Dynamic Spatial Networks & Economic Development

A thesis submitted in fulfilment of the requirements for the degree of Master of Science in the School of Mathematics, Statistics and Applied Mathematics at the National University of Ireland, Galway.

NUI Galway, September 2015

E. Staunton

Candidate: Eoghan Staunton

Supervisors: Dr. Petri Piiroinen
Dr. Srinivas Raghavendra
Dr. James Duggan

Head of School: Dr. Ray Ryan

Contents

List of Figures

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of M.Sc., is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work. I have not obtained a degree in this University, or elsewhere, on the basis of this work.

Author:_____ **ID No:**_____

Eoghan Staunton *09370803*

Acknowledgements

I would like to thank my supervisors Dr. Petri Piiroinen, Dr. Srinivas Raghavendra and Dr. James Duggan for their help and guidance. I would have never been able to finish my thesis without their support. This project was funded by the NUI Galway School of Mathematics, Statistics and Applied Mathematics and I would also like to gratefully acknowledge their financial support.

I am also grateful to have been blessed with a friendly and helpful group of fellow postgraduates in the School of Mathematics, Statistics and Applied Mathematics. In particular I would like to thank the members of the NUI Galway Modelling Research group who provided helpful input and productive critiques during our weekly meetings.

Finally, I am thankful to my family, especially my parents, for their support and patience.

Abstract

The spatial agglomeration of economic activities and the unequal growth of urban centres are major topics of interest to both economists and policymakers worldwide. We aim to model and analyse these uneven spatial economic dynamics by setting up an agent-based simulation environment. Our approach is to build mathematical models composed of multiple overlapping dynamic spatial networks across space and time. We introduce our approach to modelling the evolution and growth of spatial networks, both in terms of individual node sizes and the addition of links and nodes to the network. Motivated by official statistics related to the economy of Ireland we then apply the ideas developed in an attempt to model spatial economic dynamics in a small open economy.

Chapter 1

Introduction

1 A Brief Overview

This thesis introduces and develops dynamic network models with the ultimate aim of presenting a novel approach to modelling spatial economic development. Issues related to patterns of spatial economic development such as the spatial agglomeration of economic activities and the unequal growth of urban centres are of major interest to economists and policymakers all over the world. Irish policymakers are certainly no exception. Concerned about the imbalances caused by the concentration of Irish economic activity in the Greater Dublin Area in 2002 the government of Ireland launched a National Spatial Strategy which was designed to ensure *"more balanced social, economic and physical development between the regions"* [32]. However, the strategy failed to remedy the problems faced by the country and was scrapped in 2013. As a result these imbalances remain a major issue of concern for Irish policymakers.

Before presenting any of our models, in this first chapter we provide a quick overview of the contents of this thesis and introduce some of the basic history, language and concepts associated with networks in a mathematical sense. In Chapters 2 and 3 we introduce dynamic models of networks. In particular, Chapter 3 focuses on the growth of individual nodes within a larger network while Chapter 2 focuses on the growth of spatial networks in terms of adding nodes and links over time. Finally, Chapter 4 presents some motivating statistics regarding recent spatial economic development in Ireland and introduces an approach to modelling spatial economic development in which our model is composed of multiple dynamic, overlapping networks across space and time.

In Chapter 2 we introduce a set of algorithms for the growth of spatial networks. These networks are ones which are embedded in space and in which space plays a key role in their evolution. In that chapter we allow spatial networks to grow by adding nodes and links over time according to the rules set out by our algorithms. The spatial networks begin as a single isolated node and the positions of any further nodes added to the network are not predetermined, rather their position and the links between them are established over time as the network grows and develops. This approach was inspired by a simple and elegant model due to Kaiser and Hilgetag who were in turn initially motivated by the evolution of biological networks [41].

In Chapter 3 we investigate growth of individual nodes in a network where it is governed by a discrete time evolution operator known as a map. In defining node growth we take inspiration from a particular map known as the logistic map which was made famous by the theoretical physicist and mathematical biologist Lord Robert May in his 1976 article *Simple mathematical models with very complicated dynamics* [48]. The logistic map is often presented as the quintessential example of how complex and even chaotic behaviour can arise in simply defined systems. We analyse the long-term behaviour of node sizes in a network evolving according to our map observing interesting and complex behaviour. In particular we note both similarities and striking differences to the behaviour observed in the case of the classic logistic map.

In our final chapter, Chapter 4, we turn our attention to spatial economic

development and in particular to the problem of uneven spatial economic dynamics. The spatial agglomeration of economic activities, unequal growth of urban centres and unequal growth of industries are major topics of economic interest. In this chapter we provide evidence for all three in the Irish context since Ireland's emergence from protectionism in the 1950s. We then introduce some of the most prominent arguments put forward by economists to explain many of the important features of spatial economic development. Finally we introduce our new approach to modelling spatial economic dynamics. We apply the ideas developed in Chapters 3 and 2 alongside existing concepts from economics to build models composed of dynamic spatial networks in which each node represents an economic entity and each link a relationship.

2 Introducing Networks

Before continuing we will first provide a basic history of the birth of network theory and introduce some of the mathematical language and notation for describing networks. We will also present some examples of networks being applied in a real-world setting.

2.1 The First Network

Motivated by the famous Königsberg bridge problem the Swiss mathematician Leohnard Euler first introduced the study of *graphs* or *networks* in 1736 [27]. The inhabitants of the old Prussian city of Königsberg, now Kaliningrad, Russia, debated whether it was possible to go for a walk crossing each of the seven bridges in the city over the river Pregel exactly once with the additional requirement that the walk ends in the same place it began.

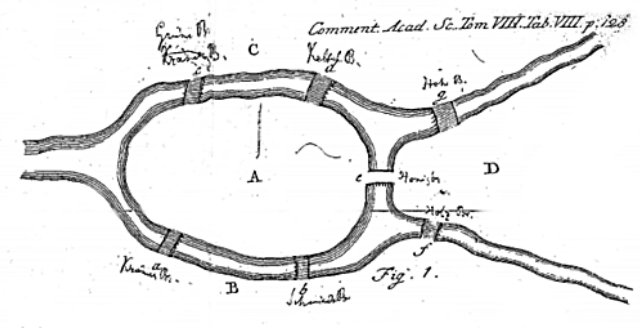


Figure 1.1: The famous bridges of Königsberg problem reproduced from [27].

Euler showed that there was no such path by using a simple, elegant approach. He reduced the map in Figure 1.1 to a series of lines, known as *edges* or *links*, representing bridges, and circles, known as *vertices* or *nodes*, representing portions of land, known as a graph or network. This network representation of

the problem is shown in Figure 1.2[1]. Using this approach Euler showed that to be able to perform a walk on a network starting and finishing on the same node and using each link exactly once is only possible if every node is connected by an even number of links. This condition was subsequently proved to be sufficient in 1873 by the German mathematician Carl Hierholzer [35].

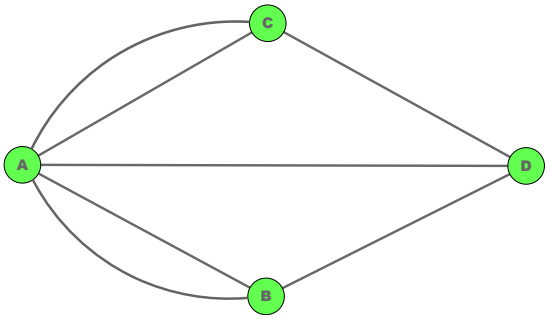


Figure 1.2: Network representation of the famous bridges of Königsberg problem.

2.2 Describing Networks

A network $N = (V, E)$ is defined as a set of vertices or nodes, $V = \{v_1, v_2, \ldots\}$, connected by a set of edges or links, $E = \{(v_i, v_j), (v_k, v_l) \ldots\}$. Networks can be directed, i.e. have links that connect nodes only one way, or undirected, i.e. have links that connect nodes both ways. In this thesis we focus on networks that are undirected, have no self-links and have at most one link between each pair of nodes in the network. A network diagram of an example of such a network is shown in Figure 1.3. The network shown has eight nodes connected by thirteen links. The node set of the network is given by

$$V = \{v_1, v_2, \ldots, v_8\} \tag{1.1}$$

and the edge set is given by

$$\begin{aligned} E \quad = \quad & \{(v_1, v_2), (v_1, v_4), (v_1, v_6), (v_2, v_5), (v_2, v_6), (v_3, v_4), (v_3, v_7), \\ & (v_3, v_8), (v_4, v_5), (v_4, v_6), (v_4, v_7), (v_5, v_6), (v_6, v_7)\}. \end{aligned} \tag{1.2}$$

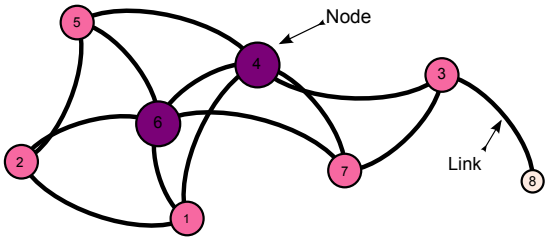


Figure 1.3: A small network composed of eight nodes and thirteen links.

[1]This network diagram and all other network diagrams in this thesis, unless otherwise stated, were produced in Gephi [13].

Another useful way a network can be described is by its *adjacency matrix* A. We define this matrix in the following way

$$A_{ij} = \begin{cases} n & \text{if there is are n links from node } j \text{ to node } i \\ 0 & \text{if node } j \text{ and node } i \text{ are not connected} \end{cases} \tag{1.3}$$

As a result of this definition an undirected network has a symmetric adjacency matrix while a directed network has an asymmetric adjacency matrix. A network with no self-links and at most one link between each pair of nodes has zeros along the diagonal and either ones or zeros elsewhere. For example the adjacency matrix of the network shown in Figure 1.3 is the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.4}$$

We can immediately see that this matrix is indeed symmetric, has entries equal to zero all along its diagonal and all other entries are either zero or one.

One important characteristic of a node that can be easily computed from the network's adjacency matrix is its *degree*. The degree, k_i, of a node, v_i is defined as the number of links it has and can be computed from the adjacency matrix as follows:

$$k_i = \sum_{j=1}^{n} A_{ij}. \tag{1.5}$$

Indeed we can compute a vector $k = (k_1, k_2, \ldots, k_n)^T$ of the degrees of every node in the network through:

$$k = A \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{1.6}$$

In our example we find that

$$k = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 5 \\ 3 \\ 5 \\ 3 \\ 1 \end{pmatrix}, \tag{1.7}$$

which can be easily verified from Figure 1.3.

2.3 Examples of Real-World Networks

Networks are a very useful way of representing connections and interactions between different parts of a system. Networks have been applied by scientists in a wide variety of fields, including mathematics, computer science, physics, biology, economics and social science, to many different problems. Representing a system as a network allows researchers to apply the many tools previously developed for networks to analyse the system and its properties. Here we will present two interesting and well-known examples of networks. The first is a simple social network with no spatial element and the second is a spatial transportation network.

2.3.1 Social Network: The Erdős Collaboration Network

The prolific and influential Hungarian mathematician Paul Erdős (1913-1996) authored over 1,500 mathematical articles in his lifetime, more than any other mathematician in history. He worked on a broad range of topics including number theory, set theory, analysis, probability theory and combinatorics (including network or graph theory). A comprehensive list of Erdős' publications in graph or network theory can be found on the European Mathematical Information Service website [62]. One important contribution by Erdős to the study of networks came during the 1950's when in a series of eight papers he and his fellow Hungarian Alfréd Rényi introduced the most basic of network models, the random network. The Erdős-Rényi random network model $N(n, p)$ generates a network by taking n nodes and placing an edge between any two nodes with the fixed independent probability p. One of the most important questions addressed by Erdős and Rényi was finding the threshold probabilities associated with the appearance of certain properties such as having a subgraph with m nodes and l edges in a given network $N(n, p)$.

Erdős was completely devoted to mathematics and lived a large portion of his life as a mathematical pilgrim, a nomad with no home and no conventional position, travelling to meet collaborators all over the world. Most of his articles were co-written and he was famous for having many collaborators. According to the last count by the Erdős Number Project [60] the number stands at 511. Erdős even dreamed of collaborating with the great mathematicians of the past. He is quoted as saying:

> *"My mother said, 'Even you, Paul, can be in only one place at one time.' Maybe soon I will be relieved of this disadvantage. Maybe, once I've left, I'll be able to be in many places at the same time. Maybe then I'll be able to collaborate with Archimedes and Euclid"* [64].

Erdős's phenomenal output and huge number of collaborators has given rise to the notion of an *Erdős number* and according to the Erdős Number Project the first published source about this idea was a 1969 article by Casper Goffman in The American Mathematical Monthly, volume 76, page 791. One's Erdős number describes one's *collaborative distance* from Erdős based on co-authorship.

Erdős is defined to have an Erdős number of zero. Anyone else's Erdős number is one greater than the lowest Erdős number of one's coauthors. In other words, any of Erdős's direct coauthors are said to have an Erdős number of one. Anyone other than Erdős himself who has collaborated on a paper with someone with an Erdős number of one, but not with Erdős himself, has an Erdős number of two, and so on. If there is no chain of coauthorships connecting someone with Erdős, then that person's Erdős number is said to be infinite.

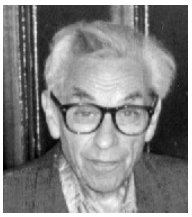


Figure 1.4: Paul Erdős. [42]

Expanding this idea we may construct a collaboration network of all scientific researchers where each node u represents a researcher and two nodes u and v are connected by the link (u, v) if they have published at least one article together. In such a network the Erdős number of the researcher represented by the node u is is given by the minimum number of links between connected nodes one must traverse starting from the node u to arrive at the node p representing Paul Erdős. If no such path exists u does not have a finite Erdős number. Examining the various properties of such networks can help us to understand the nature of relationships between academics.

Figure 1.5 shows a subnetwork of this network containing all of those researchers with an Erdős number of at most 1, i.e. Erdős himself and all of his direct collaborators, using data available online from the Erdős Number Project [60]. The white node at the centre of the network represents Erdős himself. All other nodes in the network represent researchers with an Erdős number of one and are given a size based on their degree. The colour of these nodes is based on their modularity class, i.e. well connected groups of researchers within the network as a whole.

One interesting feature of the topology of this network (and indeed of many social networks) is the high level of clustering observed at 63.5 per cent, i.e. the average fraction of pairs of a researcher's collaborators who have also collaborated with one another, compared to the network's low density of 1.6 per cent, i.e. the fraction of all possible edges present in the network. This reflects the presence of smaller local communities within the network in which a higher than average number of people have collaborated with one another. This level of clustering is particularly high due to the fact that all nodes in the network are by definition connected to the node p representing Erdős. However, removing this node from the network we still see a high level of clustering at 34.6 per cent compared to the density of the network at 1.2 per cent.

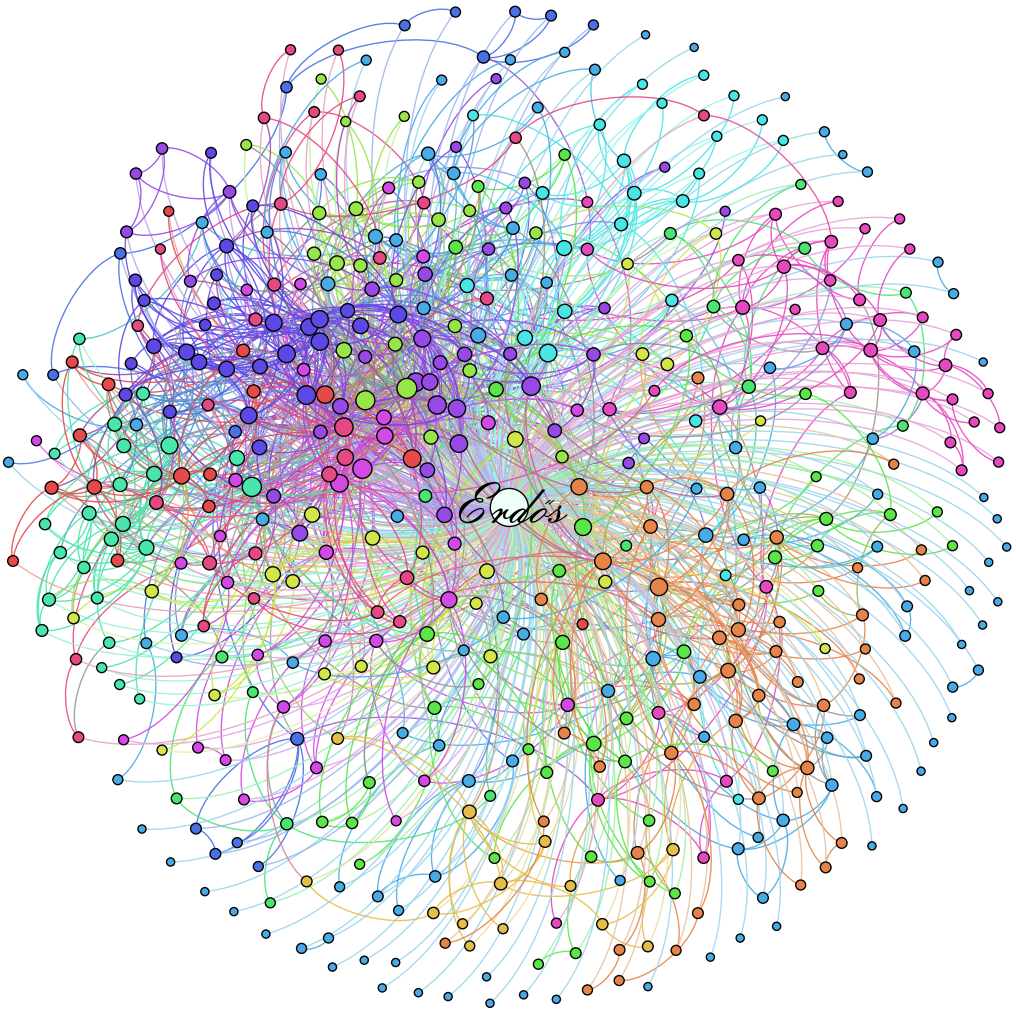


Figure 1.5: The collaboration network of those with an Erdős number of at most 1 constructed using data from [60].

2.3.2 Spatial Network: The US Airline Network

Systems in the real world which can be represented by networks very often have nodes which are embedded in two or three dimensional space. Examples include transportation and distribution networks, telecommunications networks, neural networks and power grids. For networks such as these and many others space is relevant and important as the topology of the network alone does not contain all of the information associated with the system. One important consequence of space on networks is that there is often a cost associated with long edges, this in turn can have dramatic effects on the topological and spatial structure of these networks.

Understanding how spatial constraints affect the structure of networks is important in terms of understanding many real-world systems. Indeed it is an important focus of this thesis. The models of network growth we introduce in Chapter 2 model the growth of networks over time as they react to varying spatial constraints. These networks are developed with the ultimate aim of creating a model of spatial economic development where the relationship between space and the growth of an economic network are naturally of utmost importance in terms of both how spatial constraints influence growth and how spatial considerations affect how our results should be interpreted.

Figures 1.6 and 1.7 show spatial network diagrams of the structure of the airline route network in the United States of America (USA) constructed using data from [63]. Each node in the spatial network represents an airport and its spatial location is fixed at the location of that airport. Two airports are connected by an link in the network if there is an airline route between them. Analysing the topology of this network can help us to better understand how the real-world airline network in the USA functions effectively.

It is clear that the spatial element of the network has had a large impact on its structure. It can be seen in Figure 1.6 that the network has grown to have a hub and spoke like topology with a few well connected hubs that have high degree and a high *centrality*. (Here centrality can be thought of as a measure of importance for nodes, we will return to look at the concept of centrality in more detail in Chapter 3.) The other nodes, i.e. the spokes, have few, usually short, edges and low centrality. These low degree nodes benefit from having connections to hubs, shortening the total distance in terms of edges to other nodes in the system. In other words the airline network is constructed in such a way to reduce the number of connecting flights required for any one journey. Small airports have connections with only a small number of airports (usually nearby), however, it is possible in most cases to get a direct flight from any airport to the nearest hub and continue your journey from there.

Figure 1.7 shows the network partitioned by modularity and reinforces this point. Nodes are coloured by class and sized by their degree. Again we see that the spatial element of the network has a significant impact, with the classes roughly forming regional clusters. Small airports are generally grouped in a community of airports with a major hub that they are connected to and these major hubs serve to connect the various communities (modularity classes).

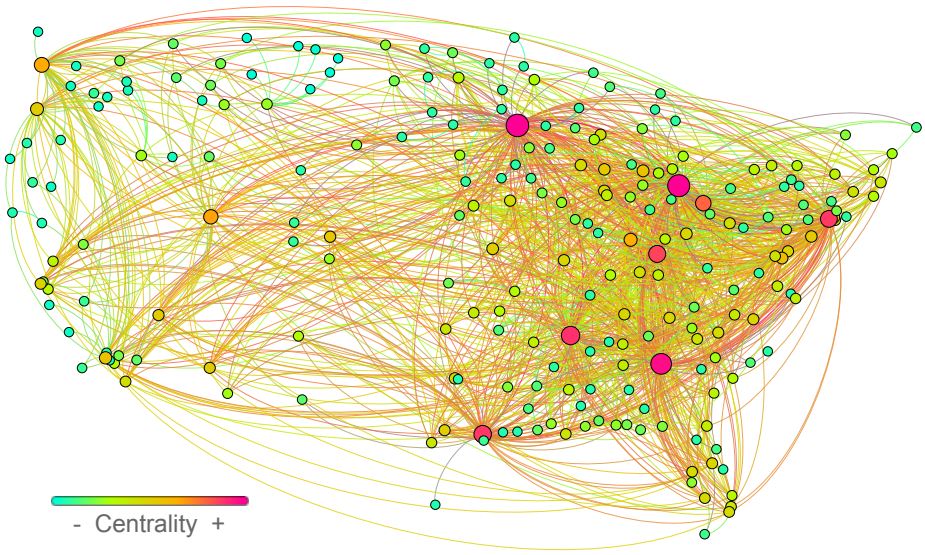


Figure 1.6: The US Airline Network with nodes sized by degree and coloured by eigenvector centrality.

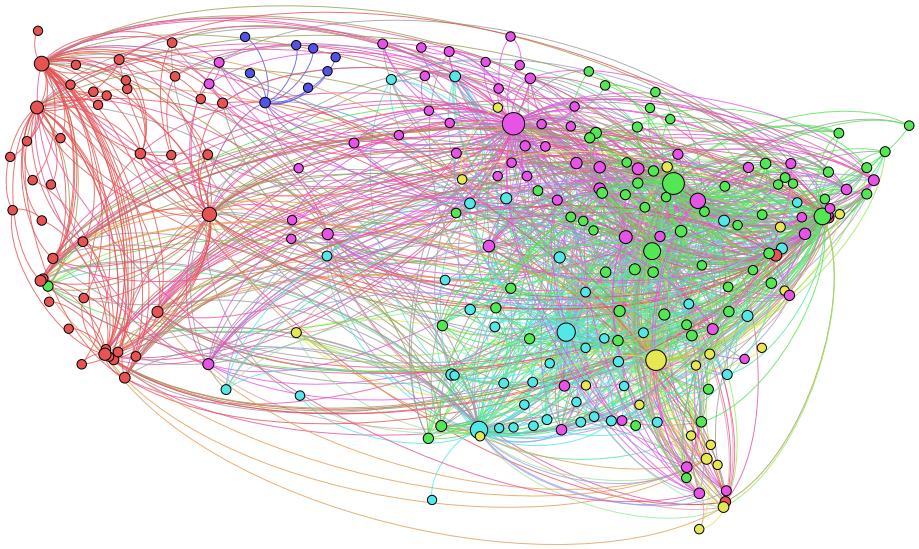


Figure 1.7: The US Airline Network with nodes sized by degree and coloured by modularity class.

Chapter 2

Spatial Network Growth

1 Introduction

In this chapter we introduce a set of algorithms for generating spatial networks. Our ultimate aim is to model uneven spatial economic dynamics using a series of overlapping spatial networks across space and time. The methods introduced and examined in this chapter will form the basis of the economic models introduced in Chapter 4.

Many models of network formation rely only on the network topology of the network in question and have no spatial element. These models are useful for many real-world networks in which spatial distances are not overly important or entirely irrelevant such as the world wide web or citation networks. One example of such a model is the famous *"preferential attachment"* model of Barabási and Albert [6] in which new nodes are more likely to form links with nodes that have a higher degree. This model produces an undirected network with a power-law degree distribution. Power-law degree distributions have been observed in many real-world networks including citation networks, the world wide web and protein-protein interaction networks. The Barabási-Albert model has been cited as an explanation for these observed distributions. It does so by adding nodes one by one, forming links with the existing set of vertices with probability proportional to the nodes' current degree.

In other real-world networks, however, the spatial element cannot be overlooked due to the prohibitive cost or time associated with travelling, communicating or establishing connections over long distances. Space is relevant and the network topology alone does not contain all of the information that is important to the network. Transportation and distribution networks depend quite clearly on distance due to fuel costs, time costs, construction costs etc. Power grids and many communication networks also depend on distance. Regions in the brain that are closer together have a higher probability of being connected due to the material and energy costs associated with longer axons [17].

Spatial networks, therefore, are very important in current problems in many fields ranging from urbanisation to biology and epidemiology. In our analysis we first look at algorithms for the growth of a single spatial network, $N = (V, E)$ over time. These networks consists of a set of vertices or nodes, V, and a set of edges or links, E, which connect them. Each node $v \in V$ is given a fixed spatial location. In one of the models we examine in this chapter individual nodes are also given sizes which may evolve over time by a discrete process. Later we examine a model for the growth of a pair of overlapping spatial networks.

The vast majority of spatial-growth algorithms that have been developed and studied in the past predetermine the position of all nodes before adding connections between these nodes based on the spatial distance between them or choose the position of new nodes independent of the existing network and again have only the connections rely on spatial distances [24, 71]. In reality, these two methods of placing nodes are essentially equivalent. Another approach that has been used is to draw together connected nodes using an *a posteriori* pulling algorithm to create spatial clusters [61]. Our approach is different however. We grow our spatial networks beginning with a single isolated node, the positions

of any further nodes added to the network are not predetermined, rather their position and the links between them are established over time as the network develops. This approach is inspired by the work of Kaiser and Hilgetag [41]. Our models are extensions of their original model which is both simple and elegant and our simulations are run in Matlab [47]. Before introducing those models we will first introduce some previously studied spatial network growth models.

1.1 Examples of Existing Models of Spatial Growth

One example of a previously studied spatial network growth model is an extension of the Barabási-Albert preferential attachment model [6] which expands the original model by taking spatial distances into consideration during link formation. In this model we begin with a small and complete spatial network of m nodes and proceed to add nodes and links to the network as follows:

1. Add a new node v_i at a random position with uniform distribution.
2. The new node v_i forms links with m existing nodes with the probability that one of the links connects v_i to the existing node v_e proportional to

$$P_1 \propto (k_e)^\sigma f(d(v_i, v_e)), \tag{2.1}$$

   where k_e is the degree of v_e, f is a decreasing function and $d(v_i, v_e)$ is the Euclidean distance between the two nodes. The values of $P_1(v_e)$ are calculated such that the probability measure is well defined and in particular $\sum_e P_1(v_e) = 1$.
3. We repeat this process until the desired number of nodes M is reached.

Just like the case where distance is not considered as a factor in edge formation, this model can generate networks with power-law degree distributions which have been observed in many real-world networks. Yook *et al.* [71] showed it was possible to recreate some of the standout features of the internet with this model by choosing $\sigma = 1$ and $f(d(v_i, v_e)) = d(v_i, v_e)^{-1}$.

Another example of a spatial network growth model is the model of Gastner and Newman [30]. This model was proposed for building optimal distribution networks. The spatial distribution of the nodes in the network is set from the outset. We then proceed as follows:

1. Each potential edge between a pair of nodes, v_i and v_j, in the network is assigned an effective length

$$\bar{l}(v_i, v_j) = (1 - \delta) l(v_i, v_j) + \delta, \tag{2.2}$$

   where δ is a constant which allows us to adjust the balance between the fixed and variable costs of forming an edge and $l(v_i, v_j)$ is the shortest Euclidean distance between the two nodes measured along the network's links.

2. Links are then added to the network in order to minimize the sum

$$\sum_{i<j} A_{ij} l(v_i, v_j) + \gamma \sum_{i<j} w_{ij} \bar{l}(v_i, v_j), \tag{2.3}$$

where A is the adjacency matrix of the network, $\gamma > 0$ is a constant and $w_{ij} > 0$ is a weight which represents the traffic between v_i and v_j. This minimization represents an attempt to minimize and balance the construction cost of the network, represented by $\sum_{i<j} A_{ij} l(v_i, v_j)$, and the cost for users to travel across the network, represented by $\sum_{i<j} w_{ij} \bar{l}(v_i, v_j)$.

By adjusting the values of δ and γ in this model, and placing nodes in areas of high population density, the authors were able to use this spatial growth model to produce networks similar to both the highway and airline networks in the United States of America.

1.2 Some Network Metrics

Before we continue, in order to analyse the networks produced by new growth processes introduced and analysed in this chapter, we now introduce four important network metrics for a network $N = (V, E)$. The first of these metrics is particular to spatial networks and allows us to get a quick impression of the spatial concentration of the nodes in the network.

Definition 1.2.1. The *average node distance*, AND, of a spatial network, $N = (V, E)$, is the mean of the Euclidean distances, $d(v_i, v_j)$, between all pairs of nodes $v_i, v_j \in V$.

Our next two metrics give us a measure of the level of connectivity in a network. The *density* of a network gives us a measure of the global connectivity of the network.

Definition 1.2.2. The *density*, D, of a network, $N = (V, E)$, is the percentage of the total possible edges present in the network

$$D = \frac{2|E|}{|V|(|V| - 1)}. \tag{2.4}$$

The *clustering coefficient* of a network, on the other hand, gives us a measure of the local connectivity of the network. It is a measure of how connected the neighbourhoods in a network are.

Definition 1.2.3. The *clustering coefficient* [52], cc, of a node $v_i \in V$ is given by

$$cc(v_i) = \frac{number\ of\ connected\ pairs\ of\ neighbours\ of\ v_i}{number\ of\ pairs\ of\ neighbours\ of\ v_i}. \tag{2.5}$$

The clustering coefficient, C, of a network $N = (V, E)$ is simply the average of $cc(v_i)$ for all $v_i \in V$.

The final metric we will introduce, *average shortest path length*, relies on the concept of a network *path*. It is a measure of the average number of links we need to traverse to get from one node to another in a network.

Definition 1.2.4. A *path* in a network is a sequence of nodes $\{v_1, v_2, \ldots, v_m\}$ such that every consecutive pair of nodes $\{v_i, v_{i+1}\}$ are connected by a link (v_i, v_{i+1}). The *length* of such a path is the number of links traversed along the path. A *shortest path* is a path between two nodes such that no shorter path exists. If no such path exists we say that the path length between the two nodes is infinite.

Definition 1.2.5. The *average shortest path length*, ASP, of a network is the average of the lengths of the shortest paths between every pair of nodes in that network.

1.3 Simple Growth - The Kaiser-Hilgetag Algorithm [41]

The approach of Kaiser and Hilgetag [41] is to begin with some initial connected network. For our purposes we begin with one node at the centre of the unit square. Nodes are then added to the network until we reach the desired number of nodes, M. Connections between nodes are established based on the spatial distance between them. Throughout our analysis in this chapter the distance we consider will be the Euclidean distance. However, our distance metric need not be limited to Euclidean distance nor our network to Euclidean space. We may consider for example measures of similarity such as coexport probabilities which give a measure of the similarity of industrial products [34], an idea we further investigate in Chapter 4.

The following development algorithm describes the process for adding nodes and links to the network:

1. New node position is given by a uniform random variable on the unit square.

2. The new node v_i connects with each existing node v_e with probability

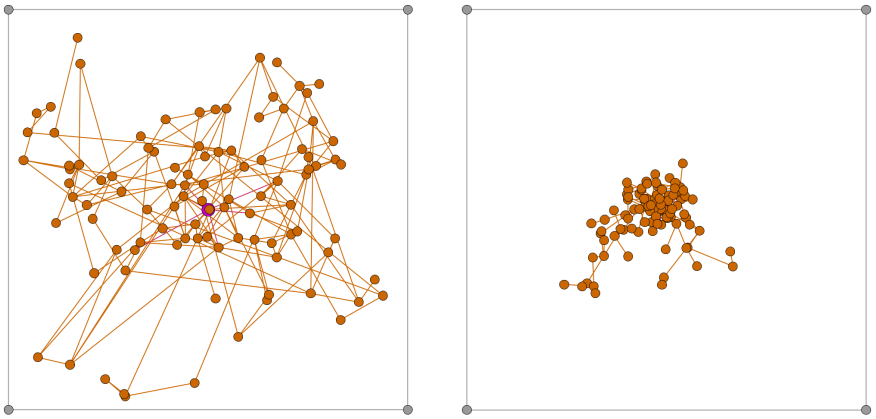
$$P_1 = \beta e^{-\alpha d(v_i, v_e)}, \tag{2.6}$$

   where $d(v_i, v_e)$ is the Euclidean distance between the two nodes, $0 < \beta \leq 1$ and $\alpha > 0$. Here β simply scales the general probability of link formation. Increasing β makes the formation of a link between two nodes in any given spatial locations more likely. The parameter α on the other hand regulates the dependence of link formation on the spatial distance between nodes. Note that the given probability measure is well defined. The formation of each possible link between the new node v_i and each existing node v_e is a separate and independent Bernoulli random variable with probability of success, i.e. link formation, p, where $0 < p = P_1 = \beta e^{-\alpha d(v_i, v_e)} \leq 1$, since $0 < \beta \leq 1$ and $\alpha > 0$, and probability of failure q, where $0 \leq q = 1 - p < 1$.

3. The new node survives, i.e. is kept, if and only if it forms a connection with at least one existing node. If no connections are formed the node is

discarded. This step is particularly important, it makes the survival of new nodes dependent on the spatial layout of the existing nodes.

4. We repeat this process until the desired number of nodes, M, is reached.



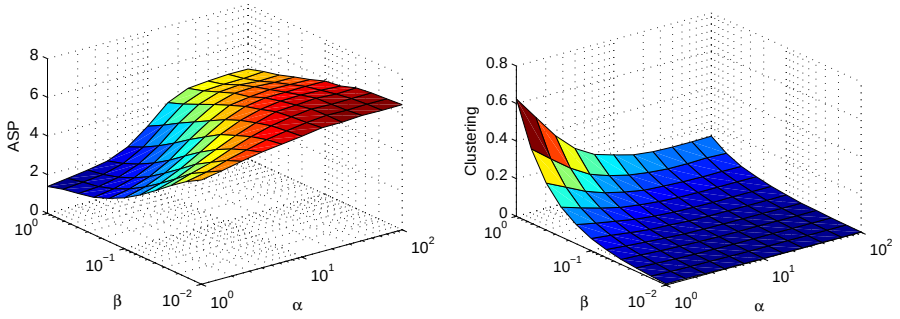
(a) N_1^S: $\alpha = 15$ and $\beta = 0.5$. (b) N_2^S: $\alpha = 50$ and $\beta = 0.2$.

Figure 2.1: Network diagrams of two 100-node spatial networks generated by the algorithm outlined above for different values of α and β. The boundaries of the unit square are shown in grey and the initial node is highlighted in pink.

Figure 2.1 shows two examples of 100-node spatial networks generated by the Kaiser-Hilgetag algorithm. We can see that different choices of α and β can produce spatial networks with significantly different structures and properties. The most striking difference is immediately clear by observation. The higher α value chosen for the network in Figure 2.1b causes a higher level of spatial concentration of nodes and lower mean distance between nodes in the network. This is due to the fact that, according to our algorithm, the probability, $P_1 = \beta e^{-\alpha d(v_i, v_e)}$, of a new node connecting with an existing node decays exponentially with distance scaled by α. This means that in the case of N_2^S new nodes must be closer to the existing network in order to be added for higher values of α. The average node distance of N_1^S is 0.3774 (2.1a) while the average node distance of N_2^S is 0.1233 (2.1b), indicating the higher level of spatial clustering in N_2^S that is clear from observation in Figure 2.1.

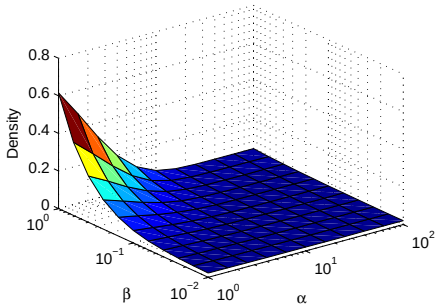
The average shortest path length of N_1^S is 4.4638 while the average shortest path length of N_2^S is higher at 5.3457. This means that on average we must traverse more links to get from one node to another in N_2^S than in N_1^S. The network N_1^S has a higher clustering coefficient than N_2^S, 0.0639 compared to 0.0502, meaning that on average nodes in N_1^S have more complete neighbourhoods than those in N_2^S. Finally, N_1^S also has higher density, 0.032 compared to 0.028, and so has a higher number of edges and degree of connectivity. These differences can be attributed, at least in part, to the higher β value used when generating N_1^S. This causes any possible link to have a higher probability of being formed, leading to more links in the overall network generated.

Figure 2.2 gives an overview of the properties of networks generated using the Kaiser-Hilgetag algorithm with different values for α and β. For each parameter combination we simulated 300 networks and then took the mean value of each network metric over the 300 networks.

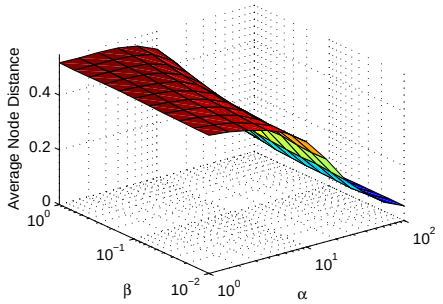


(a) Average Shortest Path Length

(b) Clustering Coefficient



(c) Network Density



(d) Average node distance

Figure 2.2: Comparison of the properties of 100-node spatial networks generated using different values of the parameters α and β. Each data point represents the average for 50 networks.

We see that the density and clustering are highest for low values of α and high values of β, this is unsurprising as this maximises the probability, $P_1 = \beta e^{-\alpha d(v_i,v_e)}$, of any link forming, increasing the number of links in the network and so increasing both local and global connectivity. It is also interesting to note that while density decreases rapidly for all values of β as α increases, clustering remains, relatively speaking, quite high as short links are still formed, increasing local connectivity. Average node distance decreases rapidly with increasing α while average shortest path length is lowest for those networks with high probabilities of edge formation where α is low and β is high.

Kaiser and Hilgetag found that their model was able to generate *small-world* networks, networks which maintain a comparable average shortest path length to random networks while displaying a higher clustering coefficient, which is not thought to be possible in the spatial graph model in which positions are chosen randomly *before* edge formation [70]. They also showed, amongst other things,

that the model was able to produce spatial networks that displayed similar properties to real-world spatial networks such as the German highway system.

We can and will extend this model in what follows to allow for more complicated growth processes. We will investigate what effects these extensions have on the properties of the networks that are generated and how the resulting networks compare to those originally investigated by Kaiser and Hilgetag. These new ideas and processes will then be used when developing our models of spatial economic development in Chapter 4.

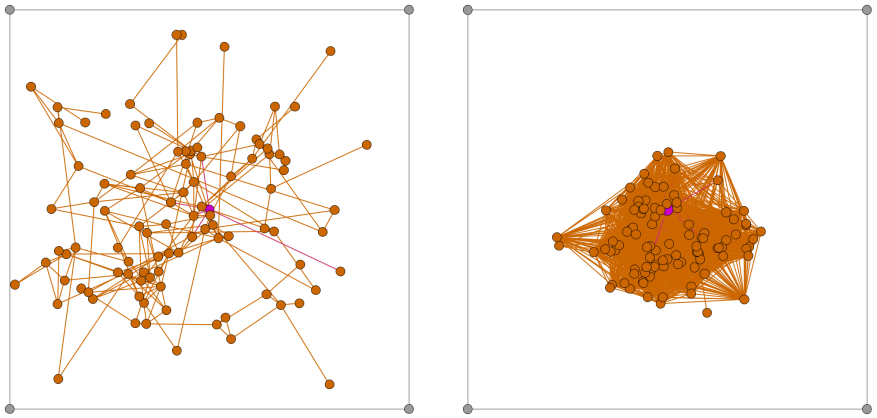
2 Allowing Mature Nodes to Connect

In the development algorithm presented in Section 1.3 the only time at which it was possible for two nodes to form a connection was at the point where one of the nodes was being added to the network for the first time. In this section we introduce a modification to the original algorithm that allows us to deal with this issue. In the extension we allow existing nodes to attempt to form connections with other existing nodes every T timesteps. The probability of establishing such a connection is also based on the distance between the two nodes. In a way this can be thought of as a combination of the approach of Kaiser and Hilgetag, and an approach where the position of our nodes is predetermined. Nodes added do not have a predetermined position, however every T timesteps, i.e. every T iterations of our algorithm, we look to add links to a network of $m < M$ nodes with predetermined positions based on the distance between them.

2.1 Development Algorithm and Sample Networks

We again begin with one node at the centre of the unit square and proceed as follows:

1. New node position is given by a uniform random variable on the unit square.
2. The new node v_i connects with each existing node v_e with probability
$$P_1 = \beta e^{-\alpha d(v_i, v_e)}. \tag{2.7}$$
3. The new node survives, i.e. is kept, if and only if it forms a connection with at least one existing node. If no connections are formed it is discarded.
4. Every T timesteps we allow connections to form between any pair of existing nodes λ, γ with probability
$$P_2 = Be^{-Ad(\lambda,\gamma)}, \tag{2.8}$$
where $0 < B \leq 1$ and $A > 0$.
5. We repeat this process until the desired number of nodes, M, is reached.



(a) N_1^M: $\alpha = 15$, $A = 50$ and $\beta = 0.5$, $B = 0.2$. (b) N_2^M: $\alpha = 50$, $A = 15$ and $\beta = 0.2$, $B = 0.5$.

Figure 2.3: Network diagrams of two 100-node spatial networks generated by the algorithm outlined above with $T = 20$ for different values of α, A, β and B. The boundaries of the unit square are shown in grey and the initial node is highlighted in pink.

Figure 2.3 shows two examples of 100-node spatial networks generated by this network development algorithm. We can see that different choices of parameter values can produce spatial networks with significantly different structures and properties. We have chosen the same values for α and β used for generating the networks N_1^S and N_2^S shown in Figure 2.1 of Section 1.3 which used the original Kaiser-Higetag algorithm.

Allowing for the establishment of connections between mature nodes every $T = 20$ iterations of the algorithm has not made any major impact on the spatial distribution of the nodes in the networks generated. The higher α value chosen for the network in Figure 2.3b again causes a higher level of spatial concentration of nodes and lower mean distance between nodes in the network. The average node distance of N_1^M is 0.3553 while the average node distance of N_2^M is 0.1767. In N_1^S and N_2^S the corresponding values were 0.3774 and 0.1233 respectively.

Allowing for the establishment of connections between mature nodes every $T = 20$ iterations of the algorithm has made a dramatic difference to other network properties in the case of N_2^M. The fact that A is significantly smaller than α for this network is the most prominent reason that these differences exist. Recall that α regulates the dependence of link formation for new nodes on their spatial distance from existing nodes and so a high α in this case results in tightly clustered nodes. On the other hand A for regulates the dependence of link formation between existing mature nodes nodes on the spatial distance between them, a low A in the case of this network means there is a high probability of forming links between mature pairs of nodes in the spatially dense network every T timesteps. The high value of α also means it requires more iterations

of the algorithm to produce a 100-node network. This means that there are more chances for connections between mature nodes to form. The fact that B is greater than β for this network also compounds the effect. On the other hand, the effect of allowing connections to form between mature nodes on the network properties of N_1^M is relatively small as $A > \alpha$ and $B < \beta$.

The average shortest path length of N_1^M is 3.7758 which is slightly shorter than the ASP of N_1^S which was 4.638. On the other hand, the average shortest path length of N_2^M is now lower than N_1^M at 1.3632 which is significantly lower than the corresponding ASP of 5.3457 for N_2^S due to the increased number of links in the network generated by our new algorithm.

The density of N_1^M is 0.0400, this is not a major increase on the density of 0.032 of N_1^S however the clustering coefficient of N_1^M is almost double that of N_1^S, 0.1244 compared to 0.0639. This may be because many of the links added between mature nodes every $T = 20$ timesteps are likely to be between nodes which both have links to the same node. If nodes u_1 an u_2 are attached to the same node v they are likely to be spatially close to eachother as they both needed to be close to v in order to form the initial connection. In this case every $T = 20$ timesteps their proximity means that they will have a relatively high probability of forming a connection.

The effect is more dramatic in the case of N_2^M due to the relative sizes of α & A and β & B. A high α causes nodes to be highly spatially concentrated. A low A then allows mature nodes to form links relatively easily. Both the density an clustering coefficient of N_2^M are significantly higher than the corresponding values for N_2^S. N_2^M has a density of 0.6436 and a clustering coefficient of 0.7755 compared to values of 0.028 and 0.0502 respectively for N_2^S.

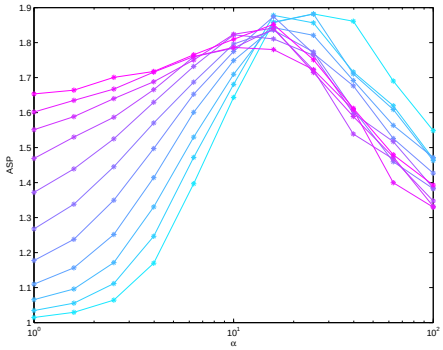
2.2 Results

In this section we investigate the effect of different parameter value combinations on the spatial networks generated by our algorithm outlined in Section 2.1. Our analysis focuses on the changes in the four metrics defined in Section 1.2. For each parameter combination we simulated 50 100-node networks (in a very small number of cases we reduced this to 30 networks for high values of α due to computational constraints). Of particular interest to us are the effects of the interactions between different values of α and β, of changing T, of different ratios of α and A and of different ratios of β and B on the properties of the spatial networks produced by our algorithm. We look first at the interactions between different values of α and β and how allowing for the formation of links between mature nodes every T timesteps affects the properties of the network generated.

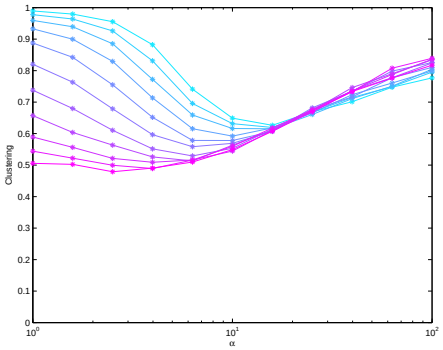
2.2.1 Varying α and β

For simplicity when examining how different combinations of α and β affect the properties of the spatial networks generated by our algorithm we take $A = \alpha$ and $B = \beta$. This means that the probability of forming a link between mature nodes is equal to the probability of forming such a link when the younger of the two nodes was first added to the network. Our approach will be to first examine

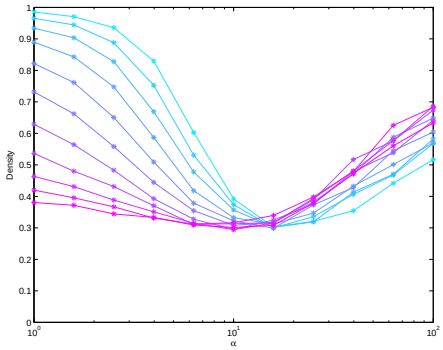
the most extreme case of allowing mature nodes to form connections with other mature nodes. We set $T = 1$, allowing the formation of links between mature nodes every timestep.



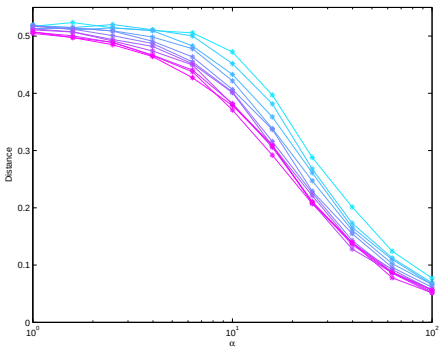
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.



(d) Average node distance of the generated networks for different combinations of α and β.

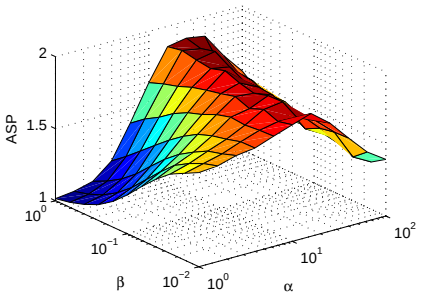
Figure 2.4: Comparison of the properties of networks generated by our algorithm for different values of α and β, allowing connections between mature nodes every $T = 1$ timesteps. We plot in increasingly magenta for lower β.

We take 11 values of α between 1 and 100, 11 values of β between $\frac{1}{100}$ and 1, and generate multiple spatial networks using our algorithm for each parameter combination. In Figure 2.4 and Figure 2.5 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination.

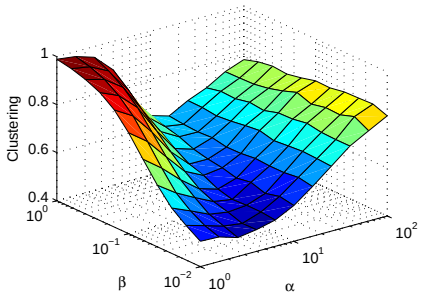
We first note that the relationship between the average node distance of the generated networks and the values of α and β used to generate those networks is strikingly similar to the relationship seen when analysing the original Kaiser-Hilgetag algorithm, which did not allow mature connections, as seen in Figure 2.2. The relationship between the value of β used when generating a network and

the average node distance of that network is positive. Increasing β increases the probability, $P_1 = \beta e^{-\alpha d(v_i,v_e)}$, of a new node connecting with an existing node independent of the distance between them, making new nodes of all distances to the existing network more likely to be added.

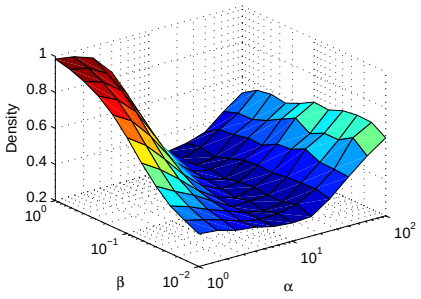
The $\log(\alpha)$ - average node distance relationship is sigmoidal. This relationship is negative, the average node distance of the network decreases as we increase α. This is due to the fact that, according to our algorithm, the probability, $P_1 = \beta e^{-\alpha d(v_i,v_e)}$, of a new node connecting with an existing node decays exponentially with distance scaled by α. This means that as we increase α new nodes must be closer to the existing network in order to be added. The decrease in average node distance is relatively slow as we initially increase α on the log scale this decrease in average node distance becomes more rapid as we increase α before slowing again as $\alpha \to \infty$. The initial slow rate of decrease of average node distance as we increase $\log(\alpha)$ is due to the fact as $\alpha \to 0$ average node distance approaches a limit imposed by the finite space in which our network is located (in this case the unit square). Similarly as $\alpha \to \infty$ average node distance approaches its lower bound 0.



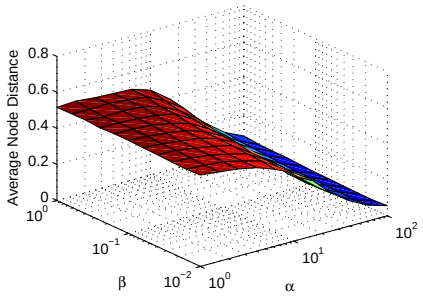
(a) *ASP* of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.



(d) Average node distance of the generated networks for different combinations of α and β.

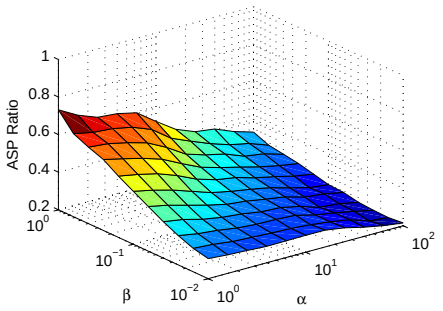
Figure 2.5: Comparison of the properties of networks generated by our algorithm for different values of α and β, allowing connections between mature nodes every $T = 1$ timesteps.

Next we examine the relationship between the ASP of the generated networks and the values of α and β used to generate those networks. For low values of α there is a negative relationship between β and ASP. For large values of α this relationship becomes more complicated and appears to begin to reverse. This is a feature not seen in the networks generated with the original Kaiser-Hilgetag algorithm In those networks the negative relationship between the ASP of the network generated and the value of β used remained negative for all values of α analysed. For all values of β, ASP initially exhibits an increasing relationship with α before reaching some maximum and falling again as $\alpha \to 100$. This is another feature of the relationship not seen in the networks generated with the original Kaiser-Hilgetag algorithm where the relationship between ASP and α was strictly increasing for all values of β.

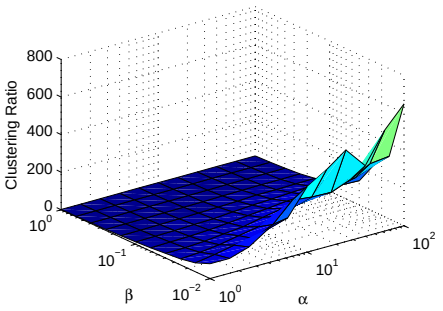
Both the clustering coefficient and the density, our measures of local and global connectivity, respectively, of the networks generated have similar relationships with the values of the parameters α and β used to generate them. For low values of α there is a positive relationship between β and clustering as well as between β and density. For large values of α these relationships becomes more complicated and appear to begin to reverse. Again these features are not seen in the networks generated with the original Kaiser-Hilgetag algorithm In those networks the positive relationship between the clustering and density of the network generated and the value of β used was positive for all values of α analysed. For all values of β, clustering and density initially exhibit a decreasing relationship with α before reaching some minimum and increasing again as $\alpha \to 100$. These are also features of the relationships not seen in the networks generated with the original Kaiser-Hilgetag algorithm where the relationships were strictly decreasing.

For low values of α, decreasing β has the effect of reducing the probability that any two nodes will connect either when one node is initially added to the system or later when the two try to connect as mature nodes. This reduces the number of links in the network and as a result likely increases the average number of links that must be traversed to get from one node to another, i.e. the network's ASP. However, this effect is somewhat offset by the increased number opportunities given to mature nodes to connect as a result of the increased number of iterations of our algorithm that are required to produce a network of M nodes due to the decreased β value. This effect becomes more pronounced as α increases which also results in an increase in the number of iterations required to produce a network of M nodes. For a sufficiently high α the relationship between β and ASP appears to begin to reverse as a result.

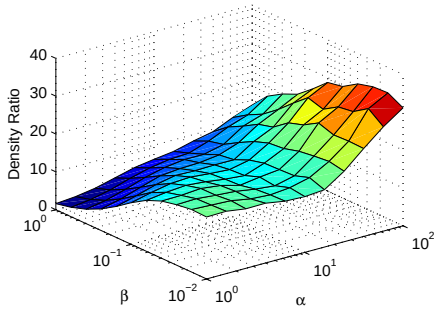
This also explains the changes in the relationships between β and the clustering coefficient of the network generated and the relationship between β and the density of the network generated. Both the clustering coefficient and the density of a network are likely to increase as more links are added to the network.



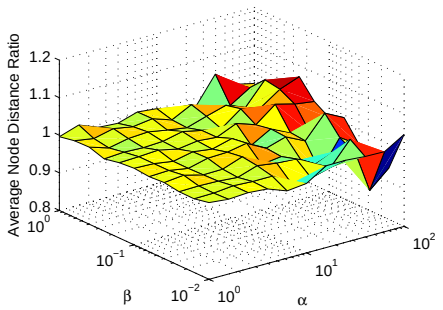
(a) Ratio of the ASP of the generated networks for different combinations of α and β to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and β to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of α and β to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(d) Ratio of the average node distance of the generated networks for different combinations of α and β to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.6: Ratios of the properties of networks generated by our algorithm for different values of α and β allowing connections between mature nodes every $T = 1$ timesteps to the same properties for networks generated by the Kaiser-Hilgetag algorithm

A similar explanation allows us to understand the novel relationship between α and the three network metrics in question. As α increases the likelihood of forming a link with a new node distant to the existing network, or between two distant mature nodes decreases. This initially results in an increase in ASP and a fall in density and clustering as the number of links in the network falls. As α increases further the increase in the number of iterations required to produce a network of M nodes causes the formation of mature links becomes more influential on the properties of the networks generated. Not only does the increase in the number of iterations increase the number of chances for new links to form but the likelihood of mature links forming is higher than a link

from a new node to an existing one due to the fact that nodes in the network are already highly spatially concentrated while new nodes are given random positions in space that may be distant to the existing network. In this way we may also explain why clustering increases faster than density as α increases. Mature links are more likely to form between neighbours than the average pair of nodes in the network they are likely to be spatially close to eachother as they both needed to be close to the node they are both connected to v in order to form the initial connection.

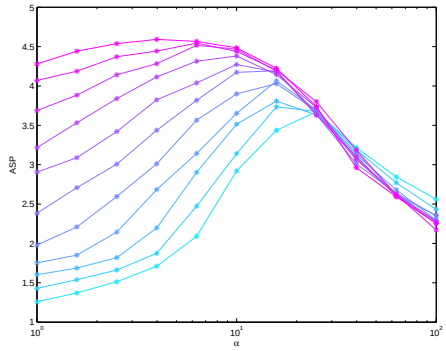
In order to compare the networks generated by our new algorithm, which allows the formation of links between mature nodes every T iterations, to the networks generated by the original Kaiser-Hilgetag algorithm, which allows links to form between nodes only when one of the nodes is first being added to the network, we have plotted the ratios of the value of each of the four network metrics for a network generated by the new algorithm to the corresponding value of that metric for a network generated by the original algorithm in Figure 2.6. We first note that this extension of the original algorithm has no effect on the average node distance of the spatial network produced. For all values of α and β considered the ratio of the average node distance of networks produced by the extended algorithm to the average node distance of networks produced by the original algorithm ≈ 1. This is natural as the spatial location of each node is entirely determined when it is first added to the network and is not influenced in any way by the addition of links later on.

For the other three metrics we note that the extension to the model has the largest effect (our ratio is furthest from 1) when α is large and β is small. This is, as discussed earlier, due to the fact $P_1 = \beta e^{-\alpha d(v_i, v_e)}$ is minimised by maximising α and minimizing β. This causes us to require more iterations of our algorithm per node added, and as a result more opportunities for mature nodes to form links with other mature nodes. On the other hand the effect is smallest for small α and large β.

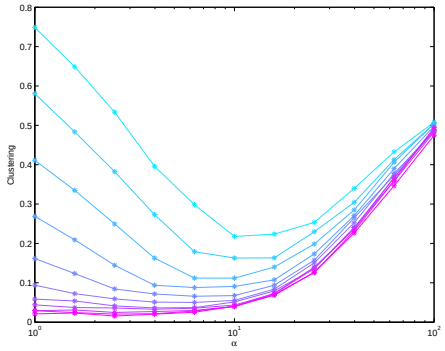
For every value of α and β the extension of the algorithm results in more links being present in the generated network than the corresponding network generated by the original algorithm. This results in a lower ASP and so the ASP ratio is less than 1 for all values of α and β. It also causes the density and clustering coefficient of the resulting network to be higher and so these ratios are both greater than 1 for all values of α and β. The effect is more pronounced in the case of the clustering coefficient. Local connectivity, in both a spatial and topological sense is increased as both initial connections and connections between mature nodes are more likely to form over short spatial distances.

We now examine a less extreme case in which links between mature nodes are allowed to form but only on selected timesteps. We set $T = 100$, allowing the formation of links between mature nodes every 100 timesteps, and look at how different combinations of α and β affect the properties of the spatial networks generated by our algorithm. For simplicity we once again take $A = \alpha$ and $B = \beta$. We take the same 11 values of α between 1 and 100 and 11 values of β between $\frac{1}{100}$ and 1 as above, and again generate multiple spatial networks using our algorithm for each parameter combination. In Figure 2.7 and Figure

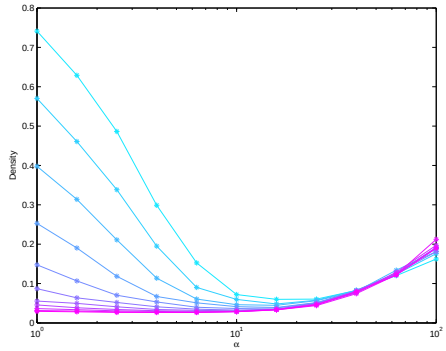
2.8 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination.



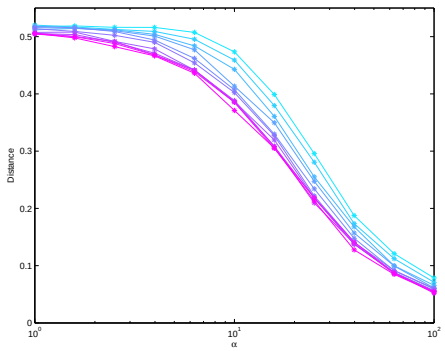
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.



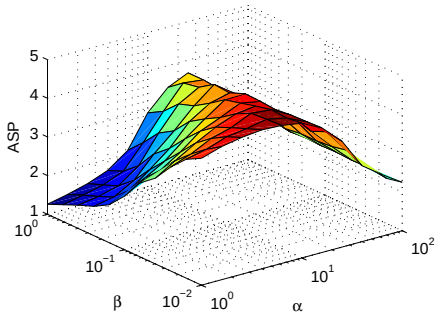
(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.7: Comparison of the properties of networks generated by our algorithm for different values of α and β, allowing connections between mature nodes every $T = 100$ timesteps. We plot in increasingly magenta for lower β.

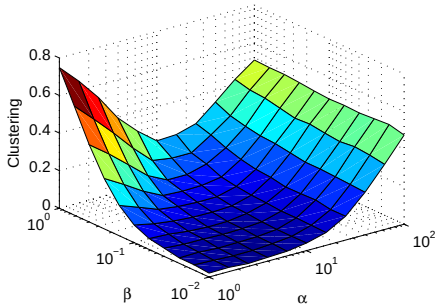
As we saw in Figure 2.4 the relationship between the average node distance of the generated networks and the values of α and β used to generate those networks is once again essentially identical to the relationship seen when analysing the original Kaiser-Hilgetag algorithm, which did not allow mature connections, as seen in Figure 2.2. This is again expected as the spatial distribution of the nodes in any network generated by our extended algorithm depends in no way on the formation of links between pairs of mature nodes.

When examining the relationship between the ASP, clustering coefficient and density of the generated networks and the values of α and β used to generate those networks we again see some of the novel features that were seen for the $T =$

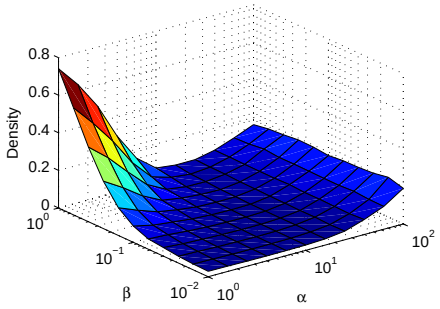
1 case analysed above that are not seen when we prohibit pairs of mature nodes from forming links at any point. The ASP again increases to some maximum as we increase α for all values of β before decreasing. Both the clustering coefficient and the density decrease to some minimum before increasing again as we increase α. The mechanisms by which this occurs is the same as in the $T = 1$ case although the effect is less pronounced as we don't allow mature nodes to form links with other mature nodes as often.



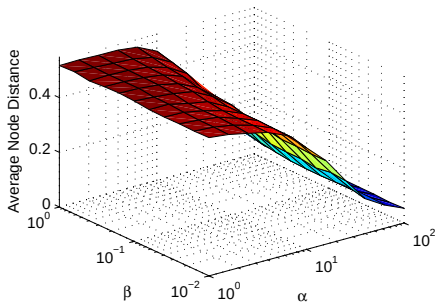
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



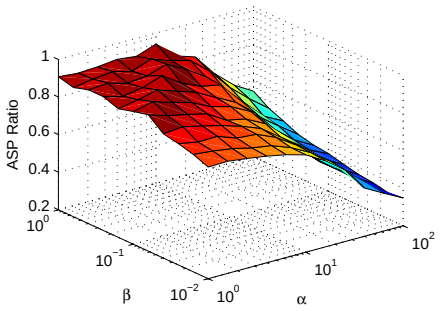
(c) Density of the generated networks for different combinations of α and β.



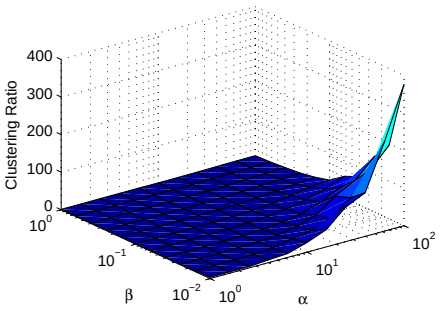
(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.8: Comparison of the properties of networks generated by our algorithm for different values of α and β, allowing connections between mature nodes every $T = 100$ timesteps.

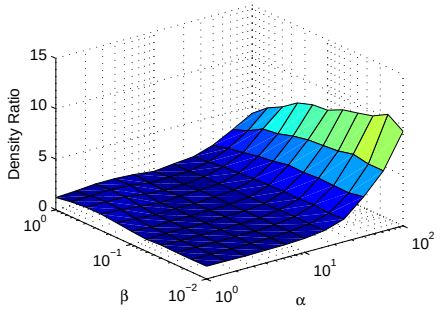
For low values of α there is a negative relationship between β and ASP. For large values of α this relationship becomes more complicated and appears to begin to reverse however this is not as clear as in the $T = 1$ case as there are not as many chances for links between pairs of mature nodes to form. We hypothesise that increasing α further would see a clearer reversal of this relationship as the effect of adding mature nodes becomes more pronounced with the increased opportunities to connect that arise as a result.



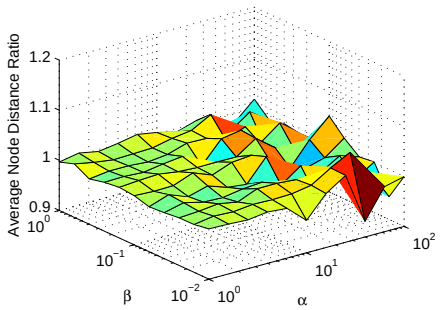
(a) Ratio of the ASP of the generated networks for different combinations of α and β to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and β to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of α and β to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(d) Ratio of the average node distance of the generated networks for different combinations of α and β to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

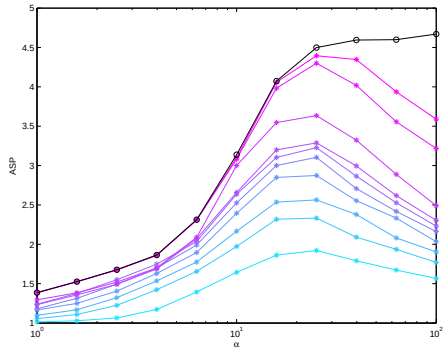
Figure 2.9: Ratios of the properties of networks generated by our algorithm for different values of α and β allowing connections between mature nodes every $T = 100$ timesteps to the same properties for networks generated by the Kaiser-Hilgetag algorithm

Again, as in the $T = 1$ case, the clustering coefficient and the density, our measures of local and global connectivity respectively, of the networks generated have similar relationships with the values of the parameters α and β used to generate them. For low values of α there is a positive relationship between β and clustering and between β and density. For large values of α these relationships becomes more complicated and again appear to begin to reverse although the reversal is not as clear or complete as the $T = 1$ case. we believe this reversal would again become clearer with further increases in α.

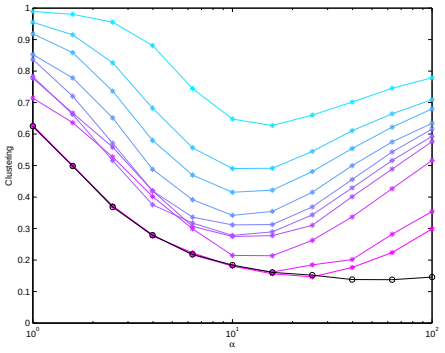
Examining the ratios in Figure 2.9 we again find confirmation that allowing links to form between pairs of mature nodes does not affect the spatial distri-

bution of nodes in the network. The ratio of average node distance in networks generated by our algorithm to networks generated by the original algorithm ≈ 1. The effect of allowing connections between mature nodes every $T = 100$ timesteps, as in the $T = 1$ case, is most pronounced for large α and small β. At these points the ASP, clustering coefficient and density ratios are furthest from 1. Allowing the possibility of mature connections either does not affect the network topology or increases clustering and density and decreases ASP for all combinations of α and β. The clustering coefficient is more affected than density as connections between mature nodes are always more likely between nodes which are spatially close which are in turn likely to be neighbours.

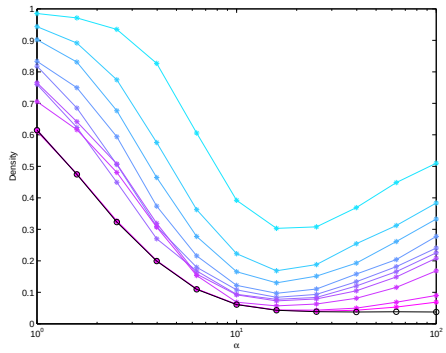
2.2.2 Varying α and T



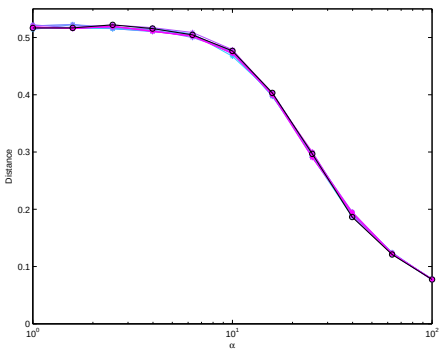
(a) *ASP* of the generated networks for different combinations of α and T.



(b) Clustering coefficient of the generated networks for different combinations of α and T.



(c) Density of the generated networks for different combinations of α and T.



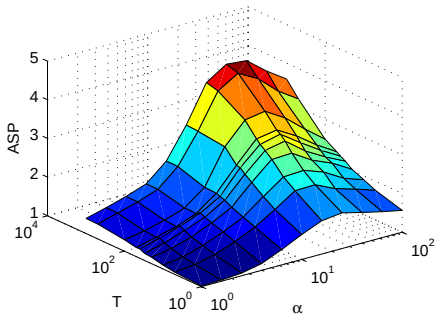
(d) Average node distance of the generated networks for different combinations of α and T.

Figure 2.10: Comparison of the properties of networks generated by our algorithm for different values of α and T. We set $\beta = 1$ and plot in increasingly magenta for higher values of T. The relationship between each of the metrics and α for the original Kaiser-Hilgetag algorithm is plotted in black.

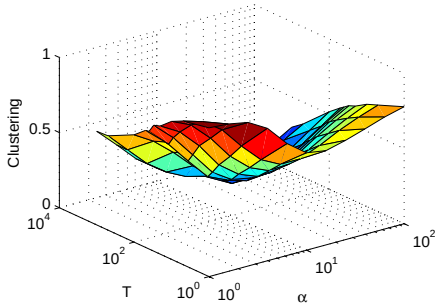
In this section we examine how different combinations of the parameter α, which affects the spatial range of links, and the parameter T, which regulates how often mature nodes may attempt to form links with other mature nodes, affect the properties of the spatial network generated by our algorithm. Throughout this analysis we will hold β constant, taking $\beta = 1$. As in Section 2.2.1 we will, for simplicity, take $A = \alpha$ and $B = \beta$, making the probability of forming a link between mature nodes equal to the probability of forming such a link when the younger of the two nodes was first added to the network.

We take 11 values of α between 1 and 100, 10 values of T between 1 and

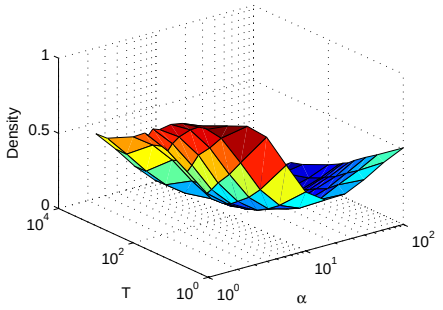
1000, and generate multiple spatial networks using our algorithm for each parameter combination. In Figure 2.10 and Figure 2.11 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination.



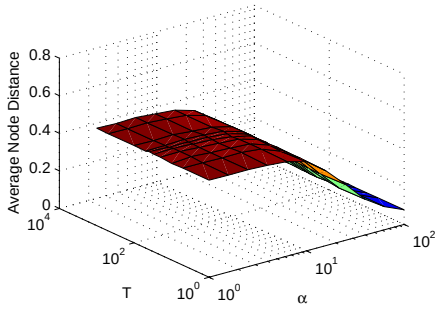
(a) ASP of the generated networks for different combinations of α and T.



(b) Clustering coefficient of the generated networks for different combinations of α and T.



(c) Density of the generated networks for different combinations of α and T.



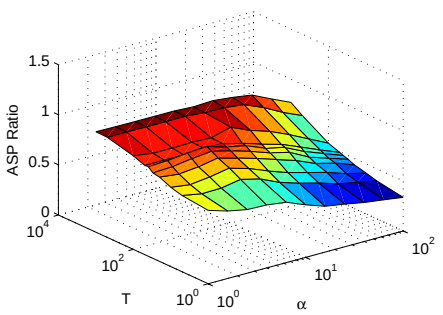
(d) Average node distance of the generated networks for different combinations of α and T.

Figure 2.11: Comparison of the properties of networks generated by our algorithm for different values of α and T when $\beta = 1$.

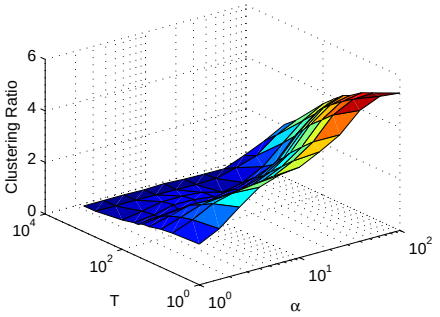
For large values of T and low values of α we see that allowing mature connections has no influence on the topology of the network generated. This is due to the fact that for low values of α a 100-node spatial network is generated by relatively few iterations of the network generation algorithm. This prevents nodes from having many chances to develop connections between pairs of mature nodes. However, provided T is not too large or α is not too small we see that allowing mature connections has a significant influence on the topology of the network generated.

Unsurprisingly however, in Figure 2.10d we once again see that the relationship between the mean distance between nodes in the generated network and α is entirely independent of our choice of T. The spatial position of new

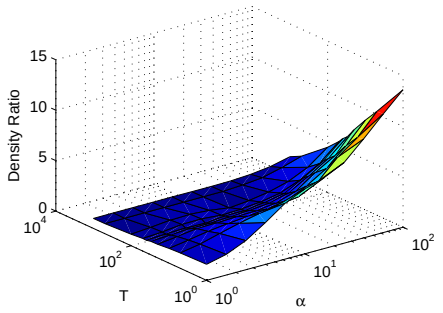
nodes relative to the existing network does not depend in any way, directly or indirectly on T.



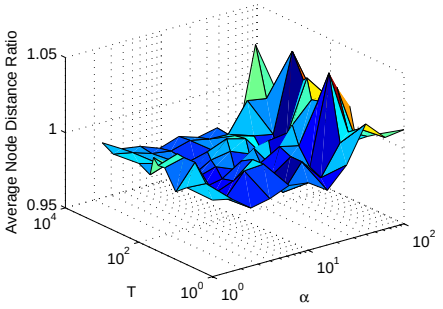
(a) Ratio of the ASP of the generated networks for different combinations of α and T where $\beta = 1$ to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and T where $\beta = 1$ to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of α and T where $\beta = 1$ to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(d) Ratio of the average node distance of the generated networks for different combinations of α and T where $\beta = 1$ to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.12: Ratios of the properties of networks generated by our algorithm for different values of α and T when $\beta = 1$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

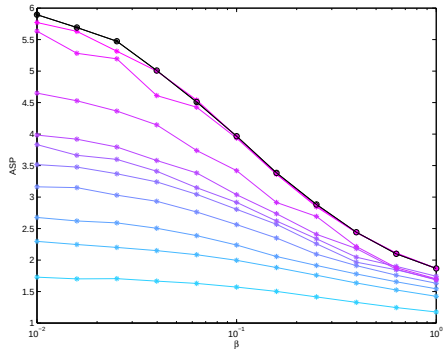
For all values of T analysed we again see the novel relationships between α and ASP, clustering and density observed in Section 2.2.1 that were not seen in the analysis of the original algorithm when links between pairs of mature nodes were prohibited. As we increase α ASP increases to a maximum before decreasing while both clustering and density decrease to minima before increas-

ing. The maximum attained by ASP is lowest and the minima attained by clustering and density are highest when $T = 1$ and pairs of mature nodes have the maximum possible number of opportunities to form links. The minima of density and clustering are achieved at increasing values of α as T increases. Interestingly the value of α for which the maximum ASP is achieved appears to remain approximately equal as T is increased. Further analysis would be required to confirm this.

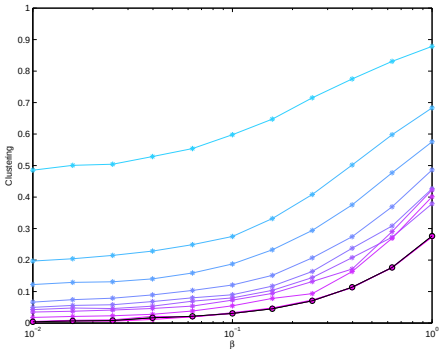
In order to compare the networks generated by our new algorithm to the networks generated by the original Kaiser-Hilgetag algorithm, which allows the formation of links between mature nodes every T iterations, we have plotted the ratios of the value of each of the four network metrics to the corresponding value of that metric for a network generated by the original algorithm, which allows links to form between nodes only when one of the nodes is first being added to the network, in Figure 2.12.

We confirm once again that our extension does not affect the spatial distribution of the nodes in the networks generated. We also note that for the highest values of T and low to intermediate values of α the extension has no effect on the topology of the network. This is indicated by ASP, clustering coefficient and density ratios equal to 1 and is due to the fact that for low α we require few iterations to generate a 100-node network and so if T is high pairs of mature nodes will have little or no opportunities to attempt to form links. In the case of all three metrics we find that the effect of the extension to the original algorithm as the maximum effect (we observe ratios furthest from 1) when α is maximised and T is minimised. In all cases the extension has either no effect or increases density and clustering while decreasing ASP (ratio > 1 and ratio < 1 respectively).

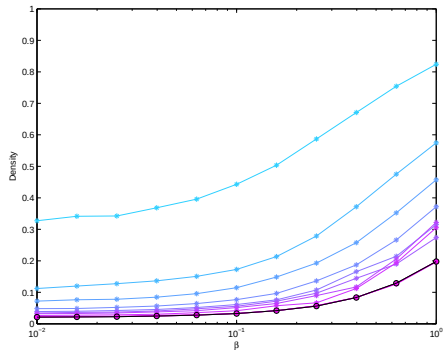
2.2.3 Varying β and T



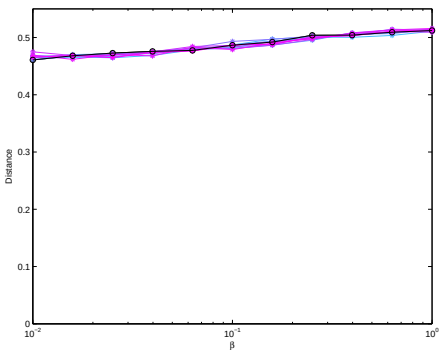
(a) *ASP* of the generated networks for different combinations of β and T.



(b) Clustering coefficient of the generated networks for different combinations of β and T.



(c) Density of the generated networks for different combinations of β and T.

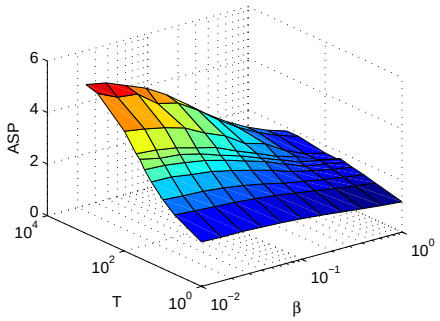


(d) Average node distance of the generated networks for different combinations of β and T.

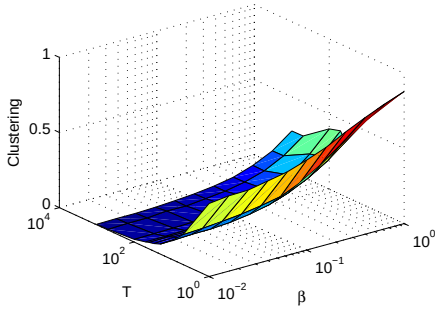
Figure 2.13: Comparison of the properties of networks generated by our algorithm for different values of β and T. We set $\alpha = 4$ and plot in increasingly magenta for higher T. The relationship between each of the metrics and β for the original Kaiser-Hilgetag algorithm is plotted in black.

In this section we examine how different combinations of the parameter β, which affects the probability of any given link forming, and the parameter T, which regulates how often mature nodes may attempt to form links with other mature nodes, affect the properties of the spatial network generated by our algorithm. Throughout this analysis we will hold α constant, taking $\alpha = 4$. As in Section 2.2.1 we will, for simplicity, take $A = \alpha$ and $B = \beta$, making the probability of forming a link between mature nodes equal to the probability of forming such a link when the younger of the two nodes was first added to the network.

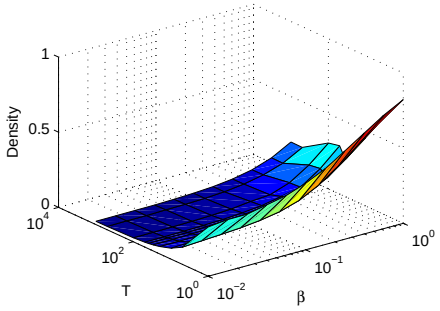
We take 11 values of β between $\frac{1}{100}$ and 1, 10 values of T between 1 and 1000, and generate multiple spatial networks using our algorithm for each parameter combination. In Figures 2.13 and 2.14 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination.



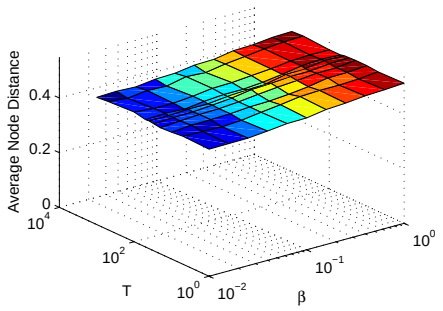
(a) ASP of the generated networks for different combinations of β and T.



(b) Clustering coefficient of the generated networks for different combinations of β and T.



(c) Density of the generated networks for different combinations of β and T.



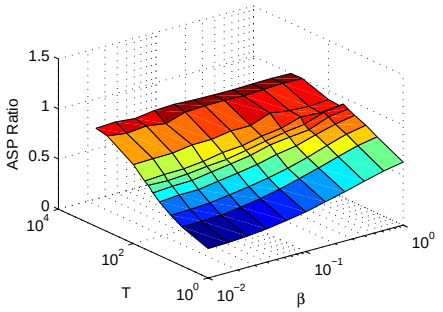
(d) Average node distance of the generated networks for different combinations of β and T.

Figure 2.14: Comparison of the properties of networks generated by our algorithm for different values of β and T when set $\alpha = 4$.

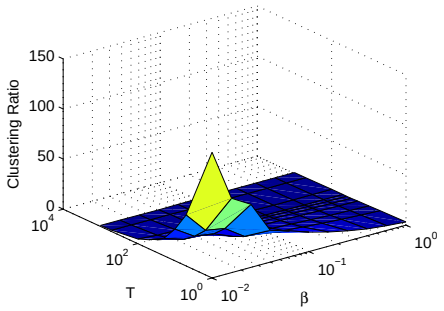
For large values of T and β we see that allowing mature connections has no influence on the topology of the network generated. For large values of β a 100-node spatial network is generated by relatively few iterations of the network generation algorithm as all new nodes have a relatively high probability of forming at least one link with he existing network. This prevents nodes from having many chances to develop connections between pairs of mature nodes. For smaller values of β and T we see that allowing mature connections does have a significant influence on the topology of the network generated. The spatial distributions of the nodes in the network on the other hand is not affected.

For $\alpha = 4$, the relationship between β and the ASP of the network generated

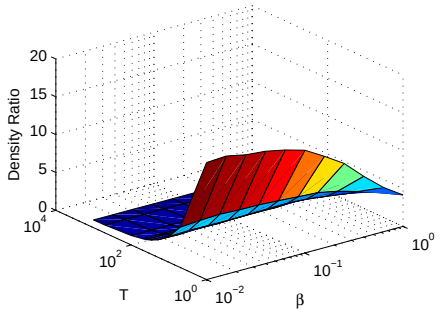
by our algorithm is negative for all choices of T. The relationships between β and the clustering coefficient of the network generated by our algorithm and between β and the density of the network generated are positive for all values of T. We also observe a positive relationship between ASP and T and negative relationships between clustering and T and density and T.



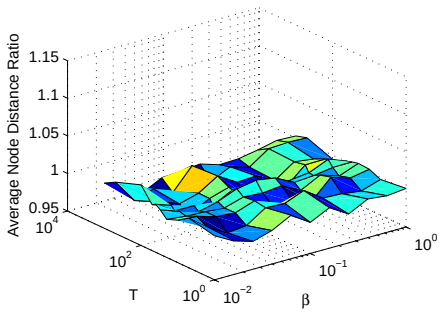
(a) Ratio of the ASP of the generated networks for different combinations of β and T where $\alpha = 4$ to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of β and T where $\alpha = 4$ to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of β and T where $\alpha = 4$ to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



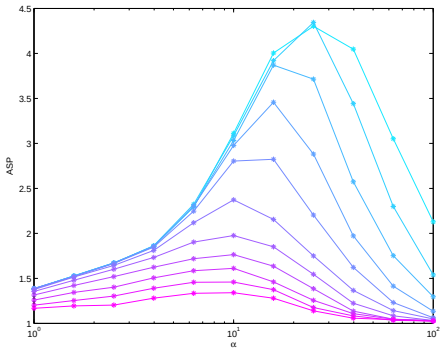
(d) Ratio of the average node distance of the generated networks for different combinations of β and T where $\alpha = 4$ to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.15: Ratios of the properties of networks generated by our algorithm for different values of β and T when $\alpha = 4$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

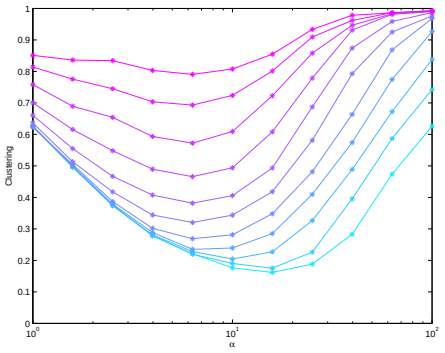
The ratios plotted in Figure 2.15 allow us to compare the properties of networks generated by our new algorithm, which allows connections to form

between pairs of mature nodes to the properties of networks generated by the original algorithm in which this type of behaviour is prohibited. We see that our extension has the largest effect on he topology of the networks generated for low values of T and low values of β as in this case we maximise the opportunities for connections to form between mature nodes. For high values of T and β there is no effect and the ASP, clustering coefficient and density ratios ≈ 1 in these cases. Clustering and density are increased by the extension while ASP is reduced.

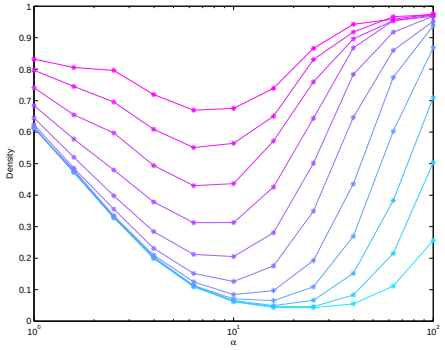
2.2.4 Varying α and A



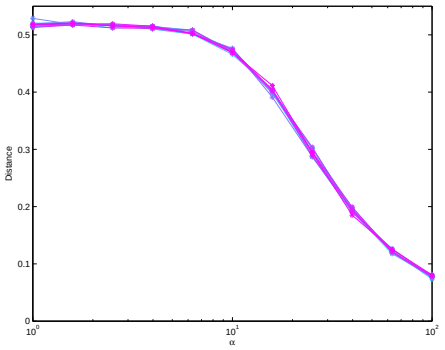
(a) ASP of the generated networks for different combinations of α and A.



(b) Clustering coefficient of the generated networks for different combinations of α and A.



(c) Density of the generated networks for different combinations of α and A.



(d) Average node distance of the generated networks for different combinations of α and A.

Figure 2.16: Comparison of the properties of networks generated by our algorithm for different values of α and A. We set $\beta = B = 1$, $T = 20$ and plot in increasingly magenta for lower A.

In our analysis so far we have limited our investigation to cases where $A = \alpha$ and $B = \beta$. In this section we investigate how the relationship between our two spatial range parameters α and A affect the topology of the networks generated by our algorithm. In particular we are interested in how varying the ratio between α and A affects each of our four spatial network metrics. For simplicity we consider β, B and T to be a constant for all the networks we generate. We set $\beta = B = 1$ and $T = 20$.

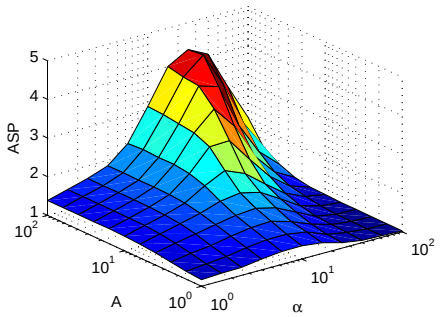
We take 11 values of α between 1 and 100, the same 11 values for A, and generate multiple spatial networks using our algorithm for each parameter combination. In Figures 2.16 and 2.17 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination. The average node distance is unaffected by changes in A as this parameter influences only the formation of links between pairs of mature nodes and not the placement of new nodes. Average node distance retains its decreasing sigmoidal relationship with $\log(\alpha)$ that we have observed throughout our analysis of this algorithm and the original algorithm.

Holding A constant as we increase α the ASP of the network generated by our algorithm increases to some maximum before decreasing. This relationship is significantly different from the positive relationship between α and the ASP of networks generated by the original algorithm which did not allow new links to form between pairs of mature nodes. However, it is similar to the relationship seen in Section 2.2.1. In that case we took $A = \alpha$ and also observed that as we increased α (and therefore A) the ASP initially increased to some maximum before decreasing. We also see that holding A constant as we increase α both the clustering coefficient and the density of the network initially decrease before reaching some minimum and increasing thereafter. Again these relationships differ significantly from the relationships between α and clustering and density for the original algorithm but bear some resemblance to the relationships between α and clustering and density for the algorithm allowing links between mature nodes when we take $A = \alpha$.

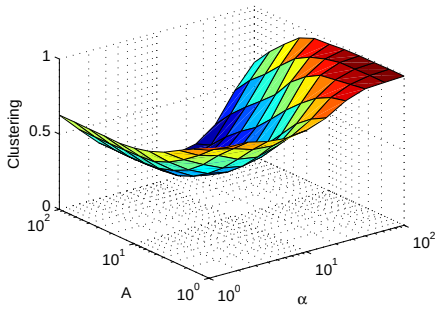
We can explain these novel relationships in a similar way to the explanation given for the corresponding relationships when we set $A = \alpha$. Initially as α increases the likelihood of forming a link with a new node distant to the existing network, or between two distant mature nodes decreases. This initially results in an increase in ASP and a fall in density and clustering as the number of links in the network falls. As α increases further the increase in the number of the formation of links between mature nodes becomes more influential on the properties of the networks generated. There are two main factors at play that cause the formation of these links to becomes more influential. The first of these factors is the increased number of iterations required to generate a 100-node network which gives pairs mature nodes more opportunities to form links. The second is the increase in the ratio of α to A, this results in the nodes in the network generated to be more spatially concentrated but since we hold A constant the spatial constraint on the formation of links between mature nodes does not change resulting these links becoming more likely to form. These added links between mature nodes increase the overall number of links in the resulting

network and for a sufficiently high value of α sees the ASP of the network begin to reduce and the density and clustering increase.

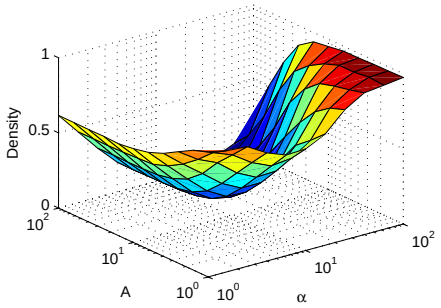
Decreasing A results in a higher $\alpha : A$ ratio and increases the effect. We see that for all values of α a decrease in A results in a lower ASP, a higher clustering coefficient and a higher density for the network generated. In particular we see that when $A = 1$ we see that as α nears 100 the effect is so strong that almost all possible links in the network are formed, density and clustering near their upper bound 1 and ASP nears its lower bound 1 as a result.



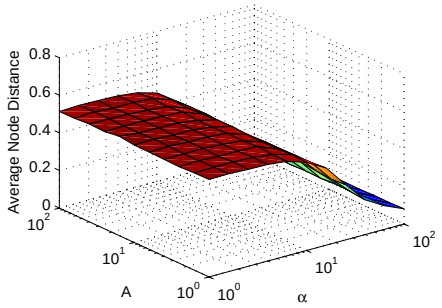
(a) ASP of the generated networks for different combinations of α and A.



(b) Clustering coefficient of the generated networks for different combinations of α and A.



(c) Density of the generated networks for different combinations of α and A.

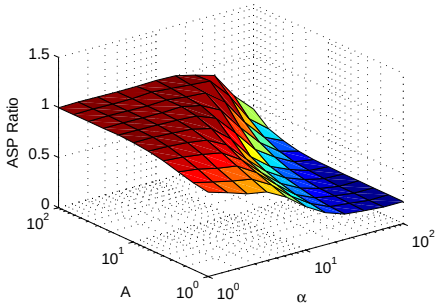


(d) Average node distance of the generated networks for different combinations of α and A.

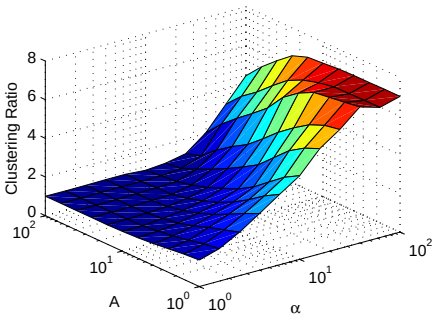
Figure 2.17: Comparison of the properties of networks generated by our algorithm for different values of α and A where $\beta = B = 1$ and $T = 20$.

The ratios plotted in Figure 2.18 allow us to compare the properties of networks generated by our new algorithm to the properties of networks generated by the original algorithm for different combinations of the parameters α and A. We see that when A is large and α is small, i.e. when the $\alpha : A$ ratio is small, the extension has no effect on the topology of the networks generated (ASP, clustering and density ratios equal to 1). On the other hand we see that our extension has the largest effect on he topology of the networks generated for large

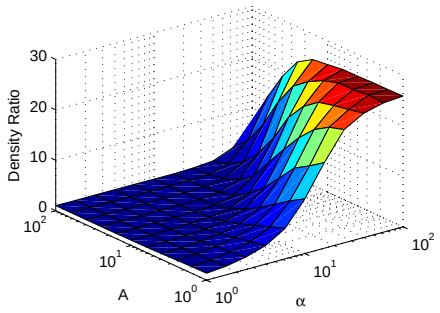
values of α and low values of A as in this case we maximise the opportunities for connections to form between mature nodes, we spatially concentrate the nodes by means of high α and we make it very probable for mature nodes which are spatially concentrated to form links by having low A. Where there is an effect, clustering and density are increased by the extension while ASP is reduced.



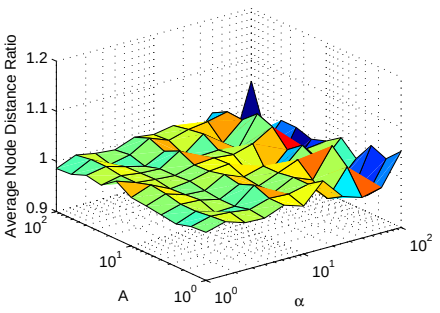
(a) Ratio of the ASP of the generated networks for different combinations of α and A where $\beta = B = 1$ and $T = 20$ to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and A where $\beta = B = 1$ and $T = 20$ to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of α and A where $\beta = B = 1$ and $T = 20$ to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

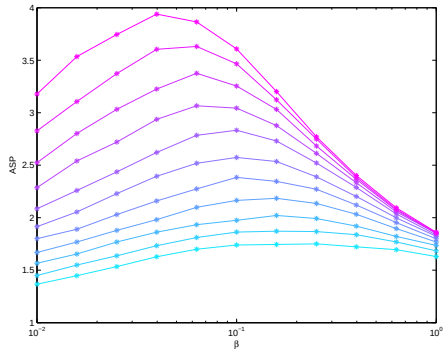


(d) Ratio of the average node distance of the generated networks for different combinations of α and A where $\beta = B = 1$ and $T = 20$ to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

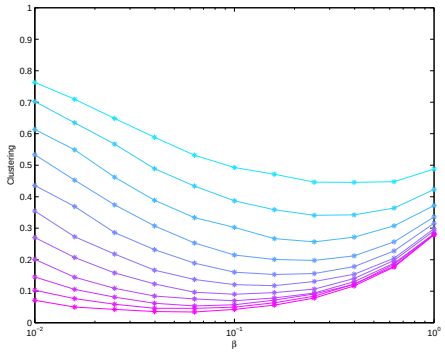
Figure 2.18: Ratios of the properties of networks generated by our algorithm for different values of α and A when $\beta = B = 1$ and $T = 20$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

2.2.5 $\beta : B$ ratio

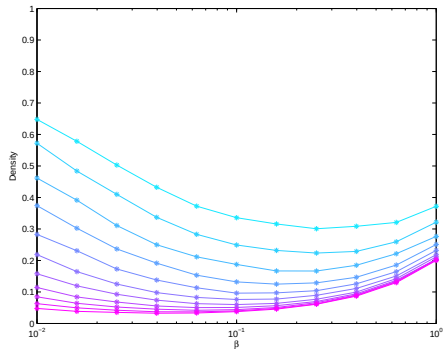
In this section we investigate how the relationship between our two scaling parameters β, which linearly scales the probability of link formation between a new node and an existing node, and B, which linearly scales the probability of link formation between two mature nodes, affect the topology of the networks generated by our algorithm. Recall that so far we have restricted our investigations to the case where $\beta = B$. In particular we are interested in how varying the ratio between β and B affects each of our four spatial network metrics. For simplicity we consider α, A and T to be a constant for all the networks we generate. We set $\alpha = A = 4$ and $T = 20$.



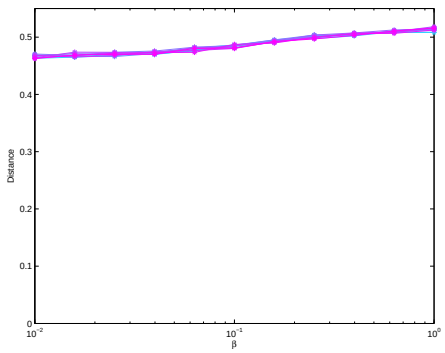
(a) *ASP* of the generated networks for different combinations of β and B.



(b) Clustering coefficient of the generated networks for different combinations of β and B.



(c) Density of the generated networks for different combinations of β and B.



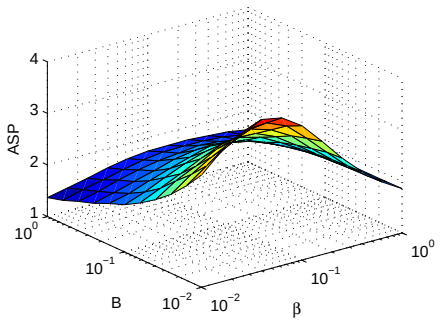
(d) Average node distance of the generated networks for different combinations of β and B.

Figure 2.19: Comparison of the properties of networks generated by our algorithm for different values of β and B where $\alpha = A = 4$, $T = 20$ and plot in increasingly magenta for lower B.

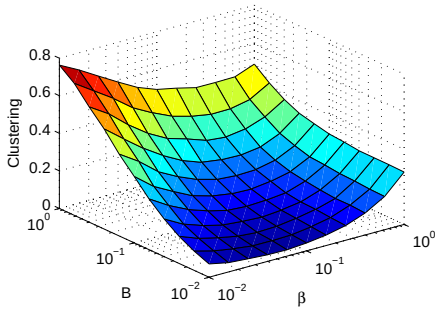
We take 11 values of β between $\frac{1}{100}$ and 1, the same 11 values for B, and

generate multiple spatial networks using our algorithm for each parameter combination. In Figures 2.19 and 2.20 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination. The average node distance is unaffected by changes in A as this parameter influences only the formation of links between pairs of mature nodes and not the placement of new nodes. Average node distance retains its increasing relationship with β that we have observed throughout our analysis of this algorithm and the original algorithm.

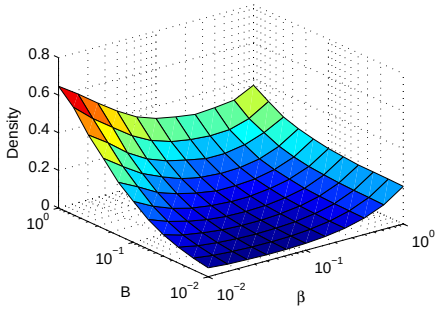
For this value of α ($\alpha = 4$), when examining both the original algorithm in Section 1.3 and the algorithm which allows mature links with $\beta = B$ in Section 2.2.1 we found that there was a positive relationship between both the clustering coefficient and the density of the networks generated and β and a negative relationship between ASP and β. Here, however, this is not the case. Holding B constant we find that increasing β sees both density and clustering fall to some minimum before increasing while ASP increases to some maximum before decreasing.



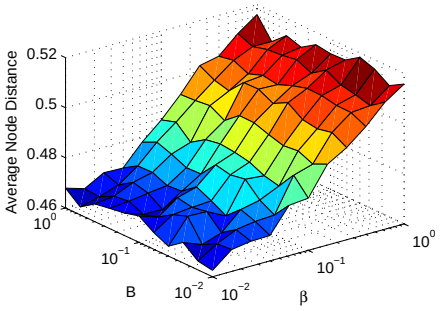
(a) ASP of the generated networks for different combinations of β and B.



(b) Clustering of the generated networks for different combinations of β and B.



(c) Density of the generated networks for different combinations of β and B.



(d) Average node distance of the generated networks for different combinations of β and B.

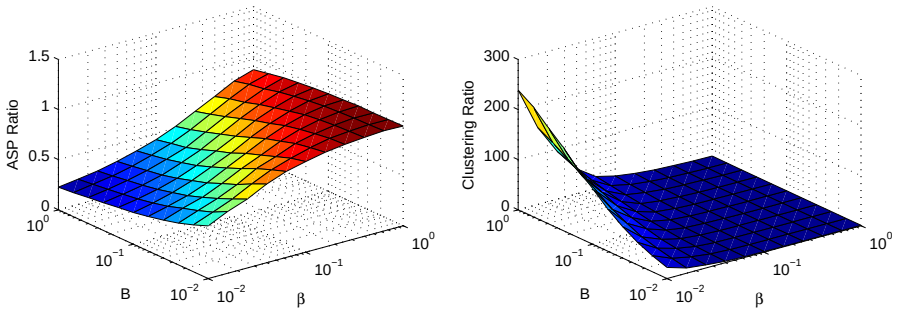
Figure 2.20: Comparison of the properties of networks generated by our algorithm for different values of β and B. We set $\alpha = A = 4$, $T = 20$.

As we increase β we see an increase in the number of links in the network that will be added when connecting a new node to the existing network. Added links will increase the density of the network and also tend to decrease the networks ASP and increase its clustering coefficient. However, increasing β also results in a fall in the number of links formed between mature nodes if we hold B constant. This is because as we increase β we reduce the number of iterations required to generate the network and therefore connection opportunities for mature nodes while not increasing the probability of a link forming between two mature nodes each time they attempt to connect (which they do so every $T = 20$ iterations).

Initially as we increase β the loss of links formed between mature nodes outweighs the extra links formed by new nodes. This results in a falling density and clustering coefficient and a rising ASP. However as β increases further the extra links formed by new nodes begin to outweigh the losses of links formed between mature nodes and density and clustering begin to rise once more as ASP falls.

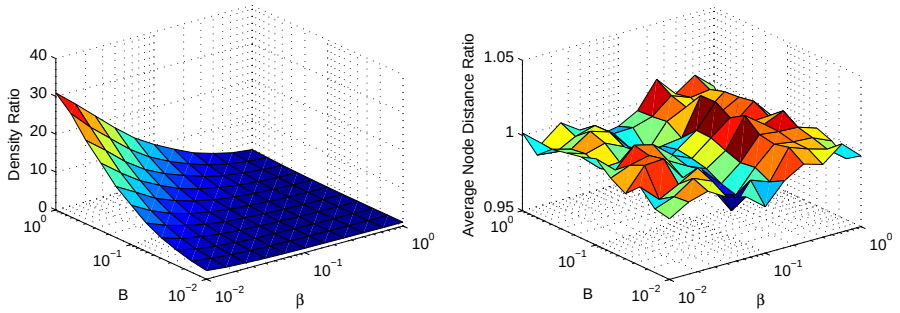
For all values of β we see a positive relationship between B and both the density and the clustering coefficient of the networks generated and a negative relationship between B and the ASP of the networks generated. This is due to the fact that an increase in B results only in an increased probability of forming connections between pairs of mature nodes and has no influence on the links formed when nodes are first added to the network.

Using the ratios plotted in Figure 2.21 we can compare the properties of networks generated by our new algorithm to the properties of networks generated by the original algorithm for different combinations of the parameters β and B. For large β and small B, i.e. when the $\beta : B$ ratio is large, the extension has no effect on the topology of the networks generated, this is due to the fact that in these cases links between mature nodes which are not previously connected are very unlikely to form and will have relatively very few opportunities to attempt to connect. On the other hand we see that our extension has the largest effect on he topology of the networks generated for large values of B and small values of β as in this case we maximise the opportunities for connections to form between mature nodes and the small β results in many opportunities to attempt to connect.



(a) Ratio of the ASP of the generated networks for different combinations of β and B where $\alpha = A = 4$ and $T = 20$ to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

(b) Ratio of the clustering coefficient of the generated networks for different combinations of β and B where $\alpha = A = 4$ and $T = 20$ to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of β and B where $\alpha = A = 4$ and $T = 20$ to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

(d) Ratio of the average node distance of the generated networks for different combinations of β and B where $\alpha = A = 4$ and $T = 20$ to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.21: Ratios of the properties of networks generated by our algorithm for different values of β and B when $\alpha = A = 4$, $T = 20$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

3 Considering Node Growth

In some systems, including the economic systems we will introduce in Chapter 4 where nodes may represent firms in an economy, it may be interesting to not only model the growth of the network but to also allow for the growth of individual nodes. We model node growth as a discrete process and each iteration of our algorithm is considered as a timestep.

In general, our model gives that nodes with a higher weighted degree, i.e. those that are connected to many other large nodes will grow larger than those with a low weighted degree. The size, $x_i(n)$ of an individual node v_i at time n is given by

$$x_i(n) = r_i(n)x_i(n-1)\left(1 - x_i(n-1)\right), \tag{2.9}$$

where

$$r_i(n) = 1 + \rho w_i(n), \tag{2.10}$$

and

$$w_i(n) = \frac{\sum\limits_{v_j \in V:(v_1,v_j)\in E} x_j(n)}{\sum\limits_{\forall v_j \in V} x_j(n)} \tag{2.11}$$

is the normalised weighted degree of the node at time n and $\rho \in \mathbb{R}$ is taken such that $x_i(n)$ has one stable fixed point. In particular, for our analysis in this section we will take $\rho = 1$. This discrete process, which we call the *network logistic map*, is further investigated in Chapter 3.

3.1 Development Algorithm

We now consider node size as an important factor in attracting new neighbours. Larger nodes are more likely to develop connections with new nodes than smaller ones. This may be thought of as a type of preferential attachment.

We again begin with one node at the centre of the unit square and proceed as follows:

1. New node position is given by a uniform random variable on the unit square, and new node size is given by a uniform random variable between 0 and the mean size of all nodes currently in the system.

2. The new node v_i connects with each existing node v_e with probability

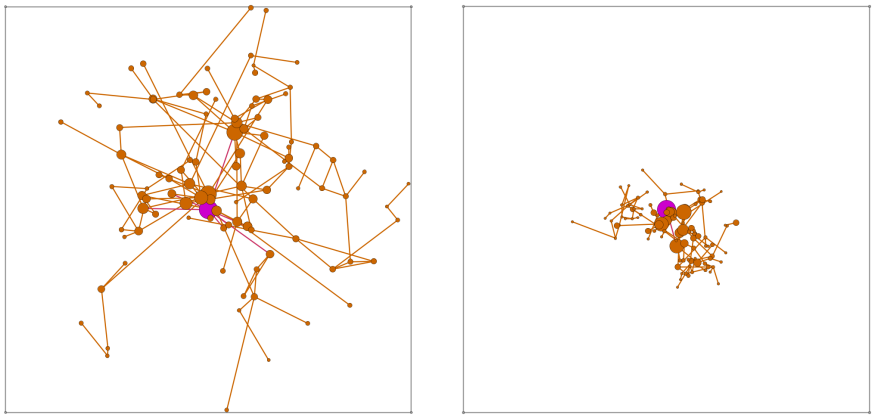
$$P_1 = \beta\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right]e^{-\alpha d(v_i,v_e)}, \tag{2.12}$$

   where $\bar{x}_e(n)$ is the normalised size of v_e at the time of the addition of v_i and L is a constant such that $0 \leq L \leq 1$.

3. The new node survives, i.e. is kept, if and only if it forms a connection with at least one existing node. If no connections are formed it is discarded.

4. Update individual node sizes according to equation (2.9).

5. Repeat this process until the desired number of nodes is reached.



(a) N_1^G: $\alpha = 15$ and $\beta = 0.5$. (b) N_2^G: $\alpha = 50$ and $\beta = 0.2$.

Figure 2.22: Network diagrams of two 100-node spatial networks generated by the algorithm outlined above for different values of α and β with $L = 0.5$. The boundaries of the unit square are shown in grey and the initial node is highlighted in pink. Nodes sizes are given by their x_i value when the final node v_{100} was added.

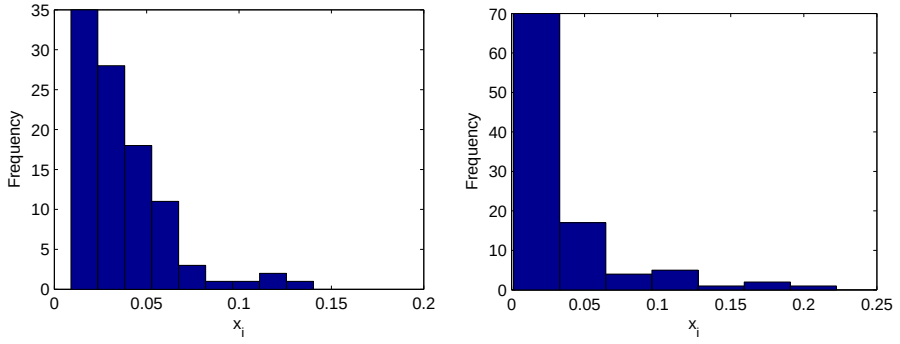
Figure 2.22 shows two examples of 100-node spatial networks generated by this network development algorithm. We can see that different choices of parameter values can again produce spatial networks with significantly different structures and properties. We have chosen the same values for α and β used for generating the networks in Figures 2.1 and 2.3. The new algorithm has not had any major visible impact on the spatial distribution of the nodes in the networks generated. The higher α value chosen for the network in Figure 2.22b once again causes a higher level of spatial concentration of nodes and lower mean distance between nodes in the network. The average node distance of N_1^G is 0.3392 while the average node distance of N_2^G is 0.1307. Recall that the original Kaiser-Hilgetag algorithm, investigated in Section 1.3, with the same parameter values generated networks with average node distances of 0.3774 and 0.1233 respectively, while the algorithm which allowed the establishment of mature connections, investigated in Section 2, with the same parameter values generated networks with average node distances of 0.3553 and 0.1767.

Allowing node growth and preferential attachment in the way specified by our algorithm has resulted in sparse networks in both cases. For large networks $\frac{\ln(1+x_e(n))}{\ln(2)}$ can be very small, resulting in small P_1 and the establishment of fewer links. The effect is similar to choosing small β for the original network generation algorithm discussed in Section 1.3. However, allowing node growth also has other effects, making nodes with high degree grow larger and then in turn attract more links.

The average shortest path length of N_1^G is 4.8251 while the average shortest path length of N_2^G is higher at 6.2471. N_1^G has a higher clustering coefficient than N_2^G, 0.0556 compared to 0.0348. Finally, N_1^G also has slightly higher

density, 0.0260 compared to 0.0219, although both are close to the lower bound 0.02.

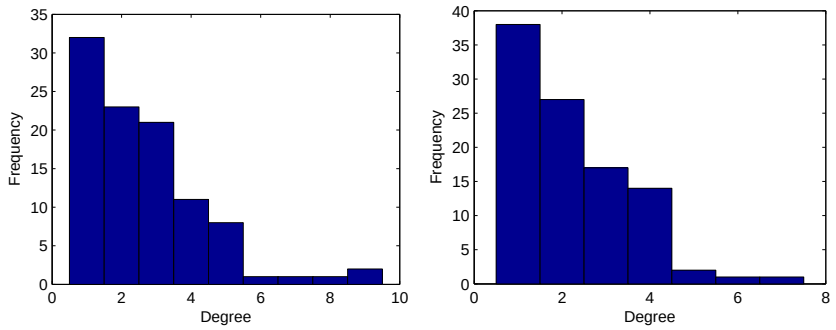
In Figure 2.23 we have plotted the distributions of the node sizes of both networks after the final iteration of our algorithm. In both cases we see that nodes with smaller sizes are seen more frequently. The minimum node size in N_1^G is 0.0090 and the maximum is 0.1403. For N_2^G the minimum is 0.0015 while the maximum is 0.2223. N_1^G has a higher mean node size than N_2^G, 0.0376 versus 0.0317, and a significantly lower standard deviation, 0.0243 versus 0.0412.



(a) Node sizes histogram for N_1^G: $\alpha = 15$ and $\beta = 0.5$.

(b) Node sizes histogram for N_2^G: $\alpha = 50$ and $\beta = 0.2$.

Figure 2.23: Histograms of the node sizes of N_1^G and N_G^2 when the final node v_{100} was added.



(a) Degree distributions of N_1^G: $\alpha = 15$ and $\beta = 0.5$.

(b) Degree distributions of N_2^G: $\alpha = 50$ and $\beta = 0.2$.

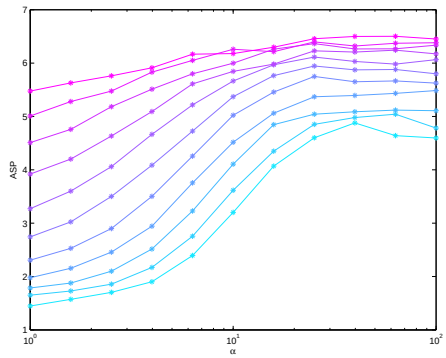
Figure 2.24: Degree distributions of N_1^G and N_G^2 .

We see in Figure 2.24 that the distribution of the node sizes of both networks is quite closely mirrored by their degree distribution. This is a consequence of the discrete process we use to model the growth of individual nodes in this set-up. Nodes with higher degrees have a higher $r_i(n)$ value, causing them to grow to be larger.

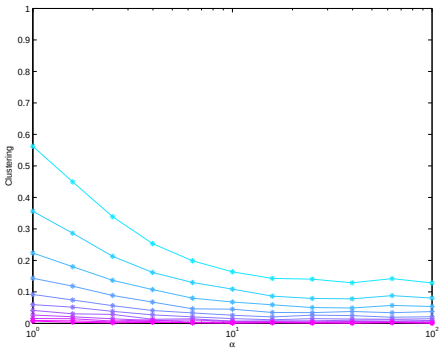
3.2 Results

In this section we investigate the effect of different parameter value combinations on the spatial networks generated by our algorithm. Our analysis again focuses on the changes in the four metrics defined in Section 1.2. For each parameter combination we simulated 50 100-node networks. In particular we are interested in how the relationships between α and β and the properties of the networks our algorithms change as we vary L making individual node size more and less important when forming links.

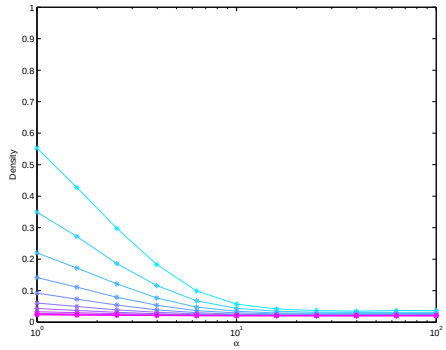
3.2.1 Varying α and β



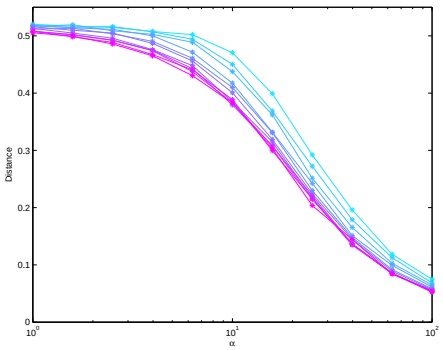
(a) *ASP* of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



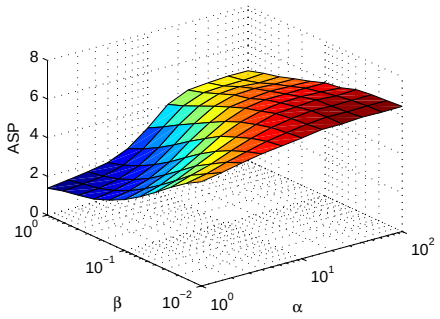
(c) Density of the generated networks for different combinations of α and β.



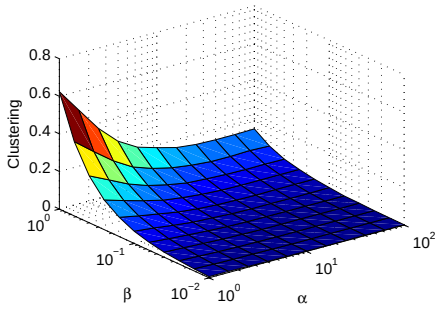
(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.25: Comparison of the properties of networks generated by our algorithm for different values of α and β, with $L = 0.1$. We plot in increasingly magenta for lower β.

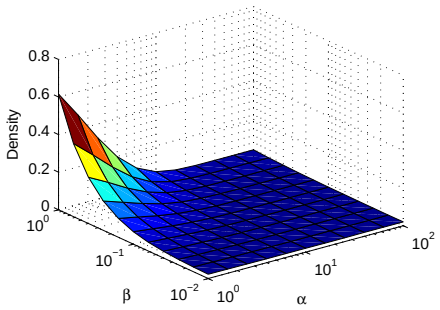
We look first at the interactions between different values of α and β and how allowing node size to have an impact on link formation probabilities affects the properties of the network generated. In our first example we look at a case where node size has relatively little influence. We set $L = 0.1$ and take 11 values of α between 1 and 100, 11 values of β between $\frac{1}{100}$ and 1, and generate 50 spatial networks using our algorithm for each parameter combination. In Figure 2.25 and Figure 2.26 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination.



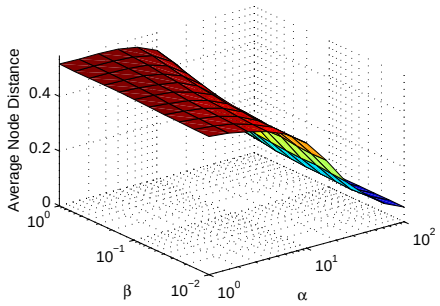
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.



(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.26: Comparison of the properties of networks generated by our algorithm for different values of α and β, with $L = 0.1$.

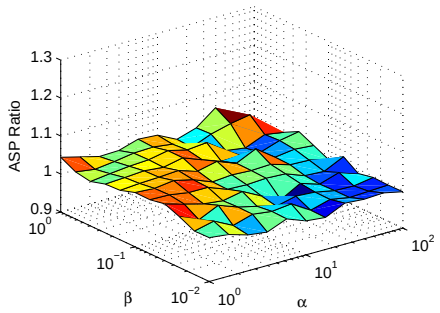
The relationships between both α and β and the average node distance of the spatial networks generated by our algorithm look very similar to the relationships seen for the original algorithm which did not consider node growth in Section 1.3 and the relationships seen for the algorithm which allowed connections to form between mature nodes but also did not consider node growth in Section 2.2.1. The relationship between β and average node distance is positive

while the relationship between $\log(\alpha)$ and average node distance is sigmoidal and negative.

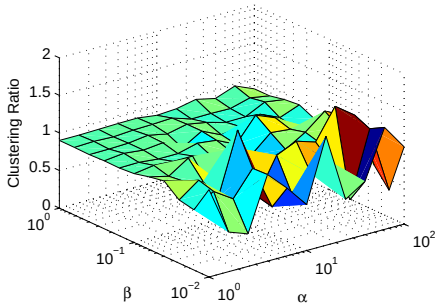
The relationships between the density of the networks generated and the clustering coefficient of the networks generated and α and β also look similar to the relationships we observed for the original algorithm. Density, a measure of the global connectivity of the network, falls as we increase α, making it more difficult for distant nodes to form connections and also decreases as we decrease β making it more difficult for any given pair of nodes to connect. Local connectivity, measured by the clustering coefficient has a similar relationship with α and β, although a relatively high clustering level is maintained for high α and high β as local connectivity is less affected than global connectivity by an increase in α.

The relationship between the ASP of the networks generated and α and β also look similar to the relationships we observed for the original algorithm. In general we see the ASP of the network increase as we increase α. However for small values of β where we generate networks with low density we note that the rate of increase of ASP as we increase α seems to be slower than in the case of the original algorithm. Perhaps the slight tendency to favour connections with larger nodes causes us to choose links better suited to maintaining short average shortest path lengths in sparse networks.

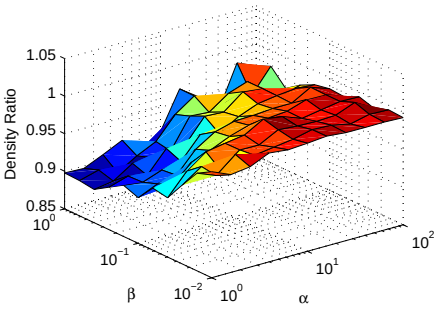
In order to compare the networks generated by our new algorithm to the networks generated by the original Kaiser-Hilgetag algorithm we have plotted the ratios of the value of each of the four network metrics for a network generated by the new algorithm to the corresponding value of that metric for a network generated by the original algorithm in Figure 2.27. These ratios confirm that when we choose a small L, in this case we have chosen $L = 0.1$, considering node growth has a negligible impact on the spatial distribution or topology of the network produced with most ratios being close to 1 for all values of α and β considered. The biggest impact it seems is on the density of the network produced for large values of β and small values of α. These networks are slightly less dense than in the case of the original algorithm. This is because $\frac{\ln(1+\bar{x}_e(n))}{\ln(2)} < 1$ and so our extension has a similar effect to reducing β in the original algorithm.



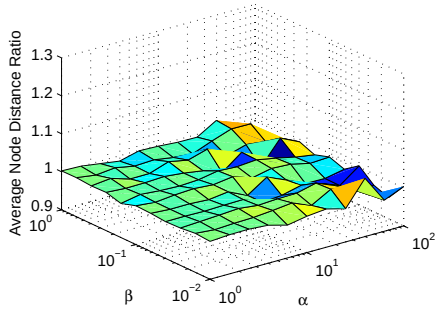
(a) Ratio of the ASP of the generated networks for different combinations of α and β to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and β to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



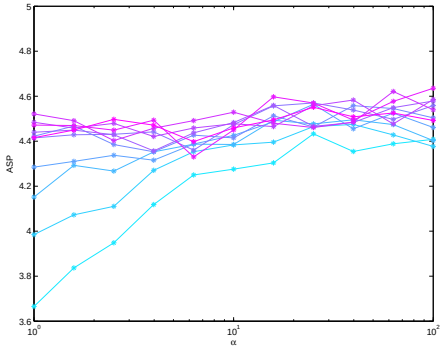
(c) Ratio of the density of the generated networks for different combinations of α and β to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



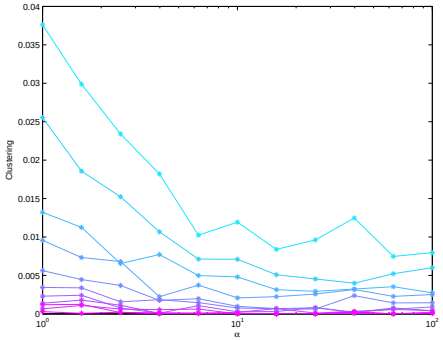
(d) Ratio of the average node distance of the generated networks for different combinations of α and β to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.27: Ratios of the properties of networks generated by our algorithm for different values of α and β with $L = 0.1$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

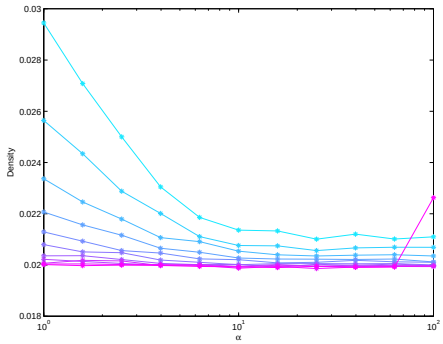
We now look at a more extreme case. We allow node size to have its maximum influence on link formation and so on the network generated by setting $L = 1$. We take the same 11 values of α and 11 values of β and again generate spatial networks for each parameter combination. In Figures 2.28 and 2.29 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated.



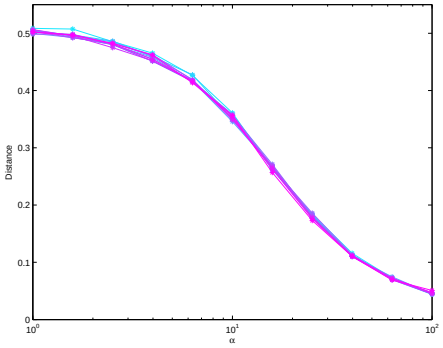
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.



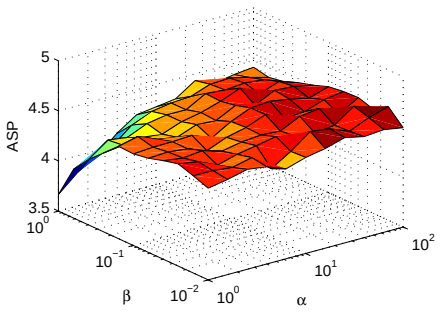
(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.28: Comparison of the properties of networks generated by our algorithm for different values of α and β, with $L = 1$. We plot in increasingly magenta for lower β.

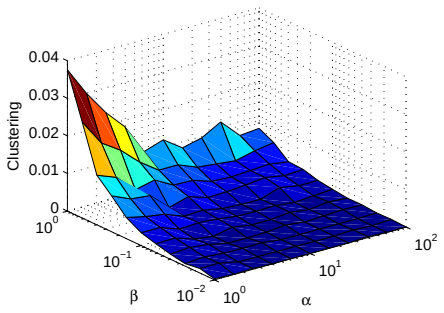
The density of the networks generated and their clustering coefficients maintain negative relationships with α and positive relationships with β however their values are reduced throughout compared to the densities and clustering coefficients of networks produced by the original algorithm with the same values of α and β. This is due to the fact that the connection probability for any two nodes in the $L = 1$ case for our new algorithm $P_1 = \beta \left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)} \right) e^{-\alpha d(v_i,v_e)}$ is significantly less than the corresponding value for the original algorithm $Q_1 = \beta e^{-\alpha d(v_i,v_e)}$. Average node distance retains its negative sigmoidal relationship with average node distance while β now scaled by $\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}$ has little effect.

Especially for lower values of β the relationships between ASP and α and β are not so clear. For high values of β where density is highest there is a positive relationship between ASP and α. For lower values of β where the number of

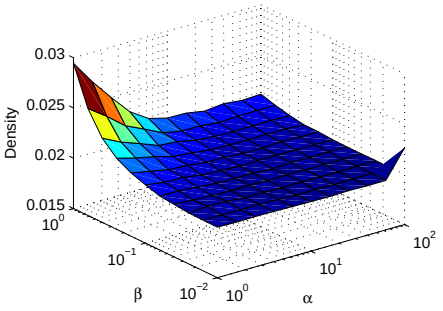
links in the network is close to the minimum the relationship is unclear.



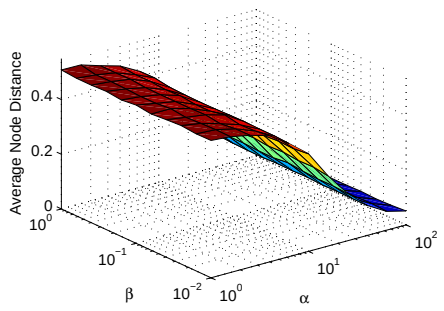
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.

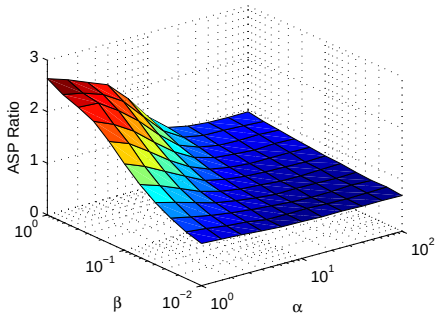


(d) Average node distance of the generated networks for different combinations of α and β.

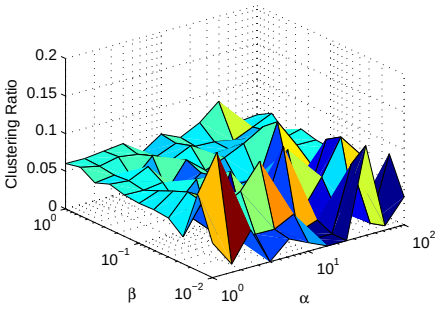
Figure 2.29: Comparison of the properties of networks generated by our algorithm for different values of α and β, with $L = 1$.

In order to achieve a greater understanding of these relationships we refer to the ratios plotted in Figure 2.30. For high values of β and low values of α we see that the ASP of networks generated by this algorithm is significantly higher than the ASP of networks generated by the original algorithm investigated in Section 1.3 which did not consider node size during link formation (ratios > 1). We can attribute this to the reduced density of these networks compared to the corresponding networks generated by the original algorithm (ratios < 1). For high values of α and low β on the other hand we find the ASP is significantly reduced. For these values the density in both the networks produced by the original algorithm and our new algorithm is close to the minimum and so there are no density effects on ASP, instead the main driver of difference between the ASPs of the networks generated by the two algorithms is the nodes that we prefer to attach to in each case. The only criteria in the case of the original algorithm is spatial proximity, in the new algorithm the main criteria is node

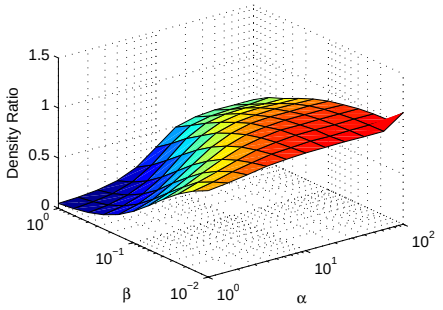
size with spatial proximity a secondary criteria. This leads to formation of a few highly connected nodes in the case of the new network as large nodes attract more and more connections reducing ASP.



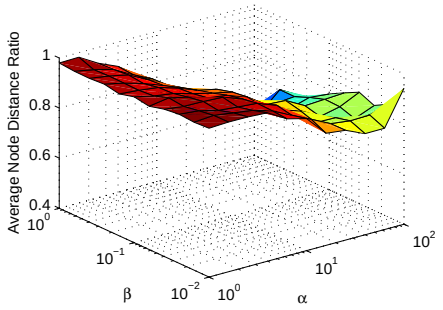
(a) Ratio of the ASP of the generated networks for different combinations of α and β to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and β to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



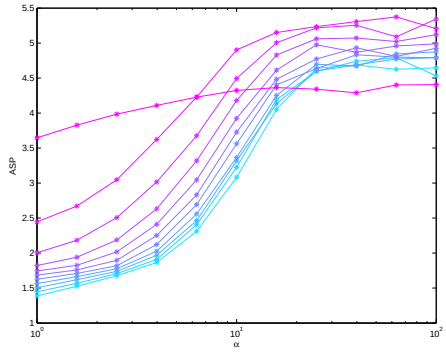
(c) Ratio of the density of the generated networks for different combinations of α and β to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



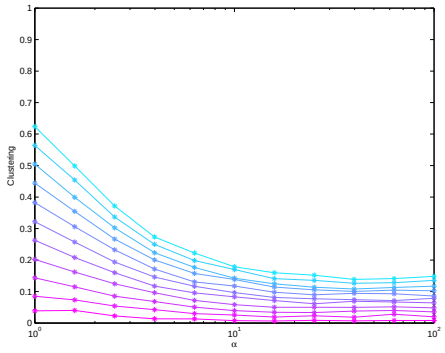
(d) Ratio of the average node distance of the generated networks for different combinations of α and β to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.30: Ratios of the properties of networks generated by our algorithm for different values of α and β with $L = 1$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

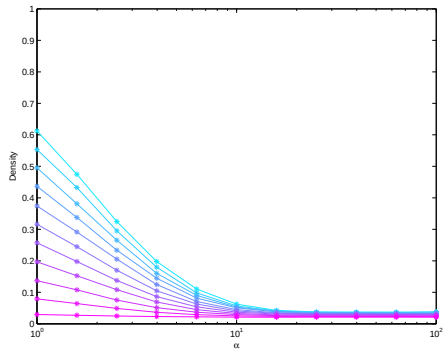
3.2.2 Varying α and L



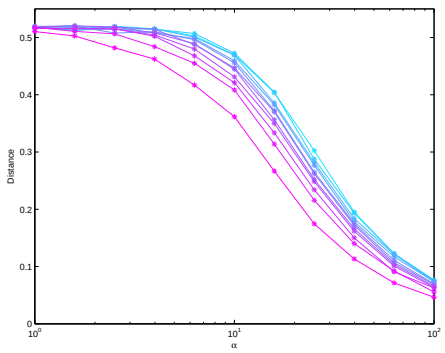
(a) *ASP* of the generated networks for different combinations of α and L.



(b) Clustering coefficient of the generated networks for different combinations of α and L.



(c) Density of the generated networks for different combinations of α and L.



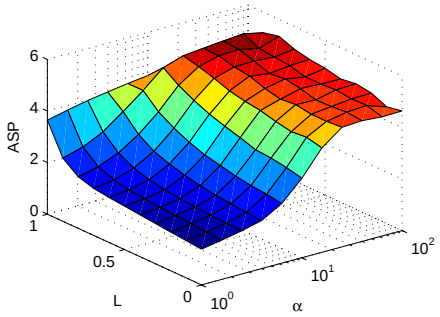
(d) Average node distance of the generated networks for different combinations of α and L.

Figure 2.31: Comparison of the properties of networks generated by our algorithm for different values of α and L, with $\beta = 1$. We plot in increasingly magenta for higher L.

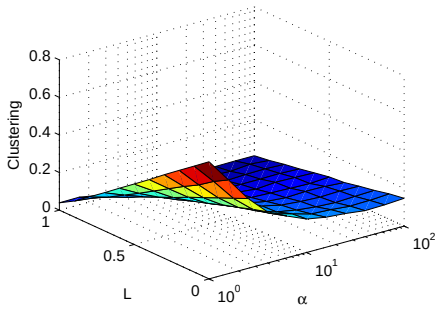
We now investigate the interactions between different values of α and L and how varying the influence of node size on link formation affects the properties of the networks generated. We set $\beta = 1$ and take 11 values of α between 1 and 100, 11 values of L between 0 and 1, and generate 50 spatial networks using our algorithm for each parameter combination. In Figure 2.31 and Figure 2.32 we can see the mean of the *ASP*, clustering coefficient, density and average node distance of the networks generated for each parameter combination.

We first note that α maintains a negative relationship with the density of the network generated, the clustering coefficient of the network generated and the average node distance of the network generated. We also see that for this

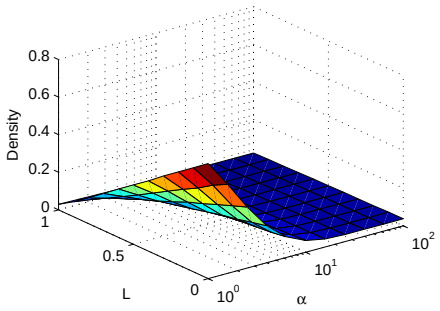
value of β, $\beta = 1$, for every value of α investigated L has a negative relationship with density and the clustering coefficient. We hypothesise that the major cause of this relationship is the fact that $\frac{\ln(1+\bar{x}_e(n))}{\ln(2)} < 1$ and so one of the effects of increasing L is equivalent to reducing β in the original algorithm making the formation of any given link less likely. This reduces the total number of links in the network and as a result the global connectivity, i.e. the density, of the network and is also likely to result in a reduction in local connectivity, i.e. the clustering coefficient, of the network.



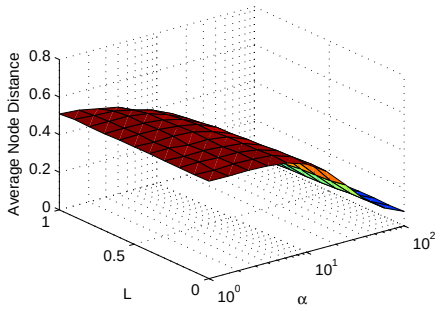
(a) *ASP* of the generated networks for different combinations of α and L.



(b) Clustering coefficient of the generated networks for different combinations of α and L.



(c) Density of the generated networks for different combinations of α and L.



(d) Average node distance of the generated networks for different combinations of α and L.

Figure 2.32: Comparison of the properties of networks generated by our algorithm for different values of α and L, with $\beta = 1$.

This reduction in $P_1 = \beta\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right] e^{-\alpha d(v_i,v_e)}$ due to an increase in L also results in a reduced average node distance. This reduced probability forces new nodes to be closer to the existing network in order to form a link. The effect is compounded by the preferential attachment feature of our new algorithm. New nodes are more likely to form links with the original node from which the network was grown or nodes added early in the development of

the network which are more likely to be nearby than newer nodes on the spatial periphery. This is because these nodes are more likely to have a higher degree as they have had more opportunities to form links with nodes being added to the network over time and hence have a lager node size.

For $\beta = 1$ and all values of L investigated the relationship between the ASP of the networks generated by our algorithm is generally positive. However for high values of L and α where the density of the networks is close to or equal to the lower bound 0.02 this ASP does not change significantly as we increase α. For low values of α the relationship between L and ASP is positive, an increase in L reduces the number of links in the network increasing the ASP. For high values of α where the density of the networks produced is low for all values of L choosing high L can lead to a reduced ASP. In these cases a high L causes node size to be a large factor in link formation, nodes with a high node size attract more links causing their size to increase further and attract even more links. This leads to networks with a small number of high degree *hubs*, which in turn leads to relatively low ASP.

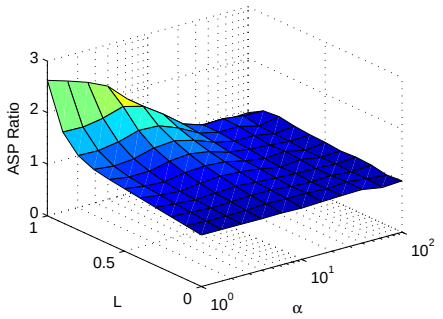
The ratios plotted in Figure 2.33 allow us to compare the networks generated by this algorithm to the original algorithm. We note that as expected the ratios for all four metrics are approximately equal to 1 for $L = 0$ since in this case our algorithm reduces to the original algorithm in therms of node placement and link formation. The largest effects in all four cases are seen when we allow node size to have the largest influence on link formation by setting $L = 1$. The effect on average node distance is largest on sparse networks when α is large and all new nodes are drawn to connect to nodes with high degree and therefore node size at the centre of the network.

For $L > 0$ we see that the density of the networks generated by our algorithm is reduced compared to the original algorithm (ratio < 1). The effect is most pronounced for small α where networks generated by our original algorithm with $\beta = 1$ are very dense. Increasing L reduces the probability of link formation between any two nodes, $P_1 = \beta \left[(1 - L) + L \left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right] e^{-\alpha d(v_i, v_e)}$, and so reduces the number of links formed. For higher α the networks generated by the original algorithm are quite sparse, with densities close to the minimum 0.02, in this case the effect is less pronounced as we can only bring our density closer to the minimum.

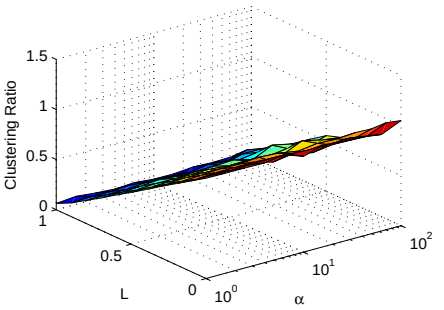
The clustering coefficient of the networks produced when $L > 0$ is also reduced compared to networks generated by the original algorithm. In this case the effect is pronounced for all α. Recall that for $\beta = 1$ in the original algorithm clustering remains, relatively speaking, quite high for high α as short links are still formed, increasing local connectivity. However, as $L \to 1$ node size dominates the criteria for link formation and so local connectivity does not remain high even though $\beta = 1$. For this reason for all alpha we see our clustering coefficient ratio approaches 0 as $L \to 1$ for all α.

For large L and small α the ASP of the networks produced is increased compared to networks generated by the original algorithm. The reduced density due to high L is the main influence in his case. As we increase α the density

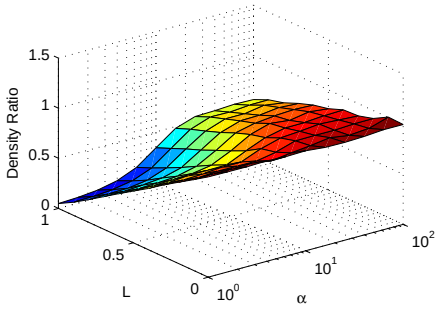
although still reduced is less affected. Now the tendency for new nodes to form links with nodes of a higher node size and so create a small number of highly connected hubs in the network reduces the ASP of the networks generated compared to those generated by the original algorithm.



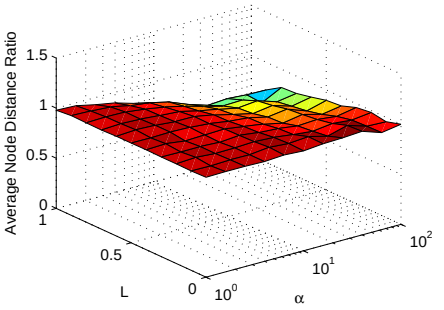
(a) Ratio of the ASP of the generated networks for different combinations of α and L when $\beta = 1$ to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and L when $\beta = 1$ to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



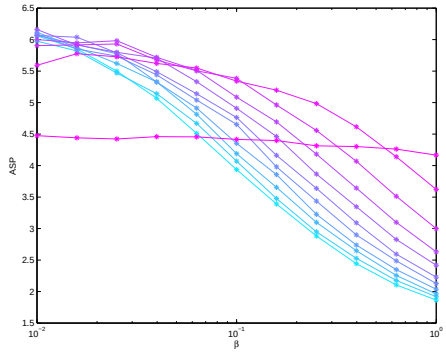
(c) Ratio of the density of the generated networks for different combinations of α and L when $\beta = 1$ to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



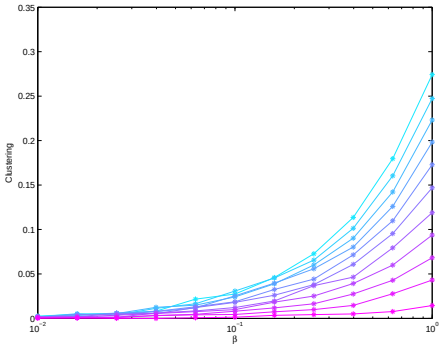
(d) Ratio of the average node distance of the generated networks for different combinations of α and L when $\beta = 1$ to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.33: Ratios of the properties of networks generated by our algorithm for different values of α and L with $\beta = 1$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

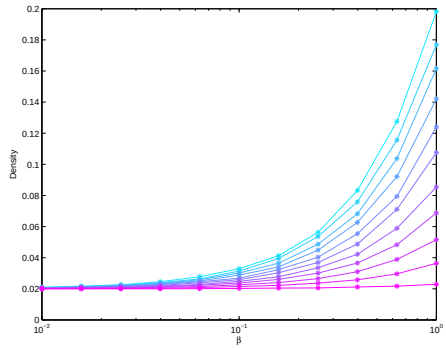
3.2.3 Varying β and L



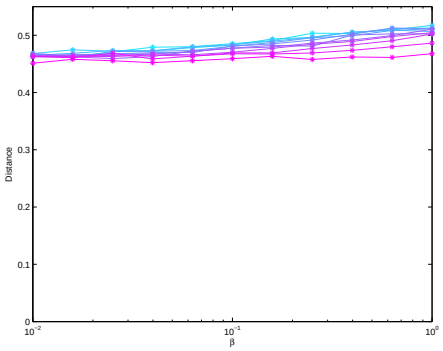
(a) ASP of the generated networks for different combinations of β and L.



(b) Clustering coefficient of the generated networks for different combinations of β and L.



(c) Density of the generated networks for different combinations of β and L.



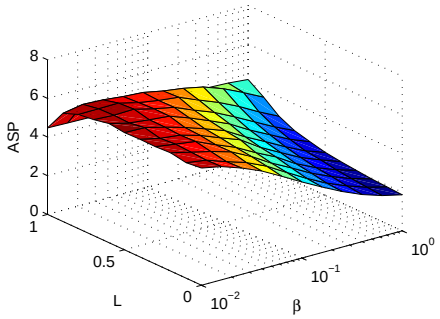
(d) Average node distance

Figure 2.34: Comparison of the properties of networks generated by our algorithm for different values of β and L, with $\alpha = 4$. We plot in increasingly magenta for higher L.

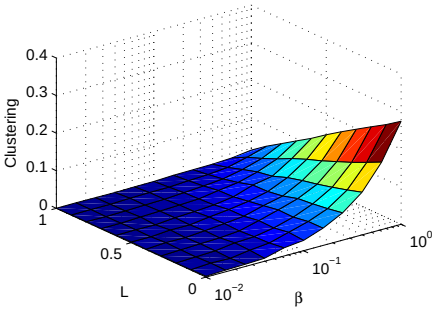
Finally, we will investigate how the interactions between β and L affect the properties of the spatial networks produced by our algorithm. We set $\alpha = 4$ and take 11 values of β between $\frac{1}{100}$ and 1, 11 values of L between 0 and 1, and generate 50 spatial networks using our algorithm for each parameter combination. In Figure 2.34 and Figure 2.35 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination.

We see that for all values of L and $\alpha = 4$, β has an positive relationship with the clustering coefficient, density and average node distance of the network produced. For small β and high L both clustering and density are close to their respective minimums 0 and 0.02. For all values of β investigated and $\alpha = 4$

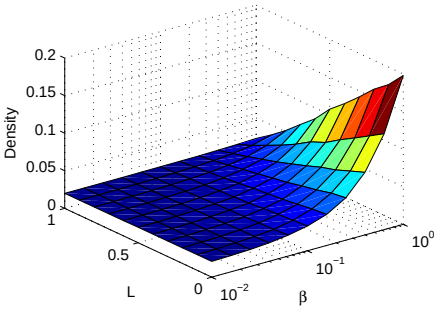
L has negative relationship with clustering, density and average node distance. All these relationships can be explained by the fact that an increase in L or decrease in β reduces the probability of link formation between any two nodes $P_1 = \beta\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right] e^{-\alpha d(v_i,v_e)}$.



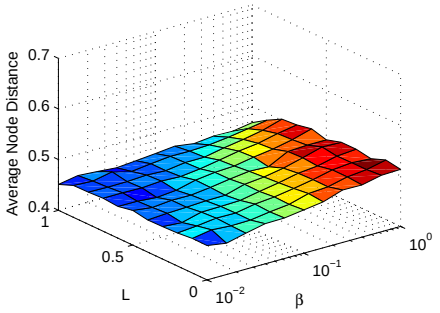
(a) ASP of the generated networks for different combinations of β and L.



(b) Clustering coefficient of the generated networks for different combinations of β and L.



(c) Density of the generated networks for different combinations of β and L.

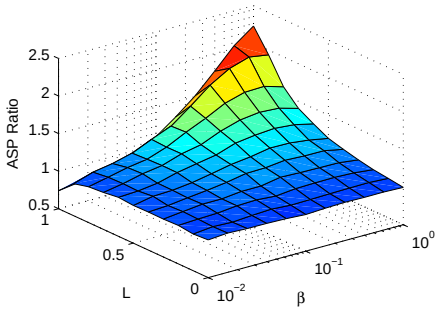


(d) Average node distance of the generated networks for different combinations of β and L.

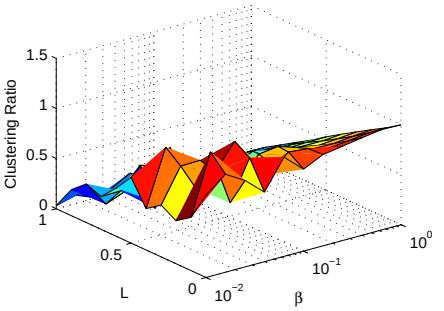
Figure 2.35: Comparison of the properties of networks generated by our algorithm for different values of β and L, with $\alpha = 4$.

The most interesting relationship is between L and the ASP of the network generated. For the highest density networks investigated, generated when $\beta = 1$ the relationship between ASP and L is positive as any increase in L results in a significant fall in the number of links in the network generated leading to an increased ASP. As we reduce β the density of all networks reduces and approaches its minimum. This makes the falls in density due to an increase in L less significant and the relationship begins to reverse. For $\beta = \frac{1}{100}$ the density of the network produced for all values of $L \approx 0.02$ and any change in L has an almost insignificant impact on the number of links in the network. Now the higher tendency to generate networks with a hub and spoke structure for higher

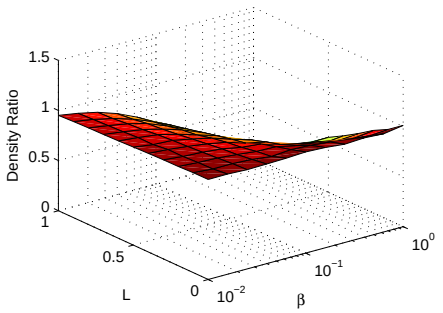
L results in a shorter ASP for networks generated in these cases.



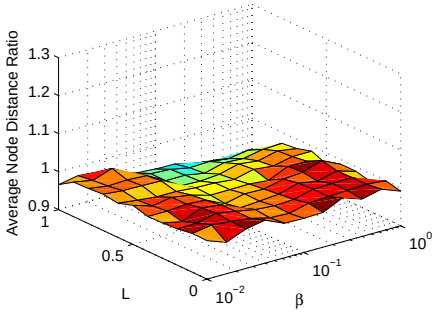
(a) Ratio of the ASP of the generated networks for different combinations of β and L when $\alpha = 4$ to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of β and L when $\alpha = 4$ to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of β and L when $\alpha = 4$ to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(d) Ratio of the average node distance of the generated networks for different combinations of β and L when $\alpha = 4$ to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.36: Ratios of the properties of networks generated by our algorithm for different values of β and L with $\alpha = 4$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

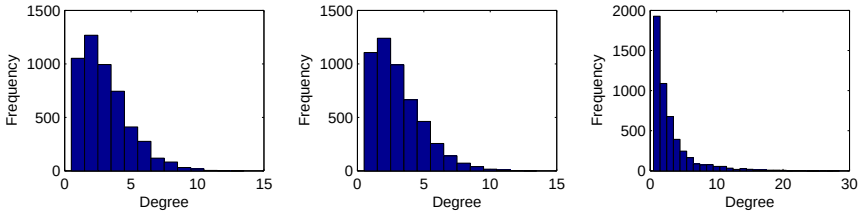
The ratios in Figure 2.36 confirm that the networks generated by this algorithm differ most significantly from those generated by the original Kaiser-Hilgetag algorithm, which did not consider node growth or the influence of node size on link formation, when $L = 1$ and this influence is at its maximum. The density and clustering coefficients of networks generated are both reduced compared to the case of the original algorithm when $L > 0$. For β the effect is less pronounced for small β as in the case of the original algorithm the density of the

networks generated for these values of β was already close to the minimum possible. In the case of clustering the effect is significant for all β as node size rather than spatial distance now dominates link formation probabilities, significantly reducing the formation of links between neighbours. For large β and high L the ASP of the relatively sparse network produced by our new algorithm is longer than the corresponding, relatively dense, network produced by the original algorithm. For small β and high L the topology of the sparse network produced by our algorithm favouring larger nodes during link formation has a smaller ASP than the corresponding sparse network produced by the original algorithm.

3.2.4 Degree Distributions

In this Section we present an example of how different values of L affect the structure of the degree distributions of the networks generated by our algorithm and how these changes in the degree distribution can lead to shorter average shortest path lengths for networks of equal densities as we increase the influence of node size on link formation. We saw evidence of this behaviour for extremely sparse networks in Sections 3.2.1, 3.2.2 and 3.2.3, we will present some evidence now for slightly more dense networks.

For our example we take $\alpha = 20$ and generate 50 100-node networks for $L = 0$, $L = 0.5$ and $L = 1$. For the $L = 0$ case (which reduces us to the original algorithm in terms of node placement and link formation) we take $\beta = 0.5$ and the mean density of the 50 networks generated is 0.0309. For $L = 0.5$ and $L = 1$ we choose values of β in order to give us the same mean density over the 50 networks generated. If this value of β was greater than 1 we took $P_1 = \max\{1, \beta\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right] e^{-\alpha d(v_i,v_e)}\}$ in place of $\beta\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right] e^{-\alpha d(v_i,v_e)}$.

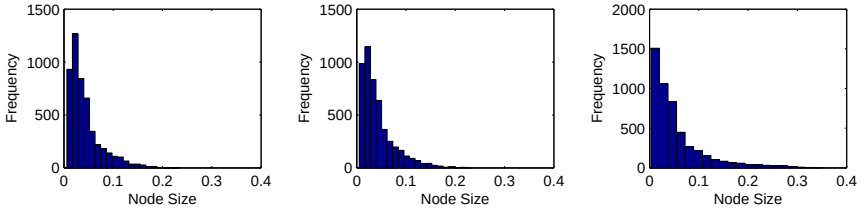


(a) $L = 0$ (b) $L = 0.5$ (c) $L = 1$

Figure 2.37: Degree distribution of 50 100-node networks generated by our algorithm for different values of L, where $\alpha = 20$ for all values of L, $\beta = 0.5$ for $L = 0$ and β is chosen to produce networks of the same mean density for $L = 0.5$ and $L = 1$.

Increasing L from 0 to 0.5 sees a very slight change in the degree distribution of the networks produced. An increased number of nodes now have degree 1 while the number of highly connected nodes of degree > 10 also slightly increases. The mean ASP length over the 50 networks generated also shows a

slight fall from 4.8583 to 4.7315. Increasing L further to 1 sees a more significant visual change in the degree distribution of the networks generated. Most nodes now have degree 1 and the maximum degree seen increases significantly from 13 to 28 this is consistent with a *hub and spoke* topology in which very small number of nodes have a high degree, the hubs, while the rest have a small degree, the spokes. This topology allows the networks generated to have a significantly lower mean ASP of 3.8272. Figure 2.38 allows us to see that the distribution of node sizes quite closely resembles the distribution of degrees of nodes. Generally speaking, as we discussed in Chapter 3, for the discrete process we have used to model node growth for this algorithm leads to higher node sizes for nodes with a higher degree.



(a) $L = 0$ (b) $L = 0.5$ (c) $L = 1$

Figure 2.38: Node size distribution of 50 100-node networks generated by our algorithm for different values of L, where $\alpha = 20$ for all values of L, $\beta = 0.5$ for $L = 0$ and β is chosen to produce networks of the same mean density for $L = 0.5$ and $L = 1$.

4 Overlapping Networks

We now look to model the growth of two overlapping spatial networks, the *pioneer network* and the overlapping *settler network*.

The Pioneer Network, $\mathbf{N_p} = (\mathbf{V_p}, \mathbf{E_p})$, In this analysis we will consider only cases where the growth of the *pioneer network* is completely independent of the overlapping *settler network*. It grows according to one of the simple development algorithms outlined in Section 1.3, Section 2 or Section 3 choosing parameters independent of those chosen for the settler network.

The Settler Network, $\mathbf{N_s} = (\mathbf{V_s}, \mathbf{E_s})$, The *settler network* also develops over time and its growth is affected by the structure of the *pioneer network*.

4.1 Settler Network Development Algorithm

In the most simple case, which we investigate here, we allow the pioneer network to affect only the initial spatial location of potential nodes.

We again begin with one node from each network in the unit square. We allow the pioneer network to develop according to one of the algorithms detailed

in Sections 1.3 - 3. At the same time our settler network develops according to the following algorithm:

1. For each new node we define a truncated Gaussian mixture distribution with probability density function

$$f(x) = \sum_{v \in V_p} w_v f_v(x), \tag{2.13}$$

   where the weights w_v are given by the normalised node size of v and $f_v(x)$ is the probability density function of a truncated Gaussian distribution[1] on the unit square with mean μ_v given by the spatial location of v and covariance matrix $\Sigma_v = \sigma^2 I$. From this distribution we choose our new node position.

2. The new node v_i connects with each existing node v_e with probability

$$P_1 = \beta_s e^{-\alpha_s d(v_i, v_e)}, \tag{2.14}$$

   where $d(v_i, v_e)$ is the Euclidean distance between the two nodes, $0 < \beta_s \leq 1$ and $\alpha_s > 0$.

3. The new node survives, i.e. is kept, if and only if it forms a connection with at least one existing node. If no connections are formed it is discarded.

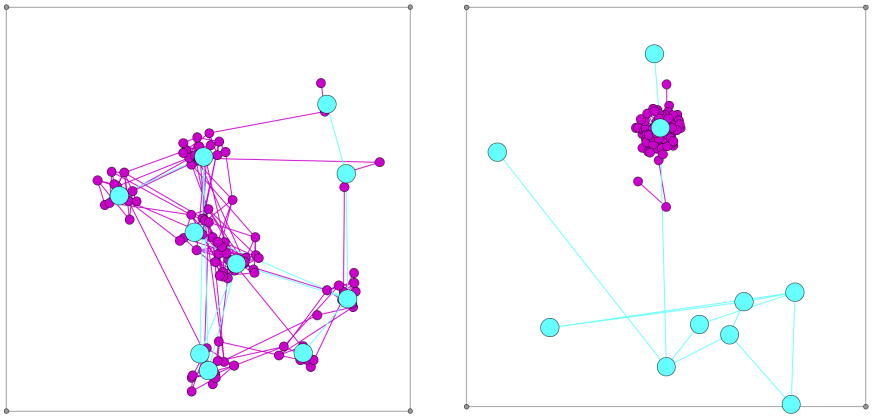
4. We repeat this process until the desired number of nodes is reached.

For simplicity, we choose to generate our pioneer networks in this section using the original, simple Kaiser-Hilgetag algorithm, detailed in Section 1.3. Figure 2.39 shows two examples of 100-node settler networks generated by this network development algorithm, along with their corresponding 10-node pioneer networks. Both pioneer networks were generated using the algorithm discussed in Section 1.3 grown from an initial node at $(0.5, 0.1)$ with $\alpha = \alpha_p = 3$ and $\beta = \beta_p = 0.4$. We have chosen the same values for α_s and β_s as were given to α and β when generating the networks in Figures 2.1, 2.3 and 2.22. We can immediately see striking differences between the spatial distribution of the nodes in N_1^O and N_2^O and the corresponding networks in Figures 2.1, 2.3 and 2.22 in Sections 1.3, 2.1 and 3.1 respectively.

In Figure 2.39a we can see that the relationship between the pioneer network and the settler network has resulted in the formation of multiple clearly distinct spatial clusters in the settler network N_1^O. Spatial clusters of this kind were not clearly visible in the case of N_1^S (see Figure 2.1a) where we used the original Kaiser-Hilgetag algorithm, N_1^M where we allowed mature connections to form, or N_1^G where we considered the growth of individual nodes. Despite the clearly different spatial structure of N_1^O the new algorithm has not had such a major effect on the average node distance of N_1^O. The significant change has been in the distribution of node distances rather than the mean of all node distances.

[1] In our simulations we generate pseudo-random vectors drawn from the truncated Gaussian distribution using code by Benham from the mathworks.com file exchange [14].

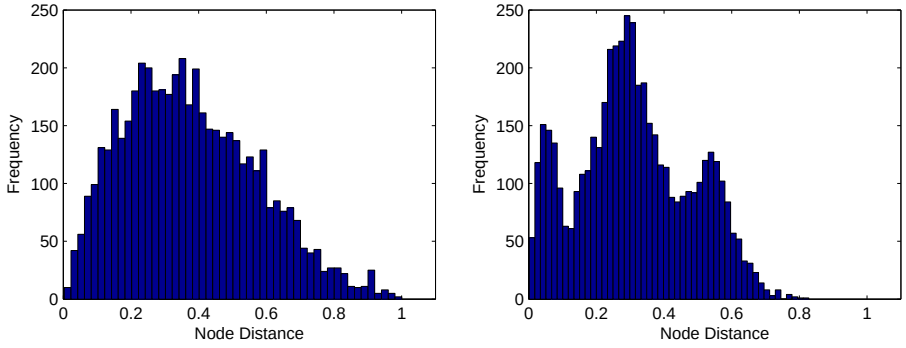
The average node distance of N_1^O is 0.3120 compared to 0.3774 for N_1^S, 0.3553 for N_1^M and 0.3392 for N_1^G.



(a) N_1^O: $\alpha_s = 15$ and $\beta_s = 0.5$.

(b) N_2^O: $\alpha_s = 50$ and $\beta_s = 0.2$.

Figure 2.39: Network diagrams of two 100-node settler (pink, small nodes) networks, N_1^O and N_2^O, generated by the algorithm outlined above for different values of α_s and β_s with $\sigma^2 = 0.001$. For both networks we have chosen ten-node pioneer (cyan, large nodes) networks , N_1^P and N_2^P, generated by development algorithm outlined in Section 1.3 grown from an initial node at $(0.5, 0.1)$ with $\alpha_p = 3$ and $\beta_p = 0.4$. The boundaries of the unit square are shown in grey.



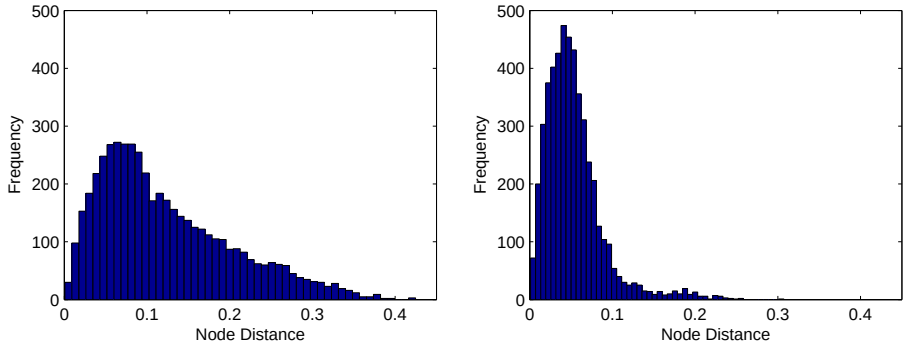
(a) Distribution of the distances between pairs of nodes in the network N_1^S generated by the original Kaiser-Hilgetag algorithm and analysed in Section 1.3.

(b) Distribution of the distances between pairs of nodes in the settler network N_1^O seen in Figure 2.39a.

Figure 2.40: Node distance distributions for N_1^S and N_1^O.

Figure 2.40 shows the frequency distribution of the distances between each pair of nodes in the networks N_1^S and N_1^O. The distribution for N_1^S is unimodal with a slightly long right-hand tail. This is consistent with a spatial network whose nodes are distributed in a single spatial cluster as we observed in Section 1.3. The distribution for N_1^O by contrast is trimodal. These multiple modes are a consequence of the multiple clearly distinct spatial clusters seen in the settler network. There are a high number of short distances between pairs of nodes which are in the same spatial cluster, a high number of medium distances between pairs of nodes which are in the nearby spatial clusters and a high number of long distances between pairs of nodes in distant spatial networks.

In Figure 2.39b we can see that, despite the presence of the pioneer network N_2^P, multiple spatial clusters have not formed in the settler network N_2^O. This may be attributed to the relative sizes of α_s & α_p and β_s & β_p. The settler network has a very high α_s and a low β_s. This makes it very difficult for new nodes to be added in the vicinity of a distant node in the pioneer network, whose nodes are widely spread across the unit square as a result of its low α_p, in order to form a new spatial cluster. We do however see a significantly higher level of spatial concentration in N_2^O than was seen in N_2^S, N_2^M, or N_2^G as nodes in the settler network N_2^O cluster around a single node of the pioneer network N_2^P. The average node distance of N_2^O is 0.0532 compared to 0.1233 for N_2^S, 0.1767 for N_2^M and 0.1307 for N_2^G. Figure 2.41 shows the frequency distribution of the distances between each pair of nodes in the networks N_2^S and N_2^O. We can see that both distributions are unimodal, consistent with with spatial networks whose nodes are distributed in a single spatial cluster. The distribution of N_2^O. has both a lower spread of distances and a lower mode.



(a) Distribution of the distances between pairs of nodes in the network N_2^S generated by the original Kaiser-Hilgetag algorithm and analysed in Section 1.3.

(b) Distribution of the distances between pairs of nodes in the settler network N_2^O seen in Figure 2.39b.

Figure 2.41: Node distance distributions for N_2^S and N_2^O.

The formation of multiple spatial clusters in N_1^O has also resulted in a network which is highly connected locally but has few links between spatial clusters. This has caused N_1^O to have a relatively high clustering coefficient while maintaining quite a low density. The clustering coefficient of N_1^O is 0.1450 and its density is 0.0484. N_1^S, generated with the original simple algorithm has a clustering coefficient of 0.0639 and a density of 0.032. The average shortest path length of N_1^O is 3.8063, lower than the corresponding value of 4.638 for N_1^S.

The high level of spatial concentration in N_2^O has also had effects on its network topology. The proximity of the added nodes has resulted in the formation of a higher number of links than in the corresponding network generated by the original algorithm. N_2^O has a slightly higher density and higher clustering coefficient than N_2^S, a density of 0.0416 compared to 0.028 and a clustering coefficient of 0.0680 compared to 0.0502.The average shortest path length of N_2^O is 3.4337 compared to a value of 5.3457 for N_2^S.

4.2 Results

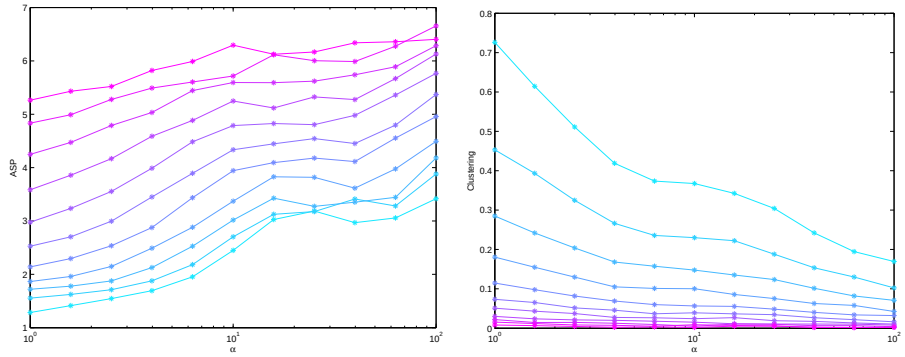
In this section we investigate the effect of different parameter value combinations on the spatial settler networks generated by our algorithm described in Section 4.1. Our analysis again focuses on the changes in the four metrics defined in Section 1.2. For each parameter combination we simulated 50 100-node networks. Of particular interest to us are the effects of the interactions between different values of α and β, and of changing σ^2 on the properties of the settler networks produced by our algorithm. For simplicity we have generated the 10-node pioneer network in every case using the original Kaiser-Hilgetag algorithm, described in Section 1.3 with $\alpha = 3$, $\beta = 0.4$ and a single starting node with position $(0.4, 0.5)$.

4.2.1 Varying α and β

We first look at the effect of different combinations of α and β on the settler networks generated when $\sigma^2 = 0.001$. As we have done previously, we take 11 values of α between 1 and 100, 11 values of β between $\frac{1}{100}$ and 1, and generate multiple spatial networks using our algorithm for each parameter combination. In Figure 2.42 and Figure 2.43 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination.

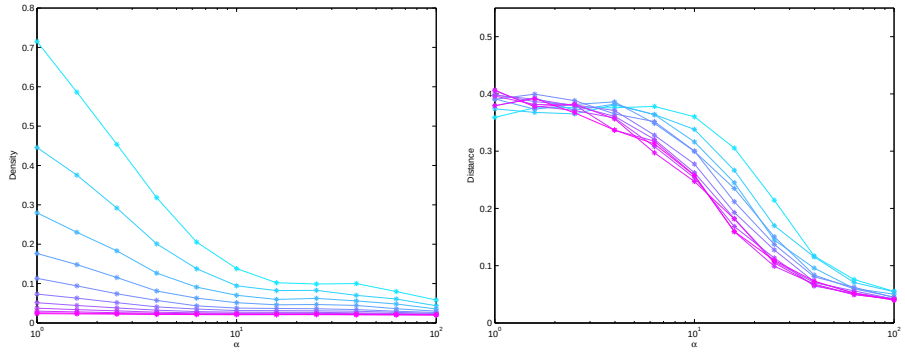
The relationship between $\log(\alpha)$ and the average node distance of the settler network generated by our algorithm is sigmoidal, similar to the relationships seen in our analysis of the original Kaiser-Hilgetag algorithm in Section 1.3, our extension allowing connections between mature nodes in Section 2 and our extension which considered growth of individual nodes in Section 3. This relationship is negative, the average node distance of the network decreases as we increase α. This is again due to the fact that, according to our algorithm, the probability, $P_1 = \beta e^{-\alpha d(v_i, v_e)}$, of a new node connecting with an existing node decays exponentially with distance scaled by α. This means that as we increase

α new nodes must be closer to the existing network in order to be added. The decrease in average node distance is relatively slow as we initially increase α on the log scale this decrease in average node distance becomes more rapid as we increase α before slowing again as $\alpha \to \infty$. The initial slow rate of decrease of average node distance as we increase $\log(\alpha)$ is due to the fact as $\alpha \to 0$ average node distance approaches a limit imposed by the finite space in which our network is located and the truncated Gaussian mix distribution with small σ^2, generated by a small, finite pioneer network from which the spatial location of nodes we attempt to add is chosen. Once again average node distance tends to 0 as $\alpha \to \infty$.



(a) *ASP* of the generated networks for different combinations of α and β.

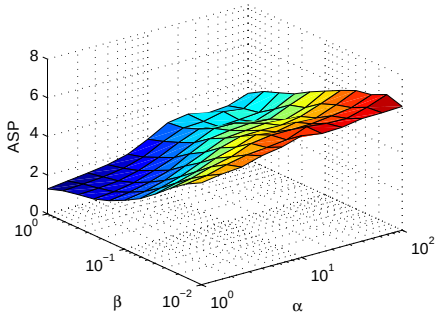
(b) Clustering coefficient of the generated networks for different combinations of α and β.



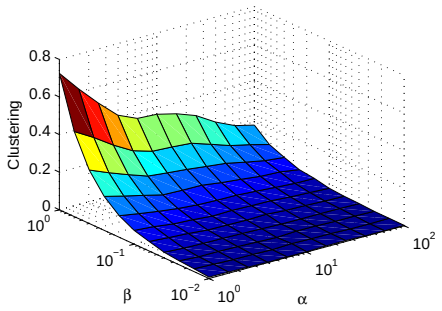
(c) Density of the generated networks for different combinations of α and β.

(d) Average node distance of the generated networks for different combinations of α and β.

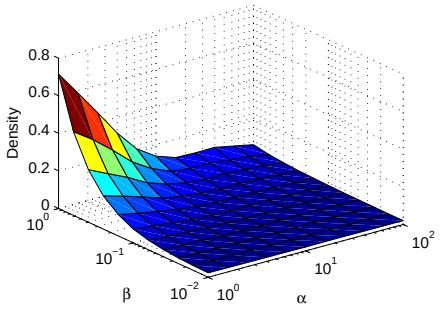
Figure 2.42: Comparison of the properties of the settler networks generated by our algorithm for different values of α and β, $\sigma^2 = 0.001$. We plot in increasingly magenta for lower β.



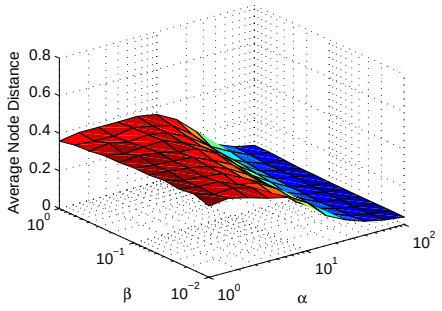
(a) *ASP* of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.



(d) Average node distance of the generated networks for different combinations of α and β.

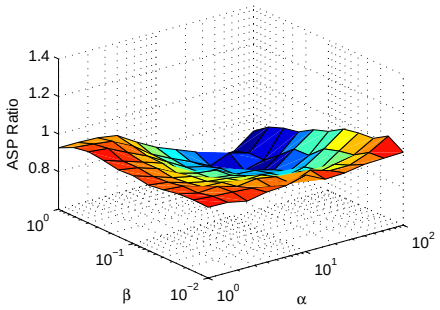
Figure 2.43: Comparison of the properties of the settler networks generated by our algorithm for different values of α and β, $\sigma^2 = 0.001$.

The relationship between β and the average node distance of the network generated is not so clear. Previously we have seen a positive relationship between β and average node distance caused by the fact that links of all lengths are more likely to form as we increase β, increasing the number of spatially distant nodes added to the network during its formation. This positive relationship is observed for intermediate and high values of α, however, for low values of α the relationship between β and average node distance is not seen to necessarily be positive. This perhaps due to the fact that for low values of α and high β almost every node we attempt to add to our network will form a link and be added as a result. Since the spatial distribution of the nodes we attempt to add will be relatively concentrated, due to the truncated Gaussian mix with small σ^2 from which we choose their locations, this will not result in an increase in average node distance as was the case when we chose these locations from a uniform distribution.

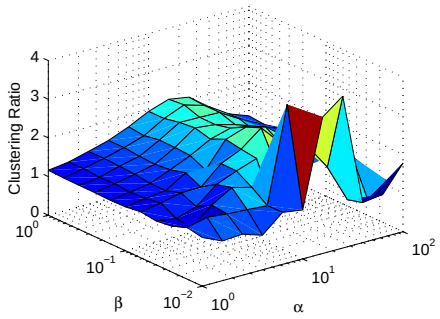
As was the case for the original Kaiser-Hilgetag algorithm, there is generally a positive relationship between β and density with every potential link being

more likely to form as we increase β. For low β density nears its lower bound 0.02. A similar relationship is observed between β and the clustering coefficient of the generated settler network. For most values of α we also see that the higher density in the network due to an increased β causes the ASP of the generated settler networks to be reduced.

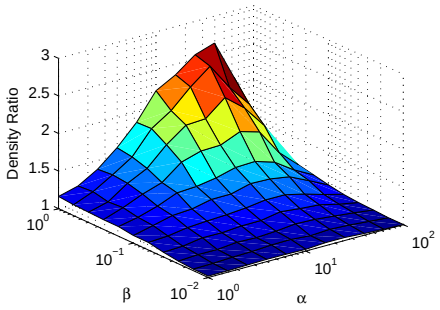
As we increase α we generally see a decrease in density and clustering and an increase in ASP. However the relationships, are not entirely monotonic. Perhaps the intermediate fluctuations are due to switches in the number of spatial clusters that the settler network can sustain as α increases.



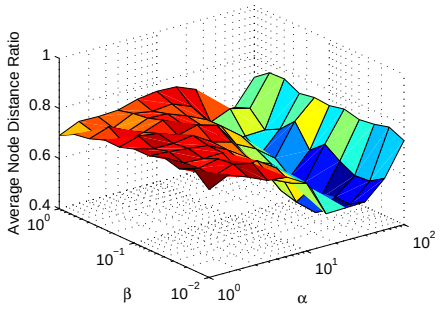
(a) Ratio of the ASP of the generated networks for different combinations of α and β to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and β to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of α and β to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(d) Ratio of the average node distance of the generated networks for different combinations of α and β to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

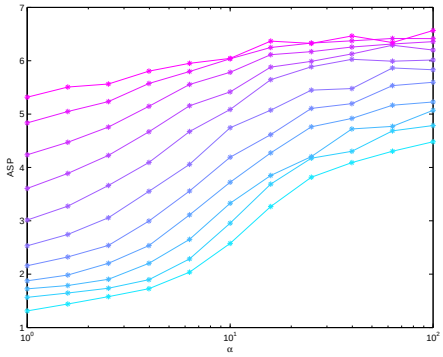
Figure 2.44: Ratios of the properties of settler networks generated by our algorithm for different values of α and β with $\sigma^2 = 0.001$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

In order to compare the 100-node settler networks generated by our new algorithm, which chooses the spatial location of potential new network nodes based on a truncated Gaussian mix distribution generated by a pioneer network, to the original Kaiser-Hilgetag algorithm,which chooses these locations from a uniform distribution, we have plotted the ratios of the value of each of the four network metrics for a network generated by the new algorithm to the corresponding value of that metric for a network generated by the original algorithm in Figure 2.44.

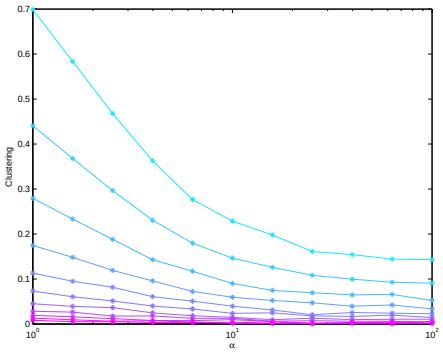
For all three topological metrics we note that the extension to the model appears to have the largest effect (our ratio is furthest from 1) for intermediate values of α and large values of β. These differences are driven by the increased number of links, i.e. increased density, in the networks generated by our new algorithm for these values of α and β compared to the corresponding networks generated by the original algorithm. We hypothesise that this increased density is the result of the increased number of links formed by the average new node added to the settler network due to its position being chosen very close to a cluster of spatially concentrated nodes as a result of the presence of the pioneer network. This effect is most pronounced for intermediate values of α as for these values a small number of highly spatially concentrated and highly connected clusters of nodes may be sustained by the network, resulting in far higher spatial concentration and density than in the corresponding networks generated by the original algorithm. For lower values of α this effect is less as we see the formation of a high number of highly spatially concentrated clusters of nodes which are themselves widely spaced across the unit square. For higher values of α the effect is less as the single highly spatially concentrated cluster which forms around a single pioneer node in this case is quite similar to the highly concentrated spatial cluster that forms in the absence of the pioneer network.

The effects of the formation of these spatial clusters due to the pioneer network can also explain the differences in average node distances between networks generated using the two different methods. In all cases due to the low σ^2 value our average node distance is lower for networks generated using the new algorithm. The effect is largest for intermediate values of α when a small number of spatial clusters form in the case of the new algorithm and nodes are relatively widely distributed in the case of the original algorithm.

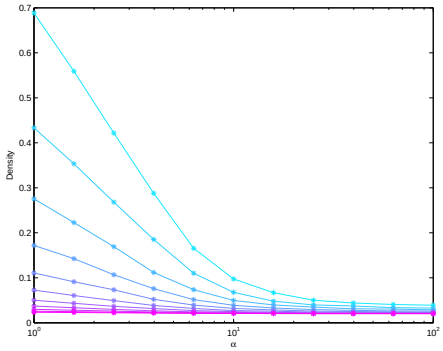
We now look to see what changes, if any, occur in these relationships if we increase σ^2 to 0.01. The relationship between $\log(\alpha)$ and the average node distance of the settler network generated by our algorithm is once again negative and sigmoidal. As in the $\sigma^2 = 0.001$ case the relationship between β and the average node distance of the network generated is not entirely clear for low values of α. The positive relationship between average node distance and β does however emerge earlier in this case than in the $\sigma^2 = 0.001$ case.



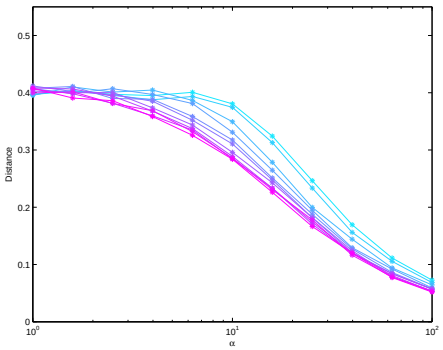
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



(c) Density of the generated networks for different combinations of α and β.

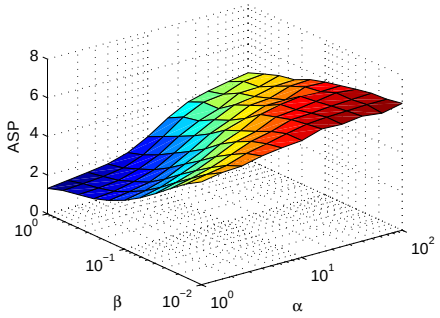


(d) Average node distance of the generated networks for different combinations of α and β.

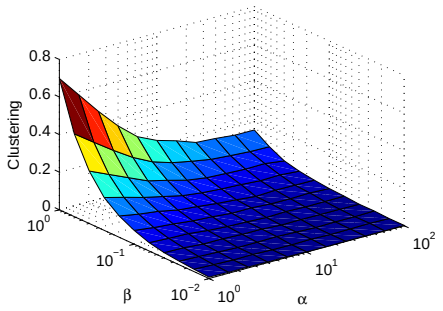
Figure 2.45: Comparison of the properties of the settler networks generated by our algorithm for different values of α and β, $\sigma^2 = 0.01$. We plot in increasingly magenta for lower β.

Again, as was the case for the original Kaiser-Hilgetag algorithm, there is generally a positive relationship between β and density with every potential link being more likely to form as we increase β. For low β density nears its lower bound 0.02. A similar relationship is observed between β and the clustering coefficient of the generated settler network. For most values of α we also see that the higher density in the network due to an increased β causes the ASP of the generated settler networks to be reduced.

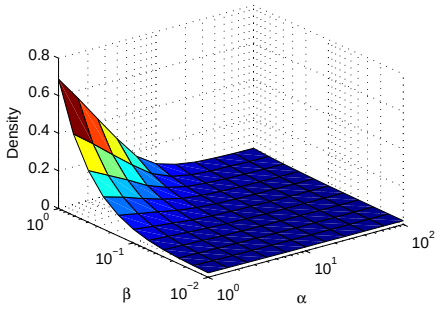
As we increase α we generally see a decrease in density and clustering and an increase in ASP. However the relationships, are not entirely monotonic. The intermediate fluctuations however, do seem to be less significant than those seen in the $\sigma^2 = 0.001$ case.



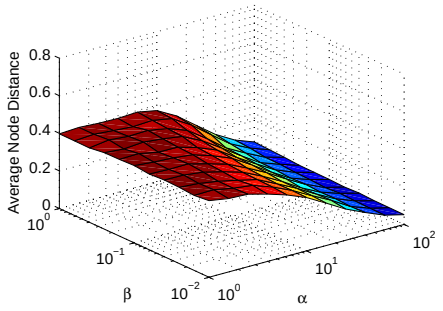
(a) *ASP* of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



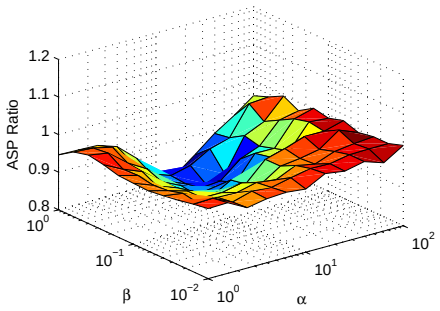
(c) Density of the generated networks for different combinations of α and β.



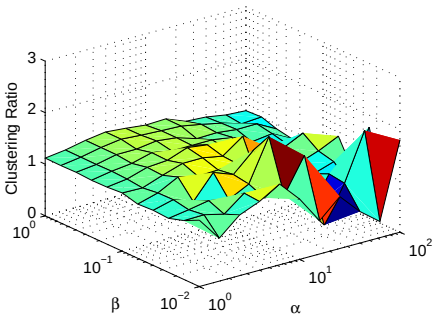
(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.46: Comparison of the properties of the settler networks generated by our algorithm for different values of α and β, $\sigma^2 = 0.01$.

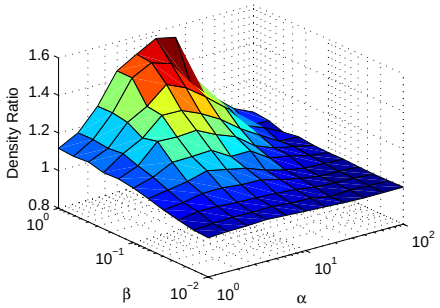
The ratios plotted in Figure 2.47 allow us to compare the networks generated by this algorithm when $\sigma^2 = 0.01$ to the corresponding networks generated by the original algorithm. We see that for large values of α the networks generated are the same and the pioneer-settler network structure has no effect and so we see ratios for all four metrics ≈ 1. This is due to the fact that the spatial constraints put on the position of new nodes in being forced to be close to the pioneer network are far less constraining than those imposed by the need for new links to form between new nodes and the existing network when α is high. For lower values of α we also see that the effect is less than in the $\sigma^2 = 0.001$ case, for example the maximum density ratio observed is ≈ 1.6 while in the $\sigma^2 = 0.001$ case we saw in Figure 2.44 the maximum density ratio observed was ≈ 2.6.



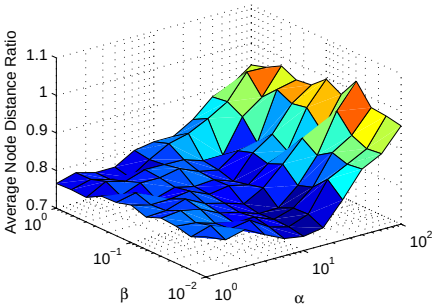
(a) Ratio of the ASP of the generated networks for different combinations of α and β to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and β to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



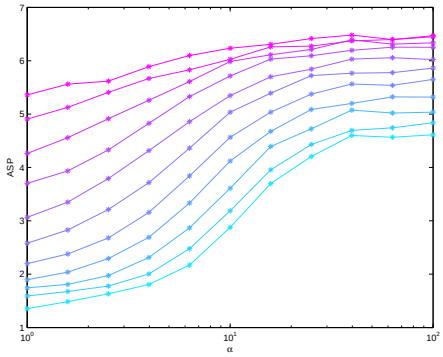
(c) Ratio of the density of the generated networks for different combinations of α and β to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



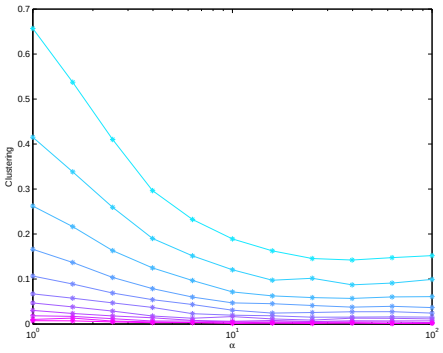
(d) Ratio of the average node distance of the generated networks for different combinations of α and β to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.47: Ratios of the properties of settler networks generated by our algorithm for different values of α and β with $\sigma^2 = 0.01$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

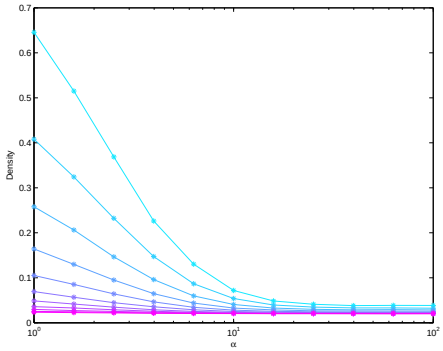
As we increase σ^2 further we would expect to see he resulting networks to resemble even more closely those generated by the original algorithm as when we increase σ^2 the truncated Gaussian mix distribution becomes approximately equal to a uniform distribution. We now look at the $\sigma^2 = 0.1$ case and see that this is indeed what occurs.



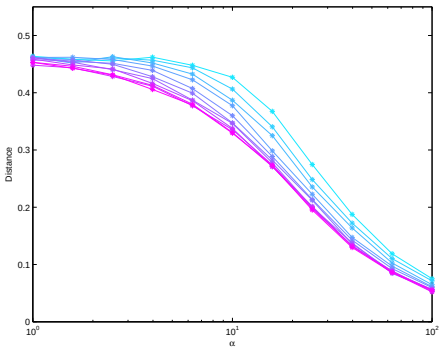
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



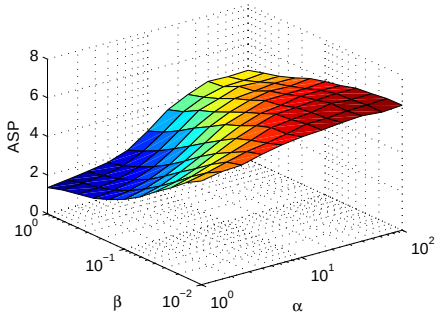
(c) Density of the generated networks for different combinations of α and β.



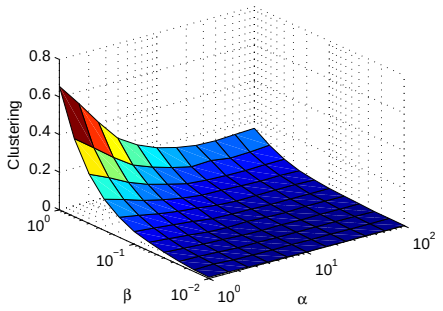
(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.48: Comparison of the properties of the settler networks generated by our algorithm for different values of α and β, $\sigma^2 = 0.1$. We plot in increasingly magenta for lower β.

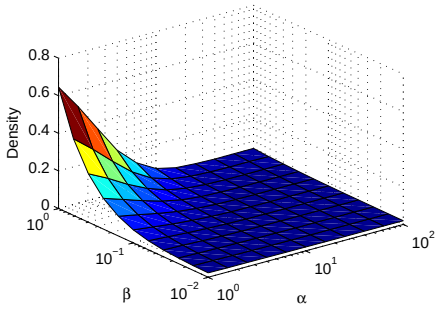
In Figure 2.48 and Figure 2.49 we can see the mean of the ASP, clustering coefficient, density and average node distance of the networks generated for each parameter combination. There now seems to be a positive relationship between β and the average node distance of the network generated for almost every value of α as was the case for the original algorithm as seen in Section 1.3. The maximum average node distance observed is now also closer to that seen for the original algorithm than for either the $\sigma^2 = 0.001$ or the $\sigma = 0.01^2$ case. The relationships between α and β and the three topological networks also look very similar to those seen for the original algorithm.



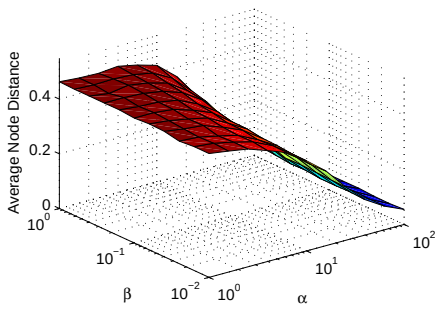
(a) ASP of the generated networks for different combinations of α and β.



(b) Clustering coefficient of the generated networks for different combinations of α and β.



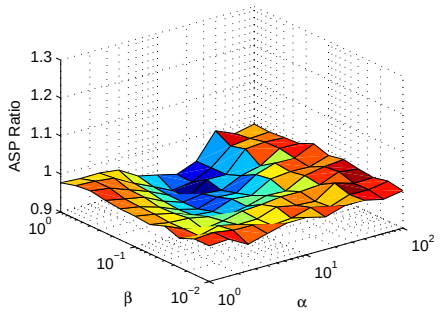
(c) Density of the generated networks for different combinations of α and β.



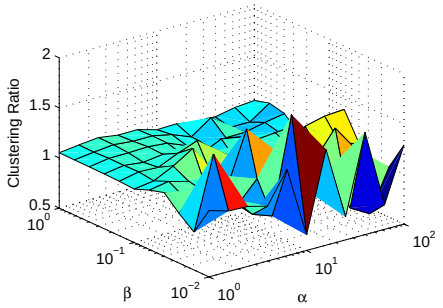
(d) Average node distance of the generated networks for different combinations of α and β.

Figure 2.49: Comparison of the properties of the settler networks generated by our algorithm for different values of α and β, $\sigma^2 = 0.1$.

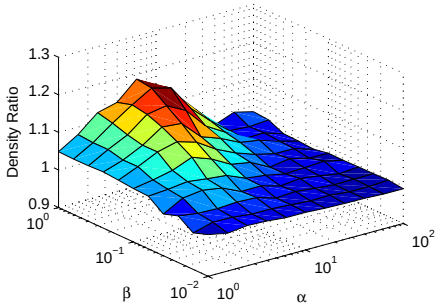
The ratios plotted in Figure 2.50 allow us to more easily compare the networks generated by this algorithm when $\sigma^2 = 0.1$ to the corresponding networks generated by the original algorithm. We see that for values of $\alpha > 10$ the networks generated are essentially the same and the pioneer-settler network structure has no effect and so we see ratios for all four metrics ≈ 1. For values of $\alpha < 10$ any effect of the pioneer-settler structure is quite small with all four ratios remaining relatively close to 1. For example the maximum density ratio is less than 1.2, the minimum ASP ratio is greater than 0.9 and the minimum average node distance ratio is ≈ 1.



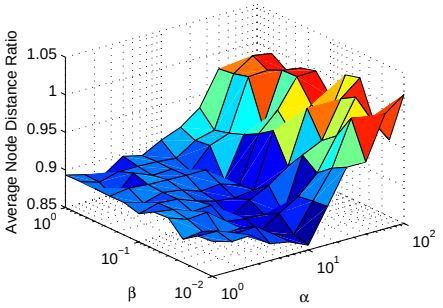
(a) Ratio of the ASP of the generated networks for different combinations of α and β to the ASP of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(b) Ratio of the clustering coefficient of the generated networks for different combinations of α and β to the clustering coefficient of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(c) Ratio of the density of the generated networks for different combinations of α and β to the density of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.



(d) Ratio of the average node distance of the generated networks for different combinations of α and β to the average node distance of networks generated by the original Kaiser-Hilgetag algorithm with the same combinations of α and β.

Figure 2.50: Ratios of the properties of settler networks generated by our algorithm for different values of α and β with $\sigma^2 = 0.1$ to the same properties for networks generated by the Kaiser-Hilgetag algorithm

5 Conclusion

In this chapter we introduced three new models of spatial network growth. These new models are adaptations and extensions of a model due to Kaiser and Hilgetag [41] which was initially motivated by a wish to analyse the growth and structure of biological networks. We allow spatial networks to grow by adding nodes over time whose positions are not predetermined rather their position and the links between them are established over time based on spatial and other constraints as the network grows and develops.

The first extension we introduced in this chapter allows for links to be established between pairs of existing nodes in the network rather than limiting link formation between pairs of nodes to the moment one of the pair is first added to the network. The second extension we introduced allowed nodes to grow according to a map, in the examples analysed here this map is the network logistic map which we will introduce formally and analyse in more detail in Chapter 3, and allowed link formation to depend on node size. The final extension looked at the case of overlapping spatial networks in which the growth and structure of one of the networks, known as the pioneer network, affects the development of the other, known as the settler network. All three extensions were shown to have significant effects on the structure of the networks generated, both in terms of topology and spatial distribution.

It is possible to combine two or more of these extensions in the same model of spatial network growth. However, we do not examine the results of doing so in any detail here. We believe it would be both useful and interesting to do so in the future. Indeed in Chapter 4 our models of the economy incorporates the ideas of all three of the extensions analysed in this chapter. We introduce an approach to modelling the economy as a series of overlapping spatial networks where links can develop between pairs of existing nodes, and nodes grow over time according to a map. There is also great scope to develop these extensions further, for example it is possible to have an element of feedback in the case of the overlapping networks where both networks have an effect on the spatial distribution of the nodes in the other network or to have more complicated models of node growth in which perhaps the structure and spatial distribution of one network has an effect on the growth of nodes in another.

Chapter 3

Node Growth

1 Introduction

In this chapter we investigate growth of individual nodes in a network, $N = (V, E)$. We simulate our systems throughout in Matlab [47]. In our model the growth of nodes in the network is governed by a discrete time evolution operator known as a *map*. Maps are a specific type of dynamical system that are well suited to our purposes. Before continuing we first introduce some general information about the analysis of dynamical systems and in particular the analysis of maps and their long-term behaviour.

1.1 Dynamical Systems

A dynamical system is made up of a set of states, X, known as the state space and a dynamical rule that specifies the immediate future of all state variables, in terms of past states. A dynamical system can have discrete or continuous time. A deterministic system with discrete time is defined by a *map*.

Definition 1.1.1. A deterministic evolution operator, f, with discrete time, $T = \mathbb{N}$, and a continuous state space, $X \subseteq \mathbb{R}^m$, is called a map,

$$f : X \to X,$$

where the state evolution is defined by $x(n+1) = f(x(n))$ for $x(n) \in X$, $n \in T$.

A dynamical system with continuous time is essentially the limiting case of discrete-time systems with smaller and smaller updating times. A deterministic system with continuous time is defined by a *flow*.

Definition 1.1.2. A deterministic evolution operator, ϕ, with continuous time, $T = \mathbb{R}$, and a continuous state space, $X \subseteq \mathbb{R}^m$, is called a flow,

$$\phi : X \to X,$$

where evolution is defined by $\frac{d}{dt}x(t) = \phi(x(t))$ for $x(t) \in X$, $t \in T$.

For our purposes it is useful to consider systems with discrete time and so our focus will be on maps.

1.1.1 Fixed Points, Periodic Orbits & Stability

Of particular interest in the analysis of any dynamical system that we may consider is the concept of *homeostasis*, *equilibrium* or *steady state*. We wish to find the points of the system at which there is no change. In the context of a map, a fixed point or steady state, $x^* \in X$, of the system is some value that satisfies

$$x(n+1) = x(n) = x^*, \tag{3.1}$$

i.e a point at which there is no change in the system from time n to time $n+1$. For a map where the evolution is defined by $x(n+1) = f(x(n))$, a fixed point x^* satisfies the relation

$$x^* = f(x^*). \tag{3.2}$$

We are able to distinguish between two different types of fixed points. If all points sufficiently close to x^* evolve even closer to x^* under iteration, then we call x^* a *stable fixed point*, an *attractor* or a *sink*. On the other hand if the converse is true and all points sufficiently close to x^* evolve away from $x*$ under iteration, then we call x^* an *unstable fixed point*, a *repellor* or a *source*. A good analogy is that a ball balanced on the tip of the peak of a mountain is unstable as it will roll down the mountain and away from the peak if it is moved even slightly, while a ball resting at the bottom of a valley is stable as when it is moved slightly it will roll back down to the bottom of the valley to its original resting place. More formally, we have the following two definitions where $||x|| = \sqrt{x_1^2 + x_2^2 + \dots x_m^2}$ is the Euclidean length of the vector $x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$.

Definition 1.1.3. Let f be a map and $f(x^*) = x^*$. If there is an $\epsilon > 0$ such that for all $x \in \{x \in X : ||x - x^*|| < \epsilon\}$, $f^k(x) \to x^*$ as $k \to \infty$, where f^k denotes the kth composition of the map f, then x^* is a stable fixed point.

Definition 1.1.4. Let f be a map and $f(x^*) = x^*$. If there is an $\epsilon > 0$ such that there exists $K \in \mathbb{N}$ such that for all $k \geq K$ and all $x \in \{x \in X : ||x - x^*|| < \epsilon\}$ other than x^* itself $f^k(x) \notin \{x \in X : ||x - x^*|| < \epsilon\}$ then x^* is an unstable fixed point.

These definitions lead us to the following result for one-dimensional maps through a process of *linearisation*.

Theorem 1.1.1. *[1] Let f be a differentiable map on $X \subseteq \mathbb{R}$, $f(x(n)) = x(n+1)$. Given a fixed point x^*, i.e. $f(x^*) = x^*$, then*

1. *if $|f'(x^*)| < 1$, x^* is a stable fixed point,*
2. *if $|f'(x^*)| > 1$, x^* is an unstable fixed point,*

where

$$|f'(x^*)| = \lim_{x \to x^*} \left| \frac{f(x) - f(x^*)}{x - x^*} \right| = \lim_{x \to x^*} \left| \frac{f(x) - x^*}{x - x^*} \right|$$

(magnification factor).

As we wish to model the growth of individual nodes in networks with more than one node we will mainly deal with higher dimensional maps. In this case we arrive at an analogous result which requires the definition of the *Jacobian matrix* of a map.

Definition 1.1.5. Let $f = (f_1, f_2, \dots, f_m)$ be a map on $X \subseteq \mathbb{R}^m$, and let $x \in \mathbb{R}^m$. The Jacobian matrix of f at x, which we denote $\mathbf{J}(x)$, is the matrix

$$\mathbf{J}(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_m}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_m}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_m}(x) \end{pmatrix}$$

of partial derivatives evaluated at x.

We can now examine the properties of the eigenvalues of the Jacobian matrix of f, evaluated at a fixed point x^*, in order to determine the stability of the fixed point.

Theorem 1.1.2. *[1] Let f be a differentiable map $X \subseteq \mathbb{R}^m$, $f(x(n)) = x(n+1)$. Given a fixed point x^*, i.e. $f(x^*) = x^*$, then*

1. *if for each eigenvalue λ_i of $\mathbf{J}(x)$, $|\lambda_i| < 1$, x^* is a stable fixed point,*
2. *if for some eigenvalue λ_i of $\mathbf{J}(x)$, $|\lambda_i| > 1$, x^* is an unstable fixed point.*

For a 2-dimensional system it is sufficient to test whether the condition

$$2 > 1 + c > |b| \tag{3.3}$$

is satisfied, where $\lambda^2 - b\lambda + c = 0$ is the characteristic equation of $\mathbf{J}(x^*)$.

We are also interested in the periodic orbits of a map. Instead of remaining at the same point in a periodic orbit a map will oscillate between a finite set of points, returning to the same points in the same order each time.

Definition 1.1.6. Let f be a map on $X \subseteq \mathbb{R}^m$. We call p a periodic point of period k if $f^k(p) = p$, and k is the smallest such positive integer. The orbit with the initial point p is called a periodic orbit of period k.

Just as in the case of fixed points it is important to distinguish between stable periodic orbits, which we will return to after a small perturbation away, and unstable periodic orbits where a small perturbation away will be magnified under iteration. Using the fact that a periodic point of period k for f is a fixed point for the map f^k we can easily define what it means for a periodic orbit to be stable.

Definition 1.1.7. Let f be a map and p a period k point with corresponding period-k orbit $\{p, f(p), f^2(p), \ldots, f^{k-1}(p)\}$.

1. The period-k orbit of p is stable (an attractor) if p is a stable fixed point for f^k.
2. The period-k orbit of p is unstable (a repellor) if p is an unstable fixed point for f^k.

Again using the fact that a periodic point of period k for f is a fixed point for the map f^k we can apply Theorem 1.1.1 and Theorem 1.1.2 along with the chain rule to arrive at the following results for one and higher-dimensional maps.

Theorem 1.1.3. *[1] Let f be a differentiable map on $X \subseteq \mathbb{R}$, $f(x(n)) = x(n+1)$. Given a periodic point of period k for f, p, then the corresponding period-k orbit $\{p_1, p_2, p_3, \ldots, p_k\}$*

1. *is stable if $|f'(p_1)f'(p_2)\ldots f'(p_k)| < 1$,*

2. *is unstable if* $|f'(p_1)f'(p_2)\dots f'(p_k)| > 1$.

Theorem 1.1.4. *[1] Let* f *be a differentiable map* $X \subseteq \mathbb{R}^m$, $f(x(n)) = x(n+1)$. *Given a periodic point of period* k *for* f, p, *then the corresponding period-*k *orbit* $\{p_1, p_2, p_3, \dots, p_k\}$

1. *is stable if for each eigenvalue* λ_i *of the matrix product,* $\mathbf{J}(p_1)\mathbf{J}(p_2)\dots\mathbf{J}(p_k)$, $|\lambda_i| < 1$,

2. *is unstable if for some eigenvalue* λ_i *of the matrix product,* $\mathbf{J}(p_1)\mathbf{J}(p_2)\dots\mathbf{J}(p_k)$, $|\lambda_i| > 1$.

1.1.2 Bifurcations and Bifurcation Diagrams

Bifurcation is the name given to a qualitative change in the phase portrait of a dynamical system under parameter variation. In other words, bifurcation refers to changes in behaviour such as a switch from fixed point to an oscillation between two points or other significant changes such as changes to higher periodicities, quasi-periodicity and *chaos*. The term was introduced to the world by the *"father of dynamical systems"*, French mathematician, Henri Poincaré.

Bifurcation diagrams are useful diagrams which provide a summary of the essential dynamics of a system. They show the appearance, evolution and disappearance of attracting sets under parameter variation. For a map with parameter r, $f_r(x)$, a bifurcation diagram can be produced by choosing an initial value for x and calculating the orbit of x under $f_r(x)$ for values of r in an appropriate range. We ignore the first N iterates for some large $N \in \mathbb{N}$ iterates and plot the orbit. The plotted points then approximate either fixed or periodic attractors or other attracting sets such as quasi-periodic or chaotic attractors. This type of bifurcation diagram is sometimes referred to as a brute-force bifurcation diagram.

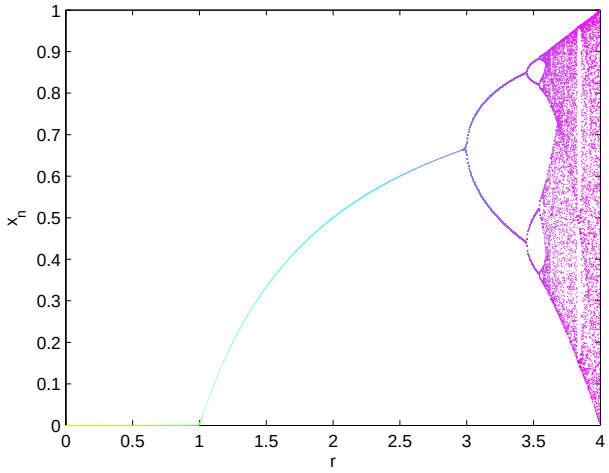


Figure 3.1: Bifurcation diagram for the map $f_r(x) = rx(1-x)$ for values of $r \in (0, 4]$.

In Figure 3.1 we give an example of a bifurcation diagram of the map $f_r(x) = rx(1-x)$ for values of $r \in (0, 4]$. This map is known as the *logistic map*. We see for example that this map has a stable fixed point for $r = 2$ and at $r = 3.25$ the diagram shows the two points of a period-two attractor. For values of r closer to 4 we see that the orbit appears to fill a subinterval of $[0, 1]$ randomly. This represents a chaotic attractor. Chaos, unlike fixed or periodic behaviour, is quite hard to describe. Essentially a chaotic orbit is one that is bounded, non-periodic and displays sensitive dependence on the initial conditions. In other words even a slight perturbation away from our initial condition will produce a significantly different orbit.

We will return to analyse the dynamics of the map $f_r(x)$, known as the logistic map, in more detail in Section 2. Now, we concern ourselves with giving a more precise definition of chaos.

1.1.3 Lyapunov Exponents

In order to quantify chaos we introduce the concept of *Lyapunov exponents* [45]. Chaos, as mentioned in Section 1.1.2, can be considered, in essence, as a highly sensitive dependence on initial conditions. Lyapunov exponents measure the exponential divergence of nearby trajectories and so can give us a quantitative measure of chaos. We will consider a strictly positive maximal Lyapunov exponent as the definition of chaos.[1]

Consider a system allowed to evolve from two slightly differing initial points, x_0 and $x_0+\epsilon_0$, where $|\epsilon_0| \lll 1$. Let ϵ_n be the separation between the two points after n iterations, then

$$|f^n(x_0 + \epsilon_0) - f^n(x_0)| = |\epsilon_n| \approx |\epsilon_0| e^{\mu n}. \tag{3.4}$$

Here μ the Lyapunov exponent which gives the average rate of divergence. If μ is negative we observe convergence of nearby trajectories; while if μ is positive we observe divergence of slightly separated trajectories. Divergence of nearby trajectories implies that the evolution of the system is sensitive to the initial conditions. We say that a system is chaotic if it has a positive Lyapunov exponent. Taking logarithms we find

$$\mu \approx \frac{1}{n} \log \left| \frac{f^n(x_0 + \epsilon_0) - f^n(x_0)}{\epsilon_0} \right|. \tag{3.5}$$

Taking the limit as $\epsilon_0 \to 0$ this becomes

$$\begin{aligned} \mu &\approx \frac{1}{n} \log \left| \frac{df^n}{dx} \right| \\ &= \frac{1}{n} \log |(f^n)'(x)| \end{aligned}$$

[1] A stricter definition can be found in [1] and would require us to ensure orbits are not asymptotically periodic, however, in this thesis we will limit ourselves to examining Lyapunov exponents for indications of chaotic behaviour.

$$= \frac{1}{n}\log\left|\prod_{k=0}^{n-1} f'(x(k))\right|$$
$$= \frac{1}{n}\sum_{k=0}^{n-1}\log|f'(x(k))|\,. \tag{3.6}$$

Finally taking the limit as $n \to \infty$ we obtain an expression for the Lyapunov exponent

$$\mu = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\log|f'(x(k))|\,. \tag{3.7}$$

For m-dimensional maps there are m Lyapunov exponents, since stretching can occur along each of the m axes. An m dimensional initial volume develops on average as

$$V = V_0 e^{(\mu_1+\mu_2+\ldots+\mu_m)n}. \tag{3.8}$$

1.2 Logistic Maps

When defining node growth in our networks we take inspiration from a particular map known as the logistic map, given by

$$x(n+1) = \lambda x(n)(1-x(n)), \quad x(n)\in[0,1]\,,\ n\in T,\ \lambda\geq 0. \tag{3.9}$$

The logistic map was made famous by the theoretical physicist and mathematical biologist Lord Robert May in his 1976 article *Simple mathematical models with very complicated dynamics* [48]. In it he showed that even simple and deterministic maps such as the logistic map can exhibit a surprising array of dynamical behaviour, including chaos. The logistic map itself is named after the corresponding differential equation

$$\frac{dx}{dt} = \lambda x(1-x), \quad \lambda\geq 0, \tag{3.10}$$

that was first published by Pierre François Verhulst in 1845 [68]. This differential equation can be seen as a model of the growth of a population whose reproductive rate is density-regulated due to environmental constraints.

We model the growth of each node in a network over time using a modified form of the logistic map. The size, y_i, of an individual node, $v_i \in V$, at time $n+1$ is given by

$$y_i(n+1) = r_i(n)y_i(n)\left(1-\frac{y_i(n)}{Y}\right), \quad Y>0, \tag{3.11}$$

where the constant Y represents the maximum possible size of a node. This set-up allows us to vary our control parameter. In place of a constant control parameter λ we take a function $r_i(n)$ of time which varies between nodes.

For simplicity we consider the ratio of each node's size to the maximum possible size, $x_i(n) = \dfrac{y_i(n)}{Y}$, and so (3.11) reduces to

$$x_i(n+1) = r_i(n)x_i(n)(1-x_i(n)) \quad x_i(n)\in(0,1]\,. \tag{3.12}$$

2 Growth of a Single Isolated Node

To develop an understanding of the classic logistic map we first we look at the growth of a single isolated node v. We take $r(n)$ to be a constant $r \in (0, 4]$ and so our growth model for the size, $x(n)$, of v reduces to the logistic map

$$x(n+1) = rx(n)(1 - x(n)) = f(x(n)). \tag{3.13}$$

Let us now analyse the long-term dynamics of this system.

2.1 Analysis

2.1.1 Fixed Points

As is the case when analysing any dynamical system the fixed points of the system are of particular interest. The fixed points of the logistic map are given by x^* where

$$\begin{aligned} x^* &= rx^*(1 - x^*) \implies \\ x^* &= rx^* - rx^{*^2} \implies \\ 0 &= rx^* - x^* - rx^{*^2} \implies \\ 0 &= rx^* \left(\frac{r-1}{r} - x^* \right) \implies \\ x_a^* = 0, \qquad & x_b^* = \frac{r-1}{r}. \end{aligned} \tag{3.14}$$

For $r < 1$ the fixed point $x_b^* = \dfrac{r-1}{r}$ is negative and so it does not exist. For $r < 1$ it is in a sense *unphysical*.

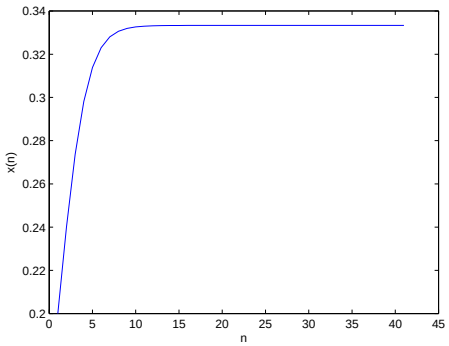


Figure 3.2: Node size converging to the fixed point $x_b^* = \frac{1}{3}$ for $r = \frac{3}{2}$.

In Figure 3.2 we see an example of the convergence of the node size $x(n)$ to the fixed point $x_b^* = \frac{r-1}{r} = \frac{1}{3}$ for $r = \frac{3}{2}$ as $n \to \infty$. The convergence is monotonic and it appears that for this value of r the fixed point x_b^* is *stable*.

An overall picture of the stability of the two fixed points, $x_a^* = 0$ and $x_b^* = \frac{r-1}{r}$, for all values of $r \in (0, 4]$ is of interest. We wish to know, for each value of r, whether each of these fixed points is a stable attractor, sometimes known as a *sink*, or an unstable repellor, sometimes known as a *source*. To do this we examine the derivative of $f(x) = rx(1-x)$, given by

$$f'(x) = r - 2rx. \tag{3.15}$$

Any fixed point, x^*, of the map given by $x(n+1) = f(x(n))$ is stable when

$$|f'(x^*)| < 1. \tag{3.16}$$

First we examine the stability of the fixed point $x_a^* = 0$. We have that $x_a^* = 0$ is stable for

$$|f'(x_a^*)| = |f'(0)| = |r| < 1 \tag{3.17}$$

We only consider cases with $r > 0$. And so the fixed point $x_a^* = 0$ is stable for

$$0 < r < 1. \tag{3.18}$$

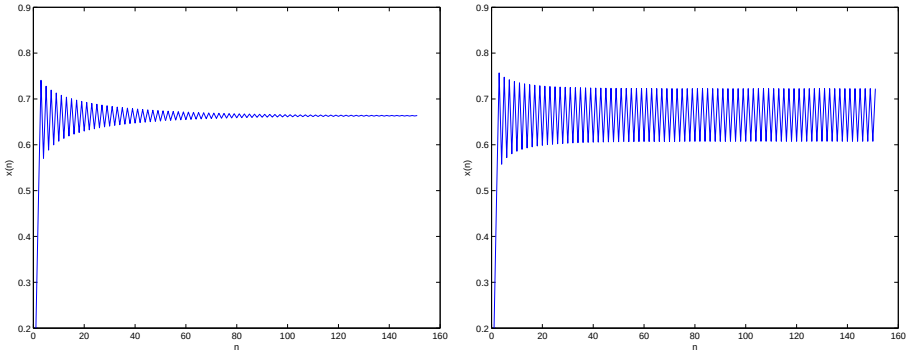
We now examine the stability of the second fixed point, ${}_b x^* = \frac{r-1}{r}$. We have that $x^* = \frac{r-1}{r}$ is stable for

$$|f'(x_b^*)| = \left| f'\left(\frac{r-1}{r}\right)\right| = \left| r - \frac{2r^2 - 2r}{r}\right| = |2 - r| < 1 \tag{3.19}$$

Since $x_b^* = \frac{r-1}{r}$ is unphysical for $r < 1$ we only consider cases with $r > 1$. And so $x_b^* = \frac{r-1}{r}$ is stable for

$$1 < r < 3: \ = r_c. \tag{3.20}$$

The following Figures show the behaviour of $x(n)$ for different values of r about the critical value r_c.



(a) Node size over time for $r = 0.99r_c = 2.97$.

(b) Node size over time for $r = 1.01r_c = 3.03$.

Figure 3.3: Behaviour of $x(n)$ for different values of r about the critical value r_c.

In Figure 3.3a we take a value for r just less than the critical value r_c and we observe slow oscillatory convergence to the fixed point. This is in contrast to the rapid monotonic convergence seen in Figure 3.2 for a lower value of r. In Figure 3.3b we take a value for r just greater than the critical value $r_c = 3$ and we observe that there is no convergence to the fixed point. Instead $x(n)$ appears to settle into a period-two orbit.

2.1.2 Period-two Orbits

We now wish to find all the period-2 orbits of the logistic map. The period-2 points of the map given by $x(n+1) = f(x(n))$ are the fixed points of $f(f(x)) = f^2(x)$, and so are found by solving $x^* = f^2(x^*)$.

$$\begin{aligned} x^* &= f^2(x^*) = rf(x^*)(1 - f(x^*)) \implies \\ x^* &= r(rx^*(1-x^*))(1-(rx^*(1-x^*))) \implies \\ 0 &= x^*(r^2(1-x^*))(1-(rx^*(1-x^*))) - 1). \end{aligned} \tag{3.21}$$

We know x^* is a root since $x_a^* = 0$ is a fixed point of the map and hence a period-two point. Ignoring $x_a^* = 0$ we find

$$\begin{aligned} r^2(1-x^*))(1-(rx^*(1-x^*))) - 1 &= 0 \implies \\ r^3(x^*)^3 - 2r^3(x^*)^2 + (r^3+r^2)x^* + (1-r^2) &= 0 \implies \\ r^3(x^*)^3 - 2r^3(x^*)^2 + (r^3+r^2)x^* + (1-r^2) &= 0 \implies \\ (x^*)^3 - 2(x^*)^2 + \frac{r+1}{r}x^* + \frac{1-r^2}{r^3} &= 0 \implies \\ \left(x^* - \frac{r-1}{r}\right)\left((x^*)^2 - \frac{r+1}{r}x^* + \frac{r+1}{r^2}\right) &= 0. \end{aligned} \tag{3.22}$$

Again we know $x^* - \frac{r-1}{r}$ is a root since $x^* = \frac{r-1}{r}$ is a fixed point of the map and hence a period-two point. Ignoring $x^* = \frac{r-1}{r}$ we find

$$\begin{aligned} (x^*)^2 - \frac{r+1}{r}x^* + \frac{r+1}{r^2} &= 0 \implies \\ \left(x^* - \frac{r+1+\sqrt{(r+1)(r-3)}}{2r}\right)\left(x^* - \frac{r+1-\sqrt{(r+1)(r-3)}}{2r}\right) &= 0. \end{aligned} \tag{3.23}$$

Our period-two points are therefore

$$x_+^* = \frac{r+1+\sqrt{(r+1)(r-3)}}{2r}, \quad x_-^* = \frac{r+1-\sqrt{(r+1)(r-3)}}{2r}. \tag{3.24}$$

These period-two points exist only for $r > 3 = r_c$. For $r < 3$ x_+^* and x_-^* are complex. In Figure 3.3b we see an example of this period-two behaviour for $r = 1.01r_c = 3.03$.

The stability of the period-two orbit, (x_+^*, x_-^*), found above is also of interest. We can investigate the stability of this periodic orbit by again examining the

derivative of $f(x) = rx(1-x)$, $f'(x) = r - 2rx$. The periodic orbit (x^*_+, x^*_-) of the map given by $x(n+1) = f(x(n))$ is stable for

$$\left|f'(x^*_+)f'(x^*_-)\right| < 1. \tag{3.25}$$

In our case this gives that

$$\begin{aligned}
\left|(r - 2rx^*_+)(r - 2rx^*_-)\right| &= \\
\left|(r - \frac{2r}{2r}\left(r + 1 + \sqrt{(r+1)(r-3)}\right))(r - \frac{2r}{2r}\left(r + 1 - \sqrt{(r+1)(r-3)}\right))\right| &= \\
\left|\left(-1 - \sqrt{(r+1)(r-3)}\right)\left(-1 + \sqrt{(r+1)(r-3)}\right)\right| &= \\
|1 - (r+1)(r-3)| &= \\
\left|r^2 - 2r - 4\right| < 1. &
\end{aligned} \tag{3.26}$$

So the periodic orbit is stable when the following two inequalities are satisfied

$$r^2 - 2r - 4 > -1 \quad \& \quad r^2 - 2r - 4 < 1. \tag{3.27}$$

The first inequality gives

$$\begin{aligned}
r^2 - 2r - 3 &> 0 \implies \\
(r-3)(r+1) &> 0.
\end{aligned} \tag{3.28}$$

Since we have that $r > 0$ this gives us that

$$r > 3. \tag{3.29}$$

The second inequality gives

$$\begin{aligned}
r^2 - 2r - 5 &< 0 \implies \\
(r - \left(1 + \sqrt{6}\right))(r - \left(1 - \sqrt{6}\right)) &< 0 \implies \\
r &< 1 + \sqrt{6}.
\end{aligned} \tag{3.30}$$

Taking these two together we find our period-two orbit is stable for

$$3 < r < 1 + \sqrt{6} \approx 3.4494897. \tag{3.31}$$

2.1.3 Period-two Doubling to Chaos

As further in-depth analysis becomes increasingly cumbersome we will not present the methods for finding any higher period orbits analytically. As r increases beyond $1+\sqrt{6}$ a stable period-four orbit appears. As r increases further we observe a *period-two doubling* effect with stable orbits of period $8, 16, 32, \ldots$ appearing. These successive bifurcations occur at an increasing rate. In fact the ratio of distances between successive values of r at which these bifurcations

occur converges to a constant δ defined by Feigenbaum in 1978 [28]. If the i^{th} bifurcation occurs at r_i then

$$\delta = \lim_{i \to \infty} \frac{r_i - r_{i-1}}{r_{i+1} - r_i} = 4.669201609\ldots. \tag{3.32}$$

This constant has a remarkable universality property. Its value is the same for all unimodal maps with a quadratic maximum. It can finally be shown that the period-doubling bifurcations come successively closer, eventually accumulating at the point r_∞ where an infinite number of bifurcations occur

$$r = r_\infty = 3.5699456718\ldots. \tag{3.33}$$

As r increases beyond r_∞ we observe regions of chaotic behaviour, interrupted by intervals of periodic behaviour known as periodic windows. In Figure 3.4 we see an example of this chaotic behaviour for $r = 1.3r_c = 3.9$.

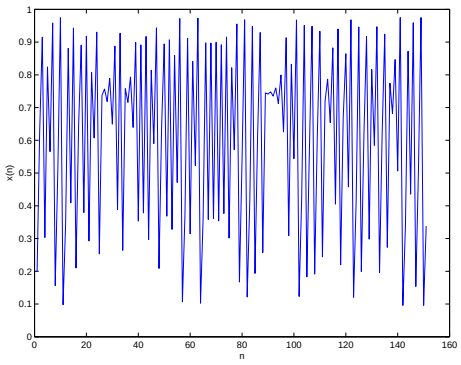


Figure 3.4: Chaotic behaviour of node size over time for $r = 1.3r_c = 3.9$.

2.1.4 Bifurcation Diagram

The bifurcation diagram in Figure 3.5 shows the long-term behaviour of the logistic map model for values of $r \in (0, 4]$. This diagram confirms our earlier results numerically. For $0 < r < 1$ we see that $x^* = 0$ is a global attractor for the system, as we increase r beyond 1 this attractor becomes unstable. For $1 < r < 3$ $x^* = \frac{r-1}{r}$ is a global attractor for the system which becomes unstable when we increase r beyond the critical value $r_c = 3$. After this value a period-doubling cascade occurs, we first see the appearance of a period-two attractor, which is stable for $3 < r < 1 + \sqrt{6}$ followed by a stable period-four attractor and so on as we move towards chaos.

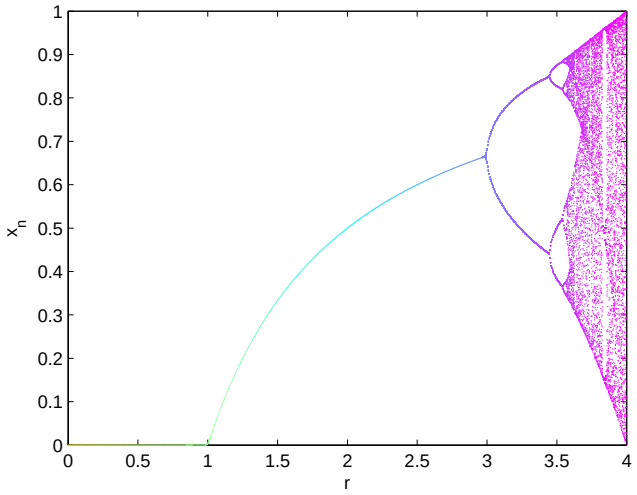


Figure 3.5: Bifurcation Diagram for the Logistic Map.

2.1.5 Lyapunov Exponents

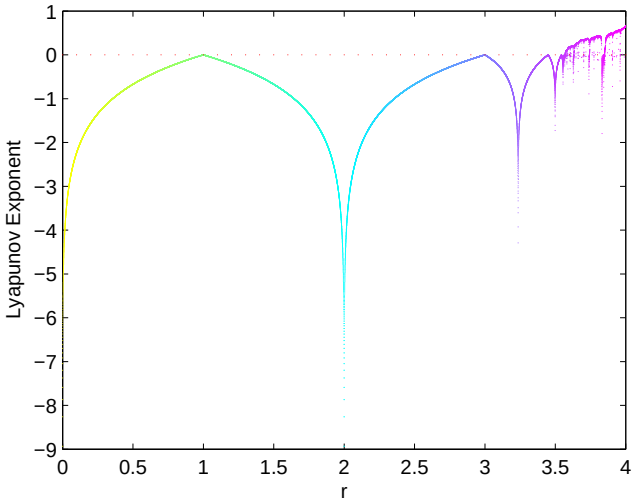
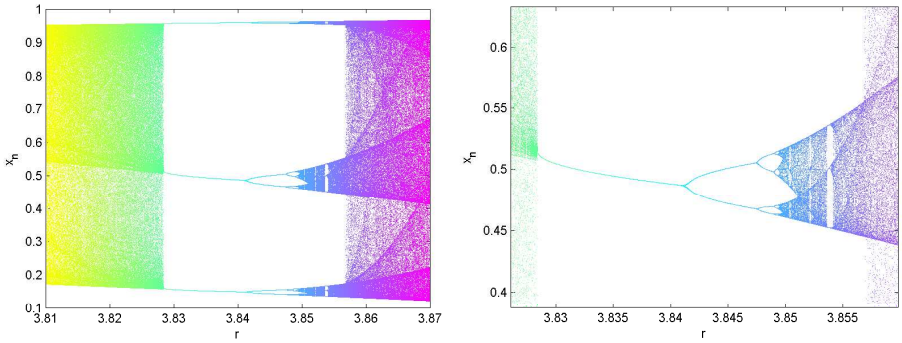


Figure 3.6: Numerically Calculated Lyapunov Exponents for the Logistic Map for $0 < r \leq 4$.

Using the method outlined in Section 1.1.3 we numerically estimated the Lyapunov exponent of the logistic map for values of $r \in (0, 4]$. We plot the

numerically estimated exponents in Figure 3.6. We see positive Lyapunov exponents for many values of $r > r_\infty$, the accumulation point defined in Equation (3.33), indicating chaotic behaviour in this region. We also see some negative Lyapunov exponents in this range corresponding to periodic windows interspersed amongst the chaos. Figure 3.7 allows us to take a closer look at the largest of these periodic windows.



(a) Bifurcation Diagram for the logistic map when $r \in [3.81, 3.87]$.

(b) Close up of the behaviour of a single branch of the original period 3 orbit.

Figure 3.7: Magnification of the bifurcation diagram in the region of the period 3 window.

At first a stable period 3 orbit emerges from the chaos. We then observe period doubling on the return to chaos. In fact when magnified the bifurcation diagram for each branch once period doubling begins mirrors the overall bifurcation diagram for r in the interval $[3, 4]$. This can be seen in Figure 3.7b.

3 Growth of Nodes in a Complete Network

Having familiarised ourselves with the dynamics of the classic logistic map in Section 2 we now present our modified logistic map for a network. We consider connected networks $N = (V, E)$ with $m \geq 2$ nodes. As we discussed in Section 1 we allow our control parameter to vary between nodes and over time. For each node $v_i \in V$ we let the control function $r_i(n)$ depend on the weighted degree of v_i at time n. We set

$$r_i(n) = 1 + \rho w_i(n) = 1 + \rho \left(\frac{\sum_{j:(v_i,v_j)\in E} x_j(n)}{\sum_{j=1}^{m} x_j(n)} \right) \tag{3.34}$$

where $\rho > 0$ and $w_i(n)$ is the normalised weighted degree of v_i at time n. Then for all $i \in \{1, 2, \ldots, m\}$ we have that

$$x_i(n+1) = \left(1 + \rho \left(\frac{\sum_{j:(v_i,v_j)\in E} x_j(n)}{\sum_{j=1}^{m} x_j(n)}\right)\right) x_i(n)(1 - x_i(n)),\ x_i(n) \in (0, 1] \quad (3.35)$$

We call this map the *network logistic map of* N. In what follows, in order to avoid confusion, we will refer to the classic logistic map discussed in Section 2 as the *single logistic map*. It is intuitive to expect that the network logistic map of any network N will bear some similarities to the single logistic map. In particular we expect that the map will also display an interesting and diverse range of dynamical behaviour. In the following sections we investigate these dynamics.

We begin our investigations by analysing the the growth of nodes in a network, N, that is fully connected or *complete*, i.e. for $i \neq j$ $(v_i, v_j) \in E$. We are interested mainly in the long-term behaviour of the system and any *bifurcations* the system undergoes as we vary the parameter ρ.

3.1 A Network With Two Nodes

To begin our analysis of the network logistic map we look at the simplest case of a complete network. This network has only two nodes, v_1 and v_2, connected by a single link (v_1, v_2). We will refer to this network as N_2^C. A network diagram of this type of network is given in Figure 3.8.

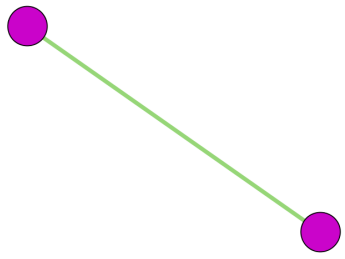


Figure 3.8: Network diagram of a complete two-node network

For the network logistic map of N_2^C we have that

$$x_1(n+1) = r_1(n)x_1(n)\,(1 - x_1(n)) \quad (3.36)$$

where our control function

$$r_1(n) = 1 + \rho\left(\frac{x_2(n)}{x_1(n) + x_2(n)}\right). \quad (3.37)$$

Similarly we have that,

$$x_2(n+1) = r_2(n)x_2(n)\left(1 - x_2(n)\right) \tag{3.38}$$

where

$$r_2(n) = 1 + \rho\left(\frac{x_1(n)}{x_1(n) + x_2(n)}\right). \tag{3.39}$$

3.1.1 Fixed Points

First let us find every fixed point of the network logistic map of N_2^C. The fixed points for $x_1(n)$ are given by x_1^* where

$$\begin{aligned} x_1^* &= r_1(n)x_1^*(1 - x_1^*) \implies \\ x_1^* &= r_1(n)x_1^* - r_1(n)(x_1^*)^2 \implies \\ 0 &= r_1(n)x_1^* - x_1^* - r_1(n)(x_1^*)^2 \implies \\ 0 &= r_1(n)x_1^*\left(\frac{r_1(n) - 1}{r_1(n)} - x_1^*\right). \end{aligned} \tag{3.40}$$

We have that $x_1(n) > 0$ and so our fixed point is given by

$$x_1^* = \frac{r_1(n) - 1}{r_1(n)} = \frac{1 + \rho\frac{x_2(n)}{x_1^* + x_2(n)} - 1}{1 + \rho\frac{x_2(n)}{x_1^* + x_2(n)}} = \frac{\rho\frac{x_2(n)}{x_1^* + x_2(n)}}{\frac{x_1^* + x_2(n) + \rho x_2(n)}{x_1^* + x_2(n)}} = \frac{\rho x_2(n)}{x_1^* + (1+\rho)x_2(n)} \tag{3.41}$$

This is stationary only when $x_2(n)$ is a constant i.e. when $x_2(n)$ is also at its fixed point x_2^* and so

$$x_1^* = \frac{\rho x_2^*}{x_1^* + (1+\rho)x_2^*}. \tag{3.42}$$

Similarly we find the fixed point for $x_2(n)$ is given by

$$x_2^* = \frac{\rho x_1^*}{x_2^* + (1+\rho)x_1^*}. \tag{3.43}$$

Developing (3.42) further gives

$$\begin{aligned} x_1^* &= \frac{\rho x_2^*}{x_1^* + (1+\rho)x_2^*} \implies \\ \rho x_2^* &= x_1^*(1+\rho)x_2^* + (x_1^*)^2 \implies \\ 0 &= (x_1^*)^2 + (1+\rho)x_1^*x_2^* - \rho x_2^*. \end{aligned} \tag{3.44}$$

Similarly, developing (3.43) gives

$$0 = (x_2^*)^2 + (1+\rho)x_1^*x_2^* - \rho x_1^*. \tag{3.45}$$

Subtracting (3.45) from (3.44) gives

$$(x_1^*)^2 - (x_2^*)^2 + \rho(x_1^* - x_2^*) = 0 \implies$$

$$(x_1^* - x_2^*)(x_1^* + x_2^* + \rho) \quad = \quad 0. \tag{3.46}$$

So we have that $x_1^* = x_2^*$ or $x_1^* = -(x_2^* + \rho)$. This second solution is *unphysical* as it implies that x_1^* is negative since $\rho > 0$ and $x_2^* > 0$. Therefore we have that $x_1^* = x_2^*$. Combining this and (3.42) gives

$$x_1^* = \frac{\rho x_2^*}{x_1^* + (1+\rho)x_2^*} = \frac{\rho x_1^*}{x_1^* + (1+\rho)x_1^*} = \frac{\rho x_1^*}{(2+\rho)x_1^*} = \frac{\rho}{2+\rho} = x_2^*. \tag{3.47}$$

In Figure 3.9 we see an example of the node sizes in N_2^C rapidly converging to this fixed point monotonically for $\rho = \frac{1}{2}$. If we look at Figure 3.9 it certainly

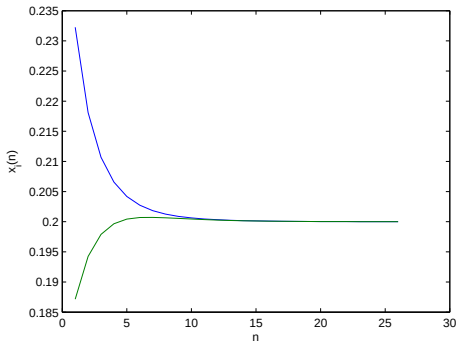


Figure 3.9: Node sizes converging to the fixed point $x^* = \frac{1}{5}$ for $m = 2$ and $\rho = \frac{1}{2}$.

appears that the fixed point $(x_1^*, x_2^*) = \left(\frac{\rho}{2+\rho}, \frac{\rho}{2+\rho}\right)$ is stable for $\rho = \frac{1}{2}$. We now wish to investigate the stability of this fixed point in detail. Combining (3.36) and (3.37) gives

$$\begin{aligned} x_1(n+1) &= \left(1 + \rho\left(\frac{x_2(n)}{x_1(n)+x_2(n)}\right)\right) x_1(n)\,(1 - x_1(n)) \\ &= \left(1 + \rho\left(\frac{x_2(n)}{x_1(n)+x_2(n)}\right)\right) (x_1(n) - (x_1(n))^2) \\ &= (x_1(n) - (x_1(n))^2) + \rho\left(\frac{x_2(n)}{x_1(n)+x_2(n)}\right) (x_1(n) - (x_1(n))^2) \\ &= f(x_1(n), x_2(n)) \end{aligned} \tag{3.48}$$

or

$$f(x_1, x_2) = (x_1 - x_1^2) + \rho\left(\frac{x_2}{x_1 + x_2}\right)(x_1 - x_1^2) \tag{3.49}$$

Similarly,

$$x_2(n+1) = g(x_1(n), x_2(n)) \tag{3.50}$$

where

$$g(x_1, x_2) = (x_2 - x_2^2) + \rho\left(\frac{x_1}{x_1 + x_2}\right)(x_2 - x_2^2) \tag{3.51}$$

We can investigate the stability of our fixed points by examining the Jacobian. The partial derivatives of the functions f and g given in (3.49) and (3.51) respectively are

$$\frac{\partial f}{\partial x_1} = (1-2x_1) + \rho\left[(x_1 - x_1^2)\frac{-x_2}{(x_1+x_2)^2} + \frac{x_2}{x_1+x_2}(1-2x_1)\right], \tag{3.52}$$

$$\frac{\partial f}{\partial x_2} = \rho\frac{x_1^2(1-x_1)}{(x_1+x_2)^2}, \tag{3.53}$$

$$\frac{\partial g}{\partial x_1} = \rho\frac{x_2^2(1-x_2)}{(x_1+x_2)^2}, \tag{3.54}$$

and

$$\frac{\partial g}{\partial x_2} = (1-2x_2) + \rho\left[(x_2 - x_2^2)\frac{-x_1}{(x_1+x_2)^2} + \frac{x_1}{x_1+x_2}(1-2x_2)\right]. \tag{3.55}$$

Evaluating at our fixed point $(x_1^*, x_2^*) = \left(\frac{\rho}{2+\rho}, \frac{\rho}{2+\rho}\right)$ we find

$$\frac{\partial f}{\partial x_1}|_{(\frac{\rho}{2+\rho},\frac{\rho}{2+\rho})} = \frac{\partial g}{\partial x_2}|_{(\frac{\rho}{2+\rho},\frac{\rho}{2+\rho})} = \frac{4-\rho^2-\rho}{2(2+\rho)} \tag{3.56}$$

and

$$\frac{\partial f}{\partial x_2}|_{(\frac{\rho}{2+\rho},\frac{\rho}{2+\rho})} = \frac{\partial g}{\partial x_1}|_{(\frac{\rho}{2+\rho},\frac{\rho}{2+\rho})} = \frac{\rho}{2(2+\rho)}. \tag{3.57}$$

So the Jacobian is

$$\mathbf{J}\left(\frac{\rho}{2+\rho}, \frac{\rho}{2+\rho}\right) = \begin{pmatrix} \frac{4-\rho^2-\rho}{2(2+\rho)} & \frac{\rho}{2(2+\rho)} \\ \frac{\rho}{2(2+\rho)} & \frac{4-\rho^2-\rho}{2(2+\rho)} \end{pmatrix} \tag{3.58}$$

and the characteristic equation is

$$\lambda_2 - \frac{4-\rho^2-\rho}{2+\rho}\lambda + \frac{(\rho-2)(\rho^2+2\rho-4)}{4(2+\rho)} = 0. \tag{3.59}$$

The condition for stability is therefore

$$2 > 1 + \frac{(\rho-2)(\rho^2+2\rho-4)}{4(2+\rho)} > \left|\frac{4-\rho^2-\rho}{2+\rho}\right| \tag{3.60}$$

Using the fact that $\rho > 0$ this reduces to

$$16 + 8\rho > \rho^3 - 4\rho + 16 > 4\left|4-\rho-\rho^2\right| \tag{3.61}$$

which reduces further to

$$0 > \rho^3 - 12\rho > 4\left(\left|4-\rho-\rho^2\right| - 4 - 2\rho\right). \tag{3.62}$$

Taking each inequality separately this gives that

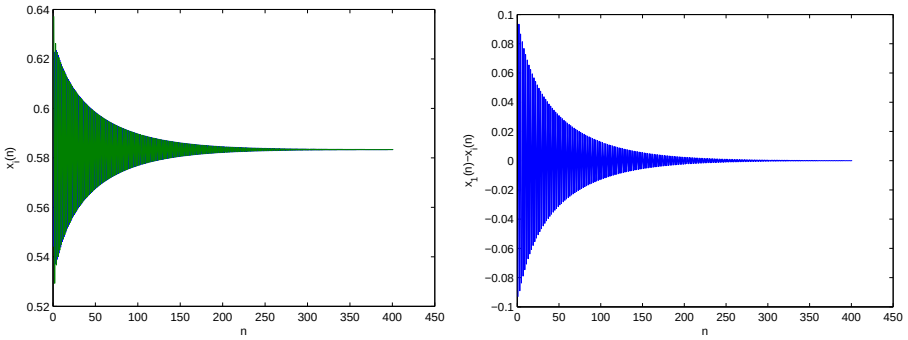
$$\rho^3 - 12\rho < 0 \implies \rho^2 < 12 \implies 0 < \rho < 2\sqrt{3} \tag{3.63}$$

and

$$\begin{aligned} \rho^3 - 12\rho &> 4\left(\left|4 - \rho - \rho^2\right| - 4 - 2\rho\right) \implies \\ \rho^3 - 4\rho + 16 &> 4\left|4 - \rho - \rho^2\right| \implies \\ \rho^6 - 8\rho^4 + 32\rho^3 + 16\rho^2 - 128\rho + 196 &> 16(16 - 8\rho - 7\rho^2 + 2\rho^3 + \rho^4) \implies \\ \rho^6 - 24\rho^4 + 128\rho^2 &> 0 \implies \\ \rho^4 - 24\rho^2 + 128 &> 0 \implies \\ (\rho^2 - 16)(\rho^2 - 8) &> 0 \end{aligned} \tag{3.64}$$

Taking (3.63) and (3.64) together gives that the fixed point $(x^*, x^*) = \left(\frac{\rho}{2+\rho}, \frac{\rho}{2+\rho}\right)$ is stable when

$$0 < \rho < 2\sqrt{2} = \rho_{c_2}. \tag{3.65}$$



(a) Node size over time in a complete network with $m = 2$ and $\rho = 0.99\rho_{c_2}$.

(b) Difference in node size over time in a complete network with $m = 2$ and $\rho = 0.99\rho_{c_2}$.

Figure 3.10: Node behaviour in a complete network with $m = 2$ and $\rho = 0.99\rho_{c_2} \approx 2.80$.

Figures 3.10 - 3.12 show the behaviour of $x_i(n)$ for different values of ρ about the critical value $\rho_{c_2} = 2\sqrt{2}$. In Figure 3.10a we take a value for ρ just less than the critical value ρ_{c_2} above which the fixed point (x^*, x^*) becomes unstable. We plot the dynamics of a network with initial node sizes randomly chosen from a uniform distribution on $(0, 1)$, $x_1(0) \in (0, 1)$ and $x_2(0) \in (0, 1)$, and observe slow oscillatory convergence to (x^*, x^*) as $n \to \infty$. The dynamics of $x_1(n)$ are plotted in blue while the dynamics of $x_2(n)$ are plotted in green. Both oscillate about the fixed point x^* as they converge. This behaviour is in contrast to the rapid monotonic convergence we saw earlier in Figure 3.9 for a lower value of ρ, $\rho = \frac{1}{2}$. It is interesting to note that while oscillating about the fixed point

before converging the two node sizes are out of phase. This can be seen by examining Figure 3.10b. Here we show the difference between the two node sizes over time, $x_1(n) - x_2(n)$. If the two node sizes moved perfectly together we would have $x_1(n) - x_2(n) = 0$, instead we observe a difference in sizes that oscillates between positive and negative values before converging to 0 as the node sizes converge to (x^*, x^*). This indicates that at every timestep the node with the largest size alternates between v_1 and v_2, with the difference between the two sizes becoming smaller at each step until they reach $\left(\frac{\rho}{2+\rho}, \frac{\rho}{2+\rho}\right)$.

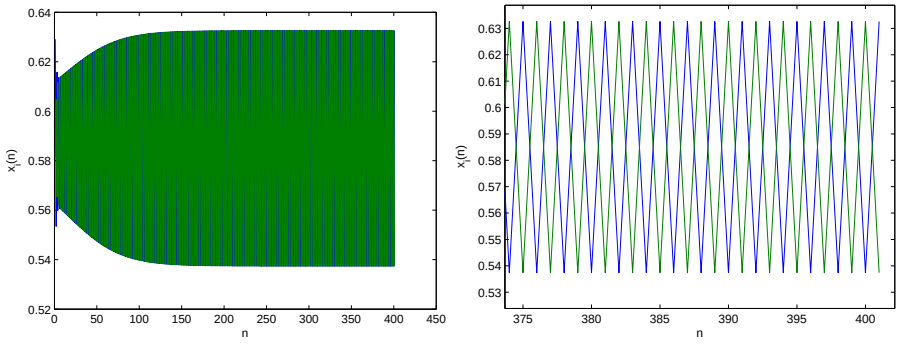


Figure 3.11: Node size over time in a complete network with $m = 2$ and $\rho = 1.01\rho_{c_2} \approx 2.86$.

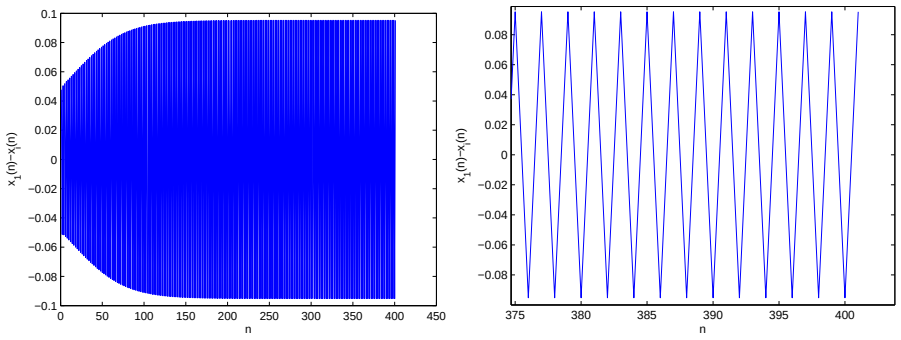


Figure 3.12: Difference in node size over time in a complete network with $m = 2$ and $\rho = 1.01\rho_{c_2} \approx 2.86$.

In Figure 3.11 we take a value for ρ just greater than the critical value ρ_{c_2} below which the fixed point (x^*, x^*) is stable. We again plot the dynamics of a network with initial node sizes randomly chosen from a uniform distribution on $(0, 1)$, $x_1(0) \in (0, 1)$ and $x_2(0) \in (0, 1)$. The dynamics of $x_1(n)$ are plotted in blue while the dynamics of $x_2(n)$ are plotted in green. The panel on the left shows the dynamics of the system from time $n = 0$ to 400 while the panel on the right gives us a more detailed view of the dynamics for $n = 375$ to 400 after

they settle into a regular pattern. We observe that in this case, unlike both of the cases we have examined so far, there is no convergence to the fixed point $\left(\frac{\rho}{\rho+2}, \frac{\rho}{\rho+2}\right)$ as $n \to \infty$. Instead both node sizes $x_1(n)$ and $x_2(n)$ appear to settle into a period-two orbit, oscillating between the same two values, $x_-^* \approx 0.539$ and $x_+^* \approx 0.631$. In Figure 3.12 we plot the difference between the two node sizes, $x_1(n) - x_2(n)$ over time. The panel on the left again shows the dynamics from time $n = 0$ to 400 while the panel on the right gives us a more detailed view for $n = 375$ to 400 after they settle. We see that the difference between the two node sizes settles into a period-two orbit about 0. This, together with the right hand panel of Figure 3.11 allow us to see that that the orbits of the two node sizes are out of phase. As $n \to \infty$ when $x_1(n) = x_-^*$, $x_2(n) = x_+^*$ and vice versa.

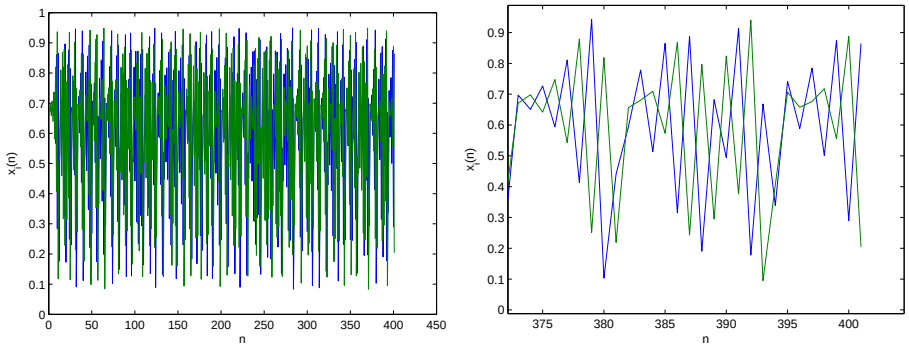


Figure 3.13: Node size over time in a complete network with $m = 2$ and $\rho = 1.5\rho_{c_2} \approx 4.24$.

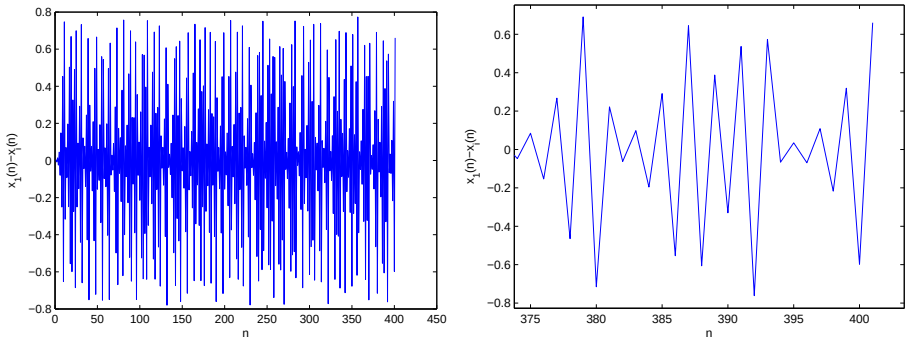


Figure 3.14: Difference in node size over time in a complete network with $m = 2$ and $\rho = 1.5\rho_{c_2} \approx 4.24$.

In Figures 3.13 and 3.14 we see evidence of chaotic behaviour in the system for values of the parameter ρ significantly larger than ρ_{c_2}. We again plot the dynamics of a network with randomly chosen initial node sizes $x_1(0) \in (0, 1)$

and $x_2(0) \in (0, 1)$. The dynamics of $x_1(n)$ are plotted in blue while the dynamics of $x_2(n)$ are plotted in green. The panel on the left shows the dynamics of the system from time $n = 0$ to 400 while the panel on the right gives us a more detailed view of the dynamics for $n = 375$ to 400. After a large number of timesteps the dynamics of the system do not appear to settle into any recognisable pattern. It is interesting to note that not only do the sizes between the two nodes appear to experience chaos but there is evidence in Figure 3.14, where we plot the difference between the two node sizes, $x_1(n) - x_2(n)$ over time, that their relationship is also chaotic. Just like the node sizes themselves, this difference does not appear to settle into any recognisable pattern.

3.1.2 Towards Chaos

Finding further attractors analytically becomes increasingly cumbersome. However, numerically we can see that although the initial behaviour of the network logistic map is very similar to the behaviour of the single logistic map, the routes the two systems take to chaos are quite different.

When analysing the single logistic map in Section 2 we observed that as we increased the control parameter, r, beyond $r_c = 3$ a period doubling cascade to chaos occurred. However, this is is not the case when dealing with the network logistic map of N_2^C. As we increase ρ beyond $\rho_c = 2\sqrt{2}$ we do not always observe the same qualitative behaviour. At first we do see the appearance of a stable period-two attractor as is the case in the single logistic map. However, in the single logistic map as we increase r beyond $1 + \sqrt{6}$ a period four attractor appears, in the case of the network logistic map of N_2^C in place of a period-four attractor after our next bifurcation point $\rho \approx 3.65$ a quasi-periodic attractor with period-two appears. In Figure 3.15 we see a nice example of this striking behaviour for $\rho = 3.7$. The subsequent bifurcation sees the appearance of a period ten attractor before cascading to seemingly chaotic behaviour. Later by analysing the Lyapunov exponents of the system we will be able to quantify chaos and show that the behaviour of the system in this region is indeed chaotic.

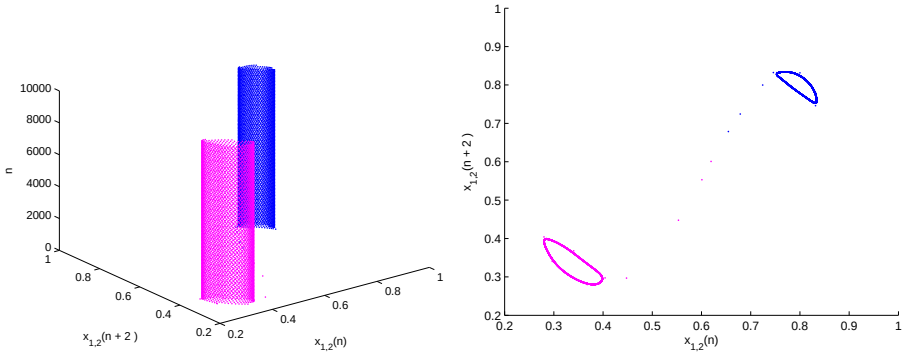
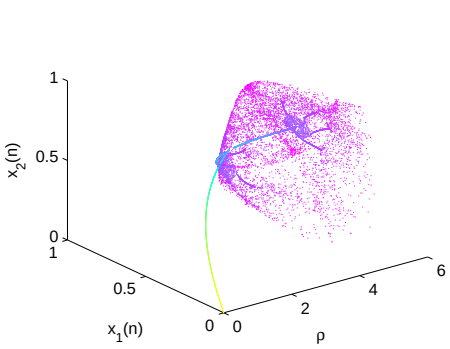


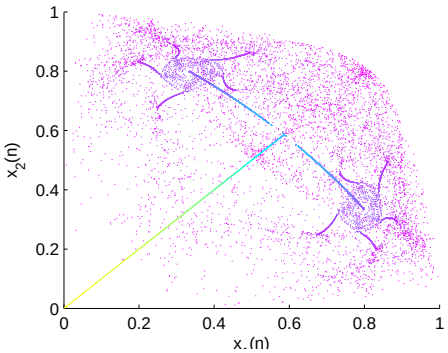
Figure 3.15: Quasi-periodic behaviour of $x_1(n)$ (magenta) and $x_2(n)$ (blue) in a complete two node network for $\rho = 3.7$.

3.1.3 Bifurcation Diagrams

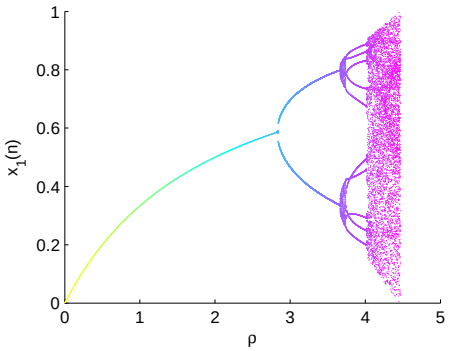
The bifurcation diagrams in Figure 3.16 allow us to analyse and observe the long-term behaviour of the network logistic map model for $m = 2$ and values of $\rho \in (0, 4.47]$. These bifurcation diagrams confirm our earlier analytical and numerical results. For $\rho < 2\sqrt{2}$ we see a single stable fixed point given by $\left(\frac{\rho}{2+\rho}, \frac{\rho}{2+\rho}\right)$. For $\rho > 2\sqrt{2}$ this fixed point becomes unstable and a period-two attractor appears. The two node sizes oscillate out of phase between the same two values for each value of ρ for which the period-two attractor is stable. This period-two attractor remains stable until the next bifurcation, at this point a quasi-periodic attractor of period two appears. The subsequent bifurcation sees the appearance of a period-ten attractor before cascading to chaos.



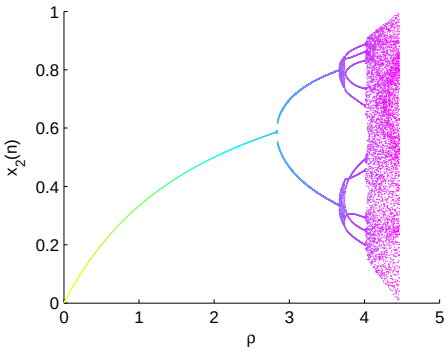
(a) Bifurcation diagram for a complete network with $m = 2$.



(b) $x_1(n)$ vs $x_2(n)$.



(c) Bifurcation diagram projection for $x_1(n)$.



(d) Bifurcation diagram projection for $x_2(n)$.

Figure 3.16: Bifurcation diagrams for a complete network with $m = 2$. Colour corresponds to ρ value.

3.1.4 Lyapunov Exponents

We have numerically estimated the maximum Lyapunov exponent of the network logistic map of N_2^C for values of $\rho \in (0, 4.47]$. Our estimates were obtained by implementing a version of an algorithm due to Eckmann and Ruelle in Matlab [25]. We plot the numerically estimated exponents in Figure 3.17. We see that the maximum Lyapunov exponent of the network logistic map of N_2^C is negative for values of $\rho \lessapprox 3.65$. This indicates the non-chaotic behaviour of the map for these values of ρ. We note that the maximum Lyapunov exponent is converging to zero for the range of values of ρ for which stable quasi-periodic behaviour was observed in Section 3.1.3. After this the maximum Lyapunov exponent becomes negative again until ρ exceeds $\rho_{\infty_2} \approx 4.02$.

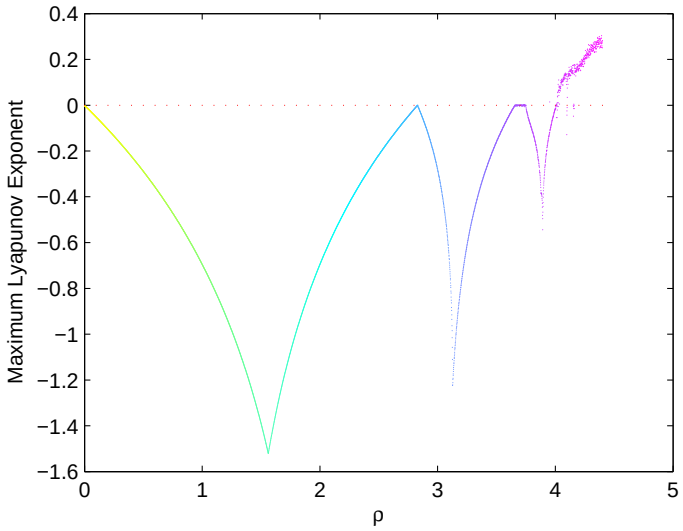
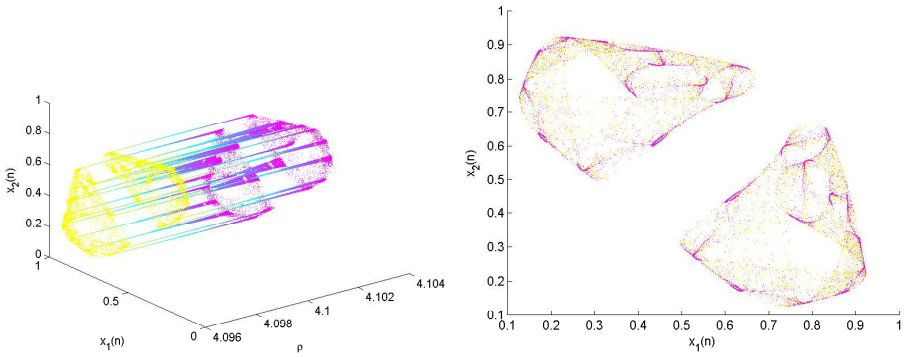


Figure 3.17: Numerically Calculated Maximum Lyapunov Exponents for the Network Logistic Map of a Complete Network with $m = 2$.

Positive Lyapunov exponents are then observed for values of $\rho > \rho_{\infty_2} \approx 4.02$ indicating chaotic behaviour in this region as expected from our earlier analysis and bifurcation diagrams. However for some values of ρ in this range we also observe negative Lyapunov exponents. These negative values correspond to periodic windows interspersed amongst the chaos.

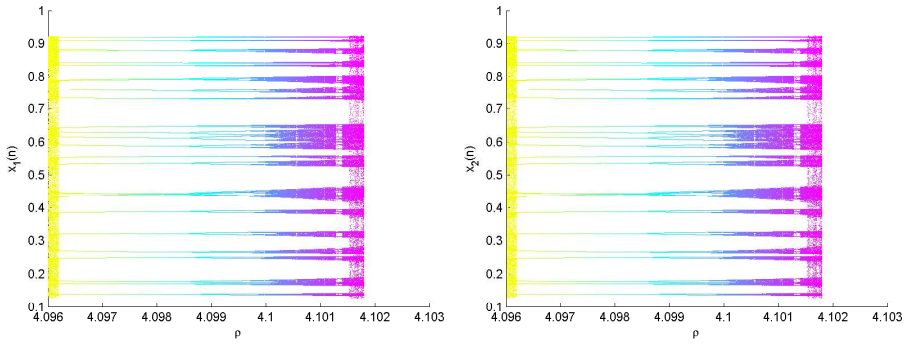


(a) Bifurcation diagram for $x_1(n)$ and $x_2(n)$.

(b) $x_1(n)$ vs $x_2(n)$. Colour corresponds to ρ value.

Figure 3.18: Bifurcation diagrams for $x_1(n)$ and $x_2(n)$ for $\rho \in [4.096, 4.103]$.

Figures 3.18, 3.19 and 3.20 allow us to take a closer look at one of these periodic windows. At first a stable period 26 orbit emeges from the chaos. We then observe period doubling-like behaviour on the return to chaos. In Figure 3.20 we magnify one branch of the period 26 orbit in order to see this period doubling behaviour clearly. We note the similarities between the behaviour of this branch and the single logistic map.



(a) Bifurcation diagram projections for $x_1(n)$ for $\rho \in [4.096, 4.103]$.

(b) Bifurcation diagram projections for $x_2(n)$ for $\rho \in [4.096, 4.103]$.

Figure 3.19: Bifurcation diagram projections for $x_1(n)$ and $x_2(n)$ for $\rho \in [4.096, 4.103]$.

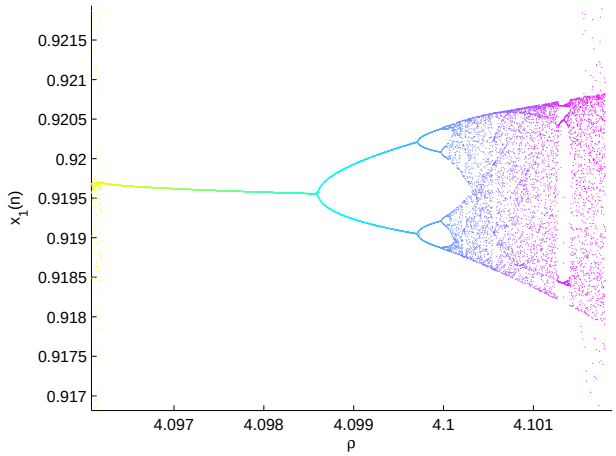


Figure 3.20: Close up of the behaviour of a single branch of the original period 26 orbit shown in Figure 3.19a.

3.2 A Network With m Nodes

We would now like to generalise the results of Section 3.1 for a fully connected or complete network $N = (V, E)$ with $m \geq 2$ nodes. Here N has node set $V = \{v_i\}_{i=1}^{m}$ where each node v_i has a corresponding size $x_i(n)$ at time n and link set $E = \{(v_i, v_j) : i \neq j,\ i, j \in \{1, 2, \ldots, m\}\}$. We will refer to a complete m node network as N_m^C.

The network logistic map of N_m^C is given by

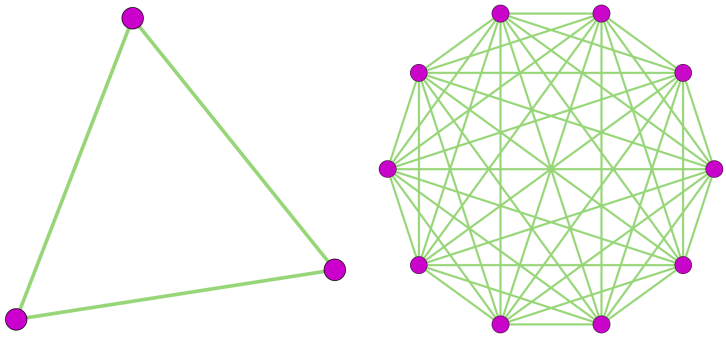
$$x_i(n+1) = r_i(n)x_i(n)\left(1 - x_i(n)\right), \tag{3.66}$$

where, following the definition given in (3.121), we have that

$$r_i(n) = 1 + \rho\left(\frac{\sum_{j\neq i} x_j(n)}{\sum_{j=1}^{m} x_j(n)}\right) \tag{3.67}$$

for all $i \in \{1, 2, \ldots, m\}$.

Throughout our analysis we will refer to two examples to illustrate the behaviour of a complete m-node network, namely $m = 3$ and $m = 10$. Network diagrams for these two sample networks are shown in Figure 3.21.



(a) A complete network with 3 nodes.

(b) A complete network with 10 nodes.

Figure 3.21: Network diagrams of our two sample networks, N_3^C and N_{10}^C .

3.2.1 Fixed Points

Again we begin our analysis of the system by finding all of the fixed points of the system. For all $i \in \{1, 2, \ldots, m\}$ the fixed points for $x_i(n)$ are given by x_i^* where

$$\begin{aligned} x_i^* &= r_i(n)x_i^*(1-x_i^*) \implies \\ x_i^* &= r_i(n)x_i^* - r_i(n)(x_i^*)^2 \implies \\ 0 &= r_i(n)x_i^* - x_i^* - r_i(n)(x_i^*)^2 \implies \\ 0 &= r_i(n)x_i^*\left(\frac{r_i(n)-1}{r_i(n)} - x_i^*\right). \end{aligned} \tag{3.68}$$

Since $x_i(n) > 0$ this means that our fixed points, x_i^*, for $x_i(n)$, are given by

$$\begin{aligned} x_i^* = \frac{r_i(n)-1}{r_i(n)} &= \frac{1+\rho\left(\dfrac{\sum_{j\neq i} x_j(n)}{m}\Big/\sum_{j=1} x_j(n)\right)-1}{1+\rho\left(\dfrac{\sum_{j\neq i} x_j(n)}{m}\Big/\sum_{j=1} x_j(n)\right)} = \frac{\dfrac{\rho\sum_{j\neq i} x_j^*}{m}\Big/\sum_{j=1} x_j^*}{\dfrac{\sum_{j=1}^{m} x_j^* + \rho\sum_{j\neq i} x_j^*}{m}\Big/\sum_{j=1} x_j^*} \\ &= \frac{\rho\sum_{j\neq i} x_j^*}{\sum_{j=1}^{m} x_j^* + \rho\sum_{j\neq i} x_j^*} = \frac{\rho\sum_{j\neq i} x_j^*}{x_i^* + (1+\rho)\sum_{j\neq i} x_j^*}. \end{aligned} \tag{3.69}$$

Rearranging (3.69) gives

$$(x_i^*)^2 + x_i^*(1+\rho)\sum_{j\neq i} x_j^* - \rho\sum_{j\neq i} x_j^* = 0. \tag{3.70}$$

For every pair of nodes in the complete network, v_a, v_b, with $a, b \in \{1, 2, \ldots, m\}$ we have

$$(x_a^*)^2 + x_a^*(1+\rho)\sum_{j\neq a} x_j^* - \rho\sum_{j\neq a} x_j^* = 0, \tag{3.71}$$

$$(x_b^*)^2 + x_b^*(1+\rho)\sum_{j\neq b} x_j^* - \rho\sum_{j\neq b} x_j(n) = 0. \tag{3.72}$$

(3.71) - (3.72) gives

$$\begin{aligned}
(x_a^*)^2 - (x_b^*)^2 + (x_a^* - x_b^*)(1+\rho)\sum_{\substack{j\neq a\\ j\neq b}} x_j^* + \rho\left(\sum_{j\neq b} x_j^* - \sum_{j\neq a} x_j^*\right) &= 0 \implies \\
(x_a^*)^2 - (x_b^*)^2 + (x_a^* - x_b^*)(1+\rho)\sum_{\substack{j\neq a\\ j\neq b}} x_j^* + \rho(x_a^* - x_b^*) &= 0 \implies \\
(x_a^* - x_b^*)\left(x_a^* + x_b^* + \rho + (1+\rho)\sum_{\substack{j\neq a\\ j\neq b}} x_j^*\right) &= 0. \qquad (3.73)
\end{aligned}$$

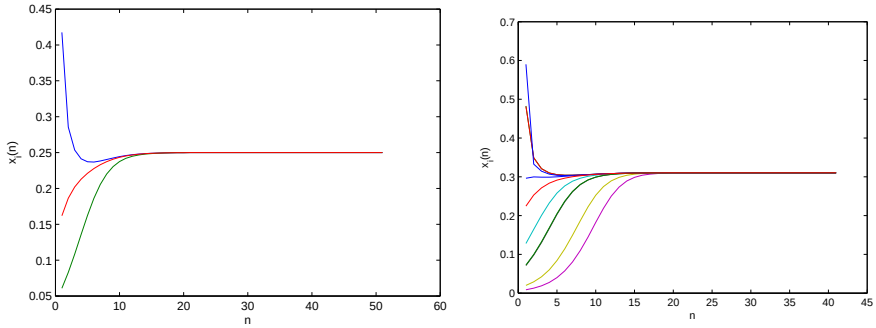
So we have that $x_a^* = x_b^*$ or $x_a^* = -\left(x_b^* + \rho + (1+\rho)\sum_{\substack{j\neq a\\ j\neq b}} x_j^*\right)$. This second solution is *unphysical* as it implies that x_a^* is negative since $\rho > 0$ and $x_i(n) > 0$ for all i, n. and so we have $x_a^* = x_b^*$. This implies that $x_i^* = x^* > 0$ for all $i \in \{1, 2, \ldots, m\}$ and so

$$x^* = \frac{\rho\sum_{i=1}^{m-1} x^*}{x^* + (1+\rho)\sum_{i=1}^{m-1} x^*} = \frac{(m-1)\rho x^*}{mx^* + (m-1)\rho x^*} = \frac{(m-1)\rho}{m + (m-1)\rho}. \tag{3.74}$$

In Figure 3.22 we see examples of the node sizes in N_3^C and N_{10}^C converging to this fixed point monotonically for $\rho = \frac{1}{2}$.

It is interesting to note that

$$\lim_{m\to\infty} x^* = \lim_{m\to\infty} \frac{(m-1)\rho}{m + (m-1)\rho} = \frac{\rho}{1+\rho} \tag{3.75}$$



(a) Converging to the fixed point $x^* = \frac{1}{4}$ for $m = 3$ and $\rho = \frac{1}{2}$.

(b) Converging to the fixed point $x^* = \frac{9}{29}$ for $m = 10$ and $\rho = \frac{1}{2}$.

Figure 3.22: Monotonic convergence of node sizes to the fixed point in an m node complete network.

and

$$\frac{\rho}{1+\rho} = \rho \sum_{k=0}^{\infty} (-\rho)^k, \quad |\rho| < 1. \tag{3.76}$$

In fact, this should be expected since for a complete network with m nodes we have that

$$\lim_{m\to\infty} r_i(n) = \lim_{m\to\infty} 1 + \rho w_i(n) = 1 + \rho. \tag{3.77}$$

Now referring to Section 2 we see that the fixed point in the simple logistic map with constant control parameter r was given by $\dfrac{r-1}{r}$ and $\dfrac{(1+\rho)-1}{(1+\rho)} = \dfrac{\rho}{1+\rho}$.

If we look at Figure 3.22 it certainly appears that the fixed point $(x^*, x^*, \ldots, x^*) = \left(\frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho}\right)$ is stable for $\rho = \frac{1}{2}$ for the logistic network map of both N_3^C and N_{10}^C. We know wish to investigate the stability of this fixed point in detail. Combining (3.66) and (3.67) gives

$$x_i(n+1) = \left(1 + \rho \left(\frac{\sum\limits_{j\neq i} x_j(n)}{\sum\limits_{j=1}^{m} x_j(n)}\right)\right) x_i(n)\,(1 - x_i(n)) = f_i(x_1(n), x_2(n), \ldots, x_m(n)) \tag{3.78}$$

For all $i \in \{1, 2, \ldots, m\}$

$$f_i(x_1, x_2, \ldots, x_m) = \left(1 + \rho \left(\frac{\sum\limits_{j\neq i} x_j}{\sum\limits_{j=1}^{m} x_j}\right)\right) x_i\,(1 - x_i) \tag{3.79}$$

We can investigate the stability of the fixed points by examining the Jacobian. We have

$$\frac{\partial f_i}{\partial x_i} = (1-2x_i)\left(1+\rho\left(\frac{\sum_{j\neq i} x_j}{\sum_{j=1}^{m} x_j}\right)\right) - x_i(1-x_i)\rho\left(\frac{\sum_{j\neq i} x_j}{\left(\sum_{j=1}^{m} x_j\right)^2}\right) \tag{3.80}$$

and for $j \neq i$

$$\frac{\partial f_i}{\partial x_j} = \rho x_i(1-x_i)\left(\frac{\sum_{j=1}^{m} x_j - \sum_{j\neq i} x_j}{\left(\sum_{j=1}^{m} x_j\right)^2}\right) = \frac{\rho(x_i^2 - x_i^3)}{\left(\sum_{j=1}^{m} x_j\right)^2} \tag{3.81}$$

Evaluating at the fixed point $(x_1^*, x_2^*, \ldots x_m^*)$ we find

$$\frac{\partial f_i}{\partial x_i}|_{(\frac{(m-1)\rho}{m+(m-1)\rho}, \frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho})} = \frac{m^2-(m-1)^2\rho^2-(m-1)\rho}{m(m+(m-1)\rho)} \tag{3.82}$$

and for $i \neq j$

$$\frac{\partial f_i}{\partial x_j}|_{(\frac{(m-1)\rho}{m+(m-1)\rho}, \frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho})} = \frac{\rho}{m(m+(m-1)\rho)}. \tag{3.83}$$

So the Jacobian is

$$\mathbf{J}\left(\frac{(m-1)\rho}{m+(m-1)\rho}, \frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho}\right) =$$

$$\begin{pmatrix} \frac{m^2-(m-1)^2\rho^2-(m-1)\rho}{m(m+(m-1)\rho)} & \frac{\rho}{m(m+(m-1)\rho)} & \cdots & \frac{\rho}{m(m+(m-1)\rho)} \\ \frac{\rho}{m(m+(m-1)\rho)} & \frac{m^2-(m-1)^2\rho^2-(m-1)\rho}{m(m+(m-1)\rho)} & \cdots & \frac{\rho}{m(m+(m-1)\rho)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\rho}{m(m+(m-1)\rho)} & \frac{\rho}{m(m+(m-1)\rho)} & \cdots & \frac{m^2-(m-1)^2\rho^2-(m-1)\rho}{m(m+(m-1)\rho)} \end{pmatrix} \tag{3.84}$$

$\mathbf{J}$ has $\mathbf{v}_1 = (1, 1, \ldots, 1)^T$ as an eigenvector with corresponding eigenvalue

$$\lambda_1 = \frac{m^2-(m-1)^2\rho^2-(m-1)\rho}{m(m+(m-1)\rho)} + (m-1)\frac{\rho}{m(m+(m-1)\rho)}.$$

$\mathbf{J}$'s other eigenvectors are given by $(-1, 1, 0, \ldots, 0)^T, (-1, 0, 1, \ldots, 0)^T, \ldots,$ $(-1, 0, 0, \ldots, 1)^T$ and all have the corresponding eigenvalue

$$\lambda_2 = \frac{m^2-(m-1)^2\rho^2-(m-1)\rho}{m(m+(m-1)\rho)} - \frac{\rho}{m(m+(m-1)\rho)}.$$

For stability we require $|\lambda_1| < 1$ and $|\lambda_2| < 1$. First

$$\begin{aligned} |\lambda_1| &< 1 \implies \\ \left|\frac{m^2-(m-1)^2\rho^2-(m-1)\rho+(m-1)\rho}{m(m+(m-1)\rho)}\right| &< 1 \implies \\ \left|\frac{m^2-(m-1)^2\rho^2}{m(m+(m-1)\rho)}\right| &< 1 \end{aligned} \tag{3.85}$$

Using the fact that $\rho > 0$ and $m \geq 2$ this becomes

$$-(m(m+(m-1)\rho) < m^2-(m-1)^2\rho^2 < m(m+(m-1)\rho. \tag{3.86}$$

Since we know $\rho > 0$ and $m \geq 2$ the second inequality is always satisfied. Examining the other inequality we find

$$\begin{aligned} m^2-(m-1)^2\rho^2 &> -(m(m+(m-1)\rho) \implies \\ m^2-(m-1)^2\rho^2 &> -m^2-m(m-1)\rho \implies \\ 2m^2-(m-1)^2\rho^2+m(m-1)\rho &> 0 \implies \\ (m-1)\rho^2-m\rho-\frac{2m^2}{m-1} &< 0 \implies \\ \left(\rho-\frac{m+\sqrt{m^2+8m^2}}{2(m-1)}\right)\left(\rho-\frac{m-\sqrt{m^2+8m^2}}{2(m-1)}\right) &< 0 \implies \\ \left(\rho-\frac{2m}{m-1}\right)\left(\rho+\frac{m}{m-1}\right) &< 0 \implies \\ \left(\rho-\frac{2m}{m-1}\right) &< 0 \implies \\ \rho &< \frac{2m}{m-1}. \end{aligned} \tag{3.87}$$

Since $\forall m \geq 2 : \frac{m}{m-1} > 0$. Now

$$\begin{aligned} |\lambda_2| &< 1 \implies \\ \left|\frac{m^2-(m-1)^2\rho^2-(m-1)\rho-\rho}{m(m+(m-1)\rho)}\right| &< 1 \implies \\ \left|\frac{m^2-(m-1)^2\rho^2-m\rho}{m(m+(m-1)\rho)}\right| &< 1 \end{aligned} \tag{3.88}$$

Using the fact that $\rho > 0$ and $m \geq 2$ this becomes

$$-(m(m+(m-1)\rho) < m^2-(m-1)^2\rho^2-m\rho < m(m+(m-1)\rho. \tag{3.89}$$

Since we know $\rho > 0$ and $m \geq 2$ the second inequality in 3.89, $m^2-(m-1)^2\rho^2-m\rho < m(m+(m-1)\rho$ is always satisfied. Examining the other inequality in 3.89, $-(m(m+(m-1)\rho) < m^2-(m-1)^2\rho^2-m\rho$, we find

$$\begin{aligned} m^2-(m-1)^2\rho^2-m\rho &> -(m(m+(m-1)\rho) \implies \\ m^2-(m-1)^2\rho^2-m\rho &> -m^2-m(m-1)\rho \end{aligned} \tag{3.90}$$

and so

$$\begin{aligned}(m-1)^2\rho^2 - m(m-1)\rho + m\rho - 2m^2 &< 0 \implies \\ (m-1)^2\rho^2 + (2m-m^2)\rho - 2m^2 &< 0 \implies\end{aligned}$$

$$\begin{aligned}\left(\rho - \frac{m^2-2m+\sqrt{(2m-m^2)^2+8m^2(m-1)^2}}{2(m-1)^2}\right)\left(\rho - \frac{m^2-2m-\sqrt{(2m-m^2)^2+8m^2(m-1)^2}}{2(m-1)^2}\right) &< 0 \implies \\ \left(\rho - \frac{m^2-2m+\sqrt{12m^2-20m^3+9m^4}}{2(m-1)^2}\right)\left(\rho - \frac{m^2-2m-\sqrt{12m^2-20m^3+9m^4}}{2(m-1)^2}\right) &< 0 \implies \\ \rho - \frac{m^2-2m+\sqrt{12m^2-20m^3+9m^4}}{2(m-1)^2} &< 0 \implies \\ \frac{m^2-2m+\sqrt{12m^2-20m^3+9m^4}}{2(m-1)^2} &> \rho. \qquad (3.91)\end{aligned}$$

We can say this since $\forall m \geq 2 : \frac{m^2-2m-\sqrt{12m^2-20m^3+9m^4}}{2(m-1)^2} < 0$. Taking (3.87) and (3.91) together with the fact that for all $m \geq 2$

$$\frac{m^2-2m+\sqrt{12m^2-20m^3+9m^4}}{2(m-1)^2} < \frac{2m}{m-1}, \qquad (3.92)$$

gives that this fixed point is stable when

$$0 < \rho < \frac{m^2-2m+\sqrt{12m^2-20m^3+9m^4}}{2(m-1)^2} := \rho_{c_m}. \qquad (3.93)$$

Note that

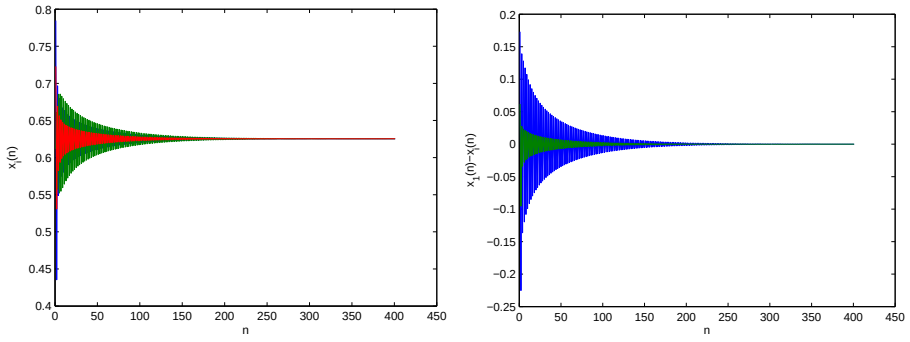
$$\lim_{m\to\infty} \rho_{c_m} = \lim_{m\to\infty} \frac{m^2 - 2m + \sqrt{12m^2 - 20m^3 + 9m^4}}{2(m-1)^2} = 2. \tag{3.94}$$

Again this should be expected as for a complete network with m nodes we have that

$$\lim_{m\to\infty} r_i(n) = \lim_{m\to\infty} 1 + \rho w_i(n) = 1 + \rho \tag{3.95}$$

where ρ is a constant. Referring to Section 2 we see that the critical value of the constant control parameter r in the single logistic map is given by $r_c = 3$ and taking $1 + \rho_c = r_c = 3$ gives that $\rho_c = 2$.

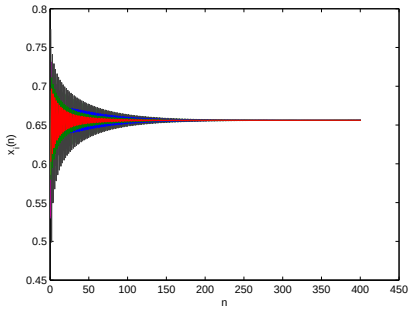
We now examine the behaviour of $x_i(n)$ for different values of ρ about the critical value ρ_{c_m} above which the fixed point $(x^*, \ldots, x^*)$ becomes unstable according to our analysis. In Figures 3.23 and 3.24 we take a value for ρ just less than the critical value ρ_{c_m}. We plot the dynamics of node sizes in a 3-node complete network in 3.23 and a 10-node complete network in 3.24, choosing the initial node sizes $x_i(0)$ randomly from the interval $(0, 1)$. In Figure 3.23a and 3.24a we plot the dynamics of each individual $x_i(n)$ as $n \to \infty$ in a different colour. Just like in the two-node case, in both cases we observe slow oscillatory convergence to $(x^*, \ldots, x^*)$. We also see that again while oscillating the node sizes in each system are not all in phase or equal. In Figures 3.23b and 3.24b we have plotted the dynamics of the difference, $x_1(n) - x_i(n)$, between each node size and the size of v_1, $x_1(n)$, as $n \to \infty$, in different colours. If the node sizes moved perfectly together we would have $x_1(n) - x_i(n) = 0$ for all i. Instead, in both the 3-node and the 10-node network we observe node-size differences that oscillate between positive and negative values before converging to 0 as the node sizes converge to $\left(\frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho}\right)$.



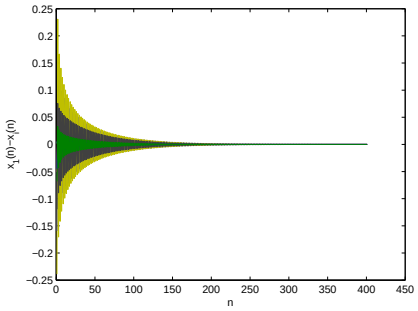
(a) Node size over time in a complete network with $m = 3$ and $\rho = 0.99\rho_{c_m}$.

(b) Difference in node size over time in a complete network with $m = 3$ and $\rho = 0.99\rho_{c_m}$.

Figure 3.23: Node size over time in a complete network with $m = 3$ and $\rho = 0.99\rho_{c_m} \approx 2.50$.



(a) Node size over time in a complete network with $m = 10$ and $\rho = 0.99\rho_{c_m}$.



(b) Difference in node size over time in a complete network with $m = 10$ and $\rho = 0.99\rho_{c_m}$.

Figure 3.24: Node size over time in a complete network with $m = 10$ and $\rho = 0.99\rho_{c_m} \approx 2.12$.

In Figures 3.25 to 3.28 we take a value for ρ just greater than the critical value ρ_{c_m}, below which the fixed point $(x^*, \ldots, x^*)$ is stable. Again we plot the dynamics of node sizes in a 3-node complete network and a 10-node complete network, choosing the initial node sizes $x_i(0)$ randomly from the interval $(0, 1)$. For both networks we observe behaviour resembling what was seen in the 2-node network in Section 3.1.1. We observe that there is no convergence to $(x^*, \ldots, x^*)$ as $n \to \infty$. Instead in both cases all of the $x_i(n)$'s appear to settle into one of two period-two orbits which are out of phase. The right-hand panels of Figure 3.25 and Figure 3.27 allow us to get a closer look at this settled behaviour. In Figure 3.25 we see that in this case the sizes of the nodes in the three-node network settle into two period-two orbits with significantly different oscillation ranges. In Figure 3.27 the oscillation ranges of the two observed orbits have quite similar oscillation ranges.

In Figure 3.26 and Figure 3.28 we have plotted the dynamics of the difference, $x_1(n) - x_i(n)$, between each node size and the size of v_1, $x_1(n)$, as $n \to \infty$. In the three-node case shown in Figure 3.26 we can see from the right-hand panel that neither $x_2(n)$ nor $x_3(n)$ is equal to $x_1(n)$ as $n \to \infty$. Instead $x_2(n) = x_3(n)$ and the difference $x_1(n) - x_2(n)$ is periodic with period 2, oscillating between a positive value and a negative value. In the ten-node case shown in Figure 3.28 we can see from the right-hand panel that some other node sizes, $x_i(n)$ are equal to $x_1(n)$ as $n \to \infty$ since $x_1(n) - x_i(n) = 0$ for some $i \neq 1$. All other $x_j(n) \neq x_1(n)$ are equal as $n \to \infty$ and the difference $x_1(n) - x_j(n)$ is again periodic with period 2, oscillating between a positive value and a negative value.

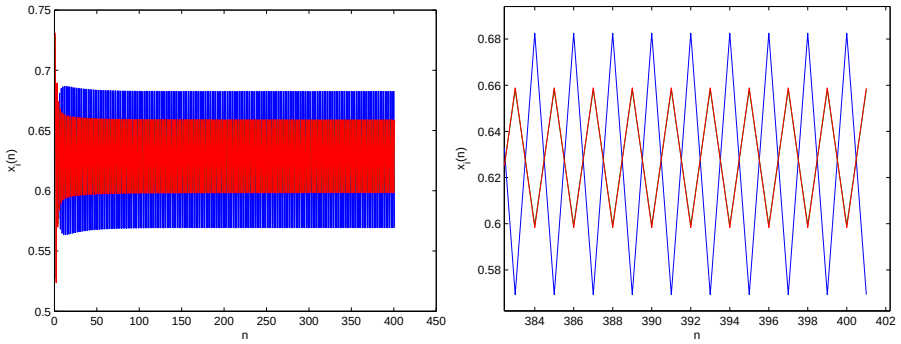


Figure 3.25: Node size over time in a complete network with $m = 3$ and $\rho = 1.01\rho_{c_m} \approx 2.55$.

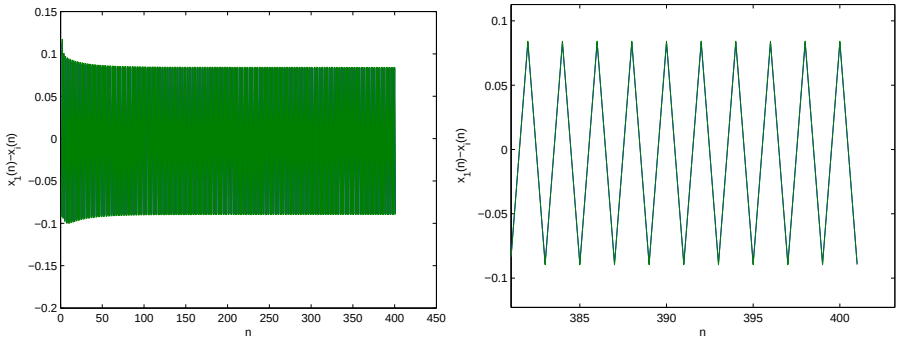


Figure 3.26: Difference in node size over time in a complete network with $m = 3$ and $\rho = 1.01\rho_{c_m} \approx 2.55$.

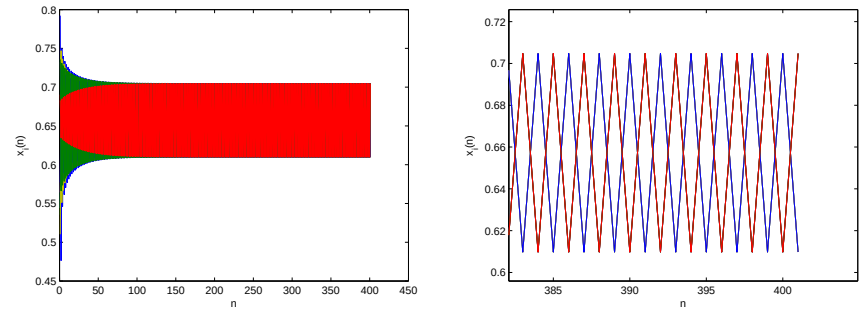


Figure 3.27: Node size over time in a complete network with $m = 10$ and $\rho = 1.01\rho_{c_m} \approx 2.16$.

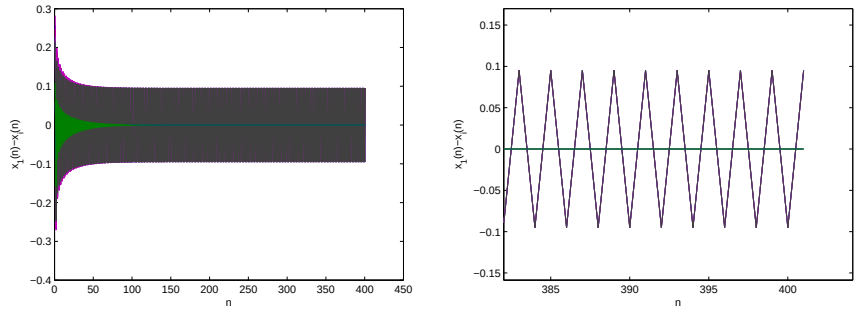


Figure 3.28: Difference in node size over time in a complete network with $m = 10$ and $\rho = 1.01\rho_{c_m} \approx 2.16$.

Increasing ρ further sees these period-two attractors become unstable in both cases. For sufficiently large values of ρ we observe what appears to be chaotic behaviour, we confirm that this behaviour is indeed chaotic in Section 3.3. Examples of this chaotic behaviour for both the three-node and the ten-node complete network are shown in Figures 3.29 to 3.32. As $n \to \infty$ the dynamics of the $x_i(n)$s in both cases do not settle into any regular recognisable pattern. This is shown in Figure 3.29 for the three-node case and in Figure 3.31 for the ten-node case. We can see from the right-hand panels in both figures that node size dynamics to not fall into any recognisable pattern even after 400 timesteps. It is interesting to note that in the three-node case the relationship between the different node sizes for this value of ρ is not entirely chaotic. In the right-hand panel of Figure 3.30 we can see that as $n \to \infty$, although the behaviour of both $x_1(n)$ and $x_3(n)$ is chaotic, their difference $x_1(n) - x_3(n)$ which is shown in green converges to 0 and they become equal. The relationship between these two node sizes and $x_2(n)$ on the other hand appears to be chaotic, with their difference $x_1(n) - x_2(n)$ settling into no recognisable pattern as $n \to \infty$.

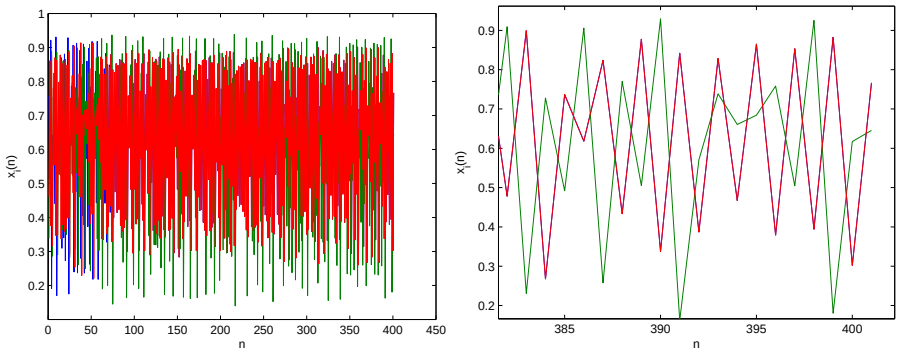


Figure 3.29: Node size over time in a complete network with $m = 3$ and $\rho = 3.5$.

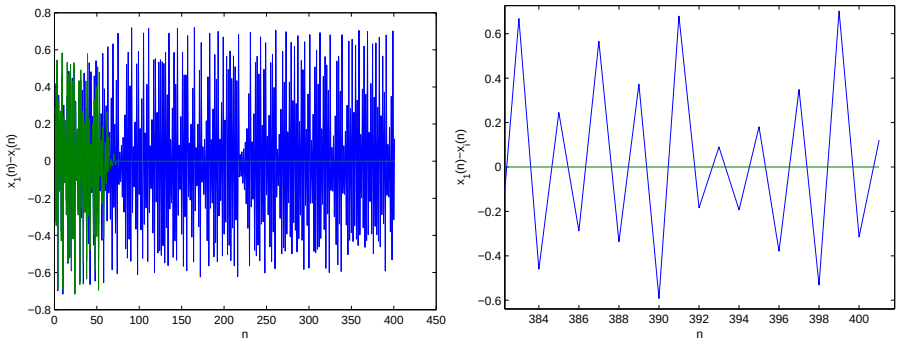


Figure 3.30: Difference in node size over time in a complete network with $m = 3$ and $\rho = 3.5$.

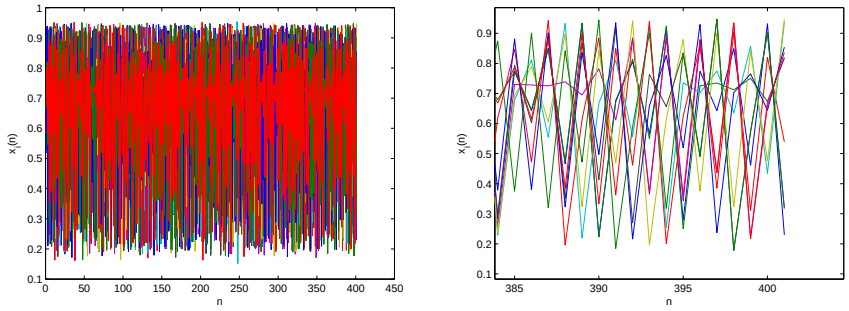


Figure 3.31: Node size over time in a complete network with $m = 10$ and $\rho = 1.4\rho_{c_m} \approx 3.00$.

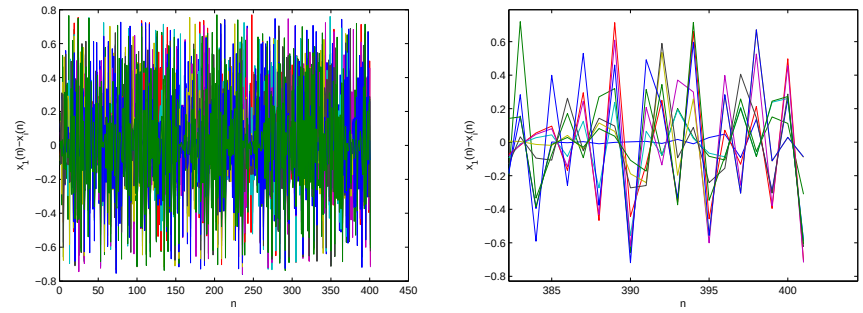
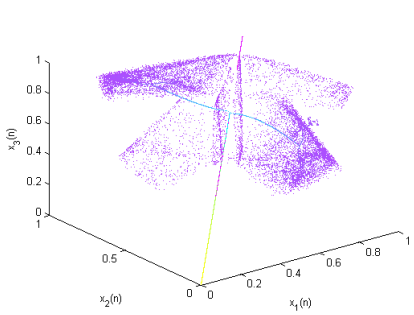


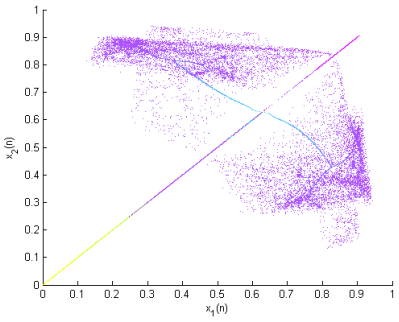
Figure 3.32: Difference in node size over time in a complete network with $m = 10$ and $\rho = 1.4\rho_{c_m} \approx 3.00$.

3.2.2 Bifurcation Diagrams

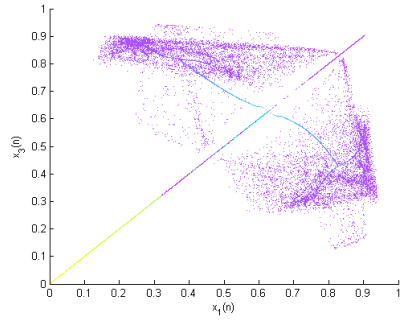
The first set of bifurcation diagrams in this section show the long-term behaviour of the complete network logistic map model for $m = 3$ and values of $\rho \in (0, 3.91]$. Beyond $\rho = \rho_{c_3}$ our numerical work has shown there are 3 separate attractors with different basins of attraction. In the first part of our analysis we limit ourselves to tracking one of these attractors. Later we try to show all 3 attractors in the same bifurcation diagram. In reality however, when considering a complete network in which the nodes do not have any specific properties attached to them other than size these three attractors are one and the same. The three attractors can be generated by a single attractor by taking three simple permutations of x_1, x_2 and x_3.



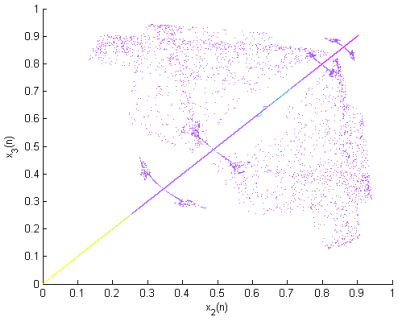
(a) Long-term behaviour of the relationship between $x_1(n)$, $x_2(n)$ and $x_3(n)$.



(b) $x_1(n)$ vs $x_2(n)$.



(c) $x_1(n)$ vs $x_3(n)$.

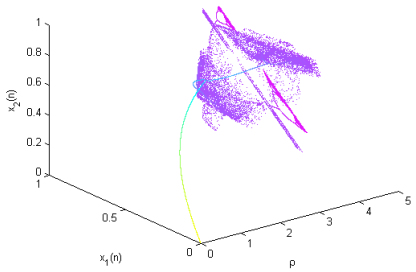


(d) $x_2(n)$ vs $x_3(n)$.

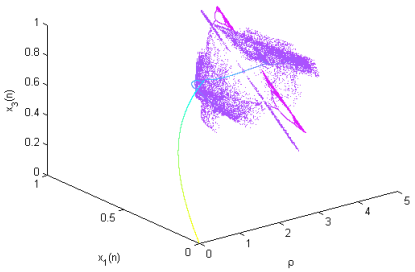
Figure 3.33: Long-term behaviour of the relationship between node sizes in a complete network with $m = 3$, colour corresponds to ρ value.

By examining Figure 3.33 and Figure 3.34 we can see that after the bifurcation at ρ_{c_3} node sizes are absorbed by a period-two orbit. In this example the period-two behaviour of $x_1(n)$ has the wider oscillation range while $x_2(n)$ and $x_3(n)$ have an identical, narrower oscillation range. The two sets of nodes also

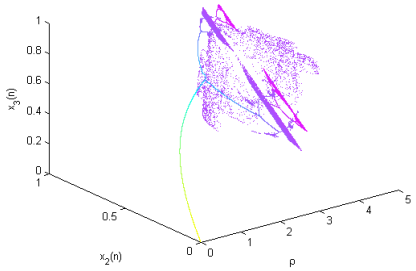
oscillate out of phase.



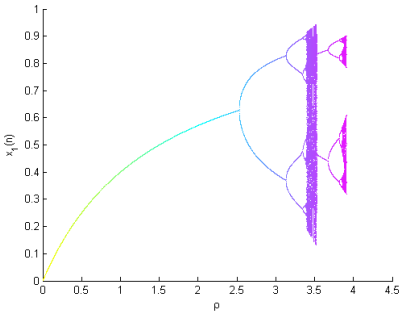
(a) Bifurcation diagram for $x_1(n)$ and $x_2(n)$.



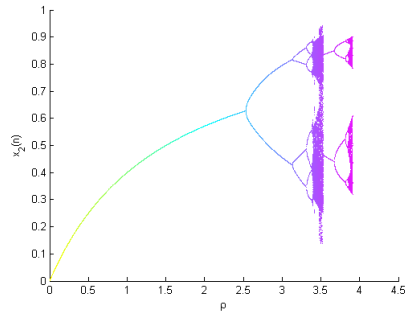
(b) Bifurcation diagram for $x_1(n)$ and $x_3(n)$.



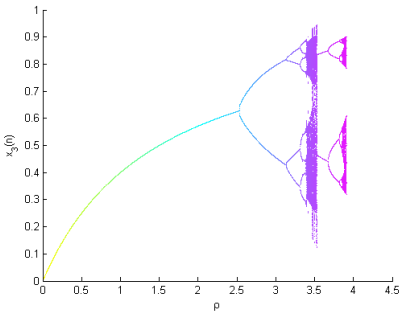
(c) Bifurcation diagram for $x_2(n)$ and $x_3(n)$.



(d) Bifurcation diagram for $x_1(n)$.



(e) Bifurcation diagram for $x_2(n)$.



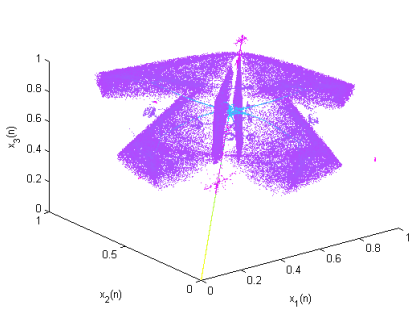
(f) Bifurcation diagram for $x_3(n)$.

Figure 3.34: Bifurcation diagrams for a complete network with $m = 3$, colour corresponds to ρ value.

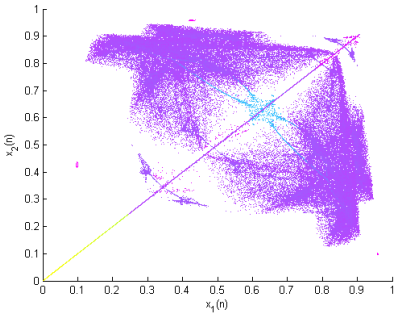
The next bifurcation sees another period-doubling bifurcation and a subsequent cascade to chaos, much like the behaviour of the single logistic map. This contrasts with the behaviour of the network logistic map of N_2^C where we saw

a very different route to chaos via a quasi-periodic attractor and a period-ten attractor. From chaos a period-two attractor emerges before once again experiencing a period-doubling cascade to chaos. In this region it appears that all three nodes behave identically.

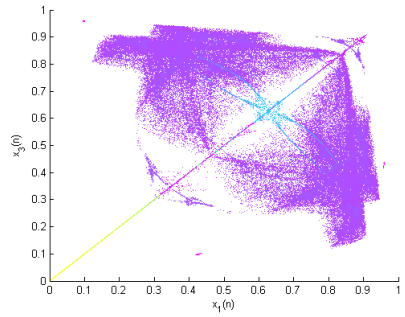
In Figure 3.35 and Figure 3.36 we show all attractors. The three separate period-two attractors after the bifurcation at ρ_{c_3} are generated by choosing any pair of the three node sizes to behave like $x_2(n)$ and $x_3(n)$ are shown to behave in Figure 3.33 and Figure 3.34 and allowing the remaining node to behave like $x_1(n)$. Interestingly, after the descent to chaos it appears that only one attractor emerges from the chaos and all nodes behave identically for subsequent values of ρ. We will return to further investigate this behaviour in Section 3.4 where we show that all three nodes do indeed behave identically for these values of ρ.



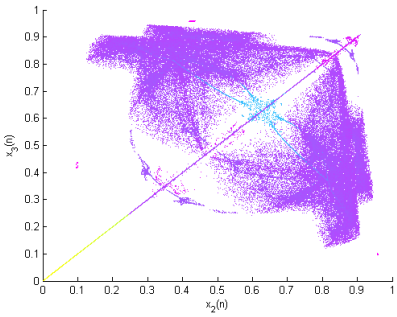
(a) Long-term behaviour of the relationship between $x_1(n)$, $x_2(n)$ and $x_3(n)$.



(b) $x_1(n)$ vs $x_2(n)$.

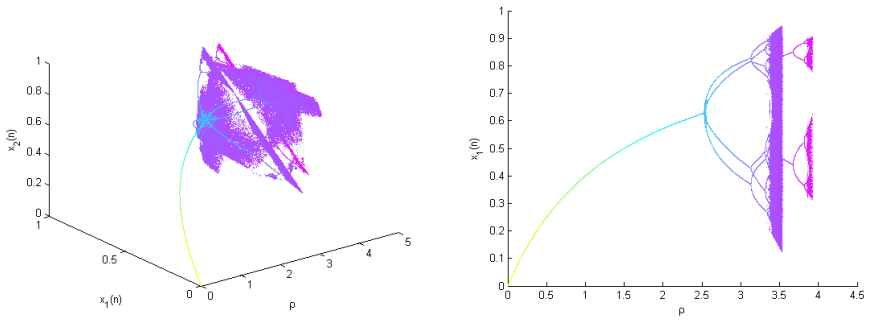


(c) $x_1(n)$ vs $x_3(n)$.



(d) $x_2(n)$ vs $x_3(n)$.

Figure 3.35: Long-term behaviour of the relationship between node sizes considering all attractors in a complete network with $m = 3$, colour corresponds to ρ value.

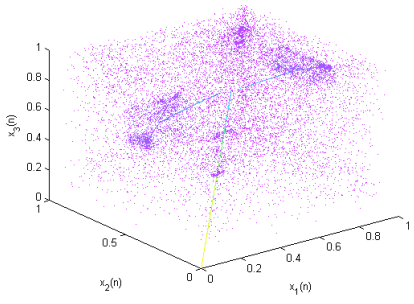


(a) Bifurcation diagram for $x_1(n)$ and $x_2(n)$.

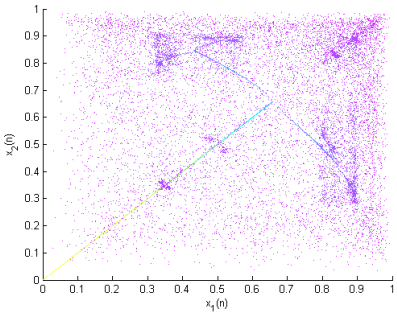
(b) Bifurcation diagram for $x_1(n)$.

Figure 3.36: Bifurcation diagrams showing all attractors for a complete network with $m = 3$, colour corresponds to ρ value.

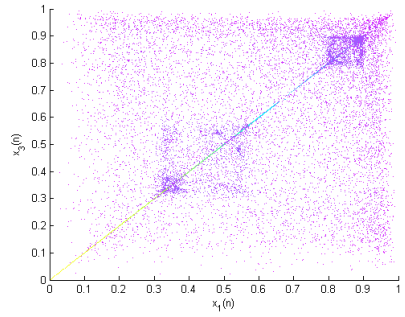
The bifurcation diagrams that follow in this section show the long-term behaviour of the complete network logistic map model for $m = 10$ and values of $\rho \in (0, 3.]$. Beyond $\rho = \rho_{c_m}$ there are $\sum_{i=0}^{\lfloor \frac{m}{2} \rfloor} \binom{m}{i}$ separate period-two attractors with different basins of attraction for a complete network with $m \in \{4, 5, 6, \ldots, 10\}$ nodes where $\lfloor x \rfloor$ is the integer floor function (In fact we believe this is indeed the case for all $m \geq 4$). In the first part of our analysis we limit ourselves to tracking one of these attractors, focusing on the behaviour of three node sizes $x_1(n)$, $x_2(n)$, $x_3(n)$. $x_1(n)$ and $x_2(n)$ are in phase while $x_3(n)$ is out of phase. Later we try to show all attractors. Only $1 + \lfloor \frac{m}{2} \rfloor$ of these are truly different period-two attractors the others are simply given by permuting the x_is.



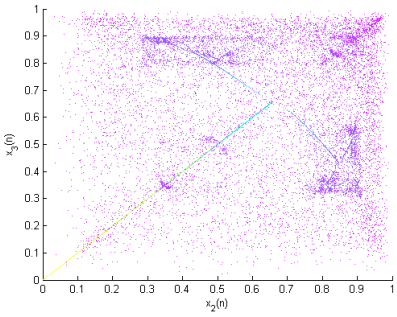
(a) Long-term behaviour of the relationship between $x_1(n)$, $x_2(n)$ and $x_3(n)$.



(b) Long-term behaviour of the relationship between $x_1(n)$ and $x_2(n)$.

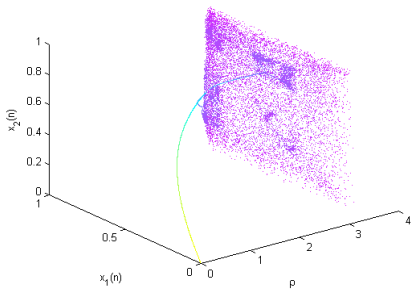


(c) Long-term behaviour of the relationship between $x_1(n)$ and $x_3(n)$.

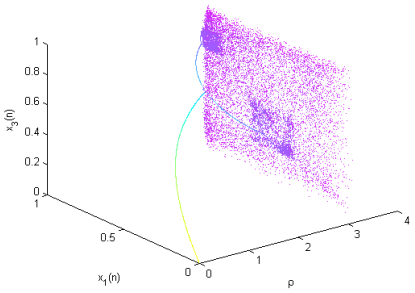


(d) Long-term behaviour of the relationship between $x_2(n)$ and $x_3(n)$.

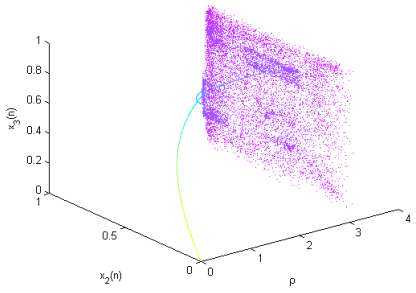
Figure 3.37: Long-term behaviour of the relationship between node sizes in a complete network with $m = 10$, colour corresponds to ρ value.



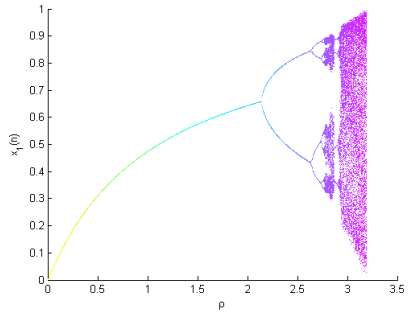
(a) Bifurcation diagram for $x_1(n)$ and $x_2(n)$.



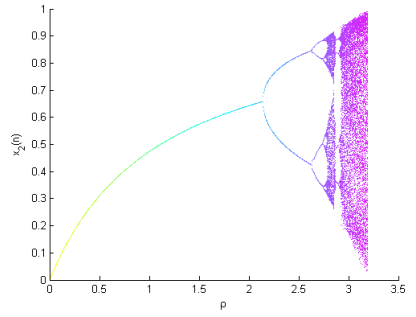
(b) Bifurcation diagram for $x_1(n)$ and $x_3(n)$.



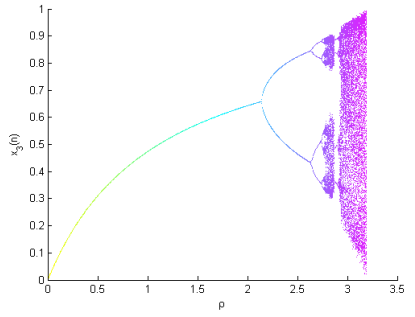
(c) Bifurcation diagram for $x_2(n)$ and $x_3(n)$.



(d) Bifurcation diagram projection for $x_1(n)$.



(e) Bifurcation diagram projection for $x_2(n)$.

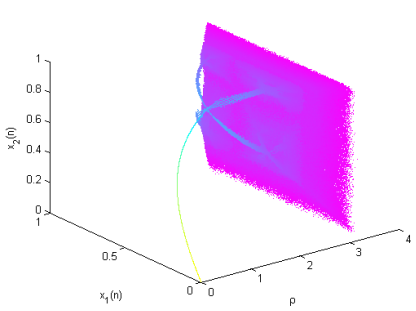


(f) Bifurcation diagram projection for $x_3(n)$.

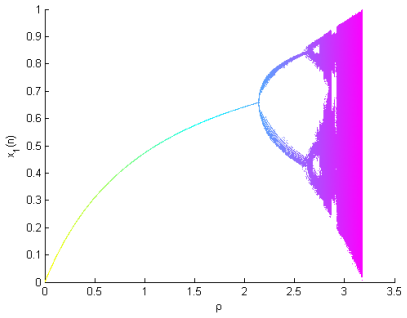
Figure 3.38: Bifurcation diagrams for $x_1(n)$, $x_2(n)$ and $x_3(n)$ in a complete network with $m = 10$, colour corresponds to ρ value.

Examining Figure 3.37 and Figure 3.38 we can see that after the bifurcation at $\rho_{c_{10}}$ node sizes are absorbed by a period-two orbit. In this example we focus on three node sizes, $x_1(n)$, $x_2(n)$ and $x_3(n)$. The period-two behaviour of $x_1(n)$ and $x_3(n)$ is identical while $x_2(n)$ has a different oscillation range and is out of

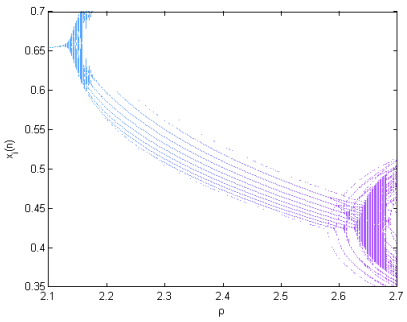
phase with $x_1(n)$ and $x_3(n)$. Just as in the three-node case node sizes are split into two sets that move together out of phase with the other. Subsequent bifurcations see a period-doubling cascade to chaos. the chaos itself is interrupted by a large periodic window which begins with a period four attractor and cascades once again to chaos.



(a) Bifurcation diagram for two nodes in a ten node complete network showing all attractors.



(b) Bifurcation diagram projection for a single node in a ten node complete network showing all attractors.



(c) A close up look at all period-two attractors.

Figure 3.39: Bifurcation diagrams showing all attractors in a complete network with $m = 10$

In Figure 3.39 we show bifurcation diagrams for two typical nodes in N_{10}^C with all attractors. Our focus is on the period-two behaviour of the system after ρ_{c10} where we see the appearance of several period-two attractors, some becoming stable later than others. In Figure 3.39c it can be seen that each node can fall into one of ten different stable period-two orbits. Each of these orbits corresponds to being in a set S of $s \in \{1, 2, \ldots, 10\}$ nodes in the system whose node sizes move exactly together, i.e for all $v_i, v_j \in S$ $x_i(n) = x_j(n)$ as $n \to \infty$. The other $10 - s$ node sizes also fall into a stable period-two orbit moving exactly together out of phase with the nodes sizes in S, i.e for all $v_i, v_j \notin S$

$x_i(n) = x_j(n)$ as $n \to \infty$.

3.3 Lyapunov Exponents

We have numerically estimated the maximum Lyapunov exponent of the network logistic map for different values of ρ for $m = 3$. Our estimates were again obtained by implementing a version of an algorithm due to Eckmann and Ruelle in Matlab. We plot the numerically estimated exponents in Figure 3.40.

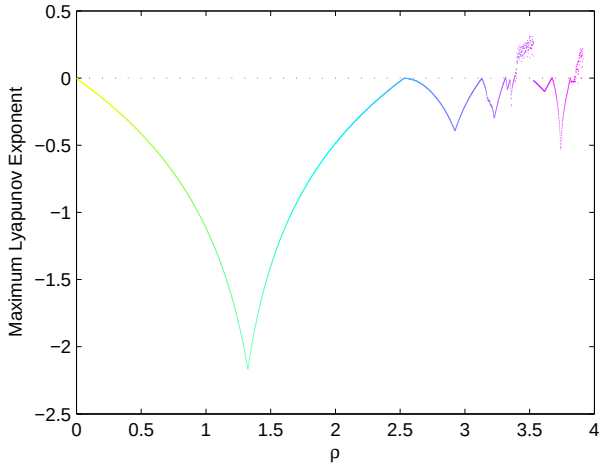


Figure 3.40: Maximum Lyapunov Exponents for an Complete 3 Node Network.

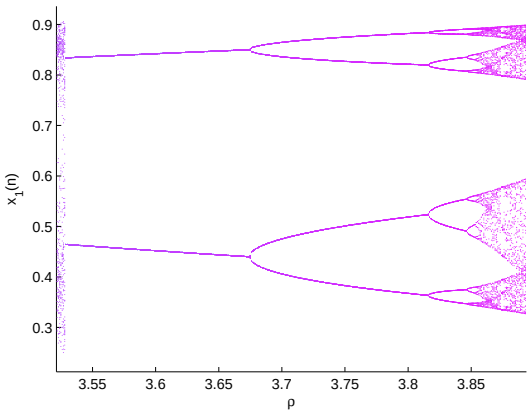


Figure 3.41: A close up look at a projection of the large periodic window seen in the case of a complete network with $m = 3$ nodes.

Positive Lyapunov exponents are observed for values of $\rho > \rho_{\infty_3} \approx 3.39$

indicating chaotic behaviour in this region. However we observe negative Lyapunov exponents for $\rho \in (3.53, 3.86)$. These negative values correspond to a large periodic window which is one of the striking features of the bifurcation diagrams in Figures 3.34 and 3.36. Figure 3.41

We have also numerically estimated the maximum Lyapunov exponent of the network logistic map for different values of ρ for $m = 10$. We plot the numerically estimated exponents in Figure 3.42. We again see evidence of chaotic behaviour for large values of ρ and periodic windows amongst the chaos. One of these periodic windows is again interesting due to its size and is easy to see with the naked eye in Figure 3.38.

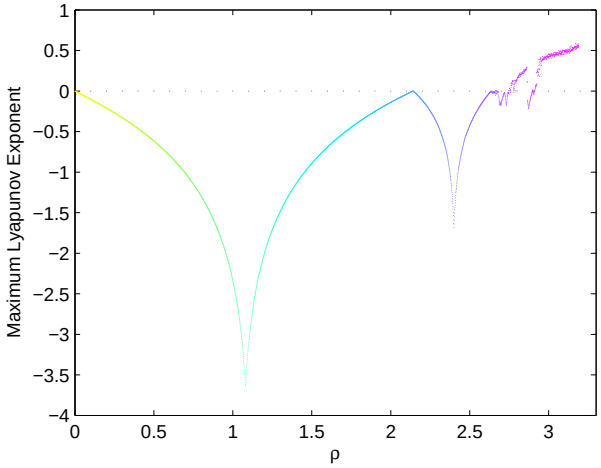


Figure 3.42: Maximum Lyapunov Exponents for an Complete 10 Node Network.

3.4 An Invariant Set

Definition 3.4.1. An **invariant set** of a map (1.1.1) is a subset $A \subset X$ such that $x \in A$ implies $f(x) \in A$. It then follows that $f^n(x) \in A$ for all $n \in \mathbb{N}$.

For every complete network N there exists an invariant set of the corresponding network logistic map, L, given by

$$L = \{x = (x_1, x_2, \ldots, x_m) = t(1, 1, \ldots, 1) : t \in (0, 1)\}. \tag{3.96}$$

For any $x(n) \in L$ we also have that $x(n+1) \in L$. This means that if we choose any initial condition $x(0)$ such that $x(0) \in L$ we will remain on L as $n \to \infty$.

Consider a complete network with m nodes such that $x(n) = (x_1(n), x_2(n), \ldots, x_m(n)) \in L$. The long-term behaviour of each individual $x_i(n)$ is the same as the long-term behaviour given by the single logistic map with control parameter $r = 1 + \rho\left(\frac{m-1}{m}\right)$. This is because for all $x(n) \in L$ and $i \in \{1, 2, \ldots, m\}$ we have

$$r_i(n) = 1 + \rho w_i(n) = 1 + \rho \left(\frac{\sum\limits_{j:(x_i,x_j)\in E} x_j(n)}{\sum\limits_{j=1}^{m} x_j(n)} \right) = 1 + \rho \left(\frac{\sum\limits_{\substack{j=1 \\ j\neq i}}^{m} x_j(n)}{\sum\limits_{j=1}^{m} x_j(n)} \right)$$
$$= 1 + \rho \left(\frac{(m-1)t}{mt} \right) = 1 + \rho \left(\frac{m-1}{m} \right). \quad (3.97)$$

We are interested in the fixed points and periodic orbits of the invariant set, L, and their stability. We consider their stability in terms of both L and the overall system. Any fixed point or orbit that is stable in terms of the overall system is of course also stable in terms of L, however the converse is not necessarily true.

3.4.1 The Attractors of The Invariant Set

Using the information above we can easily analyse the long-term dynamics of points on the invariant set. For any point $x(n) = (x_1(n), x_2(n), \ldots, x_m(n))$ on L we have that $x(n) = y(n)(1, 1, \ldots, 1)$ where $y(n) \in (0, 1)$ and

$$y(n+1) = \left(1 + \rho \left(\frac{m-1}{m} \right) \right) y(n)(1 - y(n)). \quad (3.98)$$

Referring to Section 2 we can immediately see that the fixed point of $y(n)$ is given by

$$y^* = \frac{(m-1)\rho}{m + (m-1)\rho} \quad (3.99)$$

and so the fixed point of the invariant set is $\left(\frac{(m-1)\rho}{m+(m-1)\rho}, \frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho} \right)$. Since $y^* = \frac{(m-1)\rho}{m+(m-1)\rho}$ is a stable fixed point of $y(n)$ for $0 < \rho < \frac{2m}{m-1} = \rho_{c_m}^{SM}$. We have that $\left(\frac{(m-1)\rho}{m+(m-1)\rho}, \frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho} \right)$ is stable in terms of the invariant set for values of ρ in this range.

Again referring to Section 2 we find that the period-two orbit of the map given in 3.98 is given by

$$(y_+^*, y_-^*) = \left(\frac{2 + \rho\left(\frac{m-1}{m}\right) + \sqrt{\rho^2\left(\frac{m-1}{m}\right)^2 - 4}}{2\left(1 + \rho\left(\frac{m-1}{m}\right)\right)}, \frac{2 + \rho\left(\frac{m-1}{m}\right) - \sqrt{\rho^2\left(\frac{m-1}{m}\right)^2 - 4}}{2\left(1 + \rho\left(\frac{m-1}{m}\right)\right)} \right) \quad (3.100)$$

and exists ony for $\rho > \frac{2m}{m-1}$. This orbit is stable for $\frac{2m}{m-1} < \rho < \frac{\sqrt{6}m}{m-1}$. This means that $\left((y_+^*, y_+^*, \ldots, y_+^*), (y_-^*, y_-^*, \ldots, y_-^*) \right)$ is a period-two attractor which is stable in terms of the set for those values of ρ.

Continuing in the same manner it is possible to find expressions for further attractors and bifurcation points in terms of ρ and m as the system period doubles to chaos along the set, for example the accumulation point after which chaos occurs is given by $\rho_{\infty}^{SM} = (2.569946\ldots)\frac{m}{m-1}$. However, the rest of our analysis will focus on the stability of the period one and period-two attractors in terms of the overall system.

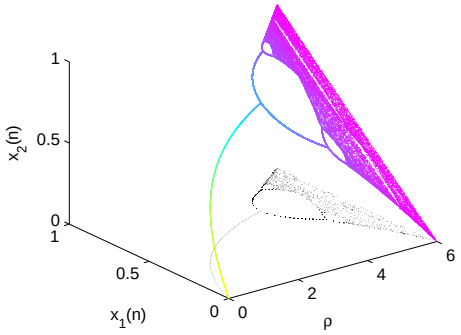


Figure 3.43: Bifurcation diagram for a two node system beginning on the invariant set in colour. The logistic map with control parameter $r = 1 + \rho\left(\frac{m-1}{m}\right)$ in black.

3.4.2 The Overall Stability of These Attractors

We have already seen in Section 3.2 that $\left(\frac{(m-1)\rho}{m+(m-1)\rho}, \frac{(m-1)\rho}{m+(m-1)\rho}, \ldots, \frac{(m-1)\rho}{m+(m-1)\rho}\right)$ is a stable fixed point of the network logistic map of a complete network with m nodes for

$$0 < \rho < \frac{m^2 - 2m + \sqrt{12m^2 - 20m^3 + 9m^4}}{2(m-1)^2} = \rho_{c_m} < \frac{2m}{m-1} = \rho_{c_m}^{SM}. \quad (3.101)$$

It is interesting that although in general this fixed point becomes unstable earlier in terms of the overall system, these two bifurcation points converge as $m \to \infty$.

$$\lim_{m\to\infty} \rho_{c_m} = \lim_{m\to\infty} \frac{m^2 - 2m + \sqrt{12m^2 - 20m^3 + 9m^4}}{2(m-1)^2} = 2 = \lim_{m\to\infty} \frac{2m}{m-1} = \lim_{m\to\infty} \rho_{c_m}^{SM}. \quad (3.102)$$

To test the stability of the period-two fixed point we need to examine the eigenvalues of the matrix given by the product

$$\mathbf{J}(y_+^*, y_+^*, \ldots, y_+^*)\mathbf{J}(y_-^*, y_-^*, \ldots, y_-^*) = \begin{pmatrix} ac + (m-1)bd & ad + bc + (m-2)bd & \ldots & ad + bc + (m-2)bd \\ ad + bc + (m-2)bd & ac + (m-1)bd & \ldots & ad + bc + (m-2)bd \\ \vdots & \vdots & \ddots & \vdots \\ ad + bc + (m-2)bd & ad + bc + (m-2)bd & \ldots & ac + (m-1)bd \end{pmatrix} \quad (3.103)$$

where

$$\begin{aligned} a &= (1-2y_+^*)\left(1+\rho\left(\frac{m-1}{m}\right)\right)-(1-y_+^*)\rho\left(\frac{m-1}{m^2}\right), \\ b &= \frac{\rho}{m^2}(1-y_+^*), \\ c &= (1-2y_-^*)\left(1+\rho\left(\frac{m-1}{m}\right)\right)-(1-y_-^*)\rho\left(\frac{m-1}{m^2}\right), \\ d &= \frac{\rho}{m^2}(1-y_-^*). \end{aligned} \tag{3.104}$$

By observation we find that $\mathbf{J}(y_+^*, y_+^*, \ldots, y_+^*)\mathbf{J}(y_-^*, y_-^*, \ldots, y_-^*)$ has $\mathbf{v}_1 = (1, 1, \ldots, 1)^T$ as an eigenvector with corresponding eigenvalue

$$\lambda_1 = ac + (m-1)(ad + bc + (m-1)bd). \tag{3.105}$$

$\mathbf{J}(y_+^*, y_+^*, \ldots, y_+^*)\mathbf{J}(y_-^*, y_-^*, \ldots, y_-^*)$'s other eigenvectors are given by $(-1, 1, 0, \ldots, 0)^T$, $(-1, 0, 1, \ldots, 0)^T$, $\ldots$, $(-1, 0, 0, \ldots, 1)^T$ and all have the corresponding eigenvalue

$$\lambda_2 = (a-b)(c-d). \tag{3.106}$$

For stability we require $|\lambda_1| < 1$ and $|\lambda_2| < 1$. First

$$\begin{aligned} |\lambda_1| &< 1 \implies \\ \left|\frac{m^2\rho^2 - 2m\rho^2 - 5m^2 + \rho^2}{m^2}\right| &< 1 \implies \\ \frac{\left|m^2\rho^2 - 2m\rho^2 - 5m^2 + \rho^2\right|}{m^2} &< 1 \implies \\ \left|(m-1)^2\rho^2 - 5m^2\right| &< m^2. \end{aligned} \tag{3.107}$$

Using the fact that $m \geq 2$ and $\rho > 0$ this gives us that

$$\begin{aligned} -m^2 < \quad (m-1)^2\rho^2 - 5m^2 \quad &< m^2 \implies \\ 4m^2 < \quad (m-1)^2\rho^2 \quad &< 6m^2 \implies \\ \frac{4m^2}{(m-1)^2} < \quad \rho^2 \quad &< \frac{6m^2}{(m-1)^2} \implies \\ \frac{2m}{m-1} < \quad \rho \quad &< \frac{\sqrt{6}m}{m-1}. \end{aligned} \tag{3.108}$$

Notice that this is the same range of values of ρ for which this period-two orbit is an attractor for the invariant set. The second inequality gives us that

$$|\lambda_2| < 1$$

$$\begin{aligned} |\lambda_2| &< 1 \implies \\ \left|\frac{1}{4}\left(\frac{1}{m^2(m\rho + m - \rho)^2}\right)\right| & \end{aligned}$$

$$\begin{aligned}&\times\left|((2\gamma m^2\rho+2\gamma m^2-3\gamma m\rho+2m^2\rho+m\rho^2+2m^2-2\rho m-\rho^2)\right|\\&\times\left|(2\gamma m^2\rho+2\gamma m^2-3\gamma m\rho-2m^2\rho-m\rho^2-2m^2+2\rho m+\rho^2))\right| \quad < \quad 1\end{aligned} \tag{3.109}$$

where $\gamma=\sqrt{\dfrac{m^2\rho^2-2m\rho^2-4m^2+\rho^2}{m^2}}$.

For $m=2$ for example using the fact that $\rho>0$ this gives that

$$4<\rho<\infty. \tag{3.110}$$

Taking the inequalities in (3.108) and (3.110) together this gives that the period-two orbit given by $\big((y_+^*,\ldots,y_+^*),(y_-^*,\ldots,y_-^*)\big)$ is not stable for any value of ρ in the overall sense.

For $m=3$ we find that

$$\begin{aligned}&-\frac{3}{4}+\frac{9}{8}\sqrt{2}+\frac{3}{8}\sqrt{22+20\sqrt{2}}<\rho\\&<-\frac{3}{4}+\frac{1}{8}\Omega+18\sqrt{6}\left(\frac{(\chi^{\frac{2}{3}}+72\chi^{\frac{1}{3}}+129)\Omega+828\chi^{\frac{1}{3}}}{\chi^{\frac{1}{3}}\Omega}\right)^{\frac{1}{2}}\end{aligned} \tag{3.111}$$

where

$$\Omega=\sqrt{\frac{-6(2079+24\sqrt{3777})^{\frac{2}{3}}+216(2079+24\sqrt{3777})^{\frac{1}{3}}-774}{(2079+24\sqrt{3777})^{\frac{1}{3}}}}$$

and $\chi=2079+24\sqrt{3777}$.

Taking the inequalities in (3.108) and (3.111) together this gives that the period-two orbit given by $\big((y_+^*,\ldots,y_+^*),(y_-^*,\ldots,y_-^*)\big)$ is an attractor for

$$3.500167883\ldots=-\frac{3}{4}+\frac{9}{8}\sqrt{2}+\frac{3}{8}\sqrt{22+20\sqrt{2}}<\rho<\frac{3\sqrt{6}}{2}=3.674234614\ldots. \tag{3.112}$$

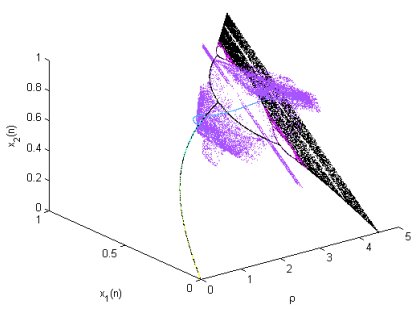
In Figure 3.44 we look at the case of a complete network with three nodes and have plotted bifurcation diagrams for both the overall system and the invariant set. Notice that our earlier results are confirmed numerically by these diagrams. Both have the same fixed point attractor to begin with however this attractor becomes unstable earlier in the overall case. Initially the period-two attractor for the invariant set is not stable in the case of the overall system. We see that after a period of chaos in the overall system the period-two attractor of the invariant set becomes stable in the overall sense. In fact all attractors for the overall system after this point before the map leaves its domain are attractors on the invariant set.

For $m=10$ it can be shown that

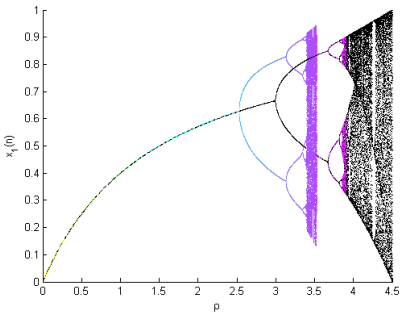
$$|\lambda_2|<1 \implies 2.269119885\cdots<\rho<2.810648520\ldots. \tag{3.113}$$

Taking the inequalities in (3.108) and (3.113) together we see that the period-two attractor for the invariant set is an attractor in the overall sense for

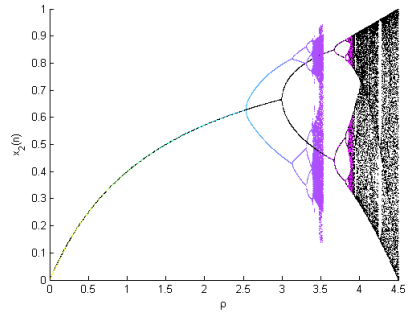
$$2.269119885\cdots<\rho<\frac{10\sqrt{6}}{9}=2.721655269\ldots. \tag{3.114}$$



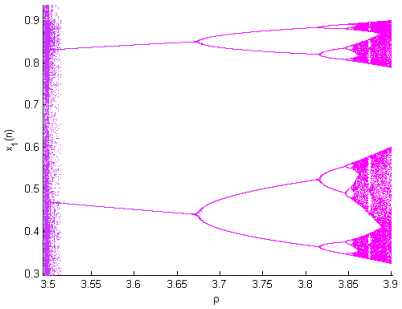
(a) Bifurcation diagram for $x_1(n)$ and $x_2(n)$ in a three-node complete network in colour with the bifurcation diagram for the invariant set in black.



(b) Bifurcation diagram projection for $x_1(n)$ in a three-node complete network in colour with the bifurcation diagram for the invariant set in black.



(c) Bifurcation diagram projection for $x_2(n)$ in a three-node complete network in colour with the bifurcation diagram for the invariant set in black.



(d) Close up look at the point at which the period-two attractor of the stable maniold becomes stable in the overall sense.

Figure 3.44: Bifurcation diagrams for a three-node complete network and the invariant set.

In Figure 3.45 we look at the case of a complete network with ten nodes and have plotted a bifurcation diagram projection for a typical node in the system showing all period-two attractors. We see that the period-two attractor of the invariant set becomes stable in terms of the overall system for $\rho > 2.269119885\ldots$ and this attractor corresponds to the period-two attractor for which all node sizes move together.

In the case of a 100 node network we find that the period-two attractor for the invariant set is an attractor in the overall sense for

$$2.023642493\cdots < \rho < \frac{100\sqrt{6}}{99} = 2.474232065\ldots. \tag{3.115}$$

This is almost the entire range of values of ρ for which the attractor is stable in

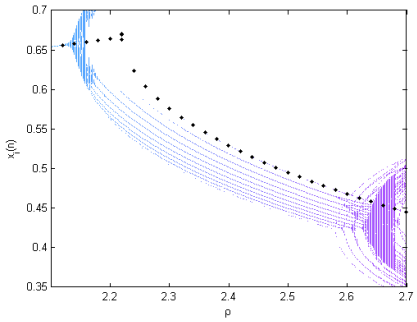


Figure 3.45: A close up look at all period-two attractors for a node in a complete network with ten nodes in colour, overlayed with the period-two attractor for the invariant set in black.

the sense of the invariant set given by

$$2.020202020\cdots = \frac{200}{99} < \rho < \frac{100\sqrt{6}}{99} = 2.474232065\ldots. \tag{3.116}$$

3.5 The Limiting Case

Since many real-world networks are large, i.e. have a large number of nodes, it is interesting to see how the system behaves as $m \to \infty$. We have seen already that in the limiting case of a complete network

$$\lim_{m\to\infty} x^* = \lim_{m\to\infty} \frac{(m-1)\rho}{m+(m-1)\rho} = \frac{\rho}{1+\rho} \tag{3.117}$$

and

$$\lim_{m\to\infty} \rho_{c_m} = 2 \tag{3.118}$$

corresponding to the single logistic map seen in Section 2 with control parameter $r = 1+\rho$. In fact, in the limiting case, since

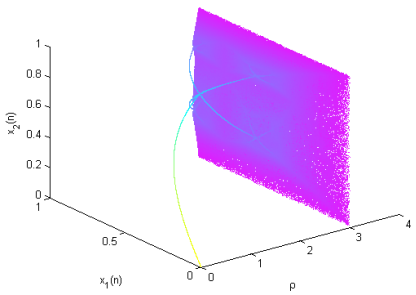
$$\begin{aligned}\lim_{m\to\infty} r_i(n) = \lim_{m\to\infty} 1+\rho w_i(n) &= \lim_{m\to\infty} 1+\rho\left(\frac{\sum\limits_{j:(x_i,x_j)\in E} x_j(n)}{\sum\limits_{j=1}^{m} x_j(n)}\right) \\ &= \lim_{m\to\infty} 1+\rho\left(\frac{\sum\limits_{\substack{j=1\\ j\neq i}}^{m} x_j(n)}{\sum\limits_{j=1}^{m} x_j(n)}\right) = 1+\rho, \end{aligned} \tag{3.119}$$

our entire system reduces to a series of m independent logistic maps describing the behaviour of the m node sizes, each with control parameter $r = 1 + \rho$

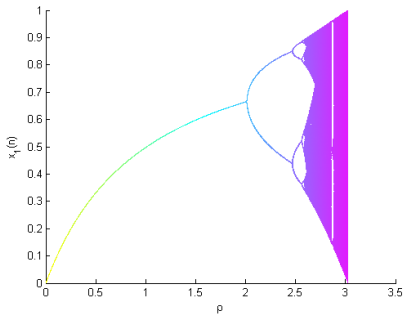
$$x_i(n+1) = (1+\rho)x_i(n)(1-x_i(n)). \tag{3.120}$$

3.5.1 An Example of a Complete Network With $m \ggg 1$

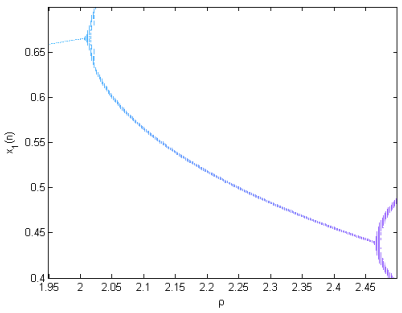
In this section we examine the long-term behaviour of the logistic network map for N_{100}^C to illustrate the convergence of the network logistic map of a complete network with m nodes to the behaviour described above as $m \to \infty$. In Figure 3.46 we can see that all attractors for very value of ρ in a complete network with $m = 100$ nodes are converging to a single attractor. This single attractor corresponds to the attractor in a single logistic map with control parameter $r = 1 + \rho$.



(a) Bifurcation diagram for two typical nodes in a complete network with $m = 100$ showing all attractors.



(b) Bifurcation diagram projection for a typical node in a complete network with $m = 100$ showing all attractors.



(c) A close up look at all period-two attractors for a typical node in a complete network with $m = 100$.

Figure 3.46: Bifurcation diagrams for a complete network with $m = 100$ showing all attractors.

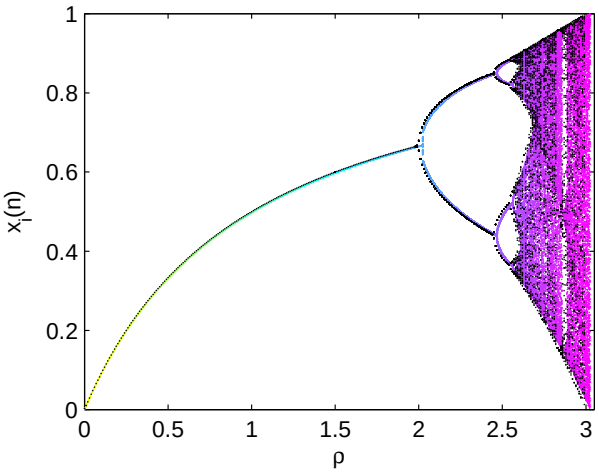


Figure 3.47: Bifurcation diagram for a typical $x_i(n)$ in a 100 node complete network in colour, overlayed with the bifurcation diagram for the single logistic map with control parameter $1 + \rho$ in black.

In Figure 3.47 we have plotted the bifurcation diagrams for a node growing according to both the single logistic map with control parameter $r = 1 + \rho$ described by Equation (3.120) and a typical node from a 100 node complete network growing according to the network logistic map described in Section 3.2. It is immediately clear that the behaviour in both cases is extremely similar.

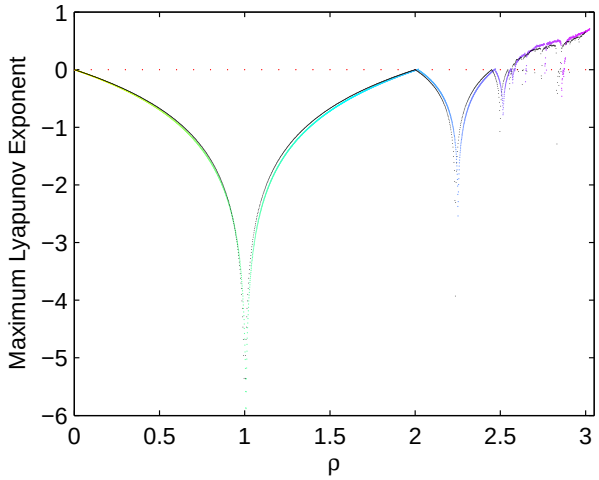


Figure 3.48: Maximum Lyapunov exponent for the 100 node complete network logistic map in colour, overlayed with the Lyapunov exponent for the single logistic map with control parameter $1 + \rho$ in black.

The major difference between the behaviour of the two systems is that each bifurcation in the 100 node system occurs slightly later than the corresponding bifurcation in the single logistic map. As m increases these bifurcation points converge.

Figure 3.48 allows us to compare the maximum Lyapunov exponent for the two systems. Again we can see that the behaviour of the two systems seems to be converging as predicted by our analysis in Section 3.5 and that the bifurcations occur for slightly lower values of ρ in the network case than the corresponding single case.

4 Growth of Nodes in an Incomplete Network

It is also interesting to consider more complicated, incomplete but connected networks. In fact, when applying this model of node growth later, in general the networks we will deal with will be incomplete. In this section we will look at various examples of incomplete networks. Again we set

$$r_i(n) = 1 + \rho w_i(n) \tag{3.121}$$

where $\rho > 0$ and $w_i(n)$ is the normalised weighted degree of v_i at time n.

4.1 A Network With 3 Nodes

We begin our analysis by first considering the simplest possible case. In the simplest case an incomplete network has three nodes v_1, v_2 and v_3 and two links (v_1, v_2) and (v_1, v_3). We call the node with two links, v_1, the *central node* or *hub node* of the network.

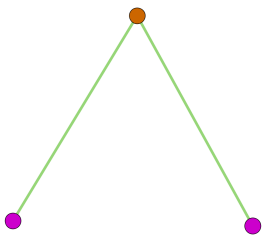


Figure 3.49: An incomplete network with three nodes. v_1 is coloured orange while v_2 and v_3 are coloured pink.

For this network configuration we then have

$$x_1(n+1) = r_1(n)x_1(n)\left(1 - x_1(n)\right) \tag{3.122}$$

where

$$r_1(n) = 1 + \rho\left(\frac{x_2(n) + x_3(n)}{x_1(n) + x_2(n) + x_3(n)}\right). \tag{3.123}$$

Similarly,

$$x_2(n+1) = r_2(n)x_2(n)\left(1 - x_2(n)\right) \tag{3.124}$$

where

$$r_2(n) = 1 + \rho\left(\frac{x_1(n)}{x_1(n) + x_2(n) + x_3(n)}\right). \tag{3.125}$$

$$x_3(n+1) = r_3(n)x_3(n)\left(1 - x_3(n)\right) \tag{3.126}$$

where

$$r_3(n) = 1 + \rho\left(\frac{x_1(n)}{x_1(n) + x_2(n) + x_3(n)}\right). \tag{3.127}$$

4.1.1 Fixed Points

Just as we saw in the complete cases studied in Section 3for all $x_i(n)$ we find that we have fixed points at $x_i^* = \frac{r_i(n)-1}{r_i(n)}$.

$$\begin{aligned} x_1^* &= \frac{r_1(n)-1}{r_1(n)} = \frac{1+\rho\left(\frac{x_2(n)+x_3(n)}{x_1^*+x_2(n)+x_3(n)}\right)-1}{1+\rho\left(\frac{x_2(n)+x_3(n)}{x_1^*+x_2(n)+x_3(n)}\right)} \\ &= \frac{\rho\left(\frac{x_2(n)+x_3(n)}{x_1^*+x_2(n)+x_3(n)}\right)}{\frac{x_1^*+(1+\rho)(x_2(n)+x_3(n))}{x_1^*+x_2(n)+x_3(n)}} = \frac{\rho(x_2(n)+x_3(n))}{x_1^*+(1+\rho)(x_2(n)+x_3(n))}, \end{aligned} \tag{3.128}$$

$$\begin{aligned} x_2^* &= \frac{r_2(n)-1}{r_2(n)} = \frac{1+\rho\left(\frac{x_1(n)}{x_1(n)+x_2^*+x_3(n)}\right)-1}{1+\rho\left(\frac{x_1(n)}{x_1(n)+x_2^*+x_3(n)}\right)} \\ &= \frac{\rho\left(\frac{x_1(n)}{x_1(n)+x_2^*+x_3(n)}\right)}{\frac{x_2^*+x_3(n)+(1+\rho)x_1(n)}{x_1(n)+x_2^*+x_3(n)}} = \frac{\rho x_1(n)}{x_2^*+x_3(n)+(1+\rho)x_1(n)} \end{aligned} \tag{3.129}$$

and

$$\begin{aligned} x_3^* &= \frac{r_3(n)-1}{r_3(n)} = \frac{1+\rho\left(\frac{x_1(n)}{x_1(n)+x_2(n)+x_3^*}\right)-1}{1+\rho\left(\frac{x_1(n)}{x_1(n)+x_2(n)+x_3^*}\right)} \\ &= \frac{\rho\left(\frac{x_1(n)}{x_1(n)+x_2(n)+x_3^*}\right)}{\frac{x_2(n)+x_3^*+(1+\rho)x_1(n)}{x_1(n)+x_2(n)+x_3^*}} = \frac{\rho x_1(n)}{x_2(n)+x_3^*+(1+\rho)x_1(n)}. \end{aligned} \tag{3.130}$$

So

$$\begin{aligned} x_2^* &= \frac{\rho x_1^*}{x_2^*+x_3^*+(1+\rho)x_1^*} = x_3^* \\ \implies x_2^* &= \frac{\rho x_1^*}{2x_2^*+(1+\rho)x_1^*} \end{aligned} \tag{3.131}$$

and

$$x_3^* = \frac{\rho x_1^*}{2x_3^* + (1+\rho)x_1^*}. \tag{3.132}$$

We also find

$$x_1^* = \frac{2\rho x_2^*}{x_1^* + 2(1+\rho)x_2^*} \tag{3.133}$$

Developing (3.133) further gives

$$(x_1^*)^2 + 2(1+\rho)x_1^* x_2^* - 2\rho x_2^* = 0. \tag{3.134}$$

Developing (3.131) further gives

$$2(x_2^*)^2 + (1+\rho)x_1^* x_2^* - \rho x_1^* = 0. \tag{3.135}$$

Similarly we find

$$2(x_3^*)^2 + (1+\rho)x_1^* x_3^* - \rho x_1^* = 0. \tag{3.136}$$

(3.134) - (3.135) gives

$$\begin{aligned} 0 &= (x_1^*)^2 + \left((1+\rho)x_2^* + \rho\right) x_1^* - 2x_2^*(x_2^* + \rho) \\ \implies x_1^* &= \frac{-\left((1+\rho)x_2^* + \rho\right) \pm \sqrt{\left((1+\rho)x_2^* + \rho\right)^2 + 8x_2^*(x_2^* + \rho)}}{2} \end{aligned} \tag{3.137}$$

We take

$$x_1^* = \frac{-\left((1+\rho)x_2^* + \rho\right) + \sqrt{\left((1+\rho)x_2^* + \rho\right)^2 + 8x_2^*(x_2^* + \rho)}}{2} \tag{3.138}$$

since the other option is unphysical. Combining (3.138) and (3.131) we find

$$\begin{aligned} x_2^* &= \frac{-\rho\left[\dfrac{\left((1+\rho)x_2^* + \rho\right) - \sqrt{\left((1+\rho)x_2^* + \rho\right)^2 + 8x_2^*(x_2^* + \rho)}}{2}\right]}{2x_2^* - (1+\rho)\left[\dfrac{\left((1+\rho)x_2^* + \rho\right) - \sqrt{\left((1+\rho)x_2^* + \rho\right)^2 + 8x_2^*(x_2^* + \rho)}}{2}\right]} \\ &= \frac{-\rho\left[\left((1+\rho)x_2^* + \rho\right) - \sqrt{\left((1+\rho)x_2^* + \rho\right)^2 + 8x_2^*(x_2^* + \rho)}\right]}{4x_2^* - (1+\rho)\left[\left((1+\rho)x_2^* + \rho\right) - \sqrt{\left((1+\rho)x_2^* + \rho\right)^2 + 8x_2^*(x_2^* + \rho)}\right]} \end{aligned} \tag{3.139}$$

which has a positive real solution for $\rho > 0$ corresponding to the single fixed point $(x_1^*, x_2^*, x_3^*) = (x_1^*, x_2^*, x_2^*)$ which is stable for $0 < \rho \lessapprox 2.66$. As it is extremely cumbersome we will not present any further analytical investigation of this system, instead we will focus on using numerical work to give us a better understanding of the long-term dynamics.

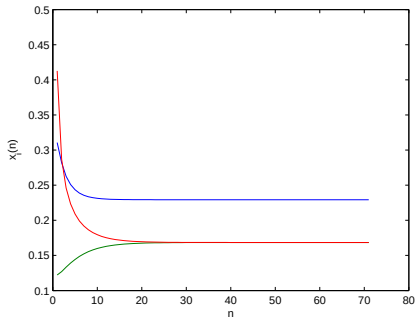


Figure 3.50: Node sizes converging to the fixed point (x_1^*, x_2^*, x_2^*) for $\rho = \frac{1}{2}$ in an incomplete 3-node network. $x_1(n)$ is plotted in blue, $x_2(n)$ is plotted in red and $x_3(n)$ is plotted in green.

4.1.2 Long-Term Relationships

Before looking at explicit bifurcation diagrams let us first examine at the long-term relationship between $x_1(n)$, $x_2(n)$ and $x_3(n)$ for varying values of the parameter ρ.

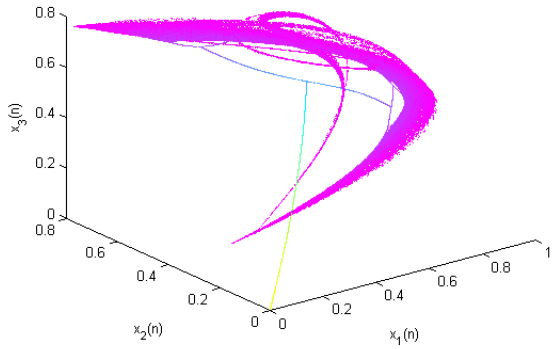


Figure 3.51: Bifurcation Diagram $x_1(n)$ vs $x_2(n)$ vs $x_3(n)$. Colour corresponds to ρ value.

In Figure 3.52 we observe a simple relationship between the long-term values of $x_1(n)$ and $x_2(n)$ for lower values of ρ. As ρ increases this relationship becomes far more interesting. We also see that $x_2(n) = x_3(n)$ as $n \to \infty$. Intuitively, since $r_2(n) = r_3(n)$ and we know for the single logistic map if we fix r all initial conditions in $(0, 1)$ will be absorbed by the same attractor, this is not particularly surprising. This means that the more simple bifurcation diagrams examined later in Section 4.1.3 are sufficient to show the relationship between

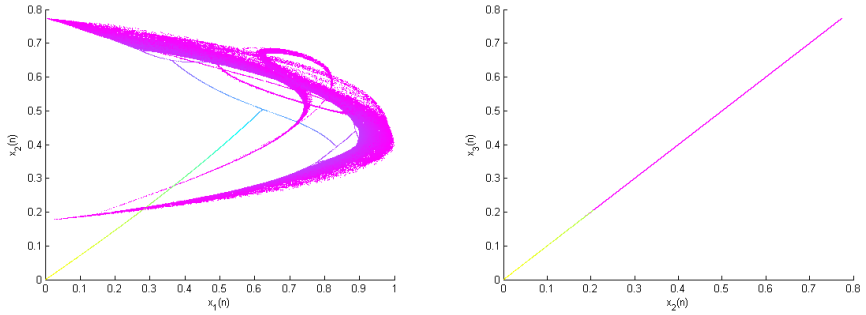


Figure 3.52: Bifurcation Diagram $x_1(n)$ vs $x_2(n)$ and $x_2(n)$ vs $x_3(n)$. Colour corresponds to ρ value.

the value of ρ and the long-term behaviour of the system.

4.1.3 Bifurcation Diagrams

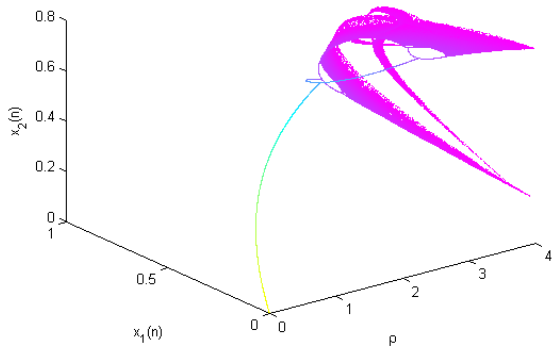


Figure 3.53: Bifurcation diagram for an incomplete 3 node network.

Finding further attractors analytically becomes increasingly messy. Numerically we can see that the initial behaviour of the network logistic map is very similar to the behaviour of the single logistic map. We also observe that in this case, unlike the most simple case of a complete network investigated in Section 3.1, there is a period doubling cascade to chaos as ρ increases. We note however that of the three nodes only the *hub node*, v_1, experiences chaotic behaviour which fills the interval $(0, 1)$. The value of ρ_c, the value for which the fixed point becomes unstable and we see our first period doubling, is also lower than in the corresponding complete three node network case.

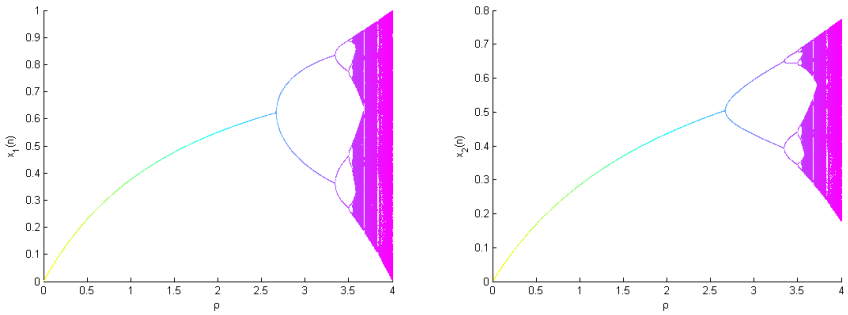


Figure 3.54: Bifurcation diagram projections for $x_1(n)$ and $x_2(n)$.

4.1.4 Lyapunov Exponents

We have again numerically estimated the maximum Lyapunov exponent of the network logistic map for different values of ρ for this incomplete network. We plot the estimates in Figure 3.55. Positive Lyapunov exponents are observed for many values of $\rho > \rho_{\infty_{3I}} \approx 3.54$ indicating chaotic behaviour in this region. However for values of ρ in this range we also observe negative Lyapunov exponents. These negative values correspond to periodic windows interspersed amongst the chaos.

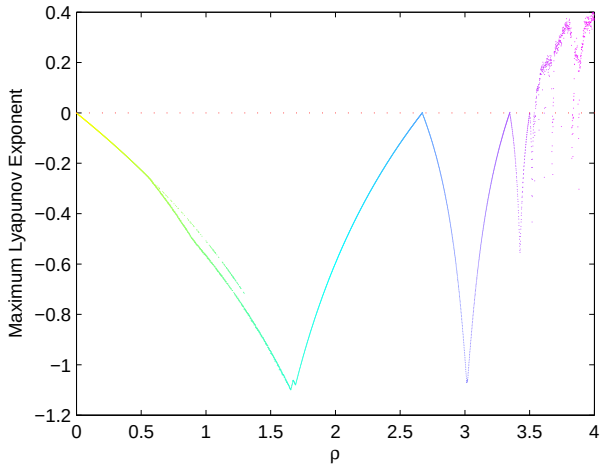


Figure 3.55: Maximum Lyapunov Exponents for an Incomplete 3 Node Network.

4.2 Star Networks

The network described in Section 4.1 is a *star network*. In star networks all nodes are connected to a *central node* or *hub node* by links and these are the

only links present in the network. In this section we will examine star networks for $m = 10$ and $m = 100$. For convenience we will always label the central or hub node v_1. Let us first examine the $m = 10$ case.

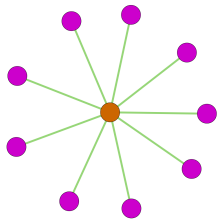
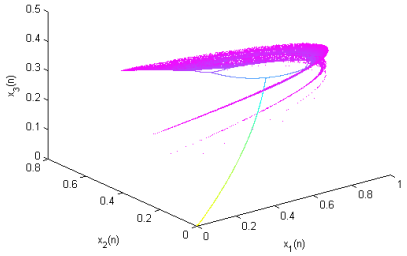
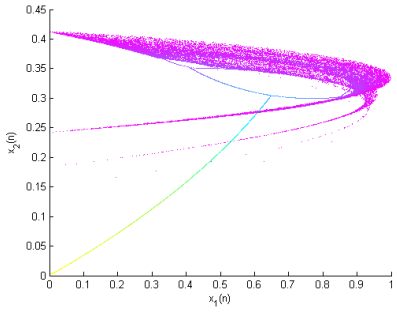


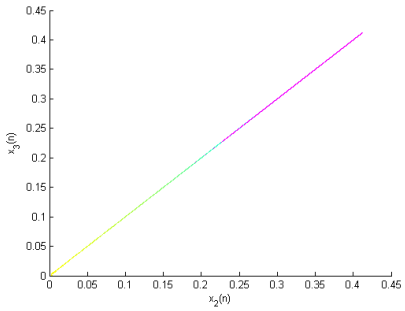
Figure 3.56: Network diagram of a 10 node star network. v_1 is coloured orange while all other nodes are coloured pink.



(a) Long-term behaviour of the relationship between $x_1(n)$, $x_2(n)$ and $x_3(n)$.



(b) $x_1(n)$ vs $x_2(n)$.

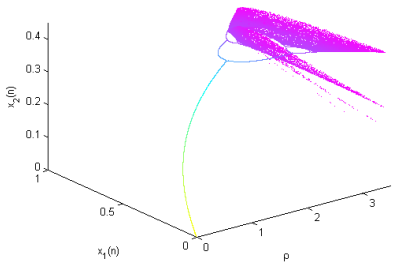


(c) $x_2(n)$ vs $x_3(n)$.

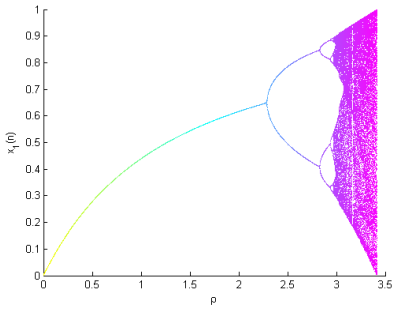
Figure 3.57: Long-term behaviour of the relationship between node sizes in a star network with $m = 10$, colour corresponds to ρ value.

In Figure 3.57 we have plotted the long-term behaviour of the relationship between $x_1(n)$, $x_2(n)$ and $x_3(n)$. Here it is clear that asymptotically $x_2(n) =$

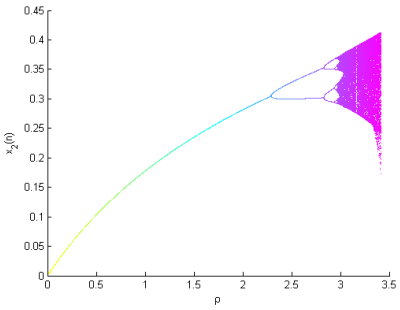
$x_3(n)$ and in fact just as in the case of the three node star network $x_i(n) = x_j(n)$ asymptotically for all $i, j \neq 1$. This is again easy to understand because for all $i, j \neq 1$ we have that $r_i(n) = r_j(n)$ for all values of n. This is true for all star networks of any size if we label the central node v_1.



(a) Bifurcation diagram for $x_1(n)$ and $x_2(n)$.



(b) Bifurcation diagram projection for $x_1(n)$.



(c) Bifurcation diagram projection for $x_2(n)$.

Figure 3.58: Bifurcation diagrams for $x_1(n)$ and $x_2(n)$ in a star network with $m = 10$, colour corresponds to ρ value.

With this knowledge we can get a complete understanding of the long-term dynamics of the system using the bifurcation diagrams in Figure 3.58. Note that v_1 is the dominant node in the network, growing to a larger steady-state size than the other two nodes for $\rho \lessapprox 2.3$ and having a wider range of oscillation for the range of values of ρ for which the system undergoes periodic behaviour. We also see that the chaotic orbit of $x_1(n)$ is wider than the corresponding chaotic orbits of $x_2(n)$ and $x_3(n)$. This node seems to drive the dynamics of the overall system. Its behaviour looks quite similar to that of a single logistic map period doubling to chaos which eventually fills the interval $(0, 1)$. All other nodes also experience a period-doubling cascade to chaos, however, their steady-state values are lower for each value of ρ before the first bifurcation and their chaotic attractor does not fill the interval $(0, 1)$ for any value of ρ.

Also note that the value for ρ_c, the value at which the fixed point becomes unstable and a stable period-two attractor appears is greater than 2.25. This is in turn greater than the corresponding value in a complete 10-node network as analysed in Section 3.2.

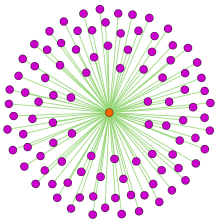
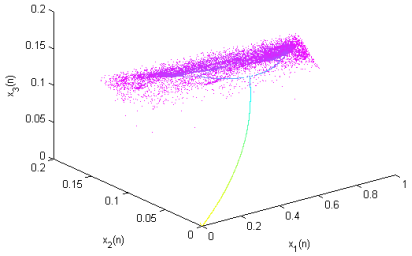
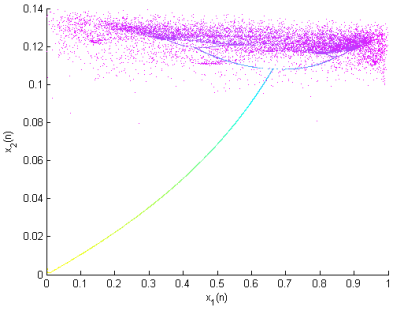


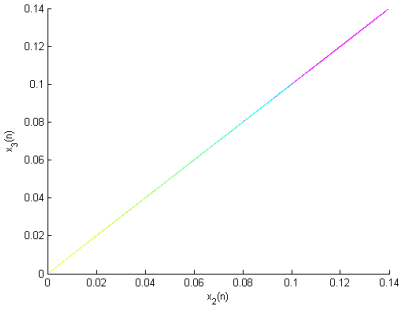
Figure 3.59: Network diagram of a 100 node star network. v_1 is coloured orange while all other nodes are coloured pink.



(a) Long-term behaviour of the relationship between $x_1(n)$, $x_2(n)$ and $x_3(n)$.



(b) $x_1(n)$ vs $x_2(n)$.

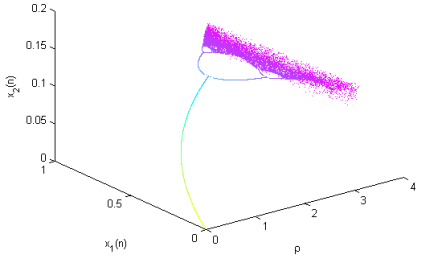


(c) $x_2(n)$ vs $x_3(n)$.

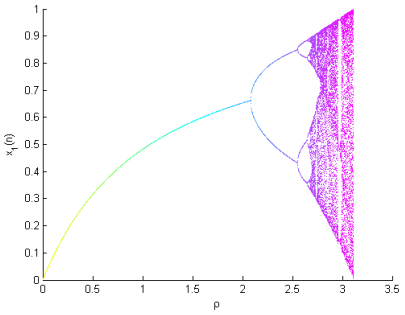
Figure 3.60: Long-term behaviour of the relationship between node sizes in a star network with $m = 100$, colour corresponds to ρ value.

Now let us consider the $m = 100$ node example. Again the most interesting long-term relationship between the $x_i(n)$s is between $x_1(n)$ and any other $x_j(n)$. This is because as before $x_i(n) = x_j(n)$ asympotically for all $i, j \neq 1$. this can be seen in Figure 3.60.

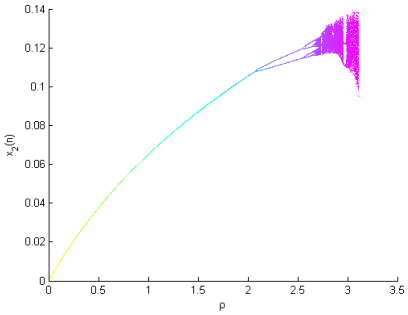
With this knowledge we can get a complete understanding of the long-term dynamics of the system using the bifurcation diagrams in Figure 3.61. Again, as in the ten-node case we note that $x_1(n)$ is the dominant node size having larger steady-state values and wider periodic oscillation ranges. Its behaviour looks quite similar to that of a single logistic map period doubling to chaos which eventually fills the interval $(0, 1)$. All other nodes also experience a period-doubling cascade to chaos, however, their steady-state values are lower for each value of ρ before the first bifurcation and their chaotic attractor does not fill the interval $(0, 1)$ for any value of ρ.



(a) Bifurcation diagram for $x_1(n)$ and $x_2(n)$.



(b) Bifurcation diagram projection for $x_1(n)$.



(c) Bifurcation diagram projection for $x_2(n)$.

Figure 3.61: Bifurcation diagrams for $x_1(n)$ and $x_2(n)$ in a star network with $m = 100$, colour corresponds to ρ value.

Also note that the value for ρ_c, the value at which the fixed point becomes unstable and a stable period-two attractor appears is greater than 2.05. This is in turn greater than the corresponding value in a complete 100 node network

as analysed in Sections 3.2 and 3.5.1.

4.3 More Incomplete Networks

In this section we look at the behaviour of the network logistic map for a network with 100 nodes generated using the algorithm described in Chapter 2. We number the nodes by the order in which they were added to the network by our algorithm. Due to the large size of these systems it is necessary to focus most of our in analysis on a small portion of the nodes. We analyse three examples and focus on the long-term behaviour of 3 separate nodes in each case. For both networks we will look in detail at the behaviour of $x_1(n)$, the size of the original node from which the network was grown, as this node is naturally very important in the system. Our criteria for choosing the other two nodes will be based on their *centrality*. We will choose the two nodes with the highest and lowest centrality in the network after v_1.

4.3.1 Centrality

Centrality measures allow us to quantify the importance of each node to the overall network structure. There are of course many different ways in which a node could be considered important to the network and each may lead to a separate centrality measure. The most simple centrality measure is degree centrality, in this measure the centrality of each node is simply its degree.

We use a more sophisticated centrality measure popularly known as *PageRank*, which is the name that was given to it by the internet web search company Google [52]. This centrality measure is a variation of *Katz centrality*. Katz centrality allows nodes to derive centrality not only from its own degree, but also from its neighbours. In PageRank centrality the centrality a node derives from its neighbours is proportional to their centrality divided by their degree.

Definition 4.3.1. The PageRank centrality of node v_i is given by

$$c_i^{PR} = \alpha \sum_{j:v_j \in N} A_{ij} \frac{c_j^{PR}}{d_j} + \beta \tag{3.140}$$

where d_j is the degree of node v_j, A is the adjacency matrix of the network and α, β are constants.

We choose 0.85 as the value of our free parameter α, the value used by Google [52]. By convention we take $\beta = 1$,

4.3.2 Example 1 - Bifurcation Diagrams

The first network we analyse, N_1, is shown in Figure 3.62. This network has a density of 0.1032 and an average clustering coefficient of 0.1563. The average shortest path length between any two nodes in the network is 2.3388 and the mean degree of an individual node is 10.2200. Recall that both the average

clustering coefficient and average shortest path length of a network are defined in Chapter 2. The node with the highest centrality, 14.7284, is v_2, the node with the lowest centrality, 1.6170, is v_{96}. v_2 is the node with the minimum shortest path length to v_1 while v_{96} has the maximum shortest path length to v_1.

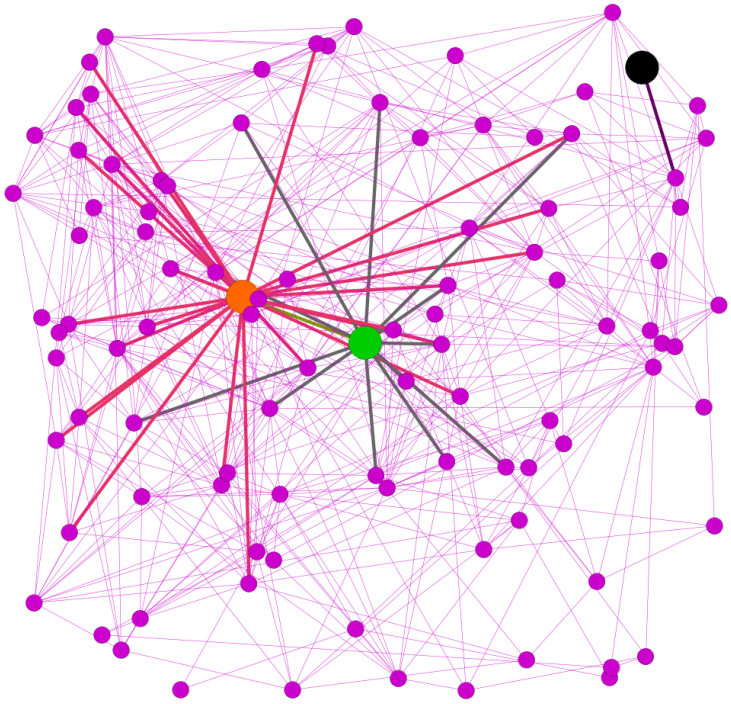
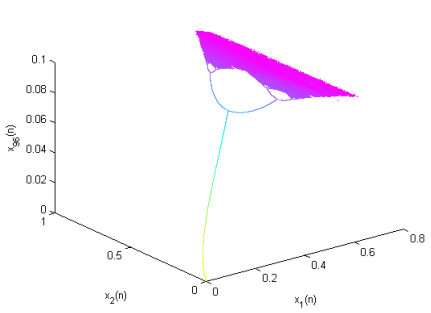
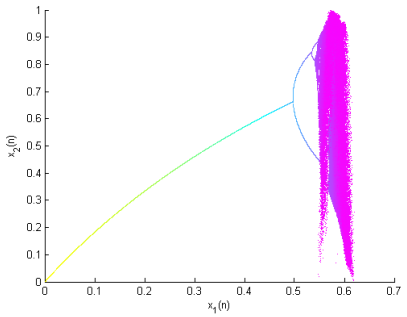


Figure 3.62: Network Diagram of N_1 highlighting the three nodes of interest and the edges attached to those nodes. In the diagram v_1 is coloured green, v_2 is coloured orange and v_{96} is coloured black.

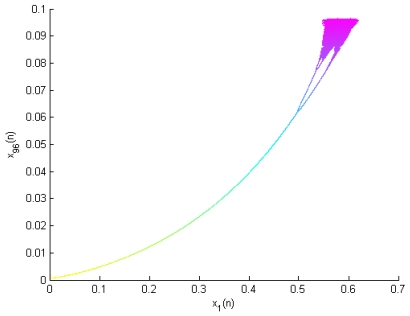
In Figure 3.63 we look in detail at the long-term relationships between the node sizes $x_1(n)$, $x_2(n)$ and $x_{96}(n)$. The only edge between these three nodes connects v_1 to v_2, the shortest path length between both other pairs of nodes is 3. We see that the long-term relationships between the node sizes are far from trivial and vary between each pair of nodes. We also note that the direct link between v_1 and v_2 does not necessarily imply a simpler relationship between the long-term behaviour of their sizes under the network logistic map. Before the switch to period-two behaviour we see that the fixed point for nodes with higher centrality is higher.



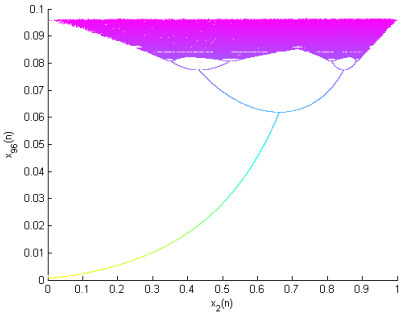
(a) Long-term behaviour of the relationship between $x_1(n)$, $x_2(n)$ and $x_{96}(n)$.



(b) $x_1(n)$ vs $x_2(n)$.



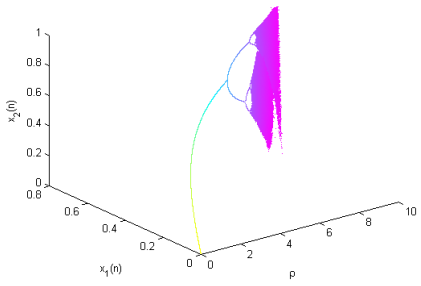
(c) $x_1(n)$ vs $x_{96}(n)$.



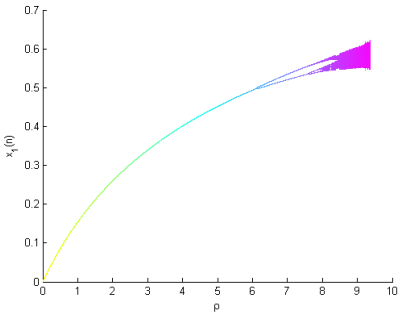
(d) $x_2(n)$ vs $x_{96}(n)$.

Figure 3.63: Long-term behaviour of the relationship between node sizes in a random network with $m = 100$, colour corresponds to ρ value.

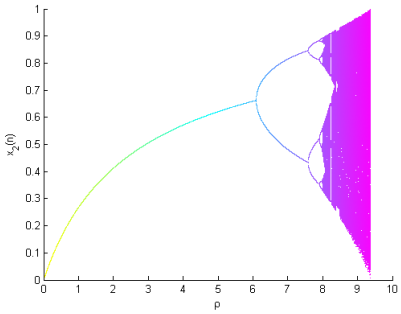
In Figure 3.64 we examine bifurcation diagrams for $x_1(n)$ and $x_2(n)$. The shortest path length between these two nodes is 1, i.e. there is a direct link between the two, v_1 has degree 13 while v_2 has degree 26. We see that $x_2(n)$, the node size of the node with the maximum centrality in N_1 has the higher steady-state node size for all values of ρ before the first bifurcation at the critical value of ρ, $\rho_c \approx 6.1$. After this first period-doubling bifurcation both $x_1(n)$ and $x_2(n)$ undergo a period doubling cascade to chaos. While the behaviour of $x_2(n)$ looks quite similar to the behaviour of the single logistic map the dynamics of $x_1(n)$ appear *'squashed'*. We hypothesise that this occurs because in the case of the node with high centrality, v_2, the main cause of the period doubling and chaotic dynamics experienced as ρ increases is the magnitude of its corresponding $r_2(n)$. In the case of the single logistic map an increase in the control parameter is also what pushes the system to chaos. On the other hand we believe the main cause of the period doubling and chaotic dynamics of $x_1(n)$ is not the magnitude of $r_1(n)$ but rather the chaotic behaviour of $r_1(n)$ itself, which is driven in large by the changes in $x_2(n)$.



(a) Bifurcation Diagram for $x_1(n)$ and $x_2(n)$.

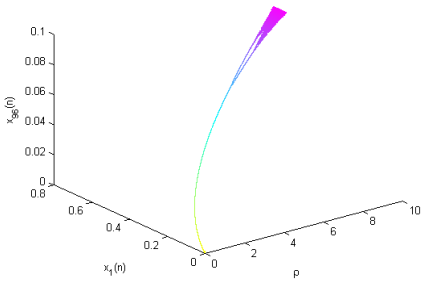


(b) Bifurcation Diagram projection for $x_1(n)$.

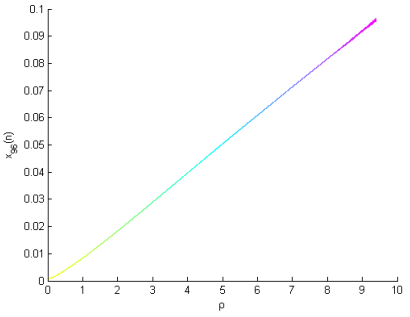


(c) Bifurcation Diagram projection for $x_2(n)$.

Figure 3.64: Bifurcation diagrams for $x_1(n)$ and $x_2(n)$.



(a) Bifurcation Diagram for $x_1(n)$ and $x_{96}(n)$.



(b) Bifurcation Diagram projection for $x_{96}(n)$.

Figure 3.65: Bifurcation diagrams for $x_1(n)$ and $x_{96}(n)$.

In Figure 3.65 we have given bifurcation diagrams for $x_1(n)$ and $x_{96}(n)$.

The shortest path length between these two nodes is 3, v_{96} has degree 1. We again see that node with lower centrality has a lower fixed point node size value for values of ρ below ρ_c and 'squashed' dynamics thereafter. The squashing of the dynamics is even more pronounced in this case as v_{96} is the node with the minimum centrality in N_1.

In Figure 3.66 we have plotted the bifurcation diagram for the mean of all node sizes in N_1. We have also plotted the bifurcation diagram for a single logistic map given by

$$x(n+1) = \left(1 + \rho\left(\sum_{i=1}^{100} \frac{d_i}{100^2}\right)\right) x(n)(1 - x(n)) \tag{3.141}$$

where d_i is the degree of v_i. We can see that this map can give us a very crude estimate of the mean node size of the network. This is quite a natural idea, in the network logistic map $r_i(n) = 1 + \rho w_i(n)$ where $w_i(n)$ is the normalised weighted degree of v_i. Now in order to estimate the mean of all $x_i(n)$ we are replacing the normalised weighted degree of $v_i(n)$ with the mean normalised *unweighted* degree over the whole network.

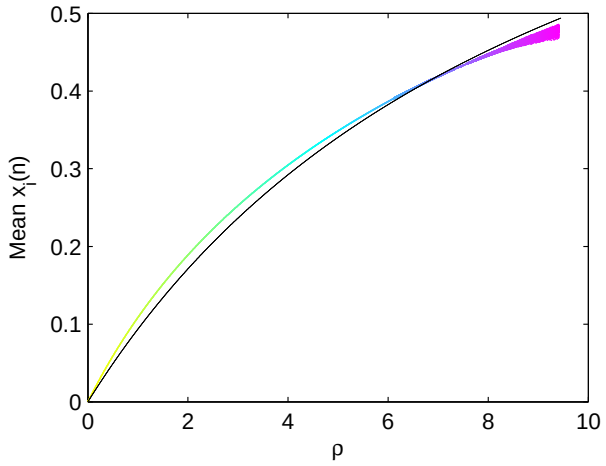


Figure 3.66: Bifurcation diagram for the size of an average node in N_1 in colour, bifurcation diagram for a single logistic map with $r = 1 + \rho\left(\sum_{i=1}^{100} \frac{d_i}{100^2}\right)$ in black.

4.3.3 Example 2 - Bifurcation Diagrams

Our second network, N_2, is shown in Figure 3.67. This network has a density of 0.1515 and an average clustering coefficient of 0.2612. The average shortest path length between any two nodes in the network is 2.0473 and the mean degree of an individual node is 15. The node with the highest centrality, 11.9441, is v_{33}, the node with the lowest centrality, 1.7125, is v_{87}. The shortest path length between any two of these nodes is the same.

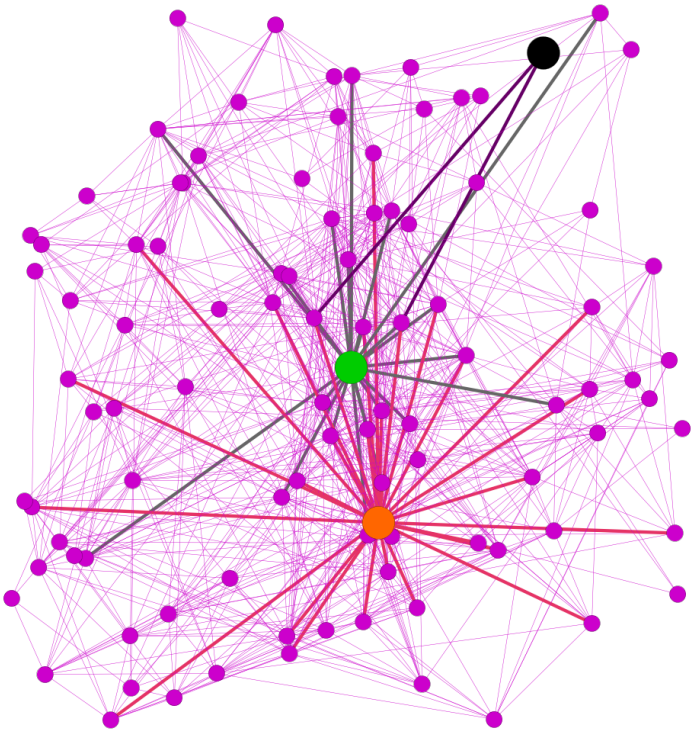
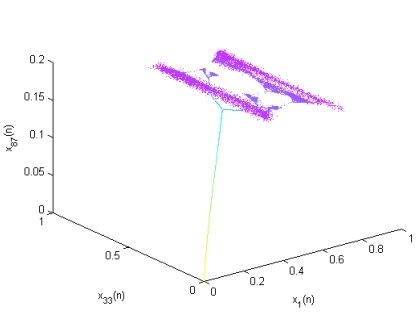
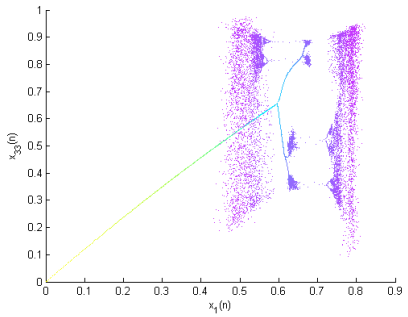


Figure 3.67: Network Diagram of N_2 highlighting the three nodes of interest and the edges attached to those nodes. In the diagram v_1 is coloured green, v_{33} is coloured orange and v_{87} is coloured black.

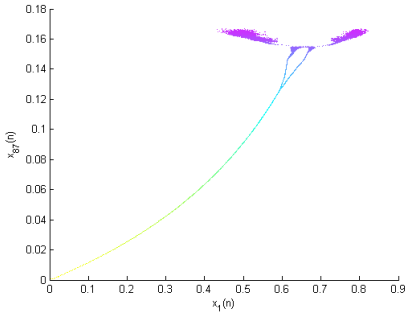
In Figure 3.68 we look in detail at the long-term relationships between the node sizes $x_1(n)$, $x_{33}(n)$ and $x_{87}(n)$. The shortest path length between these any pair of these nodes is 2. We see that the long-term relationships between the node sizes are are again complex in nature. We note that, just as in the case of N_1, before the switch to period-two behaviour the fixed point for nodes with higher centrality is higher.



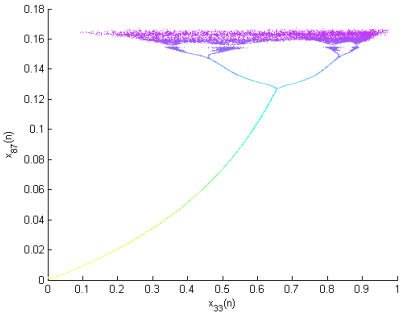
(a) Long-term behaviour of the relationship between $x_1(n)$, $x_{33}(n)$ and $x_{87}(n)$.



(b) Long-term behaviour of the relationship between $x_1(n)$ and $x_{33}(n)$.



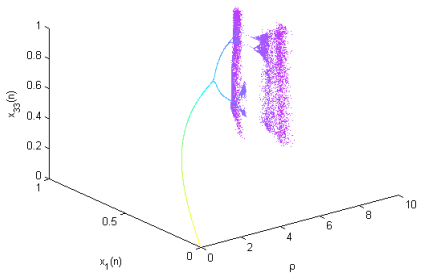
(c) Long-term behaviour of the relationship between $x_1(n)$ and $x_{87}(n)$.



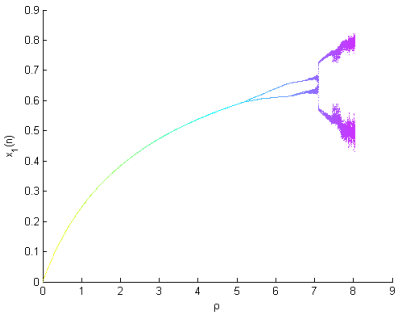
(d) Long-term behaviour of the relationship between $x_{33}(n)$ and $x_{87}(n)$.

Figure 3.68: Long-term behaviour of the relationship between node sizes in a random network with $m = 100$, colour corresponds to ρ value.

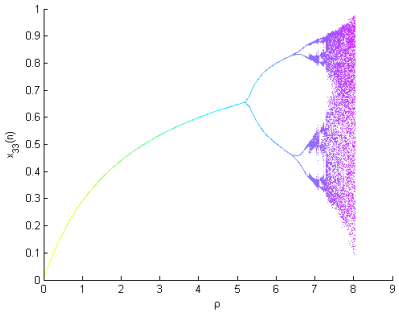
In Figure 3.69 we examine bifurcation diagrams for $x_1(n)$ and $x_{33}(n)$. The shortest path length between these two nodes is 2, v_1 has degree 22 while v_{33} has degree 30. Again, we see that $x_{33}(n)$, the node size of the node with the maximum centrality in N_2 has the higher steady-state node size for all values of ρ before the first bifurcation at the critical value of ρ, $\rho_c \approx 5.2$. We also see that the dynamics of $x_1(n)$ appear 'squashed' once more. This supports our hypothesis that qualitative changes in the long-term behaviour of node size is driven in large part by the magnitude of $r_i(n)$ for nodes v_i with high centrality while the largest cause for changes in behaviour of nodes v_j with low centrality is the changes in behaviour of the nodes' $r_j(n)$.



(a) Bifurcation diagram for $x_1(n)$ and $x_{33}(n)$.

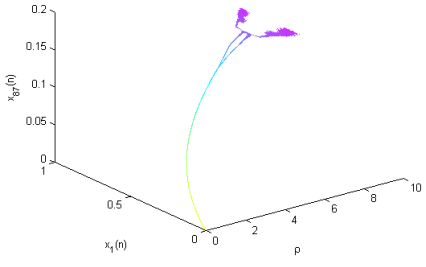


(b) Bifurcation diagram projection for $x_1(n)$.

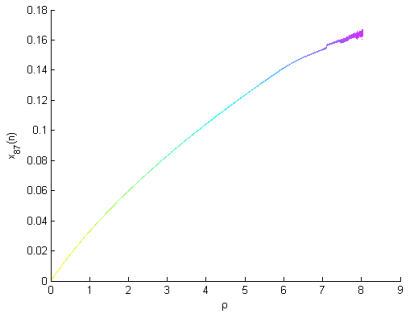


(c) Bifurcation diagram projection for $x_{33}(n)$.

Figure 3.69: Bifurcation diagrams for $x_1(n)$ and $x_{33}(n)$, colour corresponds to ρ value.



(a) Bifurcation diagram for $x_1(n)$ and $x_{87}(n)$.



(b) Bifurcation diagram projection for $x_{87}(n)$.

Figure 3.70: Bifurcation diagrams for $x_1(n)$ and $x_{87}(n)$, colour corresponds to ρ value.

In Figure 3.70 look at bifurcation diagrams for $x_1(n)$ and $x_{87}(n)$. The shortest path length between these two nodes is 2, v_{87} has degree 2. We again see that the node with lower centrality has a lower fixed point node size value for values of ρ below ρ_c and 'squashed' dynamics thereafter. The squashing of the dynamics is even more pronounced for v_{87} than it was for v_{33}. We argue that this is because as v_{87} is the node with the minimum centrality in N_2 its bifurcations after this point are driven almost entirely by the behaviour, rather than the magnitude, of $r_{87}(n)$.

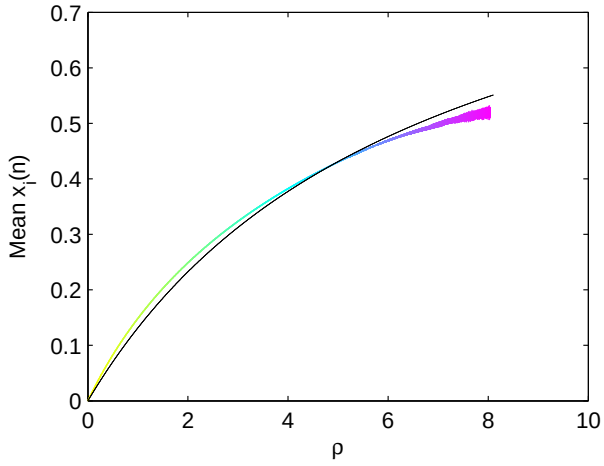


Figure 3.71: Bifurcation diagram for the size of an average node in N_2 in colour, bifurcation diagram for a single logistic map with $r = 1 + \rho\left(\sum_{i=1}^{100} \frac{d_i}{100^2}\right)$ in black.

Finally, in Figure 3.71 we again see that the map given by $x(n+1) = \left(1 + \rho\left(\sum_{i=1}^{100} \frac{d_i}{100^2}\right)\right) x(n)(1 - x(n))$ can give us a crude estimate of the long-term dynamics of the mean node size in N_2.

4.3.4 Example 3 - Bifurcation Diagrams

Our third network, N_3, is shown in Figure 3.72. N_3 has a density of 0.0978 and an average clustering coefficient of 0.3115. The average shortest path length between any two nodes in the network is 3.0240 and the mean degree of an individual node is 9.68. The node with the highest centrality, 12.2328, is v_{19}, the node with the lowest centrality, 2.2098, is v_{70}. There are no links directly connecting any two of these nodes.

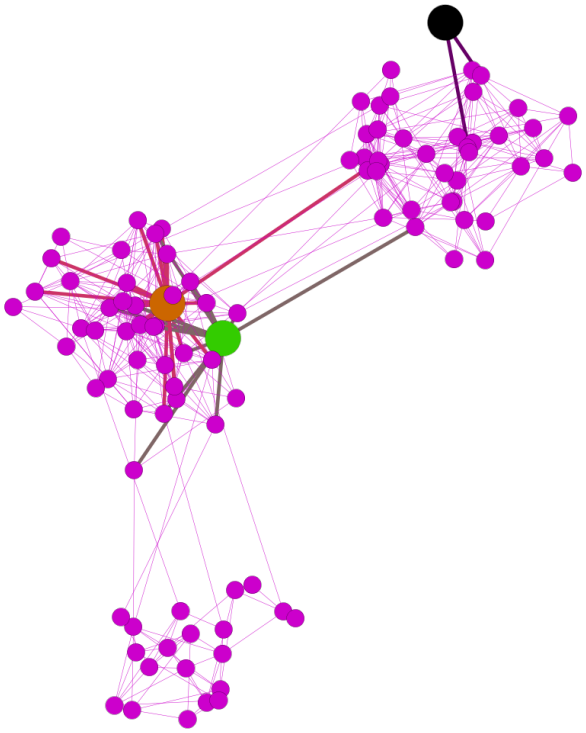
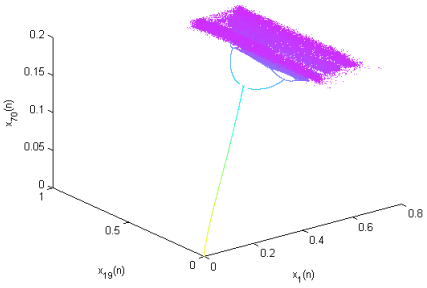
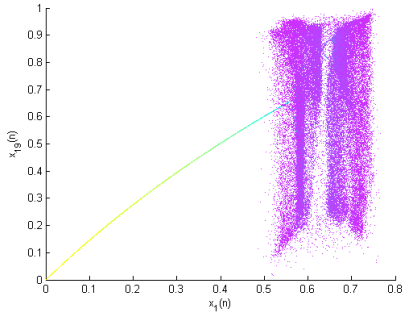


Figure 3.72: Network Diagram of N_3 highlighting the three nodes of interest and the edges attached to those nodes. In the diagram v_1 is coloured green, v_{19} is coloured orange and v_{70} is coloured black.

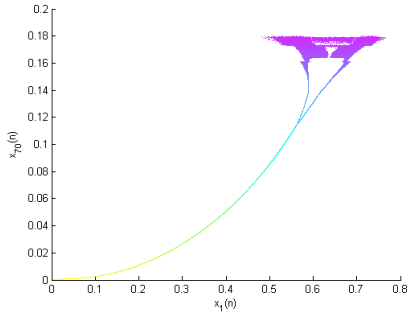
In Figure 3.73 we look in detail at the long-term relationships between the node sizes $x_1(n)$, $x_{19}(n)$ and $x_{70}(n)$. The minimum shortest path length between any two of these three nodes is 2, between v_1 and v_{19}. The shortest path length between both other pairs is 4. We see that the relationships between the node sizes are complex, especially after the first bifurcation. We also note that the shorter path length between v_1 and v_{19} does not result in a simpler relationship between the long-term behaviour of their sizes under the network logistic map. Before the switch to period-two behaviour we see that the fixed point for nodes with higher centrality is higher.



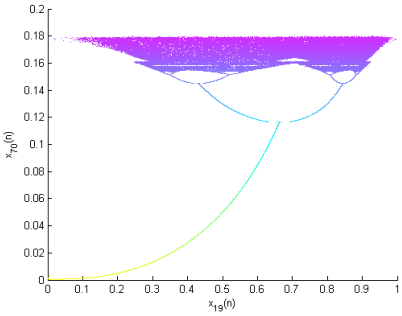
(a) Long-term behaviour of the relationship between $x_1(n)$, $x_{19}(n)$ and $x_{70}(n)$.



(b) Long-term behaviour of the relationship between $x_1(n)$ and $x_{19}(n)$.



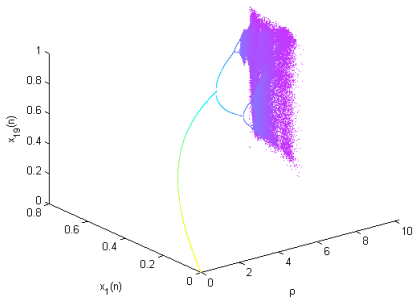
(c) Long-term behaviour of the relationship between $x_1(n)$ and $x_{70}(n)$.



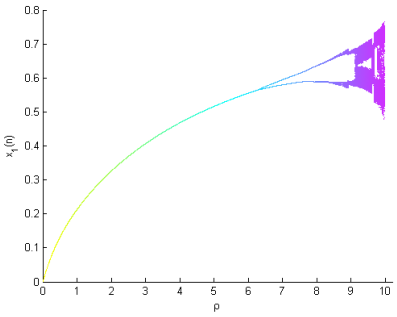
(d) Long-term behaviour of the relationship between $x_{19}(n)$ and $x_{70}(n)$.

Figure 3.73: Long-term behaviour of the relationship between node sizes in a random network with $m = 100$, colour corresponds to ρ value.

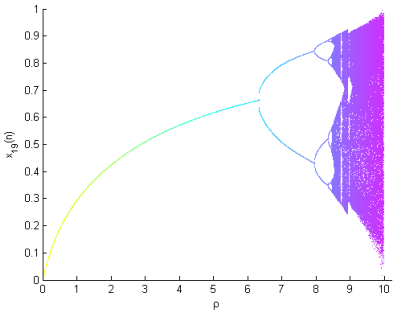
Figure 3.74 shows bifurcation diagrams for $x_1(n)$ and $x_{19}(n)$. The shortest path length between these two nodes is 2, v_1 has degree 15 while v_{19} has degree 22. Figure 3.75 shows bifurcation diagrams for $x_1(n)$ and $x_{70}(n)$. The shortest path length between these two nodes is 4, v_{70} has degree 2. Despite the significantly different network topology of N_3 we see that the striking features observed in N_1 and N_2 are also present here. The node with the highest centrality, v_{19}, dominates. It has the highest fixed point value for all values of $\rho < \rho_c \approx 6.3$ and the widest dynamics for values of $\rho > \rho_c$. The node with the lowest centrality v_{70} has dynamics so 'squashed' beyond ρ_c that the period-two dynamics are almost invisible to the naked eye.



(a) Bifurcation diagram for $x_1(n)$ and $x_{19}(n)$.

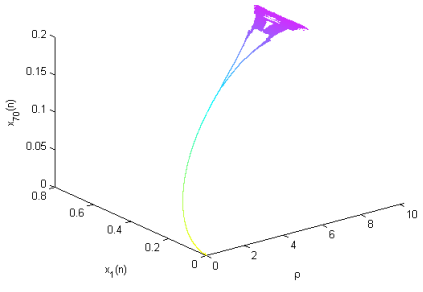


(b) Bifurcation diagram projection for $x_1(n)$.

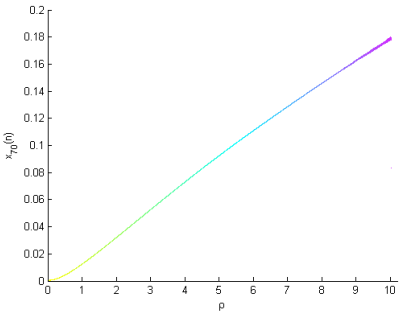


(c) Bifurcation diagram projection for $x_{19}(n)$.

Figure 3.74: Bifurcation diagrams for $x_1(n)$ and $x_{19}(n)$, colour corresponds to ρ value.



(a) Bifurcation diagram for $x_1(n)$ and $x_{70}(n)$.



(b) Bifurcation diagram projection for $x_{70}(n)$.

Figure 3.75: Bifurcation diagrams for $x_1(n)$ and $x_{70}(n)$, colour corresponds to ρ value.

Finally, in Figure 3.76 we once more observe that the map given by $x(n+1) = \left(1+\rho\left(\sum_{i=1}^{100}\frac{d_i}{100^2}\right)\right)x(n)(1-x(n))$ can once more give us a crude estimate of the long-term dynamics of the mean node size in N_3.

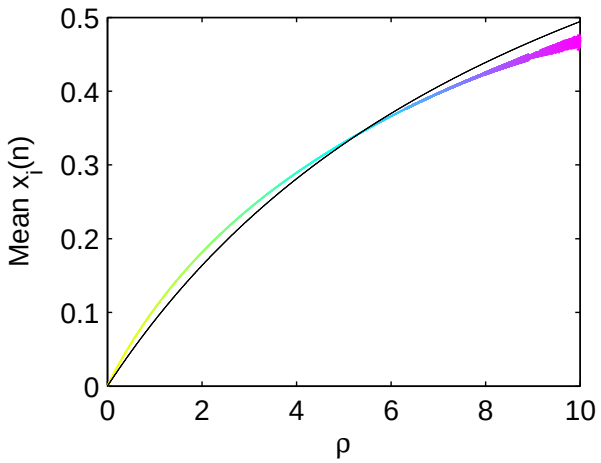


Figure 3.76: Bifurcation diagram for the size of an average node in N_3 in colour, bifurcation diagram for a single logistic map with $r = 1+\rho\left(\sum_{i=1}^{100}\frac{d_i}{100^2}\right)$ in black.

5 A Lower Bound for ρ_c

Of particular interest to us is the critical value ρ_c for any network. This is the value above which our fixed point becomes unstable and the dynamics of the entire system become far more complicated. In this section we hypothesise that there exists a lower bound for ρ_c. This will be useful to us as later when applying this type of node growth in our model. Knowledge of this bound will allow us to bound our parameter ρ to ensure realistic model dynamics.

Note that for any network $N = (V, E)$

$$w_i(n) = \frac{\sum_{j:(x_i,x_j)\in E} x_j(n)}{\sum_{j=1}^{m} x_j(n)} \leq 1. \tag{3.142}$$

Using this fact we have that for all i and n:

$$r_i(n) = 1 + \rho w_i(n) \leq 1 + \rho. \tag{3.143}$$

We argue that a switch to period-two behaviour in the system begins at lastest when any node v_i has a corresponding $r_i(n)$ with a steady-state $r_i(n) > 3$. This is of course the point at which period doubling begins in the single logistic map.

In fact, this is exactly the steady-state value of $r_i(n)$ for which this bifurcation occurs that is approached in the limiting case of a complete network where $m \to \infty$. In this case ρ_c approaches 2 as argued in Section 3.2.1. We also have that since for a complete network with m nodes

$$\rho_c = \frac{m^2 - 2m + \sqrt{12m^2 - 20m^3 + 9m^4}}{2(m-1)^2} > 2, \tag{3.144}$$

where 2 represents a lower bound for ρ_c in the case of a complete network. We argue that this bound is in fact satisfied by all networks. A complete network maximises connectivity and hence steady-state node size and the level of coupling in the system. This forces the bifurcation to occur earliest in these networks. In Figure 3.77 we have plotted the bifurcation diagrams for typical nodes in twenty 20-node networks with randomly generated adjacency matrices and we can see none of the systems switch to period-two behaviour before a complete 20-node network or in particular before ρ exceeds our proposed lower bound.

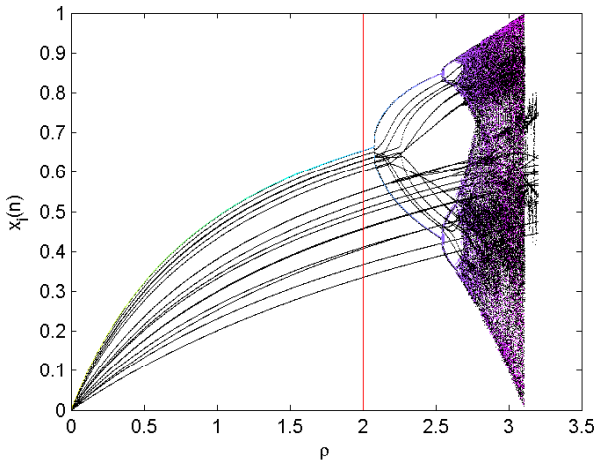


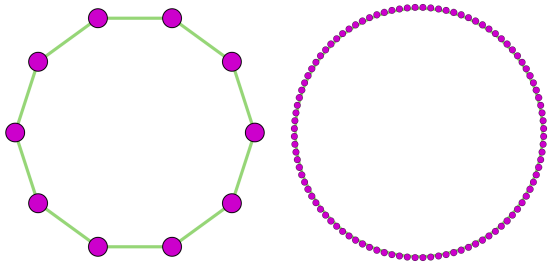
Figure 3.77: Bifurcation diagram projections for a typical node in twenty 20-node networks with randomly generated adjacency matrices in black, bifurcation diagram projection for a typical node in 20-node complete network highlighted in colour showing ρ_c is lowest for this network. Our proposed lower bound for ρ_c in red.

For less dense networks, like those examined in Section 4.3 ρ_c is in fact often significantly larger than this hypothesised lower bound. This can be clearly seen in Figure 3.64 for example where $\rho_c > 6$ for a network with a density of 0.1032. In this section we also saw evidence that the node with the highest centrality in the system drives the system to undergo period-doubling bifurcations. This is further reason to believe that a network in which every node has maximised its possible centrality by maximising its connections period doubling will occur earlier than in any other network with the same number of nodes. Finally, we

saw in Section 4.2 that a star network with m nodes has a higher value of ρ_c than the corresponding complete network.

5.1 Networks With Large ρ_c

To further illustrate this let us look at two examples of *chain networks*. In these networks $E = \{(v_i, v_j) : i = j+1 \mod m\}$, so for example v_1 is connected to v_2 and v_m, v_2 is connected to v_3 and v_1 and so on. These networks have both low density and low maximum degree, we would therefore expect these networks to have a large value for ρ_c.

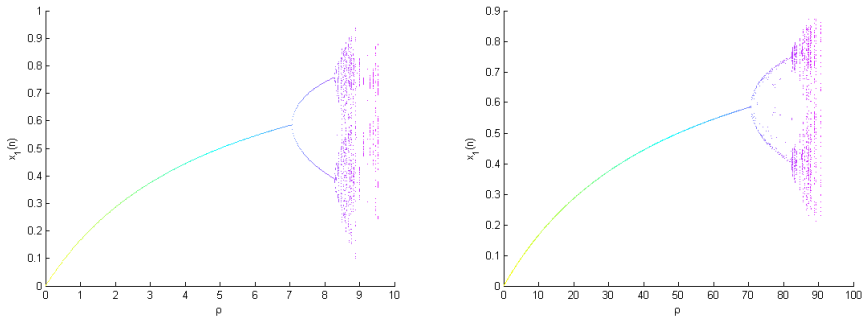


(a) A chain network with $m = 10$ nodes.

(b) A chain network with $m = 100$ nodes.

Figure 3.78: Network diagrams for two chain networks.

Examining the bifurcation diagrams of these networks we see that this is indeed the case. In Figure 3.79 we see that the $\rho_c \approx 6.87$ for $m = 10$ and $\rho_c \approx 70.35$ for $m = 100$. These values are significantly higher than our proposed lower bound 2 and also the corresponding values for a complete network given by $\frac{10}{81}(4 + \sqrt{178}) \approx 2.14$ and $\frac{100}{9801}(49 + \sqrt{22003}) \approx 2.01$ respectively.



(a) Bifurcation diagram projection for x_1 for a chain network with $m = 10$ nodes.

(b) Bifurcation diagram projection for x_1 for a chain network with $m = 100$ nodes.

Figure 3.79: Bifurcation diagrams for our two chain networks.

6 A New Centrality Measure?

In Section 4.3.1 we introduced the concept of a centrality measure. The main goal of any centrality measure is to quantify in some way the importance of each node in a network. Identifying the most important node in a network can be extremely useful. Many centrality measures have their origins in social network theory, however, there are applications in many fields, from biology to computer science and finance. Knowledge of the most important node in a system is useful to search engines such as *Google* as it allows them to rank their pages, it is useful in controlling power grids, epidemics & the dissemination of information, protecting systems such as computer networks from attack and identifying important proteins.

6.1 Logistic Centrality

We propose that the network logistic map of a network, N, can itself provide us with a new centrality measure. If we restrict our view to the range of values of ρ for which N has a single stable fixed point we have seen throughout our analysis in Chapter 3 that the more *important* nodes have always had the largest steady-state value. In Section 3 we proved that for a complete network, a network where all nodes are of equal importance, the steady-state value for each node under the network logistic map is equal. It is easy to show numerically that this is also the case for the chain networks examined in Section 5.1. In Section 4.1 and Section 4.2 we saw that the hub node of a star network has a high steady-state value while all other nodes have an equal lower steady-state value. In Section 4.3 we repeatedly observed the node with the highest PageRank centrality of the three nodes examined having the highest steady-state value and also driving the dynamics. We use this as motivation for defining a new centrality measure. We refer to this new centrality measure as *logistic centrality* and define it as follows.

Definition 6.1.1. The logistic centrality of any node v in a network N is given by the steady-state node size of v under the network logistic map of N with parameter $\rho = 1$.

6.2 Comparing Logistic Centrality to Established Measures

We now examine and compare this centrality measure to three well established centrality measures, *degree centrality*, *Katz centrality* and *PageRank centrality*, in the context of the three 100 node networks N_1, N_2 and N_3 investigated in Section 4.3. Recall that PageRank centrality is a variation of Katz centrality in which a node derives centrality from its neighbours proportional to their centrality divided by their degree and was formally defined in Definition 4.3.1. We define degree centrality and Katz centrality as follows.

Definition 6.2.1. The degree centrality of any node v in a network N is the degree of v, the number of links connected to v.

Definition 6.2.2. The Katz centrality of node v_i in a network N is given by

$$c_i^K = \alpha \sum_{j:v_j \in N} A_{ij} c_j^K + \beta, \tag{3.145}$$

where A is the adjacency matrix of N and α, β are constants. We choose $\alpha = 0.95\kappa^{-1}$ where κ is the largest eigenvalue of A. By convention we take $\beta = 1$.

The purpose of any centrality measure is to quantify the importance of any node to the network it is a part of. As such, it is the ranking of nodes under any centrality measure that is important and not the value given for the centrality of of any one node by the measure. Keeping this in mind, we examine the differences and similarities between the ranking of nodes in N_1, N_2 and N_3 under logistic centrality and our established centrality measures.

First, in Table 3.1 we have outlined for which networks our three established centrality measures agree with our new centrality measure, logistic centrality, in picking the most and least important nodes in each network. In many cases these two nodes may be the most important to identify. Where the measures do not agree with our measure we have provided the alternative ranking (in order of increasing centrality) for the node given by the established measures in parentheses. We see that all four measures agree in many cases but there are differences. Where the centrality measures differ we see that for the most part the difference in the ranking given is not large. Notably, the Katz centrality measure and our new measure of centrality agree on the ranking of the most and least important nodes for all three networks except in one case. The node in N_3 ranked least important by logistic centrality is ranked the third least important by Katz centrality. It is in ranking the node of minimum centrality for this network that we see most disagreement across all four centrality measures. We also see that Katz centrality is closest to agreement with logistic centrality in this case.

	Node of Maximum Centrality			**Node of Minimum Centrality**		
	Degree	***Katz***	***PageRank***	***Degree***	***Katz***	***PageRank***
N_1	✓	✓	✓	✗ (3)	✓	✓
N_2	✗ (99)	✓	✗ (99)	✗ (2)	✓	✗ (3)
N_3	✓	✓	✓	✗ (16)	✗ (3)	✗ (23)

Table 3.1: Agreement of established centrality measures with logistic centrality in choosing the most and least important nodes for the three 100-node networks, N_1, N_2 and N_3, described in Section 4.3. Where the established measures do not agree with logistic centrality the alternative ranking (in order of increasing centrality) for the node given by the established measures is given in parentheses.

In Table 3.2 we provide the mean ranking difference between logistic cen-

trality and the three established measures, Δ, given by

$$\Delta = \sum_{i=1}^{m} \frac{\left|R_i^L - R_i^O\right|}{m}, \tag{3.146}$$

where R_i^L is the ranking from 1 to m of v_i in order of increasing logistic centrality and R_i^O is the ranking from 1 to m of v_i in order of increasing centrality under an alternative established centrality measure. Again we see the Katz centrality measure shows the most agreement with logistic centrality for all three networks. PageRank centrality shows the largest difference. In fact, although we have not shown mean ranking differences between the pairs of established centrality measures here, the mean ranking difference between the Katz centrality measure and the logistic centrality measure is the smallest of any pair for all three networks.

Network	**Mean Ranking Difference, Δ**		
	Degree	***Katz***	***PageRank***
N_1	7.16	0.82	8.98
N_2	4.56	0.72	5.46
N_3	13.02	2.64	19.98

Table 3.2: Mean ranking difference across all nodes between our established centrality measures and logistic centrality for the three 100-node networks, N_1, N_2 and N_3, described in Section 4.3.

Evidence of the similarity in the ranking of nodes given by logistic centrality and Katz centrality presented in Table 3.1 and Table 3.2 is not entirely unexpected. Katz centrality gives a node, v_i, in a network, N centrality based on the number of neighbours it has and the centrality of those neighbours. Logistic centrality allocates centrality to nodes in much the same way. The logistic centrality of a node, v_i, in a network, N, is its steady-state size under the network logistic map of N with $\rho = 1$. This steady-state size is higher for a higher steady-state value of $r_i(n)$,

$$r_i^* = 1 + \frac{\sum_{j:(v_i,v_j)\in E} x_j^*}{\sum_{j=1}^{m} x_j^*}, \tag{3.147}$$

i.e. if the steady-state size of its neighbours is higher.

7 Conclusion

In this chapter we introduced a new dynamical system on a network to describe the growth of nodes over time. We call this map the network logistic

map as it was inspired by the classic logistic map made famous by Lord Robert May in 1976 [48]. Like the classic logistic map, the archetypal example of how complex and chaotic behaviour can arise in simple systems, despite being a relatively simple system we observed some very complex and interesting behaviour during our analysis. This behaviour included some behaviour not seen in the case of the classic logistic map such as quasiperiodicity.

We observed that the topology of the network had a massive influence on the behaviour of the system under the network logistic map. We focused our attention on the special case of a complete network in which all pairs of nodes are linked. We found that below a certain threshold value of the parameter ρ, $\rho_c = \dfrac{m^2 - 2m + \sqrt{12m^2 - 20m^3 + 9m^4}}{2(m-1)^2} > 2$ where m is the number of nodes in the network, there is a single stable fixed point in the system where all nodes are of equal size. Beyond this threshold the dynamics of the system become far more complicated and we observed multiple bifurcations, attractors, periodic orbits and even chaotic behaviour.

In networks with other topologies we also observed a single fixed point attractor below a threshold value of ρ and we hypothesised that this threshold value is greater than two for all networks. At this fixed point nodes of higher degree with neighbours of high degree have higher equilibrium sizes. For this reason we argue that the network logistic map has the potential to provide us with a new centrality measure and compared it to selected existing centrality measures.

Chapter 4

Spatial Economic Development

1 Introduction

In this chapter we turn our attention to the problem of spatial economic development. Spatial agglomeration of economic activities, unequal growth of urban centres and unequal growth of industries are major topics of economic interest. For example, policy makers and economists struggle to fully understand why economic activity tends to concentrate in one urban centre, such as Dublin in the Irish context, over others and how this development can be influenced [67, 32]. Policymakers and economists also have a strong interest in developing an understanding of the mechanisms behind the formation of specialised industrial clusters, such as Galway's medtech cluster or Cork's pharmaceutical cluster, and in predicting and promoting the emergence of future clusters [66, 31, 67, 22]. In Section 2 we examine Ireland's economic development since its emergence from protectionism in the 1950s and introduce the arguments put forward by economists to explain some important features of spatial economic development.

Recently, network theory and agent-based simulation have been used to analyse economic dynamics. One of the major limitations of these models is that the nodes are simply individual entities and do not capture the multiple attributes that these entities may have because of the interplay between the socio-economic, political and historical factors in the areas in which they are located. We believe that the characteristics and dynamics of each node may be determined by multiple overlapping networks across space and time and by the links between these networks. In Section 4 we introduce a set of spatial economic development models which will help us to describe and analyse uneven spatial economic dynamics. These models apply some of the ideas which were developed and introduced in Chapters 3 and 2 with this purpose in mind.

2 Motivating Statistics and Concepts

Before presenting our models of spatial economic development, we first introduce some interesting motivating statistics for the Irish case. Ireland is a small open economy whose economy has seen significant growth and structural changes in recent decades. This growth has not been evenly distributed across the country. Large urban centres such as Dublin, Cork and Galway have seen the most growth and the establishment of specialised industrial clusters and it has been a key goal of policymakers to halt the continued concentration of economic activity and population in the Greater Dublin Area. In this section we present some official statistics from national and international bodies such as the Central Statistics Office (CSO), the Higher Education Authority (HEA) and the United Nations (UN) which help to illustrate the changing composition and size of the Irish economy. We also introduce some economic concepts regarding the spatial agglomeration of economic activities.

2.1 The Evolution of Irish GDP

We first look at the evolution of Ireland's Gross Domestic Product (GDP) in recent years. GDP is defined by the Organisation for Economic Co-operation and Development (OECD) as the standard measure of the value of final goods and services produced by a country during a period minus the value of imports. GDP reflects the level of economic activity and productivity in a country and so is often used as a barometer of economic health. In Figure 4.1 we have plotted the annual GDP figures from the World Bank [5] for Ireland for the years 1960 to 2012 in current United States Dollars (US$).

Ireland experienced slow economic growth in the post-war 1950s and early 1960s despite the post-war boom in the rest of Western Europe at the time due to a slow dismantling of a policy of *protectionism*. Protectionism refers restricting international trade, often done with the intent of protecting local businesses and employment from foreign competition, through methods such as import tariffs, quotas and subsidies or tax cuts for local businesses. It is now widely accepted that the cost of protectionism policies outweigh the benefits and protectionism hurts the economy of the country that imposes it [15]. As a result of its protectionism policies Ireland had an annual growth rate of less than 2 per cent during 1950s compared around 6 per cent in the rest of Western Europe where such policies had already been abandoned [8].

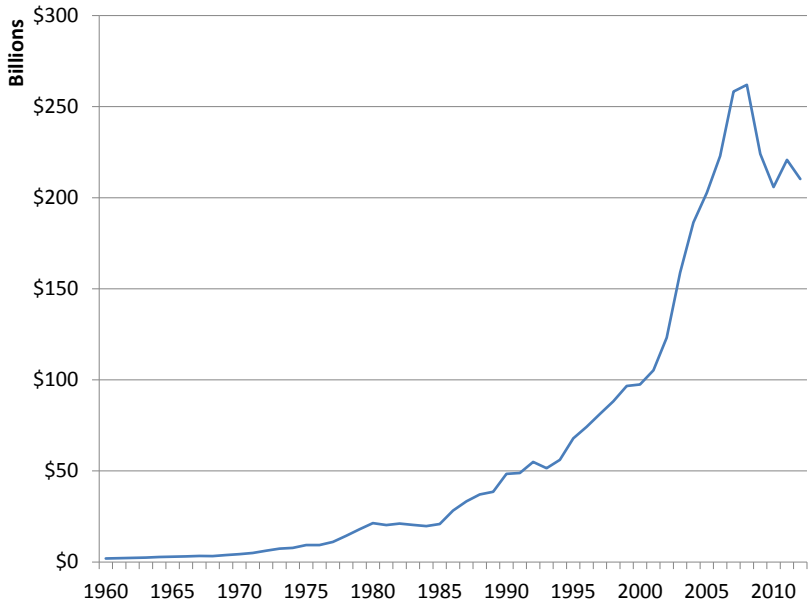


Figure 4.1: Evolution of Irish annual GDP in billions of current US$ between 1960 and 2012. Source: [5].

By the late 1960s and early 1970s Ireland had almost completely emerged from protectionism. Ireland came to a free-trade agreement with its primary

trading partner the United Kingdom (UK) in 1966 and together with the UK and Denmark joined the EU in 1973. Ireland's entry to the European Union (EU), at the time known as the European Economic Community (EEC), did not result in a sudden expansion of the Irish economy. In 1973 the world experienced the first of the 1970s oil shocks hindering growth. A second oil shock in 1979 ensured Ireland's growth remained relatively slow.

The early 1980s saw a sharp increase in world interest rates, this along with a recession in the UK, which reduced Irish exports and the number of Irish of working age emigrating, further affected Ireland's economy. Unemployment rose rapidly forcing up social welfare payments and causing national debt to spiral out of control and Ireland to enter recession. Throughout the 1980s Ireland's economy also struggled to adapt to the new free market. As the 1980s continued Walsh, Whelan and Barry [69, 8] argue that there were several forces working in different directions on the Irish economy. Many firms in internationally tradeable sectors after the end of protectionism could not adjust to the new free market and become exporters or compete with imports. They declined slowly and had not fully left the economy until the beginning of the 1990s and the so-called *Celtic Tiger* era. However due to foreign direct investment (FDI), new foreign firms were entering the Irish economy and indigenous firms sprung up to supply them. Firms that had traditionally been exporters also fared well.

In the late 1980s and early 1990s Ireland benefited from a series of positive economic shocks. Changing government policies in taxation and the establishment of *social partnership*, a process used to negotiate and achieve consensus on a range of social and economic policy issues such as wage determination, improved the public finances and bolstered Ireland's competitiveness. With assistance from the EU Structural Fund infrastructure was also improved further helping competitiveness. Consumer spending, construction and business investment all rose and Ireland saw rapid growth. FDI was a particularly important portion of this business investment. The development of the Single Market in the late 1980s led to a massive increase in FDI in the EU as a whole. For example the amount of investment by US firms in the EU between the early and the late 1980s doubled while Ireland's share of these investments quadrupled over this period [8].

According to Central Statistics Office (CSO) statistics Ireland was the first country in the EU to officially enter a recession related as the Financial crisis hit in 2008. Ireland was severely impacted by the global recession and along with Portugal, Italy, Greece and Spain was one of the worst affected EU member states. A collapse in the housing industry, whose boom helped fuel the Celtic Tiger, and a domestic banking crisis forced Ireland into a deep recession and a bailout from the European Central bank (ECB) and International Monetary Fund (IMF). Under EU-IMF supervision Ireland put in place austerity measures in order to control spiraling debt and has begun in recent years to slowly recover.

2.2 The Evolution of Irish Exports

As a small open economy exports are extremely important for Ireland's recovery and future development. In a speech on October 12 2011 Olli Rehn, at the time European Commissioner for Economic and Monetary Affairs and the Euro, who had a key role in Ireland's bailout, stated that he was confident that the *"process of export-led recovery will succeed in Ireland"* [57].

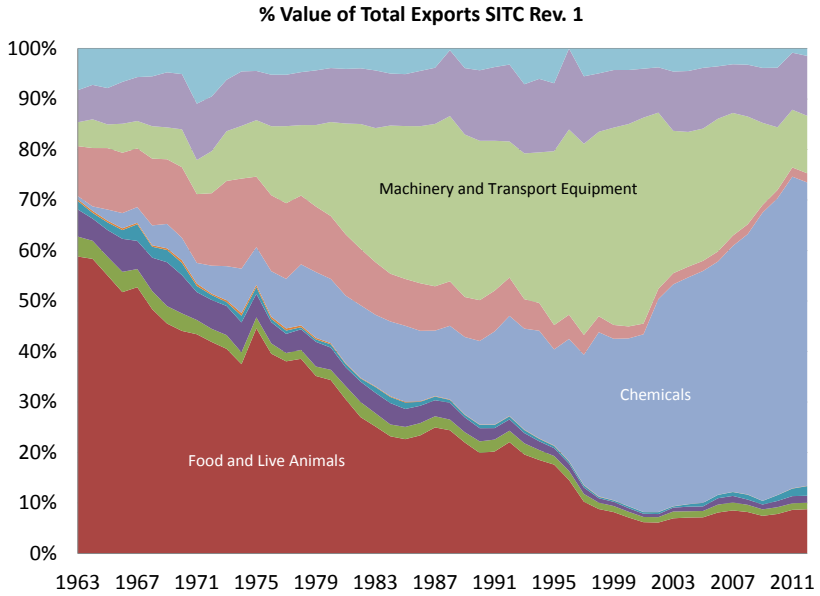


Figure 4.2: The changing composition of Irish exports by Standard International Trade Classification between 1963 and 2012. Source: [21].

Ireland's export sector, like its economy in general, has evolved significantly since the abandonment of protectionist policies. The composition and the value of Irish exports have undergone huge changes in recent decades. We present evidence of these changes through statistics from the United Nations Commodity Trade Statistics Database. The changing composition of Irish exports between 1963 and 2012 is plotted in Figure 4.2. Reading from bottom to top the areas represent the percentage of total Irish exports in

- 0 - Food and live animals
- 1 - Beverages and tobacco
- 2 - Crude materials, inedible, except fuels
- 3 - Mineral fuels, lubricants and related materials
- 4 - Animal and vegetable oils, fats and waxes

- 5 - Chemicals and related products
- 6 - Manufactured goods classified chiefly by material
- 7 - Machinery and transport equipment
- 8 - Miscellaneous manufactured articles
- 9 - Commodities and transactions not classified elsewhere.

These product classifications are the classifications given by the Standard International Trade Classification (SITC). According to the UN [49], the SITC is used for compiling international trade statistics on all merchandise entering international trade, and to promote international comparability of international trade statistics. The commodity groupings of SITC reflect (a) the materials used in production, (b) the processing stage, (c) market practices and uses of the products, (d) the importance of the commodities in terms of world trade, and (e) technological changes. The majority of countries and international organizations continue to use SITC for various purposes, such as study of long-term trends in international merchandise trade and aggregation of traded commodities into classes more suitable for economic analysis of trade.

We have highlighted the three categories of exported product we believe have undergone the most significant and interesting changes in the Irish context between 1963 and 2012 in Figure 4.2. Those are: food and live animals, chemicals and machinery and transport equipment. In the 1960s Ireland's economy was largely low tech and had a very large agricultural sector. Agricultural exports dominated with food and live animals accounting for over 50 per cent of Irish exports up until the late 1960s. Since then food and live animal's share of Irish exports has seen a steady decline up until the early 2000s where the share has settled at around 10 per cent of Irish exports.

The percentage of total exports attributed to machinery and transport equipment grew steadily from the late 1970s until the early 2000s and it became the largest contributor to Irish exports during this time. From 2001 on machinery and transport equipment saw its share of Irish exports decline rapidly and this decline continued right up until 2012. In the early 1960s Ireland's chemical exports were almost non existent, their share of Irish exports grew slowly until the mid 1990s. At this point the percentage of chemical exports began to grow rapidly and by 2012 accounted for over 60 per cent of total Irish exports.

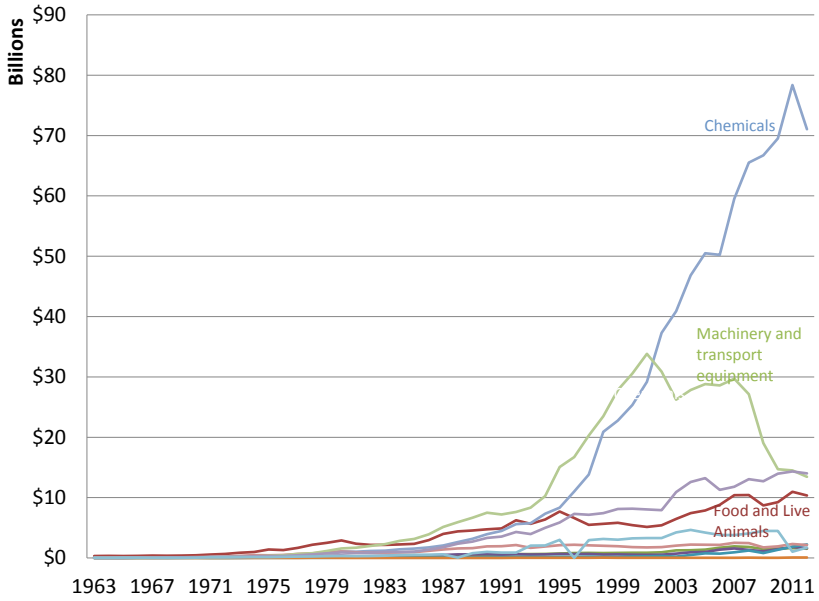


Figure 4.3: The changing value of Irish exports in billions of US$ by Standard International Trade Classification between 1963 and 2012. Source: [21].

Figure 4.3 shows the evolution of the US$ value of the same 10 classes of Irish exports as above, again between the years 1963 and 2012. We have also highlighted exports of food and live animals, machinery and transport equipment and chemicals here. It is notable that while the share of agricultural exports has declined rapidly since the early 1960s, in US$ terms exports in this area have seen generally slow but steady growth. This is true even when the figures are adjusted for inflation. Their declining share of total exports may be attributed to the far more rapid growth in other areas rather than a decline in the actual US$ value. US$ exports of machinery and transport equipment overtook exports of food and live animals in 1983 and continued to grow at a far higher rate until 2001 reaching a peak US$ value of almost 34 billion in 2001. Since then this export sector has seen a considerable decline. The value of machinery and transport equipment sold by Irish exporters fell to around 13.5 billion US$ by the year 2012, less than half its peak value reached in the year 2001. Until the early 1990s Ireland exported a relatively small amount of chemicals. The US$ value of chemicals exported did not overtake the US$ value of food and live animals exported until 1993. During the 1990s and 2000s this sector saw explosive growth in exports, overtaking machinery and transport equipment as the largest contributor to Irish exports in 2002 as its US$ value grew from less than 4 billion in 1990 to a peak of over 78 billion in 2011.

We now examine the evolution of Irish exports in these three classes over the same period in more detail, taking the finer 2-digit level SITC breakdown in each case.

Food and Live Animals: We can see in Figure 4.4 the growth experienced by different categories of Irish food and live animals exports. We have highlighted the two largest categories in terms of US$ exports in 2012, meat and meat preparations and dairy products and eggs, and the largest category in terms of US$ exports in 1963, live animals. Reading from highest to lowest in 2012 the time series plotted in Figure 4.4 represent the US$ value of Irish exports in

- 01 - Meat and meat preparations
- 02 - Dairy Products and eggs
- 04 - Cereals and cereals preparations
- 03 - Fish and fish preparations
- 09 - Miscellaneous food preparations
- 00 - Live animals
- 05 - Fruit and vegetables
- 08 - Feed stuff for animals excluding unmilled cereals
- 07 - Coffee, tea, cocoa, spices and manufactures thereof
- 06 - Sugar, sugar preparations and honey.

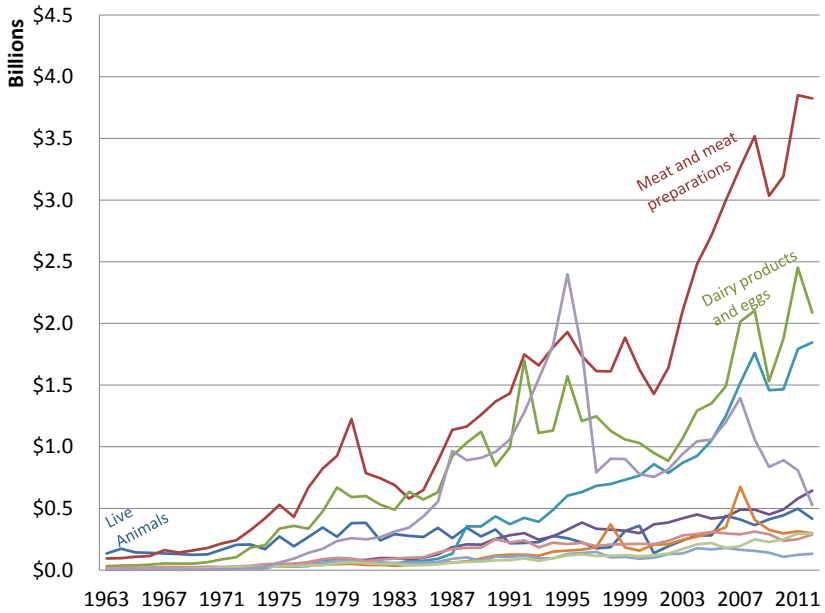


Figure 4.4: The evolution of Irish food and live animal exports in billions of US$ broken down by their two-digit SITC classification between 1963 and 2012. Source: [21].

In general we see that there has been slow growth in this sector since 1963. Meat and meat preparations overtook live animals as the subclass with the largest value of US$ exports in the late 1960s and has remained the top subclass ever since, excepting a spike in miscellaneous food preparations exports in the mid 1990s, which is possibly attributable to reporting and classification issues, and 1984 when it was slightly outstripped by exports of dairy products and eggs. There has been very little growth in exports in live animals throughout this period resulting in a fall to the sixth most important subclass by 2012. Dairy products and eggs exports were larger than exports of live animals for the first time in 1974 and, excepting the periods described above, have been the second largest subclass of exports in US$ terms since.

Machinery and Transport Equipment: Figure 4.5 shows the evolution of Irish exports of different subclasses of machinery and transport equipment between 1963 and 2012. The two-digit SITC classification breaks this class into only three different subclasses

- 71 - Machinery, other than electric
- 72 - Electrical machinery, apparatus and appliances
- 73 - Transport equipment.

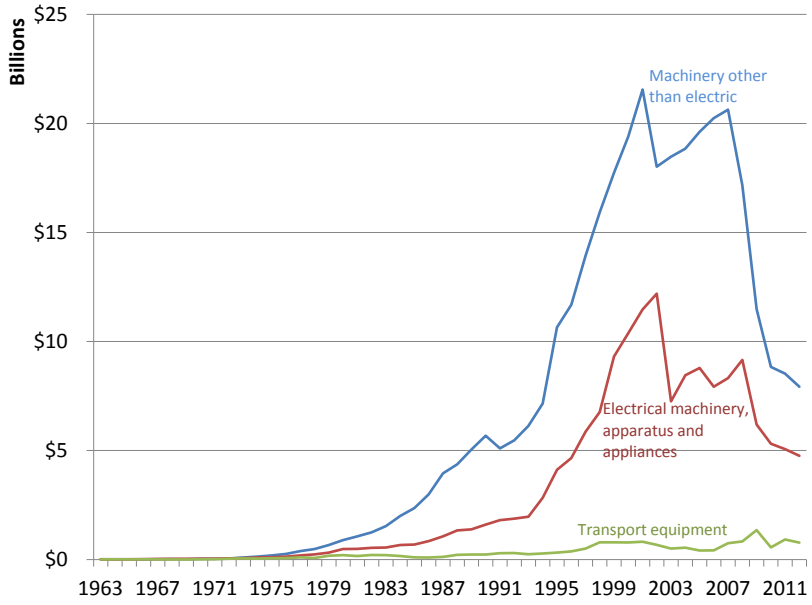


Figure 4.5: The evolution of Irish machinery and transport equipment exports in billions of US$ broken down by their two digit SITC classification between 1963 and 2012. Source: [21].

We saw in Figure 4.2 and Figure 4.3 that Ireland's exports of machinery and transport equipment began to grow quite rapidly during the 1970s, continuing to grow strongly until the late 1990s and early 2000s before entering a period of decline. It becomes clear when examining Figure 4.5 that these changes were driven by changes in exports of machinery other than electric and exports of electrical machinery, apparatus and appliances with exports of transport equipment exhibiting little change during this period.

Within these categories the bulk of Ireland's productivity is/was in computer hardware, office machines, data processing equipment etc. This sector went through a quite dramatic rise and fall in Ireland between the 1970s and the present day. In 1971 Digital Equipment Corporation, a pioneer in the minicomputer industry, set up a large manufacturing plant in Ireland. They were followed by several other leading minicomputer companies as he 1970s continued. However, as PCs increased in popularity demand for minicomputers collapsed many of these plants were closed during the 1980s. Apple was the first company to set up a PC assembly plant in Ireland in 1980. They were followed by many other major PC and computer component manufacturers such as Dell, Intel, IBM, Logitech and the homegrown Horman Electronics. 33 per cent of PCs sold in Europe in 1999 were manufactured in Ireland and at one stage roughly 90 per cent of the mice sold in Europe were manufactured in Ireland [10].

The decline in exports in these subsectors from the late 1990s on was as a result of firms such as Apple, IBM and Intel shifting relatively low-tech labour-intensive activities from Ireland to lower wage economies. Of the five microcomputer companies in 1998, by 2002 only Dell and Apple were still assembling microcomputers, and Apple's system assembly operation was seriously downsized [10]. Companies such as Logitech who manufactured a large portion of the mice produced in Ireland and keyboard manufacturers Keytronics, Mitsumi and Alps also closed their doors. The sector suffered another blow when Dell closed its manufacturing plant in Ireland in 2009 [23].

Chemicals: Figure 4.6 shows the growth experienced by different categories of Irish chemicals exports between 1963 and 2012. We have highlighted the two largest categories in terms of US$ exports in 2012, medicinal and pharmaceutical products and chemical elements and compounds, in the graph. Reading from highest to lowest in 2012 the top five time series plotted in Figure 4.6 represent the US$ value of Irish exports in

- 54 - Medicinal and pharmaceutical products
- 51 - Chemical elements and compounds
- 55 - Perfume materials, toilet & cleansing preparations
- 59 - Chemical materials and products not elsewhere specified
- 58 - Plastic materials etc.

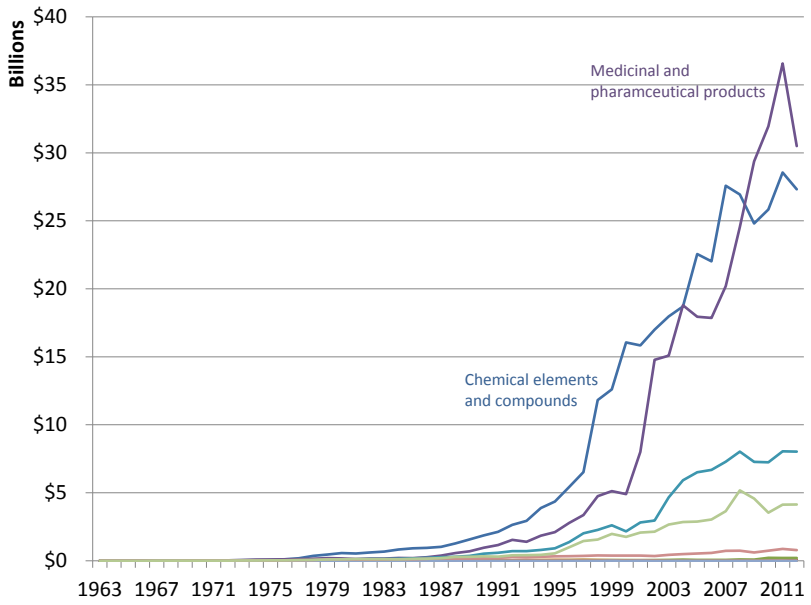


Figure 4.6: The evolution of Irish food and chemical exports in billions of US$ broken down by their two digit SITC classification between 1963 and 2012. Source: [21].

Both our highlighted subclasses of chemicals exports have experienced incredible growth in recent decades. Until the 1960s there was essentially no pharmaceutical industry in Ireland [29]. Despite this, due to phenomenal growth Irish exports of pharmaceutical and medicinal products exceeded 35 billion US$ in 2011. The fine chemicals industry in Ireland began to grow significantly in the 1970s after the Industrial Development Authority (IDA) began to encourage inward FDI in the sector [18]. Growth was first seen in exports of chemical elements and compounds with growth of exports of medicinal and pharmaceutical products following soon after. More than 120 pharmaceutical companies now have a presence in Ireland. 13 of the world's top 15 companies, including Pfizer, Novartis, AstraZeneca and GlaxoSmithKline, have substantial Irish operations and 5 of the top 12 medicines globally are manufactured in Ireland [3].

2.3 Ireland's Changing Economy and the Evolution of Irish Unemployment

In Figure 4.7 and Figure 4.8 we examine the evolution of two subclasses of exports from each of our three major classes of interest along with seasonally adjusted unemployment in Ireland between 1983 and 2012. During the recession of the 1980s the unemployment rate in Ireland was very high, reaching 17 per cent in 1986, and remained so up until the early 1990s.

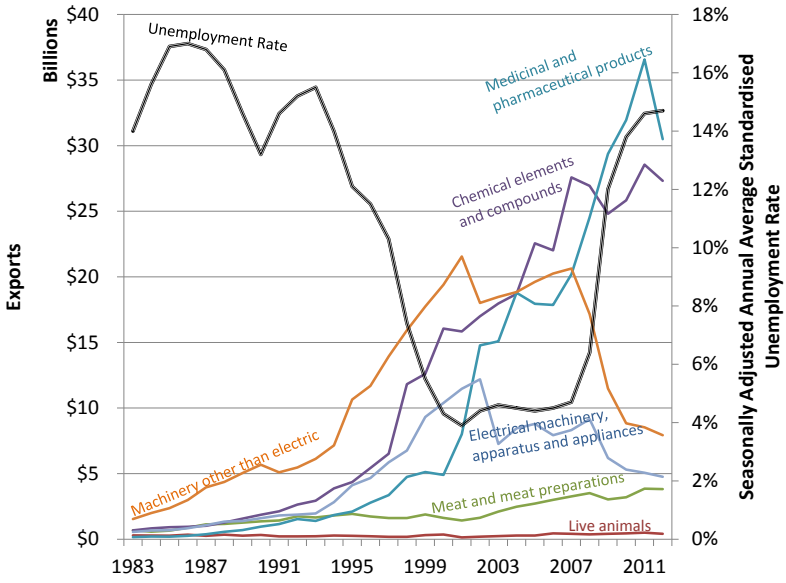


Figure 4.7: The evolution of selected Irish exports in billions of US$ and the seasonally adjusted annual average standardised unemployment rate between 1983 and 2012. Sources: [21] & [56].

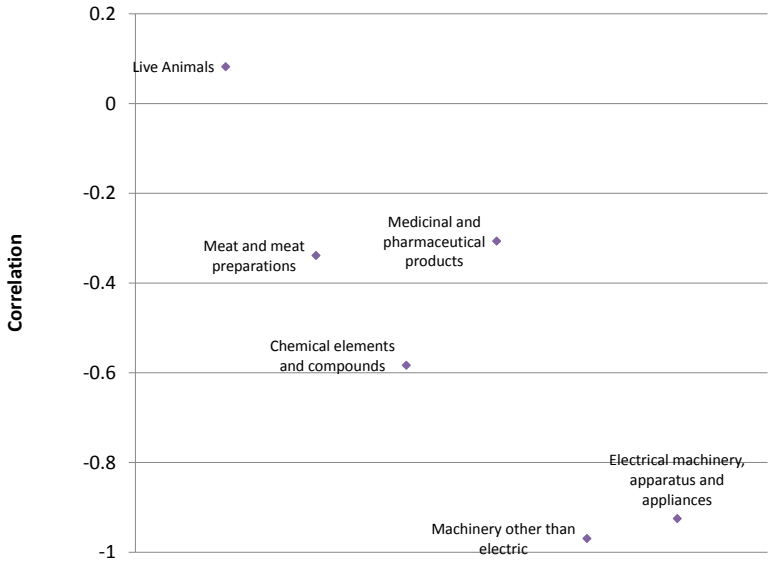


Figure 4.8: Correlations between the volume of selected Irish exports an the seasonally adjusted annual average standardised unemployment rate between 1983 and 2012. Sources: [21] & [56].

The export-driven early Celtic Tiger growth of the 1990s saw an increase in activity in labour-intensive sectors such as the manufacturing of computer hardware discussed earlier and repeated sustained falls in unemployment figures. By the year 2000 the seasonally adjusted average annual unemployment rate had fallen to 4.3 per cent. Economists consider a rate above 0 per cent the full-employment unemployment rate, usually averaging between 5 and 6.5 per cent for OECD countries according to OECD analysis [54], at this rate there is no demand deficient unemployment, the only unemployment is is caused by the time taken to find a suitable job in a suitable location. From 2000 to 2007 unemployment remained stable at around 4.5 per cent, virtually full-employment, and the economy continued to grow rapidly. However, at this point the nature of Ireland's growth had changed from the stable export led growth of the 1990s to a property price and construction bubble. The boom sustained employment and output growth until 2007 but Ireland lost wage competitiveness and saw its export sectors shifting relatively low-tech labour-intensive activities from Ireland to lower wage economies [36]. When the global financial crisis struck in 2008 Ireland entered a deep recession with many labour intensive industries fueled by the bubble, particularly in construction and related industries, suffering the most. From 2008 until 2012 Ireland saw sharp increases in the unemployment rate, rising to 14.7 per cent in 2012. Despite the economic problems in Ireland during this period exports of chemicals, especially in the pharmaceutical sector continued to grow strongly. However, growth in these industries, which are not particularly labour intensive, did little to halt the rise in unemployment.

Correlation does of course not imply causation. However, looking at the correlations between the US$ value of different subclasses of Irish exports and the unemployment rate plotted in Figure 4.8 does help to provide some insight. The correlations of around -1 between the unemployment rate and the exports of labour-intensive machinery sectors reflect the fact that the export-driven growth of the 1990s led to a fall in unemployment while the decline in these sectors, when not masked by the credit-fueled construction boom of the early to mid 2000s, along with other factors saw a sharp increase in unemployment. The correlations closer to 0 in the chemicals sector reflect that this is a high-tech sector which is not particularly labour intensive and that it did not suffer badly in the crash. The correlations close to 0 in the agricultural sector are due to Ireland's exports of food and live animals remaining relatively stable throughout Ireland's shift to a more modern economy and the recessions and booms since.

2.4 Spatial Concentration of Economic Activities

The changing nature, health and composition of the Irish economy affects society in many different ways. Unemployment is just one of many major issues. These changes affect different areas more profoundly than others due to the spatial agglomeration of certain economic activities. For example as Ireland opened and modernised its economy in the 1960s, 1970s and 1980s agricultural sectors became less and less important, economic activity in urban areas began to grow far more quickly than in rural areas. Some industries tend to concentrate in

industrial clusters, the OECD Economic Survey of Ireland 2013 identified three important industrial clusters in Ireland, medical devices in Galway, pharmaceuticals in Cork and computer hardware and software in Dublin [55]. Another example of a highly clustered sector in Ireland is the international financial services sector which is highly concentrated in the International Financial Services Centre (IFSC) in Dublin. Positive or negative shocks in any of these industries would have a major effect on the economy and community of the regions in which their activity is clustered.

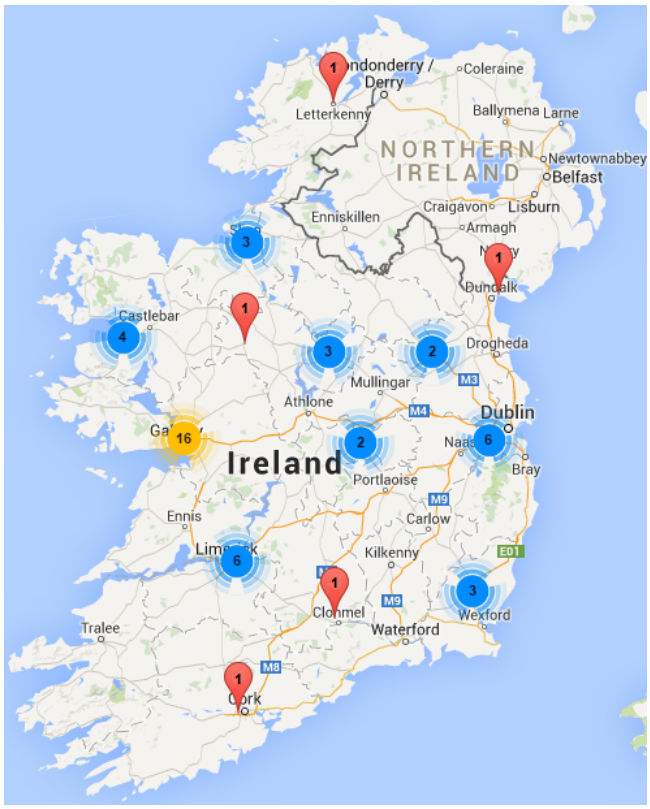


Figure 4.9: Map of Ireland highlighting the locations of medical device research and manufacturing operations. The number on each node represents the number of operaions in that location. Reproduced from [2]. Map data: ©2015 GeoBasis-DE/BKG (©2009), Google.

The IFSC is hugely important to the Dublin area. Over 500 firms are located in the IFSC, directly employing 32,700 people, contributing 7.4 per cent of Irish GDP and comprising 5 per cent of all EU 27 cross-border financial services activity [40]. Similarly the medical devices industry is vital to the economy of the west of Ireland and particularly to Galway. According to the Irish Medical Devices Association the medtech sector employs 25,000 people in Ireland which

is the highest number of people working in the industry per capita in any country in Europe [39]. Figure 4.9 maps the spatial distribution of firms involved in the contract manufacturing, research and/or design of medical devices or bringing finished medical device products to market in Ireland. It is immediately clear that the majority of activity is concentrated around Galway with 16 firms present, the next largest cluster is in the greater Dublin area where 6 firms are present.

(a) Locations of Irish drug substance operations.

(b) Locations of Irish drug product operations.

Figure 4.10: Map of Ireland highlighting the locations of pharmaceutical operations in 2003. Node size is given by total employees. Reproduced from [66].

Some of the most well-known and analysed industrial clustering takes place in Ireland's pharmaceutical industry. Figure 4.10 shows the spatial distribution of firms in the pharmaceutical industry in Ireland, showing its clustered nature. Clustering is most evident in plants producing drug substances (Figure 4.10a), the active ingredients responsible for a drugs pharmacological effect, with a large cluster in Cork and a smaller cluster in Dublin these two counties had a 74 per cent share of employment in the drug substance sub-sector in 2003 [66]. Operations producing finished products are not as spatially concentrated (Figure 4.10b). The largest cluster in this case is in Dublin, four counties each accounted for between 13 and 19 per cent of the total in 2003 [66]. Taking the pharmaceutical industry as a whole in 2003 Cork and Dublin together accounted for 45 per cent of all employment. As of 2012 almost 25,000 people are employed directly by the pharmaceutical industry [3], up from 1300 in 1972

[66]. According to the Irish Pharmaceutical Healthcare Association (IPHA) a further 25,000 jobs depend on the provision of services to the sector [3].

Particular industries are attracted to certain areas and tend to cluster for a variety of different reasons. Ireland's 2002 National Spatial Strategy (NSS) provides a good summary of many of the factors that lead to the spatial agglomeration of economic activities:

> *"Business is likely to align itself closely with local strengths, facilities, talents and skills. This can be facilitated in important ways through good links between business and third-level institutions. Clusters of similar or interrelated overseas and Irish-owned businesses will tend to form and consolidate in particular geographic areas because of the advantages available locally and the resulting synergies. "* [32]

The facilities, institutions and infrastructure in a particular area are always a factor for firms in choosing the location for a new operation. For example, for many firms, and in particular firms in high-tech sectors, links with nearby educational institutions are important. Many beneficial links between Irish third-level institutions and industry are clear. Ireland's Universities and Institutes of Technology (ITs) provide companies with a pool of highly educated and skilled potential employees. For example a BPharm/BSc degree programme was recently established by University College Cork (UCC) and the content of the programme was influenced by the desire to cater for the needs of the local pharmaceutical cluster after discussions with staff of local pharmaceutical companies [7] and half of the 25,000 people employed by Ireland's pharmaceutical industry hold a third-level qualification [3].

The OECD Economic Survey of Ireland 2013 [55] also identified the specialised training and research programmes in the medical devices field as a factor in the growth of Galway's medical devices industry. Industry/academic collaborations in terms of research are also highly important. Ireland's Centres for Science, Engineering and Technology, supported by Science Foundation Ireland (SFI), link scientists and engineers from academia and industry to work on leading-edge research. In the medical devices sector for example, global leaders Medtronic, who have been manufacturing vascular products such as stents in Ireland for over 20 years, is involved in several Irish industry/academic collaborations, including REMEDI, the Regenerative Medicine Institute based in National University of Ireland, Galway (NUIG). Another example of industry/academic collaboration in high-tech manufacturing is the Tyndall National Institute based in UCC. The Tyndall National Institute is a leading centre for semiconductor research with industry participants including IBM, Sony, Siemens and Phillips Semiconductor [38].

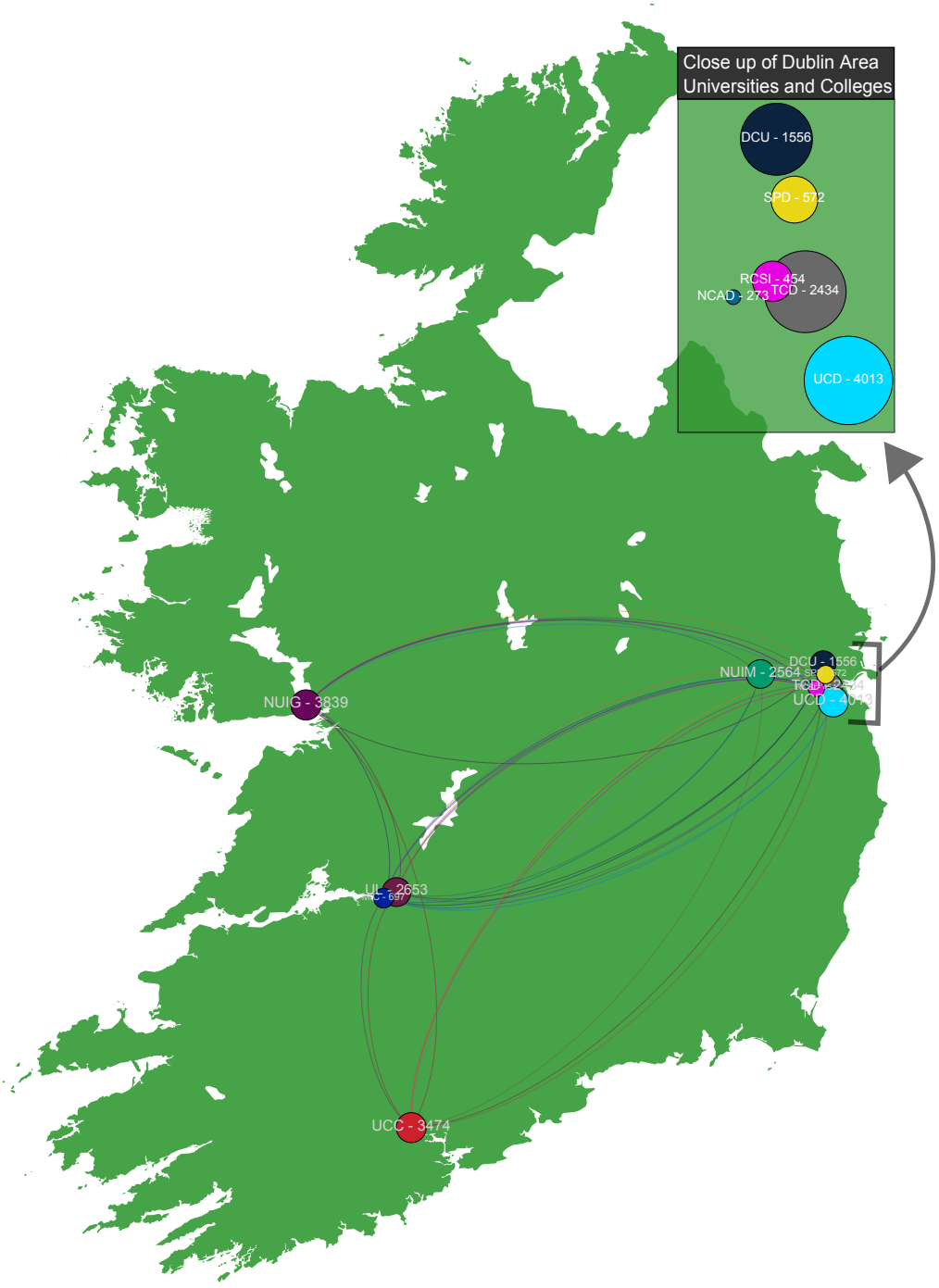


Figure 4.11: Map of Ireland highlighting the locations of all Irish universities and selected colleges. 2012 Undegraduate graduate numbers are indicated for each institution. Source: [4].

Figure 4.11 maps the locations of all Irish universities and selected colleges. The nodes representing the third-level institutions are sized by undergraduate graduate numbers according to the Higher Education Authority (HEA) in 2012 [4]. More detailed maps which include a breakdown of graduate numbers by academic area are available for the years 2009-2012 on request. The majority of graduates in 2012 graduated from universities or colleges in Dublin. University College Dublin (UCD) had the highest number of graduates among the universities at 4,013, followed by National University of Ireland Galway (NUIG) at 3839. Dublin City University (DCU) had the lowest number of graduates at 1,556. The largest college in terms of graduate output included in our map was Mary Immaculate College (MIC) in Limerick. MIC had a graduating class of 697 undergraduates in 2012. The smallest was the National College of Art and Design (NCAD) with only 273 graduates in 2012.

The *synergies* to be taken advantage of referred to in the NSS are often known as *agglomeration economies*. The idea of agglomeration economies was first introduced by Marshall in 1890 [46]. In 1937 Hoover divided these economies into two types: *localisation economies* and *urbanisation economies* [37]. Localisation economies are those introduced by Marshall, advantages that firms in a set of closely related industries, such as a pooled market for workers with specialised skills, a growing number of increasingly specialised input suppliers and technological spillovers, gain from being located in the same location, while urbanisation economies are advantages gained by all firms, regardless of sector. Galway's strong ICT sector was a factor in the growth of another similar high-tech industry in the area in the form of a medical devices cluster [11], this could be seen as an example of Marshallian localisation economies at work. In the late 1980s the EU Single Market was being developed, this led to a doubling in the amount of FDI by US firms in the EU between the early and the late 1980s. Localisation economies have been cited as an important factor in Ireland quadrupling its share of these investments over the same period [8, 12]. Most industrial clusters in urban areas are also bound to benefit from at least some level of urbanisation economies in the form of educational institutions and infrastructure [66].

It has been argued that even if these synergies between similar firms are not a factor firms may still choose to agglomerate spatially. DeCoster and Strange [20] argue that even if there are no efficiency reasons for spatial agglomeration, due to uncertainty about locations in which to invest, investors may imitate the location decisions of other firms. This happens because if one firm chooses a specific location for a new operation it signals to other firms that this location is a good one and as a result financial institutions, who may be required to provide credit for new investments, and investors view investments in operations in the same location as safer with a higher probability of success. These *demonstration effects* have also been cited as a factor in Ireland's success in attracting FDI in recent decades [12]. Surveys of foreign companies in the computer, instrument engineering, pharmaceutical and chemical sectors have indicated that location decisions are strongly influenced by the fact that other key market players have an existing presence in Ireland [9].

In 2002 the government of Ireland released the NSS, a twenty-year strategic plan designed to ensure *"more balanced social, economic and physical development between regions"* [32]. Ireland's development up to 2002 and indeed up until the present day has been marked by a spatial imbalance. The capital city Dublin and the Greater Dublin Area (GDA) have grown far quicker than other areas and as a result has problems with congestion, economic activity and as a result employment opportunities have become more concentrated in certain areas, other areas have not benefited to the same extent from Ireland's growth and remain economically weak.

The NSS aimed to counteract the concentration of economic activity in the GDA by stimulating the development of regional centres referred to as *gateways* and other urban centres referred to as *hubs*. The NSS aimed to promote specialisation by different regions through the formation of specialised industrial clusters. Policymakers in Ireland have looked to promote the formation of such clusters since the 1992 Culliton Report [19]. This report recommended the promotion of industrial clusters focused on niches of national competitive advantage in line with the ideas of Porter [59]. The NSS cited *"[s]patial clusters of international excellence [...] emerging in Ireland"* and sought to *"strengthen these areas and increase their number by supporting the formation of self-sustaining clusters of economic activity"*.

The aims of the NSS were very ambitious. Influencing patterns of spatial economic development and promoting balanced development is an extremely difficult task. The strategy failed to promote substantial economic growth in most of the identified gateways and hubs. Only Cork and Galway, driven by the electronics and medical devices sectors respectively saw the kind of growth and specialisation envisaged by the NSS. Other areas such as Sligo, Limerick-Shannon and the Midlands saw little or no benefit [67]. The NSS was scrapped by Phil Hogan TD, Irish Minister for the Environment, Community and Local Government, in February 2013 saying the *"strategy had failed"* due to a lack of resources [53]. However, lack of resources was not the only reason for the failure of the NSS. In 2013 van Egeraat and Breathnach called for *"more focused identification of existing regional strengths followed by appropriate support measures for further development of these strengths"* [67]. Balanced spatial growth remains an important issue for Irish policymakers and we believe a novel model of spatial economic development has the potential to be a useful tool in future policy decisions aimed at promoting growth and specialisation in underdeveloped regions.

3 Uses of Networks in Economics

Since the 1990s economists have become increasingly interested in the use of networks in modelling and analysing economies. A number of influential publications including Mark Granovetter's 1985 *Economic Action and Social Structure: The Problem of Embeddedness* led to this rising interest in the role of networks in economics [33]. In his article Grovonettor [33] argued against the

traditional neoclassical view of economics in which institutions and agents are viewed as independent. He described this view as an *"atomised,* undersocialised *conception of human action"*, disallowing by any hypothesis *"any impact of social structure and social relations on production, distribution, or consumption"* [33]. At the time of publication he argued that economic behaviour had been inadequately interpreted because professional economists were *"so strongly committed to atomised theories of action"* and most economic behaviour is closely embedded in social relationships [33].

One particular example we are interested in is the use of networks in development economics to analyse and assess inter-industry relatedness. Economists such as Bryce and Winter [16], Hidalgo *et al.* [34] and Neffke and Svensson Henning [50] have used co-occurrence analysis to estimate how industry similarity. Co-occurrence analysis measures the similarity between two industries by measuring how frequently two industries are found together in the same economic entity.

Neffke and Svensson Henning's [50] approach was to construct a measure to estimate industry similarity which they called *revealed relatedness* (RR). The measure derives industry similarity from the co-occurence of products that belong to different industries in the portfolios of manufacturing plants. They first counted the the number of co-occurrences between each pair of industries, then compared this number to the number of co-occurrences that would be expected between each pair based on industry level characteristics such as profitability, employment numbers and numbers of active plants in each industry. The more often two industries co-occur relative to this baseline, the higher their RR.

Hidalgo *et al.*'s [34] approach was to examine country level co-occurrences. They developed a measure of proximity between 775 SITC 4-digit level product classes based on the number of times that two products have a *revealed comparative advantage* in the same country. A product has a revealed comparative advantage in a country if that country exports more of that product as a share of total exports than the average country. Formally a product i has a revealed comparative advantage in country c if

$$RCA_{c,i} = \frac{\frac{x(c,i)}{\sum_j x(c,j)}}{\frac{\sum_d x(d,i)}{\sum_{d,j} x(d,j)}} > 1, \tag{4.1}$$

where $x(a,b)$ represents the exports of product b by country a. The proximity ϕ between product i and product j is then given by the minimum of the pairwise conditional probabilities of a country having a revealed comparative advantage in one given it exports the other, i.e.

$$\phi_{i,j} = \min\{P(RCA_{c,i} > 1 | RCA_{c,j} > 1), P(RCA_{c,j} > 1 | RCA_{c,i} > 1)\}. \tag{4.2}$$

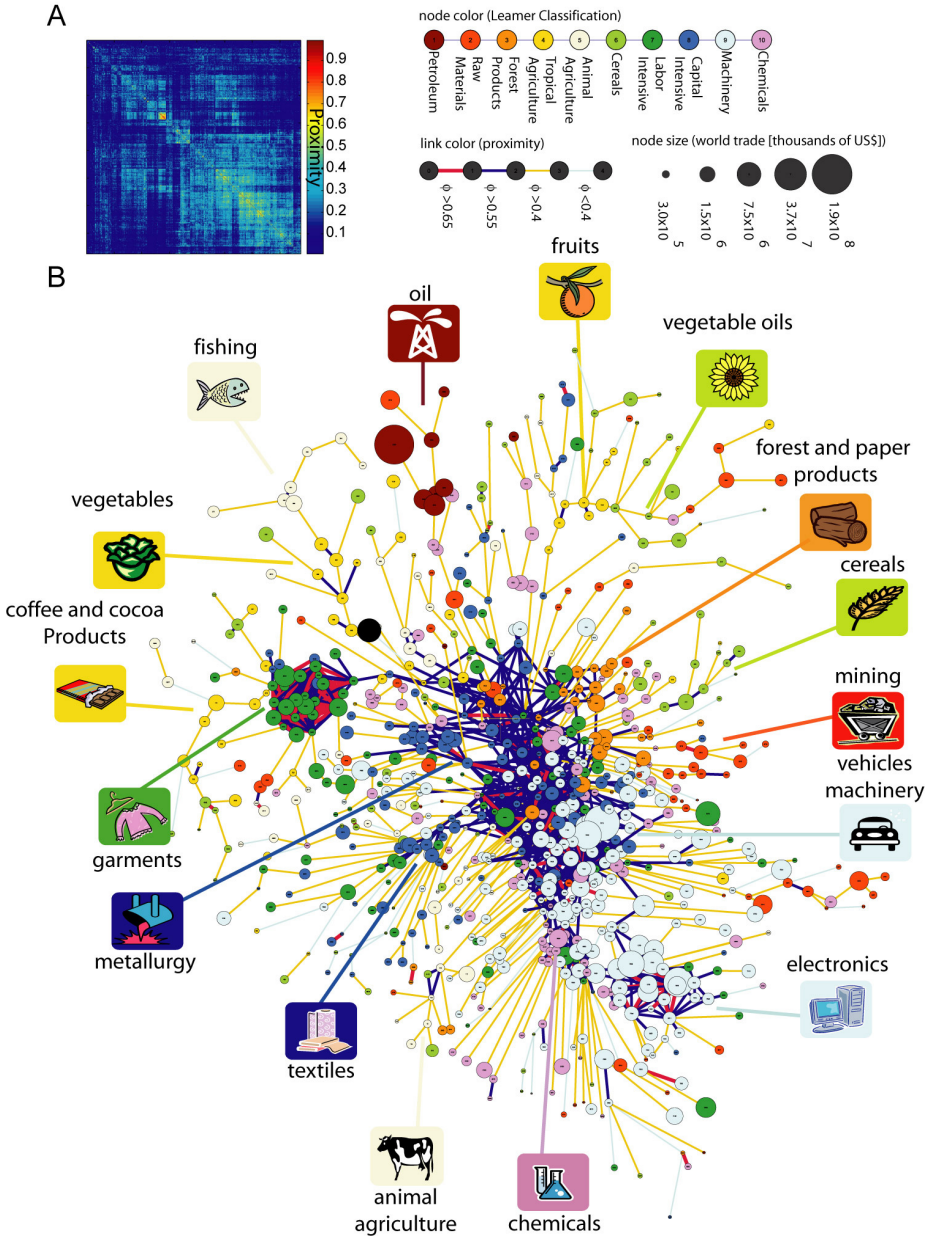


Figure 4.12: The product space of Hidalgo *et al.*
(A) Hierarchically clustered proximity matrix representing the 775 SITC-4 product classes exported in the 1998-2000 period.
(B) Network representation of the product space. Links are colour coded with their proximity value. The sizes of the nodes are proportional to world trade, and their colours are chosen according to the classification introduced by Leamer. Reproduced from [34]

Figure 4.12 shows the network of relatedness between products termed the *product space* produced by Hidalgo *et al.*'s analysis. Some products are seen to be highly connected while other products are disconnected. The distribution of proximities between product pairs, $\phi_{i,j}$, is broad and the product space network itself is quite sparse, 5 per cent of the elements of the proximity matrix (ϕ) are equal to zero, 32 per cent are smaller than 0.1, and 65 per cent take values below 0.2.

The product space is very inhomogenous. It has a clearly visible core-periphery structure featuring a dense, highly-connected core and a sparsely connected periphery. The core group of products is made up of mostly upscale, high-value products such as metal products, machinery and chemicals while products from other, mainly lower-income product classes are found in the periphery. One notably strong, dense peripheral cluster is formed by products in the garments sector and can be seen on the left of the product space network diagram in Figure 4.12B. The product clusters shown by the product space network show quite strong agreement to the product classification performed by Leamer [43] in 1984 based on the relative amount of capital, land, labour or skills required to produce each product. However, we also see a more detailed split within many industries and shows links of varying strengths between others. For example, the machinery cluster is split in two, one cluster consisting of heavy machinery and the other consisting of electronics.

Economies grow by upgrading the type of products they produce and export. Both Neffke *et al.* and Hidalgo *et al.* argue that the growth and development of countries and regions are subject to strong path dependencies [34, 51]. The future economic development of a country or region is directly linked to the products already produced in that region and the position of these products in the product space (or in Neffke *et al.*'s case the *industry space*). An economy's specialisation in the production of certain goods dictates its future ability to specialise in other areas with it being most likely that the economy will branch into industries that are closely related to the preexisting industries in the region or country. For example a country which has a revealed comparative advantage in exporting apples will probably have most of the correct conditions required to begin exporting pears.

This tendency of countries and regions to move to goods close to those they are currently specialized in may help to explain why some regions grow faster than others and different areas develop different specialised clusters. This can lead to certain areas finding it difficult to develop more competitive sophisticated sectors and failing to converge to the income levels of richer countries or regions.

Hidalgo *et al.* use a simple analogy involving monkeys moving through a forest to explain the idea:

> *"Think of a product as a tree and the set of all products as a forest. A country is composed of a collection of firms, i.e., of monkeys that live on different trees and exploit those products. The process of growth implies moving from a poorer part of the forest, where trees have little fruit, to better parts of the forest. This implies that monkeys*

> *would have to jump distances, that is, redeploy (human, physical, and institutional) capital toward goods that are different from those currently under production. Traditional growth theory assumes there is always a tree within reach; hence, the structure of this forest is unimportant. However, if this forest is heterogeneous, with some dense areas and other more-deserted ones, and if monkeys can jump only limited distances, then monkeys may be unable to move through the forest."* [34]

Since it is indeed the case that the *"forest"*, i.e. the product space, has a very heterogeneous structure as can be seen in Figure 4.12B, this structure and a country's position within it are vital to the future economic development of countries. These ideas of product/industry relatedness and the path dependency of regional economic development will be of central importance as we develop our networks of spatial economic development in the following sections.

4 Spatial Models of Economic Development

Recently network theory and agent-based simulation have been used to analyse economic dynamics. As we stated in the introduction to this chapter (Section 1), one of the major limitations in these models is that the nodes are simply individual entities and do not capture the multiple attributes that these entities may have because of the interplay between the socio-economic, political and historical factors in which they are located. In this section we introduce two simplified models of spatial economic development based on ideas we believe could address these issues. The first model, presented and analysed in Section 4.1, describes the growth of a single town or city in a small open economy. The second model, presented and analysed in Section 4.2, describes growth and development in a small country with multiple urban centres and an open economy.

In both cases our approach draws inspiration from the spatial network growth model of Kaiser *et al.* [41], presented in Section 1.3, Chapter 2, and our extensions of this model which are presented and analysed later in the same chapter. Recall that in these models we grew large networks from a small existing networks by adding nodes whose position was not predetermined, rather their position and the links between them were established over time as the network developed and were influenced by the existing network. In our models any growth in the economy of our city or state will be influenced by the structure and spatial distribution of existing economic activity.

We will also draw inspiration from the product space and industry space maps of Hidalgo *et al.* and Neffke *et al.* discussed in Section 3. Industry relatedness will have a large influence on the level of demonstration effects and localisation economies experienced by firms in our model. Only firms that are closely related will gain advantages, such as a pooled market for workers with specialised skills, a growing number of increasingly specialised input suppliers and technological spillovers from being located in the same location. Similarly, the success of a firm in a new location signals to other firms and investors

that the location is a good one only if they are in related industries. Rather than segregate economic activity in our models based on the 775 4-digit level SITC product classes investigated by Hidalgo *et al.*, in our simple model we will consider 19 sectors. The proximities between these sectors will be based on a product space analysis by staff of the World Bank for the 16th Annual Conference on Global Economic Analysis [44]. These sectors are listed in Table 4.1 and their proximities are shown in Table 4.2.

Key	**Sector**
1	Agriculture, Forestry & Hunting
2	Fishing & Fish Farming
3	Extractive Industries
4	Food and Tobacco
5	Textiles, Clothing & Leather
6	Other Manufacturing
7	Refining of Petroleum & Other Energy Products
8	Chemicals
9	Mechanical, Metallurgical & Electrical Manufacturing
10	Production and Distribution of Electricity & Water
11	Construction
12	Trade and Repair Services
13	Hotels and Restaurants
14	Transportation Services
15	Mail and Telecommunications
16	Finance and Insurance
17	Business Services
18	Government Services
19	Other Services (Non-finance)

Table 4.1: Industry Sectors.

Figure 4.13 gives a network representation of this *sector space*. The diagram was produced using GEPHI [13] and shows all links between all pairs of sectors with proximities greater than 0.25. Nodes are colour-coded by sector and links between node pairs are coloured using a mix of the two node colours. Links are weighted by their proximity value and this is graphically represented by link thickness. Again we see a core-periphery structure with a strong agriculture/food cluster in the centre and a strong manufacturing cluster to the left of the diagram. The peripheral nodes are mainly in the services industry with the node representing the financial services sector (16) being the least well connected.

Sector	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	0.57	0.45	0.69	0.51	0.48	0.34	0.23	0.32	0.51	0.43	0.58	0.40	0.58	0.49	0.11	0.26	0.23	0.22
2	0.57	1.00	0.41	0.62	0.59	0.48	0.35	0.24	0.35	0.46	0.50	0.54	0.44	0.49	0.48	0.15	0.26	0.24	0.26
3	0.45	0.41	1.00	0.38	0.34	0.28	0.26	0.23	0.33	0.40	0.30	0.35	0.20	0.39	0.35	0.26	0.23	0.26	0.18
4	0.69	0.62	0.38	1.00	0.59	0.64	0.43	0.38	0.49	0.48	0.51	0.49	0.43	0.51	0.51	0.11	0.21	0.23	0.36
5	0.51	0.59	0.34	0.59	1.00	0.64	0.40	0.34	0.46	0.49	0.54	0.40	0.38	0.54	0.46	0.12	0.18	0.18	0.28
6	0.48	0.48	0.28	0.64	0.64	1.00	0.52	0.48	0.65	0.47	0.50	0.37	0.33	0.43	0.50	0.07	0.15	0.22	0.46
7	0.34	0.35	0.26	0.43	0.40	0.52	1.00	0.53	0.51	0.40	0.36	0.29	0.28	0.34	0.38	0.08	0.11	0.19	0.36
8	0.23	0.24	0.23	0.38	0.34	0.48	0.53	1.00	0.62	0.32	0.34	0.13	0.25	0.26	0.38	0.13	0.06	0.35	0.58
9	0.32	0.35	0.33	0.49	0.46	0.65	0.51	0.62	1.00	0.42	0.42	0.23	0.33	0.36	0.40	0.16	0.05	0.19	0.51
10	0.51	0.46	0.40	0.48	0.49	0.47	0.40	0.32	0.42	1.00	0.47	0.45	0.25	0.54	0.42	0.15	0.19	0.19	0.19
11	0.43	0.50	0.30	0.51	0.54	0.50	0.36	0.34	0.42	0.47	1.00	0.29	0.40	0.49	0.50	0.12	0.16	0.28	0.44
12	0.58	0.54	0.35	0.49	0.40	0.37	0.29	0.13	0.23	0.45	0.29	1.00	0.44	0.48	0.38	0.17	0.35	0.19	0.12
13	0.40	0.44	0.20	0.43	0.38	0.33	0.28	0.25	0.33	0.25	0.40	0.44	1.00	0.28	0.42	0.13	0.33	0.35	0.38
14	0.58	0.49	0.39	0.51	0.54	0.43	0.34	0.26	0.36	0.54	0.49	0.48	0.28	1.00	0.46	0.10	0.18	0.20	0.26
15	0.49	0.48	0.35	0.51	0.46	0.50	0.38	0.38	0.40	0.42	0.50	0.38	0.42	0.46	1.00	0.17	0.23	0.35	0.38
16	0.11	0.15	0.26	0.11	0.12	0.07	0.08	0.13	0.16	0.15	0.12	0.17	0.13	0.10	0.17	1.00	0.22	0.19	0.10
17	0.26	0.26	0.23	0.21	0.18	0.15	0.11	0.06	0.05	0.19	0.16	0.35	0.33	0.18	0.23	0.22	1.00	0.38	0.07
18	0.23	0.24	0.26	0.23	0.18	0.22	0.19	0.35	0.19	0.19	0.28	0.19	0.35	0.20	0.35	0.19	0.38	1.00	0.38
19	0.22	0.26	0.18	0.36	0.28	0.46	0.36	0.58	0.51	0.19	0.44	0.12	0.38	0.26	0.38	0.10	0.07	0.38	1.00

Table 4.2: Sector proximities (ϕ). Source: [44].

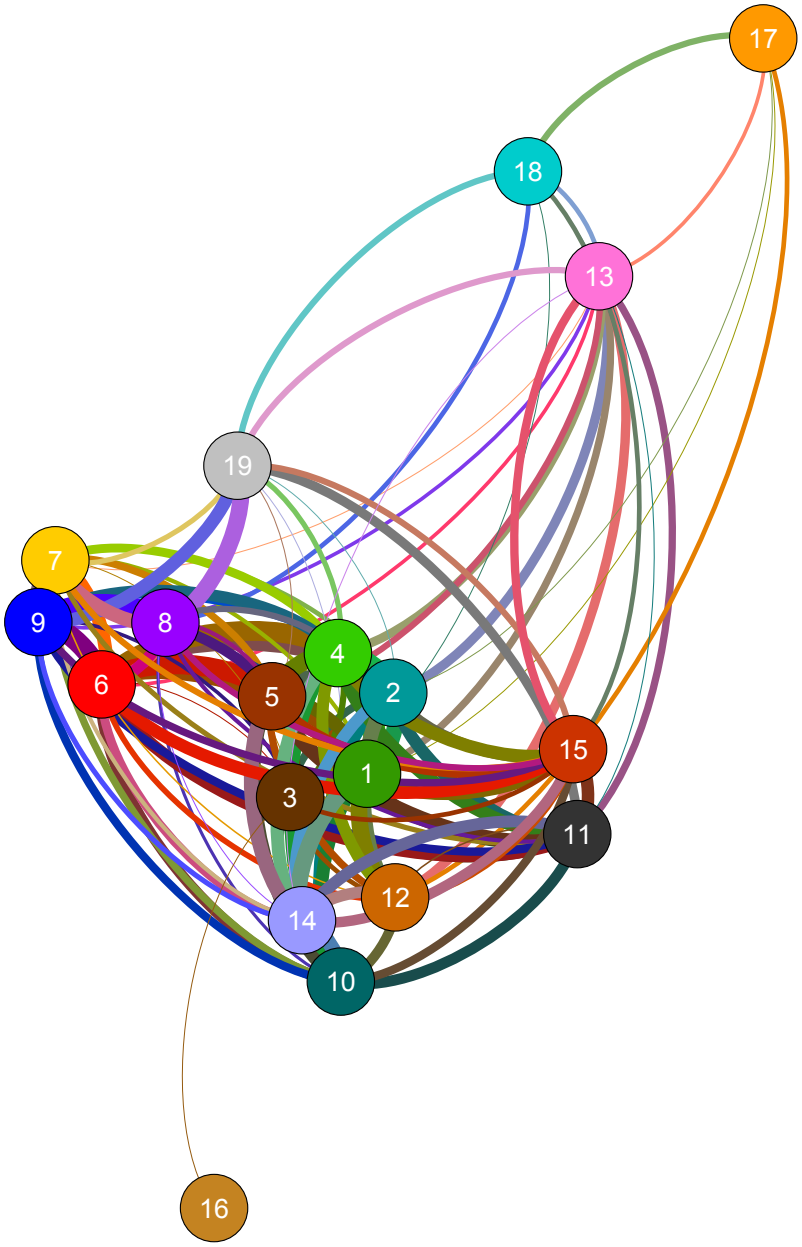


Figure 4.13: Network representation of the sector space showing all links with a proximity greater than 0.25. Links are weighted by their proximity value and coloured with a mix of the colours of the nodes they connect.

4.1 Spatial Network Model 1: A Single Urban Centre

In this section we introduce our first economic spatial network model which describes the growth and development of a single urban centre or isolated industrial cluster over time. We wish to model the changing sectoral composition and spatial distribution of economic activity in the given urban centre through the evolution of the relative sizes of firms in different sectors in the area, the entry of firms establishing new operations in the area and the exits of existing firms.

Our model is built on seven core concepts:

1. Firms considering establishing a new operation in the given urban centre or industrial cluster have taken into account the potential urbanisation economies available as a result of basing their new operation there.

2. Firms take into account the structure and composition of the existing economy in the given urban centre when evaluating a location for a new operation. New operations are established only if there are localisation economies available and/or the success of existing operations in the area signals to investors that the chosen location is a good one (demonstration effects).

3. The smaller the spatial distance between the chosen location for a new firm and an existing firm the more likely that both firms will benefit from localisation economies as a result of being located in the same area. This smaller spatial distance will also make it more likely that the existing firm will signal to the new firm that its chosen location is a good one. Firms are also more likely to benefit from localisation economies or be influenced by demonstration effect signals from firms in similar industry sectors.

4. Firms who take advantage of localisation economies and are established in a good location have the potential to grow larger.

5. Firms periodically reevaluate their links to the local economy, looking to take advantage of more synergies.

6. Firms grow towards an optimum size based on their ever changing environment.

7. Firms shut down their operations and exit the area if it becomes undesirable.

S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0	0.5621	0.7985	0.3711	0.6733	0.7340	1.0788	1.4697	1.1394	0.6733	0.8440	0.5447	0.9163	0.5447	0.7133	2.2073	1.3471	1.4697	1.5141
2	0.5621	0	0.8916	0.4780	0.5276	0.7340	1.0498	1.4271	1.0498	0.7765	0.6931	0.6162	0.8210	0.7133	0.7340	1.8971	1.3471	1.4271	1.3471
3	0.7985	0.8916	0	0.9676	1.0788	1.2730	1.3471	1.4697	1.1087	0.9163	1.2040	1.0498	1.6094	0.9416	1.0498	1.3471	1.4697	1.3471	1.7148
4	0.3711	0.4780	0.9676	0	0.5276	0.4463	0.8440	0.9676	0.7133	0.7340	0.6733	0.7133	0.8440	0.6733	0.6733	2.2073	1.5606	1.4697	1.0217
5	0.6733	0.5276	1.0788	0.5276	0	0.4463	0.9163	1.0788	0.7765	0.7133	0.6162	0.9163	0.9676	0.6162	0.7765	2.1203	1.7148	1.7148	1.2730
6	0.7340	0.7340	1.2730	0.4463	0.4463	0	0.6539	0.7340	0.4308	0.7550	0.6931	0.9943	1.1087	0.8440	0.6931	2.6593	1.8971	1.5141	0.7765
7	1.0788	1.0498	1.3471	0.8440	0.9163	0.6539	0	0.6349	0.6733	0.9163	1.0217	1.2379	1.2730	1.0788	0.9676	2.5257	2.2073	1.6607	1.0217
8	1.4697	1.4271	1.4697	0.9676	1.0788	0.7340	0.6349	0	0.4780	1.1394	1.0788	2.0402	1.3863	1.3471	0.9676	2.0402	2.8134	1.0498	0.5447
9	1.1394	1.0498	1.1087	0.7133	0.7765	0.4308	0.6733	0.4780	0	0.8675	0.8675	1.4697	1.1087	1.0217	0.9163	1.8326	2.9957	1.6607	0.6733
10	0.6733	0.7765	0.9163	0.7340	0.7133	0.7550	0.9163	1.1394	0.8675	0	0.7550	0.7985	1.3863	0.6162	0.8675	1.8971	1.6607	1.6607	1.6607
11	0.8440	0.6931	1.2040	0.6733	0.6162	0.6931	1.0217	1.0788	0.8675	0.7550	0	1.2379	0.9163	0.7133	0.6931	2.1203	1.8326	1.2730	0.8210
12	0.5447	0.6162	1.0498	0.7133	0.9163	0.9943	1.2379	2.0402	1.4697	0.7985	1.2379	0	0.8210	0.7340	0.9676	1.7720	1.0498	1.6607	2.1203
13	0.9163	0.8210	1.6094	0.8440	0.9676	1.1087	1.2730	1.3863	1.1087	1.3863	0.9163	0.8210	0	1.2730	0.8675	2.0402	1.1087	1.0498	0.9676
14	0.5447	0.7133	0.9416	0.6733	0.6162	0.8440	1.0788	1.3471	1.0217	0.6162	0.7133	0.7340	1.2730	0	0.7765	2.3026	1.7148	1.6094	1.3471
15	0.7133	0.7340	1.0498	0.6733	0.7765	0.6931	0.9676	0.9676	0.9163	0.8675	0.6931	0.9676	0.8675	0.7765	0	1.7720	1.4697	1.0498	0.9676
16	2.2073	1.8971	1.3471	2.2073	2.1203	2.6593	2.5257	2.0402	1.8326	1.8971	2.1203	1.7720	2.0402	2.3026	1.7720	0	1.5141	1.6607	2.3026
17	1.3471	1.3471	1.4697	1.5606	1.7148	1.8971	2.2073	2.8134	2.9957	1.6607	1.8326	1.0498	1.1087	1.7148	1.4697	1.5141	0	0.9676	2.6593
18	1.4697	1.4271	1.3471	1.4697	1.7148	1.5141	1.6607	1.0498	1.6607	1.6607	1.2730	1.6607	1.0498	1.6094	1.0498	1.6607	0.9676	0	0.9676
19	1.5141	1.3471	1.7148	1.0217	1.2730	0.7765	1.0217	0.5447	0.6733	1.6607	0.8210	2.1203	0.9676	1.3471	0.9676	2.3026	2.6593	0.9676	0

Table 4.3: Product distances ($d_P(a, b)$).

In our model we will segregate economic activities into the nineteen sectors listed in Table 4.1. In order to quantify the likelihood firms will benefit from localisation economies or be influenced by demonstration effect signals from firms in other sectors (concept (3) above) we introduce a new metric, the *product distance*. The product distance between two firms, f_i and f_j, denoted $d_P(f_i, f_j)$, is a measure of how different two firms are in terms of their industrial activities and is determined by what sectors the two firms' activities are in.

The examples in this chapter use product distances between sectors based on the proximities between sectors from the product space analysis by staff of the World Bank for the 16th Annual Conference on Global Economic Analysis given in Table 4.2 [44]. If f_i and f_j are firms where f_i's activities are in sector a and f_j's activities are in sector b then we take

$$d_P(f_i, f_j) = d_P(a, b) = \log(\phi_{a,b}). \tag{4.3}$$

These product distances are shown in Table 4.3.

4.1.1 Spatial Network Representation of the Economy

In this model the local economy is described by a single spatial network layer called the *industry space layer*. The industry space layer, $N_I = (V_I, E_I)$, is a spatial network representing the spatial distribution of firms in various industry sectors in, and in the vicinity of, an urban centre. Each node in the industry space layer $f \in V_I$ represents a firm. Two firms f_i and f_j in the industry space network are linked, i.e. $(f_i, f_j) \in E_I$, if they benefit from localisation economies as a result of being based close to one another or if one firm was influenced to establish their operation in the area by the demonstration effects associated with the success of the other. Links are weighted according to how similar the industrial activities of the two firms are. Pairs of firms in similar sectors establish strong links while pairs of firms in relatively unrelated sectors establish weak links.

The growth and evolution of the local economy is described by the growth and evolution of the industry space layer. This spatial network can grow and evolve in two ways:

1. Individual firms, represented by individual nodes, grow towards an optimum size based on the ever changing structure and composition of the local economy. Firms whose location is favourable and manage to take advantage of significant localisation economies will have a larger optimum size than other firms.

2. New firms may establish new operations in the area or existing firms may leave, firms may take advantage of more synergies by establishing new connections with other firms while other connections may be lost. This type of growth and evolution is described by the addition or removal of nodes and links to or from the industry space network. The probabilities of adding or removing nodes and links at a given point in time are based on the structure of the industry space at that time.

4.1.2 The Growth of an Individual Firm

The growth of individual firms in the industry space network is governed by a discrete time evolution operator known as a *map*. This map is based on the *network logistic map* introduced and analysed in Chapter 3. For each firm in the industry space $f_i \in V$ we consider the evolution of

$$x_i(n) = \frac{s_i(n)}{S_i}, \tag{4.4}$$

where $s_i(n)$ is the size of f_i at time n and $S_i > 0$ is a constant which scales the optimum size of the firm based on its sector. For an industry space network with m firms, for all $i \in \{1, 2, \ldots, m\}$, we set

$$x_i(n+1) = r_i(n)x_i(n)\left(1 - x_i(n)\right). \tag{4.5}$$

We then let the control function for our map $r_i(n)$ depend on the weighted degree of f_i at time n. This means that each individual firm's control function depends on the level of localisation economies experienced by the firm and the strength of the signal that the firm is located in a 'good' location. Formally, we set

$$r_i(n) = 1 + \rho w_i(n) = 1 + \rho \left(\frac{\sum\limits_{j:(f_i,f_j)\in E_I} l_{i,j}x_j(n)}{\sum\limits_{j=1}^{m} x_j(n)} \right) \tag{4.6}$$

where $\rho > 0$, $x_i(n)$ is the size of f_i at time n, $l_{i,j}$ is the weight of the link between f_i and f_j and $w_i(n)$ is the normalised weighted degree of f_i at time n.

4.1.3 Entry and Exit of Firms

We draw on the ideas of the spatial network development algorithms introduced in Chapter 2 to model the entry of firms to the area, the exit of existing firms and the evolution of links between firms in similar industries. We begin with some initial industry space network at the centre of a bounded space (in our examples this space will be the unit square) and proceed as follows:

1. A potential new firm's sector is chosen from a suitable probability distribution.
2. The location under consideration by the potential new firm is chosen from a uniform distribution on the unit square.
3. The potential new firm f_i is influenced by signals that this location is a good one by each existing firm f_e or recognises available localisation economies associated with choosing a location near to f_e, i.e. f_i and f_e form a link, with probability

$$\beta(\beta_i + \beta_e)\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right] e^{-\alpha_T(\alpha_i+\alpha_e)d_E(f_i,f_e)} e^{-\alpha_P d_P(f_i,f_e)}, \tag{4.7}$$

where $d_E(f_i, f_e)$ is the Euclidean distance between f_i and f_e. The parameter β is a positive constant which scales the probability of any such link forming. The parameters α_T and α_P are also positive constants that scale the overall importance of spatial distance and product distance to the formation of links between firms respectively. The parameters α_i and α_e reflect how important spatial distance is to the formation of potential links for the two individual firms, and β_i and β_e scale the probability of links forming for each firm. The values of these parameters are based on the firms' sectors. $\bar{x}_e(n)$ is the normalised size of f_e at the time of the addition of f_i to the industry space network and L is a constant such that $0 \leq L \leq 1$. If $L \neq 0$ firms are more likely to form links with larger firms, i.e. the demonstration effects and localisation economies associated with larger firms are stronger.

The link formed between the two firms is given a weight $l_{i,e}$ based on the product distance between the two firms, it is chosen from the uniform distribution on the interval $(0, \beta_m \phi_{a,b})$ where a and b are the sectors in which the firms f_i and f_e operate and $\beta_m = \min\{\beta_i, \beta_e\}$.

4. The potential new firm is established if and only if it forms a connection with at least one existing firm. If no connections are formed it is discarded.

5. Each time a new firm is established previously established links between all other firms are reevaluated. Weak links between firms are likely to only benefit the firms for a limited period of time. The weight $l_{\lambda,\gamma}$ of each link between a pair of existing firms f_λ, f_γ is compared to a threshold chosen from a uniform distribution on the interval $(0, \tau)$ where $0 \leq \tau \leq 1$ and is removed if the weight falls below this threshold. If at the end of this process any firm is entirely disconnected from the industry space it exits the area, i.e. it closes its operation in that location.

6. We allow firms to periodically attempt to increase their links to the local economy and take advantage of more synergies. Every T timesteps we allow connections to form between any pair of existing firms f_λ, f_γ with probability

$$B(\beta_\lambda + \beta_\gamma)\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_\lambda(n) + +\bar{x}_\gamma(n))}{\ln(2)}\right)\right] e^{-A_T(\alpha_\lambda+\alpha_\gamma)d_E(f_\lambda,f_\gamma)} e^{-A_P d_P(f_\lambda,f_\gamma)}. \tag{4.8}$$

The parameters B, A_T and A_P are, like β, α_T and α_P involved in the computation of P_1, positive constants which scale the probability of links forming between pairs of firms.

7. We update individual firm sizes according to equation (4.5) over G timesteps as each firm grows towards its optimum size based on its changing environment and position within the wider economy.

8. We repeat this process until the desired number of firms, M, are added.

4.1.4 Some Examples

In this section we present some examples showing the growth and evolution of firms and their relationships in different urban centres under varying conditions. Our first example shows the growth of a city of 250 firms whose industry space network is shown in Figure 4.15. Nodes are coloured by firm sector according to the table in Figure 4.14 and their size is given by relative firm size. The city has quite a diverse industry space with a strong manufacturing base to the west of the city where several mechanical, metallurgical and electrical manufacturing firms are based (blue nodes) alongside other manufacturing firms (red nodes). There is also a concentration of chemicals firms in the east of the city (purple nodes) where two distinct chemical industrial clusters are evident. These clusters are strongly connected, with firms benefiting from significant localisation economies and a desirable location. The advantages gained by operations located in these clusters have allowed them to grow towards a large optimum size.

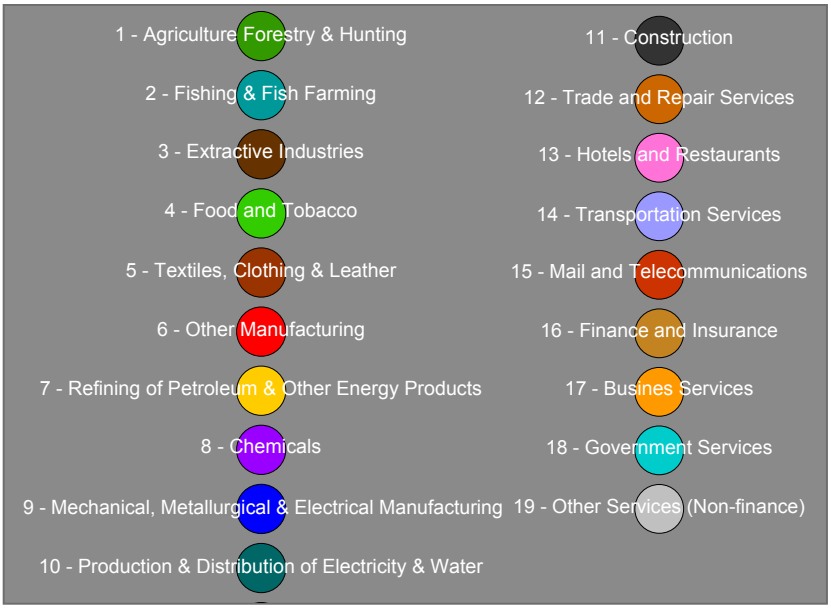


Figure 4.14: Sector colour codes.

Eighteen of the nineteen sectors listed in Figure 4.14 are represented. As is the case with most Irish urban areas, there is no activity in the refining of petroleum and other energy products. Economic activity in the city centre is dominated by trade and repair services (orange nodes), hotels and restaurants (pink nodes) and transportation services (indigo nodes). We see two small fishing operations (aqua nodes) on the eastern outskirts of the city and one small agricultural operation (dark green node) located on the northern outskirts of the city. Services operations are widely distributed throughout the city and in

particular there is quite a strong other services (including professional, technical and scientific services) sector, with 22 operations present (light grey nodes).

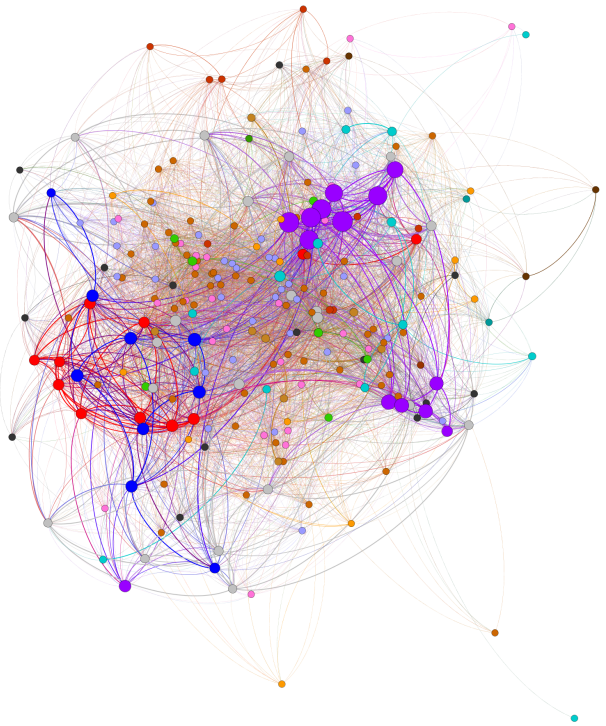


Figure 4.15: Network representation of the diverse industry space layer of a city with 250 firms.

Figure 4.16 shows the industry space network of the city shown in Figure 4.15 after adding 250 firms according to the algorithm presented in Section 4.1.3. Nodes in the industry space network representing firms are once again coloured by sector according to the table given in Figure 4.14, while node size is given by the relative size of the firm at the end of the simulation. 48 operations closed during the simulation period after becoming disconnected from the industry space. The locations of these operations are highlighted by white *ghost* nodes. In general we see that the strongest links have been formed between firms in the same sector or in similar sectors. We also note that in each sector the firm with the strongest links to the largest firms, i.e. the firm with the most desirable location taking advantage of the highest level of localisation economies, is generally the largest firm in that sector.

Examining Figure 4.16 in more detail we can observe many interesting features of the city's overall growth in terms of both the sectoral composition and the spatial distribution of its economic activities. For example we can see that the city's strong manufacturing base has continued to grow with much of this growth being concentrated near the existing operations in the west of the city. Many new operations have been established in both the mechanical, metallurgical and electrical manufacturing sector and the other manufacturing sector in this area with strong links to the existing operations.

Some chemicals firms have also been encouraged through demonstration effects and available localisation economies to establish new operations nearby. Most new chemicals operations, however, have been established near the two existing chemicals clusters in the east of the city where these effects are stronger. The cluster in the south east, which was initially smaller in terms of both firm size and number of firms, has seen particularly impressive growth. Firm numbers in the area have more than doubled from five firms to eleven firms. One would expect individual operations in this cluster to begin to rival operations in the northeastern cluster in terms of size in the future by establishing more links within the cluster and with the wider industry space increasing their optimum size and leading to growth.

We also see the emergence of a small cluster of three new firms in the other services sector (light grey nodes) strategically located approximately half way between the two large chemicals clusters. Recall that the other services sector, which includes technical and scientific services, is closely related to the chemicals sector according to our product distance metric based on the World Bank product space analysis (see Figure 4.13 and Table 4.3). In our model firms in more closely related sectors are more likely to influence eachother's location choices through demonstration effects and to benefit from localisation economies related to being based close to one another. For example in this case we could argue that scientific services sector firms and chemicals firms located in the same area are likely to have the opportunity to take advantage of potential input-output linkages. The location of this particular other services cluster allows the three firms in the other services sector to benefit from localisation economies related to being based close to eachother and close to firms in both chemicals clusters. Indeed we see that all three firms are connected and have also formed strong

links with multiple firms in both clusters.

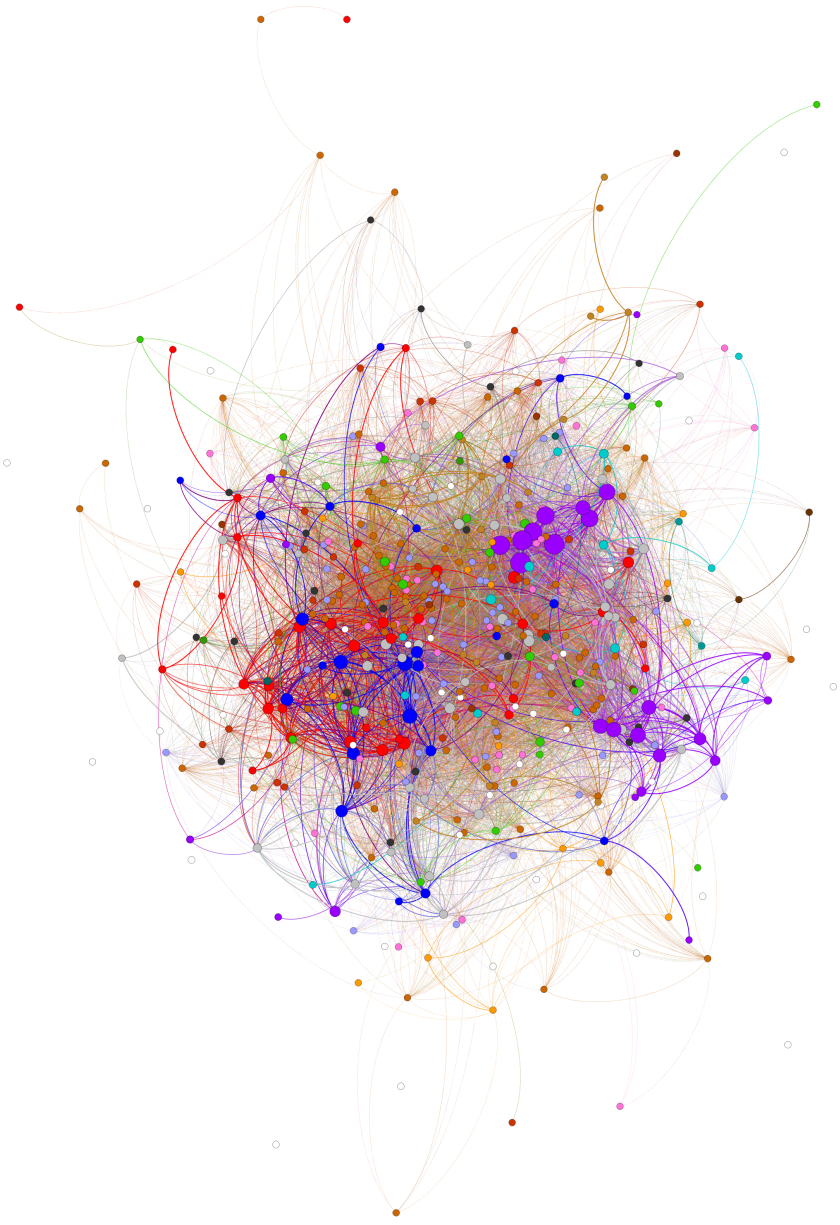


Figure 4.16: Network representation of the industry space layer of the city shown in Figure 4.15 after adding 250 firms according to the algorithm presented in Section 4.1.3.

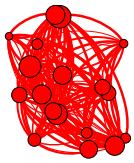
Our next examples investigate the effects of varying the importance of product distance during link formation, i.e. varying α_P and A_P, on the growth and evolution of an isolated cluster of firms. For simplicity in our examples we set $A_P = \alpha_P$. We consider the growth of three different clusters of nineteen firms, network representations of the industry space networks of each of these clusters are shown in Figure 4.17. Our examples represent three extremes in terms of the initial sectoral composition of the clusters and each has an identical, randomly generated, initial spatial distribution of firms.

The first cluster, N_I^C, is a homogenous cluster of nineteen firms in the other manufacturing sector, a sector at the core of our sector space with strong links to many other sectors (see Figure 4.13 and Table 4.2). The homogeneity of the cluster has allowed strong link to form between firms as they benefit from the localisation economies associated with being based close to one another and the favourable location. The second, N_I^D, is a diverse cluster of nineteen firms with one representative from each of the nineteen sectors we consider in our model. This cluster is quite sparse in terms of links between firms. Due to the heterogenous nature of their activities, many pairs of firms in this cluster do not benefit from one another through localisation economies or otherwise despite being based in close spatial proximity. The final cluster, N_I^P, is a homogenous cluster of nineteen firms in the finance and insurance sector, a sector which is on the periphery of our sector space with only weak links to the other eighteen sectors (see Figure 4.13 and Table 4.2). This cluster is again strongly connected due to its homogenous nature.

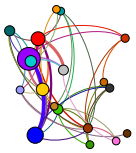
Choosing three different values of α_P, we allow each of these clusters to grow and evolve adding 19 firms according to the algorithm presented in Section 4.1.3. For our first example we choose a moderate value of α_P, $\alpha_P = 2.3$, i.e. the similarity between sectors of two firms is moderately important during link formation. This value was also used in our previous example in this section when examining the growth of a city with 250 firms. Network representations of the industry space networks of each of the three clusters shown in Figure 4.17 after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 2.3$ are shown in Figure 4.18. We observe different features in the growth and evolution of each of the clusters.

The first cluster, N_I^C, which began as a homogenous cluster of nineteen firms in the other manufacturing sector, sees slow diversification over time while continuing to expand its existing activities in the other manufacturing sector, establishing five new other manufacturing plants. This initial diversification is towards sectors which are closely related to the other manufacturing sector. The first operation established in the cluster with activities outside the other manufacturing sector is in the textiles, clothing and leather sector. This sector has strong links to the other manufacturing sector. According to our product distance metric it is the joint second closest sector to the manufacturing sector level with the food and tobacco sector and behind the mechanical, metallurgical and electrical manufacturing sector with a product distance of 0.4463 between the two sectors. The establishment of this operation and the existing other manufacturing sector operations encourages the establishment of more operations in

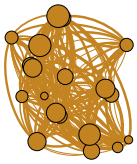
the textiles clothing and leather sector. Four of the nineteen firms added during the simulation were in this sector, three of which survived with one exiting after only forming weak links with the existing cluster.



(a) N_I^C: Network representation of an initial industry space network composed of 19 connected firms in the other manufacturing sector.



(b) N_I^D: Network representation of an initial industry space network composed of 19 connected firms, one in each of the 19 sectors listed in Table 4.1.



(c) N_I^P: Network representation of an initial industry space network composed of 19 connected firms in the finance and insurance sector.

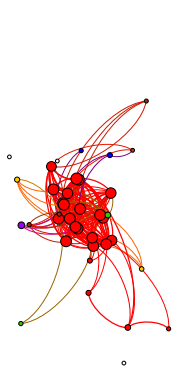
Figure 4.17: Network representations of three initial 19-firm industry space networks.

Other surviving new operations were also in sectors closely related to the other manufacturing sector including two in the most closely related mechanical, metallurgical and electrical manufacturing sector and two in the food and

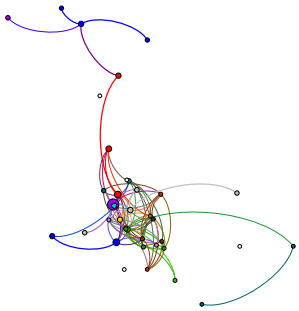
tobacco sector. Three of the new operations exited the cluster before the nineteenth was added. As previously mentioned one of these was in the textiles, clothing and leather sector, the other two were in sectors only distantly related to the other manufacturing sector, the hotels and restaurants sector and the other services sector with product distances of 1.1087 and 0.7765 to the other manufacturing sector respectively. These new operations formed weak links with the existing industry space at the outset and failed to strengthen their position before being removed.

The second cluster, N_I^D, which began as a very diverse cluster with each of the nineteen sectors represented by one of the nineteen firms saw a move towards specialisation in a few key sectors. Firms in some of the more peripheral sectors, such as the finance and insurance and business services sectors, which had only weak links to the cluster at the outset, closed their operations as this process of specialisation took place. Newly established operations on the other hand were concentrated in the closely related other manufacturing, mechanical, metallurgical and electrical manufacturing and chemicals sectors. Seven of the sixteen surviving new firms were in one of these three sectors. There was also strong growth in the other services sector with three new firms being established, perhaps benefiting from input-output linkages with the strongly growing manufacturing sector.

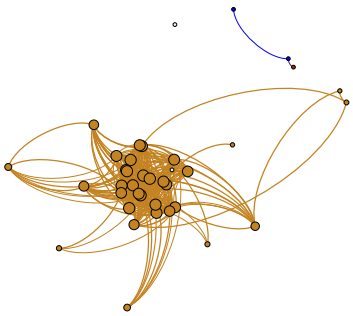
The final cluster, N_I^P, which began as a homogenous cluster of nineteen firms in the finance and insurance sector, saw very little diversification over the course of the simulation. As the finance and insurance sector is on the periphery of our sector space with only weak links to the other eighteen sectors, firms in other sectors were not strongly encouraged to locate in the cluster due to weak demonstration effect signals and the very limited potential localisation economies available. Only four of the nineteen new operations established were in sectors other than the finance and insurance sector and only three of those four firms survived. In fact, these three new firms did not even maintain a connection to the original industry space network. The original link between the three firms and the finance and insurance cluster was weak and short-lived and by the end of the simulation there were two separate components in the industry space network of the cluster. One of these components was composed of the original nineteen finance and insurance firms and the fourteen surviving new finance and insurance firms added during the simulation, the other, located to the northeast of the cluster was composed of the two mechanical, metallurgical and electrical manufacturing firms and a textiles, clothing and leather firm. Over time these firms have the potential to attract new firms in closely related sectors and to slowly begin to rival the dominant finance and insurance sector.



(a) Network representation of the industry space layer N_I^C shown in Figure 4.17a after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 2.3$.



(b) Network representation of the industry space layer of N_I^D shown in Figure 4.17b after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 2.3$.

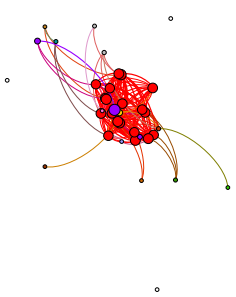


(c) Network representation of the industry space layer of N_I^P shown in Figure 4.17c after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 2.3$.

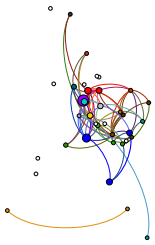
Figure 4.18: Network diagrams of the three 19-firm industry space networks shown in Figure 4.17 after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 2.3$.

For our second example we set $\alpha_P = 0$, meaning that product distance is not considered as a factor during link formation, in this case only the spatial distance between two firms influences the probability of a link forming between them. Network representations of the industry space networks of each of the three

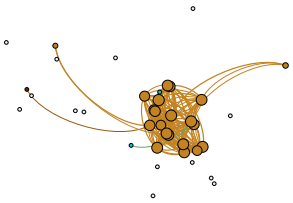
clusters shown in Figure 4.17 after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 0$ are shown in Figure 4.19.



(a) Network representation of the industry space layer N_I^C shown in Figure 4.17a after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 0$. (One firm which exited the cluster not pictured at this level of zoom.)



(b) Network representation of the industry space layer of N_I^D shown in Figure 4.17b after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 0$.



(c) Network representation of the industry space layer of N_I^P shown in Figure 4.17c after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 0$.

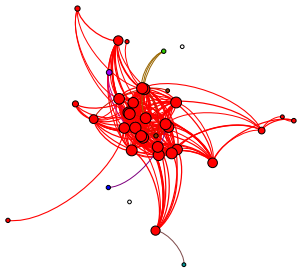
Figure 4.19: Network diagrams of the three 19-firm industry space networks shown in Figure 4.17 after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 0$.

In all three cases we see that for this value of α_P more firms exited the cluster during the course of the simulation than was the case for our moderate value. With $\alpha_P = 0$ there are low barriers to entry for firms in every sector, firms in sectors not well suited to the area are more likely to be established after forming weak links at the outset and quickly exit the area after failing to improve their situation. This could correspond to the situation where firms in all sectors are set up without due consideration of the sectoral composition of the existing local economy, leading to poor decisions and many failed startups.

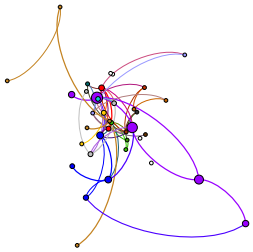
This effect is most pronounced in the case of N_I^P, new firms are quick to be established with $\alpha_P = 0$ but are likely to form only weak links with the existing firms in the peripheral finance and insurance sector. These links are later removed as they fall below the required threshold and the firms exit the area once more. Only firms added which were also in the finance and insurance sector or in the somewhat closely related government services sector survived until the end of the simulation in this case. The effect is less pronounced in the case of N_I^D, where each firm added has the potential to form a link to a firm in a sector identical to or closely related to its own sector as each of the nineteen sectors are represented from the outset. However, despite this we still see many new firms exiting before the end of the simulation as they are added after only forming weak links. The effect is least pronounced in the case of N_I^C. In this case the existing nineteen firms in the other manufacturing sector have the potential to form strong links with new firms in most sectors as the sector is at the core of our sector space and is closely related to many other sectors. As a result, despite not considering the sectoral composition of the existing economy during their establishment, many new firms manage to form at least one strong link and survive to the end of the simulation.

As in the $\alpha_P = 2.3$ case we again see the beginning of a process of diversification for N_I^C towards sectors closely related to the other manufacturing sector. Three of the surviving fifteen new operations are in the chemicals sector while two are in the other services sector. We also see evidence of specialisation in N_I^D again, with a move towards sectors at the core of our sector space, closely related to many other sectors. By the end of the simulation three of the surviving 27 firms were in the agriculture forestry and hunting sector and three in the mechanical, metallurgical and electrical manufacturing sector. Finally, we again see some slight diversification in the case of N_I^P. However, hampered by the peripheral nature of the existing economic activity in the area, startups in sectors other than finance and insurance find it difficult to form the necessary strong links with the local industry space needed to survive and soon exit once more. As a result the only other sector present in the cluster by the end of the simulation is the government services sector with two representatives.

For our final example we set $\alpha_P = 5$ corresponding to the case where product distance is an extremely important factor during link formation for firms in all nineteen sectors. Network representations of the industry space networks of each of the three initial clusters shown in Figure 4.17 after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 5$ are shown in Figure 4.20.



(a) Network representation of the industry space layer N_I^C shown in Figure 4.17a after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 5$.



(b) Network representation of the industry space layer of N_I^D shown in Figure 4.17b after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 5$.

(c) Network representation of the industry space layer of N_I^P shown in Figure 4.17c after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 5$. (One firm which exited the cluster not picured at this level of zoom.)

Figure 4.20: Network diagrams of the three 19-firm industry space networks shown in Figure 4.17 after adding 19 firms according to the algorithm presented in Section 4.1.3 with $\alpha_P = 5$.

For this high value of α_P we observe less diversification for both N_I^C and N_I^P during the course of the simulation than was seen for either the moderate $\alpha_P = 2.3$ case or the extreme $\alpha_P = 0$ case. In fact, in the case of N_I^P, a

homogenous cluster of firms in the peripheral finance and insurance sector, we see no diversification. All nineteen of the firms added during the simulation, including one which exited before the end were also in this sector. The peripheral nature of this sector in terms of our sector space made it very difficult for this cluster to attract firms in other sectors for this value of α_P. In the case of N_I^C we see some diversification towards firms closely related to the other manufacturing sector which is at the core of our sector space. However, this process of diversification is slower than for the other two values of α_P examined. Only six of the surviving firms at the end of the simulation are in sectors other than the manufacturing sector in the $\alpha_P = 5$ case compared to eleven in the $\alpha_p = 2.3$ case and fifteen in the $\alpha_P = 0$ case. We also see evidence of a stronger tendency towards specialisation in key areas in the case of N_I^D. At the end of the simulation fourteen of the 34 surviving firms were in the chemicals sector (5 firms), the finance and insurance sector (5 firms) and the mechanical, metallurgical and electrical manufacturing sector (4 firms).

In all three cases we see that, for this value of α_P, less firms exited the cluster during the course of the simulation than was the case for our moderate value. With $\alpha_P = 5$ there are high barriers to entry for firms in every sector, firms in sectors not well suited to the area are not likely to be established at the outset and only firms which have the potential to form strong links to the existing industry space are added. This corresponds to the situation where firms in all sectors are only established if the sectoral composition of the existing local economy gives strong evidence that they are likely to survive and have the potential to benefit from significant localisation economies.

4.2 Spatial Network Model 2: A Small Open Economy

Our second model extends the ideas of our first model in an attempt to model the growth and development of a small open economy composed of multiple towns and cities over time. As before, we wish to model the changing sectoral composition and spatial distribution of economic activity which is driven by the evolution of the relative sizes of firms in different sectors in the small open economy, the entry of firms establishing new operations and the exits of existing firms.

The model is built on seven core concepts which are closely related to those associated with our model describing the growth and development of a single urban centre:

1. Firms looking to establish a new operation in the country take into account the any urbanisation economies available and are more likely to consider locating a new operation in an area where more urbanisation economies can be taken advantage of.

2. Firms take into account the structure and composition of the existing economy in the vicinity of a potential location when evaluating its potential. New operations are established only if there are localisation economies available and/or the success of existing operations in the area signals to

the investors that the chosen location is a good one (demonstration effects).

3. The smaller the spatial distance between the chosen location for the new firm and an existing firm the more likely that both firms will benefit from localisation economies as a result of being located in the same area and that the existing firm will signal to the new firm that the chosen location is a good one. Internal transport infrastructure such as highways or railways effectively reduce the spatial distance between the areas they service. Firms are also more likely to benefit from localisation economies or be influenced by demonstration effect signals from firms in similar industries.

4. Firms who take advantage of localisation economies and urbanisation economies and have a good location have the potential to grow larger. Increased worldwide demand for products or services in a firm's sector also increase the firm's potential size.

5. Firms periodically reevaluate their links to the local economy, looking to take advantage of more synergies.

6. Firms grow towards an optimum size based on their ever changing environment.

7. Firms shut down their operations and exit an area if it becomes undesirable.

4.2.1 Travel Distance

In order to take into account the effects of the presence of internal transport infrastructure on the relationships between firms and institutions in our small open economy we introduce a new metric, the *travel distance*. The travel distance gives us the effective distance between two entities taking into account the efficiencies gained by using high quality internal transport infrastructure. In the examples we present only road infrastrucure is considered. We assume that the roads in the space are of varying quality and split them into two categories, the primary road network, consisting of national routes and motorways, and the regular road network, consisting of all other roads. We take travel along the regular road network as the base level of efficiency for travelling from one point to another and assume that travelling along the primary road network is more efficient.

Before formally defining the travel distance between two nodes in our model, let us first introduce the *road distance* which measures the effective distance between two points along the road network using at least one primary road. The *road distance* between two nodes in our model f and g is given by

$$d_R(f,g) = d_E(f,R_f) + \xi d_r(R_f,R_g) + d_E(R_g,g) \tag{4.9}$$

where R_f and R_g are the closest points to f and g on the primary road network respectively, $\xi < 1$ and $d_r(R_f,R_g)$ is the shortest distance along the primary

road network between R_f and R_g. We assume that agents will use the primary road network to travel between two points only if it reduces the effective distance between them. The *travel distance*, $d_T(f, g)$, between any two nodes in our model f and g is then given by

$$d_T(f, g) = \min\{d_E(f, g), d_R(f, g)\}, \tag{4.10}$$

where $d_E(f, g)$ is the Euclidean distance and $d_R(f, g)$ is the road distance between f and g.

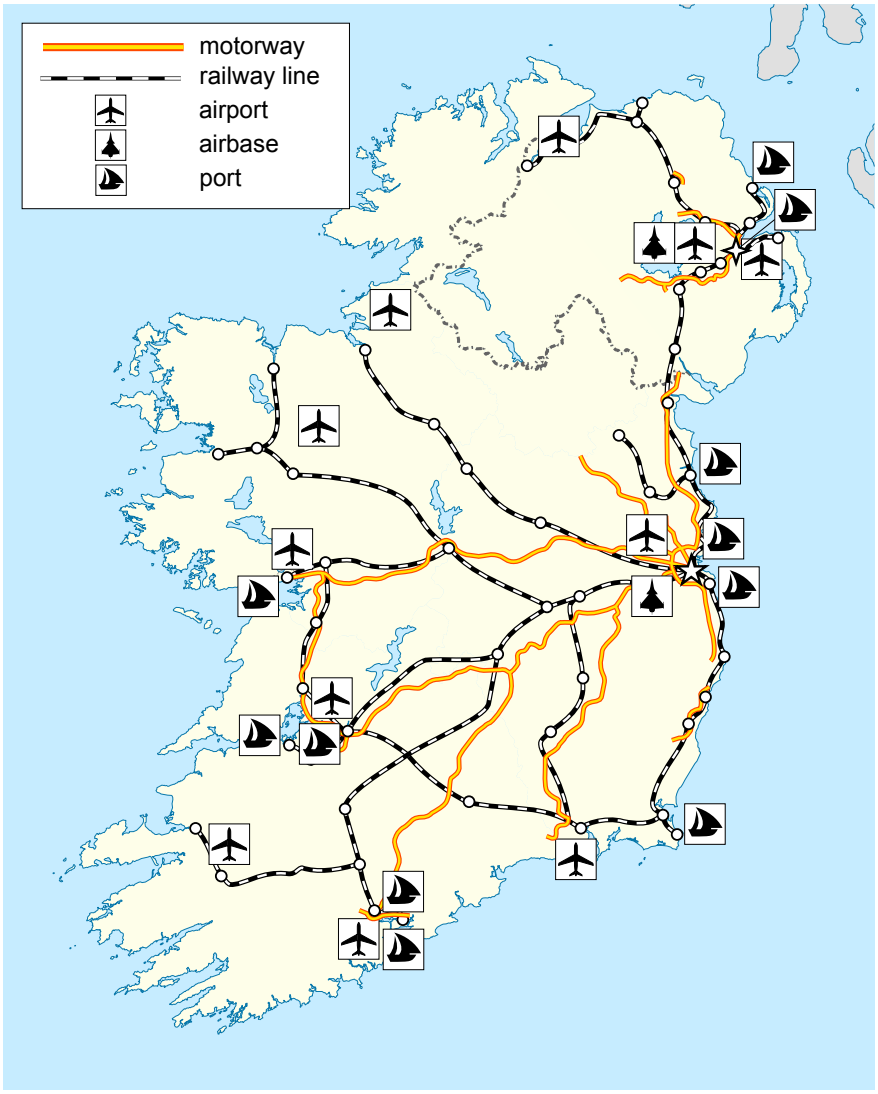


Figure 4.21: Map of Ireland's motorway network and other important transport infrastructure. [65]

The structure and layout of a country's internal transport links can be an important factor in the pattern of its spatial economic development. In many countries certain areas are served far better by national transport infrastructure than others, with the capital often being favoured. Ireland is a good example of this, the country's motorway network is composed of radial system of motorways linking the capital city Dublin with other cities and large urban areas, as can be seen from the map in Figure 4.21. Ireland's rail network also emanates radially from Dublin. For the examples later in this section we use the Irish motorway network as inspiration for the layout of our road network.

4.2.2 Spatial Network Representation of the Economy

In this model the wider economy is described by a multiple overlapping spatial network layers. These layers represent the spatial distribution of firms in the economy and the spatial distribution of sources of urbanisation economies and important infrastructure. For simplicity we will consider only one layer representing sources of urbanisation economies and important infrastructure, the *skill space layer*.

The skill space spatial network layer, $N_S = (V_S, E_S)$, is a spatial network which represents the spatial distribution of third-level educational institutions in the small open economy. Each node in the skill space layer $h \in V_S$ represents a third-level institution and has a graduate output vector $o_h(n)$ composed of the graduate output of the institution in ten disciplines listed in Table 4.4. We assume that graduates and researchers from different disciplines have different skills and research in different disciplines gives results which are more relevant to certain industry sectors than others. We assign each firm $f_i \in V_i$ a discipline-relevance vector d_i, which gives a weight between 0 and 1 to each discipline based on its importance to firms in that sector.

Discipline
Science
Social Science, Business and Law
Humanities and Arts
Education
Broad Programmes
Combined
Services
Health and Welfare
Agriculture and Veterinary
Engineering, Manufacturing and Construction

Table 4.4: Academic Disciplines.

Again, we are most interested by the growth and evolution of the industry space layer, $N_I = (V_I, E_I)$, a spatial network representing the spatial distribution of firms in various industry sectors throughout the economy and the links

between them. The growth and evolution of the local economy is described by the growth and evolution of the industry space layer. This spatial network can grow and evolve in two ways:

1. Individual firms, represented by individual nodes, grow towards an optimum size based on the ever changing structure and composition of the wider economy. Firms whose location is favourable and manage to take advantage of significant localisation economies will have a larger optimum size than other firms. Significant urbanisation economies will also boost a firm's optimum size. In this simplified model we will specifically consider urbanisation economies associated with third-level institutions such as those discussed in Section 2.4. Proximity to a third-level institution with activity in relevant disciplines in terms of graduate output and research is beneficial to a firm. (We assume that the level of graduate output in each discipline in an institution is proportional to its research activity in the same discipline.)

2. New firms may establish new operations in the area or existing firms may leave, firms may take advantage of more synergies by establishing new connections with other firms while other connections may be lost. This type of growth and evolution is described by the addition or removal of nodes and links to or from the industry space network. The probabilities of adding or removing nodes and links at a given point in time are based on the structure of both the industry space and skill space at that time.

4.2.3 The Growth of an Individual Firm

The growth of individual firms in the industry space network in this model is also governed by a map based on the network logistic map. For each firm in the industry space $f_i \in V$ we again consider the evolution of

$$x_i(n) = \frac{s_i(n)}{S_i}, \tag{4.11}$$

where $s_i(n)$ is the size of f_i at time n and $S_i > 0$ is a constant which scales the optimum size of the firm based on its sector. For an industry space network with m firms, for all $i \in \{1, 2, \ldots, m\}$, we set

$$x_i(n+1) = r_i(n)x_i(n)\left(1 - x_i(n)\right). \tag{4.12}$$

We then let the control function for our map $r_i(n)$ depend on three factors:

1. The level of localisation economies experienced by the firm and the strength of the signal from other nearby firms in similar industries that the firm is located in a 'good' location.

2. The level of benefit gained from urbanisation economies in the form of educational institutions.

3. The level of worldwide demand for products or services in the firm's sector.

We set

$$r_i(n) = \lambda_l r_i^l(n) + \lambda_s r_i^s(n) + \lambda_y r_i^y(n), \tag{4.13}$$

where

$$r_i^l(n) = 1 + \rho_l w_i(n) = 1 + \rho_l \left(\frac{\sum\limits_{j:(f_i,f_j)\in E_I} l_{i,j} x_j(n)}{\sum\limits_{j=1}^{m} x_j(n)} \right) \tag{4.14}$$

represents the level of localisation economies,

$$r_i^s(n) = 1 + \rho_s \left(\frac{\sum\limits_{j:h_j\in V_S} (d_i \cdot o_j) e^{-\alpha_S d_T(f_i,h_j)}}{\sum\limits_{j:h_j\in V_S} \mathbf{1} \cdot o_j} \right), \tag{4.15}$$

represents the level of urbanisation economies and

$$r_i^y(n) = 1 + \rho_y Y(n) \tag{4.16}$$

represents worldwide demand for products or services in the firm's sector. Here ρ_l, ρ_s and ρ_y are positive constants and λ_l, λ_s and λ_y are positive such that $\lambda_l + \lambda_s + \lambda_y = 1$, which reflect the relative importance of each of the three factors to the firm f_i based on its sector.

4.2.4 Entry and Exit of Firms

We again draw on the ideas of the spatial network development algorithms introduced in Chapter 2 to model the entry of firms, the exit of existing firms and the evolution of links between firms in similar industries. We begin with some initial industry space network in our small open economy and proceed as follows:

1. A potential new firm's sector is chosen from a suitable probability distribution.

2. The location under consideration by the potential new firm is chosen from a probability distribution generated by the skill space spatial network layer. Firms will consider locations for new operations close to third-level institutions with activity in relevant disciplines. For our examples we use a uniform-plus-truncated-Gaussian mixture distribution on our bounded space. For each new firm we define a uniform-plus-truncated-Gaussian mixture distribution with probability density function

$$p(x) = \Omega_0 u(x) + \sum_{h\in V_S} \Omega_h p_h(x). \tag{4.17}$$

The weights Ω_0 and Ω_h depend on the graduate output o of the institution h and the firm's sector. These weights are scaled such that $\sum_j \Omega_j = 1$. $u(x)$ is the probability density function of a uniform distribution on the space, while $p_h(x)$ is the probability density function of a truncated Gaussian mixture on the space with mean μ_h given by the spatial location of h and covariance matrix $\Sigma_h = \sigma^2 I$ where σ^2 depends on the firm's sector.

3. The potential new firm f_i is influenced by signals that this location is a good one by each existing firm f_e or recognises available localisation economies associated with choosing a location near to f_e, i.e. f_i and f_e form a link, with probability

$$\beta(\beta_i + \beta_e)\left[(1-L) + L\left(\frac{\ln(1+\bar{x}_e(n))}{\ln(2)}\right)\right] e^{-\alpha_T(\alpha_i+\alpha_e)d_T(f_i,f_e)} e^{-\alpha_P d_P(f_i,f_e)}. \tag{4.18}$$

Here the parameter β is a positive constant which scales the probability of any such link forming. The parameters α_T and α_P are also positive constants, these scale the overall importance of spatial distance and product distance to the formation of links between firms respectively. The parameters α_i and α_e reflect how important spatial distance is to the formation of potential links for the two individual firms and β_i and β_e scale the probability of links forming for each firm. The values of these parameters are based on the firms' sectors. $\bar{x}_e(n)$ is the normalised size of f_e at the time of the addition of f_i to the industry space network and L is a constant such that $0 \leq L \leq 1$. If $L \neq 0$ firms are more likely to form links with larger firms, i.e. the demonstration effects and localisation economies associated with larger firms are stronger.

The link formed between the two firms is given a weight $l_{i,e}$ based on the product distance between the two firms, it is chosen from the uniform distribution on the interval $(0, \beta_m \phi_{a,b})$ where a and b are the sectors in which the firms f_i and f_e operate and $\beta_m = \min\{\beta_i, \beta_e\}$.

4. The potential new firm is established if and only if it forms a connection with at least one existing firm. If no connections are formed it is discarded.

5. Each time a new firm is established previously established links between all other firms are reevaluated. Weak links between firms are likely to only benefit the firms for a limited period of time. The weight $l_{\lambda,\gamma}$ of each link between a pair of existing firms f_λ, f_γ is compared to a threshold chosen from a uniform distribution on the interval $(0, \tau)$ where $0 \leq \tau \leq 1$ and is removed if the weight falls below this threshold. If at the end of this process any firm is entirely disconnected from the industry space it exits the area, i.e. it closes its operation in that location.

6. We allow firms to periodically attempt to increase their links to the local economy and take advantage of more synergies. Every T timesteps we allow

connections to form between any pair of existing firms f_λ, f_γ with probability

$$B(\beta_\lambda+\beta_\gamma)\left[(1-L)+L\left(\frac{\ln(1+\bar{x}_\lambda(n)++\bar{x}_\gamma(n))}{\ln(2)}\right)\right]e^{-A_T(\alpha_\lambda+\alpha_\gamma)d_T(f_\lambda,f_\gamma)}e^{-A_P d_P(f_\lambda,f_\gamma)}. \tag{4.19}$$

The parameters B, A_T and A_P are, like β, α_T and α_P involved in the computation of P_1, positive constants which scale the probability of links forming between pairs of firms.

7. We update individual firm sizes according to equation (4.12) over G timesteps as each firm grows towards its optimum size based on its changing environment and position within the wider economy.

8. We repeat this process until the desired number of firms, M, are added.

4.2.5 An Example

In this section we present an example showing the growth and evolution of a small open economy over time in terms of the sectoral composition and spatial distribution of firms operating within its borders and their relationships. A map highlighting the layout of the country's primary road network (in blue) and the locations of third-level institutions (in pink) is shown in Figure 4.22 while a map highlighting the initial spatial and sectoral distribution of firms is shown in Figure 4.23.

The map in Figure 4.22 shows that the small open economy being modelled has a radial primary road network with roads emanating from its largest urban centre, perhaps the capital city, located in the east of the country. These roads connect other major urban centres, such as the second largest city located in the south and two small cities in the west of the country, to the capital city. Some smaller urban centres are not serviced by the country's primary road network including two towns in the north west of the country. This situation is very similar to the situation in Ireland which is mapped in Figure 4.21, where the motorway network has a radial layout with motorways emanating from the capital city Dublin in the east of the country.

The layout of third-level institutions in our examples was also inspired by the Irish case. Two large third-level institutions with a combined output of 4700 graduates per year are located in the capital in the east while most of the other large urban centres in the country are also home to a third-level institution. The largest of these other third-level institutions produces 3900 graduates per year and is located in the city directly to the west of the capital. One small town in the north west also hosts a small third-level institution producing 1000 graduates per year.

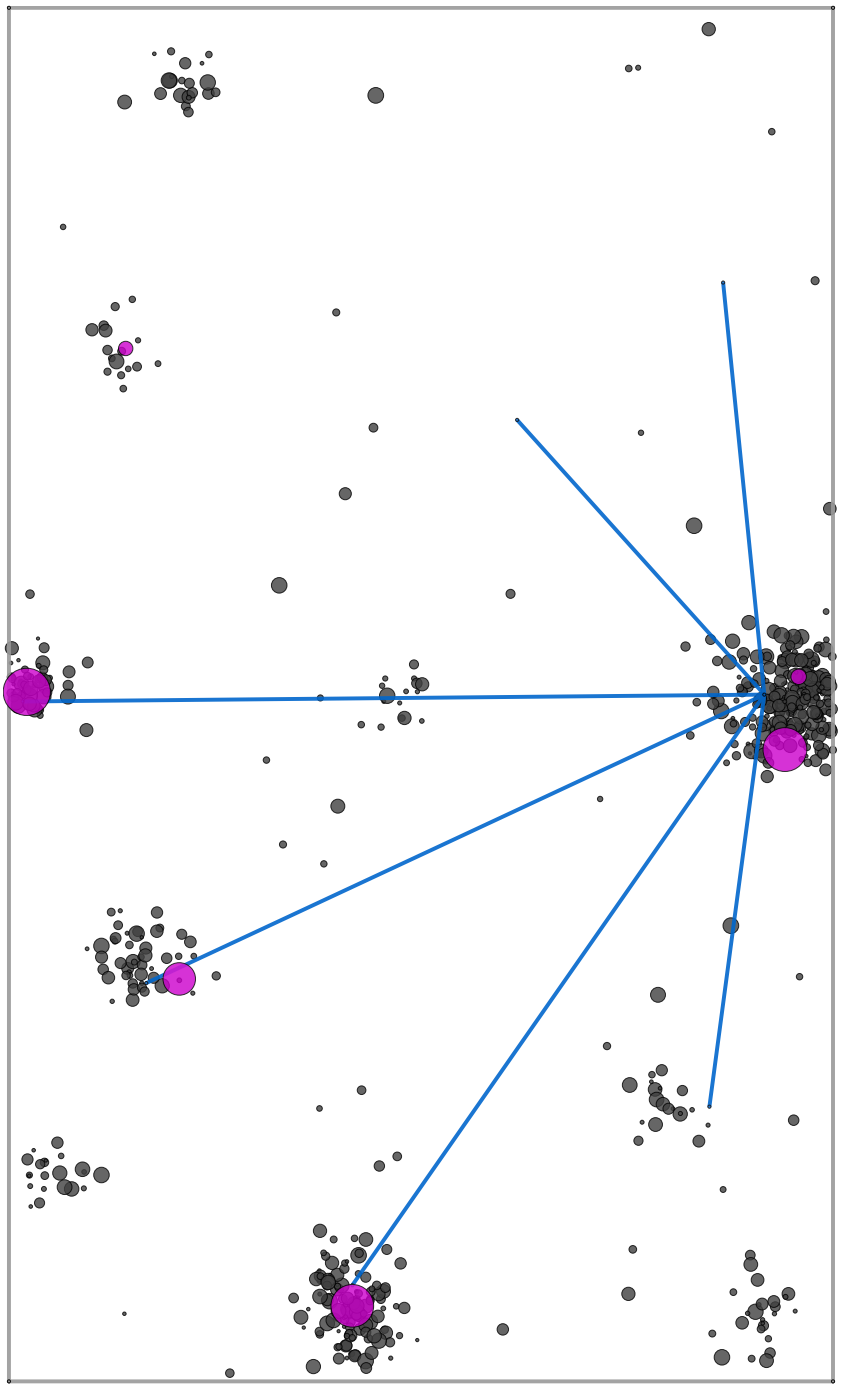


Figure 4.22: Map detailing the layout of infrastructure and firms in the small open economy. Third-level institutions are shown in pink and are scaled by their graduate output, the primary road network is shown in blue and firm locations are shown in grey.

The map in Figure 4.23 shows the distribution of the 615 firms initially present in our small open economy. As in the examples in Section 4.1.4 each node represents a firm and nodes are coloured by sector according to the colour codes given in Figure 4.14. Examining this map we can see that the country under consideration has a strong agricultural base with a widespread distribution of agriculture, forestry and hunting firms (dark green nodes) and ten urban centres of varying size and sectoral composition. In terms of firm numbers the trade and repair services sector (orange nodes), which includes both wholesale and retail trade, is quite naturally the leading sector, accounting for 19.22 per cent of firms, followed by the hotels and restaurants sector (pink nodes) accounting for 11.95 per cent of firms.

Some of the urban centres have strong industrial clusters of firms from one or more closely related sectors or display noticeable specialities in a few key sectors. For example, the economy of the most northerly small town in the northwest, which has no third-level institution and is isolated from the country's primary road network, is dominated by the trade and repair sector and the hotels and restaurant sector. Perhaps this town, supported by its isolation and the natural beauty of surrounding areas, relies on the tourist industry to drive its economy. The town in the south west which also has no third-level institution and is isolated from the primary road network is home to firms mainly in the food and tobacco sector. The economies of the country's larger towns and cities also display interesting sectoral compositions. For example, the capital city in the east, due to historical factors and supported by strong infrastructure, is strong in many areas. Notably it has a particularly strong finance and insurance sector with many large finance and insurance firms clustered in the city centre, no other city in the country is home to such a cluster. The second largest city in the south has a very strong chemicals sector concentrated near the northern fringe of the city (purple nodes) and the city directly to the west of the capital is home to several large mechanical, metallurgical and electrical manufacturing firms (blue nodes).

Figure 4.24 maps the industry space network of the small open economy shown in Figure 4.23 after adding 585 firms according to the algorithm presented in Section 4.2.4. Nodes in the industry space network representing firms are once again coloured according to their sector (see Figure 4.14), while node size is given by the relative size of the firm at the end of the simulation. 142 different firms closed their operations in the small open economy during the simulation period and the locations of these operations are marked by white *ghost* nodes. In terms of the general growth of the industry space we can immediately see that the strongest links between nodes have been formed between firms in the same sector or in similar sectors and that the longest links have been formed between nodes in separate urban centres connected by a primary road. We can also see that the majority of the new firms established during the simulation are located in, or in the vicinity of, existing urban centres.

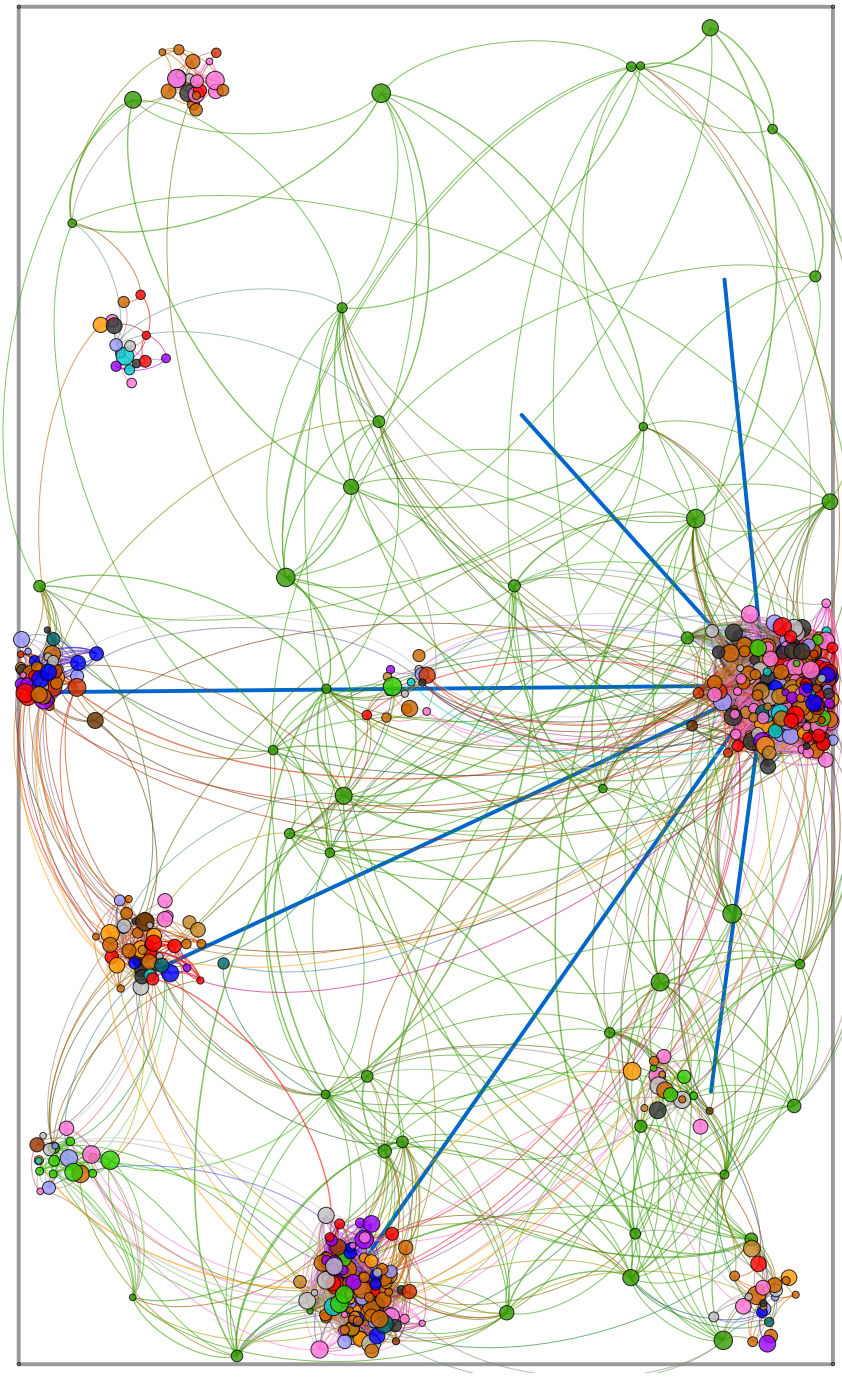


Figure 4.23: Network representation of the industry space layer of a small open economy with 615 firms.

Two notable exceptions to this are firms in the agriculture, forestry and hunting sector (dark green nodes) and the fishing and fish farming sector (aqua nodes). New firms in these sectors were established in rural areas more sparsely populated by other firms. These more remote locations are a result of two factors, the distribution from which locations for potential new firms in these two sectors are chosen and the potential of firms in these sectors to form links with distant. First, the high values of Ω_o and σ^2 in the uniform-plus-truncated-Gaussian mixture distributions from which we choose potential locations for a new firm f_γ in one of these sectors result in a more uniform spread of potential new locations for firms in these sectors. Second, the low values of α_γ associated with potential new firms f_γ in these sectors allow them to form links with other firms over long spatial distances.

Examining Figure 4.24 in more detail we can observe many interesting features of the country's overall growth in terms of both the sectoral composition and the spatial distribution of its economic activities. Towns and cities with large third-level institutions well connected by the primary road network to the rest of the country displayed the most growth.

The Capital City: In particular, the capital city in the east, home to two third-level institutions and connected to all of the other major cities in the country by the primary road network, saw strong growth in several sectors. The presence of two third-level institutions in this city help it to attract firms, especially those in more skill based sectors, to consider it as a potential location. When evaluating these potential locations investors are likely to be be encouraged by the demonstration effects and the potential localisation economies associated with being located close to strong existing firms to establish a new operation in the city. Since this city has a very diverse existing economy with strong firms in almost every sector this is true for firms in all sectors. The country's radial primary road network which emanates from the city makes it more feasible for new firms to form links with distant firms in rural areas and other towns and cities aswell as making it easier to form links with firms close to the network within the city itself.

The capital continued to experience strong growth in the finance and insurance sector. Thirteen new firms in this sector were established in the city during the course of the simulation with only three closing their operations during the same period. Growth in this sector in the capital far outstripped its growth in the rest of the country. Only nine new finance and insurance firms were established outside the capital during the simulation, five of these along with three of the initial finance and insurance firms based outside the capital had closed their operations by the end of the simulation.

Another area which saw particularly strong growth in the capital is the chemicals sector. At the outset the chemicals sector was not one of the standout sectors in the capital's economy. Despite its size only six firms were based in the capital (purple nodes), one of which was the smallest chemicals firm in the country, at the time. In contrast to this the second largest city in the south, which was far smaller in terms of firm numbers, was home to nine chemicals firms at the outset including the largest chemicals firm in the country.

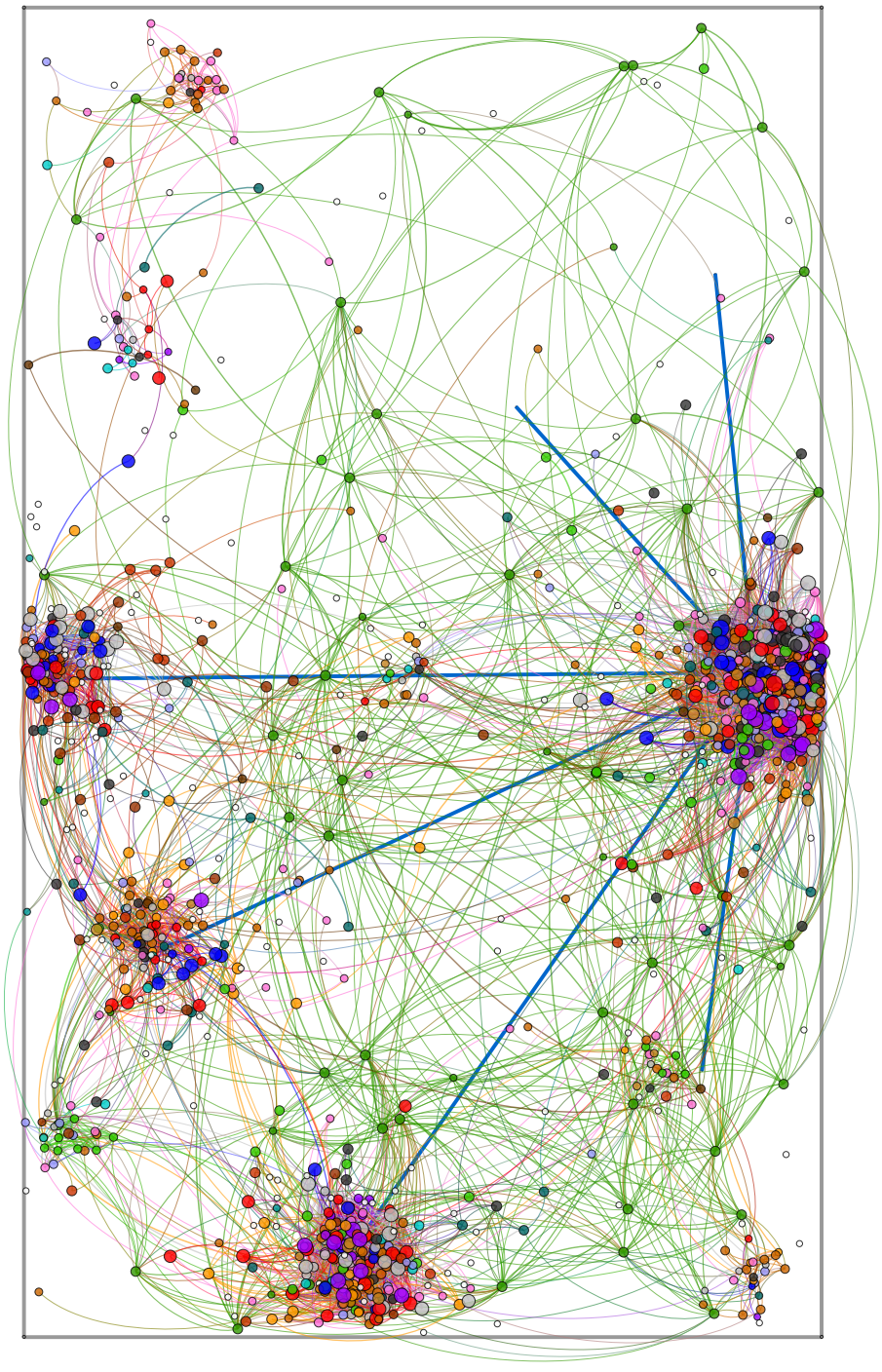


Figure 4.24: Network representation of the industry space layer of the small open economy shown in Figure 4.23 after adding 585 firms according to the algorithm presented in Section 4.2.4.

However, during the course of the simulation the capital experienced explosive growth in this sector due to the favourable conditions for firms the sector. Chemicals firm numbers in the capital more than tripled during the course of the simulation, fourteen new chemicals operations were opened in the city and there were no closures in this sector during the same period. One of the new operations also grew to become the largest chemicals firm in the country by the end of the simulation period. These firms were drawn to choose the city as a potential location by the presence of the two third level institutions outputting graduates and performing research in relevant disciplines. In fact, much of this growth was concentrated near to the larger of the city's two third-level institutions in the south of the city. 46 per cent of this institution's 3300 graduates graduate in either science or engineering, manufacturing and construction, two disciplines which, according to the parameters chosen for this simulation, are particularly relevant to firms in the chemicals sector. A smaller cluster of three chemicals firms was also established in the north east of the city close to the city's second third-level institution, 29 per cent of this institutions 1100 graduates were in either science or engineering, manufacturing and construction.

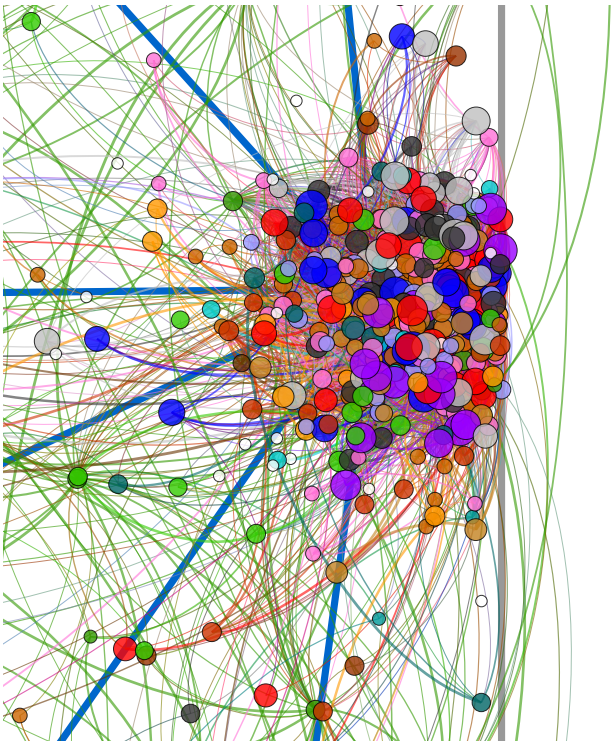


Figure 4.25: Magnified view of the industry space of the capital city at the end of the simulation period.

Many of the chemicals firms considering potential locations for new opera-

tions in the capital were reassured it was indeed a good location through demonstration effects associated with successful existing firms in related sectors such as the mechanical, metallurgical and electrical manufacturing sector or those associated with previously established chemicals firms. These firms and other may also have recognised potential localisation economies associated with being based close to such firms.

The Second City: The chemicals sector in the second largest city in the south of the country also continued to grow with ten new chemicals firms establishing operations in the area during the simulation period. The success of the chemicals sector in the city can be clearly seen in Figure 4.26. These firms were taking advantage of the urbanisation economies associated with the local third-level institution, from which 48 per cent of the 3500 graduates were in either science or engineering, manufacturing and construction, and the localisation economies associated with being based in a city in which many successful chemicals firms had already been established. Eight of these new firms were successful and grew to be among the leading chemicals firms in the country by the end of the simulation while the other two closed their operations. The continued success of the city's chemicals sector was an important factor in the growth of other sectors within the city including the other manufacturing sector and the other services sector which includes technical and scientific services. These two sectors are both closely related to the chemicals sector closely related to the chemicals sector according to our product distance metric based on the World Bank product space analysis (see Figure 4.13 and Table 4.3). Firms in these three sectors are likely to benefit from being based in the same location through factors such as technological spillovers and input output linkages.

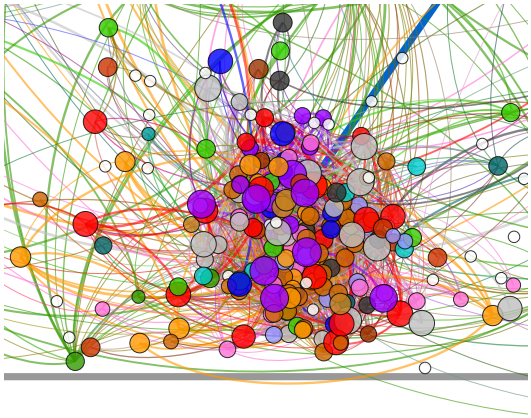


Figure 4.26: Magnified view of the industry space of the second city at the end of the simulation period.

The Western Cities: The city directly to the west of the capital city in the east grew strongly, attracting many firms in *'smart economy'*, in particular

firms in the chemicals sector (purple nodes), the other services sector (light grey nodes) and the mechanical, metallurgical and electrical manufacturing sector (blue nodes). The dominance of these sectors in the city's economy at the end of the simulation period is clearly visible in the map of its industry space shown in Figure 4.27. According to the parameters we have chosen for our model for such *'smart economy'* sectors being based close to a third-level institution which focuses on relevant disciplines is of major importance. As this city is home to the largest third-level institution in the country producing 3900 graduates each year it was well placed to attract investors in these sectors to consider it as a potential location for new operations.

Since these sectors are closely related according to our product distance metric, the city's strengths in sectors such as the mechanical, metallurgical and electrical manufacturing sector at the outset and growing strengths in all three sectors during the simulation period made it possible for firms considering the city to form strong links with existing related firms and establish themselves. Firms in these sectors were reassured through demonstration effect signals from the existing firms that the location was a good one and were further encouraged to locate in the city in order to take advantage of a pooled market for workers with specialised skills, input-output linkages and technological spillovers.

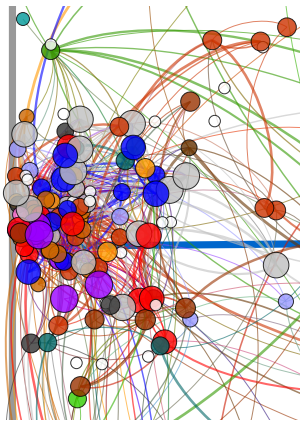


Figure 4.27: Magnified view of the industry space of the city directly to the west of the capital city at the end of the simulation period.

The city also experienced significant growth in the mail and telecommunications sector. Notably this growth included an entirely new industrial cluster on the north eastern outskirts of the city where three interconnected mail and telecommunications firms (red-orange nodes) were established during the simulation period. Two of the firms in this cluster also have links to a small second cluster growing to the west of the city. Firms in this cluster have taken advantage of its location next to the primary road network to form links with related firms in the city centre and in the small town in the midlands.

During the simulation period an important manufacturing cluster formed

near the third-level institution on the southern fringes of the city in the south west of the country as can be seen in Figure 4.28. Five new mechanical, metallurgical and electrical manufacturing firms and three new other manufacturing firms were established in the area during this time. These firms grew to become some of the largest in the city, their strategic location allowed them to take advantage of the localisation economies associated with being clustered together and the urbanisation economies associated with being based close to the third-level institution.

The business services sector was also an important growth area for the city's economy. Eight successful new firms in the business services sector (yellow-gold nodes) were established in the area during the simulation period. The city's third-level institution was a major factor in attracting these firms to the area. 35 per cent of this institution's 2600 graduates graduate in either social science, business and law or services, two disciplines which, according to our model, are the most relevant to firms in the business services sector. These firms benefited from the city's location forming links not only with other firms within the city but also with firms in the neighbouring cities closely situated to its north and south.

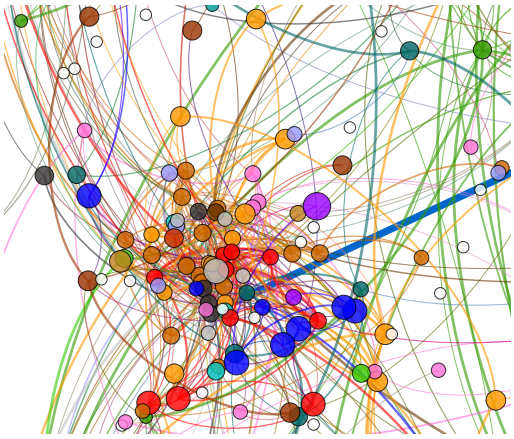


Figure 4.28: Magnified view of the industry space of the city to the south west of the capital city at the end of the simulation period.

Small Towns and Rural Areas: Smaller towns and rural areas saw less growth with firms in general preferring to locate in major urban centres where they could take advantage of more potential localisation and urbanisation economies. There was also a significantly lower density of links in the north west region in general with less interactions between firms over medium to long distances due to the lack of service by the primary road network in the area.

The country's most northerly town experienced slow growth working to its strengths by continuing to expand in the sectors related to the tourist industry. Figure 4.29 allows us a closer look at the town's industry space. Six of the seven new firms established in the area which were still open at the end of the

simulation period were in sectors likely to rely directly on tourism. Three of these new firms were in the hotel and restaurant sector (pink nodes), two in the trade and repair services sector (orange nodes) and one in the transportation services sector (indigo node). The other firm to establish a successful operation in the town during the period was a business services firm which established strong links with firms in both the hotel and restaurant sector and the trade and repair services sector perhaps benefiting from input-output linkages with these firms.

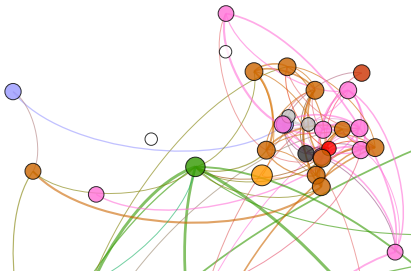


Figure 4.29: Magnified view of the industry space of the country's most northerly town at the end of the simulation period.

The town home to a third-level institution in the north west experienced the most significant growth of all the towns in the country during the simulation period. Much of this success in attracting new firms can be attributed to the presence of the small third-level institution. As can be seen in Figure 4.30 the town was successful in attracting firms in more skill-based sectors including two in the mechanical, metallurgical and electrical manufacturing sector (blue nodes) and two in the other manufacturing sector (red nodes). The existing cluster of three firms in the other manufacturing sector was also an important factor in this growth, three of these new firms formed links with firms in the cluster. The towns proximity to the city in the west was also a factor with the mechanical, metallurgical and electrical manufacturing in the south of the town forming an important link with a firm based there.

The town also saw growth in other areas. A new firm in the food and tobacco sector (lime green node) was established in the south east of the town and grew to be a major player in the local economy with significant links to two other manufacturing firms and several nearby firms in the agriculture, forestry and hunting sector, likely taking advantage of input-output linkages with the latter. Two new firms in the trade and repair services sector and one in the hotel and restaurant sector opened their doors for business for the first time during the simulation period while a third government services operation was established in the town with strong links to the existing two operations. Finally, two firms in the extractive industries sector (dark brown nodes) were also established nearby.

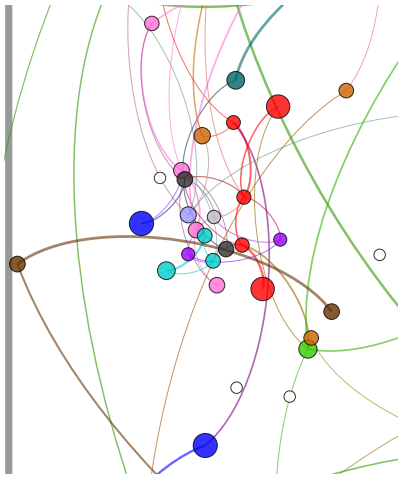


Figure 4.30: Magnified view of the industry space of the town home to a third-level institution in the north west of the country at the end of the simulation period.

Figure 4.31 shows the industry space of the town in the south west at the end of the simulation. Only five new operations were established in this town during the simulation period, a town which was dominated by firms in the food and tobacco sector (lime green nodes) at the outset. One of these firms, located in the town centre, was in the textiles, clothing and leather sector (brown node), a sector closely related to the food and tobacco sector according to our product distance metric. This firm formed strong links with a nearby firm in the food and tobacco sector and the existing textiles, clothing and leather firm in the town.

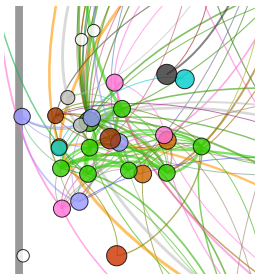


Figure 4.31: Magnified view of the town in the south west of the country at the end of the simulation period.

A transport services (indigo node) firm opened an operation to the west of the city forming strong links with the three existing operations in this sector located in the town, a new government services operation (cyan node) was established in the north of the city with a strong link to the existing operation and a mail and telecommunications firm (red-orange node) opened in the south. The only other firm to establish an operation in the town during the period was

a construction firm. This firm has no links with firms in the town itself, instead taking advantage of the towns proximity to the city just to the north it has two strong links to firms there.

The town in the far south east of the country did not prosper during the simulation period. The town saw the closure of an existing finance and insurance firm in the north of the town, marked by a white node in Figure 4.32, and only one of the new firms established in the town was still open at the end of the period. This firm was a trade and repair services firm located in the west of the town. The town's distance from all of the country's third-level institutions made it difficult for it to attract firms in many sectors to consider it as a potential location for a new operation. The town's sectoral composition did not make it a particularly attractive location for those sectors who were more likely to consider it as a potential location for new firms in terms of demonstration effect signals and localisation economies. This is in contrast to the country's most northerly town where, despite similar isolation, specialising in sectors for which close proximity to a third-level institutions is not vitally important related to the tourism industry allowed it to grow.

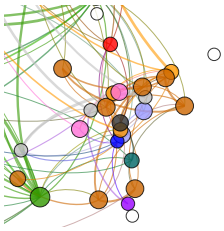


Figure 4.32: Magnified view of the industry space of the town in the far south east of the country at the end of the simulation period.

The town just south of the capital had similar problems also finding it difficult to attract new firms to set up in the area. However, its proximity to the primary road network and the capital helped it to be slightly more successful. A map of its industry space is shown in Figure 4.33. None of the firms based in the town at the beginning of the simulation period closed down while five new firms were established, two of which stayed open throughout the period and grew to become the two largest firms in the town. Both of these firms relied on links to firms in other urban centres in order to be successful. One of the firms, a mail and telecommunications firm (red-orange node) in the south of the town, took advantage of the town's location next to the primary road to the capital to form a strong link with a firm on the capital's southern outskirts. The new construction firm (dark grey node) in the west formed a link with a distant firm near the country's second city in the south of the country.

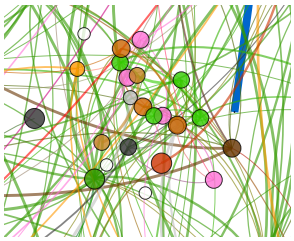


Figure 4.33: Magnified view of the industry space of the town south of the capital at the end of the simulation period.

The town in the midlands grew slowly, stretching along the primary road running through the heart of the town. The road, along with the town's location between the capital and the quickly growing city in the west, was a major factor in its growth. The town's chemical plant closed down during the simulation period however eight new operations were established in the area during the same period with five of those surviving until the end of the simulation. The largest growth sector for the town was the clothing, textiles and leather sector (brown nodes). Three new clothing, textiles and leather firms were opened, relying heavily on the primary road network for their success. Two of these firms chose locations to the west of the town and formed strong links to firms in the city in the west, the other chose a location to the east and made a strong link to a firm in the capital. The new business services firm (yellow-gold node) in the north of the town and the construction firm in the south also formed links with firms in other large urban centres with the support of the primary road network.

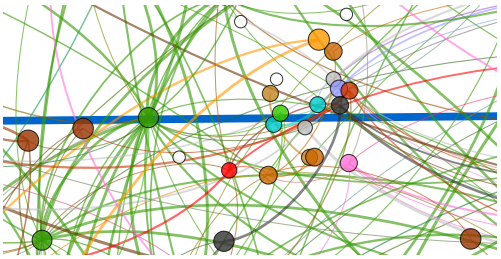


Figure 4.34: Magnified view of the industry space of the town in the midlands at the end of the simulation period.

Overall we saw that the spatial dynamics of economic growth and development in the small open economy were very uneven over the simulation period. Urban centres grew at different rates with the success of different areas in attracting new firms and the success of existing firms depending on a number of different factors including the existing sectoral composition of the local economy or access to nearby infrastructure and the interplay between those factors. As we only considered one source of urbanisation economies - proximity to third-level institutions, towns without a third-level institution grew very slowly. If

other infrastructure and institutions were taken into account such disadvantages are unlikely to have such a profound effect. We also saw that towns and cities tended to develop specialities in certain sectors or strengthen existing ones, these specialities varied from urban centre to urban centre depending on the attributes associated with its spatial location. In particular, a region, city or town was more likely to develop a speciality in an sector related to the existing strengths of its local economy in terms of the product distance metric and that could be catered to if necessary by a nearby third-level institution with activity in relevant disciplines.

5 Conclusions

Uneven spatial and sectoral economic dynamics are areas of considerable interest for policymakers. Our model provides a novel new approach to modelling such dynamics with the potential to give policymakers an insight into how to solve and analyse problems related to such uneven dynamics. For example our approach has the potential to provide promising insight into how it may be possible to promote balanced spatial growth in a country by stimulating growth in certain regions and sectors through policy measures. It could also perhaps provide in advance an insight into a country or city's likely future growth path in terms of spatial and sectoral distribution helping policymakers to provide for future for infrastructure requirements in advance.

In the Irish case for example the *"Expert Group on Future Skills Needs (EGFSN) has the task of advising Government on future skills requirements and associated labour market issues that impact on national potential for enterprise and employment growth"*, it pledges to *"discharge a central role in ensuring that labour market needs for skilled workers are anticipated and provided for"* [26]. Insights into the likely future of the sectoral composition and spatial distribution of Ireland's economy would certainly be of great interest to such groups. The promotion of growth outside the Greater Dublin Area to correct spatial imbalances in the Irish economy has been a major issue in Irish politics in recent decades. In 2002 the government of Ireland released the NSS, a twenty-year strategic plan designed to ensure *"more balanced social, economic and physical development between regions"* [32]. The NSS aimed to promote specialisation by different regions through the formation of specialised industrial clusters citing *"[s]patial clusters of international excellence [. . .] emerging in Ireland"* and seeking to *"strengthen these areas and increase their number by supporting the formation of self-sustaining clusters of economic activity"* . However, the strategy was unsuccessful and was scrapped in February 2013 and the issue of spatial imbalance in Ireland remains an important one.

In 2013 van Egeraat *et al.* called for *"more focused identification of existing regional strengths followed by appropriate support measures for further development of these strengths"* [67] while in May 2014 Jan O' Sullivan TD, Minister of State at the Department of the Environment, Community and Local Government announced that the government intended to finalise and put in place a new

national planning framework to replace the NSS by the end of 2015 [58]. The novel approach based on dynamic spatial networks presented in this chapter has the potential to help inform decisions within a framework such as this one. Our approach allows us to account for and to incorporate the multiple attributes associated with a given location. In our model the establishment of new firms and the closure or growth of existing firms is dependent on the structure of existing networks and transport links i.e. the interplay between socio-economic, political and historical factors in a given location are extremely important as we set out to address. Our examples have shown us that this approach is capable of producing interesting features of spatial and sectoral economic dynamics similar to features seen in the real world such as the unequal growth of urban centres, the unequal growth of economic sectors, the birth of new clusters of firms in similar sectors and the strengthening of existing clusters.

However, our model is far from perfect and any future work would look to address its shortcomings. The model has many parameters whose values are difficult to approximate, especially without access to the appropriate data, and requires detailed data regarding initial conditions in order to be useful. For example, for reliable results in a real world situation an accurate product distance measure and accurate data regarding firm size, sector and location would be very important. We feel that with access to the necessary data these problems could certainly be solved. Other issues arise from the fact that the model presented is extremely simplified. For example, the model neglects to account for any negative affects for a firm associated with being based close to other firms including competition for custom, employees or even higher land prices. A more comprehensive model would incorporate such effects to prevent the advantages of being based in a large city over a smaller one being overstated. Another issue arises from the fact it considers only third-level educational institutions as sources of urbanisation economies. A more accurate model would account for the impact of other important infrastructure such as ports, airports and power grids. Despite these issues we feel our approach is very promising. The framework of our model allows us to be adaptable and overcome such problems by adding extra overlapping dynamic spatial network layers and modifying relationships to represent and account for the features of the economy that we consider to be most important or wish to investigate.

Bibliography

[1] K. T. Alligood, T. D. Sauer, and J. A. Yorke. *Chaos: An Introduction to Dynamical Systems*. Springer-Verlag, New York, 1996.

[2] Irish Medical Devices Association. Medtech sector map. `http://www.imda.ie/Sectors/IMDA/Map.nsf/vPages/Medtech_sector~medtech-sector-map!OpenDocument`, 2015.

[3] Irish Pharmaceutical Healthcare Association. Pharmaceutical healthcare facts and figures 2012. 2012. URL `http://www.ipha.ie/alist/healthcare-facts-and-figures.aspx`.

[4] Higher Education Authority. Higher education statistics. `http://www.hea.ie/en/statistics/overview`.

[5] World Bank. Country: Ireland. `http://data.worldbank.org/country/ireland`, 2014.

[6] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[7] D. Barron. *Irish Times*, page 22, 7 November 2007.

[8] F. Barry. Irish economic development over three decades of EU membership. *Finance a Uver*, 53(9-10):394 – 412, 2003.

[9] F. Barry and J. Bradley. FDI and trade: The Irish host-country experience. *Economic Journal*, 107:1798–1811, 1997.

[10] F. Barry and C. van Egeraat. The eastward shift of computer hardware production: How Ireland adjusted. *NIRSA Working Paper Series*, 27, 2005.

[11] F. Barry and C. van Egeraat. The decline of the computer hardware sector: How Ireland adjusted. *Quarterly Economic Commentary - Spring*, 2008.

[12] F. Barry, H. Görg, and E. Strobl. Foreign direct investment, agglomerations and demonstration effects: An empirical investigation. *Review of World Economics/Weltwirtschaftliches Archiv*, 139(4):583–600, 2003.

[13] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 2009. URL `http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154`.

[14] T. Benham. Truncated multivariate normal. *Mathworks File Exchange*, 2012. URL `http://www.mathworks.com/matlabcentral/fileexchange/34402-truncated-multivariate-normal/content/rmvnrnd.m`. File ID: 34402.

[15] J. Bhagwati. Protectionism. In *The Concise Encyclopedia of Economics*. Library of Economics and Liberty, 2008. URL `http://www.econlib.org/library/Enc/Protectionism.html`.

[16] D. J. Bryce and S. G. Winter. A general inter-industry relatedness index. *Center of Economic Studies Discussion Papers*, pages 6–31, 2006.

[17] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10:186–198, 2009.

[18] P. Childs. The chemical industry in Cork. *Chemistry in Action*, 47, 1996.

[19] J. Culliton. A time for change: Industrial policy for the 1990s. In *Report of the Industrial Policy Review Group*. Stationery Office, Dublin, 1992.

[20] G. P. DeCoster and W. C. Strange. Spurious agglomeration. *Journal of Urban Economics*, 33:273–304, 1993.

[21] United Nations Statistics Division. UN Commodity Trade Statistics Database. `http://comtrade.un.org`.

[22] E. Doyle and C. Fanning. The role for clusters in Irish economic development policy. In *Perspectives on Irish Productivity*, pages 267–281. Forfas, Dublin, 2007.

[23] D. Duffy and E. Casey. Trends in Irish exports. *Quarterly Economic Commentary - Autumn*, 2012.

[24] K. T. D. Eames and M. J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20): 13330–13335, 2002.

[25] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57, 1985.

[26] EGFSN. Mandate of the Expert Group on Future Skills Needs for the period 2007 to 2009 as agreed by the Minister for Enterprise, Trade and Employment and the Minister for Education and Science. `http://www.skillsireland.ie/media/egfsn_mandate_2007.pdf`, 2007.

[27] L. Euler. Solutio problematis ad geometriam situs pertinetis. *Acadimae Scientiarum Imperialis Petropolitanae*, 8:128–140, 1736.

[28] M. J. Feigenbaum. Quantitative universality for a class of nonlinear transformations. *Journal of Statistical Physics*, 19(1):25–52, 1978.

[29] S. Galvin. The chemical industry in Ireland. *Chemistry in Action*, 54:9–16, 1998.

[30] M. T. Gastner and M. E. J. Newman. Optimal design of spatial distribution networks. *Physical Review E*, 74:016117, 2006.

[31] A. Gleeson, F. Ruane, and J. Sutherland. Promoting industrial clusters: Evidence from ireland. *IIIS Discussion Papers*, 89, 2005. URL `https://www.tcd.ie/iiis/documents/discussion/pdfs/iiisdp89.pdf`.

[32] Government of Ireland. *National Spatial Strategy 2002-2020: People, Places and Potential*. The Stationery Office, Dublin, 2002.

[33] M. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, 1985.

[34] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann. The product space conditions the development of nations. *Science*, 317:482–487, 2007.

[35] C. Hierholzer. Ueber die möglichkiet, einen linienzug ohne wiederholung und ohne unterbrechung zu umfahren. *Matamatische Annalen*, 6(1):30–32, 1873.

[36] P. Honohan. What went wrong in Ireland? *Report prepared for the World Bank*, May 2009. URL `http://www.tcd.ie/Economics/staff/phonohan/What%20went%20wrong.pdf`.

[37] M. Hoover. Location theory and the shoe and leather industries. *The Economic Journal*, 47(188):727–729.

[38] IDA Ireland. Ireland: A winning proposition for high value manufacturing. `http://www.idaireland.com/docs/publications/High-Value-Manufacting.pdf`, 2009.

[39] IDA Ireland. Medical technology in ireland. `http://www.idaireland.com/business-in-ireland/industry-sectors/medical-technology/`, 2014.

[40] International Financial Services Centre Ireland. Facts and figures. `http://www.ifscireland.ie/facts-figures`, 2015.

[41] M. Kaiser and C.C. Hilgetag. Spatial growth of real-world networks. *Physical Review E*, 69:036103, 2004.

[42] T. Kretts. Head of Paul Erdős, Budapest, Fall 1992. *Wikimedia Commons*, 2003. URL `http://commons.wikimedia.org/wiki/File:Erdos_head_budapest_fall_1992.jpg`. This file is licensed under the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation.

[43] E. Leamer. *Sources of Comparative Advantage: Theory and Evidence*. MIT Press, Cambridge, MA, 1984.

[44] H. Lofgren, I. Osorio-Rodarte, and M. Cicowiez. Bringing Product Space into CGE Modeling: Method and Preliminary Insights. `https://www.gtap.agecon.purdue.edu/resources/download/6564.pdf`, 2013. Presentation prepared for the 16th Annual Conference on Global Economic Analysis, Shanghai, June 12-14, 2013.

[45] A. M. Lyapunov. The general problem of the stability of motion. *International Journal of Control*, 55:531–773, 1992. An English translation by Taylor and Francis, London of the original 1892 work.

[46] A. Marshall. *Principles of Economics (8th edition, 1920)*. London, 1890. URL `http://files.libertyfund.org/files/1676/Marshall_0197_EBk_v6.0.pdf`.

[47] MATLAB. The MathWorks Inc., Natick, Massachusetts, 2010. URL `http://www.mathworks.com/products/matlab`.

[48] R. May. Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467, 1976.

[49] United Nations. Standard International Trade Classification. *Global Inventory of Statistical Standards*, 2011. URL `http://unstats.un.org/unsd/iiss/Standard-International-Trade-Classification.ashx`.

[50] F. Neffke and M. Svensson Henning. Revealed relatedness: Mapping industry space. *DRUID Working Paper Series*, (08-18), 2008. URL `http://www3.druid.dk/wp/20080018.pdf`.

[51] F. Neffke, M. Svensson Henning, and M. Ron Boschma. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. Papers in Evolutionary Economic Geography (PEEG) 0916, Utrecht University, Section of Economic Geography, October 2009. URL `http://ideas.repec.org/p/egu/wpaper/0916.html`.

[52] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, Oxford, 2010.

[53] T. O'Brien. Government scraps spatial strategy. *The Irish Times*, 12 February 2013. URL `http://www.irishtimes.com/news/government-scraps-spatial-strategy-1.1254981`.

[54] OECD. OECD NAIRU Estimates. `http://www.oecd.org/eco/outlook/2086120.pdf`, 2000.

[55] OECD. *OECD Economic Surveys: Ireland 2013*. OECD Publishing, 2013. URL `http://dx.doi.org/10.1787/eco_surveys-irl-2013-en`.

[56] Central Statistics Office. StatBank (CSO Main Data Dissemination Service). http://www.cso.ie/px/pxeirestat/statire/SelectTable/Omrade0.asp?Planguage=0.

[57] C. O'Mahony. What Olli Rehn said. *The Daily Business Post*, 12 October 2011. URL http://www.businesspost.ie/#!story/Home/News/What+Olli+Rehn+said/id/19410615-5218-4e95-8b39-824654078887.

[58] J. O'Sullivan. National Spatial Strategy Debate. http://oireachtasdebates.oireachtas.ie/debates%20authoring/debateswebpack.nsf/takes/dail2014051500040?opendocument, 2014. Speech given to Dáil Éireann, May 15, 2014.

[59] M. E. Porter. *The Competitive Advantage of Nations*. Free Press, New York, 1990.

[60] The Erdős Number Project. The Erdős Number Project Data Files. http://wwwp.oakland.edu/enp/thedata/.

[61] R. Segev, M. Benveniste, Y. Shapira, and E. Ben-Jacob. Formation of electrically active clusterized neural networks. *Physics Review Letters*, 90 (16):168101, 2003.

[62] European Mathematical Information Service. Publications in Graph Theory by Paul Erdős. http://www.emis.de/classics/Erdos/graph.htm.

[63] Unknown Source. Airlines Dataset. https://github.com/gephi/gephi/wiki/Datasets.

[64] The University of Chicago Theory Group, Department of Computer Science. Theory at U of C Salutes Paul Erdős. http://theory.cs.uchicago.edu/erdos.html.

[65] Tóraí. Ireland Transport Map. *Wikimedia Commons*, 2010. URL http://commons.wikimedia.org/wiki/File:Ireland_transport_map.svg. This file is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

[66] C. van Egeraat. Spatial concentration in the Irish pharmaceutical industry: The role of government intervention and agglomeration economies. *NIRSA Working Paper Series*, 28, 2006.

[67] C. van Egeraat, P. Breathnach, and D. Curran. Gateways, hubs and regional specialisation in the National Spatial Strategy. *Administration*, 60 (3):91–113, 2013.

[68] P. F. Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18:1–41, 1845.

[69] P. Walsh and C. Whelan. The importance of structural change in industry-for growth. *Journal of the Statistical and Social Inquiry Society of Ireland*, 29:1–32, 2000.

[70] D. J. Watts. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, Princeton, 1999.

[71] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences of the United States of America*, 99(21):13382–13386, 2002.