmore, 2006) have been found to exhibit properties of small world graphs. The graph structure with it short path length allows for a faster propagation of information, disease, and power (Watts & Strogatz, 1998).

In the analysis of the graph structure of the web, Albert et al. (1999) used a bot, a computer program that automatically navigates through the web and collects information, to explore the web's connectivity. This led to the discovery that only a few websites had a very large number of links while the majority of websites had only a few. These highly connected nodes, in this case websites, are called connectors or hubs (Decruet & Beauguitte, 2013). A network with a large fraction of nodes with a small number of links and a small fraction of nodes with a large number of links generate a distribution that follows a mathematical expression called power law (Barabasi & Albert, 1999). A power law is expressed as:

$$p(k) \sim k^{-a} \qquad (10)$$

where the probability *p* of observing a node with *k* connections is the number of connections *k* to the negative exponent α, representing the ratio of highly-connected nodes exist to nodes that are less connected. This exponent is unique to each phenomenon and is usually a number between 2 or 3. Networks with these properties are referred to as a scale free network.

2.4. Graph Theory to Measure the Structure of Complex Spatial Systems

Graph theory is used to describe observed *static* spatial network structures including transportation, social, ecological, and geophysical networks either at one point in time or across several consecutive points in time as the network structures evolve. Characterizing real spatial network structures is useful because the network structure can help to inform about the processes that take place on the networks. Specifically, network structure provides insight into the flow of people, materials, vehicles, and information between nodes. In all of the areas of application discussed below, a shift is made from traditional topological network representation and analysis to spatial network representation and analysis. This ties together geography, GISc and network science and in the end contributes to each field. This coupling between the fields is sometimes difficult to pull apart. GISc and geography provide the context, in many cases the technology and the data, and methods for spatial analysis while network science provides the representation and additional analysis.

2.4.1. Spatial Transportation Networks

The traditional space syntax for transportation networks varies, however in the interest of understanding their spatial properties, it is important that the topological and Euclidean space are identical (Batty, 2004). As a result, transportation networks like street networks are typically constructed in *planar space* where the network links represent the physical course of routes and the nodes represent the location of their associated stations, intersections, or end points. In the case of transportation networks that are *not planar* such as airlines, links *do not* represent the physical course of routes, but rather simply indicate a connection between end points exists.

The use of graph theory to mathematically characterize transportation networks is extensive ranging from street networks (Garrison, 1960; Kansky, 1969; Crucitti et al., 2006; Cardillo et al., 2008; Jiang, 2007, 2009; Lammer et al., 2006; Strano et al., 2013; Buhl et al., 2006), rail networks (Kurant and Thiran, 2006; Latora and Marchiori, 2002; Han and Lui, 2009; Xu et al., 2007), cargo ship networks (Kaluza et al., 2010; Ducruet and Notteboom, 2012) and airline networks (Guimera & Amaral, 2004; Guimera et al., 2005; Li and Cai, 2004; Barrat et al., 2005; Jia and Jiang, 2012; Han et al., 2009; Liu et al., 2010; Zhang et al., 2010) to name a few. The structure of these transportation networks is constrained by geographic space, often limiting the degree of each node. As a result, planar transportation networks have a very low average node degree $<k>$ ranging from 2 or 3 for rail-based transportation networks and 4 or 5 for street networks (Han & Liu, 2009). Non-planar transportation networks are also constrained by geographic space, although to a lesser degree, since an airport, particularly domestic airports, or cargo ship bay can only physically handle a limited number of connections (Amaral et al., 2000). Although node degree is heterogeneous, the development of hubs for any transportation is both challenging as a function of geographic space and thus costly. In addition to node degree, the structures of transportation networks are limited by the cost of connecting *distant* nodes. Building new roads, railways, or scheduling flights to distant cities cost money and time. This limits the number of long-distance links and results in local clustering between nodes that are in closer proximity. Thus, many types of transportation networks rely on other transportation networks to increase efficiency in the movement of people, objects, and vehicles from point A to B.

The structural properties of transportation networks is used to evaluate the efficiency of such networks (Jiang, 2009), understand the relationship between network structure and socio-economic characteristics (Jia & Jiang, 2012), assess overall vulnerability or resiliency of the network structure to node failure (Latora and Marchiori, 2002; Wilkenson et al., 2012), and identify the most optimal route for travel based on the shortest-path problem (Dijkstra, 1959), the maximum flow problem (Ford & Fulkerson, 1962), or the minimum-cost problem (Hillier & Lieberman, 1990) to name a few. In the above-mentioned applications, the concept of the "flow" is important, where the structure and subsequent changes to network structure over time has implications for the movement of people and transportation along the network. In the late 1990s, network science and GISc were integrated to develop spatial algorithms for proprietary GIS software that facilitated automated processes of transportation network analysis to solve these types of problems often referred to as GIS-T (Miller, 1999; Dunn & Newton, 1992). Today, algorithms for transportation network analysis are based on route optimization and are part of every day geospatial applications from delivery services, to general wayfinding, to emergency services and evacuation.

In the case that data is available, work in spatial transportation network analysis has included the study of the evolution of transportation networks over space and time. For example, the airline (worldwide and domestic) network structure varies over space and time due to geographic, economic, political, and historical factors and thus is constantly changing as links between airports appear, disappear, expand, or contract (Zhang et al., 2010; Rocha, 2009; Jia et al., 2014). Similar studies on the evolution of road and street networks have been published, however, this evolution can be very hard to trace because parts of their geometry become erased, making them difficult to trace and quantify (Gundmundsson & Mahajeri, 2013; Masucci et al., 2013).

2.4.2. Spatial Social Networks

Social networks represent the relationships or interactions between individuals where a link might signify a friendship, colleague, romantic, familial relationship or physical contact. The spatial context of social networks is not typically addressed in traditional social network research, however, as popularity of location-based services such as Facebook, Instagram, and Twitter increase and people are willing to share their geographic position with their friends and consequently their service providers, data

about the geographic position of users in combination with their social connections is now available (Scellato et al., 2011). In a spatial social network, individuals in the social network are embedded in geographic space. Typically, spatial social networks are undirected, since most relationships are reciprocated, and may model either a single or set of relationship types.

The use of graph theory to mathematically characterize spatial social networks is well documented and extensive (Yook et al., 2002; Scellato et al. 2011; Lambiotte et al., 2008, Backstrom et al., 2010, Cairncross, 2001, Alizadeh et al., 2017). As demonstrated in the previous section, the spatial structure of transportation networks is largely influenced by the cost of linking nodes that are distant from one another, often generating regional clusters and minimizing long-distance links. This finding applies to spatial social networks as well. Specifically, as the geographic distance between two people increases, the relationship requires more energy to maintain. As a result, most individuals connect instead with their spatial neighbours whereby 40% of social links are between people within 100 km (Scellato et al., 2011; Yook et al., 2002). Since only a few connections are maintained over long distances and the majority of connections are maintained over short distances, the frequency distribution of distances between all nodes in a social network produces a power law (Lambiotte et al., 2008). In fact, Backstrom et al. (2010) identified a power law that was so significant that they developed an algorithm capable of inferring the location of a user based on the location of their friends. It is anticipated that geographic space as a constraint has lessoned and will continue to do so as new technology and affordable long-distance travel reduces the cost of long-distance connections (Cairncross, 2001). In contrast to transportation networks, the number of connections that an individual may have is not constrained by physical space.

Traditional studies in social networks relied on aggregated statistical datasets in one or more static points in time and selective small-scale surveys (Hawelka et al., 2014). Today, GIScientists have the potential to harness an unprecedented access to records of user location and their interactions over time. Vehicles are equipped with sensors, people carry GPS enabled cellphones, and users are willing to report their location through social media, producing spatial social network data at a finer spatial and temporal resolutions at a large volume. Understanding and identifying key properties of real spatial social networks can be further used in the development of geospatial models

that apply this knowledge to better understand several complex spatial systems that operate as a function of underlying spatial social networks including trade (Bush, 2004; Doreian and Conti, 2012), crime (Liu, 2008; Schaefer, 2012), friendship (Mcdonald-Wallis et al. 2011), and epidemiology (Bian & Liebner, 2007; Kuperman & Abramson, 2001; Moore & Newman, 2000; Pastor-Satorras & Vespignani, 2002).

2.4.3. Spatial Ecological Networks

Ecological systems have long been analyzed based on their topological network structure where ecological entities such as individuals, species, or communities are represented as aspatial nodes and a variety of interactions between ecological entities are represented using links. Graph theory provides a mathematical tool to represent and analyze sets of ecological interactions that form food networks (Cohen 1978; Hall and Raffaelli 1993), host-parasitoid networks (Muller et al. 1999; Morris et al. 2004) and mutualistic networks (Jordano 1987; Stang et al. 2006) to name a few. Geographic space plays a large role in ecological interactions, and as a result, the inclusion of geographic space into these representations can be informative. In these representations, nodes are spatially embedded, and link together based on a variety of spatial interactions such as dispersal, proximity, or similarity between landscape features or species (Dale & Fortin, 2010). Campbell et al. (2007) argues that the spatial structure of ecological networks regulates and modifies ecological processes. Based on this idea, spatial graphs in ecology emerged in application to the movement of species between habitat patches, species evolution, the study of landscape genetics, and spatial epidemiology.

Landscape connectivity graphs represent landscape features using nodes embedded in geographic space and the functional relationship between nodes such as dispersal or movement using links (Urban & Kiett, 2001). Typically, landscape features are connected based on the maximum dispersal distance of a species of interest. The abstraction of the landscape into a network structure can help to inform the spatial dynamics of several ecological species as the “flow” between nodes and has been developed for several different species (Fortuna et al., 2006; O’Brien et al., 2006; Urban & Kiett, 2007; Campbell et al., 2007; Fortuna et al., 2009; Zetterberg et al., 2010; Pereira et al., 2011). Landscape connectivity graphs have been modified to supplement link properties such as Euclidean distance between two habitat patch nodes with cost

values, determining the paths that offer the least resistance to movement (Chardon et al., 2003; O'Brien et al., 2006, Scotti et al., 2007, Spear et al., 2010). Spatial ecological graphs in application to landscape genetics and spatial epidemiology also modify the link representation itself, where in the case of landscape genetics a link represents positive spatio-autocorrelation of landscape features that are genetically similar (Dale & Fortuna, 2010) or in the case of spatial ecological epidemiology, a link represents the spread of disease (Brooks et al., 2008, Fortuna et al., 2009).

Traditional graph theory measures such as betweenness centrality g, average clustering coefficient <C>, modularity G, and nestedness are of interest in spatial ecological networks, however, several graph theory measures were developed specifically for this area of ecological applications. Pascual-Hortal and Saura (2006) proposed a large set of measures including the integral index of connectivity, least cost path, route path diameter, dispersal likelihood, route redundancy, route vulnerability, and area connected. Other measures of note include expected cluster size (O'Brien et al., 2006), the *F metric* (Ferrari et al., 2007), and the clumpiness coefficient (Estrada and Boden, 2008). In addition to spatial network characterisation, scenario testing can be performed in order to better understand the resilience of the network in the case of node removal, useful as a tool for decision making in environmental management and conservation. For example, Fortuna et al. (2006) uses drought scenarios to predict the resilience of an amphibian population based on landscape connectivity. The representation and analysis of spatial ecological networks using graph theory is increasingly facilitated by the availability of geospatial data and geographic information systems (GIS) with built in tools for this type of analysis.

2.4.4. Geophysical Networks

Geophysical systems or Earth systems can be represented using spatial networks where nodes might represent the location of landforms, soil types, geologic formations, sources or sinks, storms, earthquakes, or fires and links might represent transport pathways, mass and energy exchange, feedback relationships, spatial adjacency, or temporal sequences (Phillips et al. 2015). Graph theory measures are often applied to characterize spatial and topological properties of the network structures and to understand the flow (Gascuel-Odoux et al., 2011; Santiago et al., 2008, Zaliapin

et al., 2010), complexity (Philips, 2013), or synchronization (Tsonis & Swanson, 2012) properties as a function of these structures.

2.5. Linking Structure and Dynamics of Complex Spatial Systems using Networks

Traditional spatial network studies in application to transportation, social, ecological, and geophysical systems focus on *describing* the static spatial network structure of these systems that shape related spatial dynamics and processes. Networks have therefore been integrated as inputs into complex process-based geospatial modelling approaches like geographic automata systems (GAS) (Torrens & Benenson, 2005) including cellular automata (CA) and agent-based modelling (ABM). GAS focus on representing local dynamics and processes of such systems over geographic space and time from which the system as a whole emerges and thus the inclusion of the network structures that shape these processes is valuable.

Human mobility networks are often represented by transportation networks such as street networks or airline networks (Yang et al., 2011), but can also be represented as the direct movement of people between landscape features or buildings (Dibble and Feldman, 2004). The inclusion of mobility networks into GAS can be useful for better representing dynamics of emergency evacuation, the spread of disease, flow of information, and traffic and congestion (Torrens, 2004; Dibble and Feldman, 2004; Balcan et al. 2009). Brachman & Dragicevic (2014) develop a Network Science Emergency Evacuation Model (NetSEEM), a geospatial application that facilitates the representation of the physical, biological, and social complexities involved in mobility during evacuation. In another example, Balcan et al. (2009) tests how variations in the structure of mobility networks including commuting networks and long-range airline traffic shape the spatio-temporal dynamics of global epidemics. Frias-Martinez et al. (2011) integrate human mobility networks into an ABM to simulate the spread of H1N1 in Mexico.

The representation and analysis of social networks and the processes occurring on or within them also lend naturally to both network and GAS modelling approaches. Burger et al. (2017) acknowledges that the observed characteristics of social networks used to inform and influence human interactions are typically absent from the ABMs that

seek to model these types of systems, prompting the integration of the two approaches for the study of social networks. Pires & Crooks (2017) integrate social networks and agent-based modelling to model the flow of information through a population to inform rioting behavior. In a similar vein, human contact networks, a network of individuals that come into contact in the same location and at the same time is the underlying structure for which processes and dynamics of disease operate. Perez & Dragicevic (2009) develop a spatially explicit model that simulates the transmission of infectious disease over a transportation network and with human contacts.

2.6. Conclusion and Future Directions

Spatial network science research leverages tools for representation and analysis of complex spatial systems from the fields of geography, GISc, and network science. These studies tend to focus on characterising static spatial network structures at one or several points in time to better understand processes and spatial dynamics of information, people, ecological species, and other objects. However, the relationship between spatial structure and spatial dynamics is non-linear where over time, the spatial dynamics and processes of these systems influence the system spatial structure. In all of the aforementioned approaches that integrate network theory and GAS, network structure is introduced to improve the representation of network dynamics and processes, but the subsequent evolution of the network structure as a function of network dynamics and processes has not been explored. Simulating the complete exchange between network dynamics and an evolving network structure can be particularly valuable to better understand and forecast real-world complex systems. This exposes the gap that calls for a new modelling framework for studying and testing the relationship between the evolution of spatial structures as a function of spatial dynamics and processes and vice versa that may be facilitated by further exploring the integration of network and GAS modelling approaches.

Increased data availability and computational efficiency may also aid in this to help understand the evolution of transportation networks, mobility networks, social networks, ecological networks, and geophysical networks across a variety of scales through the use of complex systems modelling approaches that have been postulated to operate on different spatial tessellations including networks (Torrens & Benenson, 2005). The implementation of complex systems modelling approaches and network

theory for the representation and analysis of complex spatial systems to explore and better understand the link between spatial structure and spatial processes and dynamics is not yet fully explored. These data-driven modelling approaches show even more potential in the era of big data. However, even with the promise of increased data availability, larger computer capacities, and improved knowledge about the structure and behavior of networks, developed models still face challenges with respect to all three, limiting the potential for model development and furthermore for model validation via agreement with real data so that they are suitable for use in decision-making processes. Harnessing the power of big data for both model development and validation is slowed by challenges presented by the volume (size), variety (various types of geospatial data), variability (inconsistency between data provided by different users), and veracity (attempting to combine data of different quality) and pose problems in storing, managing, processing, visualizing, and verifying the data (Laney et al., 2011). Furthermore, as data availability increases, there are a number of challenges associated with ethics, privacy, personal security, ownership, and frameworks for determining the quality of data that will need to be addressed in order for spatial network-based modelling to make use of this data.

In addition to technical challenges, there is a demand for new methods for both network integrated model development and validation as well as methods for network analysis of model outputs. Spatially explicit complex systems modelling have always faced unique challenges in model validation. This challenge is only further complicated in terms of validating the network structures and dynamics represented by these models. Furthermore, there is a demand for new network measures that are specifically geared for spatio-temporal analysis. Barthelemy (2018) has many times shared his concern that despite the large number of studies on planar networks, there is still a lack of metrics that facilitate the characterization of the spatio-temporal structure of networks.

In summary, there remain many rich avenues to be explored with respect to the integration of GISc and network science for representing and understanding of geospatial, dynamic, and complex systems. The application of graph theory provides value for future studies that apply spatial network analysis to urban, ecological, social, and geophysical systems to better understand the complex tightly coupled relationship between spatial structure and space-time dynamics.

2.7. References

Alizadeh, M., Cioffi-Revilla, C., & Crooks, A. (2017). Generating and analyzing spatial social networks. *Computational and Mathematical Organization Theory,* 23(3), 362-390.

Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*(21), 11149-11152.

Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th International Conference on World Wide Web*. Raleigh, NC: ACM.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences,* 106(51), 21484-21489.

Barabási, A. L. (2016). *Network Science*. Glasgow, UK: Cambridge University Press.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. S*cience*, *286*(5439), 509-512.

Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3-4), 590-614.

Barrat, A., Barthélemy, M., & Vespignani, A. (2005). The effects of spatial constraints on the evolution of weighted complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 05, P05003.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. In *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747-3752.

Barreira-González, P., Gómez-Delgado, M., & Aguilera-Benavente, F. (2015). From raster to vector cellular automata models: A new approach to simulate urban growth with the help of graph theory. *Computers, Environment and Urban Systems*, *54*, 119-131.

Barthélemy, M. (2011). Spatial networks. *Physics Reports,* 499(1), 1-101.

Barthélemy, M. (2018). *Morphogenesis of Spatial Networks.* Cham, Switzerland: Springer International Publishing.

Bassett, D. S., & Bullmore, E. D. (2006). Small-world brain networks*. The Neuroscientist, 12(6), 512-523.*

Batty, M. (2005). Network geography: Relations, interactions, scaling and spatial processes in GIS. *Re-presenting GIS*, 149-170.

Bian, L. (2004). A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B: Planning and Design.* 31, 381-395.

Bian, L., & Liebner, D. (2007). A network model for dispersion of communicable diseases. *Transactions in GIS*, *11*(2), 155-173.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4-5), 175–308.

Buhl, J., Gautrais, J., Reeves, N., Solé, R. V., Valverde, S., Kuntz, P., & Theraulaz, G. (2006). Topological patterns in street networks of self-organized urban settlements. *The European Physical Journal B-Condensed Matter and Complex Systems*, *49*(4), 513-522.

Burger, A., Oz, T., Crooks, A.T. and Kennedy, W.G. (2017), Generation of Realistic Mega-City Populations and Social Networks for Agent-Based Modeling. *The Computational Social Science Society of Americas Conference*, Santa Fe, NM: ACM.

Bush, S. R. (2004). Scales and sales: changing social and spatial fish trading networks in the Siiphandone fishery, Lao PDR. *Singapore Journal of Tropical Geography*, *25*(1), 32-50.

Cairncross, F. (2001). *The death of distance: How the communications revolution is changing our lives*. Brighton, MA: Harvard Business Press.

Campbell Grant, E. H., Lowe, W. H., and Fagan, W. F. (2007). Living in the branches: Population dynamics and ecological processes in dendritic networks. *Ecology Letters,* 10(2), 165–175.

Cardillo, A., Scellato, S., Latora, V., & Porta, S. (2009). Structural properties of planar graphs of urban street patterns. *Physical Review E*, *73*(6), 066107.

Chardon, J. P., Adriaensen, F., & Matthysen, E. (2003). Incorporating landscape elements into a connectivity measure: a case study for the Speckled wood butterfly (Pararge aegeria L.). *Landscape Ecology*, *18*(6), 561-573.

Cohen, J. E. (1978). *Food Webs and Niche Space.* New Jersey, NJ: Princeton University Press.

Cova, T. J., & Johnson, J. P. (2003). A network flow model for lane-based evacuation routing. *Transportation Research Part A – Policy and Practice*, 37(7), 579–604.

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E*, *73*(3), 036125.

da Rocha, L. E. (2009). Structural evolution of the Brazilian airport network. *Journal of Statistical Mechanics: Theory and Experiment*, *2009*(04), P04020.

Dale, M. R. T., & Fortin, M. J. (2010). From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 21-38.

Dibble, C., & Feldman, P. G. (2004). The GeoGraph 3D computational laboratory: Network and terrain landscapes for RePast. *Journal of Artificial Societies and Social Simulation*, *7*(1).

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269-271.

Doreian, P., & Conti, N. (2012). Social context, spatial structure and social network structure. *Social Networks*, *34*(1), 32-46.

Ducruet, C., & Beauguitte, L. (2014). Spatial science and network science: Review and outcomes of a complex relationship. *Networks and Spatial Economics*, 14(3-4), 297-316.

Ducruet, C., & Notteboom, T. (2012). The worldwide maritime network of container shipping: spatial structure and regional dynamics. *Global Networks*, *12*(3), 395-423.

Dunn, C. E., & Newton, D. (1992). Optimal routes in GIS and emergency planning applications. *Area*, 259-267.

Erdos, P., & Renyi, A. (1959). On Random Graphs I. *Selected Papers of Alfred Renyi*, *2*(1), 308–315.

Erdos, P., & Renyi, A. (1960). On the Evolution of Random Graphs. *Bulletin of the International Statistical Institute*, *34*(4), 343–347.

Estrada, E., & Bodin, Ö. (2008). Using network centrality measures to manage landscape connectivity. *Ecological Applications*, *18*(7), 1810-1825.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis (The solution to a problem relating to the geometry position). *Commentarii Academie Scientiarum Imperialis Petropolitanae*, 8, 128–140.

Fell, D. A., & Wagner, A. (2000). The small world of metabolism. *Nature Biotechnology*, *18*(11), 1121-1122.

Ferrari, J. R., Lookingbill, T. R., & Neel, M. C. (2007). Two measures of landscape-graph connectivity: assessment across gradients in area and configuration. *Landscape Ecology*, *22*(9), 1315-1323.

Ford, L. R., & Fulkerson, D. R. (1961). An out-of-kilter method for minimal cost flow problems. *Journal of the Society for the Industrial Application of Mathematics*, 9(1), 19–27.

Fortuna, M. A., and Bascompte, J. (2007). The network approach in ecology. In F. Valladares, A. Camacho, A. Elosegi, C. Gracia, M. Estrada, J. C. Senar, and J. M. Gili (Eds.), *Unity in diversity: Ecological reflections as a tribute to Margalef* (pp. 371–392). Bilbao: Fundación BBVA.

Fortuna, M. A., Gomez-Rodriguez, C., and Bascompte, J. (2006). Spatial network structure and amphibian persistence in stochastic environments. *Proceedings of the Royal Society B: Biological Sciences*, 273(1592), 1429–1434.

Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. (2011). An agent-based model of epidemic spread using human mobility and social network information. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*. Boston, MA: IEEE.

Garrison, W. L. (1960). Connectivity of the interstate highway system. *Papers in Regional Science*, *6*(1), 121-137.

Gascuel-Odoux, C., Aurousseau, P., Doray, T., Squividant, H., Macary, F., Uny, D., & Grimaldi, C. (2011). Incorporating landscape features to obtain an object-oriented landscape drainage network representing the connectivity of surface flow pathways over rural catchments. *Hydrological Processes*, *25*(23), 3625-3636.

González, P. B., Gómez-Delgado, M., & Benavente, F. A. (2015). Vector-based cellular automata: exploring new methods of urban growth simulation with cadastral parcels and graph theory. In *Proceedings of CUPUM 2015*. Cambridge, MA.

Granovetter, M. (1973). Strength of Weak Ties. *American Journal of Sociology,* 78(6), 1360–1380.

Gudmundsson, A., & Mohajeri, N. (2013). Entropy and order in urban street networks. *Scientific Reports*, *3*, 3324.

Guimera, R., & Amaral, L. A. N. (2004). Modeling the world-wide airport network. *The European Physical Journal B-Condensed Matter and Complex Systems*, *38*(2), 381-385.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, *102*(22), 7794-7799.

Hall, S. J., and Raffaelli, D. G. (1993). Food webs: Theory and reality. *Advances in Ecological Research,* 24, 187–239.

Han, C., & Liu, L. (2009). Topological vulnerability of subway networks in China. In 2009 International Conference on Management and Service Science. Wuhan, China: IEEE.

Han, D. D., Qian, J. H., & Liu, J. G. (2009). Network topology and correlation features affiliated with European airline companies. *Physica A: Statistical Mechanics and its Applications*, 388(1), 71-81.

Hillier, F. S., & Lieberman, G. J. (1990). *Introduction to Operations Research*. New York, NY: McGraw-Hill Inc.

Holland, J. H. (1996). *Hidden Order: How Adaptation Builds Complexity.* Cambridge, MA: Perseus Books.

Jia, T., & Jiang, B. (2012). Building and analyzing the US airport network based on en-route location information. *Physica A: Statistical Mechanics and its Applications*, *391*(15), 4031-4042.

Jia, T., Qin, K., & Shan, J. (2014). An exploratory analysis on the evolution of the US airport network. *Physica A: Statistical Mechanics and its Applications*, *413*, 266-279.

Jiang, B. (2007). A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, *384*(2), 647-655.

Jiang, B. (2009). Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science*, *23*(8), 1033-1048.

Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *The American Naturalist,* 129(5), 657–677.

Kaluza, P., Kölzsch, A., Gastner, M. T., & Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, *7*(48), 1093-1103.

Kansky, K. (1969) *Structure of transportation networks: relationships between network geometry and regional characteristics*. Chicago, IL: University of Chicago Press.

Kuperman, M., & Abramson, G. (2001). Small world effect in an epidemiological model. *Physical Review Letters*, *86*(13), 2909.

Kurant, M., & Thiran, P. (2006). Extraction and analysis of traffic and topologies of transportation networks. *Physical Review E*, *74*(3), 036114.

Lambiotte, R., Blondel, V. D., De Kerchove, C., Huens, E., Prieur, C., Smoreda, Z., & Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, *387*(21), 5317-5325.

Lämmer, S., Gehlsen, B., & Helbing, D. (2006). Scaling laws in the spatial structure of urban road networks. *Physica A: Statistical Mechanics and its Applications*, *363*(1), 89-95.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*, 70.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, *87*(19), 198701.

Latora, V., & Marchiori, M. (2002). Is the Boston subway a small-world network? *Physica A: Statistical Mechanics and its Applications*, *314*(1), 109-113.

Latora, V., & Marchiori, M. (2003). Economic small-world behavior in weighted networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, *32*(2), 249-263.

Latora, V., & Marchiori, M. (2007). A measure of centrality based on network efficiency. *New Journal of Physics*, *9*(6), 188.

Levinson, D. (2007). Density and dispersion: the co-development of land use and rail in London. *Journal of Economic Geography*, *8*(1), 55-77.

Lewis, T. G. (2011). *Network Science: Theory and Applications*. Hoboken, NJ: John Wiley & Sons.

Li, W., & Cai, X. (2004). Statistical analysis of airport network of China. *Physical Review E*, *69*(4), 046106.

Liu, H. K., Zhang, X. L., & Zhou, T. (2010). Structure and external factors of Chinese city airline network. *Physics Procedia*, *3*(5), 1781-1789.

Liu, L. (2008). *Artificial crime analysis systems: using computer simulations and geographic information systems: using computer simulations and geographic information systems*. Hershey, NY: IGI Global.

Macdonald-Wallis, K., Jago, R., Page, A. S., Brockman, R., & Thompson, J. L. (2011). School-based friendship networks and children's physical activity: A spatial analytical approach. *Social science & medicine*, *73*(1), 6-12.

Masucci, A. P., Smith, D., Crooks, A., & Batty, M. (2009). Random planar graphs and the London street network. *The European Physical Journal B*, *71*(2), 259-271.

Masucci, A. P., Stanilov, K., & Batty, M. (2013). Limited urban growth: London's street network dynamics since the 18th century. *PLoS One*, *8*(8), e69469.

Milgram, S. (1967). The Small World Problem. *Psychology Today*, *32*(4), 425–443.

Miller, H. J. (1999). Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis,* 31(4), 373-399.

Minor, E. S., & Urban, D. L. (2007). Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecological Applications*, *17*(6), 1771-1782.

Moore, C., & Newman, M. E. (2000). Epidemics and percolation in small-world networks. *Physical Review E*, *61*(5), 5678.

Morris, R. J., Lewis, O. T., and Godfray, H. C. J. (2004). Experimental evidence for apparent competition in a tropical forest food web. *Nature,* 428(6980), 310–313.

Muller, C. B., Adriaanse, I. C. T., Belshaw, R., and Godfray, H. C. J. (1999). The structure of an aphid–parasitoid community. *British Ecological Society,* 68(1), 346–370.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review,* 45(2), 167-256.

Newman, M. E. (2005). Power Laws: Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323-351.

O'Brien, D., Manseau, M., Fall, A., & Fortin, M. J. (2006). Testing the importance of spatial configuration of winter habitat for woodland caribou: an application of graph theory. *Biological Conservation*, *130*(1), 70-83.

O'Sullivan, D. (2002). Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems*, *4*(3), 251-274.

Pascual-Hortal, L., & Saura, S. (2006). Comparison and development of new graph-based landscape connectivity indices: towards the priorization of habitat patches and corridors for conservation. *Landscape Ecology*, *21*(7), 959-967.

Pastor-Satorras, R., & Vespignani, A. (2002). Immunization of complex networks. *Physical Review E*, *65*(3), 036104.

Pereira, M., Segurado, P., & Neves, N. (2011). Using spatial network structure in landscape management and planning: a case study with pond turtles. *Landscape and Urban Planning*, *100*(1-2), 67-76.

Perez, L., & Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, *8*(1), 50.

Phillips, J. D. (2013). Sources of spatial complexity in two coastal plain soil landscapes. *Catena*, *111*, 98-103.

Phillips, J. D., Schwanghart, W., & Heckmann, T. (2015). Graph theory in the geosciences. *Earth-Science Reviews*, *143*, 147-160.

Pires, B., & Crooks, A. T. (2017). Modeling the emergence of riots: A geosimulation approach. *Computers, Environment and Urban Systems*, *61*, 66-80.

Rahmandad, H., & Sterman, J. (2008). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science,* 54(5), 998-1014.

Santiago, A., Cárdenas, J. P., Losada, J. C., Benito, R. M., Tarquis, A. M., & Borondo, F. (2008). Multiscaling of porous soils as heterogeneous complex networks. *Nonlinear Processes in Geophysics*, *15*(6), 893-902.

Scellato, S., Noulas, A., Lambiotte, R., & Mascolo, C. (2011). Socio-spatial properties of online location-based social networks. *ICWSM*, *11*, 329-336.

Schaefer, D. R. (2012). Youth co-offending networks: An investigation of social and spatial effects. *Social Networks*, *34*(1), 141-149.

Scotti, M., Podani, J., & Jordán, F. (2007). Weighting, scale dependence and indirect effects in ecological networks: a comparative study. *Ecological Complexity*, *4*(3), 148-159.

Sole, R., & Valverde, S. (2004). *Information Theory of Complex Networks: On Evolution and Architectural Constraints*. Heidelberg, Berlin: Springer.

Spear, S. F., Balkenhol, N., FORTIN, M. J., McRae, B. H., & Scribner, K. I. M. (2010). Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, *19*(17), 3576-3591.

Stang, M., Klinkhamer, P. G. L., and Meijden, E. Van Der. (2006). Size constraints and flower abundance determine the number of interactions in a plant–flower visitor web. Oikos, 112(1), 111–121.

Strano, E., Viana, M., da Fontoura Costa, L., Cardillo, A., Porta, S., & Latora, V. (2013). Urban street networks, a comparative analysis of ten European cities. *Environment and Planning B: Planning and Design*, *40*(6), 1071-1086.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography,* 46, 234-40.

Topping, C. J., Høye, T. T., & Olesen, C. R. (2010). Opening the black box—Development, testing and documentation of a mechanistically rich agent-based model. *Ecological Modelling*, *221*(2), 245-255.

Torrens, P. M. (2004). Geosimulation, automata, and traffic modeling. *Handbooks in Transport*, *5*, 549-565.

Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science*, *19*(4), 385-412.

Tsonis, A. A., & Swanson, K. L. (2012). Review article" On the origins of decadal climate variability: a network perspective". *Nonlinear Processes in Geophysics*, *19*(5), 559-568.

Urban, D., and Keitt, T. (2001). Landscape connectivity: a graph-theoretic perspective. *Ecology,* 82(5), 1205–1218.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440.

Williams, R. J., Berlow, E. L., Dunne, J. A., Barabási, A. L., & Martinez, N. D. (2002). Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences, 99(20), 12913-12916.*

Wright, D. (2010). GISc, in Warf, B., Jankowski, P., Solomon, B.D., & Welford, M. *Encyclopedia of Geography,* Thousand Oaks, CA: SAGE Publications.

Xu, X., Hu, J., Liu, F., & Liu, L. (2007). Scaling and correlations in three bus-transport networks of China. *Physica A: Statistical Mechanics and its Applications*, *374*(1), 441-448.

Yang, Y., Roux, A. V. D., Auchincloss, A. H., Rodriguez, D. A., & Brown, D. G. (2011). A spatial agent-based model for the simulation of adults' daily walking within a city. *American Journal of Preventive Medicine*, *40*(3), 353-361.

Yook, S. H., Jeong, H., & Barabási, A. L. (2002). Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences*, *99*(21), 13382-13386.

Yule, G. U. (1925). Mathematical theory of evolution based on the conclusion of Dr. J.C. Willis*. Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character, 213(1), 21–87.*

Zaliapin, I., Foufoula-Georgiou, E., & Ghil, M. (2010). Transport on river networks: A dynamic tree approach. *Journal of Geophysical Research: Earth Surface*, *115*(F2).

Zetterberg, A., Mörtberg, U. M., & Balfors, B. (2010). Making graph theory operational for landscape ecological assessments, planning, and design. *Landscape and Urban Planning*, *95*(4), 181-191.

Zhang, J., Cao, X. B., Du, W. B., & Cai, K. Q. (2010). Evolution of Chinese airport network. *Physica A: Statistical Mechanics and its Applications*, *389*(18), 3922-3931.

Chapter 3.

Representing Spatial Systems as Evolving Networks: A Geographic Network Automata Model [1]

3.1. Abstract

Almost all real-world spatial systems can be conceptualized from the bottom-up as spatial networks where nodes representing system components are embedded in geographic space and links represent some form of relationship or interaction. Typically, spatial network analysis focuses on generating descriptive network measures obtained from static network datasets, however, more recent interest lies in the representation and analysis of evolving spatial networks that can facilitate the examination of the close coupling between spatial network structure and spatial network dynamics. Therefore, the objective of this study is to propose and implement a novel modelling framework, Geographic Network Automata (GNA), for representing and analysing complex spatial systems as evolving networks. The GNA framework is implemented and tested on a spatial network adaptation of Conway's Game of Life model and explores the structure and behavior of evolving spatial networks using three scenarios. The simulated evolving spatial network structures are quantified using graph theory measures. Results indicate that graph theory measures are dependent on spatial network size as the network grows or shrinks. The presented GNA modelling framework is both general and flexible, useful for modelling a variety of real geospatial phenomena and characterizing and exploring network structure, dynamics, and evolution of real spatial systems. The GNA modelling framework is situated within the larger framework of geographic automata systems alongside cellular automata and agent-based modelling.

3.2. Introduction

As geospatial data becomes increasingly available, networks are used as a powerful conceptual framework to represent and analyze a wide array of complex spatial

[1] A version of this chapter is submitted for publication: Anderson, T. & Dragicevic, S. (Submitted). Representing spatial systems as evolving networks: A Geographic Network Automata model. Journal of Geographical Systems.

systems from social, urban, and ecological (Scellato et al., 2011; Zhong et al., 2014; Fortuna et al., 2006). The conceptualization of complex spatial systems as networks begins from the bottom-up, where system components are represented as georeferenced nodes and interactions between components are represented as links. Sets of local interactions form the global network structure, for which the system as a whole is represented. The representation of complex spatial systems as networks offers a well-developed toolkit for analysis. Specifically, graph theory can be applied to describe the spatial structures of real phenomena and explore the tightly coupled relationship between spatial structure and spatial dynamics.

Network representations of spatial dynamics can be distinguished between dynamics *on* a network or the dynamics *of* networks (Gross and Sayama, 2009). In the former, the flow of information or materials dynamically propagate through a set of spatially arranged nodes and links. For example, in ecology, the spatio-temporal dynamics of species dispersal is highly dependent on the spatial structure of habitat features across the landscape (Fortuna et al., 2006). In an epidemiological context, the spatio-temporal dynamics of disease spread is predicted using the spatial structure of human contact networks (Bansal et al., 2010). The underlying network structure on which network dynamics propagate can either be static as in the ecological example, or dynamic as in the case of a human contact network which changes over geographic space and time. Dynamics *of* networks may operate independently of an underlying network structure. Specifically, as a function of the tight coupling and feedback between network structure and network dynamics, the structure of the network itself dynamically changes over time. For example, street network structure influences vehicle traffic, but in turn spatio-temporal patterns of traffic may damage streets, forcing their closure, or require the construction of new streets to reduce congestion.

Changes in networks over geographic space and time as function of dynamics *on* a network or the dynamics *of* networks is referred to as network evolution. In this process, nodes and links are added, removed, rewired, or undergo changes with respect to their properties over time (Smith et al., 2011). Network evolution is not well understood since detailed spatio-temporal datasets representing real phenomena that can be conceptualized as networks have been historically limited. Thus conventional spatial network analysis tends to focus on describing static spatial network structures or exploring the effect of static spatial network structures on spatial dynamics including for

example the process of gentrification on a static network of properties (O'Sullivan, 2002), ecological dispersal dynamics on landscape connectivity networks (Fortuna et al., 2006; Urban and Kiett, 2001), transportation dynamics on road networks (Decruet and Notteboom, 2012; Jiang, 2009), and epidemics on contact networks (Bansal et al., 2010; Bian and Liebner, 2007) and on airline networks (Colizza et al., 2006).

Models representing phenomena such as predator prey dynamics (Boccara et al., 1993) and human epidemics (Boccara, 1994) *as* evolving networks were developed by applying sub-rules representing network dynamics to network structures that alter the network structure itself over time. This methodology was later formalized as network automata (Sayama and Laramee, 2009; Smith et al., 2011), however these studies are not implemented on real world phenomena nor do they use geospatial data. Despite the demand for a shift from descriptive measures of spatial network structures to the study of evolving complex spatial networks that would facilitate the long-standing interest in the investigation into the link between spatial network structure and dynamics, network automata have not yet been explored in application to geospatial phenomena.

Therefore, the objectives of this study are to integrate concepts of GISc, complex systems, and network theory to 1) propose a theoretical framework of a novel modelling approach called Geographic Network Automata for the representation and analysis of complex spatial systems as evolving networks 2) demonstrate the GNA framework using a spatially explicit network version of Conway's Game of Life (Conway, 1970) and 3) develop several scenarios using the GNA framework that simulate different forms of evolving network behavior so that the structure and behavior of evolving networks can be analyzed. The Game of Life is selected as a case study to present the GNA framework because it is a well-known theoretical system that is inherently simple and operates in space and time. The original Game of Life is an automata developed by John Conway in 1970 that was designed to simulate dynamics of reproduction, death, and survival. Therefore, the Game of Life as a network representation permits the exploration of these dynamics resulting in network evolution including node addition, removal, rewiring, and changing of node properties, facilitating broader learning from its use. Furthermore, this case study facilitates a clear and simple explanation of the GNA framework that sets the stage for which the reader can envision the GNA framework in application to a real phenomenon.

Firstly, theoretical background for both the GNA modelling approach as well as the graph theory used for GNA analysis has been provided. Next, the GNA modelling framework is presented in application to the spatially explicit network version of the Game of Life using several scenarios that facilitate the exploration of changes in network structure and dynamics as they evolve over space and time. Lastly, the use of the GNA framework in application to a broad range of geospatial phenomena is discussed.

3.3. Geographic Network Automata (GNA)

This section first presents the general GNA modelling framework for network representation of real-world spatial phenomena and secondly introduces the theoretical background for how graph theory can be applied to analyze the GNA spatial network *SN* outputs.

3.3.1. GNA Modelling Framework

A geographic network automaton (GNA) is a mathematical representation of a complex system as an evolving *spatial network SN.* The major components of a GNA include the following: 1) a spatial network *SN* composed of a set of node automata *N* and a set of links *L* that represent a spatially-embedded complex system; 2) transition rules *R* that simulate system dynamics; 3) connection costs *C* that measure the resistance of the geographic space to network evolution; 4) time, where the topology of the spatial network *SN* at time *t+1* is a function of the applied transition rules *R* and connection costs *C* at time *t.*

In the conceptualization of a real-world system as an evolving spatial network *SN,* system dynamics can be represented as dynamics *on* a network or dynamics *of* a network. In either case, there exists a set of node automata with properties that are of most interest to the modeller. For example, in the network representation and analysis of spatio-temporal patterns of forest insect infestation, forest stands can be represented as nodes with one of two states, infested or not infested, and nodes with the infested state may of be more interest to the modeller. We refer to these node automata as *nodes of primary interest*. The nodes of primary interest form an evolving *spatial network of primary interest SN.*

In the case of representing dynamics *on* a network, a spatial network of primary interest *SN* is formed based on an underlying network *UN* (Figure 3.1a). Therefore, the GNA formulation includes the underlying network *UN* as an additional component and can therefore be formulated as:

$$GNA = [UN, SN, R, C, \Delta t] \qquad (1)$$

where *UN* is the underlying network, *SN* is the spatial network, *R* are the transition rules, *C* is the connection cost, and Δ*t* is the temporal resolution of the GNA.

In the case of representing dynamics *of* a network (Figure 3.1b), a spatial network of interest *SN* is formed independent of an underlying network. Therefore, the GNA can be formulated as:

$$GNA = [SN, R, C, \Delta t] \quad (2)$$

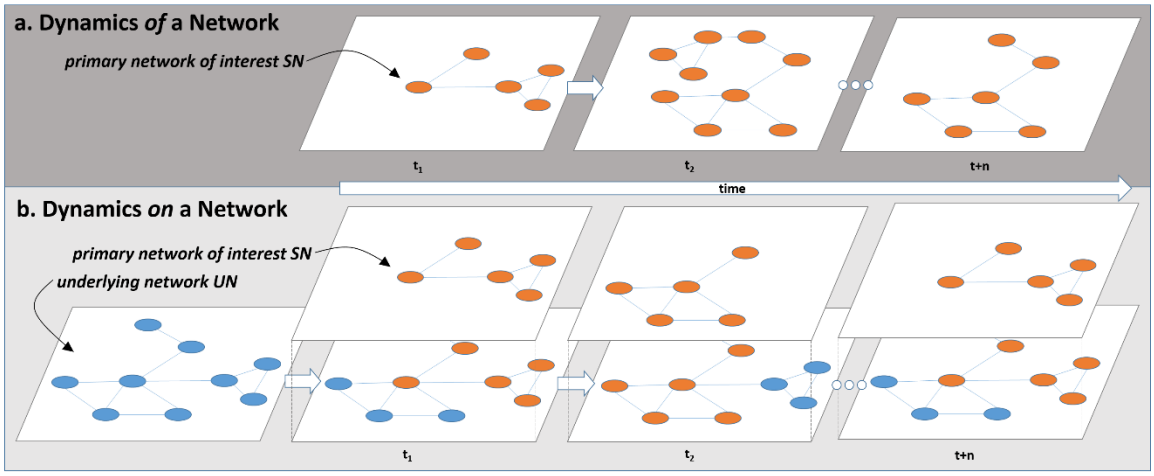

Figure 3.1. Different spatial network *SN* dynamics: (a) dynamics *of* a primary network of interest *SN* that evolves over time and (b) dynamics *on* a network *SN* where the network of interest evolves over time as a function of an underlying network *UN*.

Both the evolving spatial network *SN* generated by the GNA and the underlying network *UN* from which in some cases the spatial network *SN* might emerge is composed of a set of nodes *N* representing components of a system. Pairs of nodes are connected by links *L*, representing interactions or relationships between system components. The spatial network *SN* can be formulated as:

$$SN = [N, L] \quad (3)$$

In the case of representing dynamics *on* a network, the underlying network *UN* can also be formulated as:

$$UN = [N, L] \quad (4)$$

The dispersal of amphibians across the landscape from pond to pond is an example of dynamics *on* an underlying network *UN*. The species can only move from one pond node *i* to another pond node *j* if pond node *i* and *j* are within the maximum dispersal distance of the amphibian. The dispersal of the species is constrained by its dispersal processes which are constrained by distance and the spatial location of the ponds across the landscape. In this case, the network evolution of the spatial network *SN* is a function of the underlying network *UN*, the spatial network itself *SN*, and the matrix. The dispersal of a bark beetle across the landscape form from forest stand to forest stand is an example of a spatial network *SN* that operates independently of an underlying network *UN.* The beetle can only naturally move from forest stand *i* to forest stand *j* if it is within the maximum dispersal distance of the beetle, however, because of its ability to undergo long distance dispersal where the beetle is transported far distances through the interaction with humans, it is not limited by the spatial structure of the landscape. The beetle can therefore hop to forest stand nodes beyond its maximum dispersal distance, thus by-passing expected links that would be required to be traversed along the underlying landscape connectivity network. In this case, the spatial network evolution *SN* is a function of the dynamics of the network itself *SN* and the matrix.

It is important to note that the corresponding nodes and links in the *UN* and the *SN* are one and the same. For example, node v_i in the *UN* is exactly the same as the corresponding node v_i in the *SN*, it is only conceptualized as belonging to a different network based on what the modeller is interested in measuring. The sets of node automata *N* and links *L* in the spatial network *SN* or underlying network *UN* are further expressed as:

$$N = [v_1, v_2, \dots, v_n] \quad (5)$$

where v_i is a node in the set of nodes *N*; n is the number of nodes in *N*.

$$L = [e_1, e_2, \dots, e_m] \quad (6)$$

where e_j is a link in the set of links *L*; m is the number of links in *L*.

Each node v_i and link e_j in the set of nodes *N* and links *L* are defined by several spatial, non-spatial, and network properties (Table 3.1). Nodes *N* are defined by their spatial properties, most importantly geographic location *g* which in turn facilitates the measurement of geographic distance d_{ij} between any two nodes v_i and v_j. Depending on the type of phenomenon that the set of nodes represent, other geometric properties such as area and perimeter may be of interest. Nodes *N* are also defined by network properties such as the number and list of connections a node v_i has to other nodes in the network or node weight. Node weight *w* is a particularly useful network property that ranks the nodes importance, suitability, or preference within the set of nodes *N.*

Table 3.1. Examples of spatial, non-spatial, and network properties for both nodes *v* and links *e*

Property Type	Description	Examples of Node Properties	Examples of Link Properties
Spatial	Geometric properties pertaining to the node *v* or link *e*	location *(x, y* coordinates), area, distance from, perimeter	length, coordinates of end points, direction
Non-Spatial	Qualitative and quantitative non-spatial attributes pertaining to the node *v* or link *e*	name, ID, colour, value, type, state	name, ID, colour, value, type, state
Network	Measurements derived from network theory pertaining to the node *v* or link *e*	degree, betweenness, weight, clustering coefficient, list of neighbours	weight, list of end nodes

Links *L* are also defined by their spatial properties, which differ slightly from the spatial properties of nodes. Whereas nodes *N* are always embedded in geographic space, in most cases, links *L* are not. The exception is a planar network such as a road network, in which case the link length is of interest. Links also contain the important network property of link weight *w*, which can be used to quantify the magnitude of flow of individuals, materials, or information between nodes. Links are either unidirectional or bidirectional, meaning that flow occurs in one or both directions respectively. Both nodes *N* and links *L* have non-spatial properties which are qualitative or quantitative attributes used to describe the network node or link.

Transition rules *R* are designed to represent the real-world dynamics between system elements and determine the evolution of the spatial network *SN.* The

development of transition rules *R* is ultimately a function of whether spatial dynamics are represented *on* a network or *of* a network. In the case of representing dynamics *on* a network, transition rules *R* may be defined where the spatial network *SN* interacts with the underlying network *UN.* An example transition rule *R* demonstrating this interaction is as follows: at time step *t,* node v_i in the *SN* forms a link to node v_j in the *UN* if node v_j has some weight w_j. In the case of representing dynamics *of* a network, an example transition rule *R* demonstrating this type of evolving network is as follows: at time step *t,* if there are less than three nodes within the neighbourhood of node v_i, a new node v_j is spawned within this neighbourhood and a link forms between node v_i and node v_j. Connection cost *C* evaluates the resistance of the geographic space between nodes to both the formation of links between nodes or to the generation of new nodes. This space between nodes is referred to as the cost matrix. Resistance in this context may be a function of distance *d* between nodes or the low suitability of the matrix for connecting nodes or the spawning of new nodes.

The spatial and topological organization of the spatial network *SN* and in some cases the *UN*, specifically what nodes are connected to what nodes using which links, are recorded in an N x N adjacency table A_{SN} and A_{UN}, respectively. In these tables, the existence of a link between two nodes A_{ij} in a network is recorded using a value of 1, with the alternative recorded using a value of 0.

In the case of simulating dynamics *on* a network, applied transition rules *R* and connection costs *C* influence the underlying network *UN,* which in turn influences the spatial network *SN* and thus alter the information recorded within the adjacency matrices at each time step. Therefore, spatial network *SN* evolution with each time step is defined where the adjacency matrix at the subsequent time step *t + 1* is a function of the underlying network *UN,* the transition rules *R,* the connection cost *C,* and the adjacency matrix at the previous time *t.* In the case of simulating dynamics *of* a network, applied transition rules *R* and connection costs *C* alter the spatial network *SN* and thus alter the information recorded within the adjacency matrix at each time step. Therefore, the spatial network *SN* evolution is defined where the adjacency matrix at the subsequent time step *t + 1* is a function of the transition rules *R,* the connection cost *C,* and the adjacency matrix at the previous time *t.*

3.3.2. GNA Spatial Network Analysis using Graph Theory

The output of a GNA is a sequence of evolving spatial networks of primary interest *SN* representing a real-world phenomenon over space and time. This representation is useful because the structure of spatial networks can be described and compared and the link between spatial network structure and spatial dynamics can be explored using graph theory measures (Newman, 2003, Barthelemy, 2011, Lewis, 2011, Barabasi et al, 2016).

The observed structural and dynamical properties of real-world networks often exhibit some or many of the same properties of four well-defined theoretical graph types: regular, random, small-world, and scale-free. These types of graphs can be distinguished from one another using a few simple global graph theory measures that are able to characterize the structure of the graph as a whole including degree distribution *P(k),* average clustering coefficient <*C*>*,* and average path length <*l*> (Table 3.2).

Table 3.2. Some graph theory measures for network analysis.

Graph Theory Measure	Definition
Degree distribution *P(k)*	The number of connections a node has to other nodes in the network is a localized measure, specific to each node, and is referred to as node degree *k*. Therefore, the fraction of nodes in the network with degree k, calculated for the entire distribution of *k* is referred to as degree distribution *P(k).*
Average clustering coefficient <*C*>	Clustering coefficient *C* measures the likelihood of nodes that are connected to node i are also connected to each other. This is a localized measure specific to each node. Average clustering coefficient <C> measures the average *C* across all nodes in the network.
Average path length <*l*>	The average number of intermediate nodes and links in the shortest path between all pairs of nodes in the network is referred to as average path length <*l*>.

Networks that exhibit properties of regular graphs are composed of a set of nodes and links, where each node has the exact same number of links or degree *k*, and are often referred to as a lattice (Boccaletti et al., 2006). Alternatively, the underlying spatial structure of the original version of the Game of Life can be represented as a lattice, where all cells representing system components interact with the exact same number of their most adjacent neighbours. Since all nodes are tightly connected to their

nearest neighbours, networks with properties corresponding to regular graphs have a high clustering coefficient. The localized connections result in a long average path length between pairs of nodes in the network.

Networks that exhibit the properties of random graphs are composed of nodes that are connected to other nodes at random (Erdos and Renyi, 1959; 1960). Unlike a regular graph, nodes connect to both adjacent and distant nodes. Random graphs are defined by a degree distribution where all nodes have a similar degree. This well-defined average degree produces a *Poisson* degree distribution, characterized a peak with rapidly diminishing sides when graphed as a histogram. Since nodes are connected at random to both adjacent and distant nodes, the average clustering coefficient is very small and the average path length is very small.

Networks that exhibit properties of small-world graphs fall between regular graphs that have no randomness at all and random graphs that entirely random (Watts and Strogatz, 1998). Like a regular graph, the majority of nodes in small world graphs are connected to their nearest neighbours, however a few nodes are connected to distant nodes. This type of graph also produces a *Poisson* distribution when graphing the degree distribution as a histogram. However, small-world graphs are different than their regular or random counterparts because the few distant connections between nodes produces a high clustering coefficient <C>, but dramatically reduces the average path length <l>. Social networks typically exhibit properties of small-world graphs, where there exist only a few intermediate acquaintances between any two people in the world. Dynamics *on* a small world network, such as the spreading of information, is highly efficient.

Networks that exhibit properties of scale-free graphs are characterized by a degree distribution where a few nodes have a disproportionately large degree and the majority of nodes have a very small degree (Albert et al., 1999). This produces a *scale free* degree distribution with a low average clustering coefficient and a small average path length. Barabasi & Albert (1999) refer to networks with power law distributions as “scale free” networks because the same power law distribution remains across all scales in the network. This network structure is explained by growth and preferential attachment, meaning that as the network forms, the probability that a new link will be added to node v_i is proportional to the degree of that node and can result in the

formation of hubs with an anomalous number of links. These types of network are robust to random attack, however, the loss of a hub in a targeted attack would cause system failure (Cohen et al., 2000). Node degree is not the only factor contributing to preferential attachment (Yule, 1925) and several modifications have been proposed such as node age (Jeong et al., 2003; D'souza et al, 2007).

3.4. Game of Life GNA Model (GNA_{GOL})

In the following sections, the application of the proposed GNA framework to the spatially explicit network version of the Game of Life GNA_{GOL} is presented. The GNA_{GOL} is developed using the Java programming language in the Eclipse integrated development environment using the REcursive Porous Agent Simulation Toolkit (Repast) (Repast, 2016). Furthermore, based on the objective to explore the structure and behavior of evolving spatial networks, the GNA_{GOL} implements three scenarios to simulate typical spatial processes as represented by network fluctuation, growth, and shrinkage.

3.4.1. GNA_{GOL} Model Framework

The GNA_{GOL} model simulates system dynamics *on* an underlying random spatial network *UN*. Random spatial networks also known as random geometric graphs are created by randomly placing nodes in geographic space. Nodes are connected to other nodes if the distance between the two nodes falls within a selected distance threshold (Dall and Christensen, 2008). The distance threshold does not always need to be Euclidean as in some studies the street network distance has been used (Antonioni and Tomassini, 2012). Random geometric graphs differ from traditional random graphs because the distance threshold produces localized clustering between adjacent nodes and a lack of long-distance connections, characteristic of many spatially embedded graphs, which produces network structures with a much higher average clustering coefficient than traditional random graphs (Barthelemy, 2011). The choice to simulate dynamics *on* a network rather than *as* a network was based on the original Game of Life model that was composed of a lattice of cells that are either dead or alive, where the total number of system components do not change, but rather the state of the system

components changes. This facilitates the adaption of original Game of Life transition rules to the GNA_{GOL}.

Based on formulation (1), the GNA_{GOL} can therefore be formulated as:

$$GNA_{GOL} = [UN_{GOL}, SN_{GOL}, R, C, \Delta t] \quad (7)$$

where the GNA_{GOL} is a function of the underlying network UN_{GOL}, the spatial network of interest SN_{GOL}, the transition rules R, the connection cost C, and time Δt.

The underlying random spatial network UN_{GOL} is constructed using a set of 2000 randomly georeferenced nodes *N*. Node v_i and node v_j in the set of nodes *N* are connected by a link and thus interact if the distance d_{ij} is smaller than a given range, in this case d_{ij} *<= 1 km.* The given range, the 1km radius around node v_i is referred to as the *neighbourhood*. It can be said that the nodes v_j that are connected to node v_i as a function of this range are *neighbouring nodes* in the UN_{GOL}. Because the spatial distribution of all nodes is random rather than a regular tessellation, nodes in the UN_{GOL} do not have the same number of neighbouring nodes. This differs from the traditional formulism of the CA version of the Game of Life which essentially operates on a lattice or regular network where all cells have the exact same number of neighbours. Each node and link in the set of nodes *N* and links *L* can be defined by the properties in Table 3.3.

Table 3.3. The spatial, non-spatial, and network properties of the nodes and links in the GNA_{GOL}.

Property Type	Description	Examples of Node Properties	Examples of Link Properties
Spatial	Geometric properties pertaining to the node *v* or link *e*	location (x, y coordinates), distance *d*	coordinates of end points, direction
Non-Spatial	Qualitative and quantitative non-spatial attributes pertaining to the node *v* or link *e*	state (dead or alive)	
Network	Measurements derived from network theory pertaining to the node *v* or link *e*	degree, clustering coefficient	list of end nodes

The spatial network SN_{GOL} is composed of an evolving set of nodes that are of the state "alive". Node v_i and node v_j are connected by a link if node v_j also is of the

state “alive” and node v_j is within the neighbourhood of node v_i. In the case that a node has the state “dead”, the node is considered disconnected from the spatial network SN_{GOL}. The SN_{GOL} is the network of primary interest because it provides a working example of an evolving network that can be analyzed where nodes are added, removed, and rewired over time.

The spatial network SN_{GOL} emerges from dynamics that are implemented *on* the underlying network UN_{GOL}. In the underlying network UN_{GOL}, the state of nodes v_j in the neighbourhood of node v_i at time *t* influence the state of node v_i at time *t+1.* These local dynamics are implemented in the GNA_{GOL} using transition rules *R.* There are four transition rules *R*, which are applied to the UN_{GOL} at time *t* and determine the UN_{GOL} and the SN_{GOL} at time *t+1,* as follows:

***R1*.** To simulate the dynamics of under population, any live node v_i with *some number or fewer* of alive neighbors v_j dies and is removed from the spatial network SN_{GOL}

***R2*.** To simulate the dynamics of survival of the fittest, any alive node v_i with *exactly some number* of alive neighbors maintains their alive state and thus their place in the spatial network SN_{GOL}

***R3*.** To simulate the dynamics of over population, any alive node v_i with *some number or more* of alive neighbors v_j dies and is removed from the spatial SN_{GOL}

***R4*.** To simulate the dynamics of reproduction, any dead node v_i with *exactly some number* of alive neighbors v_j becomes an alive node and is added to the spatial network SN_{GOL}

Although the influence of the cost matrix on system dynamics is not formally explored in the traditional Game of Life, a barrier is introduced into the GNA_{GOL} to demonstrate the use of the connection cost *C* in the GNA framework. The connection cost *C* is as follows:

***C1*.** A link cannot form between node v_i and node v_j if it intersects the barrier

For any model run, the underlying network structure UN_{GOL} is always the same, although the states of the nodes change. The UN_{GOL} and subsequently the SN_{GOL} are

initialized at time t_0 where 50% of the 2000 nodes are randomly selected as “alive”. The underlying network UN_{GOL} in the GNA_{GOL} implements a synchronized node update process. First, the number of alive neighbouring nodes is calculated for each node in the underlying network UN_{GOL}. Second, the transition rules *R* are applied and the state of each node changes based on its number of live neighbours. Finally, the new number of alive nodes is calculated. If the node is alive, the node stays or becomes part of the spatial network of interest SN_{GOL} and connects to its live neighbours. The GNA_{GOL} is run for 20 iterations to allow enough time to observe the evolving spatial network SN_{GOL} behavior.

3.4.2. GNA_{GOL} Scenarios

Three scenarios where developed by adjusting the transition rules *R1-R4* of the GNA_{GOL} model to represent different spatial system behaviours as simulated evolving spatial network SN_{GOL}. Scenario 1 uses the transition rules *R* presented in Table 3.4 and Figure 3.2a to generate a spatial network that fluctuates in a cyclic pattern with constant motion close to system equilibrium. To better understand how different geographical landscapes may influence network structure, scenario 1 is implemented twice using two different geographical landscapes including one with a water barrier and the other with street network as barriers. Scenario 2 uses transition rules *R* presented in Table 3.4 and Figure 3.2b to generate a spatial network that grows and expands. Scenario 3 uses the transition rules *R* presented in Table 3.4 and Figure 3.2c to generate a spatial network that shrinks and declines in size over space and time. Scenario 2 and scenario 3 are only implemented on a landscape with a water barrier. In all three scenarios, the connection cost *C1* remains the same, where a link cannot form between node v_i and node v_j if it intersects the barrier.

There are several real-world examples of systems as networks that may exhibit these types of behaviors over space and time. In a fluctuating network, the network grows and shrinks over time as dynamics cause nodes to become disconnected, newly added, and rewired. Despite these changes to network structure, the number of nodes remains relatively the same over time. One real world system that may exhibit a fluctuating network behavior is a human contact network as spatial social network structures evolve over time generating sets of repeating structures as nodes representing people move from home to work to recreational spaces. A growing network,

where the number of nodes and links increases consistently over time, may be representative of any type of spreading phenomena like insect infestation or spread of a computer virus. A shrinking network, where nodes are continuously removed from the network may be representative of the disconnection of a landscape connectivity network composed of a set of habitat patches that are connected based on the maximum dispersal distance of a species of interest as deforestation causes the removal of forest stand nodes or drought causes the loss of pond nodes.

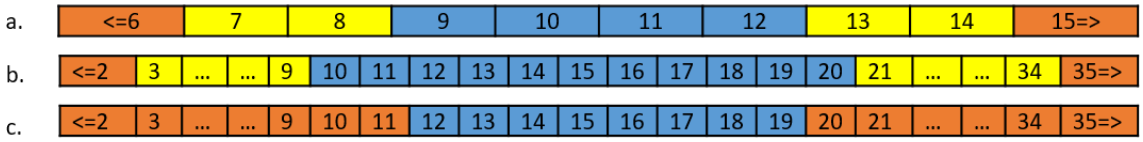

Figure 3.2. Parameterization of R1-R4 for (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3, where orange values indicate the number of neighbours that results in node death, yellow values indicate the number of neighbours that result in node survival, and blue values indicate the number of neighbours that result in node reproduction

Table 3.4. Transition rules and their parameterization specific to each scenario.

R1. To simulate the dynamics of **under population**, any live node v_i with…	Scenario 1	*6 or fewer*	… live neighbours v_i dies and is removed from the spatial network *SN.*
	Scenario 2	*2 or fewer*	
	Scenario 3	*11 or fewer*	
R2. To simulate the dynamics of **survival of the fittest**, any live node v_i with…	Scenario 1	*exactly 7, 8, 13, or 14*	… live neighbours v_i maintains their alive state and thus their place in the spatial network *SN.*
	Scenario 2	*=>3 and <=9 or =>21 and <=34*	
	Scenario 3	*N/A*	
R3. To simulate the dynamics of **over population**, any live node v_i with…	Scenario 1	*15 or more*	… live neighbors v_i dies and is removed from the spatial *SN.*
	Scenario 2	*35 or more*	
	Scenario 3	*20 or more*	
R4. To simulate the dynamics of **reproduction,** any dead node…	Scenario 1	*more than 8 or less than 13*	… live neighbors v_i becomes a live node and is added to the spatial network *SN.*
	Scenario 2	*more than 9 or less than 21*	
	Scenario 3	*more than 11 or less than 20*	

3.4.3. GNA_{GOL} Model Testing

The process of developing spatial models for representing complex systems includes several stages of model testing, typically comprised of model calibration, sensitivity analysis, and validation. The GNA_{GOL} model calibration was accomplished by using different parameter values for each rule *R1, R2, R3,* and *R4* and by verifying whether the model outputs produced meaningful behavior that corresponds to real-world spatial processes such as fluctuation, growth, or shrinkage. The GNA_{GOL} was tested for sensitivity to the structure of the underlying network *UN* as a function of the distance threshold d_{ij} that determines the neighbourhood of the nodes, in this case $d_{ij} <= 1\ km$. Therefore, we characterize the underlying network *UN* in the case that the distance threshold to connect neighbouring nodes is half that of d_{ij} where $d_{half} <= 0.5\ km$ and twice that of d_{ij} where $d_{double} <= 2\ km$ (Table 3.5). Most interesting is that doubling the distance threshold means that all nodes are connected to one third of all other nodes in the network. It is at this threshold that the clustering coefficient is impacted, meaning that the majority of nodes connected to node v_i are also connected to each other. This high clustering corresponds to a much lower average path length between any two pairs of nodes in the network.

Table 3.5. Sensitivity of the underlying network *UN* to the distance threshold d_{ij}.

Network Measure	Connectivity of Underlying Network		
	Sparse	Regular	High
Number of Nodes	2000	2000	2000
Number of Links	20554	64754	1374840
Average Clustering Coefficient	0.57	0.59	0.73
Average Degree (of all nodes)	10.85	32.44	687.42
Average Path Length	34.83	20.72	4.63

The transition rules *R* applied to the underlying network using these two alternative distance functions generate an immediate death at the first iteration to all nodes where node degree *k* is too high or too low to survive. This clearly demonstrates the tight coupling of network structure and network dynamics. Model validation was not performed as model outputs can not be compared to real spatial patterns as geospatial data was not used for this study.

3.5. Results

The output of the GNA_{GOL} are a series of spatial networks SN_{GOL} that evolve as a function of transition rules *R* that are applied to the underlying network UN_{GOL}. The choice to represent dynamics *on* a network is unique because of the interplay between the underlying network UN_{GOL} and the spatial network of interest SN_{GOL}. Specifically, the measured evolving network SN_{GOL} structure is limited by the underlying random geometric network structure UN_{GOL} and thus the GNA_{GOL} in all three scenarios produce an evolving spatial network SN_{GOL} that is *also* random. Since the degree of a node in the spatial network of interest SN_{GOL} cannot exceed the degree of the corresponding node in the underlying network UN_{GOL} and the links between nodes only exist in the case that the distance between two nodes are within the distance threshold, the underlying network UN_{GOL} does not allow for the emergence of a spatial network of interest SN_{GOL} that has properties of other graph types such as scale free or small world. Based on the random evolving spatial network structure SN_{GOL} produced by the GNA_{GOL}, the network structures observed and measured here characterize random geometric networks as they respond to dynamics that cause growth and shrinking responses. In this section, the obtained GNA_{GOL} simulation results are presented and the evolving spatial network SN_{GOL} is analyzed using graph theory measures.

3.5.1. GNA_{GOL} Simulation Results

The obtained simulation results from all three scenarios are presented in Figure 3.3. In all scenarios, following initialization, 50% of nodes are selected randomly as “alive” (Figure 3.3a).

Scenario 1

The application of the transition rules *R* designed to produce a fluctuating network initially results in several clusters that are loosely connected by thin chains of nodes (Figure 3.3b) in the water barrier landscape. At time *t,* live nodes with an ideal number of live neighbouring nodes survive onto the next iteration at time *t+1 (R2).* In addition, dead nodes with the ideal number of neighbouring live nodes at time *t* come to life at time *t+1 (R4)*. Live nodes without an ideal number of neighbouring live nodes at time *t* die from over population and under population at time *t+1*. The reproduction and

death of nodes at time *t+1* means that live nodes that have lost and gained live neighbouring nodes are at risk of dying in the following iterations. Subsequently dead nodes that have gained live neighboring nodes may have the potential of coming alive at the following iterations. These dynamics result in a delicate balance of node state, producing a spatial network *SN* composed of live nodes that evolves over time where clusters grow, shrink, merge, and re-organize over time. The network ultimately self-organizes into sets of repeating relatively stable loop-like clusters connected in a variety of ways (Figure 3.3 c, d, e). Specifically, chain and loop-like structures form as the interior of each cluster dies from overpopulation and the exterior of the cluster dies from under population, leaving the rest of the nodes in the cluster with the correct number of links survive until the next time step. In this particular landscape, the water barrier due to its location, does little to break up the overall network structure.

Comparison of Two different Landscapes. Figure 3.4 compares the evolving network structures that form based on the same transition rules for the fluctuating process, but for two different landscapes, one with a water barrier (Figure 3.4 a, b) and the other with dense street network (Figure 3.4 c, d) and for time t_5 (a, c) and t_{10} (b, d), respectively. In contrast to the landscape with the water barrier only, the landscape with more barriers, like streets in this case, have a major impact on the evolving network structure. Because the spatial interactions cannot take place with nodes that are separated by street features, the network becomes segregated into several small spatial clusters that are contained within a street block. It is more challenging for these clusters to survive, because they are limited to connect only with the nodes within their block, and as such, most clusters die or shrink over time.

Scenario 2

The application of the transition rules *R* developed specifically to simulate spatial network growth initially forms a configuration that is composed of thick clusters of nodes (Figure 3.3f). Since the transition rules *R* create an imbalance in favour of node reproduction *(R4)* and survival *(R2)*, the clusters expand over time as “dead” nodes close to the edge of clusters eventually have enough “live” neighbours that are required for them to reproduce and join the spatial network *SN* (Figure 3.3g, h, i). As such, the network as a whole grows over time as if “spreading”. In this scenario, the same loop-like configurations form where nodes internal to the cluster die from overpopulation.

Scenario 3

The application of the transition rules *R* developed specifically to simulate spatial network shrinkage initially forms a sparse set of clusters, where some clusters are connected to the larger spatial network and others are not (Figure 3.3j). This is a result of the transition rules *R* that are designed to reduce node reproduction and eliminate node survival. As a result, the network quickly shrinks until the network is reduced to a repeating sequence of relatively stable loop-like configurations. As soon as a configuration is produced that is unstable, resulting in the loss or gain of nodes, the network collapses and all nodes die from under population. The stable loop like configurations and the repeating patterns are similar to that of the patterns that are produced in the original version of the Game of Life.

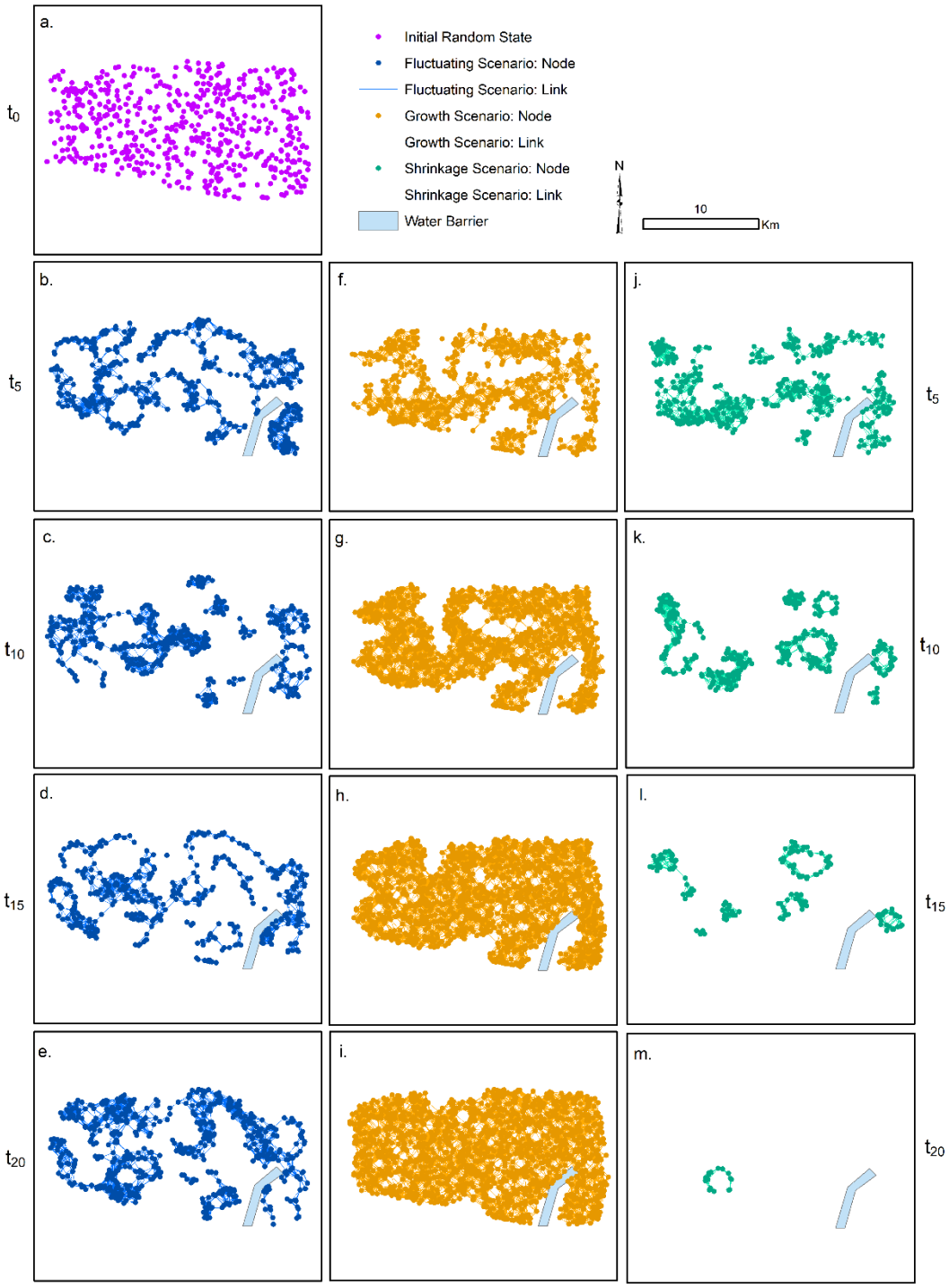

Figure 3.3. **Based on (a) the initial state, the evolving spatial networks SN_{GOL} generated from the GNA_{GOL} are presented for scenario 1 (b) t_5, (c) t_{10}, (d) t_{15}, (e) $t_{20,}$ scenario 2 (f) t_5, (g) t_{10}, (h) t_{15}, (i) $t_{20,}$ and scenario 3 (j) t_5, (k) t_{10}, (l) t_{15}, (m) t_{20}.**

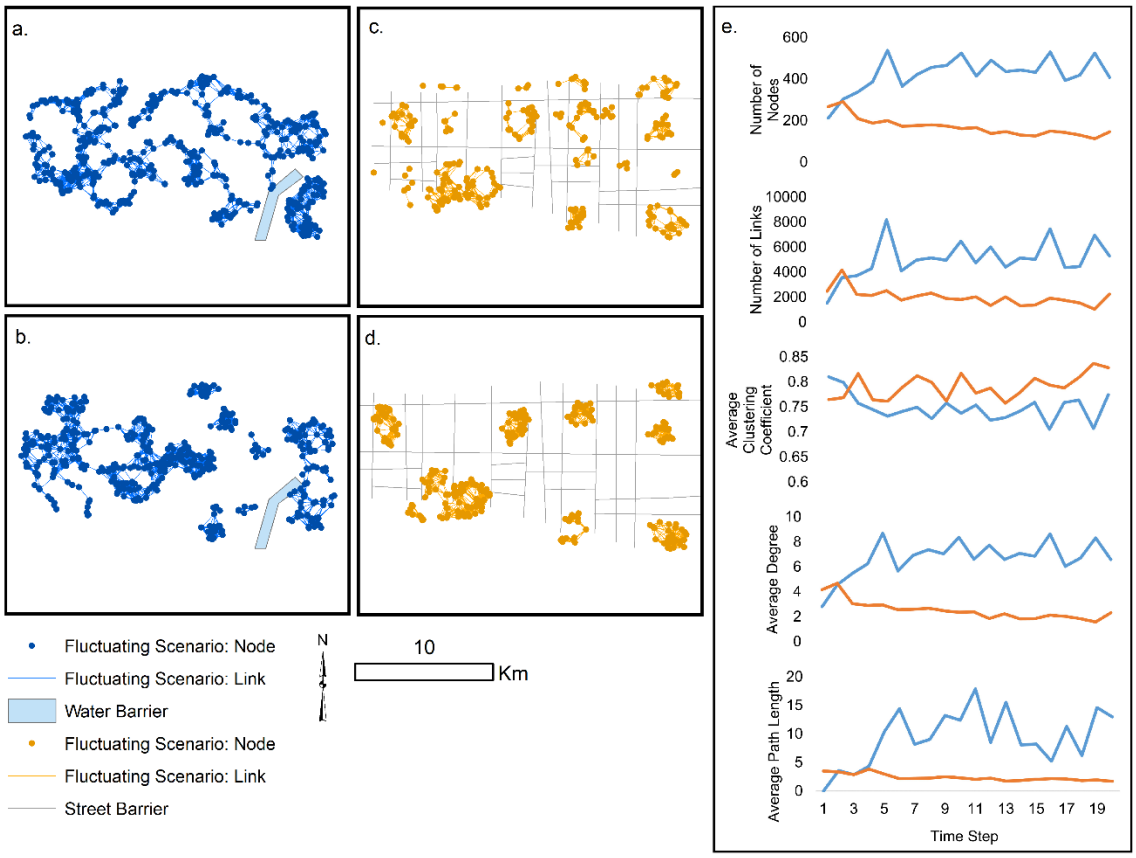

Figure 3.4. Simulation outcomes for scenario 1 fluctuating networks with (a, b) water barrier (blue) and (c, d) street barrier (orange) landscapes for time (a, c) t_5 and (b, d) t_{10} with (e) the values for the calculated number of nodes, number of links, average clustering coefficient, average degree, and average path length for each landscape and all model iterations.

3.5.2. GNA_{GOL} Spatial Network Analysis Results

The GNA_{GOL} produces a sequence of spatial networks SN_{GOL} as a function of the variations of the transition rules imposed on the UN_{GOL} in each scenario. The evolving network SN_{GOL} generated from the GNA_{GOL} can be characterized and quantified using graph theory measures to help explore the relationship between network structure and network dynamics. There are several important graph theory measures that are selected and applied based on their ability to mathematically characterize overall network structure and behavior including number of nodes *n,* number of links *m,* average degree <k>, degree distribution *P(k),* average clustering coefficient <C>, and the average shortest path length <l>. These are global measures that provide a complete snapshot of

network structure and can be used to clearly understand how the structure of a network evolves over time as a function of network dynamics.

Characterizing the Underlying Network UN

The underlying random spatial network UN_{GOL} has a node size of 2000, link size of 64754, an average clustering coefficient <C> of 0.588, an average degree <k> including all nodes that are both dead and alive of 32.44 that produces a *Poisson* degree distribution *P(k)*, and a path length of 20.72. Apart from node state, the underlying spatial network UN_{GOL} is static and does not change over time. The spatial network SN_{GOL} emerges from the applied transition rules *R* to the underlying network UN_{GOL} and evolves over time.

General Trends

The evolving spatial network SN_{GOL} is characterized by number of nodes, number of links, average clustering coefficient, average degree of alive nodes, and average path length calculated for each iteration for scenario 1 (Figure 3.5a), scenario 2 (Figure 3.5b), and scenario 3 (Figure 3.5c). In general, in scenario 1, the fluctuating network, all measures fluctuate over time as the network switches between a series of repeating stable configurations and thus the size of the network grows and shrinks over time. In scenario 2, the network growth scenario, the network grows steadily in size over time. The rate of growth is faster in early iterations and slows in later iterations as the network finds a stable configuration and less dead nodes are available to reproduce and join the network as live nodes. In scenario 3, the network shrinkage scenario, the network decreases in size over time. In this scenario, there are a few iterations where the number of nodes increases slightly before then decreasing.

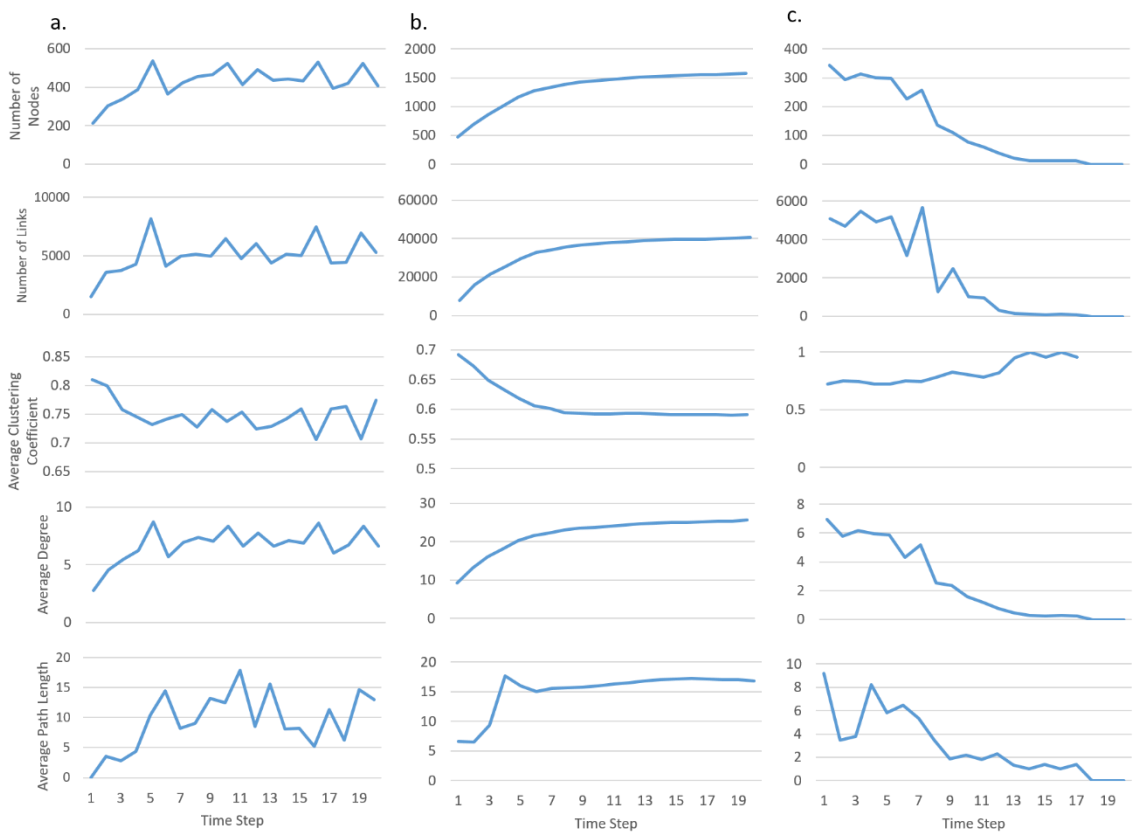

Figure 3.5. Calculated number of nodes, number of links, average clustering coefficient, average degree, and average path length for the obtained spatial network *SN* as it evolves over space and time generated from (a) scenario 1: fluctuating network, (b) scenario 2: network growth, and (c) scenario 3: network shrinkage.

Comparison of Two Different Landscapes

Figure 3.4e presents the comparison of the obtained network measures (number of nodes n, number of links m, average clustering coefficient <C>, average degree of alive nodes <k>, and average path length <l>) calculated for each iteration for scenario 1 with the water barrier landscape (in blue) and with the street barrier landscape (in orange). Because the street barriers segregate the spatial network into small spatial clusters where the number of possible connections are limited, the street barrier landscape network has fewer nodes and links, a smaller average degree <k> and a smaller average path length <l>.

Correlation between Graph Theory Measures

Table 3.6 presents the correlation between graph theory measures obtained from the generated spatial networks in scenario 1 (Table 3.6A), scenario 2 (Table 3.6B), and scenario 3 (Table 3.6C). In all three scenarios, as each of the spatial networks SN_{GOL} evolve, a strong positive correlation is maintained between the *number of nodes* and the *number of links*, the *number of nodes* and *average degre*e, and the *number of links* and *average degree*.

Scenario 1. As the fluctuating network evolves, a moderate negative correlation between *average clustering coefficient* and the *number of nodes*, *number of links,* and *average degree* is maintained. This indicates that as the size of the network increases, the clustering coefficient decreases, explained by the decreased ratio between the number of links between nodes v_j that are connected to node v_i and the number of links that could exist between them. In the fluctuating network, there appears to be a weak relationship between *average path length* and the rest of the graph theory measures, demonstrated by the weak positive correlation between *average path length* and the *number of nodes*, *number of links,* and *average degree*. Additionally, there is a weak negative correlation between *average path length* and *average clustering coefficient*.

Scenario 2. Like the spatial network SN_{GOL} produced in scenario 1, the growing spatial network SN_{GOL} structure exhibits a strong negative correlation between *average clustering coefficient* and the *number of nodes*, *number of links,* and *average degree*. This indicates that are the size of the network increases, the clustering coefficient decreases. This relationship is stronger in the growing spatial network than in the fluctuating network. Unlike the stead state network structure, there is a strong negative correlation between *average clustering coefficient* and *average path length*, where as clustering coefficient increases, path length decreases. This is a logical relationship, since as nodes become increasingly and more tightly connected, the path length between any two nodes in the network would naturally decrease. Strong positive correlations are also found to exist between *average path length* and *number of links, number of nodes, and average degree*, where as when the network size increases, the path length increases.

Scenario 3. The shrinking network structure SN_{GOL} also exhibits strong positive correlations between *average path length* and *number of links*, *number of nodes,* and

average degree. There is a moderate negative correlation between *average clustering coefficient* and *number of nodes* and *average degree*. Therefore, as the network shrinks, the average degree decreases and the average clustering coefficient increases. There is a weaker negative correlation however between *average clustering coefficient* and *number of links* and an even weaker negative correlation between *average clustering coefficient* and *average path length*.

In the scenarios of network growth and shrinkage there is also a positive correlation between number of nodes and path length. This is not the case in the fluctuating network, meaning that there may be more complex relationships between graph theory measures than observed here. For example, a long average path length may emerge from a network structure that has a unique combination of a high number of nodes *and* a higher than expected clustering coefficient.

Table 3.6. Correlation table presenting the correlation between each of the graph theory measures numbered (1) to (5) obtained from the evolving spatial network generated by (a) scenario 1: fluctuating network, (b) scenario 2: network growth, and (c) scenario 3: network shrinkage.

Scenario 1: Spatial Fluctuating Network					
	(1)	(2)	(3)	(4)	(5)
(1) Number of Nodes					
(2) Number of Links	0.89				
(3) Average Clustering Coefficient	-0.7	-0.59			
(4) Average Path Length	0.24	0.14	-0.17		
(5) Average Degree	0.98	0.9	-0.71	0.2	
Scenario 2: Spatial Network Growth					
	(1)	(2)	(3)	(4)	(5)
(1) Number of Nodes					
(2) Number of Links	0.99				
(3) Average Clustering Coefficient	-0.97	-0.98			
(4) Average Path Length	0.79	0.79	-0.82		
(5) Average Degree	0.99	0.99	-0.98	0.8	
Scenario 3: Spatial Network Shrinkage					
	(1)	(2)	(3)	(4)	(5)
(1) Number of Nodes					
(2) Number of Links	0.94				
(3) Average Clustering Coefficient	-0.73	-0.67			
(4) Average Path Length	0.74	0.94	-0.58		
(5) Average Degree	0.99	0.95	-0.72	0.74	

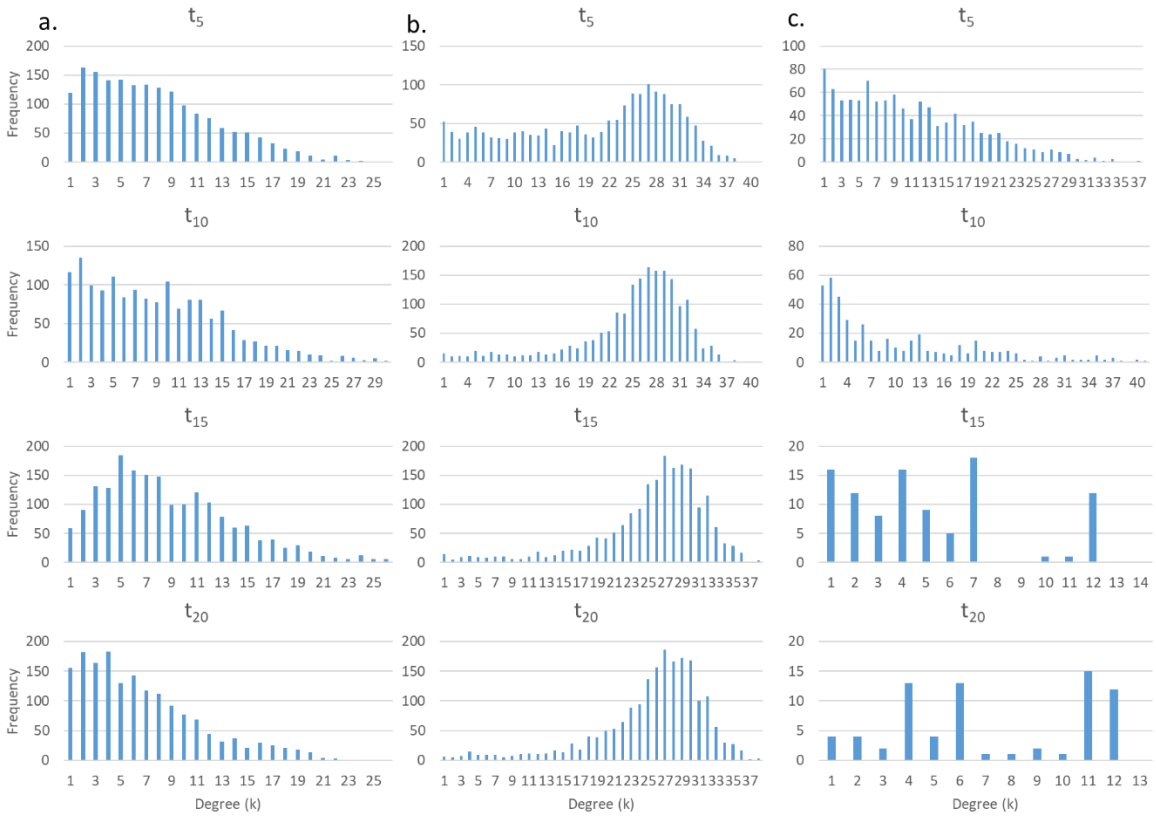

Figure 3.6. Calculated degree distribution for each evolving spatial network *SN* as it evolves over time generated from (a) scenario 1: fluctuating network, (b) scenario 2: network growth, and (c) scenario 3: network shrinkage.

Degree Distribution

When an evolving network undergoes growth (scenario 2) and network size increases, node degree increases, which produces a degree distribution with a left negative skew (Figure 3.6b). When an evolving network undergoes shrinkage (scenario 3), nodes are removed, leaving remaining nodes with a smaller degree, producing a degree distribution with a right positive skew (Figure 3.6c). Scenario 1 produces less of a consistent degree distribution, as the network undergoes both growth and shrinkage in tandem. In most cases, the network shrinks, but at some iterations, there is a shift to increased frequencies of nodes with a higher degree (Figure 3.6a).

3.6. Discussion and Conclusions

This study introduces the novel modelling framework of Geographic Network Automata (GNA) that can be used for the representation and analysis of complex spatio-temporal systems as evolving and dynamic networks. The GNA modelling framework is implemented using a spatially explicit network variant of the classic Game of Life model (Conway, 1970). Three scenarios have been implemented to represent different spatial processes as well as the influence of different geographic landscape environments. Using graph theory, the simulated evolving networks are characterized and quantified, and in this case, facilitate the exploration between evolving random spatial networks and their space-time dynamics. In general, for the evolving spatial networks SN_{GOL} simulated in all scenarios, as the number of nodes increases, the number of links and the average degree increases and the clustering coefficient decreases, supporting conclusions that graph theory measures are dependent on network size (Wijk et al., 2010). These correlations are particularly interesting because the relationship between network size and other graph measures are not well understood and rarely explored in the literature, especially in the case of spatially embedded networks. Wijk et al. (2010) mathematically calculates correlations between network size and different graph theory measures for a traditional non-spatial random network and find that increasing the number of nodes results in an increase in average path length and a decrease in clustering coefficient. This relationship appears to hold true in the case of random spatial networks as demonstrated by the GNA_{GOL} simulated spatial networks SN_{GOL}. Future work may involve further exploring these relationships between network structure and dynamics.

The Game of Life is a hypothetical spatial system, selected as a case study in order to clearly demonstrate the GNA modelling framework. Therefore, the GNA_{GOL} model does not use geospatial datasets, and thus full model validation was not performed. Model validation is seen as the degree of agreement between simulated spatial network structures and observed real-world spatial patterns, which should be evaluated using datasets that are independent of model development and calibration. Future work would require the exploration of model validation approaches suitable for comparing evolving networks with real datasets even in the face of data scarcity.

The GNA is a relatively novel and unexplored modelling approach, its framework is both general and flexible so that many types of spatio-temporal phenomena can be

simulated. In the case of the application of the GNA modelling framework to real world phenomena, the elements of the GNA including the initial network state, the underlying network, the transition rules, the connection cost, and the spatial and temporal resolution would need to be designed to properly reflect the real system and include real geospatial data for GNA development, calibration, and validation. One might consider many applications that can be conceptualized as a network where its evolution is driven by simple dynamics that occur *on* the network or *of* the network or even both. This modeling framework might be useful include urban, social, and ecological applications. For example, dynamics of social contact might be simulated as an evolving underlying network *UN* from which dynamics of the spread of influenza might emerge as a spatial network of interest *SN.* The proposed GNA modelling approach presented in this study is an entirely novel class of geographic automata (Torrens and Benenson, 2005) and represents a strong departure from the classic cell-based geographic automata models.

3.7. References

Albert, R., Jeong, H., & Barabási, A. L. (1999). Internet: Diameter of the world-wide web. *Nature*, *401*(6749), 130-131.

Antonioni, A., & Tomassini, M. (2012). Degree correlations in random geometric graphs. *Physical Review E*, *86*(3), 037101.

Bansal, S., Read, J., Pourbohloul, B., & Meyers, L. A. (2010). The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics,* 4(5), 478-489.

Barabási, A. L. (2016). Network Science. Cambridge, UK: Cambridge University Press.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.

Barthélemy, M. (2011). Spatial networks. *Physics Reports,* 499(1-3), 1-101.

Batty, M., Couclelis, H., & Eichen, M. (1997). Urban systems as cellular automata. Environment and Planning B: Planning and Design, 24, 159-164.

Bian, L. and Liebner, D. 2007. A network model for dispersion of communicable diseases. *Transactions in GIS,* 11(2), 155-173.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4-5), 175–308.

Boccara, N., & Cheong, K. (1993). Critical behaviour of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. *Journal of Physics A: Mathematical and General*, *26*(15), 3707.

Boccara, N., Roblin, O., & Roger, M. (1994). Automata network predator-prey model with pursuit and evasion. *Physical Review E*, *50*(6), 4531.

Cohen, R., Erez, K., Ben-Avraham, D., & Havlin, S. (2000). Resilience of the internet to random breakdowns. *Physical Review Letters,* 85(21), 4626.

Colizza, V., Barrat, A., Barthélemy, M., & Vespignani, A. (2006). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, *103*(7), 2015-2020.

Conway, J. (1970). The game of life. *Scientific American,* 223(4), 4.

Dall, J., & Christensen, M. (2002). Random geometric graphs. *Physical review E*, *66*(1), 016121.

Ducruet, C., & Notteboom, T. (2012). The worldwide maritime network of container shipping: spatial structure and regional dynamics. *Global Networks*, *12*(3), 395-423.

Erdos, P., & Renyi, A. (1959). On Random Graphs I. *Selected Papers of Alfred Renyi*, *2*(1), 308–315.

Erdos, P., & Renyi, A. (1960). On the Evolution of Random Graphs. *Bulletin of the International Statistical Institute*, *34*(4), 343–347.

Fortuna, M. A., Gomez-Rodriguez, C., and Bascompte, J. (2006). Spatial network structure and amphibian persistence in stochastic environments. *Proceedings of the Royal Society B: Biological Sciences,* 273(1592), 1429–1434.

Grimm, V., & Railsback, S. F. (2012). Designing, formulating, and communicating agent-based models. In *Agent-based models of Geographical Systems*. Dordrecht: Springer.

Gross, T., & Sayama, H. (2009). *Adaptive Networks*. Berlin, Heidelberg: Springer.

Heppenstall, A. J., Crooks, A. T., See, L. M., & Batty, M. (2011). *Agent-based Models of Geographical Systems*. Springer Science & Business Media.

Jiang, B. (2009). Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science*, *23*(8), 1033-1048.

Lewis, T. G. (2011). *Network Science: Theory and Applications*. Hoboken, NJ: John Wiley & Sons.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review,* 45(2), 167-256.

O'Sullivan, D. (2002). Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems*, *4*(3), 251-274.

Repast Simphony. 2016. Version 2.4. [Computer Software]. Chicago, IL: University of Chicago.

Sayama, H., and Laramee, C. (2009). Generative network automata: A generalized framework for modeling adaptive network dynamics using graph rewritings. *Adaptive Networks,* 311–332.

Scellato, S., Noulas, A., Lambiotte, R., & Mascolo, C. (2011). Socio-spatial properties of online location-based social networks. ICWSM, 11, 329-336.

Smith, D. M., Onnela, J. P., Lee, C. F., Fricker, M. D., & Johnson, N. F. (2011). Network automata: Coupling structure and function in dynamic networks. *Advances in Complex Systems,* 14(03), 317-339.

Torrens, P. M., and Benenson, I. (2005). Geographic automata systems. International Journal of *Geographical Information Science,* 19(4), 385-412.

Van Wijk, B. C., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PloS one*, *5*(10), e13701.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*(6684), 440–442.

Yule, G. U. (1925). Mathematical theory of evolution based on the conclusion of Dr. J.C. Willis*. Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character, 213(1), 21–87.*

Zhong C., Arisona S. M., Huang X., Batty, M. & Schmitt G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science,* 28(11), 2178-2199.

Chapter 4.

A Geographic Network Automata Approach for Modelling Dynamic Ecological Systems[2]

4.1. Abstract

Landscape connectivity networks are composed of nodes representing georeferenced habitat patches that link together based on a species' maximum dispersal distance. These static representations cannot capture the complexity in species dispersal where the network of habitat patch nodes changes structure over time as a function of local dispersal dynamics. Therefore, the objective of this study is to integrate geographic information, complexity, and network science to propose a novel Geographic Network Automata (GNA) modelling approach for the simulation of dynamic spatial ecological networks. The proposed GNA modelling approach is applied to the emerald ash borer (EAB) forest insect infestation using geospatial datasets from Michigan, USA and simulates the evolution of the EAB spatio-temporal dispersal network structures across a large regional scale. The GNA model calibration and sensitivity analysis are performed. The simulated spatial network structures are quantified using graph theory measures. Results indicate that the spatial distribution of habitat patch nodes across the landscape in combination with EAB dispersal processes generate a highly connected small-world dispersal network that is robust to node removal. The presented GNA model framework is general and flexible so that different types of geospatial phenomena can be modelled, providing valuable insights for management and decision-making.

4.2. Introduction

Interactions between ecological entities are commonly represented using aspatial networks such as food webs (Cohen 1978; Hall and Raffaelli 1993), host-parasitoid webs (Muller et al. 1999; Morris et al. 2004), and mutualistic webs (Jordano 1987; Stang et al. 2006). Using these representations, graph theory can be applied to mathematically

[2] A version of this chapter is accepted and in press for publication: Anderson, T. & Dragicevic, S. (In Press). A Geographic Network Automata approach for modelling dynamic ecological systems. Geographical Analysis.

characterize ecological network topology, cross-compare between network structures to find common and unique patterns, and better understand the close coupling between network structure and network dynamics (Ings et al. 2009). Although useful, these aspatial representations are unable to address complexity of the spatial structure of ecological networks (Campbell Grant et al. 2007).

Consequently, several studies apply graph theory to landscape connectivity networks to represent and analyze the connection between landscape spatial structure and species dispersal (Urban and Keitt 2001; Minor and Urban 2007; Fortuna et al. 2006; Bunn et al. 2000). In these representations, network nodes represent georeferenced habitat patches and network links represent the potential dispersal between habitat patches based on the maximum dispersal distance of the species of interest (Fortuna et al. 2006). These studies are typically motivated by species conservation and focus on better characterizing patterns of dispersal across the landscape, identifying important habitat patches and dispersal pathways, and determining the effect the removal of habitat patches by way of external factors such as deforestation or forest fires has on landscape connectivity.

Despite their potential, landscape connectivity networks represent dispersal as the movement between a fixed set of habitat patch nodes that are organized in a static spatial topology (Sayama and Laramee 2009). This representation assumes that dispersal over time takes place between the same set of habitat patch nodes and is limited to representing dispersal as a process driven only by maximum dispersal distance. However, in the real-world, dispersal processes in ecological systems are considered complex and dynamic as there are instances where the network of habitat patch nodes expands, shrinks, or changes structure as a function of species competition, adaptation, population dynamics, invasion, extinction, or seasonal migration of ecological species. These processes are of interest to ecological conservation and management, making it beneficial for representing dispersal using a dynamic network that can evolve over space and time.

Network evolution is a property of complex networks and is defined as the change in network topology as a function of the dynamics of the phenomena itself as nodes are added, rewired, or removed (Smith et al. 2011). The theory of network automata (Sayama and Laramee 2009), used for the representation of evolving network

structures, can be integrated with geospatial data and geospatial models to provide a modelling framework that can represent species dispersal as dynamic spatial networks. In dynamic spatial networks, changes in network topologies are a function of the dynamics of the network itself, rather than from external influences (Smith et al. 2011).

Therefore, the main objectives of this study are to (1) propose the theoretical framework of a novel modeling approach termed Geographic Network Automata (GNA) that integrates geographic information systems (GIS), theories of complex systems, and network automata, and (2) apply the GNA approach to model spatio-temporal dispersal patterns of ecological phenomena at the regional scale. The proposed GNA is a bottom-up approach that extends the theoretical and applied framework of geographic automata (Torrens and Benenson 2005). Specifically, the GNA generates a spatial network structure that evolves as a function of rules that are applied repetitively at each time step and govern the spatial dynamics, transformation, and emergence of network topologies. The proposed GNA modelling approach is demonstrated by simulating the evolving dispersal network of an invasive bark beetle, the emerald ash borer (EAB), using geospatial datasets covering a large-scale region in Michigan, USA (Siegert et al. 2014). The simulated dispersal network structures are quantified and analysed using graph theory measures to represent and understand the spatial dynamics of the networks of EAB dispersal processes at the regional scale.

4.3. Theoretical Framework of the Geographic Network Automata

A geographic network automaton (GNA) is a mathematical representation of a complex dynamic *spatial network SN*. Network dynamics over time are imposed through the application of transition rules *R* that *establish* connections between nodes and connection costs *C* that *restrict* connections between nodes (Figure 4.1). Transition rules *R* and connection costs *C* are applied at each model time step *(t+1)* as transition rules governing the evolution and dynamics of the spatial network *SN*.

Therefore, the GNA can be expressed as:

$$GNA = [SN, R, C, \Delta t] \quad (1)$$

where *SN* is the spatial network, *R* are the transition rules, *C* is the connection cost, and Δ*t* is the temporal increment of the GNA.

The spatial network *SN* generated by the GNA is composed of a set of nodes *N* representing system components and a set of links *L* representing an interaction or relationship between nodes and is expressed as:

$$SN = [N^t, L^t, A^t] \qquad (2)$$

where the spatial network *SN* at initial time *t* consists of a set of nodes *N,* a set of links *L,* and an adjacency matrix *A* of dimension $N \times N$ storing network topology as records of connections between pairs of nodes.

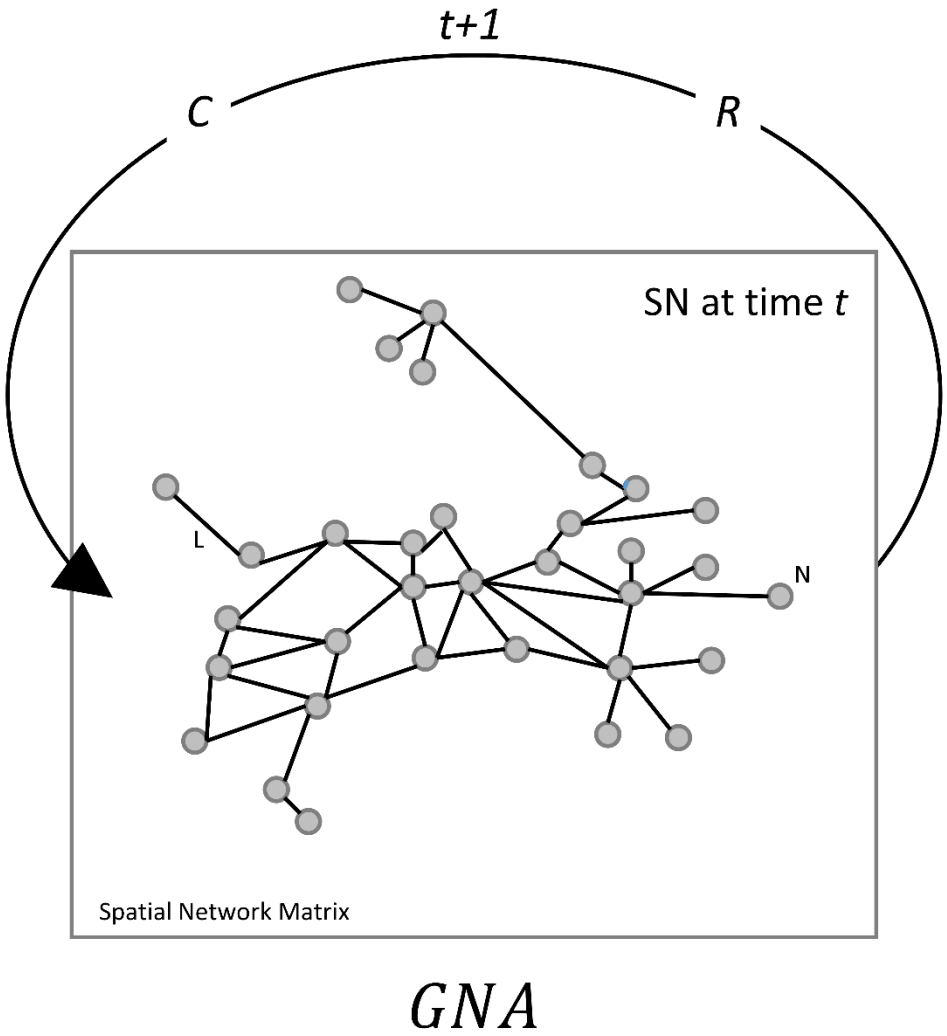

Figure 4.1. Overview of GNA model structure.

The transition rules *R* and the connection cost *C* influence the information recorded within the adjacency matrix at each time step, updating the nodes and recording changes in their connections to each other. These act as transition rules governing the evolution of the network. Therefore, network evolution with each time step is defined where the adjacency matrix at the subsequent time step *t + 1* is a function of the transition rules *R,* the connection cost *C,* and the adjacency matrix at the previous time *t,* as follows:

$$A^{t+1} = R, C(A^t) \qquad (3)$$

The sets of nodes *N* and links *L* in the spatial network *SN* are further expressed as:

$$N = [v_1, v_2, \dots, v_n] \qquad (4)$$

$$L = [e_1, e_2, \dots, e_m] \qquad (5)$$

where $v_1, v_2, \dots, v_n$ denote the individual nodes that make up the larger set of nodes *N* and $e_1, e_2, \dots, e_m$ denote the individual links that make up the larger set of links *L.*

Figure 4.2 presents an example of two simple spatial networks to demonstrate this notation, where Figure 4.2a is an undirected network, Figure 4.2c is its associated adjacency matrix, Figure 4.2b is a directed network, and Figure 4.2d is its associated adjacency matrix. Both networks in Figure 4.2 have the same number of nodes and links and can be expressed as:

$$N = [v_1, v_2, v_3]; n = 3 \quad (6)$$

$$L = [e_1, e_2]; m = 2 \qquad (7)$$

The adjacency matrix for the example of the simple spatial networks is presented in Figure 4.2d and 4.2d, where a value of 1 denotes the presence of a link between two nodes.

In the undirected network (Figure 4.2a and 4.2c), the connection between v_1 *and* v_2 *via* link e_1 and the connection between v_2 *and* v_3 via link e_2 is recorded using a value of 1 in this matrix. As an undirected network, these connections are bidirectional and thus the connection is also recorded as existing between v_2 *to* v_1 and v_3 *and* v_2. The

absence of a connection between any two nodes is recorded using a value of 0. In the directed network (Figure 4.2b and 4.2d), a directed link only exists *from v_2 to v_1* and is recorded in the matrix using a value of 1. Since this connection is unidirectional, no link is recorded *from v_1 to v_2.*

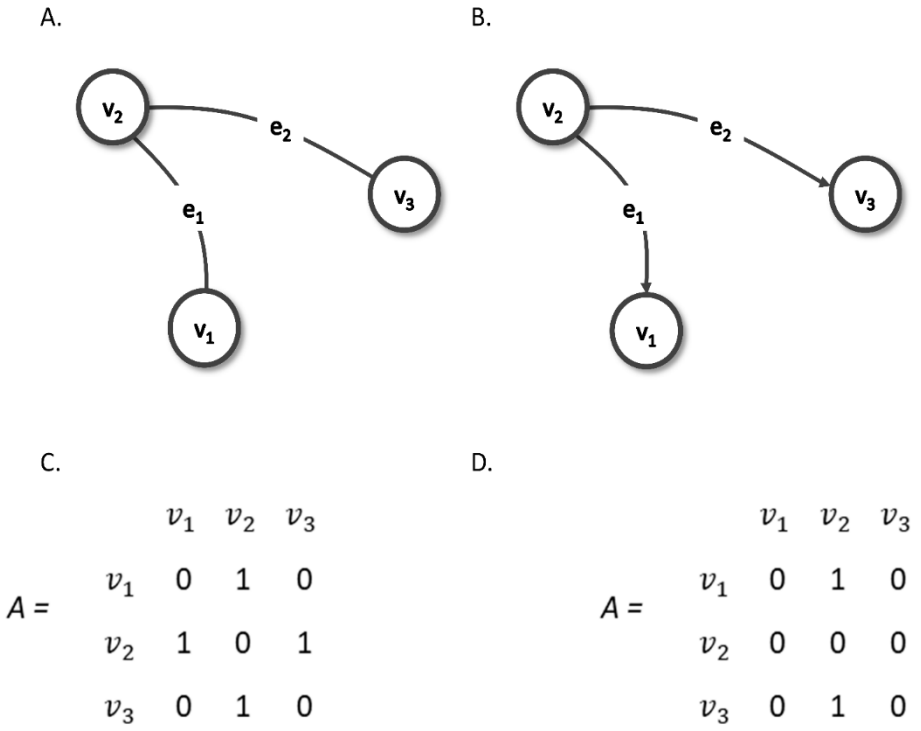

Figure 4.2. Example of simple spatial networks: (a) undirected (b) directed and the corresponding adjacency matrices (c) undirected and (d) directed.

Nodes *N* in a spatial network *SN* are embedded in geographic space and thus *always* contain the property of geographic location $g = [x, y]$, therefore, the distance *d* between them can be measured. Individual nodes v_i also contain the property of node weight *w*, network characteristics such as the number of connections to other nodes or degree *k*, and in addition, can contain several non-spatial attributes associated with the real-world system component the node v_i represents. Node weight *w* is a value assigned to each node ranking the nodes importance, suitability, or preference within the set of nodes *N* in the network. In most cases, the higher the node weight *w*, the more likely that the node will be connected to.

The links *L* in a spatial network *SN* only contain the property of geographic location *g* in the case that the network is embedded in planar space such as transportation networks. In this case, links *L* embedded in planar geographic space also contain the property of length. Additionally, links *L* may contain the property of link weight *w,* representing the magnitude of flow between nodes, for example, flow of individuals, material and information. Links *L* may be unidirectional or bidirectional, where flow occurs in one or both directions. As time *t* passes, the dynamic spatial network evolves where the number of nodes *n* and the number of links *m* change. In this process, already connected nodes make connections to new nodes (Figure 3.3a), become disconnected from the network (Figure 4.3b), or are rewired to connect to other already connected nodes (Figure 4.3c).

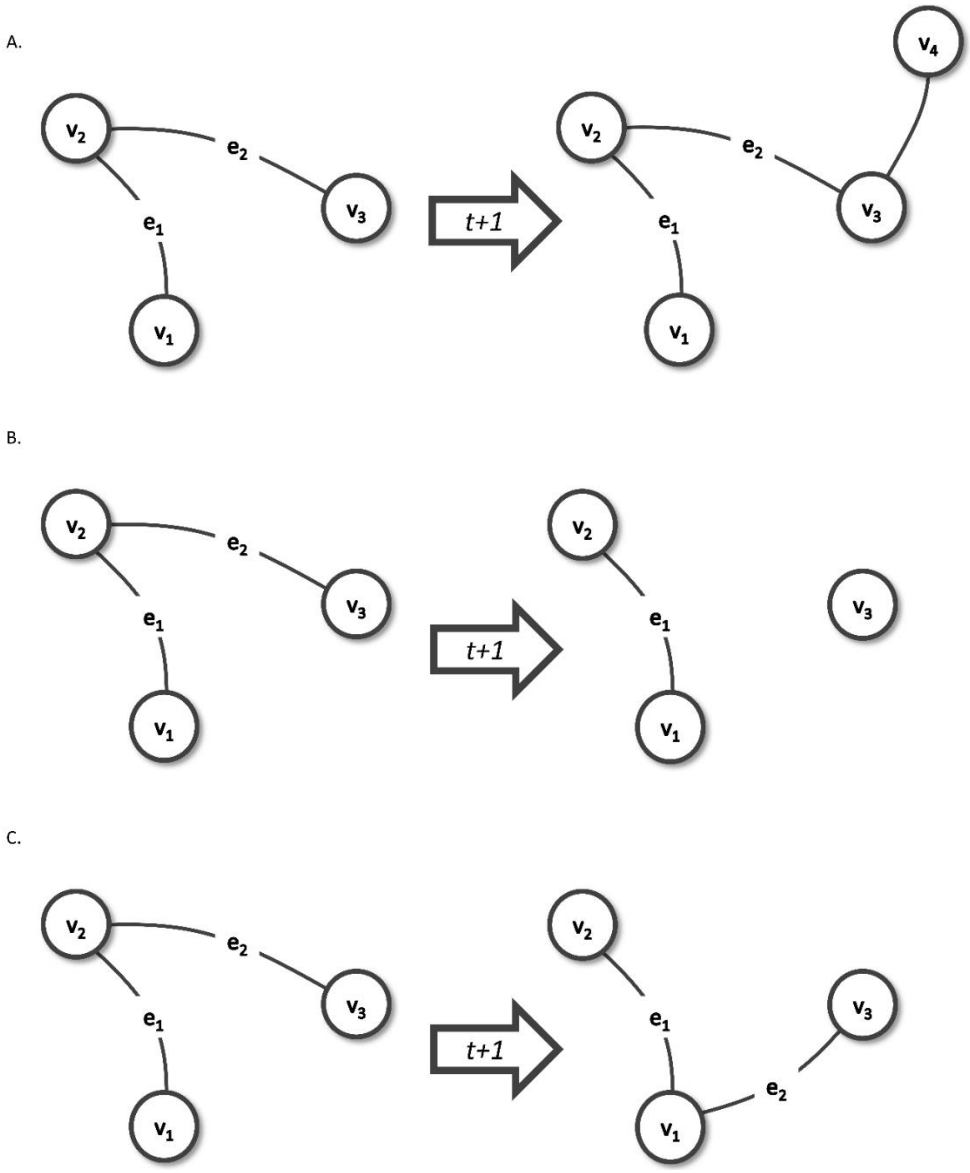

Figure 4.3. Spatial network evolution via (a) new connections, (b) node disconnection, and (c) node rewiring.

Transition rules *R* and connection cost *C* govern the formation of connections between nodes as the transition rules of the network automata and are designed to represent the real-world interactions between system elements. The transition rules *R* and connection cost *C* are part of the GNA theoretical modelling framework and ultimately determine the dynamics and evolution of the spatial network *SN.* Transition rules *R* define the likelihood of a link forming between nodes as a function of node and link properties such as directionality, weight, location, and attributes. For example, a transition rule *R* may be defined where the formation of a link from node v_i to node v_j is

more likely if node v_j is weighted heavily, is defined by a certain attribute, or is in a desirable location.

Connection cost *C* determines the resistance of the geographic space between two nodes to the formation of links. The space between two nodes is referred to as the cost matrix. Resistance in this context is a function of distance between nodes and the existence of barrier objects between nodes that impede connection. In many real-world systems, long distance connections cost time, energy, and money (Barthélemy 2011). Because the network is embedded in geographic space, making a connection between long-distance nodes or nodes with a barrier between them may be less likely or impossible.

The novel GNA modelling framework presented here is general and flexible in a manner that can be easily applied to represent complex dynamic spatio-temporal phenomena, but specifically here it is applied to dispersal networks of complex ecological systems. Particularly, this approach is useful for representing cases of dynamic spatial ecological systems where new habitat patch nodes are dispersed to, old habitat patch nodes are abandoned, or well-established dispersal paths change. The GNA framework is applied to the case study of the forest insect infestation, the emerald ash borer (EAB), for the development of the GNA_{EAB}.

4.4. GNA_{EAB} Modelling Methodology

The proposed GNA modelling framework is implemented on the case study of the invasive bark beetle, the emerald ash borer (EAB), primarily to model dynamic spatial networks generated by EAB interactions with landscape features at regional scales. The EAB spreads across its non-native landscape in search of forest stands containing ash trees to sustain larvae and after several years of infestation, the forest stands die. As such, this is a clear example of an ecological system where the number of habitat patch nodes expand, shrink, and rewire over time, making it a useful case study.

The following sections provide a brief biological background of EAB, the datasets used, the detailed description of GNA_{EAB} model development, and an overview of GNA_{EAB} model testing. The GNA_{EAB} is developed in REcursive Porous Agent Simulation Toolkit (Repast) and programmed using the Java programming language (Repast 2016).

4.4.1. Characteristics of Emerald Ash Borer Insect Infestation

The emerald ash borer (EAB), *Agrilus planipennis*, is an invasive bark beetle native to Asia (Muirhead et al. 2006). The EAB infestation was first discovered in Michigan, USA in 2002, however extensive dendrochronological surveying reveals that the EAB population may have become established in the region as early as 1997, before being discovered (Siegert et al. 2014). The EAB target and kill ash trees and have caused significant ecological, economic, and social impacts across the United States and Canada, motivating a research agenda to better understand EAB biology, life cycle, behavior, future economic impacts, climate impacts, and patterns of dispersal. Processes of EAB dispersal at the landscape scale drives EAB spatial dynamics at the greater regional scale and form the basis for the transition rules *R* and connection costs *C* implemented in the GNA_{EAB} model.

Adult EAB, with an average lifespan of 28 days, emerge in mid-June and search for food, mates, and suitable hosts to sustain larval galleries (Herms and McCullough 2014). The EAB infestation is governed via two dispersal mechanisms including: 1) local or short-distance dispersal and 2) long-distance dispersal where satellite populations become established far from the main invasion front. Eventually, local and long-distance dispersal populations grow and coalesce, increasing the overall rate of spread, a process referred to as stratified dispersal. Long-distance dispersal is not a function of the local dispersal of the EAB, but instead, is a function of human intervention, typically through the transportation of infested wood products, primarily firewood, or the inadvertent hitchhiking of mated females on travelling vehicles (Muirhead et al. 2006). Adult EAB cease activity every year in August, but larvae live under the bark of the ash trees for one to two years before emerging as adults (Herms and McCullough 2014).

Observed rates of EAB spread including local and long-distance dispersal vary widely within the literature from 12.87 km/year (Siegert et al. 2014) to 40 km/year (Straw et al. 2013), as rates of spread are study site dependent. Specifically, the observed rates of spread are a function of the spatial arrangement of the hosts at each study site, the scale at which observations are collected, and the success of infestation detection. Taylor et al. (2010) reports an observed flight distance by mated females of >3 km with 20% flying >10 km and 1% flying >20 km using flight mill tethering.

4.4.2. Study Site and Data Sets

The developed GNA_{EAB} model is implemented on datasets for the state of Michigan, USA, considered the original epicenter of EAB infestation in North America. The datasets used for the implementation of the GNA_{EAB} simulation model include the following:

1. **Location of forest stands containing ash**: GIS point data representing surveying plots (8 km^2) containing ash trees (accurate to 0.8-1.6 km) and type (white, black, and green ash) were obtained from the U.S. Forest Service (USFS). The dataset was developed during a surveying period from 2001-2005 and was funded for a 3x plot intensification providing a higher than average surveying density. For the purpose of this study, the term *forest stand* is defined as a group of trees within the plot boundaries. There are 3319 forest stands containing ash tree species. Each forest stand point identifies the presence or absence of white, black or green ash within the stand. Across the study area, white ash trees are identified in 1065 stands, black ash are identified within 1310 stands, and green ash are identified within 1340 forest stands. The dataset serves as the basis for habitat patch nodes *N* in the developed *SN* representing forest stands containing ash trees at the landscape scale. The forest stands that contain ash trees are distributed in a clustered pattern.

2. **Location of EAB epicenter:** A GIS data layer containing the original 1997 epicenter of EAB spread in Michigan, USA, identified using dendrochronological reconstruction, was developed by georeferencing and digitizing Siegert et al. (2014)'s findings.

3. **EAB year of discovery for each Michigan county**: A GIS data layer that was obtained from the Emerald Ash Borer Information Network, a collaborative effort of the USDA Forest Service and Michigan State University to provide comprehensive, accurate and timely information on the emerald ash borer.

4. **Extent of Michigan urban areas:** GIS raster files of 30m resolution containing the extent of urban areas in Michigan were obtained from the US Geological Survey (GUSGS) EarthExplorer Global Land Cover Characterization (GLCC).

5. **Other Michigan data:** GIS data layers containing the location and extent of lakes and rivers (Lake Polygons 2018), the location of campgrounds (Michigan State Park Campgrounds 2018), all major roads and highways (All Roads v17a 2018), and census tract data (Census Tracts- Population, Income 2018) for Michigan were obtained from the Michigan open data catalog.

4.4.3. GNA_{EAB} Model Overview

The GNA_{EAB} model represents spatio-temporal patterns of EAB insect infestation as a dynamic network and consists of a spatial network SN_{EAB} that evolves as a function of transition rules *R* and connection cost *C* applied at each time step. Here, the GNA framework is described in application to the EAB phenomenon. Figure 4.4 presents the workflow of the model as a useful guideline.

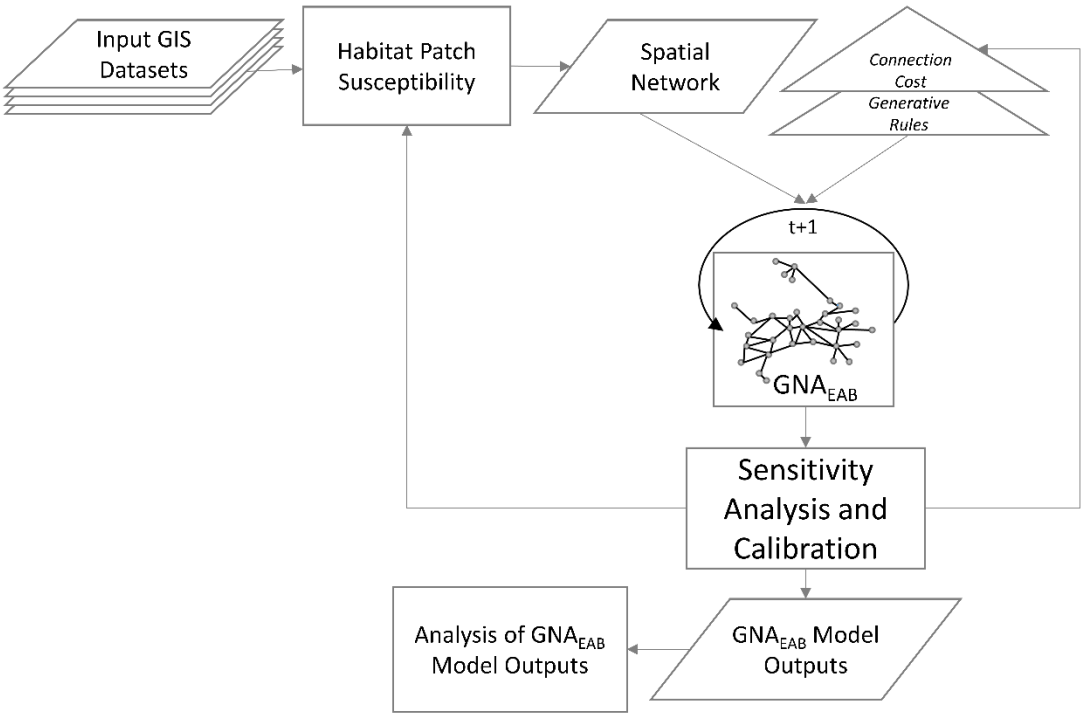

Figure 4.4. Workflow of the GNA$_{EAB}$ model of forest insect infestation.

Spatial Network Structure SN$_{EAB}$

The spatial network for representing insect infestation SN_{EAB} is defined as:

$$SN_{EAB} = [N^t, L^t, A^t] \quad (8)$$

consisting of a set of spatial network nodes *N* at time *t,* a set of spatial network links *L* at time *t.* An adjacency matrix stores the network topology at time *t.* The following sections define the spatial network with respect to representing EAB forest insect infestation dispersal.

Spatial Network Nodes N

In ecological networks, specifically for terrestrial ecological systems, it is largely agreed that the appropriate scale for representation is the landscape scale, where nodes represent habitat patches that are suitable for the species of interest (Baguette et al. 2013). Therefore, in the spatial network SN_{EAB} for representing EAB dispersal, a node v_i in a set of nodes *N* represents a forest stand that contains ash tree species. Nodes in the SN_{EAB} contain non-spatial attributes including type of ash tree species existing in the forest stand. Each habitat patch node is embedded in geographic space and thus

contains the property of location. As a function of this property, the distance between habitat patch nodes can be calculated. Nodes in the SN_{EAB} also contain the properties of node weight *w*, representing the combined susceptibility of the forest stand to EAB long-distance dispersal as a function of forest stand attributes. A node v_i is added to the set of nodes *N* at time *t* when the degree *k* of the node is >=1.

Spatial Network Links

A link e_i in a set of links *L* represents the dispersal of one or more EAB from *node* v_i to *node* v_j. In the SN_{EAB}, links are not embedded in planar geographic space since EAB are airborne and will likely not disperse in a straight line from one habitat node to another. Therefore, a link does not represent the explicit path of dispersal between two nodes, but rather simply marks a dispersal event between two nodes. Regardless, the length of links is a representation of the distance between two habitat patch nodes. Links in the SN_{EAB} also contain the property of direction where each link records dispersal *from* and *to* a habitat patch node.

Spatial Network Cost Matrix

The cost matrix in the SN_{EAB} is composed of landscape features that range from minor impediments to finite barriers in dispersal (Baguette et al.2013; Verbeylen et al. 2003). Since EAB individuals are airborne, urban structures are minor impediments that do not have a major impact on the dispersal of the EAB population as they can serve as intermediate landing points between habitat patches. Water however, provides no relief in dispersal between habitat patches and thus lakes and rivers are considered barriers in the SN_{EAB} cost matrix.

Evaluating Habitat Patch Susceptibility

The GNA_{EAB} model represents *local* and *long-distance dispersal* as two separate processes. To represent long-distance dispersal processes, habitat patch node v_j becomes connected to habitat node v_i as a function of its overall *susceptibility* to *long-distance dispersal* associated with human movement of infested sapling, contaminated firewood, and insect hitchhiking on vehicles.

Therefore, each habitat patch node is evaluated for its *long-distance dispersal susceptibility* (*LDS*). Specifically, each forest stand or habitat patch node is assigned an

LDS value as a function of the nodes (1) distance from major roads and highways, (2) distance from campgrounds, (3) distance from urban areas, and (4) proximity to high human population density. The choice of the following attributes in the evaluation of the nodes susceptibility to long-distance dispersal is justified as follows:

Distance from major roads and highways (F1): The correlation between transportation networks and satellite populations is a direct result of the transportation of infested ash materials such as nursery stock, saw timber, and wood packaging material (Prasad et al. 2010). Therefore, forest stands near roads or highways are most susceptible.

Distance from campgrounds (F2): Firewood is transported through informal pathways over long distances. As a result, firewood that is infested with EAB is nearly impossible to identify, track, and regulate (Robertson and Andow 2009). Therefore, forest stands near campgrounds are most susceptible.

Distance from urban areas (F3): It has been demonstrated that ash trees in urban areas are significantly more susceptible to EAB infestation than in any other landscape (Kovacs et al. 2010; MacFarlane and Meyer, 2005). Therefore, forest stands near urban areas are most susceptible.

Human population density (F4): The correlation between population density and the establishment of satellite populations has been demonstrated (Muirhead et al. 2006). Therefore, the probability of forest stand infestation increases as human population density increases.

The proximity of each habitat patch node to the landscape features (*F1-F4*) is evaluated and used to calculate overall *LDS*. For example, habitat patch nodes near *all* landscape features are more susceptible to long-distance dispersal than habitat patch nodes near a campground, but far from all other features. The *LDS* value is represented on a scale from 0 to 1 where an *LDS* value of 0 indicates that the forest stand node is not susceptible to long-distance dispersal and an *LDS* value of 1 indicates that the forest stand node is highly susceptible to long-distance dispersal.

The *LDS* value for each habitat patch node is used as a representation of the habitat patch node's *v* weight $w_{v,}$ as follows:

$$w_v = LDS(F1, F2, F3, F4) \qquad (9)$$

Habitat patch nodes with a higher *LDS* value obtain a higher node weight, and thus have a greater likelihood of being connected to via long-distance dispersal.

Implementing GNA$_{EAB}$ Dynamics

Transition rules *R* and connection cost *C* implemented at each model time step are designed to simulate real-world interactions between system elements and ultimately determine the dynamics and evolution of the spatial network SN_{EAB}. The model begins at t_0, corresponding to the initial introduction of the EAB in North America in year 1997. The GNA_{EAB} model represents one year of EAB infestation as three consecutive time steps. This set of time steps represents the months of June, July, and August that the EAB are active in Michigan (Herms & McCullough, 2014). The model does not represent the months that the EAB are not active and thus not spreading.

Three transition rules (*R1, R2, and R3)* as well as one connection cost *C1* are implemented to simulate EAB local dispersal, long-distance dispersal, and forest stand death in the evolution of the spatial network SN_{EAB}. The influence of the transition rules and connection cost on the adjacency matrix and thus the evolution of the network is expressed as follows in formulation (10), derived from formulation (3) from the GNA theoretical framework:

$$A_{t+1} = R1, R2, R3, C1(A_t) \qquad (10)$$

where the adjacency matrix *A* for the SN_{EAB} at the next time step, time *t+1*, is a function of the transition rules *R1, R2, R3*, the connection cost *C1*, and the adjacency matrix *A* at time *t* in the model.

Transition rule *R1* and connection cost *C1* are applied to represent the EAB local dispersal process, transition rule *R2* is applied to represent EAB long-distance dispersal, and *R3* is applied to represent forest stand death. Figure 4.5 illustrates the logic of the transition rules *R* and the cost connection *C* for network evolution.

Local Dispersal Process

In local dispersal, mated female EAB disperse up to a maximum distance of 20 km within its lifetime (Taylor et al. 2010; Kovacs et al. 2011). As such, the likelihood that dispersal will take place between two forest stands at each time step is a function of the distance between them. The process of local dispersal is represented in the GNA_{EAB} where a habitat patch node v_j becomes connected to *k = k+1* habitat patch node v_i as a function of *distance* between them. The transition rule *R* for representing the local dispersal process is as follows:

***R1*:** At time *t,* if d_{ij} <= *20 km* and k_i*>0*, then a directional link *e* is established from node v_i to v_j at time *t+1* (Figure 4.5a and b). It should be noted that this link *e* can be established between node v_i and node v_j where node v_j is not connected k_j *= 0* (Figure 4.5a) or is already connected k_j*>=1* (Figure 4.5b) to the network.

For local dispersal, connection cost *C* between habitat patch node v_i and habitat node v_j is a function of physical barriers including lakes and rivers. The connection cost *C1* is as follows:

***C1*:** A time *t+1,* if the link *e* established in *R1* intersects a river or lake landscape element, link *e* between habitat patch node v_i and habitat patch node v_j is resisted and removed (Figure 4.5c).

Long-distance Dispersal Process

It is reported that the establishment of EAB satellite populations is a result of anthropogenic factors (Robertson and Andow 2009). These satellite populations are more likely to develop in forest stands that are near roads, campgrounds, urban areas, and high human population densities. As such, in the GNA_{EAB}, long-distance dispersal between habitat patch node v_i and habitat patch node v_j is a function of 1) the likelihood that dispersal will occur from node v_i to node v_j and 2) habitat patch node weight w_v, representing the overall *LDS* of node v_j, see formulation (9). The value of the likelihood of long-distance dispersal is 0.2% and is determined using sensitivity analysis. The transition rule for representing the long-distance dispersal process is as follows:

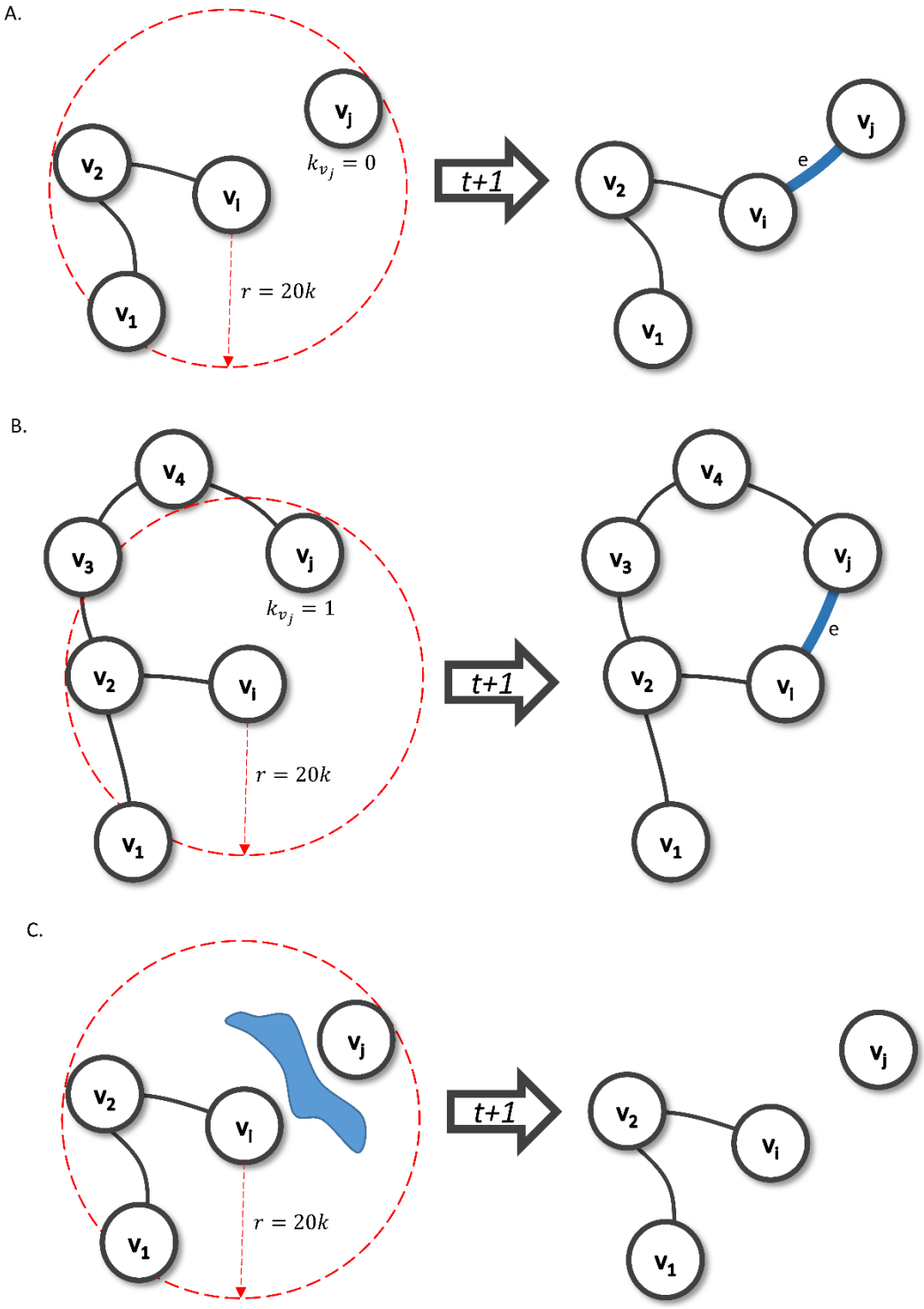

Figure 4.5. **Network evolution as a result of the application of the transition rule *R1* in the case where (a) node *j* has a degree *k=0*, (b) node *j* has a degree *k>=1*, and (c) the cost connection *C1* with barriers of lakes and rivers.**

R2: At time t+1, given the 0.2% likelihood threshold of long-distance dispersal being successful, a link *e* forms between habitat patch node v_i and the habitat patch node v_j with the highest weight w_v as a representation of *LDS*.

Death of Forest Stand Process

Until the last tree in the stand dies, EAB can continue to disperse and infest new forest stands through long-distance dispersal. Tree death typically occurs between 3 and 5 years. Forest stand death in the model is represented where habitat patch nodes and their links to other forest stands that have been infested for 5 years are removed from the network. Therefore, the transition rule for representing forest stand death is as follows:

R3: At time *t,* if the age of habitat patch node v_i is greater than 15 time-steps, habitat patch node v_i at time *t+1* and all links L connected to habitat patch node v_i are removed.

The model is initialized at t_0 representing the initial introduction of the EAB in North America in 1997 using the epicenter identified in Siegert et al. (2014)'s findings. At each time step, each of the transition rules *R* and connection cost *C* are applied.

GNA_{EAB} Model Output Analysis

The outputs of the GNA_{EAB} are a sequence of spatial networks SN_{EAB} representing EAB dispersal patterns at the regional scale over space and time. Although there is a random component in the model related to the likelihood that long-distance dispersal will occur at time step *t*, this generates very minor temporal variations across model runs. Specifically, if a forest stand with the highest *LDS* does not become infested at time step *t* as it was in a different model run, it will become infested within the next few time steps at *t+1, t+2* or *t+n.* Since the variation across time is small, an example SN_{EAB} is selected for the purpose of the presentation of the obtained results.

The SN_{EAB} generated by the GNA_{EAB} is useful because it can be directly measured using graph theory to quantitatively characterize dispersal patterns and link the spatial structure of the landscape to dispersal dynamics, presenting implications for ecological management. One of the primary objectives in network measurement is to determine the type of spatial network, specifically whether the resulting network is

random, scale-free, or small-world, each characterized by a unique networks structure. The network type can be determined by comparing the observed graph theory measures degree distribution *P(k)*, average clustering coefficient <*C*>, and average path length <*l*> (Table 4.1) to the expected measures if a network with the same number of nodes was random.

Table 4.1. Graph theory measures for network analysis.

Graph Theory Measure	Definition
Node degree *k*	A count of the number of connections node v_i has to other nodes
Degree distribution *P(k)*	A fraction of nodes in the network with degree k, calculated for the entire distribution of *k*. This is typically presented as a histogram.
Average clustering coefficient <*C*>	Clustering coefficient *C* measures the likelihood that nodes connected to node v_i are also connected to each other. When *C* = 1, all nodes connected to node v_i are also connected to each other and when *C* = *0,* none are connected to each other. Average clustering coefficient <C> measures the average *C* across all nodes in the network.
Average path length <*l*>	The average number of intermediate nodes and links in the shortest path between all pairs of nodes in the network

Therefore, graph theory is applied to quantify the spatial network SN_{EAB} simulated by the GNA_{EAB} model using the above measures. A random network model is developed, where links are randomly formed between the forest stand nodes at each time step, and the same graph theory measures are obtained from the simulated random spatial network SN_{RAND} and compared to the SN_{EAB}.

GNA_{EAB} Model Sensitivity Analysis and Calibration

Prior to the full implementation of the GNA_{EAB} model, sensitivity analysis and calibration have been performed. Specifically, sensitivity analysis is conducted to determine the effect that the changes in parameters have on GNA_{EAB} model outputs. Next, in the calibration stage, parameter values are selected based on their ability to achieve a *simulated rate of spread* that is closest to the *real-world rate of spread*. The real-world rate of spread is estimated by comparing the distance between the original 1997 epicenter and maximum distance of forest stands in each county found to be infested each year (Table 4.2). There are several years where the rate of spread is 0. In these cases, the counties between the satellite populations and the main invasion front become infested *before* the satellite populations spread to new and subsequently further

counties. As a result, the observed distance between the epicenter and the furthest forest stand does not increase.

Table 4.2. Observed real-world rate of spread from 1997 to 2011.

Year	Max Distance of Spread (km)	km/year
1997-2002	81.64	16.328
2003	232.822	151.182
2004	393.995	161.173
2005	487.978	93.983
2006	487.978	0
2007	487.978	0
2008	678.477	190.499
2009	678.477	0
2010	678.477	0
2011	763.03	84.553

Given the simplicity of the GNA modelling framework, there are few, but nonetheless important parameters tested using sensitivity analysis. Sensitivity analysis was conducted using the one-at-a-time (OAT) technique, where one parameter is adjusted within the parameter space and the other parameters remain fixed for the duration of a model run. This facilitates a better understanding of how changes in individual model parameters influence model behavior. Model parameters for which sensitivity analysis is conducted include Parameter 1: *likelihood threshold θ* used to determine the likelihood of the occurrence of long-distance dispersal, and Parameter 2: *level of importance of each forest stand attribute* used in the calculation of the overall *LDS*. Sensitivity of the model to the parameter values was evaluated using the simulated rate of EAB spread (km/year).

Parameter 1: Likelihood threshold θ. The likelihood threshold *θ* determines the likelihood of long-distance dispersal occurring from node v_i to node v_j at time *t.* At every time step, a random number generator assigns a random number between 0.00 and 100.00 to each infested forest stand node v_i where $k_i > 0$. If the generated random number is below the likelihood threshold *θ*, long-distance dispersal occurs from the forest stand node v_i to node v_j. The sensitivity of the long-distance dispersal process to the threshold values are tested using a parameter space from 0.09 to 1.1. This parameter space is selected because a likelihood threshold *θ* higher than 1.1 generates

far too many satellite populations and a likelihood threshold θ lower than 0.09 does not generate nearly enough satellite populations to match the real world rate of spread.

Model sensitivity to variations in likelihood threshold θ is presented in Figure 4.6a. Threshold values less than 0.2 produced a simulated rate of spread that was slower than observed in the real-world. In addition, the simulated maximum dispersal distance for 2011 was significantly less than the observed maximum dispersal distance for 2011. Threshold values greater than 0.2 produced a simulated rate of spread that was much faster than observed in the real-world. Furthermore, the simulated maximum dispersal distance of EAB spread for the region of Michigan was reached several years faster than in the real-world.

Parameter 2: Weighted combination of forest stand attributes. Susceptibility weights determine the influence of the proximity of the node to landscape features (F1, F2, F3, F4) in calculating its *LDS*. For example, in calculating *LDS*, a higher susceptibility weight placed on the attribute *proximity to campgrounds* will result in a significantly greater *LDS* value for nodes near campgrounds and thus a higher likelihood of becoming infested via long-distance dispersal. The GNA_{EAB} model is tested for sensitivity to variations in the weighted combination of attributes presented in Table 4.3. Specifically, each landscape feature is given a significantly greater weight (0.7) in each *LDS* calculation and the remaining weight is allocated to remaining landscape features (0.1).

Table 4.3. Variation of susceptibility weights.

Attribute	Susceptibility Weights				
	Equal	F1 Priority	F2 Priority	F3 Priority	F4 Priority
Proximity to Roads	0.25	0.7	0.1	0.1	0.1
Proximity to Campgrounds	0.25	0.1	0.7	0.1	0.1
Proximity to Urban Areas	0.25	0.1	0.1	0.7	0.1
Population Density	0.25	0.1	0.1	0.1	0.7

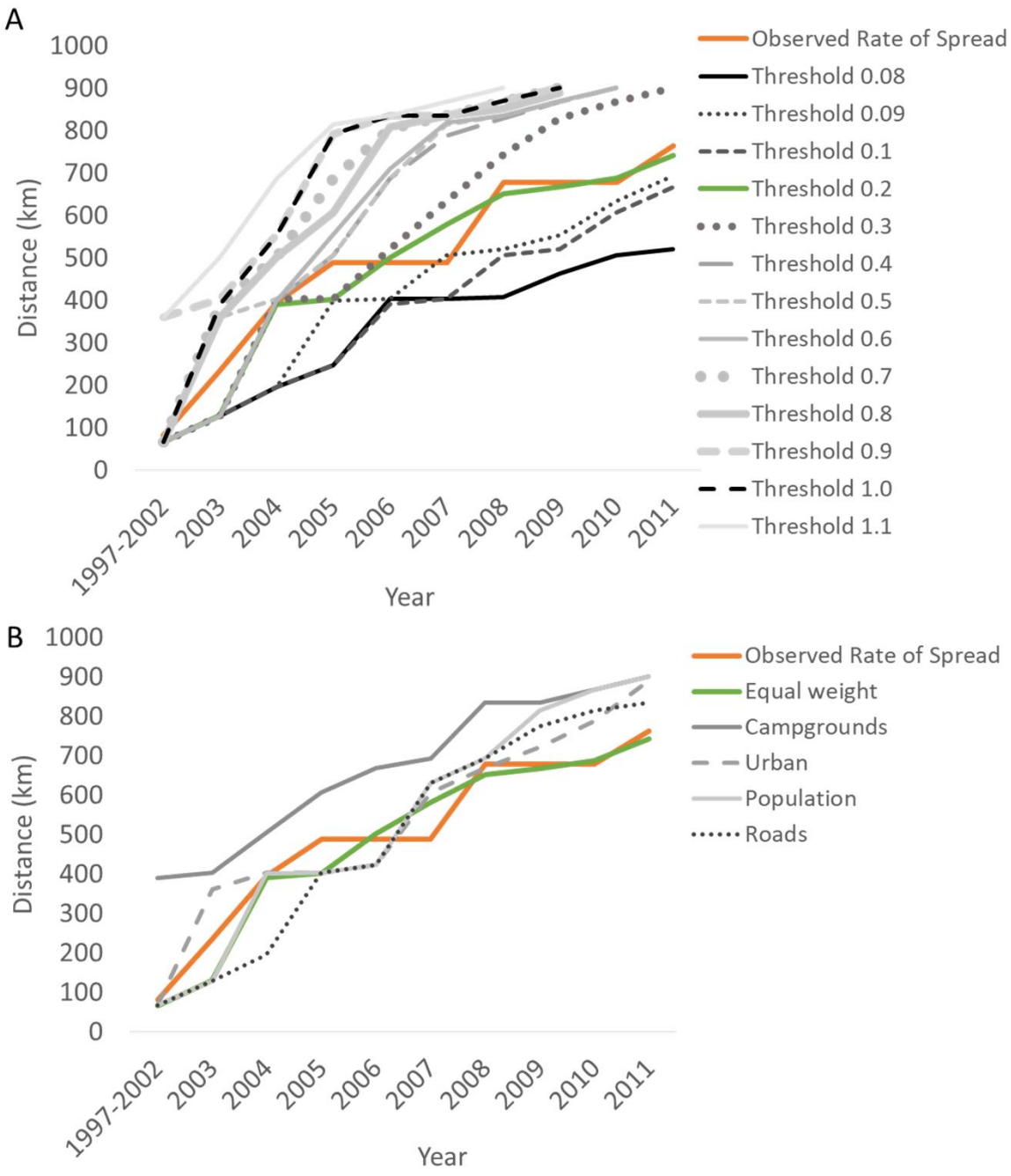

Figure 4.6. Sensitivity of GNA_{EAB} simulated rate of spread to (a) variations in likelihood threshold and (b) weighted combinations for calculating *LDS.* The observed real-world rate of spread (distance per year) is presented in orange and the simulated rate of spread using parameters that best match the observed rate of spread is presented in green.

GNA_{EAB} model sensitivity to variations in weighted combinations of forest stand node attributes used to calculate *LDS* is presented in Figure 4.6b. Although all combinations produced a relatively similar simulated rate of spread in comparison to the

rate of spread observed in the real-world, using an equal weight in the combination of all attributes in calculating the *LDS* of forest stands produced the most realistic rate of spread. Specifically, the simulated maximum dispersal distance in 2011 was more like the observed maximum dispersal distance, where all other weighted combinations simulated spread further than observed in the real-world.

Based on the outcome of the sensitivity analysis, the GNA_{EAB} model is calibrated by adjusting the two parameter values explored in sensitivity analysis, to best match the real-world rate of spread. Specifically, *Parameter 1: the likelihood threshold* is set to the value of 0.2 and *Parameter 2: the weight of attributes* is set to be equally weighted in the calculation of node *LDS*.

4.5. Results

Based on the outcomes of the sensitivity analysis and calibration, the GNA_{EAB} model parameters have been adjusted prior to generating simulation results. The obtained GNA_{EAB} simulation results are presented and the spatial networks SN_{EAB} have been analyzed using graph theory measures.

4.5.1. GNA_{EAB} Simulation Results

Figure 4.7 presents one full GNA_{EAB} model run representing the spatial network of the spread of EAB infestation across Michigan, USA from 1997 to 2011. Specifically, Figure 4.7 depicts the evolution of the generated spatial network SN_{EAB} structures from 1997-2003 (Figure 4.7a), 1997-2005 (Figure 4.7b), 1997-2009 (Figure 4.7c), and 1997-2011(Figure 4.7d). Forest stands that have died from EAB infestation by 2011 are also presented (Figure 4.7e).

Figure 4.8a presents the spatial network SN_{EAB} structure generated by the GNA_{EAB} without forest stand death for better visualization of the evolution of the network. The comparison between the GNA_{EAB} simulated spatio-temporal patterns of EAB dispersal represented as a dynamic network and the spatio-temporal patterns of EAB dispersal observed in the real-world are presented in Figures 4.8b and 4.8c. One of the most obvious spatio-temporal patterns that emerge from the GNA_{EAB} simulations is a stratified dispersal pattern, a pattern observed in real-world where the main EAB

infestation front and the satellite populations merge over time (Muirhead et al. 2006). Additionally, the GNA_{EAB} correctly simulates the establishment of several real-world satellite populations where EAB connect or “jump” to non-adjacent nodes; most notable are those infested nodes that emerge near Cheboygan and Manistee Counties in 2004, in Chippewa County in 2005, and in Houghton in 2008. Interestingly, the GNA_{EAB} model is also able to represent the lack or delay of infestation in Luce County and in the region of Marquette and Iron County as observed in the real-world. This can be explained by the low *LDS* and spatial distribution of host stands not conducive to both local and long-distance dispersal in these regions.

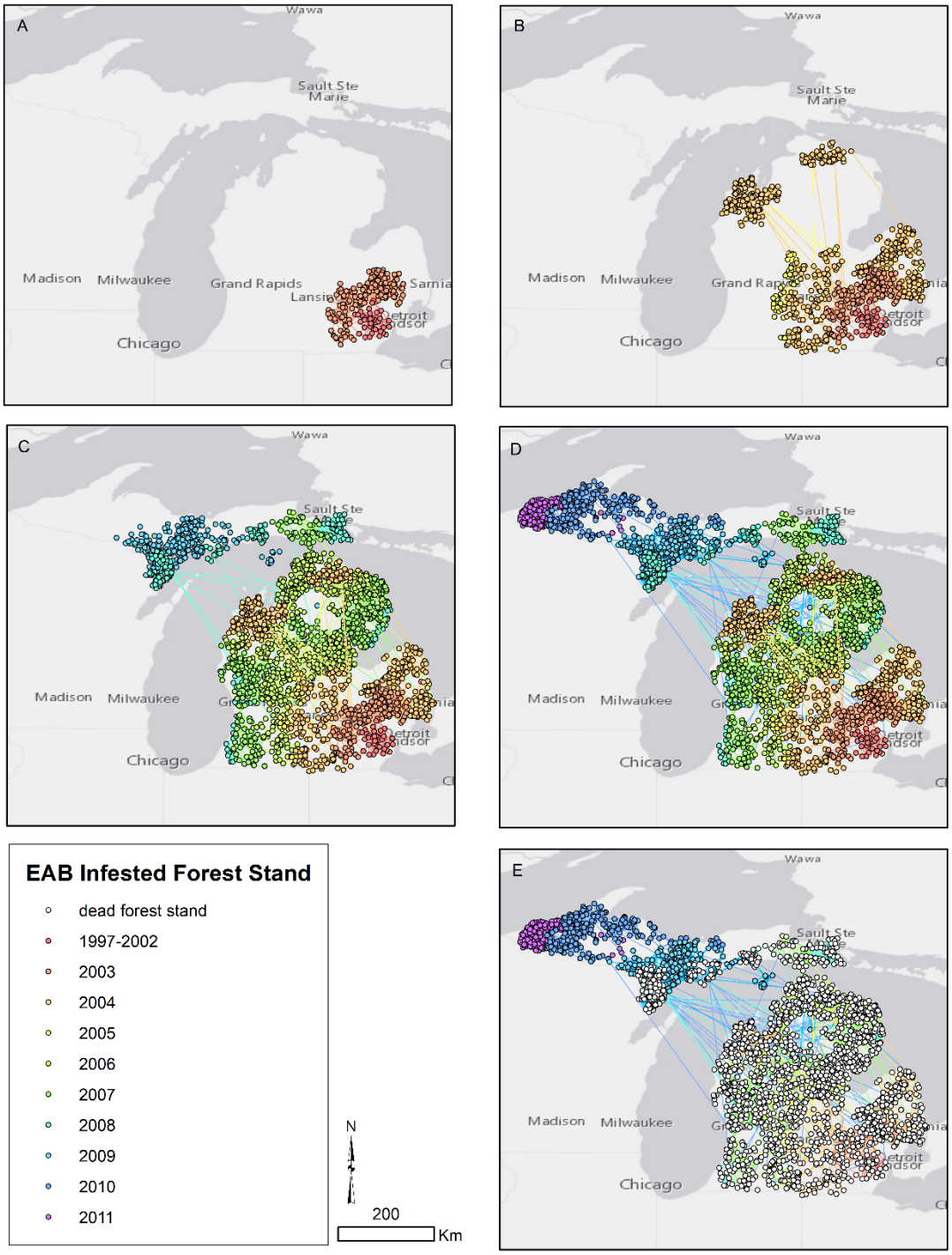

Figure 4.7. **The GNA_{EAB} model outcomes: the simulated spatial network SN_{EAB} of EAB dispersal as it evolves from year (a) 1997-2003, (b) 1997-2005, (c) 1997-2009 and (d) 1997-2011 without dead trees, and (e) the simulated forest stands with dead trees from 1997-2011.**

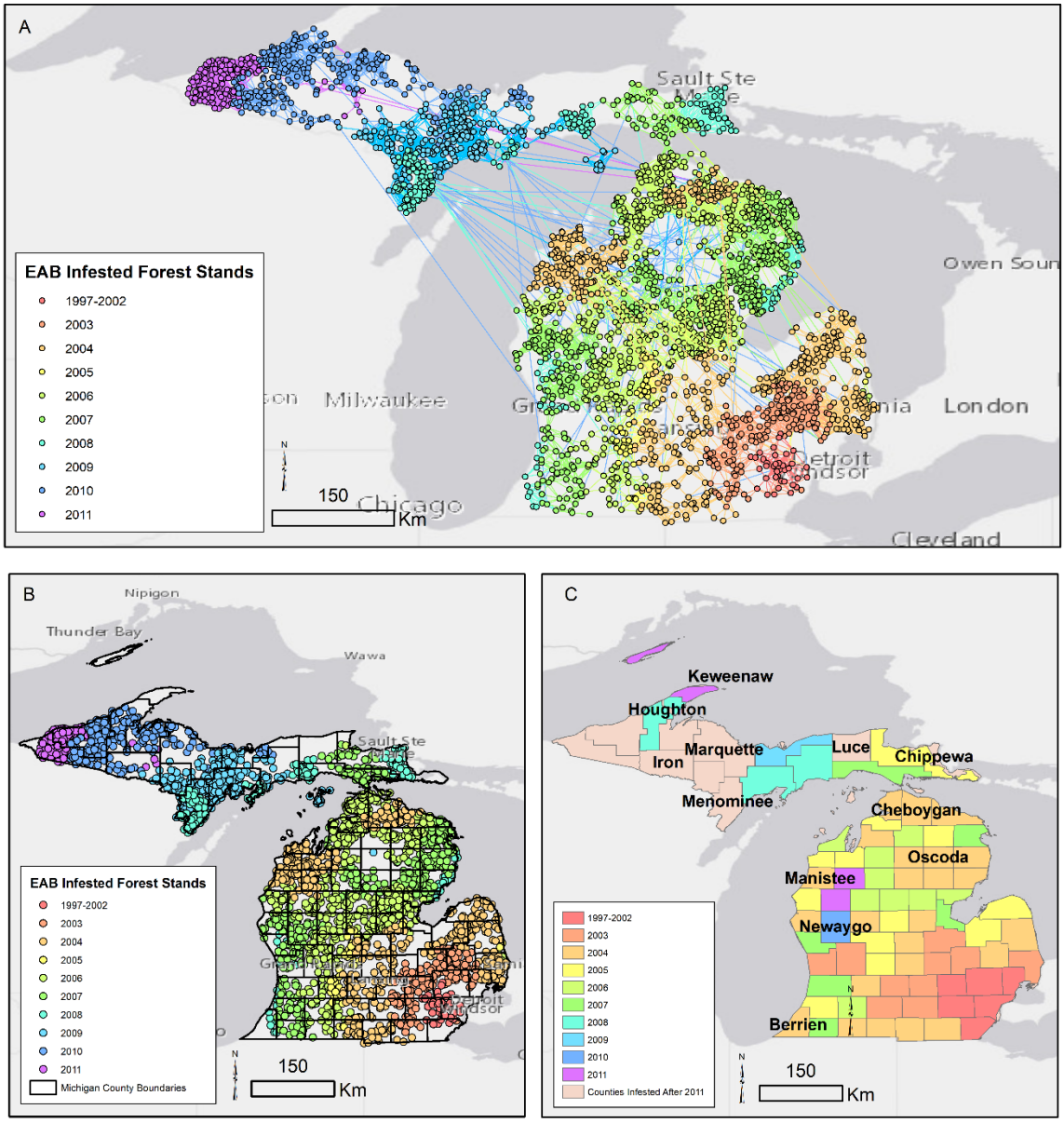

Figure 4.8. Comparison of the GNA_{EAB} simulated spatial networks of the (a) EAB dispersal SN_{EAB} from year 1997 to 2011 without forest stand death and (b) presented per each county, with (c) the observed real-world spatio-temporal patterns of EAB dispersal.

The GNA_{EAB} model simulates a satellite population in Menominee County in 2007, where in the real-world, infestation was not discovered until after 2011. The GNA_{EAB} model outputs indicate that this satellite population is established as result of a high *LDS* value generated by several campgrounds in the region. Furthermore, the GNA_{EAB} model was unable to capture the satellite populations in Manistee, Berrien, or

Oscada region in 2004. It is important to note that these inconsistencies are a function of the calculation of *LDS* and thus node weight. The model does not simulate infestation of Keweenaw County Isle Royale as a function of the LDS value. Lastly, the model outcomes indicate that the lag of infestation in Newaygo County was not captured, where in the real-world, EAB infestation is not recorded until 2010. Considering Newaygo County is bordered by two counties that were infested in 2004, the lag of infestation in this county may actually be a lag in discovery by county resources.

4.5.2. Analysis of GNA_{EAB} Model Outputs

One of the primary objectives is to characterize the structure of the generated SN_{EAB}. Networks are characterized by their structure based on measured *degree distribution P(k), average clustering coefficient <C>,* and *average path length <l>* as the discriminating features in determining network type (Table 4.4).

Table 4.4. Defining measures for each network type.

Network Type	Degree Distribution *P(k)*	Clustering Coefficient <*C*>	Average Path Length <*l*>
Random	Poisson	Low	Low
Small-World	Poisson	Higher than random	Low
Scale-Free	Power law	Low	Low

Therefore, graph theory measures including *degree distribution P(k), average clustering coefficient <C>,* and *average path length <l>* are obtained for every third model time step to characterize the SN_{EAB} structure as it evolves for each year.

At t_9 (2004), as nodes are increasingly added and re-wired to the spatial network, the SN_{EAB} degree distribution *P(k)* begins to form a normal curve, or a Poisson degree distribution. The Poisson distribution is maintained consistantly as the network evolves over time, indicating that the majority of nodes in the network have a similar degree of connectedness to other nodes (Figure 4.9). The simulated SN_{EAB} structure at its last time step t_{30} (2011) indicate that the average degree <*k*> across all nodes in the network is 26, meaning that the spatial distribution of forest stand nodes faciliates the dispersal from forest stand node v_i to an average of 26 other forest stand nodes. This suggests that the spatial distribution of forest stand nodes across the entire region is dense and well-connected, facilitating the spread of the EAB population.

The average clustering coefficient <*C*> of the SN_{EAB} increases consistantly each year as older nodes that are already part of the network connect to other older nodes (Figure 4.10a). Most connections form between local neighbouring nodes, a function of the transition rule *R1* that simulates local dispersal based on distance between nodes. This is common for many types of spatial networks as geographic space is a physical constraint where long-distance connections between two nodes are limited by energy and time. Specifically, it is more efficicient for the EAB to infest adjacent forest stands.

High clustering between local neighbouring nodes often results in a long average path length <*l*>, where several intermediate nodes exist between any two nodes in the network. However, even as the SN_{EAB} grows, the transition rule *R2* that simulates long-distance dispersal connects distant nodes and maintains a short average path length <*l*>(Figure 4.10b). The short average path length <*l*> is particuarly interesting as it indicates that EAB can move from any forest stand in the network to any other forest stand in the network in an average of 15 steps (which is equivalent to 5 years). This indicates that the landscape is well connected and that EAB can propagate quickly across the landscape.

Graph measures obtained for the SN_{EAB} at its last time step t_{30} (2011) are summarized in Table 4.5 and are compared to corresponding measures obtained from the random model SN_{RAND} to characterize SN_{EAB} network type. Both spatial networks, SN_{EAB} and SN_{RAND} generate a degree distribution that forms a Poisson curve. The average clustering coefficient $<C>_{EAB}$ obtained from the SN_{EAB} is much higher than that of the average clustering coefficient $<C>_{RAND}$ obtained from the random spatial network SN_{RAND}. Furthermore, the average path length $<l>_{EAB}$ and $<l>_{RAND}$ are short with an average length of 15 for the SN_{EAB} and 10 for the SN_{RAND}.

The generated spatial network SN_{EAB} is a small-world if $<l>_{EAB} \geq <l>_{RAND}$ and if $<C>_{EAB} \gg <C>_{RAND}$. Humphries and Gurney (2008) provide a set of equations that use $<l>_{EAB}$, $<l>_{RAND}$, $<C>_{EAB}$, and $<C>_{RAND}$ to mathematically characterize the small-world-ness of a network using the small-world coefficient S^{Δ} where if S^{Δ}>1, the network can be quantitatively characterized as a small world network. The theoretical maximum small-worldness value is 560.01, calculated using $S^{\Delta} = 0.181n$ where *n* is the number of nodes in the network. The obtained value of the small-world coefficient S^{Δ} of 131.29 indicate that the simulated SN_{EAB} is a small world network.

Table 4.5. Resulting graph theory measures from the spatial network SN_{EAB} generated by the GNA_{EAB} and the random spatial network SN_{RAND}.

Measure	SN_{EAB}	SN_{RAND}
Degree distribution *P(k)*	Poisson, positive skew	Poisson
Average clustering coefficient <*C*>	0.67	0.0034
Average path length <*l*>	15	10
Small world coefficient S^{Δ}	131.2953	N/A

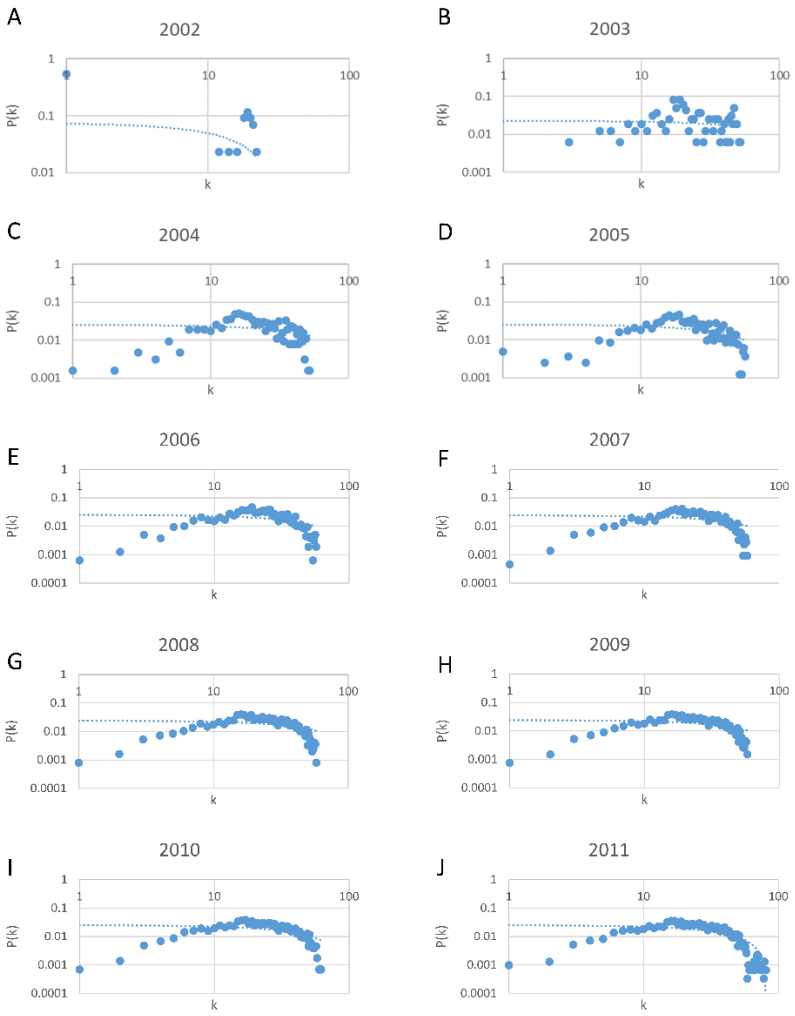

Figure 4.9. Degree distribution *P(k)* as measured for the SN_{EAB} structure for each year.

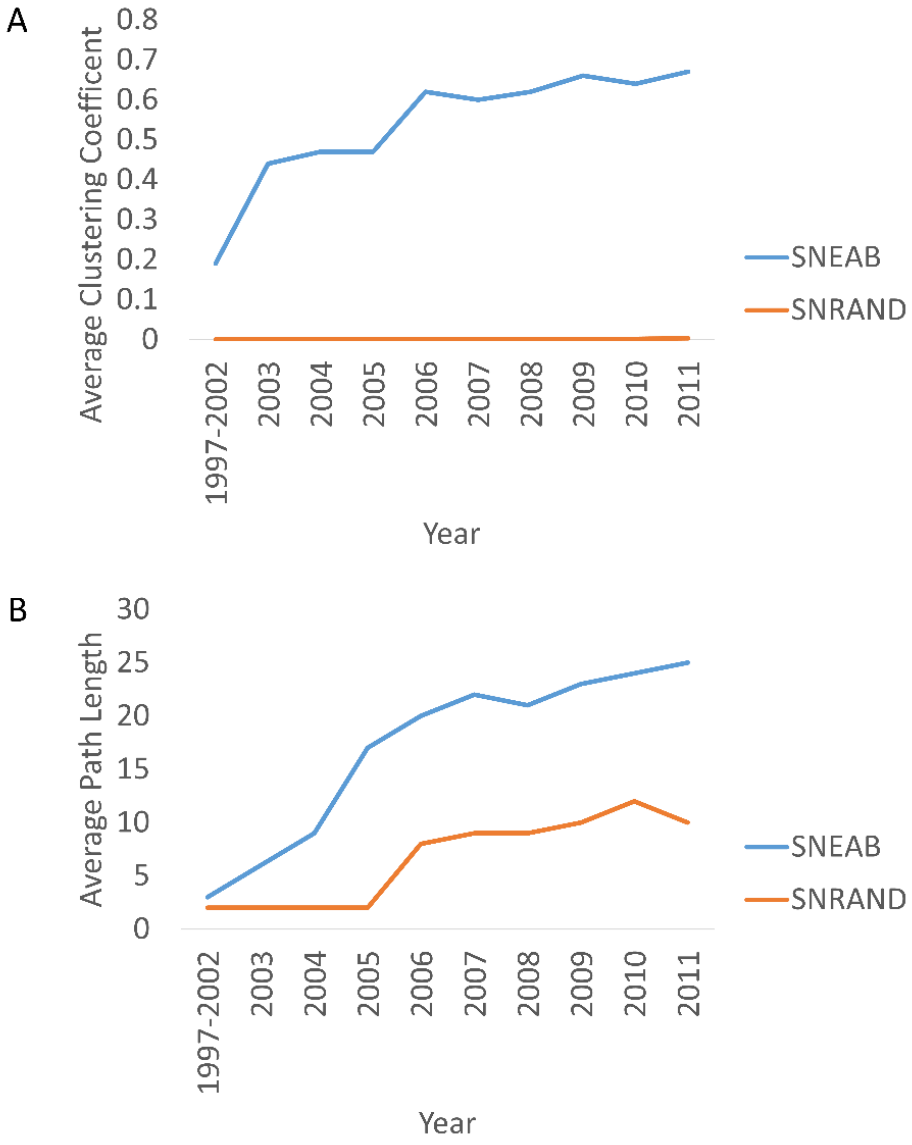

Figure 4.10. The change in (a) the average clustering coefficient <*C*> and (b) average path length <*l*> from 1997-2011.

In ecology, network robustness can give insight into the reaction to disturbances or attacks via external forces. Particularly, the results characterizing the SN_{EAB} network type can be used to understand whether eradication measures such as the removal or vaccination of forest stands will impact the connectivity of the landscape and effectively slow or stop EAB propagation. A small world network structure indicates that the spatial distribution of forest stand nodes across the landscape is highly robust to unorganized disruptions or node removals. As indicated by the high average clustering coefficient

<*C*>, the majority of forest stand nodes in the landscape are in close proximity to several other forest stand nodes, facilitating sucessful local dispersal and continued propagation, even in the event that several nodes become unsuitable hosts. This means that local eradication of EAB in a few cities or towns would have little impact on slowing the propagation of EAB across the state and a national, strategized approach would be more effective.

4.6. Discussion and Conclusion

This study introduces the novel modelling framework of Geographic Network Automata (GNA) and demonstrates that the proposed approach is a useful methodology for the representation and analysis of complex ecological networks that evolve across space and time. The spatial network approach in ecology dates to Urban and Kiett (2001), however the few studies that adopt this approach focus mainly on static landscape connectivity networks and how these static networks are impacted by node removal via external forces. The GNA approach aims to fill this methodological gap by developing a modelling framework capable of simulating spatial ecological networks where network evolution is a function of the complex systems dynamics itself. The approach uses the iterative application of transition rules to represent the addition, loss, and rewiring of nodes over space and time as a function of the phenomena's dynamics. The network can then be characterized and quantified to better understand dispersal patterns.

In some traditional network modelling approaches used for representing spreading processes such as dispersal patterns, linkages between nodes are only considered based on their adjacency. For example, cellular automata, the complex systems modelling approach under the umbrella of geographic automata systems, is well suited to modelling spreading processes, however is also typically limited to localized spreading and thus is unable to represent long-distance dispersal processes in ecological systems (Anderson and Dragicevic 2016). The flexibility of the transition rules in the GNA approach facilitates the representation of long-distance connections or "hopping" over adjacent suitable nodes, useful for representing ecological species that exhibit this particular dispersal pattern.

The proposed GNA_{EAB} simulates patterns of EAB dispersal at a relatively coarse resolution. Theoretically, the GNA framework could be applied to represent EAB dispersal at much finer spatial and temporal scales where nodes represent individual trees, links represent movement of individual EAB, and each time step represents one day. In this case, it may be necessary to include a less abstract representation of EAB, as the life cycle of the individual EAB and population dynamics may play a role in the spatio-temporal evolution of the network at this scale. This could entail the development of a spatial network agent-based model.

The GNA_{EAB} simulated patterns should reflect patterns observed in the real-world. However, it can be very difficult to collect or obtain datasets that would facilitate quantitative testing of GNA_{EAB} simulated spatial networks. Therefore, this study looks at how well the model reproduces patterns observed in the real-world to increase confidence in GNA_{EAB} parameters, processes, and generated dynamic spatial networks. As is common in modelling ecological systems, the application of the GNA methodology to other types of ecological phenomena may be subject to the same data limitations for model testing. As a relatively unexplored modelling approach, there remain limited efforts in proposing methodologies for the testing of spatial network models. This is something that would require consideration in future exploration of representing ecological phenomena as dynamic spatial networks.

The GNA framework can be further operationalized in other case studies where the dispersal network of an ecological species is not static and thus expands, shrinks, or changes structure as a function of dynamics of the species themselves (eg. invasion, competition, adaptation, evolution, seasonal migration). For example, in the case of representing changes in dispersal patterns as a function of species competition, population dynamics of the species captured by node weight and flow between nodes may play an important role in network structure and dynamics. However, the proposed GNA framework is highly general and flexible as to facilitate the representation and simulation of many complex geospatial dynamic systems that evolve over time as spatial networks. This framework can be suitable for representing many other complex spatio-temporal phenomena spreading as networks of epidemics, information, computer viruses, or cities to name a few. This proposed approach is also relevant for GISc, specifically as an extension of the framework of geographic automata systems.

4.7. References

Anderson, T., & Dragicevic, S. (2016). A geosimulation approach for data scarce environments: modeling dynamics of forest insect infestation across different landscapes. *ISPRS International Journal of Geo-Information,* 5(2), 9.

All Roads [Shapefile]. Michigan, US: State of Michigan Admin, March 10, 2018. Available: http://gis-michigan.opendata.arcgis.com/datasets/all-roads-v17 and (accessed May 22, 2018).

Augusiak, J., Van den Brink, P. J., and Grimm, V. (2014). Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. *Ecological Modelling,* 280, 117-128.

Baguette, M., Blanchet, S., Legrand, D., Stevens, V. M., and Turlure, C. (2013). Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews,* 88(2), 310–326.

Barabási, A. L. (2016). Network Science. Cambridge, UK: Cambridge University Press.

Barthélemy, M. (2011). Spatial networks. *Physics Reports,* 499(1-3), 1-101.

Bunn, A. G., Urban, D. L., and Keitt, T. H. (2000). Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management,* 59(4), 265–278.

Campbell Grant, E. H., Lowe, W. H., and Fagan, W. F. (2007). Living in the branches: Population dynamics and ecological processes in dendritic networks. *Ecology Letters,* 10(2), 165–175.

Census 2010 Tracts- Population, Income [Shapefile]. Michigan, US: Michigan DNR, April 19, 2018. Available: http://gis-michigan.opendata.arcgis.com/datasets/midnr::census-2010-tracts-population-income and (accessed May 22, 2018).

Cohen, J. E. (1978). Food Webs and Niche Space. New Jersey, NJ: Princeton University Press.

Fortuna, M. A., and Bascompte, J. (2007). The network approach in ecology. In F. Valladares, A. Camacho, A. Elosegi, C. Gracia, M. Estrada, J. C. Senar, and J. M. Gili (Eds.), Unity in diversity: Ecological reflections as a tribute to Margalef (pp. 371–392). Bilbao: Fundación BBVA.

Fortuna, M. A., Gomez-Rodriguez, C., and Bascompte, J. (2006). Spatial network structure and amphibian persistence in stochastic environments. *Proceedings of the Royal Society B: Biological Sciences,* 273(1592), 1429–1434.

Hall, S. J., and Raffaelli, D. G. (1993). Food webs: Theory and reality. *Advances in Ecological Research,* 24, 187–239.

Herms, D. A., & McCullough, D. G. (2014). Emerald ash borer invasion of North America: history, biology, ecology, impacts, and management. *Annual Review of Entomology*, *59*, 13-30.

Humphries, M. D., and Gurney, K. (2008). Network “small-world-ness”: A quantitative method for determining canonical network equivalence. *PLoS ONE,* 3(4).

Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., Woodward, G. (2009). Ecological networks - Beyond food webs. *Journal of Animal Ecology,* 78(1), 253–269.

Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *The American Naturalist,* 129(5), 657–677.

Kovacs, K. F.; Haight, R. G.; McCullough, D. G.; Mercader, R. J.; Siegert, N. W.; Liebhold, A. M. Cost of potential emerald ash borer damage in US communities, 2009–2019. *Ecological Economics,* 69, 569-578.

Lake Polygons [Shapefile]. Michigan, US: State of Michigan Admin, March 30, 2018. Available: http://gis-michigan.opendata.arcgis.com/datasets/lake-polygons and (accessed May 22, 2018).

Lewis, T. G. (2011). *Network Science: Theory and Applications*. Hoboken, NJ: John Wiley & Sons.

MacFarlane, D. W., & Meyer, S. P. (2005). Characteristics and distribution of potential ash tree hosts for emerald ash borer. *Forest Ecology and Management,* 213(1-3), 15-24.

Michigan State Park Campgrounds [Shapefile]. Michigan, US: Michigan DNR, April 19, 2018. Available: http://gis-michigan.opendata.arcgis.com/datasets/midnr::michigan-state-park-campgrounds and (accessed May 22, 2018).

Minor, E. S., and Urban, D. L. (2007). Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecological Applications,* 17(6), 1771–1782.

Morris, R. J., Lewis, O. T., & Godfray, H. C. J. (2004). Experimental evidence for apparent competition in a tropical forest food web. *Nature,* 428(6980), 310–313.

Muirhead, J. R., Leung, B., Overdijk, C., Kelly, D. W., Nandakumar, K., Marchant, K. R., and MacIsaac, H. J. (2006). Modelling local and long-distance dispersal of invasive emerald ash borer Agrilus planipennis (Coleoptera) in North America. *Diversity and Distributions,* 12(1), 71–79.

Muller, C. B., Adriaanse, I. C. T., Belshaw, R., and Godfray, H. C. J. (1999). The structure of an aphid–parasitoid community. *British Ecological Society,* 68(1), 346–370.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review,* 45(2), 167-256.

Prasad, A. M., Iverson, L. R., Peters, M. P., Bossenbroek, J. M., Matthews, S. N., Sydnor, T. D., & Schwartz, M. W. (2010). Modeling the invasive emerald ash borer risk of spread using a spatially explicit cellular model. *Landscape Ecology*, 25(3), 353-369.

Repast Simphony. 2016. Version 2.4. [Computer Software]. Chicago, IL: University of Chicago.

Robertson and Andow. Human-mediated dispersal of emerald ash borer: significance of the firewoodpathway.http://www.entomology.umn.edu/prod/groups/cfans/@pub/@cfans/@ento/documents/asset/cfans_asset_139871.pdf (accessed May 04, 2018).

Sayama, H., and Laramee, C. (2009). Generative network automata: A generalized framework for modeling adaptive network dynamics using graph rewritings. *Adaptive Networks,* 311–332.

Siegert, N. W., Mccullough, D. G., Liebhold, A. M., and Telewski, F. W. (2014). Dendrochronological reconstruction of the epicentre and early spread of emerald ash borer in North America. *Diversity and Distributions,* 20, 847–858.

Smith, D. M., Onnela, J. P., Lee, C. F., Fricker, M. D., & Johnson, N. F. (2011). Network automata: Coupling structure and function in dynamic networks. *Advances in Complex Systems*, *14*(03), 317-339.

Stang, M., Klinkhamer, P. G. L., and Meijden, E. Van Der. (2006). Size constraints and flower abundance determine the number of interactions in a plant–flower visitor web. *Oikos,* 112(1), 111–121.

Straw, N. A., Williams, D. T., Kulinich, O., and Gninenko, Y. I. (2013). Distribution, impact and rate of spread of emerald ash borer Agrilus planipennis (Coleoptera: Buprestidae) in the Moscow region of Russia. *Forestry,* 86(5), 515–522.

Taylor, R. A. J., Bauer, L. S., Poland, T. M., and Windell, K. N. (2010). Flight performance of agrilus planipennis (Coleoptera: Buprestidae) on a flight mill and in free flight. *Journal of Insect Behavior,* 23(2), 128–148.

Torrens, P. M., and Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science,* 19(4), 385-412.

Urban, D., and Keitt, T. (2001). Landscape connectivity: a graph-theoretic perspective. *Ecology,* 82(5), 1205–1218.

Verbeylen, G., De Bruyn, L., Adriaensen, F., and Matthysen, E. (2003). Does matrix resistance influence Red squirrel (Sciurus vulgaris L. 1758) distribution in an urban landscape? *Landscape Ecology,* 18(8), 791–805.

Chapter 5.

Network-Agent Based Model for Simulating the Dynamic Spatial Network Structure of Complex Ecological Systems[3]

5.1. Abstract

Non-spatial ecological networks provide insight into the organization and interaction between biological entities. More recently, biological dispersal is modelled using *spatial networks,* static sets of georeferenced habitat patches that connect based on a species' maximum dispersal distance. However, dispersal is complex, where spatial patterns at the *landscape scale* emerge from interactions between ecological entities and landscape features at much finer individual scales. Agent-based modelling (ABM) is a computational representation of complex systems capable of capturing this complexity. Therefore, this study develops a network-ABM (N-ABM) that combines network and complex systems theory to simulate complex evolving spatial networks. The developed N-ABM approach is implemented on the case study of the emerald ash borer (EAB) bark beetle using geospatial datasets in Ontario, Canada. The N-ABM generates dynamic spatial network structures that emerge from interactions between the EAB and tree agents at the individual scale. The resulting networks are analyzed using graph theory measures. Analysis of the results indicates a relationship between preferential attachment in insect host selection and the emergent scale-free network structure. The N-ABM approach can be used to represent dynamic ecological networks and provides insight into how network structure emerges from EAB dispersal dynamics, useful for forest management.

5.2. Introduction

Ecologists are largely concerned with the scientific analysis of interactions among species and their environment (Agrawal et al., 2007). To study these interactions,

[3] A version of this chapter is published: Anderson, T. & Dragicevic, S. (2018). Network-agent based model for simulating the dynamic spatial network structure of complex ecological systems. Ecological Modelling. 389, 19-32.

ecological systems are often presented as networks, where the network's nodes represent biological entities and the network's links represent some form of ecological interaction. Sets of ecological interactions are traditionally represented as non-spatial networks and are most commonly grouped into three ecological network types including food webs (Cohen, 1978; Hall & Raffaelli, 1993), host-parasitoid webs (Muller et al., 1999; Morris et al., 2004), and mutualistic webs (Jordano et al., 2003; Stang et al., 2006). Graph theory, a mathematical characterization of networks, can then be used as a tool to characterize ecological network topology, cross-compare between network structures to find common ground and uniqueness, and understand how network structures affect network dynamics and vice versa (Ings et al., 2009).

Because ecological interactions take place in geographic space and time, it is argued that the framework of ecological networks must also account for these contexts (Fortuna & Bascompte, 2007). *Landscape connectivity graphs* were one of the first efforts to represent ecological phenomena as networks in an explicitly geospatial context (Urban & Kiett, 2001). Using orthophotos or classified maps, habitat patches are abstracted as nodes that are embedded in geographic space. The formation of links between nodes signifies the potential dispersal from one habitat patch to another. Since long dispersal distances come at a greater cost (i.e. energy, time), the connectivity between two nodes is typically a function of proximity or adjacency, meaning nodes that are closer to one another have a higher probability of connecting. In these network representations, there are minimal data requirements, needing only the geographic location of habitat patches and maximum dispersal distance of the species of interest (Minor & Urban, 2007). The structure of the landscape connectivity graphs can inform species dispersal patterns, identify keystone patches that are critical to landscape connectivity, and furthermore, assess how dispersal patterns would change in response to disruptions in the network. For example, Fortuna et al. (2006) develop a network of ponds to identify the spatial structure of amphibian dispersal and the species persistence in drought. In another study, Bunn et al. (2000) compare landscape connectivity between two species that share the same habitat and find that the landscape is connected for one and disconnected for the other, presenting implications for conservation biology.

Despite the potential, network approaches that leverage graph theory for the study of ecological systems in a geospatial context mostly represent ecological systems at the *landscape species-scale* as a fixed network structure on which dispersal takes

place (Fortuna et al., 2006). Although useful, spatial ecological networks at this scale are limited in their ability to account for the complexity in emergent dispersal patterns, which often develop from local spatio-temporal dynamics between individuals and their environment. With a few exceptions, spatial ecological networks are rarely scaled down to the *individual species-scale* and as such, it is challenging to make this link between network structure and network dynamics (Dupont et al., 2011). Additionally, data capturing spatio-temporal dynamics at the individual-scale, between large sets of heterogeneous individuals, and across large spatial extents is not typically available for many types of ecological phenomena. Furthermore, field studies to acquire this data is expensive and time-consuming (Minor and Urban, 2007). As such, the development of an experimental setting to better understand the connections between network structure and network dynamics is needed (Stouffer et al., 2010).

Agent-based models (ABM) are computational representations of complex systems that explicitly represent interactions between individual entities or "agents" from which system-level behavior emerges, thus simulating ecological phenomena at a resolution that facilitates the construction of fine-scale networks (Grimm & Railsback, 2013). Developed ABMs are virtual laboratories, permitting exploration of the simulated phenomena and its associated network structure as it responds to local dynamics. ABMs have been developed to represent complex ecological systems over space and time for many ecological phenomena such as fish (Letcher et al., 1996), birds (Travis & Dytham, 1998), caribou (Semeniuk et al., 2012), and forest insect infestation such as the mountain pine beetle (Perez & Dragićević, 2010; 2011, Bone & Altaweel, 2014) and the emerald ash borer (Anderson & Dragićević, 2015; 2016), to name a few. Networks have been integrated with agent-based models (Berryman & Angus, 2010; Kirer & Çirpici, 2016); however, these studies typically focus on the topological network structure of systems rather than spatial network structures and thus do not incorporate geospatial datasets.

Therefore, this study proposes the integration of complex systems theory with network theory and geospatial datasets to represent complex spatio-temporal ecological systems using dynamic spatial networks. The main objective of this study is to develop and implement an integrated geospatial modeling approach, a network-ABM (N-ABM), to represent and analyze the dynamic spatial structure of ecological networks. Specifically, the agents' behavior and interactions form and modify the network over space and time.

Graph theory measures are used to measure and characterize the simulated spatial network structures to better understand how network structure evolves over space and time from interactions between biological individuals. The developed N-ABM approach is applied using the case study of a real forest insect infestation, the emerald ash borer (EAB) in Southern Ontario, Canada.

5.3. Networks and their Representation through Graph Theory

This section provides important network and graph theory definitions from which dynamic spatial network representations and measurement tools are derived. Foundations of graph theory are reviewed comprehensively (Barabási et al., 2016; Newman, 2003; Lewis, 2011) and for spatial graphs specifically (Barthélemy, 2011). A graph *G* is a mathematical representation of an observed spatial network *SN* reduced to a set of nodes *N* that are connected by directed or undirected links *L*. *Nodes* represent entities that make up the observed network and *links* represent the interactions between them.

Many empirically-observed spatial networks *SN* have properties corresponding to random, small-world, or scale-free graph types. Each graph type is unique in the way that it is formed and its resulting properties. Some common properties are defined in Table 5.1. The properties defined in Table 5.1 can be calculated as a local measure for each individual node or averaged across the network as a global measure.

Table 5.1. Definitions of important graph theory measures.

Graph Theory Measure	Definition
Node degree *k*	A local network measure that counts of the number of connections node *i* has to other nodes
Average node degree <*k*>	A global network measure describing the average node degree across all nodes in the network
Degree distribution *P(k)*	A fraction of nodes in the network with degree *k*
Clustering coefficient *C*	A local network measure that calculates the likelihood that nodes connected to node *i* are also connected to each other
Average clustering coefficient <*C*>	A global network measures that summarizes the average clustering coefficient across all nodes in the network
Path length l_s	A local network measure calculating the shortest path consisting of a number of nodes or links connecting two pairs of nodes in the network

Graph Theory Measure	Definition
Average path length $<l_s>$	A global network measure calculating the average number of intermediate nodes or links in the shortest path between all pairs of nodes in the network

Random graphs are modelled by linking nodes randomly based on probability *p* (Erdos & Renyi, 1959; 1960). As a result, one property of a random graph is a well-defined average node degree <*k*>, producing a Poisson degree distribution *P(k)*. The random connections that form between the set of nodes generate a low average clustering coefficient <*C*>, meaning that on average, it is unlikely that nodes connected to node *i* are also connected to each other. This lack of clustering produces a short average path length <*l*> (Boccaletti et al., 2006). Networks observed in the real world rarely exhibit properties of random graphs. However, random graphs provide a baseline for which to compare properties observed in real networks.

Some empirically-observed networks such as social networks have been found to have a much higher average clustering coefficient <*C*> and a similar average path length <*l*> than their random counterparts with the same number of nodes and the same average degree <*k*>. These networks exhibit properties of small-world graphs (Watts & Strogatz, 1998). Small world graphs are modelled as a lattice, where each node is connected to the exact same number of adjacent nodes, and a few nodes are rewired to a randomly chosen node. The establishment of just a few random connections dramatically reduces the average path length between any two nodes in the network, making the movement of material such as information, individuals, and power between nodes highly efficient.

A degree distribution unlike that of a random or small-world graph has been observed in several real networks such as the World Wide Web (Albert et al., 1999), characterized by a small fraction of nodes that have a very large number of links and a large fraction of nodes that have only a few. This degree distribution forms a power law, expressed as:

$$p(k) \sim k^{-\alpha} \qquad (1)$$

where the probability *p* of observing a node with *k* connections is the number of connections to some negative exponent called a degree exponent α. These observed

network structures exhibit the properties of scale-free graphs (Barabasi & Albert, 1999). A scale-free graph is modelled through the process of preferential attachment of new node v_j to node v_i based on node v_i's degree *k*. Specifically, new nodes prefer to link to existing nodes that have a higher degree, and thus "the rich get richer". In addition to a power law degree distribution, empirically-observed networks exhibiting properties of scale-free networks are characterized as having a shorter average path length and a similar average clustering coefficient in comparison to its random network counterpart.

In many cases, real networks can be modelled using a topological or spatial structure. The spatial structure of networks provides valuable information explaining network structure and behaviour. As geospatial data availability increases, geospatial network representations and analysis become more feasible. In a geospatial network, nodes are embedded in geographic space, defined explicitly using geographic coordinates. Geospatial networks are unique from their non-spatial, topological counterparts because any node's degree *k* is limited by physical space and thus distant connections are costly in terms of energy, money, and time. The structure and dynamics of real spatial networks have been characterized for transportation and infrastructure (Guimera & Amaral, 2004; Jiang, 2007; Watts & Strogatz, 1998) and less commonly for social (Stoneham 1977; Andris, 2016) and ecological (Fortuna et al., 2006, Pereira et al., 2011) phenomena. Graph theory-based network approaches have been proposed as useful tools in landscape ecology (Ferrari et al., 2007, Minor and Urban, 2008, Urban and Keitt, 2001), however few studies actually operationalize these approaches on real case studies using geospatial data (Andersson and Bodin, 2009, Fortuna et al., 2006, Pascual-Hortal and Saura, 2008, Zetterberg et al., 2010). Furthermore, graph theoretic approaches in landscape ecology are typically limited to static networks at the landscape scale and thus are unable to explore the spatio-temporal complexity inherent to ecological processes. Therefore, this study seeks to develop an integrated modelling approach that can facilitate the exploration of the complex dynamic spatial structure and behaviour of real ecological networks such as insect infestation. Specifically, network theory and agent-based modelling (ABM) are integrated to develop an N-ABM.

5.4. Methods

The developed N-ABM simulates a dynamic spatial network structure that emerges over space and time from interactions between agents at the individual scale (Figure 5.1). Graph theory is used to characterize and measure the obtained simulated spatio-temporal network patterns. The proposed approach is implemented and tested on the case study of the emerald ash borer (EAB) bark beetle in the Town of Oakville, Canada (Figure 5.2), first discovered in this area in 2008. The N-ABM approach is developed using the Java object-oriented programming language and Repast Simphony 2.5. (2017), a free and open source Recursive Porous Agent Simulation Toolkit (Repast). Repast Simphony is used for modelling complex adaptive systems through the development of ABMs and has a large and growing community developing a wide range of applications for social, evolutionary, industrial, and ecological simulations (North et al., 2013).

The following sections first provide an overview of the case study, including the EAB's biological background, followed by a detailed description of the development of the N-ABM and the graph measures used to characterize and analyze the N-ABM simulation outputs.

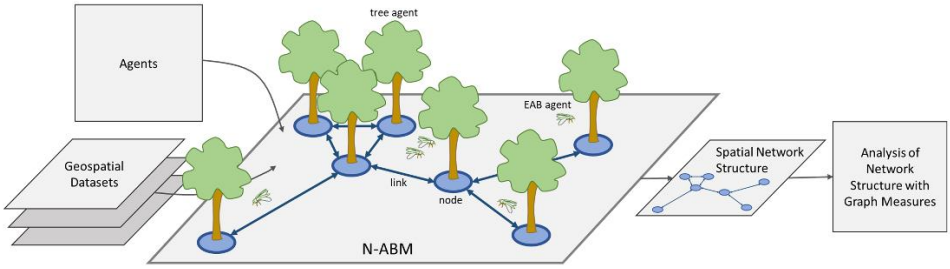

Figure 5.1. Network-agent based model (N-ABM) approach applied for emerald ash borer infestation case study.

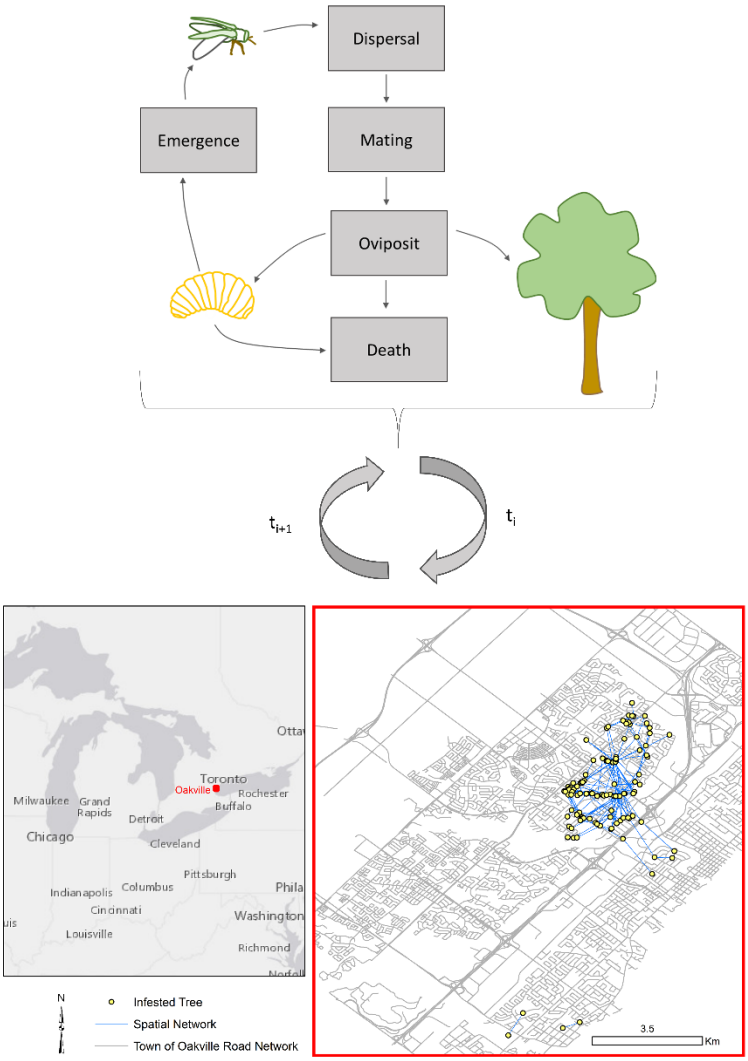

Figure 5.2. Overview of the geospatial N-ABM agents' interactions, processes, and scheduling for the simulation of dynamic spatial networks with the map of the study area in Oakville, Canada.

5.4.1. Emerald Ash Borer (EAB) Biological Background

The emerald ash borer (EAB) *Agrilus planipennis* Fairmaire (Coleoptera: Buprestidae) is an invasive phloem-feeding beetle native to countries in Asia. The EAB was first introduced into North America in the late 1990s (Siegert et al., 2014) and was discovered in 2002 in Detroit, Michigan, US and Windsor, Ontario, Canada (Straw et al., 2013). The EAB is the cause of the decline of the North American ash tree population, creating devastating ecological and economic impacts.

EAB larvae feed on ash tree phloem for one to two years before emerging as adults from under the bark of ash trees to reproduce in early June through August (Cappaert et al., 2005). Mated female EAB locally disperse in search of host trees suitable for offspring. Local dispersal is constrained by the natural flight ability of the EAB, where mated female EAB travel on average 2.8 km in 24 hours (Taylor et al., 2007). Dispersal to distances beyond the EABs' natural abilities is referred to as long-distance dispersal. This type of dispersal is facilitated by human transportation of infested wood products and firewood resulting in the establishment of satellite populations beyond the main front of infestation. As a result, ash trees that are near major roads and campgrounds are at high risk (Muirhead et al., 2006).

Mated female EAB use olfactory, tactile, and possibly auditory cues to determine the most suitable hosts for oviposition. It is believed that EAB have specific preferences for ash tree hosts of a certain type, size, location, and level of host tree stress. It has been found that ash tree types with a naturally lower resistance to insect infestations such as the green, black, and white ash, versus their blue ash counterparts, are targeted more frequently (Rebek et al., 2008; Anulewicz et al., 2008). It is also suggested that in order to sustain larval galleries, mated females prefer ash trees that are larger in size (Mercader et al., 2011) and ash trees that are closer in distance to their point of emergence (Mercader et al., 2009).

It is hypothesized that EAB host selection is influenced by volatiles emitted by stressed ash trees. Tree stress is caused by drought, woodpecker damage, wounding, and of course, insect feeding caused by larval galleries such as those produced by EAB. McCullough et al. (2009) and Tluczek et al. (2011) tested this hypothesis by girdling ash trees, a process where a 20 cm wide band of outer bark and phloem is removed, cutting off the flow of water and nutrients within the tree. They found that girdled trees captured significantly more adult EAB and had higher larval densities.

5.4.2. N-ABM

The purpose of the N-ABM is to represent the dynamic spatial network that emerges from simulated infestation dynamics. Specifically, agent interactions between EAB and ash tree agents, implemented by the ABM component of the N-ABM form the basis of the network structure that dynamically evolves over geographic space and time.

The N-ABM is implemented on datasets for the study site, the Town of Oakville, Canada, to represent the EAB infestation from June 1st, 2008 to August 31st, 2009. To account for variation in model outputs as a function of randomness incorporated into ABM processes, the model was run 50 times. The Town of Oakville (study area spanning across 138.5 km²) acquired and developed geospatial data that can in turn be used for model creation, calibration, and validation of EAB infestation dynamics. The geospatial data used in this study include the following:

1. GIS data layers of the tree inventory for the Town of Oakville with location and attribute data for all tree species (Trees, 2018);
2. GIS data layers containing (a) the location of recreational parks and campgrounds (Parks Recreation and Culture Guide, 2018) and (b) major streets for the Town of Oakville (Road Network, 2018);
3. GIS data layers representing the delimitation of actual EAB infestation according to levels of severity observed in the Town of Oakville in 2009 obtained through Forestry Services, Parks and Open Space, Town of Oakville.

Simulating Infestation Dynamics using Agents

The description of the ABM component of the N-ABM here includes some important elements from the Overview, Design concepts, and Details (ODD) protocol (Grimm et al. 2006). For a detailed description of the ABM component that includes all elements, see Anderson & Dragicevic (2016). The purpose of the ABM component of the N-ABM is to simulate interactions between EAB agents and ash tree agents. These interactions are the basis for which the tightly coupled dynamic spatial network forms and evolves over space and time. The ABM component of the N-ABM is based on an existing and validated ABM developed by Anderson & Dragićević (2016) and was previously applied to simulate the biological control of EAB in the Town of Oakville, Canada. The following sections describe how the infestation dynamics are implemented in the N-ABM as a function of agent interactions. Figure 5.2 depicts the detailed processes and scheduling pertaining to the interactions between the EAB and the ash tree agents that in turn generate the dynamic spatial networks.

Agents and State Variables

The N-ABM simulates adult EAB, EAB larvae, and ash tree hosts agents. Using biological information obtained from the literature (Table 5.2), agent individuals are

programmed with *state variables* and *agent parameters.* The *state variables* (Table 5.2a) are associated with the state of the agent at a particular model time step. The *agent parameters* (Table 5.2b) are virtual components that define the agents and shape their behavior.

Table 5.2. A description of (a) agent state variables and (b) agent parameters with references.

A. Agent State Variables	Variable	Description	
All agents	Age	Agent age in days	
	Location	Decimal degrees	
Adult EAB	Number of offspring produced	Number of offspring that have been produced by the agent	
	Fertility	Whether the agent is fertile or not	
Larvae	Sex	Sex of the larvae	
Ash Tree	Stress	Stress level associated with larval feeding as a result of infestation	
	Number of larvae	Number of larvae existing within the tree	
B. Agent Parameters	Parameter	Description	Reference
Adult EAB	Maximum flight/day	2.8 km/day	Taylor et al. (2007)
	Chance of successful fertility	82%	Rutledge & Keena (2012)
	Maximum number of offspring	Randomly selected between 60 and 90	Jennings et al. (2014)
	Survival rate of eggs	Randomly selected between 53% and 65%	Jennings et al. (2014)
Larvae	Sex ratio	1:1, 50%	Lyons & Jones (2005)
	Survival rate of larvae	Host tree defense: max 21%, disease: 3%, woodpecker: max 17%	Duan et al. (2010)
Ash Tree	Carrying capacity	44 EAB/ per m^2 of bark	Jennings et al. (2014)

Agent Process Overview and Scheduling

As depicted in Figure 5.2, the developed N-ABM simulates EAB behavior for two seasons of infestation from June 1st, 2008 (t_1), when the EAB was first introduced into the Town of Oakville, to the end of August 2009 (t_{457}). Each model time step (t_i) represents *one day* in the real world. The main processes and schedules of each agent are defined and presented in Table 5.3. Upon model initialization, the first population of female adult EAB agents emerge from an ash tree and move through their life cycle as a collection of subroutines to simulate real EAB behavior (Figure 5.2). Life cycle stages include emergence, dispersal in search of food, maturing and mating, host selection and

oviposition of EAB larvae agents, and once they have oviposited all their offspring, they die. These stages are executed as a function of the agent's age and parameters. The EAB larvae agents grow over time and if they do not die from external factors, a new generation of EAB emerge and begin their life cycles.

Host selection is a process integrated into EAB adult agent decision-making through a host selection algorithm developed by Anderson & Dragićević (2016). The host selection algorithm allows EAB agents to compare between trees within their daily flight radius and optimize their decision of which tree to infest based on their preferences at each time step. Specifically, EAB agents compare ash trees based on the tree's location, type, size, and tree stress. Ash trees are more attractive to the EAB agent in the case that they are the following: 1) a type of ash that are naturally less resistant to EAB infestation, 2) are closer to the EAB agent's location, 3) are larger, and 4) are of greater stress levels.

On average, 1% of the EAB population is transported by mechanisms of long-distance dispersal (Taylor et al 2007). In the model, long-distance dispersal is a random process where each year, 1% of the EAB population has a 30% chance of establishing successfully via long-distance dispersal. The success rate of 30% was determined through model calibration (Anderson & Dragicevic, 2015). Trees that are near major transportation networks or near recreational parks and campgrounds are susceptible to the mechanism of long-distance dispersal. Once a satellite population is established, the adult EAB agent disperses locally and the life cycle continues.

Ash tree agents are represented using tree inventory geospatial data of the Town of Oakville (Figure 5.2), containing the location, type, and size of trees. Each ash tree agent represents an ash tree. Ash tree agents record whether it is infested and the number of larvae feeding on it.

Table 5.3. Main processes and schedules of each agent.

Adult EAB Agents Processes		
Processes	Description	Scheduling
Ageing	The age of the agent is stored as a state variable.	At the initialization of an agent, agent age is equal to 0 days. With each model time step representing one day, age is increased by 1.
Emergence	EAB larvae agents emerge as adult EAB agents.	When EAB larvae agents reach the age of 340 days and if the larvae is female, they emerge as adults EAB agents.
Dispersal	Dispersal is the process whereby agents change location. Dispersal is a function of the host selection algorithm where a distance of 2.8 km/day bounds the EAB's information regarding host tree availability.	EAB agents locally disperse immediately after emergence.
Mating	EAB agent fertility is a function of the chance of fertility parameter. Mated females who are fertile are randomly assigned a maximum number of offspring.	EAB agents mate 7 days after emergence.
Oviposition	Mated fertile EAB are able to compare trees that fall within their daily flight distance radius. The comparison between trees by EAB agents is controlled by the host selection algorithm. Once selected, EAB will oviposit a random number of offspring within its maximum number of offspring onto the selected host.	If fertile, adult EAB begin oviposition 10 days after emergence.
Death	Adult EAB agents die once they have produced their maximum number of offspring.	Triggered when parameter maximum number of offspring is equal to offspring produced.
EAB Larvae Agents		
Processes	Description	Scheduling
Ageing	The age of the agent is stored as a state variable.	At the initialization of an agent, agent age is equal to 0 days. With each model time step representing one day, age is increased by 1.
Death	Larvae agents die via external factors.	External factors are applied once in the lifetime of the EAB larvae.
Tree Agents		
Ageing	The age of the agent is stored as a state variable.	At the initialization of an agent, agent age is equal to 0 days. With each model time step representing one day, age is increased by 1.
Become Infested	Ash trees become infested once an adult EAB has chosen it as a host and successfully oviposited their eggs in the tree. The number of larvae feeding on the tree proportionally increases the stress of the tree agent.	Triggered when number of larvae is greater than 1.

Agent Initialization

The model is initialized for the time t_1 on June 1st, 2008. Upon initialization, a population of EAB emerge from the North Iroquois Ridge Community (43.47 decimal degrees North and 79.69 decimal degrees West) where the beetle is thought to have first become established in the Town of Oakville (BioForest Technologies Incorporated, 2011). Although the initial number of beetle agents to emerge is random, the number falls within a threshold proportional to the carrying capacity of ash trees in the region where a maximum of 44 female adult beetles emerge per m^2 of the ash trees surface area (McCullough & Siegert, 2007). The average size of the tree in the initial location of infestation is 13 m^2, estimated by using the height and DBH recorded in the geospatial tree inventory dataset. As such, the model is initialized in 2008 with a maximum of 572 emerging female EAB agents and, with up to 41% of larval death via host tree defense, woodpecker predation, and environmental factors and disease (Duan et al., 2010), a minimum of 234 emerging female EAB agents. There are 6153 ash tree agents in the simulation.

5.4.3. Infestation Dynamics as Networks

In the developed N-ABM approach, agent interactions drive the structure of a dynamic spatial network. The programming logic and pseudocode are presented in Figure 5.3, where nodes and links in the generated spatial network correspond to EAB agents and their interactions with the forest environment. This code tightly couples the agent interactions to the generation of spatial networks. If for example, Agent A, represented by node v_i, interacts with Agent B, a new node v_j is added as a proxy for Agent B and a link *e* is created from node v_i to node v_j to represent the interaction. In addition, the link *e* stores the direction of the link (node v_i is assigned as the Start Node and *node* v_j is assigned as the End Node).

Using the case study of the EAB infestation as an example, the N-ABM generates a series of spatial networks *SN*, composed of a growing set of nodes *N* that in this case represent infested trees at *x, y* location. Directed links *L* represent the movement of a mated female EAB agent and form as it moves from infested tree node v_i to a new tree node v_j, infesting tree node v_j. Connections are made to both trees that

are not already infested and to trees that are infested and thus are already included in the infestation network.

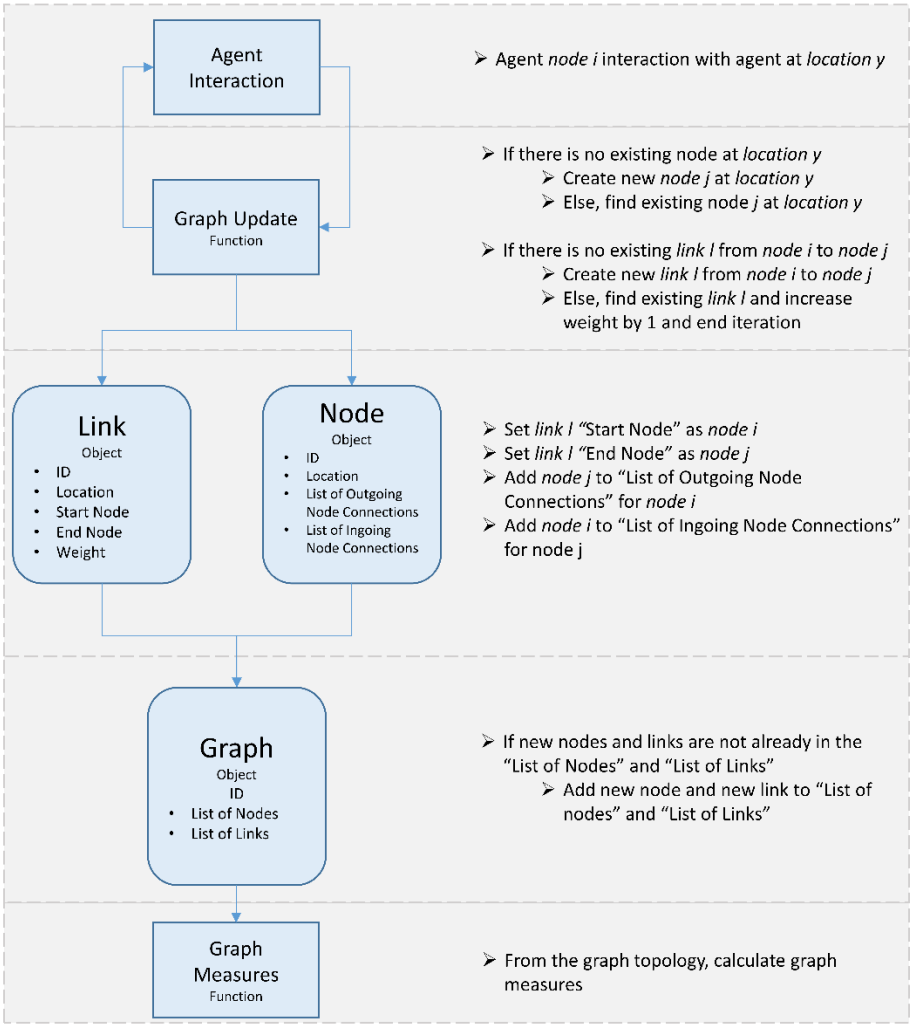

Figure 5.3. Programming logic and pseudocode developed for the N-ABM approach that couple agent interactions and the generation of the dynamic spatial network structures.

5.4.4. N-ABM Testing

The spatial networks *SN* form as a direct result of the interactions between EAB and ash tree agents over geographic space and time and as such, testing the validity of the ABM component of the N-ABM is important. The N-ABM uses the same agent reasoning and programming code from the validated ABM developed by Anderson & Dragićević (2016). To account for variation in model outputs as a function of randomness incorporated into ABM processes, the model is run 50 times. The simulated

state of each tree was determined as a function of the state of each ash tree (infested vs. not infested) in the majority of model runs. The model was validated using a binary confusion matrix that measures the agreement between the simulated state of each ash tree and the state of the corresponding ash tree as observed in the real-world for the same time-period with an overall accuracy of 72%. Additionally, a confusion matrix was used to measure the agreement between the simulated level of severity of each ash tree (low, medium, high) and the level of severity observed in the real world for the corresponding tree with an overall accuracy of 64%. The sensitivity of the model outputs to changes in parameters, most notably to changes in EAB preferences in the host-selection process, were tested using sensitivity analysis and reported in Anderson & Dragicevic, 2018.

5.4.5. Analysis of Generated Spatial Networks

Graph theory (Newman, 2003) provides the theoretical and mathematical foundation for the representation and analysis of network structures. A primary objective in network analysis using graph theory is to determine the type of network, specifically whether the observed network structure exhibits random, scale-free, or small-world properties. This can be determined using a few global network measures that characterize the network at the network-level: average node degree <k>, degree distribution *P(k),* average clustering coefficient <*C*>*,* and average path length <l>. As such, to determine network type, the above measures are calculated for the networks generated by the N-ABM and the obtained values are compared to the expected values if a network with the same number of nodes was random. Therefore, an equivalent random network model N-ABM$_{RAND}$ is developed where EAB agent host selection is programmed as a random process. The average node degree <k>, degree distribution *P(k),* average clustering coefficient <*C*>*,* and average path length <l> is calculated for the N-ABM and the N-ABM$_{RAND}$ and compared.

A detailed description of the selected graph theory measures programmed for the spatial analysis of the generated spatial network *SN* structures are as follows:

Average Node Degree

The number of nodes v_j that node v_i is connected to is referred to as the node degree *k* of v_i. In an undirected network, the average node degree <*k*> is defined as:

$$< k > = \frac{2L}{N} \qquad (2)$$

where *m* represents the number of links in the network and *n* represents the number of nodes in the network. However, not all graphs are undirected. Movement, for example, a common dynamic modelled in spatial ABMs, is directional, where EAB agents move *from* a location *to* a location. Thus, it is important to differentiate between the ingoing links k_{in} with outgoing links k_{out}. Node total degree *k* in a directed network is defined as:

$$k = k_{in} + k_{out} \qquad (3)$$

The average node degree <*k*> in a directed network is defined as:

$$< k > = \frac{L}{N} \qquad (4)$$

Degree Distribution

The degree distribution *P(k)* calculates the fraction of nodes in a network with degree *k,* $k_{in,}$ or k_{out}. Using the example of degree *k,* the fraction of nodes in a network with degree *k* can be calculated by the number of the nodes with the same degree, divided by the total number of nodes in the network *N*. For example, if there are four nodes in the network where *k = 1*, and there are a total of 10 nodes, *P(k) = 0.4*. *P(k)* can be plotted on a histogram to present the degree distribution of the network. The degree distribution is tested for goodness of fit using *powerlaw,* a Python package for analysis of heavy-tailed distributions (Alstott et al., 2014).

Average Path Length

A path $P_{i0,in}$ that connects the nodes i_0 and i_n in a graph $G = (N, L)$ is defined as an ordered collection of $n + 1$ nodes $N_P = \{i_0, i_1, i_2, \dots\ i_n\}$ and n edges $L_P = \{(i_0, i_1), (i_1, i_2) \dots\ (i_{n-1}, i_m)\}$. The shortest path length between any two nodes in the network model l_s is calculated by implementing Dijkstra's shortest path algorithm into the N-ABM (Dijkstra, 1959), an algorithm developed to find the shortest number of links connecting any two pairs of nodes in the network. Average path length <*l*> is defined as the average value of l_{ij}.

Average Clustering Coefficient

The clustering coefficient *C* measures the probability that nodes v_j connected to node v_i are also connected to each other (Watts and Strogatz, 1998). The clustering coefficient of node v_i is formulated as:

$$C(i) = \frac{E_i * 2}{k_i(k_i - 1)} \qquad (5)$$

where for node v_i of degree k_i, E_i is the number of edges among the neighbours of v_i. The average clustering coefficient is formulated as:

$$< C > \sim \frac{1}{N} \qquad (6)$$

where the brackets denote the average clustering coefficient over the network.

5.5. Results

This section presents the N-ABM simulation outcomes and the analysis of the generated spatial network structures. To account for stochasticity in generated network structures, the N-ABM is executed 50 times to ensure statistical significance, thus generating 50 different simulation outcomes. The average values and the standard error across all 50 model runs for average node degree <*k*>, degree distribution *P(k)*, average clustering coefficient <*C*>, and average path length <*l*> are calculated and presented in Table 5.4. The obtained standard error for each network measure is very small, indicating that the N-ABM generates very similar network structures across all runs.

Table 5.4. Summary of network measure results derived from (a) the single N-ABM infestation network chosen for visual presentation of results, and (b) the average value across all 50 generated infestation networks for all network measures with the associated standard error.

	A. Single Network	B. 50 Networks	
	Value	Average Value	Standard Error
Degree and Degree Distribution (k_{out})			
$<k_{out}>$	4.40 neighboring nodes	4.48 neighboring nodes	0.0146
k_{max}	64 neighboring nodes	61.22 neighboring nodes	0.4755
k_{min}	1 neighboring node	1 neighboring node	0
Degree distribution	Power law	Power law	

$p(k) \sim k^{-\alpha}$	α = 1.58	α = 1.59	0.0019
Degree and Degree Distribution (k_{in})			
$<k_{in}>$	4.48 neighboring nodes	4.33 neighboring nodes	0.0191
k_{max}	52 neighboring nodes	56.46 neighboring nodes	0.5140
k_{min}	1 neighboring node	1 neighboring node	0
Degree distribution	Power law	Power law	
$p(k) \sim k^{-\alpha}$	α = 1.59	α = 1.59	0.0020
Clustering Coefficient			
<C>	0.030	0.034	0.0002
Path Length			
$<l_s>$	11.27 nodes	11.26	0.0021

This is supported by visibly similar power law degree distributions *P(k)* across all 50 simulated network structures as presented in Figure 5.4 accompanied by minor differences in the calculated alpha exponent (Table 5.4). Large amounts of variation across several ABM outputs typically points to stochastic uncertainty (Brown et al., 2005), however the minor variation observed in network structure across 50 model runs suggests that the emergent network structure is a function of the generative mechanisms that are implemented in the model, rather than randomness. Therefore, using one network as example to present the results is a viable option as the other 49 networks have similar network structure, indicated by the similar obtained network measures for each. Based on this analysis and for the purpose of clarity and visual presentation of the generated maps with network structures, one example network is selected and presented in the following sections.

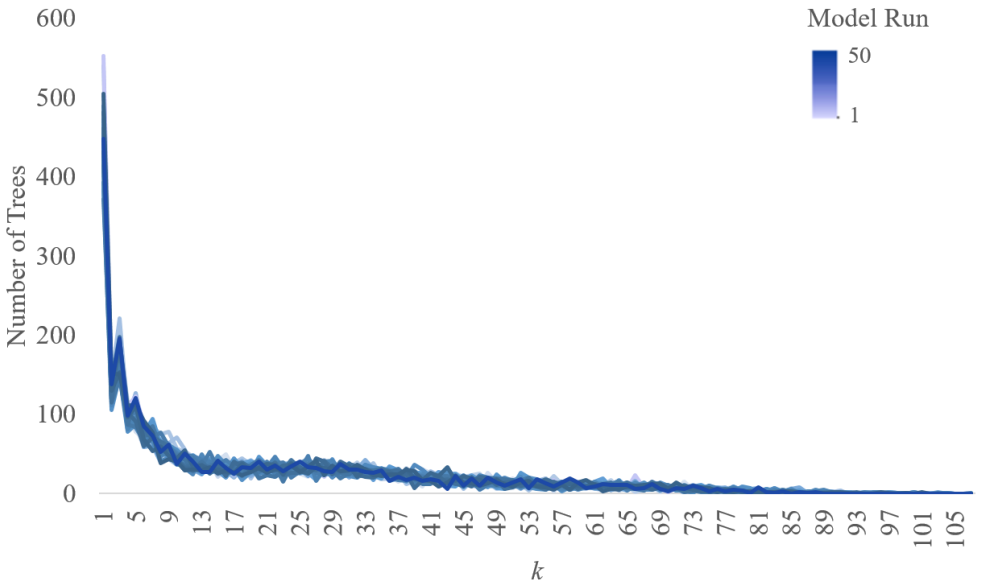

Figure 5.4. Variation of the degree distribution *P(k)* generated across all 50 model runs. The degree distribution presents the number of trees with degree *k*.

5.5.1. Simulation Results

Figure 5.5 presents one full model run of the N-ABM representing the spread of EAB infestation across the Town of Oakville, Ontario from 2008 to 2009. Specifically, Figure 5.5 shows both the infestation extent as a function of the EAB and tree agents' interactions and the generated spatial network structures for time t_i. Particularly, simulation outputs for July 2008 at t_{61} (Figure 5.5 a and b), August 2008 at t_{92} (Figure 5.5 c and d), July 2009 at t_{426} (Figure 5.5 e and f), August 2009 at t_{457} (Figure 5.5 g and h) are presented. The EAB are most active during the month of July and August, and therefore, the most notable changes in network structure are visible. The infestation extent as a function of the agent interactions (Figure 5.5 a, c, e, and g) and the network structures (Figure 5.5 b, d, f, and h) are extracted as separate geospatial layers for better visualization; however, both are derived from the same model run. Following model initialization in 2008, EAB spreads locally outward from the epicenter in the north-eastern half of the city. From the epicenter, clusters of infestation develop to the south-east and the west of the epicenter. A satellite population develops in the southern end of the city and in 2009, the second year of EAB infestation, the satellite population, and the

main infestation front merge. This is referred to as a stratified dispersal pattern, commonly observed in real-world EAB spread (Muirhead et al., 2006).

Figure 5.5. The N-ABM simulation results depicting spatial EAB infestation extent as a function of EAB and ash tree agents' interactions in the Town of Oakville between 2008 and 2009 for: (a and b) July 2008 at t_{61}; (c an d) August 2008 at t_{92}; (e and f), July 2009 at t_{426}; and (g and h) August 2009 at t_{457}.

5.5.2. Analysis of the Spatial Network Structure Results

Using the topology of the example simulated EAB infestation network in August 2009 derived from the N-ABM at the final time step t_{457}, the resulting graph theory measures are summarized in Table 5.4.

Graph Size

The EAB infestation network in August 2009 is composed of 2540 nodes, meaning that there are just over 2500 infested trees in the network. The number of links or dispersal pathways connecting these nodes or infested trees is 27,560.

Average Node Degree and Degree Distribution

The k_{out} and k_{in} measures mathematically characterize different dynamics in the infestation network. A $<k_{out}>$ value of 4.48 indicates that on average, EAB agents move from infested tree node v_i to 4.48 other tree nodes v_j. Therefore, k_{out} characterizes the *local connectivity* of each tree node v_i to other desirable trees v_j in its proximity. A $<k_{in}>$ value of 4.40 indicates that on average, EAB agents move from 4.40 other tree nodes v_j to tree node v_i. Therefore, k_{in} characterize the *desirability* of tree node v_i.

The max degree k_{max} for k_{out} is 64 and the min degree k_{min} for k_{out} is 1. The max degree k_{max} for k_{in} is 52 and the min degree k_{min} for k_{in} is 1. The values of degree do not indicate the volume of beetles that move across these links, but rather, that several beetles move along each path from node v_i to node v_j. The volume of beetles can be represented by node or link weights; however, node and link weights are not included in this network representation to maintain simplicity. The degree distribution $P(k_{out})$ and $P(k_{in})$ can provide information regarding the type of network that emerges from dynamics at the local scale. The $P(k_{out})$ and $P(k_{in})$ are plotted as histograms in Figure 5.6a and 6b respectively and as a log-log plot in Figure 5.6 c and 6d. The histograms in Figure 5.6a and 6b indicate that there are a large fraction of nodes in the network with a very small degree and a small fraction of nodes in the network with a very large degree. In the log-log plot (Figure 5.6c and Figure 5.6d), this distribution can be described as linear distributions with heavy tails.

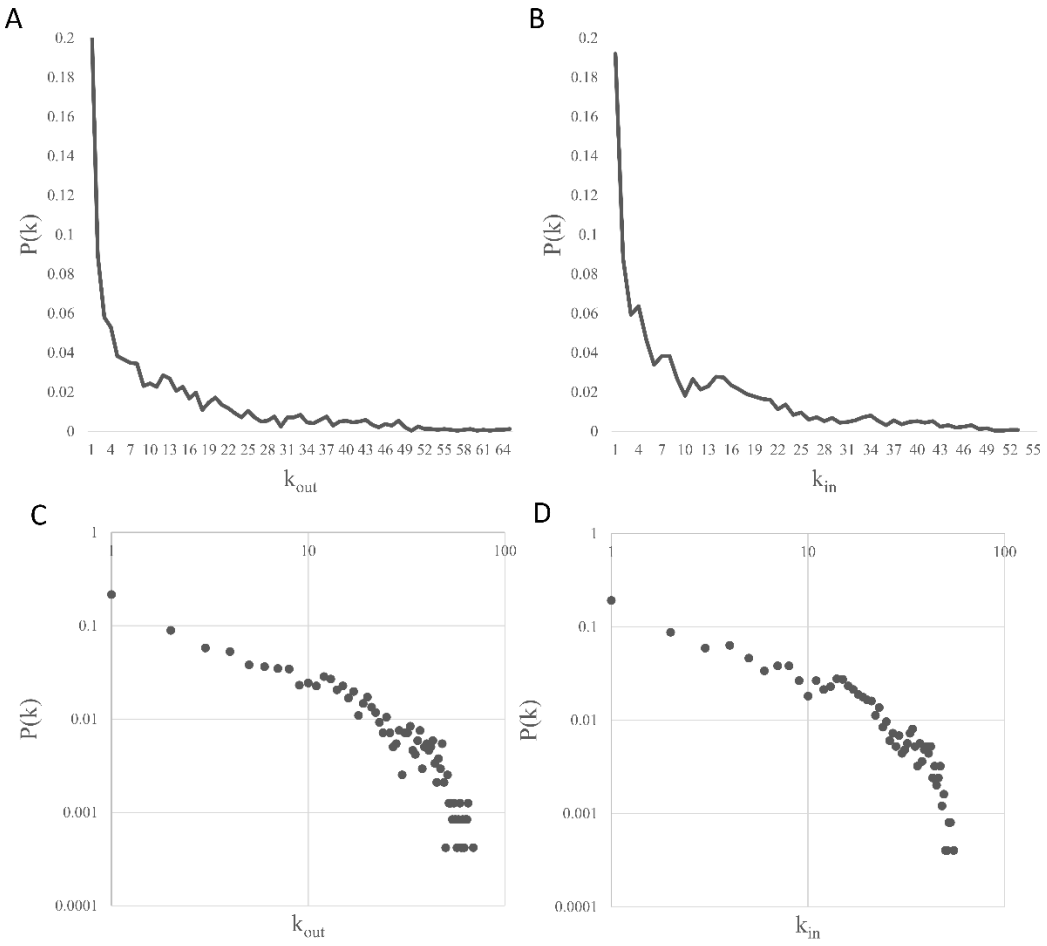

Figure 5.6. Degree distribution *P(k)* for (a) k_{out} and (b) k_{in} and degree distribution *P(k)* plotted on a log-log scale for (c) k_{out} and (d) k_{in}.

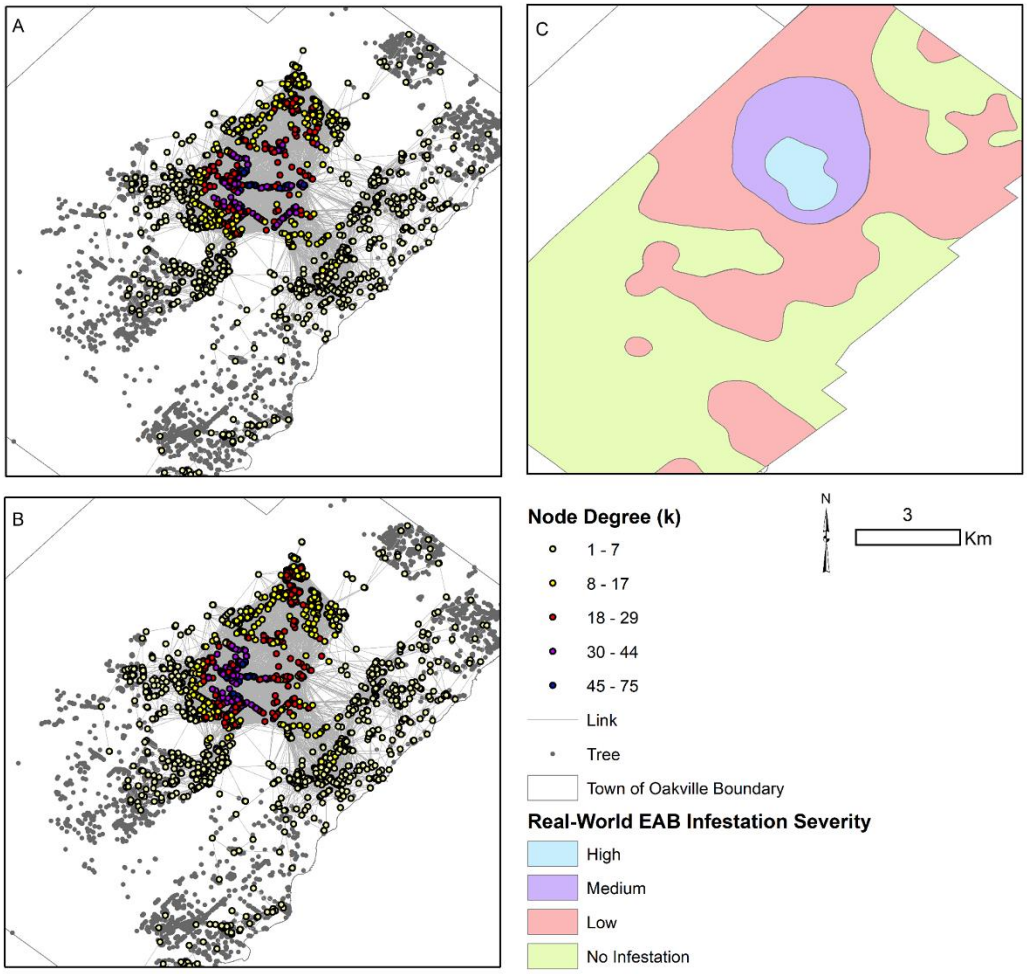

Figure 5.7. Spatial distribution of tree node degree for (a) k_{out} and (b) k_{in} demonstrating highly connected trees and less connected trees for EAB insect infestation. The simulated high degree trees correspond to the (c) regions of real-world high severity infestation.

The degree distribution for $P(k_{out})$ and $P(k_{in})$ can be described as a power law distribution. The distributions produce an alpha exponent $-\alpha$ of 1.58 and 1.59 for $P(k_{out})$ and $P(k_{in})$ respectively and a standard error sigma σ of 0.01 for both. The Kolmogorov-Smirnov distance D is also small with a value of 0.16 for both. The goodness of fit of the degree distribution between a power law and an exponential distribution is first compared, producing a log likelihood ratio R between the two candidate distributions that indicates a power law is a better fit than an exponential with a p-value of 0.0072 and 0.04 for $P(k_{out})$ and $P(k_{in})$ respectively.

The directed degrees are also presented spatially (Figure 5.7a and 5.7b). For both k_{out} and k_{in}, there is a distinguishable core region composed of ash trees that have a high node degree *k*, surrounded by a large perimeter composed of ash trees that have a low node degree *k*.

Node degree *k* appears to coincide with real-world EAB infestation severity (Figure 5.7c; Table 5.5). This is determined by classifying all tree nodes into *high, medium, low,* and *zero degree* classes based on their simulated degree. The degree of a tree is classified as *high degree* if it falls > 1.5 standard deviation, *medium degree* if from 0.5 to 1.5 of the standard deviation, and as *low degree* if it falls <0.5 of the standard deviation. All trees with a simulated node degree of zero are classified as zero degree. Using a confusion matrix approach (Congalton, 1991), the simulated degree class of each tree node is compared to *high, medium, low infestation* severity in addition to *no infestation* observed in the real world for 2009 for the same tree. The overall spatial similarity is 67% where trees with a simulated high degree and medium degree corresponds moderately well with high (18% omission and 36% commission) and with medium (33% omission and 18% commission) infestation severity in the real world respectively. This is not a model validation metric because since the network is not weighted, the tree node degree does not necessarily correlate with tree population EAB density and thus cannot be compared to infestation severity. However, it does indicate that trees that have a high degree in the infestation network are highly spatially accessible and may influence infestation severity.

Table 5.5. Confusion matrix detailing the spatial agreement of the levels of severity of infestation observed in the real world and simulated node degree.

		Real world tree infestation severity				Commission
		Not Infested	Low Infested	Medium Infested	Highly Infested	
Simulated Tree Degree	Zero Degree	2435	1556	10	0	39.15%
	Low Degree	95	1209	186	0	18.86%
	Medium Degree	0	88	478	17	18.02%
	High Degree	0	0	42	73	36.53%
Omission		3.76%	57.63%	33.25%	18.89%	67.77%

Degree Distribution across Time

Using the developed N-ABM approach, all graph theory measures can be calculated at any point in time to better understand the change in network structure as it grows and evolves. The degree distribution $P(k_{out})$ is used here as an example. In each model time step the $P(k_{out})$ of the infestation network maintains a power law. The power law in the early stages of the infestation has an exponent of 3.3. Over time, the exponent decreases before settling at 1.6 (Table 5.6). The generated power laws indicate that at the beginning of each season of EAB infestation there is a burst of new EAB dispersal where the network is expanded over geographic space and infested tree nodes with a low degree are added to the network. The season of EAB infestation continues and links form between trees that already exist in the infestation network as EAB dynamics such as attraction to trees of ash species, attraction to stressed trees, and carrying capacity begin to play a role in the network structure. This process is quantified by the changes in the power law degree distributions over time, where the power law in early months of infestation has a much higher fraction of tree nodes with a low degree than in the following months.

Table 5.6. Power law exponent $-\alpha$ over time.

Date	Model Time Step	Exponent $-\alpha$
June 2008	T_{30}	3.3
July 2008	T_{61}	2.7
August 2008	T_{92}	2.7
June 2009	T_{395}	2.1
July 2009	T_{426}	1.6
August 2009	T_{457}	1.6

Average Clustering Coefficient

The average clustering coefficient <C> is 0.03 meaning that node linkages form a "star-like" pattern. Most of the core region of infested trees is composed of trees with a very low clustering coefficient, meaning that only a few of the nodes that are connected to node v_i are also connected to each other. In total, there are only three trees with a clustering coefficient *C* of 1, where all nodes that are connected to node v_i are also connected to each other. We can present this spatially to highlight where clusters exist (Figure 5.8). The majority of the nodes with high clustering are near the perimeter. This is potentially a model edge effect, where EAB agents can 'no longer' spread due to the

spatial data's administrative town boundary and thus continue to infest trees in close proximity resulting in a higher clustering coefficient near the town administrative boundary.

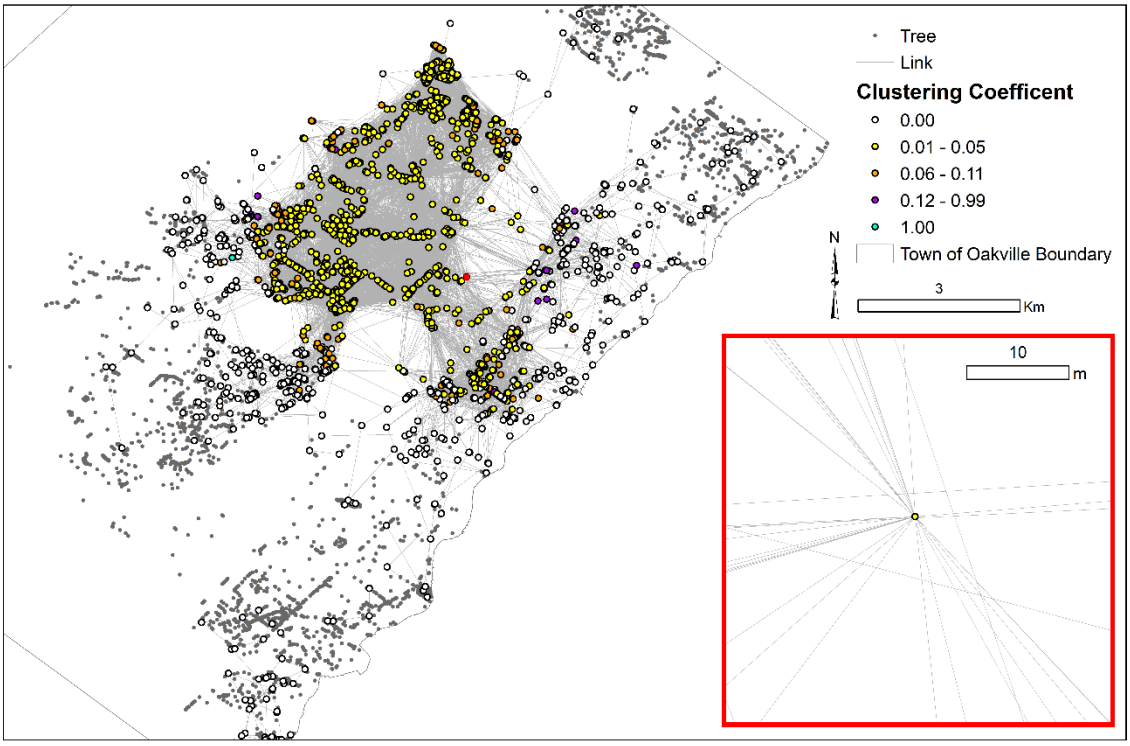

Figure 5.8. Spatial distribution of varying clustering coefficients of tree nodes. The inset map corresponds with the red dot on the main map and presents a node exhibiting the star like pattern commonly generated during the EAB host selection process.

Average Path Length

The average path length is short, with an average of 11 intermediate nodes that make up the path between any two nodes in the network. This indicates that as a function of the spatial distribution of ash tree hosts and the flight dynamics of EAB, the EAB are able to spread across long distances in a short period of time.

Using the topology of the simulated EAB infestation network in August 2009 derived from the N-ABM$_{RAND}$ at the final time step t_{457}, the resulting graph theory measures for an equivalent random network are summarized in Table 5.7.

Table 5.7. Summary of network measure results derived from the N-ABM$_{RAND}$ infestation network.

Graph Size		
	$<N>$	2,540 nodes
	$<L>$	3,788 links
Degree and Degree Distribution (k_{out})		
	$<k_{out}>$	6.4 neighboring nodes
	k_{max}	38 neighboring nodes
	k_{min}	1 neighboring node
	Degree distribution	Poisson
Degree and Degree Distribution (k_{in})		
	$<k_{in}>$	6.4 neighboring nodes
	k_{max}	17 neighboring nodes
	k_{min}	1 neighboring node
	Degree distribution	Poisson
Clustering Coefficient		
	$<C>$	0.007
Path Length		
	$<l_s>$	22.86 nodes

Comparing the obtained N-ABM analysis results for the case study of the EAB with the N-ABM$_{RAND}$ indicates that the N-ABM spatial infestation network structure is scale-free with hub and spoke architecture, formed by a power law degree distribution for both $P(k_{out})$ and $P(k_{in})$ with an exponent α of 1.58 and 1.59, respectively. Specifically, the short average path length <l> of the N-ABM at 11.27 nodes in comparison to the average path length <l> of the N-ABM$_{RAND}$ at 22.86 nodes. In addition to a power law degree distribution $P(k)$, an average path length in a network that is less than the average path length of its equivalent random network is a defining feature of a scale-free network (Barabási, 2016).

Characterization of emergent spatial network structure can provide insight into network dynamics. For example, if instead of a scale-free network, a random network structure was formed, this would indicate that the underlying processes driving the emergence of the spatial network structure are random. At the most basic level, scale-free networks form as the result of two simple generative mechanisms: growth and preferential attachment, also known as the "rich get richer" phenomenon. In preferential attachment, the probability that a new node becomes connected to an existing node is a function of the existing node's degree. The higher the degree, the more likely the node will form a connection.

To understand why the scale-free network structure emerges, the generative mechanisms included in the N-ABM can be examined. A correlation between degree and tree stress at $r^2 = 0.87$ indicates that the EAB host selection process, specifically the EAB attraction to stressed trees, is responsible for the generation of dynamics of preferential attachment and thus the emergent scale-free network structure. In the real world, trees that are preferable for oviposition are infested, resulting in the release of stress volatiles. This creates a positive feedback whereas ash trees suitable for host selection become increasingly infested, stress volatiles are released, and thus the trees are increasingly targeted for attack. Networks generated by preferential attachment tend to have high degree nodes or hubs near the center of the network with node degrees that gradually declines toward the perimeter and outer edges of the network. This emerging pattern is produced in the developed N-ABM.

Further analysis of the network structure does not indicate a significant relationship between node degree and tree size or node degree and tree type. Non-spatial scale-free models are not constrained by space, and thus commonly contain long-distance linkages to non-adjacent nodes, however, in the spatial N-ABM, long-distance links come at a cost and thus are limited to the natural dispersal distance of the EAB (Taylor et al., 2007). The lack of relationship between node degree and tree size or tree type may be a result of the heterogeneous nature of host trees in the environment and the limitation of generating long-distance links. Trees of the most attractive type in the study area may not be available within the flight distance radius allotted to the EAB agent per day, leaving agents to choose the largest or best type available to them at a closer proximity.

Based on the obtained simulation, results suggest that the emergent spatial pattern of EAB spread are primarily a function of the generative mechanisms of preferential attachment based on tree stress, tree distance, and growth or continuing spread of infestation over time. In this study, it is empirically demonstrated that in the same way scale-free networks emerge from growth and preferential attachment, patterns of EAB insect infestation emerge from host selection dynamics over time. Furthermore, the characterization of network structure can provide additional insights useful for pest management. For example, scale-free networks are particularly robust to the removal of nodes at random, meaning that to be successful in eradication, the removal of trees would need to be highly strategized.

5.6. Discussion and Conclusion

Exploring the influence of the spatial structure of the landscape on ecological processes is critical. Landscape connectivity graphs are more recently implemented for this purpose, but are typically static network representations at a single scale. As a result, landscape connectivity graphs are unable to capture the complexity in dispersal processes whereby interactions at the local scale generate spatio-temporal dispersal patterns at the larger scale. For example, in the case of insect infestation, it is important to explore how interactions between EAB and varying spatial distributions of ash tree hosts generate large-scale spatio-temporal patterns of infestation.

The inclusion of complexity in spatial ecological network approaches can characterize and quantify spatio-temporal patterns in ecological systems that might otherwise be described in a more qualitative manner, identify the underlying interactions and generative mechanisms that drive the emergence of these spatio-temporal patterns, and quantitatively link dynamics at local scales to emergent patterns at larger scales. Therefore, this study proposes the integration of the complex systems modelling approach ABM with spatial networks for the development of a network-based ABM. The approach is demonstrated in application to the case study of the forest insect infestation, EAB. The N-ABM approach facilitates the exploration between network dynamics and network structure, and in this case, links the EAB attraction to stressed trees at the individual scale to dispersal patterns at larger scales. This effectively supports findings in the literature that point to tree stress as the primary factor in host selection (McCullough et al., 2009; Tluczek et al., 2011). The approach facilitates the shift from the characterization of network structure to explaining and identifying the underlying interactions that drive the emergence of a dynamic and evolving spatial network structure.

Across all 50 model runs, stochastic processes within the N-ABM produce a distribution of unique spatial networks, however results show that each of the simulated spatial networks have a highly similar structure. This conclusion is supported by the fact that all spatial network measures for all 50 networks deviate only slightly from the mean and thus have a small standard error. As such, any selected model run would be fairly, although not perfectly, representative of the complete distribution of networks. Therefore, only one randomly selected network from one model run is presented in detail. This

serves the purpose of clarity, since it would be overwhelming to visually present all 50 network structures. In future work, it may be useful to develop an approach to summarize a distribution of spatial networks that emerge from the same local processes. Averaging the degree of each node across all 50 model runs is insufficient because it would result in a smoothing effect that would ultimately negate network degree heterogeneity, which is of great interest in network studies. In summary, as a function of the randomness of the N-ABM, the spatial pattern of EAB spread varies from model run to model run, but the network structure does not, presenting a clear link between network dynamics and network structure. Detailed analysis of the variation of the spatial patterns of EAB spread across model runs resulting from the ABM component used in the N-ABM can be found in Anderson & Dragicevic (2016; 2018).

Spatial ecological networks can be represented and analyzed at a variety of scales. Nodes can represent ecological entities from the individual level to various aggregations of individuals (community level, population level, landscape level). The modifiable aerial unit problem (MAUP) is a phenomenon where statistical outputs vary as a function of the level of aggregation in the model. This suggests that the emergent network structure may be a function of the scale at which the phenomenon is represented. However, the aim of the study is to integrate complexity into spatial ecological networks using an agent-based modelling approach, justifying the use of individual-scale representation. It would be desirable to validate the spatial networks themselves generated by the N-ABM. Unfortunately, the nature of the ecological data and current data collection tools and methods cannot provide network datasets and therefore a model of this kind cannot be validated in this context. However, the validation of the ABM agent processes and interactions integrated in the N-ABM that generate the spatial networks gives confidence that the spatial networks are being represented correctly.

In conclusion, the novel N-ABM modelling approach presented here is unique and particularly relevant for modelling complex ecological systems, as there is a demand for the exploration of dynamic ecological networks in a spatial and temporal context. The application of graph theory to the networks generated by the N-ABM helps to better understand, measure, and analyze the influence of geographic space and network structure on network dynamics as well as characterize dispersal patterns, particularly useful from an ecological management perspective. The N-ABM framework is also

highly general and flexible as to facilitate the representation and simulation of many ecological systems as dynamic evolving networks. Graph theory provides a large toolset of additional measures that can also be applied for further understanding interactions between other ecological species and the landscape, useful for ecological management or conservation. For example, graph theory measures such as betweenness centrality and link weights can help important habitat features and dispersal pathways that are essential to the connectivity of the landscape for an endangered species and thus be targeted for protection by ecological conservationists and decision makers.

5.7. References

Agrawal, A. A., Ackerly, D. D., Adler, F., Arnold, A. E., Cáceres, C., Doak, D. F., Post, E., Hudson, P.J., Maron, J., Mooney, K.A., & Power, M., Schemske, D., Stachowicz, J., Strauss, S., Turner, M.G. & Werner, E. (2007). Filling key gaps in population and community ecology. *Frontiers in Ecology and the Environment*, 5(3), 145-152.

Albert, R., Jeong, H., & Barabási, A. L. (1999). Internet: diameter of the world-wide web. *Nature,* 401(6749), 130-131.

Alstott, J., Bullmore, E. & Plenz, D. (2014). powerlaw: a Python package for analysis of heavy-tailed distributions. *PloS One*, 9.1, e85777.

Anderson, T. & Dragićević, S. (2015). An agent-based modeling approach to represent infestation dynamics of the emerald ash borer beetle. *Ecological Informatics,* 30, 97-109.

Anderson, T. & Dragićević, S. (2016) Geospatial pest-parasitoid agent-based model for optimizing biological control of forest insect infestation. *Ecological Modelling,* 337, 310-329.

Anderson, T., Dragicevic, S. (2018). Deconstructing geospatial agent-based model: Sensitivity analysis of forest insect infestation model. In Perez, L., Eun-Kyeong, K., Sengupta, R. (eds) *Agent Based Models and Complexity Science in the Age of Geospatial Big Data* (Springer), 31-44.

Andersson, E., & Bodin, Ö. (2009). Practical tool for landscape planning? An empirical investigation of network based models of habitat fragmentation. *Ecography*, *32*(1), 123-132.

Andris, C. (2016). Integrating social network data into GISystems. *International Journal of Geographical Information Science*, *30*(10), 2009-2031.

Anulewicz, A. C., McCullough, D. G., Cappaert, D. L., & Poland, T. M. (2008). Host range of the emerald ash borer (Agrilus planipennis Fairmaire) (Coleoptera: Buprestidae) in North America: Results of multiple-choice field experiments. *Environmental Entomology,* 37(1), 230–241.

Barabási, A. L. (2016). *Network Science.* Cambridge, UK: Cambridge University Press.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science,* 286(5439), 509-512.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1), 1-101.

Berryman, M. J., & Angus, S. D. (2010). Tutorials on agent-based modelling with NetLogo and network analysis with Pajek. In *Complex physical, biophysical and econophysical systems* (pp. 351-375).

BioForest Technologies Inc. (2011). *TreeAzin Systemic Insecticide: Evidence for biennial Emerald Ash Borer treatments (Agrilus plannipennis Fairmaire).* Sault Ste. Marie: Bioforest Technologies Inc. Retrieved on December 21, 2017 from http://www.bioforest.ca/documents/assets/uploads/files/en/bioforest_2011_-_treeazin_2yr_efficacy.pdf.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: structure and dynamics. *Physics Reports,* 424(4-5), 175–308.

Bone, C., & Altaweel, M. (2014). Modeling micro-scale ecological processes and emergent patterns of mountain pine beetle epidemics. *Ecological Modelling*, 289, 45–58.

Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International journal of Geographical Information Science*, *19*(2), 153-174.

Bunn, A. G., Urban, D. L., & Keitt, T. H. (2000). Landscape connectivity: a conservation application of graph theory. *Journal of Environmental Management*, 59(4), 265-278.

Cappaert, D., McCullough, D. G., Poland, T. M., & Siegert, N. W. (2005). Emerald ash borer in North America: A research and regulatory challenge. *American Entomologist,* 51(3), 152-165.

Cohen, J. E. (1978). *Food Webs and Niche Space.* New Jersey, NJ: Princeton University Press.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*,*37*(1), 35-46.

Dijkstra, E. W. (1959). A note on two problems in connection with graphs. *Numerische Mathematik*, 1(1), 269–271.

Duan, J. J., Ulyshen, M. D., Bauer, L. S., Gould, J., & Van Driesche, R. (2010). Measuring the impact of biotic factors on populations of immature emerald ash borers (Coleoptera: Buprestidae). *Environmental Entomology*, 39 (5), 1513-1522.

Dupont, Y. L., Trøjelsgaard, K., & Olesen, J. M. (2011). Scaling down from species to individuals: a flower–visitation network between individual honeybees and thistle plants. *Oikos,* 120(2), 170-177.

Erdos, P., & Renyi, A. (1959). On random graphs. *Selected Papers of Alfred Renyi*, 2(1), 308–315.

Erdos, P., & Renyi, A. (1960). On the evolution of random graphs. *Bulletin of the International Statistical Institute*, 34(4), 343–347.

Ferrari, J. R., Lookingbill, T. R., & Neel, M. C. (2007). Two measures of landscape-graph connectivity: assessment across gradients in area and configuration. *Landscape ecology*, *22*(9), 1315-1323.

Fortuna, M. A., & Bascompte, J. (2007). The network approach in ecology. In Valladares, F., Camacho, A., Elosegi, A., Gracia, C., Estrada, M., Senar, J.C. & Gili, J.M. (Eds.), *Unity in diversity: Ecological reflections as a tribute to Margalef.* (371-392). Bilbao: Fundación BBVA.

Fortuna, M. A., Gómez-Rodríguez, C., & Bascompte, J. (2006). Spatial network structure and amphibian persistence in stochastic environments. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1592), 1429-1434.

Grimm, V. & Railsback, S.F. (2013). *Individual-based Modeling and Ecology*. New Jersey, NJ: Princeton University Press.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huth, A., Jepsen, J.U., Jorgensen, C., Mooij, W.M., Buller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Rossmanith, E., Ruger, N., Strand, E., Souissi, S., Stillman, R.A. Vabo, R., Visser, U. & DeAngelis, D.L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, *198*(1-2), 115-126.

Guimera, R., & Amaral, L. A. N. 2004. Modeling the world-wide airport network. *The European Physical Journal B-Condensed Matter and Complex Systems,* 38(2), 381-385.

Hall, S. J., & Raffaelli, D. G. (1993). Food webs: theory and reality. *Advances in Ecological Research*, 24, 187-239.

Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., ... & Lauridsen, R. B. (2009). Ecological networks–beyond food webs. *Journal of Animal Ecology*, 78(1), 253-269.

Jennings, D. E., Taylor, P. B., & Duan, J. J. (2014). The mating and oviposition behavior of the invasive emerald ash borer (Agrilus planipennis), with reference to the influence of host tree condition. *Journal of Pest Science,* 87(1), 71-78.

Jiang, B. 2007. A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and its Applications,* 384(2), 647-655.

Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *The American Naturalist*, 129(5), 657-677.

Kirer, H., & Çirpici, Y. A. (2016). A Survey of Agent-Based Approach of Complex Networks. *Ekonomik Yaklasim*, *27*(98), 1-28.

Letcher, B. H., Rice, J. A., Crowder, L. B., & Rose, K. A. (1996). Variability in survival of larval fish: disentangling components with a generalized agent-based model. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(4), 787-801.

Lewis, T. G. (2011). *Network Science: Theory and Applications*. New Jersey, NJ: John Wiley & Sons.

Lyons, D. B., & Jones, G. C. (2005). The biology and phenology of the emerald ash borer. In *Proceedings, 16th US Department of Agriculture interagency research forum on gypsy moth and other invasive species* (pp. 62-63).

McCullough, D. G., & Siegert, N. W. (2007). Estimating potential emerald ash borer (Coleoptera: Buprestidae) populations using ash inventory data. *Journal of Economic Entomology*, 100(5), 1577–1586.

McCullough, D. G., Poland, T. M., Anulewicz, A. C., & Cullough, D. G. M. C. (2009). Emerald ash borer (Coleoptera: Buprestidae) attraction to stressed or baited ash trees. *Environmental Entomology*, 38(6), 1668–1679.

Mercader, R. J., Siegert, N. W., Liebhold, A. M., & McCullough, D. G. (2011). Influence of foraging behavior and host spatial distribution on the localized spread of the emerald ash borer, Agrilus planipennis. *Population Ecology*, 53(2), 271-285.

Mercader, R. J., Siegert, N. W., Liebhold, A. M., & McCullough, D. G. (2009). Dispersal of the emerald ash borer, Agrilus planipennis, in newly-colonized sites. *Agricultural and Forest Entomology,* 11(4), 421-424.

Minor, E. S., & Urban, D. L. (2007). Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecological Applications*, 17(6), 1771-1782.

Morris, R. J., Lewis, O. T., & Godfray, H. C. J. (2004). Experimental evidence for apparent competition in a tropical forest food web. *Nature*, 428(6980), 310-313.

Muirhead, J. R., Leung, B., Overdijk, C., Kelly, D. W., Nandakumar, K., Marchant, K. R., & MacIsaac, H. J. (2006). Modelling local and long-distance dispersal of invasive emerald ash borer Agrilus planipennis (Coleoptera) in North America. *Diversity and Distributions,* 12(1), 71-79.

Muller, C. B., Adriaanse, I. C. T., Belshaw, R., & Godfray, H. C. J. (1999). The structure of an aphid–parasitoid community. *Journal of Animal Ecology*, 68(2), 346-370.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167-256.

North, M. J., Collier, N. T., Ozik, J., Tatara, E. R., Macal, C. M., Bragen, M., & Sydelko, P. (2013). Complex adaptive systems modeling with Repast Simphony. *Complex Adaptive Systems Modeling*, 1(1), 1-26.

Parks Recreation and Culture Guide [Tabular Data]. Oakville Open Data, July 10, 2018. Available: https://portal-exploreoakville.opendata.arcgis.com/datasets/toak::parks-recreation-and-culture-guide and (accessed July 16, 2018).

Parunak, H. V. D., Savit, R., & Riolo, R. L. (1998). Agent-based modeling vs equation-based modeling : a case study and users' guide. *Multi-Agent Systems and Agent-Based Simulation*, Berlin, Heidelberg: Springer.

Pascual-Hortal, L., & Saura, S. (2008). Integrating landscape connectivity in broad-scale forest planning through a new graph-based habitat availability methodology: application to capercaillie (Tetrao urogallus) in Catalonia (NE Spain). *European Journal of Forest Research*, *127*(1), 23-31.

Pereira, M., Segurado, P., & Neves, N. 2011. Using spatial network structure in landscape management and planning: a case study with pond turtles. *Landscape and Urban Planning,* 100(1), 67-76.

Perez, L., & Dragićević, S. (2010). Modeling mountain pine beetle infestation with an agent-based approach at two spatial scales. *Environmental Modelling & Software*, 25(2), 223-236.

Rebek, E. J., Herms, D. a, & Smitley, D. R. (2008). Interspecific variation in resistance to emerald ash borer (Coleoptera: Buprestidae) among North American and Asian ash (Fraxinus spp.). *Environmental Entomology,* 37(1), 242–246.

Repast Simphony. (2017). Version 2.5. [Computer Software]. Chicago, IL: University of Chicago.

Road Network [Shapefile]. Oakville Open Data, June 27, 2018. Available: https://portal-exploreoakville.opendata.arcgis.com/datasets/8b09936309264feeb57cb45f8018ab14_26 and (accessed July 16, 2018).

Rutledge, C. E., & Keena, M. A. (2012). Mating frequency and fecundity in the emerald ash borer Agrilus planipennis (Coleoptera: Buprestidae). *Annals of the Entomological Society of America*, 105(1), 66-72.

Schintler, L. A., Kulkarni, R., Gorman, S. & Stough, R. (2007). Using raster-based GIS and graph theory to analyze complex networks. *Networks and Spatial Economics*, 7(4), 301-313.

Semeniuk, C. A. D., Musiani, M., Hebblewhite, M., Grindal, S., & Marceau, D. J. (2012). Incorporating behavioral–ecological strategies in pattern-oriented modeling of caribou habitat use in a highly industrialized landscape. *Ecological Modelling*, 243, 18-32.

Siegert, N. W., McCullough, D. G., Liebhold, A. M., & Telewski, F. W. (2014). Dendrochronological reconstruction of the epicentre and early spread of emerald ash borer in North America. *Diversity and Distributions*, *20*(7), 847-858.

Stang, M., Klinkhamer, P. G., & Van Der Meijden, E. (2006). Size constraints and flower abundance determine the number of interactions in a plant–flower visitor web. *Oikos*, 112(1), 111-121.

Stoneham, A. K. M. 1977. The small-world problem in a spatial context. *Environment and Planning A,* 9(2), 185-195.

Stouffer, D. B., Fortuna, M. A., & Bascompte, J. (2010). Ideas for moving beyond structure to dynamics of ecological networks. In Boccaletti, S. (Ed), *Handbook On Biological Networks* (307-328). Singapore: World Scientific Publishing.

Straw, N. A., Williams, D. T., Kulinich, O., & Gninenko, Y. I. (2013). Distribution, impact and rate of spread of emerald ash borer Agrilus planipennis (Coleoptera: Buprestidae) in the Moscow region of Russia. *Forestry,* 86(5), 515-522.

Taylor, R. A., Poland, T. M., Bauer, L. S., Windell, K. N., & Kautz, J. L. (2007). Emerald ash borer flight estimates revised. *Proceedings of the emerald ash borer/Asian longhorned beetle research and technology.* FHTET-2007-04. US Department of Agriculture Forest Service, Forest Health Technology Enterprise Team, Morgantown, West Virginia.

Tluczek, A. R., McCullough, D. G., & Poland, T. M. (2011). Influence of host stress on emerald ash borer (Coleoptera: Buprestidae) adult density, development, and distribution in Fraxinus pennsylvanica trees. *Environmental Entomology,* 40(2), 357-366.

Travis, J. M., & Dytham, C. (1998). The evolution of dispersal in a metapopulation: a spatially explicit, agent-based model. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1390), 17-23.

Trees [Shapefile]. Oakville Open Data, May 11, 2018. Available: https://portal-exploreoakville.opendata.arcgis.com/datasets/18baba686f6c445f8b942c1c9830cf0d_0 and (accessed July 16, 2018).

Urban, D., & Keitt, T. (2001). Landscape connectivity: a graph-theoretic perspective. *Ecology*, 82(5), 1205-1218.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442.

Zetterberg, A., Mörtberg, U. M., & Balfors, B. (2010). Making graph theory operational for landscape ecological assessments, planning, and design. *Landscape and Urban Planning*, *95*(4), 181-191.

Chapter 6.

A Validation Approach for Spatially Explicit Agent-Based Models using Networks

6.1. Abstract

Agent-based models (ABM) are used for the representation and analysis of a variety of complex systems. These bottom-up modelling approaches aim to represent local spatial dynamics from which observable spatial patterns at the system level emerge. In order to be useful, the validity of ABMs must be demonstrated. However, the ability of ABMs to incorporate complexity makes the process of validation challenging. Traditional validation methods that evaluate simulated patterns at the system-level are limited in their ability to reveal whether the ABM internal processes driving these emergent patterns are represented realistically. However, developed validation approaches that focus on internal model processes rely mostly on qualitative evidence. Therefore, the main objective of this study is to develop and implement a novel validation approach that integrates ABM and networks for the representation of complex systems as measureable and dynamic networks. These network representations can then be measured and compared against empirical regularities of observed real networks. This approach, called the NEtworks for ABM Testing (NEAT) is proposed and implemented to validate a theoretical ABM representing the spread of influenza in the City of Vancouver, Canada. Results demonstrate that the NEAT approach is capable to mathematically characterize the ABM processes and emergent patterns to increase confidence that the generative mechanisms included in the model are realistic. The proposed NEAT approach empowers the overall ABM development process by strengthening their validity and supporting the generation of structurally realistic ABMs.

6.2. Introduction

Complex system modelling approaches such as agent-based models (ABMs) are used to simulate a variety of complex spatio-temporal phenomena by explicitly representing small-scale interactions and processes that drive the emergence of system-level behaviour (Castle & Crooks, 2006). This bottom-up modelling approach is

alternative to top-down equation-based models that use mathematical functions to represent system-level processes and cannot easily capture the complexity resulting from individual-level heterogeneity, adaptation, local spatial and temporal variations, and subsequent non-linearity inherent to real-world complex systems (Parunak, 1998; Railsback, 2001; Torrens, 2010). More recently, ABMs are integrated with real geospatial data and geographic information systems (GIS) for the explicit representation of spatial systems (Bousquet & LePage, 2004).

ABMs can be used as an investigative tool to understand how spatial systems respond to perturbations which is useful for decision-makers (McLane et al., 2011). Naturally, decision-makers including stakeholders or policy makers will demand validity of these models in order to reduce the possibility of making unsuitable decisions and risk time and money (Augusiak et al., 2014). However, because of the complexity of ABMs, unique challenges arise with respect to model validation, stemming from confusion over to what degree model validation is even feasible, and perhaps more importantly, the use of appropriate measures to evaluate these models. Traditional ABM validation approaches compare model outputs with independent empirical data, usually at the system level (Goodall, 1972). For example, to validate an ABM simulating the spatial spread of a forest insect species over time, the degree of agreement is measured between spatial patterns of spread simulated by the model and a geospatial dataset containing observed real-world spatial patterns of spread of the forest insect for the same time period. However, ABM approaches are concerned with the discrete representation of local processes, dynamics, and generative mechanisms from which complex, system level behaviour emerges and as such evaluating patterns at the system level does little to indicate the validity of the model's structural realism i.e. the degree to which a model is capable of correctly representing the processes and dynamics driving the emergence of these patterns (Grimm & Berger, 2016). Furthermore, many ABMs incorporate stochastic processes, feedbacks, adaptation, and bifurcation into the model design, and thus will likely produce several different, yet plausible representations of the system. An ABM that closely matches real-world data may instead indicate over-fitting and is not sufficient in evaluating how well the system processes are represented (Brown et al., 2005).

In addition to approaches that test the validity of the ABMs emergent system level patterns, the development of ABM validation approaches for the evaluation of

internal model mechanisms are needed. This will ensure that processes represented in the model that generate emergent behaviour are represented correctly (Brown et al. 2005). However, this can be challenging since many ABMs are complex and heavily parameterized with a large numbers of heterogeneous autonomous agents and stochastic processes, making internal model processes difficult to understand and identify (Manson, 2007; Topping et al. 2010; Grimm and Railsback, 2012; Broeke et al. 2016). As such, the development of a framework to generate measures that describe, characterize, and help understand and identify the internal model processes that give rise to emergent system-level behaviour in ABMs may be useful.

In the same way that real world systems can be conceptualized using complex systems theory and represented using ABMs, they can also be conceptualized using network science and represented as networks (O'Sullivan, 2014). Network science is a field of science rooted in mathematics and physics that seeks to represent real world systems as a set of nodes and links and uses statistic-based network measures to both mathematically characterize the system's structure and link the system's structure and processes (Barthelemy, 2018). Network measures can also be used quantify the ABMs that aim to represent these systems by abstracting an ABM into a set of nodes representing agents and the interactions between agents as a discrete and visible set of links (Anderson & Dragicevic, 2018). Using network science measures to mathematically characterize ABM behaviour can help to identify and link internal model processes to system-level patterns. The identification and mathematical characterization of internal model processes may serve as quantitative evidence that the model is capable of reproducing the real-world system's important generative mechanisms realistically.

Therefore, the main objective of this study is to leverage spatial network analysis for the development of a robust NEtworks for Agent-based model Testing (NEAT) approach. The NEAT approach is implemented on an epidemiological network-based ABM (Epi-N-ABM) as a case study. Specifically, the simulated structure and dynamics of spatial social networks from which influenza propagates are measured using graph theory and compared against observed empirical regularities of real spatial social networks.

6.3. Overview of Approaches for Model Validation

The term “validation” is criticized as ambiguous and inconsistent (Crooks et al., 2008). Dubbed a “catch-all” term, “validation” has many meanings, and only adding to the confusion, has a number of variations. Therefore, it is important to clarify the definitions. In this study, the term validation refers to a variety of model evaluation procedures used to determine whether the model mimics the real world (Rykel, 1996; Perez et al. 2013; Liu, 2011). Traditionally, model validation can be measured using the confusion matrix approach (Congalton, 1991), kappa statistics (Cohen, 1960), fuzzy kappa statistics (Hagen 2003), a measure of relative operating characteristic (ROC) (Pontius, 2000; Fawcett, 2006), pattern based (White, 2006), and other map comparisons methods (Visser & de Nijs, 2006; Pontius et al., 2011). These approaches seek a statistically quantified agreement, often on a raster cell-by-cell basis, between independent datasets and the output of the model. Kocabas & Dragicevic (2009) use Bayesian Networks for a validation approach to compare agent-based modeling outcomes for vector-based map features. These types of model validation methods are important and used for their ability to evaluate emergent large-scale patterns.

However, using validation approaches such as these typically provides validation at a single scale and at a single point in time and this is not sufficient for evaluation of the internal parameters and processes that are driving the behaviour of the model. The correct output of system-level patterns generated by an ABM could be a result of a combination of processes and input parameters that may not be represented correctly (Augusiak et al., 2014). Furthermore, real-world phenomena and the complex systems models such as ABMs that represent them incorporate randomness, bifurcation, system feedbacks, and non-linearity. Thus, as a result of the inclusion of these factions, many ABMs produce multiple realizations or “paths”. Brown et al. (2005) demonstrates this problem by using a random output from a developed model as reference data, meaning a perfect model, but finds that when comparing future outputs with the “reference data”, the accuracies are very low. Thus, the level of agreement between system-level patterns generated by the model and those found in reality may be not be the best measure to determine the validity of complex systems models such as ABMs, and may instead promote model over fit and models that are overly specific for a certain study area.

A more recent validation approach, pattern oriented modelling (POM), compares model outcomes with independent data and patterns at multiple scales that were not used or not even known during the model design and development process (Augusiak et al., 2014; Grimm et al., 1996; Grimm & Railsback, 2012). In POM, a pattern is defined as a non-random structure found in nature and can be characterized as being strong or weak. As the number of simulated patterns emerge that match patterns observed in the real world increase, confidence in the model structure and generation of emergent properties also increases. Strong patterns are model outputs that match statistics or datasets and are quantitative in nature and are desirable. However, it may be difficult to validate with strong patterns since they may be too restrictive to account for stochasticity and uncertainty in complex systems models (Grimm & Berger, 2016). Weak patterns are qualitative and in isolation say little about the underlying structure of the model since they can be reproduced by various combinations of right and wrong generative mechanisms. However, the more weak patterns that can be reproduced in the model, the more indication there is that the underlying structure in the model is correct. Each pattern in POM can be used as a filter for which to base the grounds of rejection or acceptance of the model (Grimm & Railsback, 2012). To make matters difficult, the large number of parameters and the inclusion of stochastic processes means that ABMs are often perceived as black box models as it can be difficult to identify and understand internal processes that are generating model outputs from which to identify spatial patterns.

Additional theoretical and practical validation methods are needed that can quantitatively evaluate whether the model has correctly represented internal model processes. The integration of network theory and ABM has the potential to meet this need by discretely representing internal model processes and emergent patterns as dynamic spatial networks that are measurable using graph theory. Therefore, the NEAT approach has been developed as a way to identify discrete quantifiable patterns and processes simulated by the ABM for which to compare against empirical observations.

6.4. NEtworks for Agent-based Model Testing (NEAT) Approach

The NEAT approach is composed of three stages: 1) the integration of network theory and ABM to form a network-based ABM (N-ABM) to represent a spatial system's

structure and dynamics as simulated by the ABM as dynamic networks, based on this N-ABM representation 2) the characterization of the simulated spatio-temporal structure of the simulated networks, and 3) the comparison of the network measures obtained from the simulated network structures to network measures observed in the real-world for the same phenomena. Each network measure operates as a mathematical filter, for which to base the grounds of acceptance or rejection of the developed model. Therefore, the NEAT approach is well-situated under the larger pattern-oriented model (POM) evaluation framework.

An N-ABM is a computational representation of a complex spatio-temporal system as an evolving and dynamic spatial network *SN* (Anderson & Dragicevic, 2018). An N-ABM is composed of an ABM component to simulate spatio-temporal dynamics of a phenomena (eg. land use, mobility, or ecological dynamics) and a spatial network component, which abstracts the ABM as an evolving spatial network *SN*. Specifically, the spatial network *SN* mirrors the simulated dynamics of the developed ABM and stores both topological and spatial information that can be used for visualization and network analysis of ABM internal processes and emergent structure.

The spatial network *SN* consists of an evolving set of nodes *N* and links *L* and can be formulated as:

$$SN = [N, L] \quad (1)$$

where *SN* is the spatial network, *N* is the set of nodes and *L* is the set of links. The spatial network *SN* evolves as a function of the system dynamics simulated by the ABM resulting in the addition, removal, and rewiring or reconnection of nodes and links. Therefore, the evolving spatial network *SN* at each time step of the simulation can be formulated as SN_{t0}, SN_{t2}, SN_{t3}… SN_{tn}.

The abstraction of agents and their interactions as nodes *N* and links *L* is based on the purpose of the ABM. For example, an N-ABM representing the mobility of people in a city may abstract the simulated dynamics as a spatial network model composed of a set of nodes *N* that represent various locations in the city (i.e. stores, schools, or homes). Location nodes are connected by a set of links *L* representing the movement of individuals as agents as they move from location node v_i to new location node v_j. As the agents move to new location nodes, the new node is added to the spatial network

model. However, ABM processes may be abstracted to form several different spatial networks *SN* representations of different model dynamics. Using the same mobility example, a different set of nodes *N* may be generated to represent the agents themselves where the set of links *L* represent relationships between the nodes, forming a social network. The decision of what measures to include is a function of both the purpose of the model and the phenomena being represented.

The evolving spatial network *SN* generated by the N-ABM is measureable. Nodes *N* and links *L* are programmed to calculate and store their local network properties specific to the individual node or link as the network evolves over time. Each node v_i in the spatial network *SN* keeps a record of its connections with other nodes v_j and thus facilitate the calculation of their local network properties from which global network properties can be calculated. Each link *e* in the network also have several network properties including direction and weight. Link direction may be unidirectional, where flow or individuals, materials, information, or relationships takes place only *from* link e_i *to* link e_j or bidirectional, where the connection goes both ways. The volume of individuals, materials, information etc. is often referred to as the link weight and this information is stored in the links. Obtained network measures from the N-ABM can be compared against empirical regularities observed in the real world in order to validate that the internal model processes are being represented correctly.

6.5. Applying the NEAT Approach to an Epidemiological Agent-Based Model (Epi-N-ABM)

The proposed NEAT approach is demonstrated using a spatially-explicit epidemiological network-based ABM (Epi-N-ABM). The following section provides an overview of epidemiological modelling, the development of the Epi-N-ABM, and finally the model testing of the Epi-N-ABM model using the NEAT approach. It was not the objective that the Epi-N-ABM be a perfect representation of the real spatio-temporal patterns of disease transmission in part of the City of Vancouver, but rather to use the NEAT approach to critically evaluate the underlying processes in the Epi-N-ABM.

6.5.1. Modelling Epidemics

Traditional population-based epidemiological models such as the Susceptible-Infected (SI), Susceptible-Infected-Susceptible (SIS), and the Susceptible-Infected-Recovered (SIR) approaches operate under four basic simplistic assumptions that treat individuals as homogeneous, interactions as global, the spatial distribution of individuals as uniform, and interactions between individuals as equal, also known as homogeneous mixing (Bian et al., 2004). These traditional epidemiological models face criticism over their limited ability to represent the spread of infectious disease realistically, which would require the representation of interactions between heterogeneous individuals within the larger population and the inclusion of how these interactions change over time and geographic space. Therefore, in order to improve epidemiological modelling and overcome the limitations of traditional epidemiological models, ABM approaches have been used (Perez & Dragicevic, 2009; Rakowski et al., 2010; Frias-Martinez et al., 2011; Tian et al., 2013; Crooks & Hailegiorgis, 2014). ABM approaches are able to capture the spatio-temporal complexity in disease transmission to better represent the propagation of infectious diseases in populations realistically and in addition can easily be used to generate scenarios to answer questions such as how the size of an initial infected population impacts transmission dynamics, evaluate the effectiveness of prevention methods, and to project the number of cases for which to assign medical resources. However, for several reasons some of which are based on privacy concerns, data required for ABM creation and model testing is not typically available, especially at fine spatial and temporal resolutions. As a result, epidemiological ABMS are typically tested only using parameters collected at the system level such as the length of time of the infection and the number of cases or deaths and thus are unable to validate internal model processes (Skvortsov et al., 2011; Rakowski et al., 2010; Carpender & Sattenspiel, 2009) or alternatively, are unable to be tested at all (Frias-Martinez et al., 2011; Bian & Liebner, 2007).

6.5.2. Epi-N-ABM Simulating Influenza Dynamics using Agents

Both the Epi-N-ABM and the NEAT approach were designed and programmed using a Java based programming language and were executed using Repast Simphony 2.6 (2018). This section provides a detailed overview of the Epi-N-ABM using the

overview, design, and details (ODD) protocol (Grimm et al., 2006; 2010) and the application of the NEAT approach.

Model Overview

Purpose

The purpose of the Epi-N-ABM is twofold: (1) to develop a theoretical N-ABM representing the transmission of the influenza virus across a spatio-temporal social network in a part of the City of Vancouver, Canada and 2) to evaluate the Epi-N-ABM using the proposed NEAT approach.

Entities and State Variables

The Epi-N-ABM is composed of (1) adult agents, (2) child agents, and (3) place agents including homes, schools, and businesses. Each business plays the role as both a workplace that some agents travel to daily for work and as a location that agents visit randomly such as a grocery store, public transportation, or a restaurant. For example, one agent may treat a grocery store business as a workplace, while another treats the same grocery store as a random location. Both adult agents and child agents have the sub-class of being either a healthy agent or an infected agent. The description of the associated variables of the adult, child, and place agents are described in detail in Table 6.1.

Table 6.1. The Epi-N-ABM agent classes, sub-classes, and associated variables.

Adult and Child Agent Variables	Variable Description
ID	ID number of the agent
Location	Agent location in geographic space
List of physical interactions	A list of all the agents *y* that an agent v_i has interacted with
List of successful transmissions of infection	A list of all the agents *z* that agent v_i has infected
Local spatial social network measures	Network measures specific to agent v_i's spatial social network
Local infected network measures	Network measures specific to agent v_i's infection network
Influenza vaccination	Whether the agent has been vaccinated
Connection probability	The probability of making a connection based on the agents location
Infection probability	The probability of transmission from agent v_i to agent v_j based on the agents location
If infected, number of hours infected	The number of hours infected determines whether the infection is latent, infectious, or recovered

Period of latency	A random value selected between 1 day and 4 days
Period of contagiousness	A random value selected between the time of infection latency + 1 and 8 days
Place Agent Variables	Variable Description
ID	ID number of the place
Location	The location of the place in geographic space
Number of agents visiting	The number of agents currently at the place

Process Overview

The developed Epi-N-ABM simulation represents the transmission of a seasonal influenza virus across a spatial social network in a part of the City of Vancouver, Canada at a very fine spatial and temporal scale. The behavior of the adult and child agents are represented using a collection of subroutines that include the following: 1) travelling from home to work or school, 2) travelling to a random location, and 3) returning home. The execution of the agent subroutines is a function of the time of day in the model where each day consists of 24 hours. Each hour is represented by one model time step. One completed simulation run is composed of 270 days for a total of 6480 time steps from the initial state of hour 0 (t_0) to hour 6480 (t_{6480}).

In each of the agent subroutines, agents interact with other agents. Interactions between two agents takes place in both geographic space and time. Since influenza is airborne and thus can be transmitted over moderate distances, an interaction between two individuals, capable of transmitting influenza, does not require direct physical contact (World Health Organization, 2018). During the time of interaction in the model, if one of the agents is infected with influenza, the transmission of influenza from one infected agent to a non-infected agent may occur.

The Epi-N-ABM model structure is presented in Figure 6.1. The Epi-N-ABM is initialized at t_0. Upon model initialization, families are assigned to the home agents. Family sizes are random, where each family consists of one to two healthy adult agents and zero to two healthy child agents for an average family size of 2.5. This is consistent with the average family size in the City of Vancouver as reported by Statistics Canada (2016).

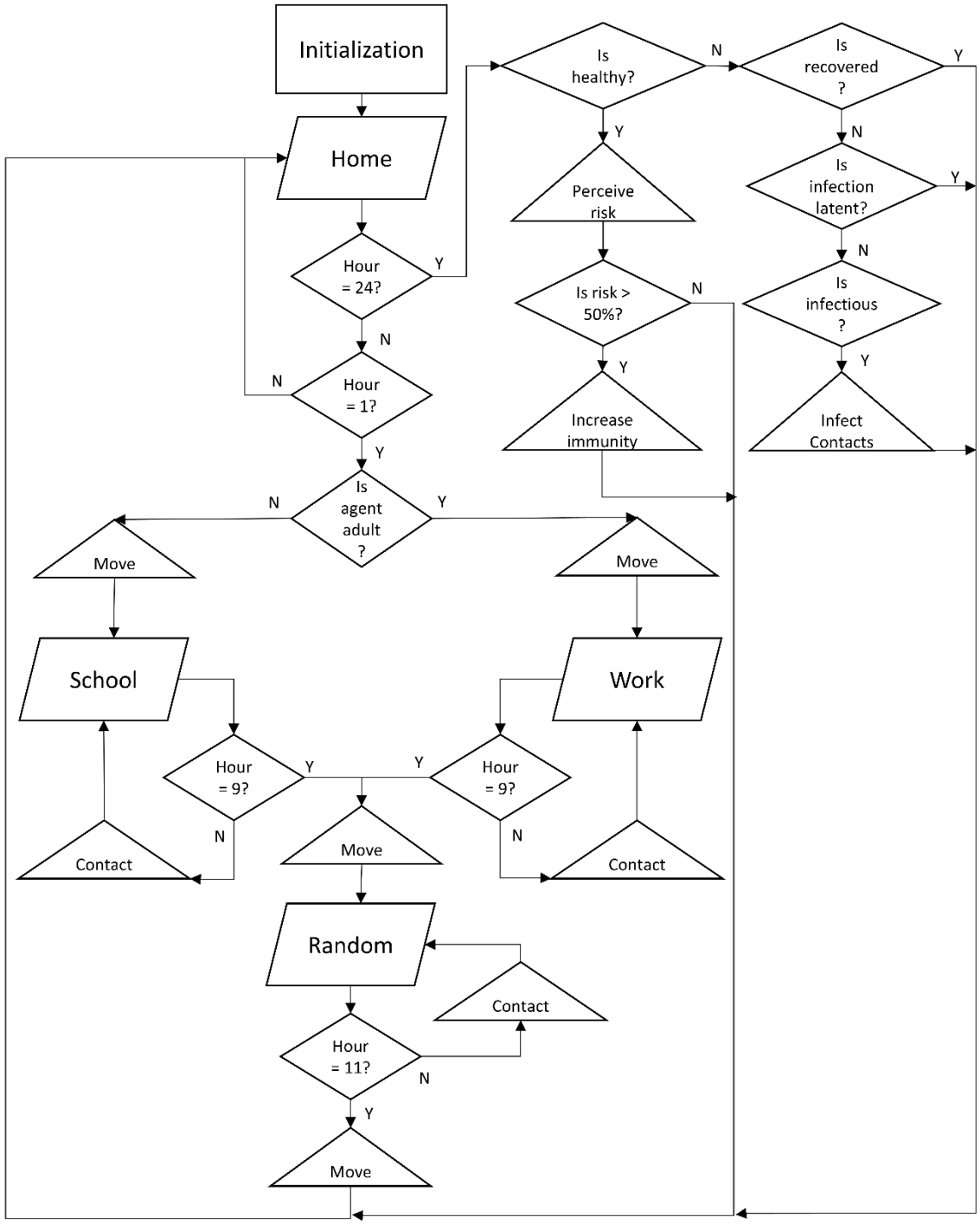

Figure 6.1. The Epi-N-ABM model structure.

Based on whether the agent is a child or an adult, each member of the family are assigned to a workplace or a school. During this process, there is a 17.2% likelihood that adult agents will find a workplace in a random location throughout the entire study area,

where distance is not a factor (Statistics Canada, 2016). Moreover, there is an 82.8% likelihood that adult agents will find a workplace in a random location that is within 21 km of their home. In this process, agents have a higher likelihood of finding a workplace that is close in distance to their home. Child agents attend the school that is the closest in distance to their homes. Once child agents have been assigned a school, they are assigned at random to a class within their school. Child agents only interact with other child agents that are in the same classroom. During initialization, 35% of the population are immunized with a vaccine to protect them against infection (Public Health Agency of Canada, 2018).

In addition, thirty healthy agents are selected at random and become infected agents. Infected agents begin with a period of latency, where their infection can not be transmitted to another agent. The number of days that an agent's infection is latent is randomly selected between one and four days (World Health Organization 2018). Following the latency period, the agent's infection becomes contagious and infected agents begin to infect other agents. The number of days that an agent's infection is contagious is randomly selected from the end of the latency period up to seven days (World Health Organization, 2018). Following the contagious period, the infected agent recovers and becomes a healthy agent that is no longer susceptible to the infection.

Each agent keeps a record of their contacts over a period of 24 hours. Within that 24 hour period, there is a 100% likelihood that each family member interacts with each member of their family. At t_1, all agents travel to work or to school. Agents remain at work or school for 8 hours, until t_8. For each hour that the agents are at work or school, each agent at the same workplace or classroom has a 5% likelihood of coming into direct contact. These interaction likelihood values were selected as a result of sensitivity analysis. At t_9, all agents travel to a random location that is also a business. This random location is intended to account for random direct interactions between agents at locations that are not the home or the workplace. Agents remain at the random location for 2 hours until t_{10}. For each hour that the agents are in a random location, each agent at the same random location have a 1% likelihood of coming into direct contact. At t_{11}, all agents travel to their respective homes.

At t_{23}, infected agents who are contagious access the record of agents that they have interacted with each other the past 24 hours. If the agent in the record of contacts

is not already recovered from the influenza or is not already infected, the agent is sorted into a new list. Next, based on the number of times an agent has come into contact with an infected agent, the infected agent's infection attacks the immune system of an agent. There is a 2.5% likelihood that infected agents will successfully infect another agent, determined during sensitivity testing. Healthy adult agents have the ability to perceive that people are ill in their environment and can increase their defense against infection by 10%. In addition, healthy agents with a vaccination have an increased defence against infection by 40% (Center for Disease Control and Prevention, 2018).

Successful transmission of the infection results in healthy agents becoming newly infected agents. Agents move through various stages where the infection is latent, contagious, and recovered. The record of contacts is cleared and all agents execute their subroutines for the next 24 hours continue as described. At t_{6480}, all agents are recovered and the simulation is complete. The Epi-N-ABM is simple, however, there are many model parameters including family size, average commute distance, percent that commute within the commute distance, the length of period of infection latency, the length of period of infection contagiousness, likelihood of direct contact β_C, and likelihood of transmission β_T. Based on the notion that several of these parameters are random, the model is executed 30 times from t_0 to t_{6480} to determine the way in which the simulation patterns converge. Since the agents exist in geographic space, the location of infected agents over time can be mapped.

Design Concepts

The Epi-N-ABM design concepts as outlined in the ODD protocol include interaction, emergence, fitness, adaptation, sensing, stochasticity, and observation. These are described in detail in Table 6.2.

Table 6.2. Epi-N-ABM design concepts as adapted from the ODD protocol.

Design Concept	Description
Interaction	Two types of interactions are incorporated into the model. These interactions take place between two healthy agents or between a healthy agent and an infected agent. These interactions can be further classified by familial interactions, interactions between coworkers and classmates, and random interactions between pairs of agents. Interactions occur in space and time and have the potential for transmission of infection. Since influenza can be transmitted over moderate distances (World Health Organization, 2018), interaction does not need to be physical.

Design Concept	Description
Emergence	From local interactions between healthy-healthy agents and healthy-infected agents, both the spatio-temporal dynamics of a spatio-temporal social network and the spatio-temporal dynamics of the spread of infection emerge.
Fitness	Adult agents have the goal to stay healthy.
Adaptation	Healthy adult agents adapt their ability to fight infection based on the perceived risk of infection in their immediate environment.
Sensing	Agent knowledge is limited to their local environment. They know where they are, what time it is, and they have the ability to sense that people around them are infected with influenza.
Prediction	Adult agents use their knowledge in order to adapt in order to meet their goal to stay healthy.
Stochasticity	There are several stochastic elements incorporated into the model. These elements include family size, the selection of work, school, and random places to visit, the selection of initial infected agents, the length of period of latency, the length of period of contagiousness, the interaction likelihood for each location, and the infection likelihood. Most stochastic elements are parameterized using a random value obtained from a threshold based on real thresholds identified within the literature.
Observation	The data that is collected from the model for analysis includes the number and ID of the contacts for each agent, the number and ID of agents that each agent infects, the location throughout time of healthy and infected agents, and the total number of agents infected over time.

Details

Initialization

The number of agents that are generated upon model initialization is random, however on average there are 5681.8 agents. Each agent is assigned an ID, their home and work or school agent, their location, and the parametrization for period of latency and period of contagiousness in the case that the agent becomes infected. Thirty of the spawned agents are selected at random to become infected. Since the number of agents that spawn and the selection of agents to become infected are random, each initialization is slightly different. Moreover, a total of 35% of the population is randomly selected as being vaccinated as estimated by the Public Health Agency of Canada (2018).

Input Data

The location of homes and businesses are created at random from a land use geospatial dataset representing a neighbourhood in downtown Vancouver. Homes are created at random within the zoning boundaries for residential land use and businesses

are created at random within the zoning boundaries for commercial and industrial land use in downtown Vancouver.

Subroutines

The Epi-N-ABM operates using a collection of subroutines including tracking the time of day, moving between place agents, making contacts, assessing risk, tracking number of days infected, becoming infected and the stages of infection, infecting contacts, and recovering. These subroutines are described in detail in Table 6.3.

Table 6.3. The Epi-N-ABM subroutines and their description.

Subroutine	Description
Tracking Time of Day	The time of day begins with hour 1. With each time step the time of day increases by 1. When the time of day at time *t* is equal to hour 24, the time of day at time *t+1* is reset back to hour 1.
Moving Between Place Agents	At hour 1, all agents move from their home agent to their work or school agent. At hour 9, all agents move from their work or school agent to a random business agent. At hour 11, all agents move from the random business agent back to their home agent.
Making Contacts	At each location, agent v_i records a list of the agents v_j that are in the same location. For each iteration, a contact likelihood β_C is applied that determines whether agent v_i interacts with each agent v_j in the list. The likelihood value is specific to the location of the agent and differs based on whether the agent is at home, work, school, or in a random location. The agents that are within the likelihood range are added to agent v_i's list of contacts. Interaction is bi-directional, meaning that once agent v_j is added to agent v_i's list of contacts, agent v_i is immediately added to agent v_j's contacts.
Assessing Risk	At hour 24, healthy adult agents assess the risk of their contacts. If more than 50% of their contacts are infected, the healthy agent can increase their immunity by 5%.
Becoming Infected	At hour 24, healthy agents are defended by their immunity or a likelihood of transmission β_T. This likelihood is based on whether the agents have been vaccinated.
Infecting Contacts	At hour 24, infected agents attack the healthy agents that are not already infected and have not recovered from a previous infection.
Recovering	Once the length of period of contagiousness specific to agent v_i has been exceeded, the infected agent recovers and becomes a healthy agent that is no longer susceptible to infection.

6.5.3. Epi-N-ABM: Simulating Influenza Dynamics as Networks

In the developed Epi-N-ABM, agent interactions drive the structure of two dynamic spatial networks including the following: an underlying spatio-temporal social network SN_{social} that evolves over a period of 24 simulated hours and a second network, a spatio-temporal infection network $SN_{infected}$ that evolves over 270 days. Both SN_{social} and $SN_{infected}$ are composed of a set of nodes *N* and a set of links *L*. In the network SN_{social}, the underlying spatio-temporal social network, both healthy and infected agents make up the set of nodes *N* in the network. A link *e* forms between agent node v_i and agent node v_j if they interact with each other. All links *L* are undirected, meaning interactions are always recorded by both agents. In $SN_{infected}$, the spatio-temporal infection network, only infected agents form the network nodes *N*. A link forms between agent node v_i and agent node v_j as agent v_i infects agent v_j. This link is directed to account for the notion that the transmission of influzena is directional whereby infection occurs *from* agent v_i *to* agent v_j , but not vice versa.

6.5.4. Epi-N-ABM Testing Using Networks

Sensitivity Analysis and Calibration

The sensitivity analysis and calibration of Epi-N-ABM were performed using network measures that were not used in the validation process. Sensitivity analysis was performed in order to understand the sensitivity of the social network SN_{social} to the interaction likelihood values β_I and the sensitivity of the infection network $SN_{infected}$ to the transmission likelihood values β_T. Sensitivity analysis in this study examines how a range of interaction likelihoods β_I and transmission likelihoods β_T affect the emergence of *average degree* <*k*>, as measured using graph theory. Degree *k* is the number of connections a single node has to other nodes. Average degree <*k*> is a global network measure that calculates the average degree for all nodes in the network.

Social Network

The sensitivity of the social network SN_{social} to the interaction likelihood β_I is evaluated using the average degree of contact $<k_{contact}>$ across all individuals in the network. Degree of contact $k_{contact}$ for an agent in the social network SN_{social} is defined as the number of interactions the agent acquires over a period of 24 hours. Based on the

literature, the expected and observed average degree of contact for an individual is 13.4 interactions per day (Mossorey et al. 2008). It is expected that the average number of interactions will be lower than the average number of interactions for children (Leung et al. 2017).

Infection Network

The sensitivity of the infection network to the transmission likelihood β_T is initially evaluated using the average degree of infection $<k_{infect}>$ of the infection network. Degree of infection k_{infect} in the infection network SN_{infected} measures the number of agents an infected agent transmits the infection to. The transmission of an infectious disease from an infected individual to other individuals is referred to as the reproduction number *R0.* The *R0* of influenza is both uncertain and dynamic as it changes as the infection runs its course in a population (Biggerstaff et al. 2014). Specifically, a value greater than 1 indicates that the infection will persist and a value less than 1 indicates that the infection will decline in the population (Biggerstaff et al. 2014). Coburn et al. (2009) estimates that the *R0* for seasonal influenza ranges between 0.9 and 2.1. The sensitivity of the infection network size $SN_{infected}$ to the transmission likelihood β_T is also tested. It is observed that 5 to 20% of the population will become infected (Biggerstaff et al. 2014).

Sensitivity Analysis and Calibration Results

Social Network

The interaction likelihood β_I values differ based on location and are a driving factor as to whether an agent interacts with other agents that are in the same location as them i.e. the same household, the same workplace or school, and the same random location (Table 6.4). It is assumed that agents will interact with all of their family members. Therefore, the likelihood β_I of interacting with family members is 100%. Interactions between individuals at work or at school implement the same interaction likelihood β_I since the two locations are similar in their nature. Agents travel to the same workplace or school each day and thus consistently interact with the same set of agents. The interaction likelihood β_I at random location is slightly lower for random locations as random interaction tends to be fleeting (Mossorey et al. 2008).

Table 6.4. The average degree of contact <$k_{contact}$> for the social network obtained using different combinations of interaction likelihoods β_I for each location.

Home β_I	Workplace β_I	School β_I	Random β_I	Average Degree of Contact <$k_{contact}$>
100%	4%	4%	1%	11.93
100%	**5%**	**5%**	**1%**	**13.493**
100%	6%	6%	1%	14.68
100%	4%	4%	2%	12.55
100%	5%	5%	2%	14.00
100%	6%	6%	2%	15.87

The interaction likelihood β_I that generated the closest fit to the expected average degree of contact <$k_{contact}$> of 13.4 individuals in 24 hours is a 100% likelihood of interacting with family members in the household, a 5% likelihood of interacting with coworkers each hour that an agent is at work, a 5% likelihood of interacting with classmates each hour that the agent is at school, and a 1% likelihood of interacting with a random agent each hour that the agent is in a random location.

Infection Network

Once all agents recover, the average degree of infection <k_{infect}>, also known as the *R0,* is calculated. This calculation includes the agents that became infected, but did not pass on the infection to another agent, meaning $k_{infect} = 0$. Table 6.5 shows the average degree of infection <k_{infect}> obtained as a function of a variety of transmission likelihoods β_T. Notably, all model runs fail to propagate using a transmission likelihood value β_T of <=1%.

Table 6.5. The average degree of infection, average degree of infection if k>0, and the average number of agent's that are infected for the infection network obtained using different combinations of transmission likelihoods β_T.

%	Average degree of infection <k_{infect}>	Average degree of infection <k_{infect}> where k_{infect} >0	Average number of agents infected
1	NA	NA	NA
2	0.57	1.36	71.85
2.5	**0.965**	**1.68**	**669.13**
3	0.98	1.73	1585.5

Parameterizing the infection network $SN_{infected}$ with a transmission likelihood value $\beta_T > 2\%$ produces an average degree of infection $<k_{infect}>$ of 0.9 and thus the infection network does not appear to be sensitive to variations in the transmission value β_T. However, it should be noted that the average degree of infection $<k_{infect}>$ is not static and changes with each iteration as influenza spreads. As the proportion of agents that are infecting new agents is no longer great enough to maintain a continued spread of infection, the average degree of infection $<k_{infect}>$ falls below 1, the infection declines, and all the agents begin to recover. As such the average degree of infection $<k_{infect}>$ may not be a reliable parameter to test and calibrate the model. Therefore, to determine whether the *R0* falls within the expected range as identified in the literature, the average degree of infection is calculated to include only agents who successfully transmitted the infection to another agent and thus have a degree k_{infect} greater than 0, formalized as $<k_{infect}>_{k>0}$. It was found that all transmission likelihoods produce an acceptable average degree $<k_{infect}>_{k>0}$ and an average $R0_{R>0}$ that falls within 0.9 and 2.1.

Because of the challenges faced in using the average degree of infection $<k_{infect}>$ as a measure for model sensitivity and calibration, a second emergent property, specifically the total number of influenza cases, is used to determine the appropriate transmission likelihood value β_T. It was found that a transmission likelihood value β_T of 2% produced far too few cases and a transmission likelihood value β_T of 3% produced far too many cases of influenza, where the percent of the agent population simulated as infected fell outside of the expected range of 5-20% (Biggerstaff et al., 2014).

Based on the sensitivity analysis results, the model is calibrated using the interaction likelihoods β_I and transmission likelihoods β_T that generated the best fit for expected average degree of contact $<k_{contact}>$ and the average degree of infection $<k_{infect}>$ for both the social network and the infection network respectively. Specifically, the interaction likelihood β_I that generated the closest fit to the expected average degree of contact $<k_{contact}>$ is a 100% likelihood of interacting with family members in the household, a 5% likelihood of interacting with coworkers each hour that an agent is at work, a 5% likelihood of interacting with classmates each hour that the agent is at school, and a 1% likelihood of interacting with a random agent each hour that the agent is in a random location. Likewise, the transmission likelihood β_T that was selected was 2.5%. This aligns with the idea that infection networks are able to persist even with a

very small transmission likelihood β_T, also known as the vanishing epidemic threshold (Barabasi 2016).

Validation

The objective of this study is to compare simulated network properties with observable empirical regularities in real world networks. Therefore, graph theory is applied to collect several network measures that characterize the simulated spatial social network SN_{social} and the simulated spatial infection network $SN_{infected}$.

Social Network

According to the literature, real-world large social networks exhibit a set of observable empirical regularities with respect to their network structure and behavior and these regularities hold, if not are enhanced, in spatial social networks (Alizadeh et al., 2017). These regularities summarized by Alizadeh et al. (2017) are outlined in Table 6a. Table 6a also lists the corresponding graph theory measures that are used to measure these regularities and that can be used as a filter to reject or accept that the simulated network contains the internal processes that are realistic enough to maintain these regularities. Therefore, to validate the social network SN_{social} the following network properties are calculated:

Table 6.6. Observed empirical regularities of (a) social networks and (b) infection networks.

a) Characteristic of Social Network	Network Property
Low network density (Wong et al., 2006)	Density *D*
The number of connections each individual has should not exceed a reasonable limit (Gilbert, 2006; Barthelemy, 2003)	Maximum degree of contact $<k_{contact}>_{max}$
The number of connections each individual has should be heterogeneous (Fisher, 1982)	Standard deviation of degree of contact $<k_{contact}>_{\sigma}$
The majority of individuals in the network have relatively small degrees and a small number of actors may have very large degrees (Wong et al., 2006; Fischer, 1982)	Degree distribution *P(k)*
Individuals with a high number of connections tend to be connected with other individuals with many connections (Bruggeman, 2008; Newman, 2002)	Assortativity *A*
A fraction of each individual's contacts should also interact with each other (Bruggeman, 2008)	Average clustering coefficient <*C*>
Presence of communities where there should be some groups in which members are highly connected with their in-group and loosely connected between groups (Newman, 2004)	Modularity *M*

a) Characteristic of Social Network	Network Property
Individuals in the network should be able to reach other via a small number of connections (Milgram, 1967)	Average shortest path length <*l*>
b) Characteristic of Infection Network	Network Property
The degree distribution should be heavily skewed, potentially exhibiting a power-law (Liljeros et al., 2001)	Degree distribution *P(k)*

Real world networks that exhibit these properties are referred to as small world networks (Watts and Strogatz, 1998). Whether the simulated social network SN_{social} is indeed a small world network is tested by statistically comparing the graph theory measures obtained from the simulated social network SN_{social} with the same graph theory measures from a simulated random network SN_{random} with the same number of nodes *N* and links *L*. Specifically, the graph theory measures used for this test include the degree distribution *P(k)*, the average clustering coefficient <C>, and the average shortest path length <*l*>. Both a small-world and random network will form a degree distribution *P(k)* with a *Poisson* curve, however in the case that the SN_{social} is indeed a small world network, $\langle l \rangle_{SNsocial} \geq \langle l \rangle_{SNrandom}$ and $\langle C \rangle_{SNsocial} \gg \langle C \rangle_{SNrandom}$ (Humphries and Gurney 2008).

Infection Network

Infection networks are a product of the social networks that they propagate on, therefore, validation of the social network SN_{social} is of the main focus. However, real infection networks on their own also exhibit observable network characteristics (Figure 6.6b). In real infection networks, the network's degree distribution *P(k)* is often heavily skewed, and occasionally follows a power law (Liljeros et al., 2001). Specifically, many individuals in the network will have a small number of transmissions and a few individuals referred to as super spreaders in the network will have a large number of transmissions (Small et al., 2007; Barabasi, 2016). It can be tested whether the simulated infection network $SN_{infected}$ exhibits these characteristics by looking at the simulated degree distribution *P(k).* The simulated degree distribution for the infection network is tested for goodness of fit using *powerlaw,* a Python package for analysis of heavy-tailed distributions (Alstott et al., 2014).

6.6. Results

6.6.1. Simulation Results

Social Network

Figure 6.2 presents an example of one full simulation representing the spatial social network SN_{social} as it evolves over 24 hours. Figure 6.2a presents the location of the homes, businesses, and the schools. Figure 6.2b presents the social network structure for the first hour of work and school at t_1. At this point, agents have only interacted with their family members. Figure 6.2c presents the social network structure at mid-day when the agents are at work and at school at t_5. Figure 6.2d presents the social network structure at the end of the day at t_8. Figure 6.2e and 6.2f present the social network structure as they visit random locations at t_9 and t_{10} respectively. Finally, Figure 6.2g presents the social network structure when the agents return home at t_{11}.

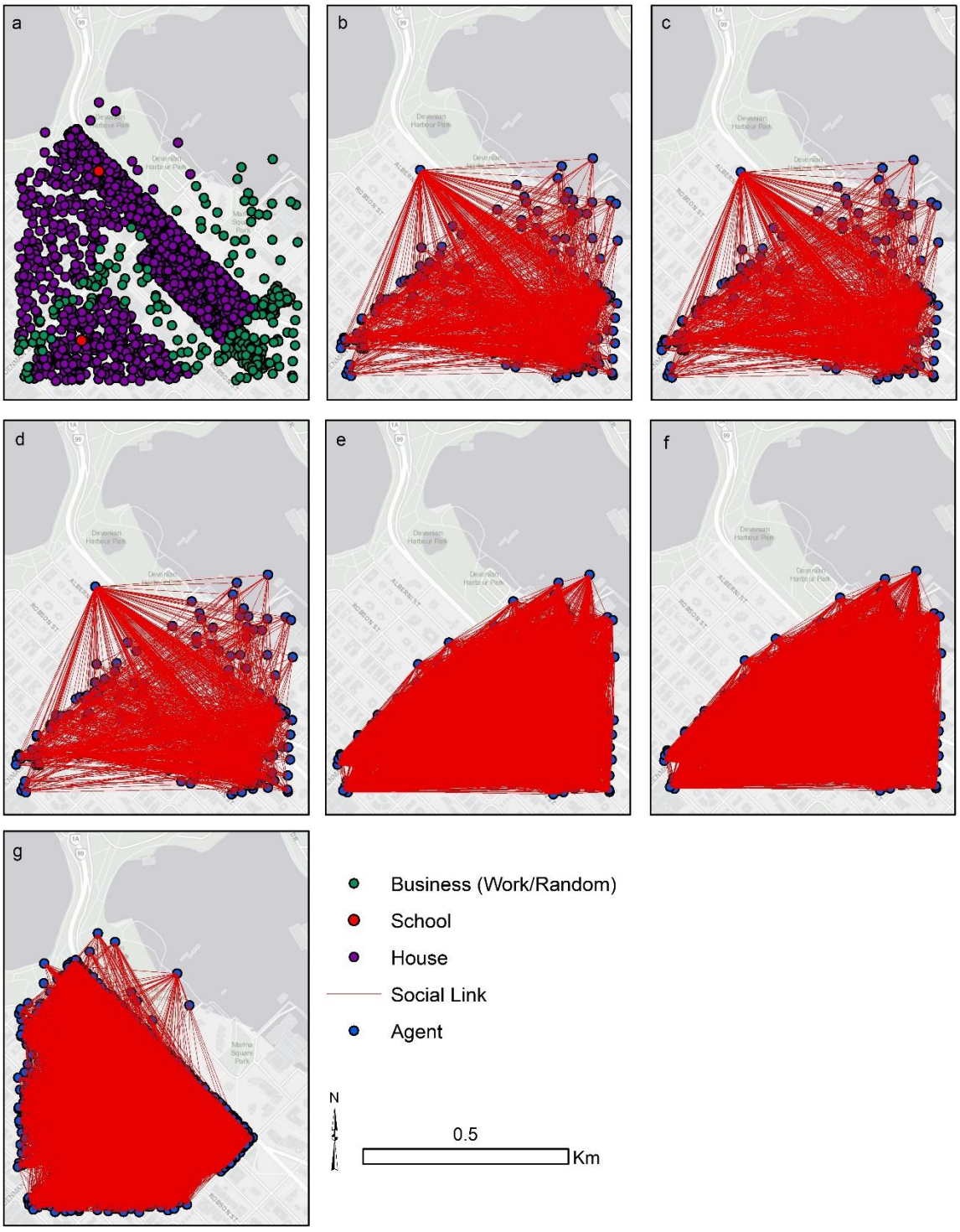

Figure 6.2. **Map of (a) locations of homes, businesses, and schools that agents travel between and the simulated social spatial network SN_{social} as it evolves over time for time steps (b)t_1, (c)t_5, (d)t_8, (e)t_9, (f) t_{10}, and (g) t_{11}.**

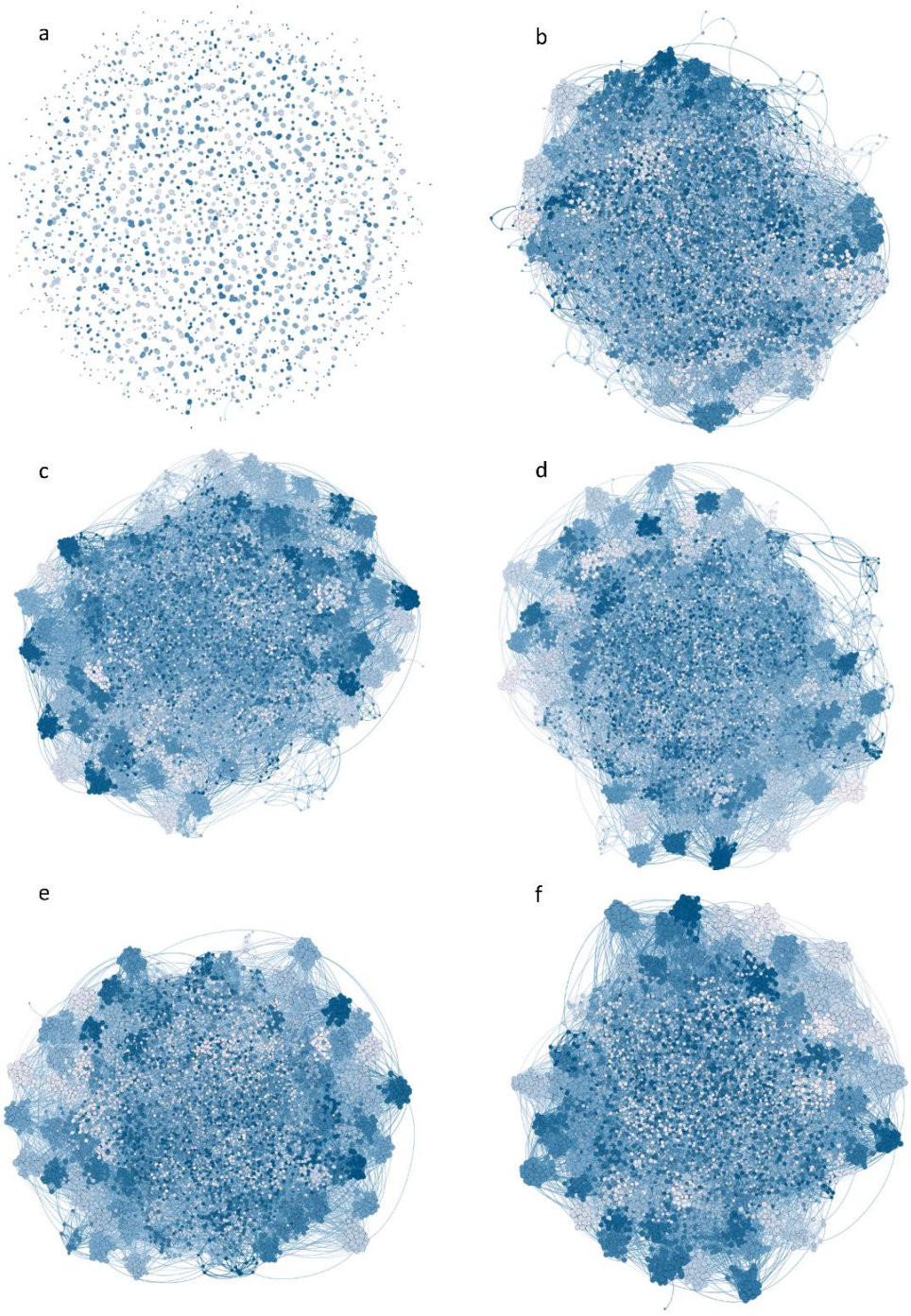

Figure 6.3. **The social network SN_{social} in a non-spatial representation for time steps t_1 (a), t_5 (b), t_8 (c), t_9 (d), t_{10} (e), and t_{11} (f).**

Large networks are particularly challenging to visualize and are typically described as “hairballs”. This term is appropriate for the characterization of large networks with a highly complex structure from which it is challenging to obtain any visual information. This problem is further challenged by representing densely connected networks in geographic space. For the purpose of improved visualization, the same example spatial social network SN_{social} in Figure 6.2 is presented in Figure 6.3 as its evolving non-spatial counterpart using the same time steps t_1 (Figure 6.3a), t_5 (Figure 6.3b), t_8 (Figure 6.3c), t_9 (Figure 6.3d), t_{10} (Figure 6.3e), and t_{11} (Figure 6.3b). The different shades of blue mark different communities or groups in which members are highly connected with their in-group and loosely connected between groups as defined in Table 6.6. These communities are a function of classrooms, workplaces, and family members. The different sizes of the nodes represent the number of contacts a node has, where larger sized nodes have more contacts and smaller sized nodes have less contacts. The non-spatial representation provides a clearer idea as to the internal processes within the model from which the network structure emerges. Table 6.7 presents the network measures obtained from this example spatio-temporal social network as it evolves over time for t_1, t_5, t_8, t_9, t_{10}, t_{11}.

Table 6.7. Spatio-temporal network measures obtained from SN_{social}

t	N	L	<C>	$<k_{contact}>$	$<k_{contact}>$ max	$<k_{contact}>$ min
1	5714	10679	1	1.94	4	0
5	5714	47789	0.24	8.37	25	0
8	5714	67149	0.37	11.75	33	0
9	5714	72455	0.41	12.67	33	0
10	5714	74984	0.38	13.11	34	0
11	5714	77453	0.36	13.34	34	1
t	$<k_{contact}>$ σ	M	C	A	D	<l>
1	0.879	0.999	1919	0.9987	0.0003	1
5	3.871	0.789	73	0.2631	0.001	5.59
8	5.29	0.844	87	0.4383	0.002	5.121
9	5.66	0.855	94	0.4781	0.002	5.026
10	5.69	0.827	93	0.4382	0.002	4.6
11	5.73	0.801	94	0.4022	0.002	4.452

At t_1, agents have come into contact with family members only. The network structure is presented in Figure 6.3a, where each family is represented by a small cluster within a set of disconnected clusters. Each family member is directly connected to each family member, producing an average clustering coefficient <*C*> of 1, a low average

degree <*k*>, a high modularity *M* made up of many tight knit communities *C,* and a very short average path length <*l*>. This also produces a high assortativity *A,* where each agent has the same degree *k* as the agents that they are currently connected to. In Figure 6.3a, larger families where $\langle k_{contact} \rangle_{max} = 4$ are located near the center of the set of clusters and smaller families and individuals with no family members where $\langle k_{contact} \rangle_{min} = 0$ are located in the surrounding area. In summary, family units form perfect network clusters.

At t_5, all agents travel to work or school and thus begin to interact with their coworkers and classmates. From this process emerges a giant network component, meaning that all agents become part of the larger network structure (Figure 6.3b). At this point, one agent remains disconnected from the giant component, where $\langle k_{contact} \rangle_{min}$, meaning the agent has no family or coworkers. The heterogeneity in the network as a function of the different family sizes becomes apparent, where the $\langle k_{contact} \rangle_{\sigma}$ and assortativity *A* decreases significantly. Since it is still early in the day and classmates and coworkers have not yet had many opportunities to interact with one another and form their own workplace and classroom communities, the average clustering coefficient <*C*>, modularity *M*, and number of communities *C* remains on the lower side. The lack of community structure is evident visually (Figure 6.3b).

Agents continue to make connections at their workplace and school until t_9. At t_{10} and t_{11} agents interact with random agents at random locations at t_{10} and t_{11}. This results in a decrease in the average clustering coefficient <*C*> since agent v_i's existing contacts consisting of family members and coworkers or classmates may never interact with the agent v_i's random connections. These random interactions are important for the overall connectivity of the social network as they act as gateways that connect different communities and reduce the overall path length of the network. As time passes, the number, size, and strength of the communities increase over time, coded with different colours in Figure 6.3 c, d, e, and f. Heterogeneity within the network increases as supported by the increase in $\langle k_{contact} \rangle_{\sigma}$ and the slight decrease in assortativity. The overall degree distribution *P(k)* is a positively skewed degree distribution that is highly peaked and is slightly bi-modal. The bi-modal nature of the degree distribution is a function of adult agents having less interactions than child agents, which corresponds with the behavior or pattern documented in the literature.

Infection Network

The infection network is technically composed only of infected agents. Links represent the transmission of influenza from one infected agent to a newly infected agent. Once an infected agent recovers, the infected agent and its associated link are removed from the infection network. This makes it challenging to both visualize and measure the network as it evolves over a long period of time, since the infected agents are only infected for a few days and then are removed from the network. Therefore, in order to present the simulation results for the infection network $SN_{infection}$, the model has been modified so that the links are not removed from the network, making it possible to keep track of the history of infection. Figure 6.4 presents one full model run representing the infection network as it grows over 270 days. The network structure is presented once the agents are returned back home and shows the link between which families infect which families over time. The infection network is denser in the highly populated location of the neighbourhood.

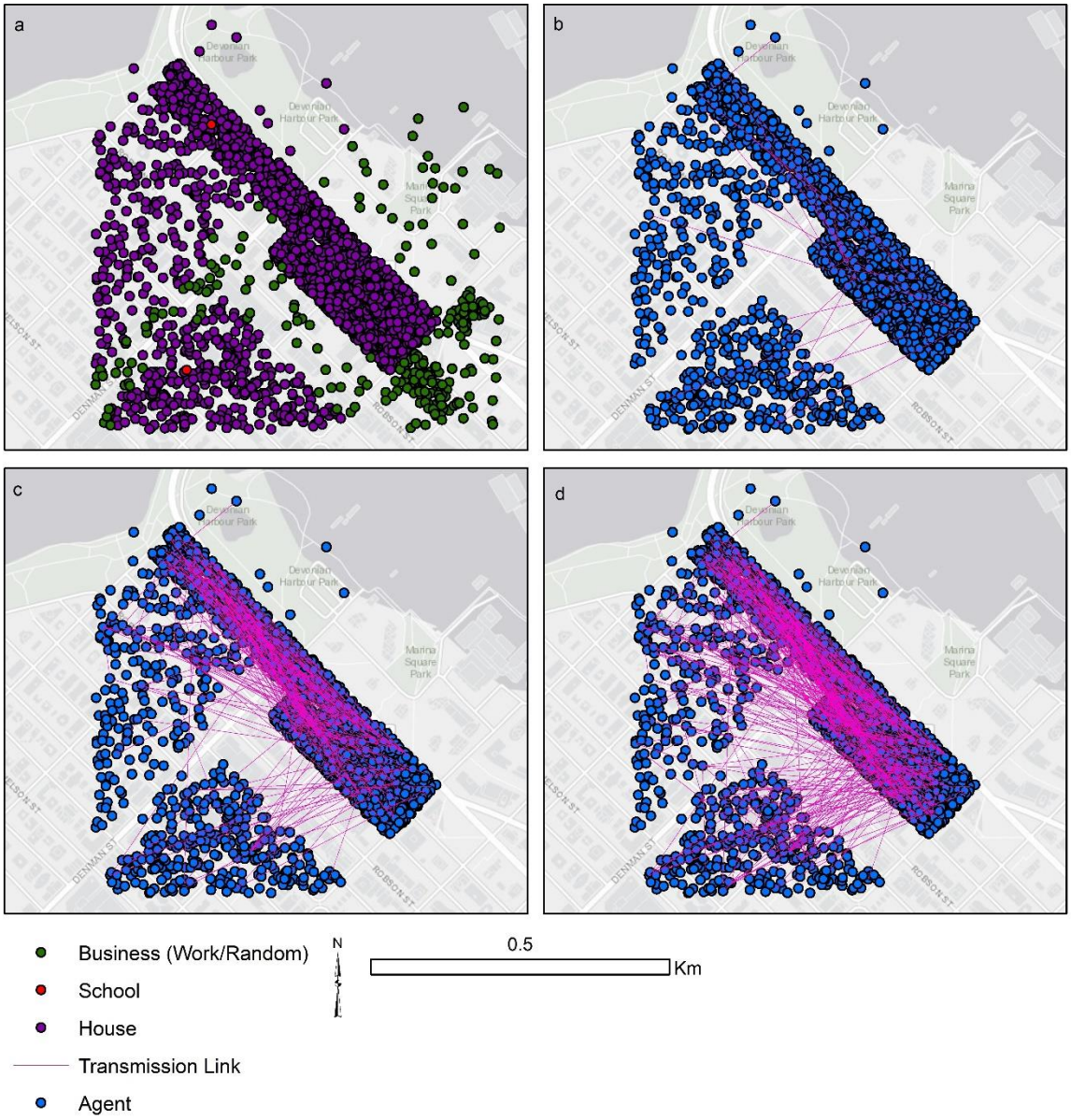

Figure 6.4. **Map of (a) spatial location of homes, businesses, and schools that the agents travel between and the simulated infection network $SN_{infected}$ as it evolves over time for time steps (b) t_{168}, (c) t_{1176}, and (d) t_{6216}.**

Figure 6.5. **The infection network $SN_{infected}$ in a non-spatial representation for time steps (a) t_{168}, (b) t_{1176}, and (c) t_{6216}.**

For the purpose of visualization, the infection network is also presented as a non-spatial representation in Figure 6.5. Table 6.8 presents the network measures obtained from this example spatio-temporal social network as it evolves over time after 7 days at t_{168}, 49 days at t_{1176}, and 259 days at t_{6216}. The model is initialized with 30 infected agents, which generate a set of disconnected infection networks. As the infection spreads, the infection networks increase in size as does the average degree of infection $<k_{infect}>$. The clustering coefficient in the infection network is always 0. This is because the agents are unable to become infected with influenza a second time. As a result, the agents v_j who are infected by agent v_i cannot infected each other. The network maintains modularity as it evolves and the powerlaw degree distribution *P(k)* becomes stronger.

Table 6.8. Spatio-temporal network measures obtained from $SN_{infected}$.

Network measure	Time *t*		
	168	1176	6216
N	54	354	582
L	34	333	561
<*C*>	0	0	0
$<k_{infect}>$	1.259	0.941	0.964
$<k_{infect}>_{max}$	4	7	7
$<k_{infect}>_{min}$	1	1	1
$<k_{infect}>_{\sigma}$	0.79	1.14	1.04
M	0.938	0.913	0.927
C	20	32	38
D	0.024	0.003	0.002
<*l*>	1.358	4.109	7.12
Apha	8.281914	3.384646	3.47124
Sigma	2.574545	0.179749	0.136037
Distance	0.124665	0.280099	0.2772
p value	0.000325	0.00017	6.36E-08

6.6.2. Validation Results

In section 5.1., the simulation results and network measures of the social network SN_{social} and the infection network $SN_{infected}$ generated by the Epi-N-ABM been presented for one model run as an example. However, due to randomness incorporated into the Epi-N-ABM, each model run is slightly different. Therefore, the model is run 30 times and

the network measures from the social network SN_{social} and infection network structure $SN_{infected}$ have been obtained and the average, standard deviation and standard error across all 30 simulated networks have been calculated. Based on the calculated standard error for each network measure, it can be concluded that each model run produces a network structure that varies with respect to size, but the network structure itself remains the same.

Social Network

Network measures have been obtained from the social network structures SN_{social} simulated in each of the 30 model runs. Table 6.9a presents the average value for each network measure, the standard deviation, and the standard error.

Table 6.9. Table 9. Network measures obtained and averaged over 30 runs for a) the spatial social network and b) its random equivalent

Network Measure	a) Network Measures for SN_{social} obtained over 30 model runs			b) Network Measures for SN_{random} obtained over 30 model runs		
	Average	Standard Deviation	Standard Error	Average	Standard Deviation	Standard Error
Number of Nodes	5681.800	47.498385	8.671979	5681.80	47.49838473	8.6719789
Number of Links	76686.233	1725.235530	314.9835	76686.2	1725.23553	314.98347
Average Clustering Coefficient	0.356	0.002187	0.000399	0.002	1.32328E-18	2.41598E-19
Average Degree	13.493	0.199093	0.036349	13.979	0.003342516	0.000610257
Average Degree Max	34.333	2.411658	0.440306	24.900	1.322223832	0.24140394
Average Degree Min	1.200	0.664364	0.121296	7.000	0	0
Average Degree St. Deviation	5.643	0.188211	0.034362	2.633	0.018675819	0.003409722
Modularity	0.801	0.002212	0.000404	0.186	0.000498273	0.000009
Number of communities	93.300	3.007462	0.549085	24.400	1.379655129	0.251889412
Assortativity	0.397	0.021963	0.004010	0.001	0.000378153	0.000007
Network Density	0.002	0.000020	0.000004	0.002	0.00002	0.000004
Path length	4.458	0.021091	0.003851	3.590	0.002034191	0.000371391
Degree distribution	Poisson	NA	NA	Poisson	NA	NA

It can be concluded that the non-imposed network characteristics as quantified by the network measures that emerge from the processes implemented in the Epi-N-ABM correspond well with observed network characteristics in real spatial social networks (Table 6.6). The simulated spatial social network SN_{social} is characterized as heterogeneous and assortative with a positively skewed degree distribution *P(k)* characteristic of a real spatial social network. The simulated spatial social network also SN_{social} exhibits a moderate clustering coefficient, meaning that agents v_j that come into contact with agent v_i also come into contact with each other. Furthermore, the simulated spatial social network SN_{social} exhibits a strong measure of modularity, which strengthens over time, indicating the presence of community structure. Finally, the average path length is short, meaning that individuals in the network are able to reach each other through a small number of connections. In order to confirm that SN_{social} is indeed a small world network, the obtained measures characterizing the SN_{social} are compared with its corresponding SN_{random} network structure (Table 6.9b). It is clear that based on the obtained network measures from both SN_{social} and SN_{random} $\langle l \rangle_{SNsocial} \geq \langle l \rangle_{SNrandom}$ and $\langle C \rangle_{SNsocial} \gg \langle C \rangle_{SNrandom}$, meaning that the SN_{social} is a small world network (Humphries and Gurney 2008).

Infection Network

Network measures have been obtained from the infection network structures $SN_{infected}$ simulated in each of the 30 model runs. Table 6.10 presents the average value for each network measure, the standard deviation, and the standard error.

As per observed properties of real transmission networks, the degree distribution *P(k)* of the simulated SN_{social} can be described as a power law distribution with an average alpha exponent -α of 3.53 and a standard error sigma σ of 0.1. The Kolmogorov-Smirnov distance *D* is small with a value of 0.2 for both. The goodness of fit of the simulated *P(k)* is compared with both a power law and an exponential distribution. The log likelihood ratio *R* between the two distributions that indicates a power law is a better fit than an exponential with a *p*-value of 0.0000000174.

Table 6.10. Network measures obtained and averaged for the infected network

Network Measure	Average	Standard Deviation	Standard Error
Number of Nodes	669.13	287.93	52.57
Number of Links	650.73	288.03	52.59
Average Clustering Coefficient	0.00	0.00	0.00
Average Degree	0.965	0.023	0.004
Average Degree Max	7.50	1.25	0.23
Average Degree Min	1.00	0.00	0.00
Average Degree St. Deviation	1.11	0.05	0.01
Modularity	0.930	0.013	0.002
Number of communities	35.23	6.25	1.14
Network Density	0.0018	0.0009	0.0002
Path length	8.32	3.00	0.55
Alpha	3.528	0.139	0.025
Sigma	0.1098	0.0213	0.0039
Distance	0.2631	0.0080	0.0015
p value	0.0000000174	0.0000000697	0.0000000127

6.7. Discussion and Conclusions

Although useful, traditional validation approaches for ABMs are unable to fully evaluate the model's internal generative mechanisms from which system level patterns emerge. Quantitative evidence that ABMs are capable to capture generative mechanisms are desirable, however internal processes are often difficult to identify and statistical methods are not flexible enough to account for complexity inherent to real-world systems (Grimm and Berger 2016). Network science, however, operates on the foundation of complexity and thus is capable of characterizing complex systems and the models that aim to represent them using a set of statistic-based network measures. Therefore, the NEAT approach was developed, capable to quantify important generative mechanisms represented by the model, reducing the black box effect common to many ABMs.

The non-imposed simulated network characteristics that emerge from the processes implemented in the Epi-N-ABM correspond with observed network characteristics in real spatial social networks and transmission networks. Interestingly, the spatial social network structure differs from the infection network structure even though the two are tightly coupled. The spatial social network structure is that of a small-

world network that can be characterized with a degree distribution *P(k)* that is normally curved with a skew to the right. The spatial infection network structure is that of a scale-free network that can be characterized by a degree distribution *P(k)* that follows a power law. This is a function of the heterogeneity across the network where some agents known as super spreaders infect an anomalously large number of agents, while the majority of agents infect only a few.

The proposed NEAT approach is subject to the same limitations with respect to the computational efficiency as the ABM it is testing. As the networks derived from the ABM increases in size and complexity, the time required to calculate the network measures increases. Furthermore, network science provides many measures for quantifying the network structure and behaviour across several scales. Apart from the measures that are included in this study in detail, there are additional spatial network measures that can be useful to characterize another geospatial ABM representing a different phenomenon.

The proposed network validation approach would be particularly useful for testing other ABM applications such as urban (transportation networks), epidemiology, and mobility modelling. The ability of developed ABMs to reproduce statistical patterns observed in the real world by network scientists would provide confidence that the internal processes represented in the model are realistic and theoretically grounded. For example, Rand *et al.* (2003) use a similar non-network based approach where they demonstrate the ability of their ABM to generate statistical patterns that are also observed in the real world including Zipf's law and Clark's law in their model of urban development.

The Epi-N-ABM is a theoretical model, one that is not grounded in real geospatial datasets representing the real socio-demographics of the City of Vancouver, Canada. As future research work, it would be interesting to initialize the model with agents with data that represent the real population structures in Metropolitan Vancouver Region using census data for the region. This would mean that real social and epidemiological network data could be used in the model validation using the NEAT approach. The proposed spatial network validation approach can be used for the validation of internal processes and emergent structures of complex systems. This approach would be most useful in combination with other ABM validation approaches, ones that validate the

spatial patterns that emerge from the internal model processes. Currently, Repast Simphony software with the Geography package for spatial ABM development contains some network functionality but does not include network measures. Therefore additional work in exploring the integration of networks and ABM may find value in the development of methods and tools that are able to automatically generate network measures, both the ones included in this paper and additional, more advanced or application specific measures, for a variety of ABMs representing spatial phenomena that can be used to give transparency and confidence to internal model processes. In summary, this study contributes to discussions and approaches that seek to advance and strengthen ABM validation methodologies and proposes thoughtful consideration for ways in which internal model processes can be evaluated and encourage the development of models that not only reproduce system level patterns, but that are also theoretically sound.

6.8. References

Alizadeh, M., Cioffi-Revilla, C., & Crooks, A. (2017). Generating and analyzing spatial social networks. *Computational and Mathematical Organization Theory,* 23(3), 362-390.

Alstott, J., Bullmore, E., & Plenz, D. (2014). powerlaw: A Python package for analysis of heavy-tailed distributions. *PloS One,* 9.1, e85777.

Anderson, T. & Dragicevic, S. (2018). Network-agent based model for simulating the dynamic spatial network structure of complex ecological systems. *Ecological Modelling.* 389, 19-32.

Augusiak, J., Van den Brink, P. J., & Grimm, V. (2014. Merging validation and evaluation of ecological models to 'evaludation': a review of terminology and a practical approach. *Ecological Modelling,* 280, 117-128.

Barabási, A. L. (2016). *Network Science*. Glasgow, UK: Cambridge University Press.

Bian, L. (2004). A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B: Planning and Design,* 31(3), 381-395.

Bian, L., & Liebner, D. (2007). A network model for dispersion of communicable diseases. *Transactions in GIS,* 11(2), 155-173.

Bousquet, F., & Le Page, C. (2004). Multi-agent simulations and ecosystem management: a review. *Ecological Modelling,* 176(3), 313-332.

Broeke, G. t., Van Voorn, G., & Ligtenberg, A. (2016). Which sensitivity analysis method should I use for my agent-based model? *Journal of Artificial Societies & Social Simulation,* 19(1).

Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science,* 19(2), 153–174.

Bruggeman, J. (2013). *Social Networks: An Introduction*. New York, NY: Routledge.

Carpenter, C., & Sattenspiel, L. (2009). The design and use of an agent-based model to simulate the 1918 influenza epidemic at Norway House, Manitoba. *American Journal of Human Biology: The Official Journal of the Human Biology Association,* 21(3), 290-300.

Castle, C. J., & Crooks, A. T. (2006). Principles and concepts of agent-based modelling for developing geospatial simulations.

Coburn, B. J., Wagner, B. G., & Blower, S. (2009). Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC medicine,* 7(1), 30.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin,* 70(4), 213.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment,* 37(1), 35-46.

Crooks, A. T., & Hailegiorgis, A. B. (2014). An agent-based modeling approach applied to the spread of cholera. *Environmental Modelling & Software*, *62*, 164-177.

Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems,* 32(6), 417–430.

Evans, A. (2012). Uncertainty and error. In Batty, M., Crooks, A.T., See, L.M., Heppenstall, A.J. (Eds). *Agent-based Models of Geographical Systems.* Netherlands: Springer.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters,* 27(8): 861-874

Fischer, C. S. (1982). *To dwell among friends: Personal networks in town and city*. Chicago, IL: University of Chicago Press.

Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. (2011). An agent-based model of epidemic spread using human mobility and social network information. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*. Boston, MA: IEEE.

Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. (2011, October). An agent-based model of epidemic spread using human mobility and social network information. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* (pp. 57-64). IEEE.

Gilbert, N. (2006). Putting the Social into Social Simulation. Keynote address to the First World Social Simulation Conference, Kyoto.

Goodall, D. W. (1972). Building and testing ecosystem models. *Mathematical Models in Ecology,* 173-194.

Grimm, V., & Berger, U. (2016). Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue. *Ecological Modelling,* 326(24), 177-187.

Grimm, V., & Railsback, S. F. (2012). Pattern-oriented modelling: a 'multi-scope'for predictive systems ecology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 367(1586), 298-310.

Grimm, V., Frank, K., Jeltsch, F., Brandl, R., Uchmański, J., & Wissel, C. (1996). Pattern-oriented modelling in population ecology. *Science of the Total Environment,* 183(1-2), 151-166.

Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science,* 17 (3), 235-249.

Humphries, M. D., and Gurney, K. (2008). Network "small-world-ness": A quantitative method for determining canonical network equivalence. *PLoS ONE*, 3(4).

Kocabas, V., & Dragićević, S. (2009). Agent-based model validation using Bayesian networks and vector spatial data. *Environment and Planning B: Planning and Design,* 36(5), 787–801.

Leung, K., Jit, M., Lau, E. H., & Wu, J. T. (2017). Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Scientific Reports,* 7(1), 7974.

Liu, F.C.S. (2011). Validation and agent-based modeling: a practice of contrasting simulation results with empirical data. *New Mathematics and Natural Computation,* 07(03), 515–542.

Manson, S. M. (2007). Challenges in evaluating models of geographic complexity. *Environment and Planning B: Planning and Design*, *34*(2), 245-260.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R. & Heijne, J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine,* 5(3), e74.

Newman, M. E. (2002). Assortative mixing in networks. *Physical Review Letters*, *89*(20), 208701.

Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*(6), 066133.

O'Sullivan, D. (2014). Spatial network analysis. *Handbook of regional science*, 1253-1273.

Parunak, H. V. D., Savit, R., & Riolo, R. L. (1998). Agent-based modeling vs equation-based modeling : a case study and users' guide. In Gilbert, N., Conte, R. & Sichman, J.S., eds. Multi-Agent Systems and Agent-Based Simulation. Berlin, Heidelberg: Springer.

Perez, L., & Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geography,* 8(1), 50.

Pérez, L., Dragićević, S., & White, R. (2013). Model testing and assessment: perspectives from a swarm intelligence, agent-based model of forest insect infestations. *Computers, Environment and Urban Systems*, *39*, 121-135.

Pontius Jr, R.G., Peethambaram, S. & Castella, J.C. (2011). Comparison of three maps at multiple resolutions: a case study of land change simulation in Cho Don District, Vietnam. *Annual Association of American Geographers,* 101(1), 45-62

Pontius, Jr, R. G. (2000). Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing,* 66(8), 1011-1016.

Railsback, S. F. (2001). Concepts from complex adaptive systems as a framework for individual-based modelling. *Ecological Modelling,* 139(1), 47–62.

Rakowski, F., Gruziel, M., Bieniasz-Krzywiec, Ł., & Radomski, J. P. (2010). Influenza epidemic spread simulation for Poland—a large scale, individual based model study. *Physica A: Statistical Mechanics and its Applications,* 389(16), 3149-3165.

Rakowski, F., Gruziel, M., Bieniasz-Krzywiec, Ł., & Radomski, J. P. (2010). Influenza epidemic spread simulation for Poland—a large scale, individual based model study. *Physica A: Statistical Mechanics and its Applications*, *389*(16), 3149-3165.

Repast Simphony. (2018). Version 2.6. [Computer Software]. Chicago, IL: University of Chicago.

Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling,* 90(3), 229–244.

Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences,* 201009094.

Skvortsov, A. T. R. B., Connell, R. B., Dawson, P., & Gailis, R. (2007). Epidemic modelling: Validation of agent-based simulation by using simple mathematical models. In *MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand*.

Small, M., Walker, D. M., & Tse, C. K. (2007). Scale-free distribution of avian influenza outbreaks. *Physical Review Letters*, *99*(18), 188702.

Tian, Y., Osgood, N. D., Al-Azem, A., & Hoeppner, V. H. (2013). Evaluating the effectiveness of contact tracing on tuberculosis outcomes in Saskatchewan using individual-based modeling. *Health Education & Behavior*, *40*(1_suppl), 98S-110S.

Topping, C. J., Høye, T. T., & Olesen, C. R. 2010. Opening the black box—Development, testing and documentation of a mechanistically rich agent-based model. *Ecological Modelling,* 221(2), 245-255.

Torrens, P. M. 2010. Agent-based models and the spatial sciences. *Geography Compass,* 4(5), 428–448.

Travers, J., & Milgram, S. (1967). The small world problem. *Phycology Today*, *1*(1), 61-67.

Visser, H., & de Nijs, T. 2006. Map comparison kit. *Environmental Modelling & Software,* 21, 346-358.

Watts, D. J., & Strogatz, S. H. 1998. Collective dynamics of “small-world” networks. *Nature,* 393(6684), 440–442.

White, R. 2006. Pattern based map comparisons. *Journal of Geographical Systems,* 8(2), 145–164.

Wong, L. H., Pattison, P., & Robins, G. (2006). A spatial model for social networks. *Physica A: Statistical Mechanics and its Applications*, *360*(1), 99-120.

Chapter 7.

Conclusions

7.1. General Conclusions

Complex systems theory provides the conceptual framework for analysis and understanding of dynamic geographic phenomena. Geographic and network modelling approaches offer new means for the representation and analysis of geographic phenomena; however, the two approaches are rarely incorporated. This dissertation argues for the integration of networks with geographic automata and for the development of a suite of network-based geographic automata modelling approaches, providing the opportunity for improved representation, analysis, testing, and communication of complex spatial systems and the models that aim to represent them.

The results can be summarized as follows: 1) the theoretical and practical foundation of integrated network-based geographic automata; 2) the application of the network-based geographic automata approaches to a variety of phenomena and the analysis of model results using network measures to describe, characterize, and better understand the structure, dynamics, and evolution of network representations; and 3) the software routines for the GNA, N-ABM, and NEAT approaches and a variety of developed functions programmed for the measurement of model results can be used for different case studies and relevant geospatial datasets.

7.2. Summary of Findings

The dissertation addresses the presented research questions by implementing a succession of novel network-based automata modelling and model testing approaches that leverage complex systems theory, GISc, and GAS, and network science for the conceptualization, representation, and analysis of complex spatial systems. The purpose of Chapter 2 is to explore research possibilities that lie at the intersection of GISc and network science. Based on this opportunity, this thesis research employs spatial networks to represent and analyze complex geographic phenomena and face challenges in 1) fully exploring the link between spatial dynamics and processes and emergent

spatial structures and 2) understanding how and why these structures evolve over space and time. This dissertation proposes that the integration of network theory and geographic automata offers the potential to fill those gaps by providing new means for representation and analysis of complex spatio-temporal systems as evolving spatial networks. At the same time, this integration offers the potential for new means to improve the robustness of the geographic automata models themselves.

The research study presented in Chapter 3 integrates networks theory, complex systems theory, and geographic automata systems theory to develop a network-based automata modelling approach called Geographic Network Automata (GNA). The GNA modelling approach can be applied to any spatial-temporal phenomena and can be operationalized by taking the following steps:

1. *Conceptualize the system of interest as a spatial network.* Based on the system of interest, determine the components that make up the system of interest that are to be represented as nodes. Determine the interactions, flows, and relationships between systems components that are to be represented as links. Define the node and link's spatial, non-spatial, and topological attributes. Determine whether the network is weighted or directional. Consider the scale at which the system(s) of interest should be represented. Identify whether the system is constrained to operating on an underlying spatial network *UN*.
2. *Identify important graph theory measures.* Determine which global or local graph theory measures are the most valuable to measure and understand the system of interest. The distribution of measures across the entire network such as degree distribution is particularly useful as it gives an overall snapshot of network structure.
3. *Define the neighbourhood.* Determine the way in which other nodes *j* might interact with node *i.* Define this interaction based on distance, weight, behavior, or probability to best capture this interaction.
4. *Development of the transition rules.* The transition rules govern interactions between a node's neighbours and results in the evolution of the network structure. The state or location of each node or the connections between two nodes may change as a result of the transition rules. Based on these transition

rules, new nodes may be added to the network, some nodes may be removed, and the connections between existing nodes may be rewired.

5. *Identify potential connection costs.* The influence of the matrix on the spatial network evolution must be identified. For example, the connection between two nodes may be limited in the case that there is a geographic barrier between the two, if the distance exceeds a certain threshold, or as a function of a cost surface analysis.
6. *Implement of the GNA.* The implementation of the GNA is developed as a programmed application containing node automata and link objects. Initial node location and attributes are often determined by input data. Links are defined as a function of the neighbourhood, transition rules, and the connection cost. Node and link automata objects store spatial, non-spatial, and topological attributes which update at each iteration. Link objects store additional information including the pairs of nodes they connect and their location as well as any flows that take place between nodes.
7. *Perform GNA calibration and sensitivity testing using an independent dataset.* The NEAT approach can be used to perform these tests.
8. *Perform GNA validation by comparing model with independent data.* The NEAT approach can be used to validate the GNA.
9. *Execute model and model scenarios.*
10. *Apply graph theory to characterize network structures (underlying or spatial).* Graph theory measures can be used to enhance the understanding of the phenomena being simulated as networks, to compare network structures as the network evolves, or to compare network structures between scenarios.

The GNA framework requires spatial data as an input to initialize node location, parameterize nodes, and to implement the network matrix and any potential geographic barriers. Data, spatial data or data from the literature, is also required to parameterize transition rules and connection cost. Datasets independent of model development are required for model testing. This can include network datasets in the case that the NEAT approach should be implemented or can include non-network based spatial datasets for more traditional methods of validation.

Chapter 3 both presents the modelling framework and applies it to a theoretical complex system, the Game of Life (Conway, 1970). Based on the network measures collected, the Game of Life system can be described as having properties of a random geometric graph. The GNA modelling approach facilitates the examination of the close coupling between spatial network structure and spatial network dynamics. Results suggest that as random network structures evolve through the process of growing and shrinking in response to local dynamics, the spatial structure of the network is altered, which has further implications for local dynamics. The GNA modelling approach facilitates the examination of the close coupling between spatial network structure and spatial network dynamics and allows for better understanding how real systems that operate on a random geometric graph structure such as social dynamics may be impacted by changes to the structure of the system. The developed approach is both general and flexible so that it can be applied to represent and analyze many real systems including urban, social, and ecological. Based on the specific use of the network data model and the development of transition rules *R* that are designed to simulate dynamics between nodes, the GNA modelling methodology is a substantial departure from the cellular automata and agent-based modelling approaches.

Chapter 4 extends the work presented in Chapter 3 by applying the GNA modelling framework to represent an important ecological system as an evolving spatial network that can be analyzed using graph theory. Specifically, Chapter 4 applies the GNA modelling framework to the case study of the emerald ash borer (EAB), an invasive bark beetle species that infests and kills ash trees. The GNA simulates the propagation of the EAB species at a coarser spatial scale across a landscape of forest stands containing ash trees. Network rules representing local dispersal processes are applied to the network of infested forest stands, forming an evolving spatial network that grows as new forest stands become infested and shrinks as old infested stands die. The approach facilitates the examination of the non-linearity and close coupling between spatial structure and spatial dynamics. The spatial organization of forest stands across the landscape influences the spread of EAB, forming the SN_{EAB}. In turn, the spread of EAB influences the structure of the SN_{EAB} as forest stand nodes are infested and die limiting dynamics of dispersal in future iterations. Results suggest that the small world spatial network structure produced by the spatial structure of the landscape and the EAB spatial dynamics has implications for EAB management. Specifically, the emergent network

structure indicates that local eradication efforts will have little impact on slowing the propagation of the EAB across the state. This study reinforces the potential use and usefulness of the GNA modelling approach for the representation and analysis of complex spatial systems.

The evolving spatial network structures of EAB forest insect infestation at finer spatial scales are explored in Chapter 5. The purpose of Chapter 5 is to integrate networks and existing geographic automata approaches such as ABM to form an N-ABM, not as inputs for the ABM, but to conceptualize the ABM differently, providing new means for understanding, analyzing, and communicating the model results. In general, the N-ABM can be operationalized to simulate a range of geospatial phenomena by implementing the following steps:

1. *Design, develop, and test an ABM.* Identify potential scenarios for knowledge discovery and decision-making.
2. *Conceptualize the system(s) of interest as represented in the ABM as a spatial network(s).* Determine the interactions, flows, and relationships between systems components that are to be represented as links. Define the node and link's spatial, non-spatial, and topological attributes. Determine whether the network is weighted or directional. Consider the scale at which the system(s) of interest should be represented as networks.
3. *Implement the N-ABM.* Implement a network model where flows, interactions, and relationships between system components in the ABM are represented as a set of links and nodes. Couple the ABM component to the network component. The programming logic and pseudocode developed for the coupling of the ABM component and the network component of the ABM is presented in Figure 5.3.
4. *Identify important graph theory measures.* The selection of network measures are dependent on the phenomena being represented. Global graph theory measures are useful for characterizing the structure of the network as a whole, particularly looking at the distribution of network measures i.e. degree distribution.
5. *Apply graph theory to measure the system(s) as spatial networks.* These measures may be calculated at one point in time, as the network evolves over time, or as the network structure responds to various scenarios.

In order to develop an N-ABM, spatial data is needed as an input to initialize agent locations and to parameterize the agents. Data, spatial data, or data from the literature, is also required to parameterize agents and agent behavior. Datasets independent of model development are required for model testing. This can include network datasets in the case that the NEAT approach should be implemented for model validation or can include non-network based spatial datasets for more traditional methods of validation.

Once again, the development of network-based automata facilities the exploration of the link between spatial structure and spatial processes, specifically by identifying the underlying interactions and generative mechanisms that drive the emergence of spatio-temporal patterns. Based on the network measures collected, patterns of EAB forest insect infestation at fine spatial scales can be described as scale-free networks. Specifically, the scale-free network structure emerges because of the preferential attachment that individual EAB has for specific types of trees. The difference between the network structure of EAB infestation at the regional scale (GNA) and the fine scale (N-ABM) can be explained by the different processes that are relevant at each scale. At the small scale, processes and dynamics including biological interactions between individual EAB and tree, lifecycle, and finer temporal scales are key drivers for patterns of spread. At a coarser spatial scale, processes such as dispersal mechanisms and the structure of the landscape as a whole are more relevant. This suggests that spatial networks are subject to the modifiable areal unit problem (MAUP) where graph theory measures calculated for the networks vary as a function of the scale at which the phenomena are represented.

The differences in network structures as a function of scale suggests that management recommendations would differ between the GNA which simulates EAB propagation at the landscape scale and the N-ABM which simulates EAB propagation and behavior at an individual scale. At a landscape scale, eradication strategies would be required to focus on the spatial structure of the landscape and how it either exacerbates or slows EAB spread based on their dispersal processes. This may involve using the GNA to test the removal of key forest stands that connect the landscape and allow for the natural propagation of the species. At the individual scale where the simulation falls within the administrative boundaries of an individual town or county, the N-ABM may better be used to strategize for biological control by testing the interactions

between a biological control agent and the EAB or for predicting financial and ecological impacts to the region.

Like all geographic automata models, the proposed network-based automata modelling approaches presented in this dissertation are challenging to test. Model testing often relies on traditional methods which compare emergent simulated spatial patterns with patterns found in reality. The integration of networks into geographic automata, not as inputs for the model, but to provide additional ways to describe, analyze, and communicate model results also has the potential for new means at model testing. Chapter 6 explores the development of a NEtworks for Agent-based model Testing (NEAT) approach. In general, the NEAT approach can be operationalized to test a variety of ABMs by taking the following steps:

1. *Design, develop, and test an ABM.* Identify potential scenarios for knowledge discovery and decision-making.
2. *Conceptualize the system(s) of interest as represented in the ABM as a spatial network(s).* Determine the interactions, flows, and relationships between systems components that are to be represented as links. Define the node and link's spatial, non-spatial, and topological attributes. Determine whether the network is weighted or directional. Consider the scale at which the system(s) of interest should be represented as networks.
3. *Implement the N-ABM.* Implement a network model where flows, interactions, and relationships between system components in the ABM are represented as a set of links and nodes. Couple the ABM component to the network component. The programming logic and pseudocode developed for the coupling of the ABM component and the network component of the ABM is presented in Figure 5.3.
4. *Identify useful network measures to characterize the system(s) of interest.* The selection of network measures are dependent on the phenomena being represented. Global graph theory measures are useful for characterizing the structure of the network as a whole, particularly looking at the distribution of network measures i.e. degree distribution.
5. *Obtain real world network measures from empirical data.* Real-world network measures should be obtained for the system of interest either from a dataset that was used independent of model development or from the literature.

6. *Obtain simulated network measures from N-ABM outputs.* The same network measures should be obtained from the simulation outputs either at one point in time or as the network evolves over time.
7. *Compare real-world network measures with corresponding simulated network measures.* Compare and critically evaluate how well the simulated network measures compare to the corresponding real-world network measures. Each network measure operates as a mathematical filter, for which to base the grounds of acceptance or rejection of the developed ABM

The NEAT approach requires real network datasets that can be characterized using a set of graph theory measures or network analysis from real datasets and observations obtained from the literature. Other non-network based spatial datasets may be used to combine the NEAT approach with more traditional validation approaches.

The NEAT approach is applied it to test an epidemiological N-ABM (Epi-N-ABM) representing the spatio-temporal transmission of influenza among a human population. The N-ABM is abstracted into both an evolving social network and an involving transmission network. The network structure that emerges from the social dynamics simulated in the Epi-N-ABM is that of a small world network. The network structure that emerges from the transmission dynamics that operate based on social dynamics is that of a scale-free network. Based on network measures that characterize the simulated social network and transmission network, the non imposed network characteristics that emerge from the processes implemented in the Epi-N-ABM correspond with observed characteristics of observed spatial social and transmission networks. The findings indicate that the NEAT approach can validate internal dynamics and spatial structures of modelled phenomena by comparing simulated network structures with empirical regularities of observed real networks. The proposed validation approach would be useful for testing other ABMs by comparing simulated network structures with observed networks. The network representation of ABM spatial processes and structures offers new means for analysis, testing, and language for which to communicate and them.

All developed modelling frameworks including the GNA, N-ABM, and the NEAT approach were implemented in Repast Simphony 2.4 (2016), 2.5 (2017), and 2.6 (2018). Repast Simphony is an open source modelling platform that operates within the integrated development environment Eclipse and allows for the import of shapefiles to

be used as model inputs. Each modelling framework was implemented using the Java programming language and utilized Repast Symphony, Apache Commons, Geotools, and Vivid Solutions open source toolkits and Java libraries. Each graph theory measure used to characterize the simulated spatial networks was programmed directly into the application as different functions that can be executed independently or in combination with other graph theory measures. These functions access the topology of the simulated network at each iteration and calculate graph theory values specific to that iteration. Repast Simphony in Eclipse offers a customizable visual interface that allows for the visualization of the spatial simulation as it runs in real time and the ability to export the simulation results including network nodes and links as georeferenced shapefiles. The exported shapefiles for nodes and links also contain an attribute table that include all attributes. For example, the attribute table for the links include a reference to the start node and the end node for each link in the dataset.

Using the exported shapefiles from the simulations that run in Repast and Eclipse, maps presented in each Chapter were designed using ArcMap 10.5 (ESRI, 2011). Node and link attribute tables can be exported and uploaded into Gephi for the aspatial visualization of networks. In Chapter 5 and Chapter 6, the scale free degree distribution of the networks obtained from the EAB N-ABM and the Epi-N-ABM is tested for goodness of fit using *powerlaw,* a Python package for analysis of heavy-tailed distributions (Alstott et al., 2014). This package uses the raw degree *k* for each node obtained from the simulations, for example in the Epi-N-ABM the number of contacts each individual has, and produces a value for the alpha exponent describing the powerlaw, the standard error, and the Kolmogorov-Smirnov distance *D* between the calculated powerlaw and the expected powerlaw. The package also includes a measure of the goodness off fit between the calculated powerlaw and other distributions like an exponential distribution or a lognormal distribution. The goodness of fit includes a p-value.

7.3. Future Directions

“Networks are everywhere”, a phrase found in many network review papers ultimately speaks to the interdisciplinary nature of the usefulness abstracting real-world phenomena to a connected network-based representation. Thus, the proposed suite of network-based automata modelling approaches has the potential to be implemented on

other geospatial applications for the representation, characterization, and analysis of a variety of complex systems. Networks are a natural fit for representing and analyzing relationships and interactions and as such, network-based automata modelling frameworks are an ideal approach for applications when interaction, relationships, dynamics, and flows between sets of components are of interest. The application potential is vast and include movement and flows of information, people, resources, money, ecological species, energy, disease, and transportation vehicles over time and across points in geographic space. Naturally, the study of spatial and non-spatial relationships between individuals is also an ideal application for these modelling approaches. In addition, the network-based automata modelling approach would be ideal to better understand interactions between two or more tightly coupled systems over space and time such interactions between as policy, social, and environmental systems.

There are, however, some technical limitations to the developed approaches. The GNA, N-ABM, and NEAT approach can be used to build and evaluate both theoretical or empirical GAS models. However, empirical models tend to require more data than theoretical GAS models data. At the moment, in a social context, network data is abundant, however remains challenging to obtain in the ecological context. Future work would focus on advancing geospatial modelling approaches to represent and understand complex spatial systems while shifting these modelling approaches into directions that are big spatial data driven.

Big spatial data, the massive amount of location-aware data that is collected every day as a result of ubiquitous computing, can be used to in complex systems modelling for both knowledge discovery and as a tool in decision-making processes (Miller & Goodchild, 2015). Big spatial data including GPS data in phones, sensor data embedded in infrastructure, remote sensors carried by airborne and satellite platforms, radiofrequency identification (RDIF) tags attached to objects, and georeferenced social network data can be incorporated into the process of model design, development, and testing from sources (Miller & Goodchild, 2015). The benefits of data-driven modelling are becoming recognized as a result of the ease of collecting, storing, processing digital data, leading to what some may call the fourth paradigm of science. Anderson (2008) refers to the fourth paradigm, or the “big data deluge”, as “the end of theory” all together, stating that the theories and models that seek to explain phenomena may be obsolete since data is now available in such fine spatial and temporal scale that explanation is not

really needed at all. This has been refuted by those who believe several challenges prevent big data from being useful. As a result, the theory and models to explain phenomena will still be necessary (Liu et al., 2016; Batty, 2013).

Big spatial data is of increased availability in urban locations, providing the opportunity to model and analyze urban phenomena. Commonly used big spatial urban data includes sensors in urban infrastructure and user-generated content; however, there exists a wide range of big urban data that is less widely leveraged including transactional data, administrative and planning data, and data that links traditional survey data with sensors. This data in combination with network-based modelling approaches is particularly relevant in application to analyzing social networks and mobility networks, but also to better understand the tight coupling between urban, social, and environmental systems. Despite the promise of big data-driven modelling, the volume, velocity, and variety of geographic data exceed capabilities of existing computational techniques, thus presenting new challenges in big spatial data management, analysis, assimilation into complex systems models, processing, scalability, and theoretical considerations. As a result, there is a need to explore further the development of new network-based modelling approaches that can overcome these limitations thus making data-driven models useful to provide the means for knowledge discovery and the effective and timely decision-making support for stakeholders, policy-makers, and planners.

Apart from technical limitations, future work may focus on overcoming some more theoretical limitations. The GNA and N-ABM modelling approaches are able to incorporate key elements inherent to complex systems including heterogeneity, interactions, adaptation, and multiple scale phenomena. However, because of their ability to capture complexity, the modelling approaches have the potential to be affected by error and uncertainties that propagate from data used as inputs and through model structure (Yeh & Li, 2006). Future work could expand sensitivity analysis to better understand how changes to inputs such as transition rules, neighbourhood configuration, simulation time, and stochastic variables amplify or reduce error propagation and uncertainty in the model. Chapter 3 to 6 use one-at-a-time (OAT) sensitivity analysis (Cariboni et al., 2007) which is a good starting point, however is limited in that it does not account for interaction effects among the inputs nor does it provide an objective measure of sensitivity (Broeke et al., 2016). Future work would benefit from further

exploring the sensitivity of the model using a sensitivity analysis method such as global sensitivity analysis (GSA) that can better account for interaction between inputs and parameters (Saltelli et al., 2008). Furthermore, it would be valuable to understand where some network structure and behavior is a function of a particular series of events or path dependence. Employing the variant-invariant method developed by Brown et al. (2005) on the developed network-based automata modelling approaches would be particularly interesting endeavors in future work.

7.4. Research Contributions

This dissertation research contributes to the scientific research domains that aim to better represent, analyze, and understand complex spatio-temporal phenomena. In general, the integration of networks into traditional GAS approaches allows for graph theory to be used for the analysis and evaluation of the models and the phenomena they aim to represent, not commonly leveraged in traditional GAS approaches. The dissertation finds that network-based geographic automata, which can be measured with graph theory, enables new ways to represent, analyze, and communicate the models and the model outputs. This facilitates the direct investigation into the relationship between spatial structure and spatial dynamics and vice versa, something that is demanded in both GIScience and network science literature.

First, this dissertation contributes to existing methodological research in Geographic Information Science (GISc), geographic automata systems (GAS), and network theory by proposing and implementing a suite of novel spatial network-based automata modelling approaches including the GNA, N-ABM, and NEAT approach. The developed GNA approach, the first of its kind, is designed to explicitly leverage network representations, network-based transition rules, and network analysis for the simulation of complex, spatio-temporal phenomena. The GNA modelling framework differs from traditional GAS including CA and ABM as a function of its uniquely explicit view of the network-based relationships and interactions between the spatial features that is represented by network links and the *NxN* adjacency matrix *A*. The GNA modelling framework places emphasis on the representation, analysis, and visualization of relational data, interactions, and flows. In a CA and ABM, rules are implemented that govern relations, interactions, and flows, but they are not represented nor measured discretely. Instead, the way in which the system responds to sets of interactions is

typically measured. The GNA offers a more flexible modelling framework than CA where nodes may be mobile with varying neighbourhood types and non-deterministic system-level behavior. In addition, the GNA offers explicit representation of interactions and thus provides “x-ray” vision for large sets of interactions between components of the system in a way that ABMs traditionally do not.

In the developed N-ABM, networks are integrated into an ABM, not as inputs for the ABM, but rather as a novel way to conceptualize the processes within ABM with the purpose of making simulation models both transparent and measurable. Specifically, the developed N-ABM approach does not seek to leverage networks as inputs, but rather finds value in networks extracted as the model output. The extraction of networks that represent ABM dynamics at one or several points in time ultimately generates a wealth of data for which can be used to analyze the model and the phenomena that the N-ABM seeks to represent. Extracted networks can then be measured using graph theory to provided additional information about the model and the phenomena it aims to represent.

The NEAT approach contributes to longstanding efforts to develop new frameworks for the testing and evaluation of GAS. The NEAT approach builds upon pattern-oriented modelling (POM) where measured graph theory measures can be compared with real graph theory measures for the same system and be used to accept or reject the validity of the ABM. The use of network measures as means for GAS model validation is novel and unique because it offers a statistical approach that is oriented towards measuring complex systems. Because of the rigidity of most extensions of statistical methods that are applied to GAS, many GAS models remain unevaluated (Manson, 2007).

The GNA, ABM, and CA modelling frameworks are presented in Figure 7.1 as approaches that are united under the larger framework of GAS (Torrens & Benenson, 2005). All three approaches aim to represent and explore phenomena using a bottom-up complex systems approach within a geographical context that places the phenomena in *x, y* coordinate space, and that can potentially be expanded to the *z* coordinate as a third spatial dimension. The development, exploration, and implementation of an approach that is explicitly network-based, geographic network automata, as presented in this dissertation helps to better place where existing integrated approaches such as graph-

CA and N-ABMs fit in this broader context. This dissertation research further contributes to existing methodological research in Geographic Information Science (GISc) and geographic automata systems (GAS) by presenting the first use of networks and graph theory for the validation of ABMs. This research contributes to the field of network theory by furthering research in the spatial properties of networks and meeting the existing demands for the development of novel modelling frameworks for studying and testing the evolution of spatial network structures as a function of their dynamics and processes and vice versa.

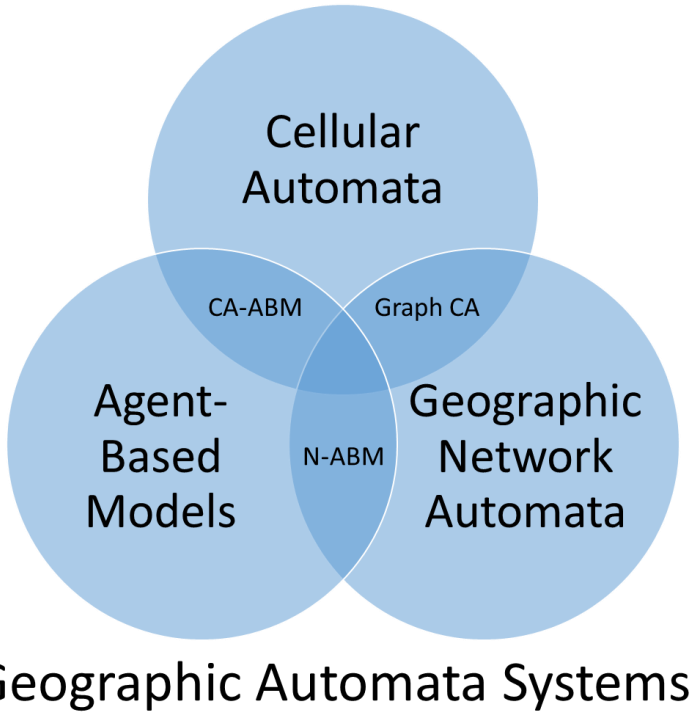

Figure 7.1. Geographic network automata in relation to other geographic automata systems models. This situates N-ABM and other integrated approaches into the larger context.

Secondly, this dissertation research contributes to the application areas of ecological and epidemiological modelling. In general, integrating network analysis into these approaches allows for the use of local or global graph theory measures to finding general rules, laws, or statistical patterns and regularities of complex phenomena. Global measures such as degree distribution can be used to statistically characterize network structure as a whole. These measurements can be used to compare between

different systems or to compare between the same system over time. The ability to statistically compare between the same system over time is crucial in order to meet the demand to understand evolving networks. Local network measures such as the degree of a node characterize individual nodes in the network and can be used to find components within the system that that are key actors in the connectivity of the network and thus facilitate the spread of information, resources, or disease and so on. This information can be used for knowledge discovery, lead to big picture understanding, understand the evolution, prediction, useful when experiments cannot be done in the real world, or decision-making purposes as investigative tools.

More specifically, the application of the network integrated automata modelling approaches to the EAB phenomenon offers both a GNA model and an N-ABM model that leverage the advantage of both complex systems modelling approaches and network modelling approaches to better understand EAB forest insect infestation. These modelling approaches can be used to represent and analyze the tight-coupling and non-linearity between landscape spatial network structure and species dispersal dynamics while also capturing the complexity inherent to forest insect infestation phenomena. This is particularly useful in decision-making for forest insect infestation eradication because the spatio-temporal patterns of the EAB species can be quantified and forecasting and the robustness of the species to eradication measures or other disruptions in the landscape at a variety of scales can be explored. This dissertation contributes to research in epidemiological modelling by further advancing and supporting the development of robust complex systems modelling approaches that represent disease transmission dynamics as a function of heterogeneities over space and time. Epidemiological modelling approaches that are flexible and offer the opportunity to be parameterized for specific case studies can be useful for to the enhancement of decision-making processes by providing a tool that can be used in forecasting and for the testing of ‘what if’ scenarios. The proposed GNA, N-ABM, and NEAT approaches are general, flexible, and widely applicable. Besides the applications explored in this dissertation, these approaches can be applied to represent complex spatial systems in the application areas of human mobility and transportation, sociology, urban planning, and environmental science.

7.5. References

Alstott, J., Bullmore, E. & Plenz, D. (2014). powerlaw: a Python package for analysis of heavy-tailed distributions. *PloS One*, 9.1, e85777.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine.* Retrieved on August 19 2016 from http://www.wired.com/2008/06/pb-theory/

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274–279.

Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International journal of geographical information science*, *19*(2), 153-174.

Cariboni, J., Gatelli, D., Liska, R., & Saltelli, A. (2007). The role of sensitivity analysis in ecological modelling. *Ecological modelling*, *203*(1-2), 167-182.

Conway, J. (1970). The game of life. Scientific American, 223(4), 4.

ESRI. (2011). ArcGIS Desktop: Release 10. [Computer Software]. Redlands, CA: Environmental Systems Research Institute.

Liu, J., Li, J., Li, W., & Wu, J. (2015). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 134-142.

Manson, S. M. (2007). Challenges in evaluating models of geographic complexity. *Environment and Planning B: Planning and Design*, *34*(2), 245-260.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, 80, 449–461.

Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

Repast Simphony. (2016). Version 2.4. [Computer Software]. Chicago, IL: University of Chicago.

Repast Simphony. (2017). Version 2.5. [Computer Software]. Chicago, IL: University of Chicago.

Repast Simphony. (2018). Version 2.6. [Computer Software]. Chicago, IL: University of Chicago.

Ten Broeke, G., Van Voorn, G., & Ligtenberg, A. (2016). Which sensitivity analysis method should I use for my agent-based model?. *Journal of Artificial Societies and Social Simulation*, *19*(1).

Torrens, P. M., and Benenson, I. (2005). Geographic automata systems. International Journal of *Geographical Information Science,* 19(4), 385-412.

Yeh, A. G. O., & Li, X. (2006). Errors and uncertainties in urban cellular automata. *Computers, Environment and Urban Systems*, *30*(1), 10-28.