Graph-based Cellular Automaton Models of Urban Spatial Processes

David Bernard O'Sullivan

The Bartlett School of Graduate Studies
Bartlett School of Architecture and Planning
University College London
University of London

A thesis submitted for the degree of Doctor of Philosophy

Abstract

The impact of spatial organisation on the dynamics of spatial change is a topic of great interest in geography. The impact of spatial form is of particular interest in the context of urban systems where the associated issues are most acutely felt.

Various concepts of space and their implications are reviewed. Two existing mathematical concepts — graphs and cellular automata — which have been used as representations of spatial systems, are discussed. Examination of the use of these concepts reveals that in urban models they have usually been applied separately to represent, respectively, structural properties, and the behaviour of dynamic processes.

It is suggested that a model which combines aspects of both approaches — the graph-based cellular automaton (*graph-CA*) — can be introduced. This enables research into relations between the spatial structure and process dynamics of systems. Conceptual tools for this investigation are discussed, drawing on previous work in cellular automata and graph theory. A measure of spatial pattern — *spatial information* — useful for characterising dynamic behaviour, is developed. The computer implementation of these concepts is described, and experiments conducted using it are reported. A tentative finding is that process dynamics can be usefully distinguished according to the extent to which they are *robust* or *fragile* under spatial change.

The application of the graph-CA model to a particular urban process gentrification is also described. Theories of gentrification are reviewed, and a model for the phenomenon developed. Rather than develop a detailed empirical model of the process, a single urban setting (in Hoxton in East London) is used to generate a number of different graph-CA models. The similarities and differences in the behaviour of these are reported and implications discussed. Attention is drawn to possible extensions of this work throughout.

For Gill with love

Acknowledgements

Like most research, this thesis owes a great deal to the environment in which it was conceived. The most heartfelt thanks must therefore go to Mike Batty for fostering the collaborative, industrious and stimulating environment that the Centre for Advanced Spatial Analysis (CASA) has become, and for his enthusiastic encouragement as my supervisor. I also thank Bill Hillier whose contributions as a supervisor have always been thought-provoking.

Thanks are also due to collaborators, colleagues and friends at CASA: Muki Haklay, Bin Jiang, Mark Thurstain-Goodwin, Paul Torrens, Alasdair Turner, Alex Aurigi, Joanna Barros, Ruth Conroy, Jake Desyllas, Martin Dodge, Simon Doyle, Chiron Mottram, Sanjay Rana, Sarah Sheppard, Naru Shiode and Andrew Smith. A special word of thanks also goes to Nick Green and Steve Marshall for alcohol-fuelled evenings of inspiration.

I also thank my parents Mary and Denis for their constant support and encouragement. And the final word — as always — must go to my ever-tolerant, ever-loving wife Gill... who got me into this mess in the first place, but without whom I'd have no reason to get out of it.

Contents

List of Figures

List of Tables

Part I

Contexts

Chapter 1

Introduction

> "It is only shallow people who do not judge by appearances. The true mystery of the world is the visible not the invisible..." (Oscar Wilde, quoted in Keiller 1999, page 5)

1.1 Overview

The 'urban environment' now extends across distance and time to encompass (almost) the globe. The flows of financial information and of capital across time-zones and continents are managed from only a few world cities. Much of the world's wealth originates outside cities, in oil, mineral and agricultural resources, yet inexorably finds it way to them, so that perhaps fifty or a hundred cities worldwide dominate world markets, economies, politics and futures. The dramatic increases in urban sizes witnessed in the old world, in the nineteenth and twentieth centuries, are being played out on a scale orders of magnitude greater in the developing world today. And just as the industrial city dominated the twentieth century, the 'informational city' looks set to dominate the twenty-first. Even as that transformation changes the meaning of the word 'city', there is little sign that the 'new economy' is going to make much difference to the pre-eminence of densely populated urban regions, across many spheres of human activity.

Yet, there is something reassuringly everyday about cities still. How a city functions from minute-to-minute, day-to-day and year-to-year depends as much on how it is organised in space and time, as on its position in the global system. Even a global city must touch down on earth in a real location, in real buildings and streets. How well those buildings and streets work to promote and sustain the activities of urban life is vital to the welfare of any city. The relationships between the city as a single coordinated whole, with a role in some wider regional, national or global order, and the mundane pieces of 'stuff' — bricks and mortar — which constitute its visible elements are elusive, and a constant concern of urban planners and architects. The effects which form has on function, the visible on the invisible, the *spatial structure* on *system dynamics* are important concerns for those who wish to design cities. Moudon (1997, page 11) gives a good indication of the approaches which will concern us here:

> "1. Urban form is defined by three fundamental physical elements: buildings and their related open spaces, plots or lots, and streets.
>
> "2. Urban form can be understood at different levels of resolution. Commonly four are recognised, corresponding to the building/lot, the street/block, the city and the region.
>
> "3. Urban form can only be understood historically since the elements of which it is comprised undergo continuous transformation and replacement.
>
> "Thus form, resolution, and time constitute the three fundamental components of urban morphological research."

These issues — at this scale — increasingly concern geographers too, as new technologies and rich data sources enable new ways of exploring and attempting to understand these relationships at scales from the local to the global.

The present work seeks to explore these elements and their relationships by combining two mathematical tools — graphs and cellular automata (CA) — to represent different aspects of change in cities. The resulting model is a suitable vehicle for investigating the relations between aspects of the visible and invisible in cities, understood as *structure* and *dynamics.* The approach also enables exploration of the effects of spatial structure on spatial dynamics in a general sense, and this is the major focus of the thesis.

Broadly speaking, a graph (or network) representation can capture many of the structural properties of urban spatial arrangement. These structural properties give rise to patterns of accessibility, and as a result of land use. Demand for development in some areas — and decay or disuse in others — is powerfully conditioned by such spatial patterns, and gives rise to the pressures which lead to urban change. A graph representation consists of some partition of the city into spatial elements. These elements are variously related to one another. The graph has the elements as vertices ('nodes') and the pattern of relations as edges ('links'). Graph theoretical measures of the resulting representation allow its structural features to be determined and described with some precision. This sort of approach appears repeatedly in the urban morphology literature. Alexander (1964, 1965), Atkin (1974*c*), Hillier & Hanson (1984), Krafta (1994, 1996), Krüger (1979*a*,*b*), Martin & March (1972) and Steadman (1983) all provide examples of the use of graph-based structural measures to accurately represent or describe various aspects of urban spatial structure, understood at the micro-scale of the elements to which Moudon refers. However, these methods remain resolutely static.

Cellular automata are simple — usually grid-based — formal systems, in which dynamic change is represented grid-cell by grid-cell, as a simple deterministic mapping from the current state of a cell and its neighbours to the state which the cell will change to next. A CA model of a city focuses on dynamics and can be used to investigate urban dynamics and processes of change (Batty & Xie 1996, Couclelis 1985, Xie 1996). This has become an increasingly common modelling approach, and has been used to some effect in post-dicting the growth of urban areas (Clarke, Hoppen & Gaydos 1997, White & Engelen 1997, are useful, near operational examples). Among the many advantages of CA models are the simplicity of their construction, their evocation of the relationship between local and global changes, and the fact that they can achieve all this without calling on the whole apparatus of gravity models, spatial interaction models, and the like (Batty 1976*b*, see). In most cases, however, a CA representation does not attend to the detail of the urban spatial arrangement, and depends on the introduction of layers of abstraction between elements of the visible urban fabric, and the grid cells required by the formalism.

A relaxation of the rather strict conditions which define a cellular automaton has been proposed for geographic and urban modelling (Couclelis 1997, Takeyama & Couclelis 1997). Such 'relaxed' representations do not require that cells or neighbourhoods be identical, as in a grid. It is a rather obvious observation (not often made) that the relations between the cells in such a model constitute a graph. I therefore develop the non-grid based CA approach into *graph-based cellular automata* (graph-CA or GCA), so that urban spatial structure and process of dynamic urban change can be explored together, and in relation to one another. A central concern of this thesis is the use of this approach to search for relationships between *overall* system structure, and *overall* system dynamics. This can be seen as a contribution to general ideas on spatial dynamics, and as a new angle on persistent concerns in geographical modelling with understanding the general relationships which may hold between spatial structure, spatial dynamics, and spatial outcomes. This investigation is pursued at a general level in part II and in relation to a specific model of a particular urban process in part III.

This same framework might also enable a more systematic development of numerous variants of the CA theme in urban modelling, in a less *ad hoc* manner than hitherto. In time this may allow more systematic thinking about the relationships between the properties and behaviour of wholly abstract cellular urban models and their increasingly distant near-operational cousins. The potential for developing detailed CA models which are 'aware' of their spatial structure (through graph theoretic measures) is also promising. After all, the wholly local character of evolution in strict CA is somewhat unconvincing — since the agents of change in urban settings, property developers, retailers, commercial interests and city wide authorities are usually free to effect change anywhere in a city as they see fit, and are not constrained solely by local considerations. By building detailed graph-CA representations of the urban structure, it may become possible to develop more subtle models of urban processes than has been attempted hitherto. A possible benefit of this approach is that regions within a city modelled in this way can be structurally described and related to each other. That is, similar regions can be identified and examined to see if they evolve in similar ways. This is one aspect of an exploration of the relationship between spatial forms or structure and system dynamics which is enabled by this model framework, although it is not a focus of this thesis. Possibly, subtle formations not previously described could be found. This in turn might allow the development of new theoretical ideas about development processes in cities.

Such extensions of the basic concept are for the future. Returning to the present work, I seek to investigate the possibility of building graph-based CA model of urban development. The background material — on spatial concepts and models, on graphs and cellular automata, and on urban morphology — is presented in part I in chapters 2, 3, and 4, where relevant ideas are identified and explained. Part II introduces and formalises the graph-CA model, and develops methods and concepts for its investigation in chapter 5. In chapter 6 an implementation which has been developed is described, and chapter 7 reports some experiments which demonstrate a methodology for exploring spatial dynamics using these ideas and tools. In part III this modelling environment is further developed and its application to the gentrification phenomenon is described over four chapters dealing respectively with the theoretical background to gentrification, model exploration, a description of the case study area, and a description of model outcomes. The thesis concludes in chapter 12 with a review of many of the findings of this work, and a discussion of some of their implications.

1.2 Detailed outline of the thesis

1.2.1 Part I: Contexts

Chapter 2 considers the 'problem' of space — what it is philosophically, and practically — and how it has been represented in geography, and in geographical information systems (GIS) and geographical information science (GISci), in particular.[1] The intention here is to establish a theoretical and philosophical basis for the current work. The chapter finishes with a consideration of more recent treatments of space, particularly *proximal space*. This lays the foundations for a connection to be established between graph representations and cellular automata as different spatial representations.

Chapter 3 is an overview of both graphs and cellular automata which are classic representations of any (not necessarily spatial) system. Particular attention is drawn to how these alternative representations may shed light on, respectively, the 'static' (structural) and 'dynamic' (emergent behavioural) aspects of situations.

Chapter 4 reviews the micro-scale, predominantly technical, urban morphology and modelling literature to give an idea of the ways in which different representations of space and spatial relations have been used to understand city form and function. Consideration of this literature demonstrates how the static and dynamic views of urban form have often been separate, and related to, respectively, the architectural/planning and geographical/historical approaches to the subject. This further reinforces the message of chapter 3 concerning the often separate views taken of spa-

[1] I have tried to use these two abbreviations systematically throughout this thesis, but it is sometimes hard to tell which is more appropriate in particular cases. Additionally, whereas 'GIS' generally refers to 'GIS technology', GISs may be use to indicate 'various geographical information systems'.

tial form as distinct from process.

1.2.2 Part II: Model development

Building on the material in the previous chapters, chapter 5 describes the proposed new spatial model, combining both graph and cellular automata ideas — the graph-CA or *irregular CA* — which allows the structure and dynamics of spatial processes to be examined together. This chapter also develops the methodologies used in the remainder of part II to pursue this type of investigation.

Chapter 6 is something of a digression, which describes the overall architecture of the modelling environment which has been developed for exploring the ideas of the previous chapter, and (in an extended form) for building an application of the graph-CA model in part III.

Chapter 7 reports experimental investigations of the graph-CA model, intended to uncover some general properties of dynamic spatial systems. Some tentative conclusions are drawn from this research.

1.2.3 Part III: Application

Chapter 8 is an overview of research and theory in the field of gentrification studies, which is an essential prelude to the application of the graph-CA model framework to that topic. Gentrification has only rarely been modelled in spite of its significance in the recent urban geography literature. This review draws out some of the more promising ideas on which a graph-CA model might be based, particularly Smith's (1979*b*) *rent gap hypothesis.* The rent gap is then operationalised as the basis for the *Gentrification* model, the mechanics of which are described in detail.

In chapter 9 this model is explored in a small abstract grid-based environment, in order to develop a feel for its spatial operation and dynamics. This leads to a 'reasonable' set of parameter settings for the model, which are used when it is applied to the real world case study.

The case study area, Hoxton in east London, is introduced in chapter 10, and the difficulties of gathering sufficiently detailed data to run the model are discussed. The admittedly rough and ready approach taken to solving these problems is described, thus clarifying the relatively limited aim of model feasibility assessment and exploration, which is the current intent.

Chapter 11 describes the behaviour of the *Gentrification* model of Hoxton, for a limited range of cases. The focus here is still on the effects of spatial structure on spatial dynamics, and in particular on the different ways in which the relevant spatial relations might be conceptualised. Thus, even though the application of the model remains fairly abstract, its practicability and usefulness as a vehicle for developing intuitions and raising research questions is nevertheless demonstrated.

Finally, an attempt is made in the concluding chapter 12 to draw together the whole range of topics, issues and questions raised by all of the foregoing material. Suggestions as to the future development of this research are integrated into this discussion.

Chapter 2

Space: definitions and models

Whatever else they may be, one of the most important aspects of cities is how they organise space for economic, social, political and other purposes. Harvey (1973, page 23) says that

> "[a]ny general theory of the city must somehow relate the social processes in the city to the spatial form which the city assumes."

This is easily said, and clearly correct, but not so easily achieved! Hence, understanding the spatial form of cities, must be central to investigations into their function and growth. Furthermore, any such investigations must be based on some sort of conceptualisation of the nature of space.

This chapter examines this problem in terms of the difficulties of conceptualising space. This is because the current work is primarily concerned with developing a methodology for investigating the spatial aspects of social processes in the city. This might be described as a contribution to the development of what Harvey terms 'spatial consciousness' in sociological thought. Hillier (1996) has used the phrase 'developing tools to think with' (echoing the title of Waddington's, 1977 book) and this seems to describe a similar process and need.

Section 2.1 examines space from various perspectives which are relevant to the approach which is developed in the remainder of this work, and in section 2.2 the framework adopted for further development is described. This clarifies the *proximal* conception of space which is the basis for model development in this thesis, and also its relation to the pragmatically *realist* stance adopted in this work.

2.1 What is space?

Whatever space is, it is everywhere these days! Urban space, geographic space, architectural space, virtual space, cyberspace — that old chestnut — or information space, body space, mental space, cognitive space, 'the space of flows', geographic space, psychological space, dream space, symbolic space... the list is potentially endless. The sheer multiplicity of spaces in contemporary academic discourse is overwhelming (Benko & Strohmayer 1997, Soja 1989, for example). Lefebvre (1991) suggests that this multiplication of spaces began after Kant's designation of space as a *synthetic a priori*, an essential structuring framework for any knowledge, and therefore essentially ungraspable. Space was thereby left to the mathematicians and physical scientists to use as an explanatory aspect of nature — an attribute of things which may serve to explain behaviour. The mathematicians proceeded to invent "an 'indefinity' [...] of spaces: non-Euclidean spaces, curved spaces, n-dimensional spaces [...] and so on." (Lefebvre 1991, page 2).[1] This 'indefinity' prompted philosophers to ponder what such an array of spaces might mean, and so space became a mental construct rather than a thing in itself. This also clouded the mapping between human understanding (knowledge) of phenomena, and the phenomena themselves: if so many spaces were possible, indeed existed, how could any of them function as a basis for the understanding of the phenomena which unfolded in space?

Arguably, we are no further on now than then. The contemporary multiplicity of spaces is a reflection of the confusion in the concept, and, in some contexts, of its near meaninglessness. For my current purposes, this is not really a problem. The important step is to attempt to classify approaches to space, and to then choose some ensemble of approaches which matches our purpose. Such a choice can then be made with some awareness of the chosen approach's advantages and limitations.

Sack (1980) investigates spatial concepts at some length, and produces a categorisation which is worth examining further, before we consider the approach adopted in the current work in more detail. Sack's framework is useful in locating different approaches to space, and in understanding their overlaps, differences and relationships. It also hints at which kinds of approach to space may be appropriate in trying to understand different kinds of spatial phenomena. Figure 1 is based on a diagram

[1] The cynical might suggest that Lefebvre was at least as industrious a producer of 'spaces' as anybody he suggests, mathematicians included! Recent work by Unwin (2000) suggests that Lefebvre's pre-eminence as one of the theoretical geographers' favourite French theorists may soon recede.

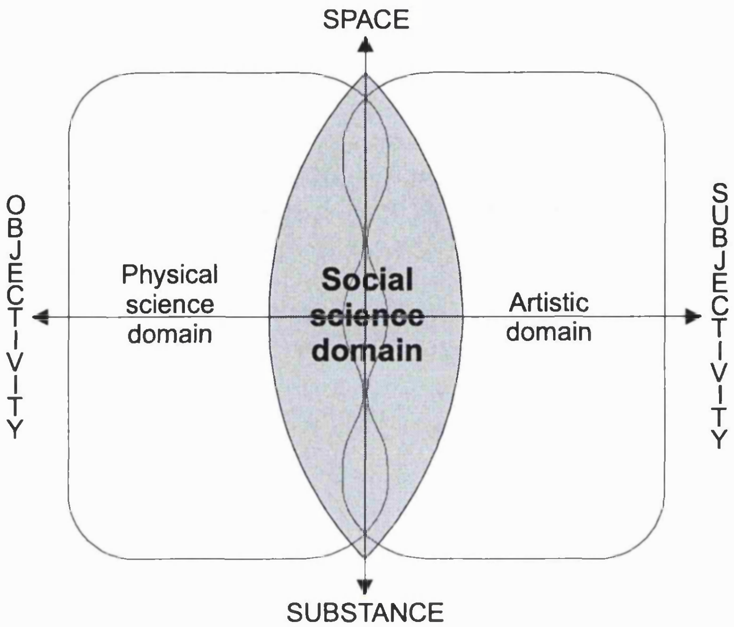


Figure 1 Sack's (1980) conceptual surface of space.

presented by Sack. He contends that different conceptions of space may be characterised by the extent to which they separate space from substance, and by their location along a spectrum from objective to subjective views. He is careful to state that "the objective end of the axis should be thought of as containing elements of the subjective and vice versa." (Sack 1980, page 24), so that there is no such thing as an entirely objective, nor entirely subjective view of space.

At this point it is worth noting that this denial of the possibility of an entirely subjective view of space is consistent with philosophical realism: there *is* something *out there*, space is not just a mental construct by which we understand the world. If it were, then all conceptions of space would necessarily be subjective. Such a view does not require us to believe that 'space' is a substance in itself, so much as that it has certain consistent topological and geometric properties (if we retrace our steps we end up back where we started, objects retain certain internal spatial relations under translation through space, and so on). Conceptions of space which regard themselves as objective, yet subscribe to the Kantian notion of the *synthetic a priori* are objective only about our perception of space, not about space itself. Sack explicitly claims to be a realist (Sack 1997, page 264, note 1). He has been criticised (see, for example, Sayer

1992, page 285, note 47) for a very abstract sort of realism, in that his rather schematic review of spatial concepts fails to relate any of the concepts to the social or political contexts in which they have been developed. This criticism seems to be well founded. However, for my current purpose it is unimportant, since Sack's contribution is to provide a framework for organising ideas about space.

Returning to figure 1, Sack argues that the more abstract the thinking we are doing about space, then the greater the extent of the conceptual surface. Thus, abstract conceptual systems are more reflective and more aware of their position in this framework. In contrast, unsophisticated spatial thinking — say a child's view of the world — is incapable of accommodating the more abstract thought systems of science or art. Sack goes on to consider the various general positions on this surface. He describes scientific, social scientific, the childhood, the practical, and the mythical/magical views of space. The picture which emerges is of a multiplicity of spatial concepts, each with its distinctive taxonomy of spatial elements. Couclelis (1992*a*) draws on Sack's work to produce a useful enumeration of these taxonomies as they relate to some of the more prevalent geographic approaches to space (see table 1).

Sack and Couclelis are both particularly interested in the socio-economic conception of space. Sack's approach draws attention to the particular complexity of this domain, given its position in the conceptual space, encompassing both scientific/objective and experiential/objective modes. This complexity is missing from the 'socio-economic' column in Couclelis's scheme, which focuses on more analytic accounts in socio-economic theory, but is implied by her inclusion of two further columns enumerating elements in allied domains. Although this taxonomy is clearly not (or could not be?) complete, it does at least give a good sense of the way in which different conceptions of space may influence subsequent analysis and discussion, and

Mathematical	**Socio-economic**	**Behavioural**	**Experiential**
Point	Location	Landmark	Place
Line	Route	Path	Way
Area	Region	District	Territory
Plane	Plain	Environment	Domain
Configuration	Distribution	Spatial layout	World

Table 1 Some spatial terminologies (*source*: Couclelis 1992*a*).

are therefore significant. In the sections which follow I consider some of these spatial concepts in turn, before outlining a general framework used in the remainder of this work.

2.1.1 Philosophical and physical science conceptions

Preceding much of the debate in other fields of enquiry is the philosophical debate concerning space. This is not an area which I wish to explore in great detail, since it has only limited relevance to space as it exists in terrestrial cases. However, it is in philosophical enquiry that the substance–space axis of Sack's conceptual space is most pertinent. The point of contention has always been whether space exists in and of itself, external to objects (or substance) which exist in space. This debate seems to be essentially unresolvable. Indeed it remained unresolved (or at any rate, there was no philosophical consensus) until the years after Newton's theory of gravity and his contributions to 'natural philosophy'.

Newton proposed an absolute model of space, in which space is conceived as an effectively empty container in which objects are located. Object locations can be described relative to some fixed frame of reference using a co-ordinate system. Thus,

> "It is not true that a Vacuum is nothing; it is the Place of Bodies; it is Space; it hath Properties; it is extended in Length, Breadth and Depth,..." (Voltaire 1738; 1967, page 180)

Leibniz, on the other hand proposed a relational model of space in which space only has meaning as a way of describing relations between objects. Space has no existence independent of objects. Eventually, for the essentially pragmatic reason that Newton's model worked, the absolute view of space was to dominate scientific and philosophic conceptions of space until the twentieth century (Gray 1989, pages 177–178). Such absolute views of space hold a great deal of sway in Western culture and thought. As an example, Western concepts about land and land-ownership are based on the idea of the possibility of defining divisions of land with reference to some external and fixed frame of reference — a Newtonian view. Kant's previously mentioned definition of space as a *synthetic a priori*, a conceptual category used by the mind to order concepts, canonised this approach (Sklar 1974).

This view only started to collapse after Einstein's theory of relativity gained general acceptance. Einstein (1960, page 155) was clear on the subject:

> "There is no such thing as empty space, i.e. space without field. Space-time does not claim existence on its own, but only as a structural quality of the field." (quoted in Couclelis 1992*b*, page 70)

The impact of Einstein's theories and mathematical abstractions of space on philosophy has been profound, so much so that the current mainstream philosophical wisdom points to relational Leibnizian conceptions (Grunbaum 1970, Reichenbach 1958, Sklar 1974). Nevertheless, a spirited debate continues and Nerlich (1994) makes a coherent argument for the independent existence of space. By drawing on ideas from topology and relativity theory, he argues that space exists and is manifest in the manner in which 'the shape of space' constrains the kinds of spatial configurations of objects which are possible. A flavour of the arguments is provided by thinking about what it is that makes a right hand different from a left hand. Those who argue for the relational view claim that it is the relation of the hand to the body, whereas Nerlich argues that it is the topological relations between parts of the hands. The argument then boils down to whether or not it is some property of space itself which makes 'handedness' possible.

2.1.2 Absolute and relational space in geography and GIS

The academic nature of such philosophical arguments at the terrestrial scale of geography, urban planning and architecture is apparent. All the same, echoes of these debates are to be heard even there.[2]

Concepts of space are fundamental to geography and hence to geographical information systems (GIS) and geographical information science (GISci). The conceptualisation of space was particularly contested in geography in the 'quantitative revolution' of the 1960s. At that time, an attempt to place geography on a more 'scientific' footing was made (Chorley & Haggett 1967, Haggett 1979, Haggett & Chorley 1969, Harvey 1969, for example). Geography was to move from description of the particular to prediction or modelling of general patterns in spatial activity. This introduced a tension into geography, between those who remained focused on understanding the

[2] Although, arguably, some fundamental properties of space are highly relevant at terrestrial scales. Thus, it is a topological property of space — the fact that two objects cannot usually occupy the same point in space simultaneously — which, combined with the political concept of the nation-state makes it difficult for the notions of joint sovereignty, and dual citizenship to gain much credence (cf. Northern Ireland, the European Union, and the notorious 'West Lothian Question' in connection with devolution in the United Kingdom).

particularity of given situations or regions, and those concerned to build largely abstract models of spatial processes, and thus to build geographical theories of general applicability. Such models were almost invariably built on mathematical — specifically geometric — concepts about space, and the closely related socio-economic concepts listed in table 1.

This tension can be characterised as a conflict between an absolute and a relational model of space. Hartshorne (1961) makes the connections between mid-twentieth century geography and Kant's view of space explicit. Where history's interest is in chronology, geography's is in 'chorography'; and where the historian focuses on arbitrarily defined periods in history, the geographer focuses on the region, an arbitrarily defined spatial unit, which can nevertheless be usefully examined in isolation. This approach is necessarily based on an absolute view of space, as the connection with Kant makes clear.

In contrast, the spatial processes approach of quantitative geography results in models in which distance is best represented by a variety of metrics. The appropriate metric depends on the problem under investigation (Harvey 1969, page 210). This insight arises fairly quickly in the economic-geography of location analysis where it is natural to think of distance as a cost which is clearly non-isotropic, and therefore not in principle compatible with an absolute model of space. The 'discovery', or at any rate application, of various metrics is in some ways comparable to the mathematicians' derivation of alternatives to Euclidean geometry in the nineteenth century, and seems to have had a similarly confusing impact on geography as a discipline. As compensation, geographers started to appreciate the multiple ways in which any spatial problem can be approached. These debates continue today, as can be seen in a theme issue of *Environment and Planning A* (Philo 1998), and in an entertaining — if depressingly bad-tempered — exchange between Openshaw (1991, 1992) and Taylor & Overton (1991).

The GISci research community, which has its roots in quantitative geography, has also been concerned to define space for itself, in a bid to formalise the concept of space in computer models. In another paper on this general theme Couclelis (1992*b*) suggests that the (then) long-running raster-vector debate in GISci had its roots in similarly conflicting views of space. Vector models embody an absolute conception of space. Raster models, on the other hand, embody a relational conception. This point is arguable — many would view the raster-vector debate as being about no more than technical and pragmatic issues of system implementation — and in more recent work Couclelis (1997) seems to change her view, implying that raster and vector GISs are both rooted in an absolute conception of space, as evidenced by their common reliance on the geo-referenced location. Seen from the perspective of the philosophical and scientific debates outlined in the previous section, the raster and vector views seem to be more usefully distinguished in terms of their positions relative to Sack's substance–space axis. Thus, vector models regard space as empty with distinct objects at definable locations, whereas raster models have continuously variable space which has substance at all locations (in the sense that it has measurable attributes at all locations). It is ironic that most GISs should embody an absolute model of space, given their origins in the distinctly relational spatial concepts of the quantitative geography tradition. However, it is difficult to conceive of a GIS implementation which does not rely at some (low) level on an absolute conception of space, in that it is by relating phenomena to some metric frame that a GIS distinguishes itself from other information systems.

The real issue is what spatial model is built on top of the underlying reference frame. Current GISs often do little more than identify trivial topological relations between polygons, edges and points, thus embodying the mathematical (and, to an extent) socio-economic concepts of table 1. In the next section more recent approaches within GISci research which attempt to incorporate some of the conceptions of space towards the subjective end of Sack's scheme are considered.

2.1.3 Cognitive conceptions of space

Almost contemporary with the above developments in quantitative geographic conceptions of space were developments in the emerging field of environmental behaviour. Seminal work in this field was done by Kevin Lynch who was one of the earliest proponents of the cognitive (or mental) map. This concept proposes that individuals maintain a mental representation of external space — the mental, or cognitive, map — which they use to move around in that space. Lynch's (1960) *The Image of the City* made the understanding of this mental representation as a *map* explicit, by its methodology, which involved interviews and sketch maps drawn by interviewees. The end result of the interview process was a map which was in some sense a collective image of the city: "There seems to be a public image of the city which is the overlap of many individual images." (Lynch 1960, page 46)

Lynch's public images of the city were presented as maps consisting of five major elements: *paths, nodes, landmarks, edges*, and *districts* (compare Couclelis's behavioural taxonomy in table 1). Such a division of urban space is clearly at some remove from an 'objective' approach based on land parcels, buildings and administrative or census boundaries. The subsequent impact of Lynch's work on urban planning has probably been less significant than on geography, environmental psychology and artificial intelligence. Tuan (1977) presents a geographic approach to space which is sensitive to the 'the perspective of experience'. It is not clear that such an approach can be reconciled with any one approach to space, as the many works on this theme in 'humanistic geography' attest (see Entrikin 1991, for example). However, this is not the point; rather, it is that as we move towards the subjective end of Sack's conceptual surface, conceptions of space rapidly multiply.

Even from the more empirical perspective of behavioural geography it is clear that the multiplication of spaces is inevitable. Golledge & Stimson (1997, pages 224–266) provide a comprehensive overview of the behavioural approach in geography and the importance of the cognitive map concept is acknowledged. Determining ways of actually drawing individual or collective cognitive maps was a major focus, and in some cases led to the refinement of new methods such as multi-dimensional scaling. Another focus was to substitute distances as measured on cognitive maps into behavioural models and compare the outcomes to those obtained using Euclidean distances. A variety of relations between objective measures of distance and perceived distance have been found (Sadalla & Magel 1980, Sadalla & Staplin 1980, for example). This work clearly involves the investigation of non-Euclidean spaces, but no generally applicable alternative geometry has been found. As a result these spaces are invariably relational spaces.

The GISci research community, partly influenced by these developments in geography, has recognised the need to incorporate a capability to handle these perceptual and multiple views of space into its models. Egenhofer & Mark (1995) explicitly called for such models to be formalised. Shariff, Egenhofer & Mark (1998) is an example of where these developments lead, and the extensive lists therein demonstrate the inherent complexity of any attempt to build an ontology of subjective spatial relations.

Moving from perceived spatial relations to perceptual space itself, Couclelis (1992*b*) outlines a taxonomy of perceptual space which may assist in discussing these concepts. She suggests that four spaces may be distinguished based largely on actual object scale:

1. *A-spaces* are those which contain everyday objects (table-top space).
2. *B-spaces* contain larger everyday objects, which must be assembled from multiple views.
3. *C-spaces* contain landscapes perceivable from a single-point but "otherwise not directly accessible to sensorimotor experience." (Couclelis 1992*b*, page 72).
4. *D-spaces* are regions beyond any direct experience, and seem to be equated with the mental spaces of cognitive maps.

Another classification, suggested by Montello (1993), also proposes four types of space, distinguished this time by the *projective* size of the space relative to the observer (that is, not by the actual extent of the space):

1. a *figural space* is smaller in projection than the human body. This has commonly been called 'table-top' space, and is a scale at which the properties of objects may be apprehended by manipulation and examination from all angles.
2. a *vistal space* is larger than the human body but may be "visually apprehended from a single location without appreciable locomotion" (Montello 1993, page 315). The external form of a house (or an elephant) exists in a vistal space.
3. *environmental space* is projectively larger than the body, and surrounds it. It cannot be fully understood without moving around in it. This is the space in which neighbourhoods and cities exist.
4. *geographical space* is much larger than the body, so much so that it cannot readily be appreciated even by large scale movement. This is the space of countries and continents, which are often only appreciated by their reduction to vistal or figural spaces by means of, respectively, vantage points or maps.

Kuipers (1982) presents yet another possible taxonomy, developed in the context of building representations of space for artificial intelligence systems. Again, the point is not to choose any one of these classifications as 'correct' but to note how in moving towards the subjective we find many alternative conceptions of space. It is clear that cities as vistal/environmental spaces (to use Montello's classification) are exemplars of the need for such multiple conceptions of space in social science.

As an example of where cognitive models of space might lead in the GIS setting, Gold (1992) and Edwards (1993) both propose a topological model based on neighbourhoods derived using Voronoi proximity polygons (Okabe, Boots & Sugihara 1994). Crucially, for these writers, the resulting mathematically-based model has cognitive resonance:

> "[inexperienced users] find the [spatial relations implied by the Voronoi model] to be consistent with their common sense. While in no way a proper psychological study, I find this easy grasp of concept suggestive of similarity with underlying human thought processes." (Gold 1992, page 227)

and

> "the Voronoi model of space is closer to a 'qualitative' view of space than existing models used by GIS." (Edwards 1993, page 204)

One of the ideas behind the current work is to use ideas and concepts in the urban morphology literature as another source of new spatial models for GISci. This thrust will become clearer in part III where I build a spatial model in a particular application and study area. Some of the relevant concepts from urban morphology are reviewed in section 4.3 (pages 73ff.).

2.2 Space in the present work

It is clear from the foregoing that a wide variety of spatial concepts could be brought to bear when building spatial models to investigate processes in cities. In the previous section we have considered existing spatial models relative to a conceptual framework developed in geography for relating subjective and objective models to one another. We have paid particular attention to a long-standing division rooted in physical science and philosophy between absolute and relational models of space. To combine these approaches, I want to introduce yet another conception of space: *proximal space*. Before doing that, it is important that I be clear about my philosophical position, with respect both to space and to academic inquiry more generally.

2.2.1 Philosophical basis: *critical realism*

Like Sack, I am a realist. The most persuasive account of a realist philosophy for me is Collier's (1994) presentation of Bhaskar's (1998) rather densely and obscurely explained *critical realist* position. The most important insight of critical realism is provided by asking — and answering — the question: 'how are experiments possible?' (Collier 1994, page 31). Bhaskar suggests the answer is that experiments are possible because there are underlying structures and mechanisms in the world which have causal power. These structures are real and intransitive (that is their existence is independent of any investigation into their nature) so that experiments are repeatable, and become a feasible way of interrogating nature, to improve our understanding.

More crucially, experiments are *necessary* because the world is an open system in which many structural components are always acting in concert so that no clear understanding of the structural causal mechanisms can be reached without isolating individual mechanisms. An experiment helps us to understand individual causal structures and effects by isolating them from other effects with which they normally coexist. Bhaskar further argues that the mechanisms of the world are inherently layered, which gives rise to the distinctive hierarchy of disciplinary division in the academy from sub-atomic physics, through solid-state physics and chemistry, to biochemistry, psychology, sociology and so on. This perspective is particularly well described by Harvey & Reed (1996), who also point to the consistency of such a view with the notion of *emergence* in complex systems.[3] An emergent phenomenon is one whose existence is not readily predicted from an understanding of the mechanisms operating in the substrate in which it occurs. Further, such a phenomenon's behaviour may be governed by laws which, while they are consistent with the laws in the substrate, may be identified and have explanatory force of their own. Perhaps the best example of this is the way in which the chemical behaviour of organic molecules (DNA, RNA, amino acids, proteins) gives rise to the biological/biochemical behaviour of primitive life-forms (in chapter 3 we will encounter a more abstract example of emergence in the complex behaviour of certain cellular automata, see page 58).

From this perspective, space can be regarded as a structural property of causal mechanisms in any kind of system, physical, chemical, biological, social, economic or

[3]The relationships of the present work to the 'sciences of complexity' are many and varied, so much so that the entire thesis could probably be re-written with its ideas set in that context.

whatever. Space is the medium in which the causal mechanisms operate, and to that extent, may shape the particular conjunctions of forces and mechanisms which may combine in any particular situation. However, against this, many mechanisms may be relatively independent of spatial constraints, so that the structuring effect of space is likely to be weak. Writing from the same perspective, Sayer (1992) is critical of both absolute and relational approaches. He argues that absolute models of space are unsatisfactory since they attribute effects (the 'friction of distance') to something which is nothing but empty space (vacuum). Relational models, on the other hand, lack coherence and are prone to 'subjectivism' whereby relations depend on your point of view. His interesting observation is that, in general, social (and other non-spatial relations) are necessary, whereas spatial relations are contingent to explanation. In other words most social situations could have many spatial realisations. For example many different spatial manifestations of the organisation of a factory, or a city are possible. This is not to say that space has no effects, because the particular spatial form which some social system takes may have effects which reinforce or undermine that social system. This may be a partial explanation of why more ambitious theoretical schemes (Harvey 1990, Soja 1989, for example) which seek to "reassert space in critical social theory" — to paraphrase the sub-title of Soja's influential book — have often failed to produce equally impressive practice. Sayer's conclusion is that in social science, relational models of space must be considered, but that there is little which can be said in general about the effects of space, which must be studied and understood in particular cases.

As will become clear, while I agree with this position in general, I think it may be possible to make some general statements about the kinds of processes (suitably defined) which can occur in particular kinds of spatial system (again, suitably defined). Chapter 7 represents a small contribution to a research program focused in this direction.

2.2.2 The proximal model of space

The philosophical position outlined above draws particular attention to the spatial structure of phenomena, and to the causal mechanisms which may be operating in any particular situation. This makes the *proximal* model a particularly appropriate conception of space for the present work.

In the proximal conception of space (Couclelis 1997, Takeyama 1996, Takeyama & Couclelis 1997), the fundamental element is the *neighbourhood*. A neighbourhood is defined by relations of *nearness* between spatial elements, and nearness in turn depends on both (spatial) *adjacency* and (functional) *influence*. Clearly a Voronoi conception of space is an example of such an approach, where adjacency is essential and neighbourhoods are contiguous. However, the proximal conception is more general, precisely because it allows non-contiguous neighbourhoods, based on relations of influence between elements. It also allows the incorporation of functional and spatial relations, and of fuzzy concepts and phenomena. This seems a rich vein worth further exploration. That it has developed in the context of a discussion of cellular automata models for urban and regional planning is significant, and strongly influences the methods adopted here for developing the idea.

Allowing that a proximal conception of space represents a useful addition to existing approaches on what bases might spatial models be constructed? Various metrics seem reasonable: the Voronoi approach is one such; Okabe et al. (1994) present a large family of augmented Voronoi-type approaches, and any of these could be used as the basis of a proximal model. Extensions of the basic Voronoi concept of proximity polygons are based on different metric systems. This is illustrated in figure 2, where a variety of spatial partitions based on different metrics are constructed based on the same set of point locations.

Ideas from urban morphology (again, see section 4.3, pages 73ff. for details) represent another set of ideas about how proximal models might be constructed. Atkin (1974*a*,*b*, 1975), Hillier & Hanson (1984), Krafta (1994, 1996), Krüger (1979*a*,*b*) and Steadman (1983) all offer non-metric methods for building up sets of spatial adjacency — hence proximal — relations, in urban environments. Examples of the sorts of proximal models which might result from this approach are shown in figure 3. The idea behind figure 3 is that a particular spatial arrangement of elements may give rise to many proximal spatial models, depending on which relations between the elements we decide are interesting. The main illustration shows a small segment of some urban environment, with the approximate centroids of street segments between junctions and buildings represented as black and white circles respectively. Various possible relations between these two sets of elements are also shown. In the street network, or *graph*, nodes representing street segments are connected to nodes represent-

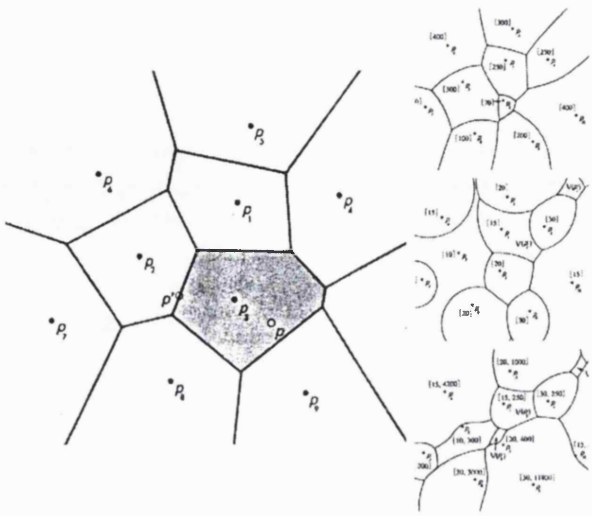


Figure 2 A simple Voronoi partition and three variants (*source*: Okabe et al. 1994).

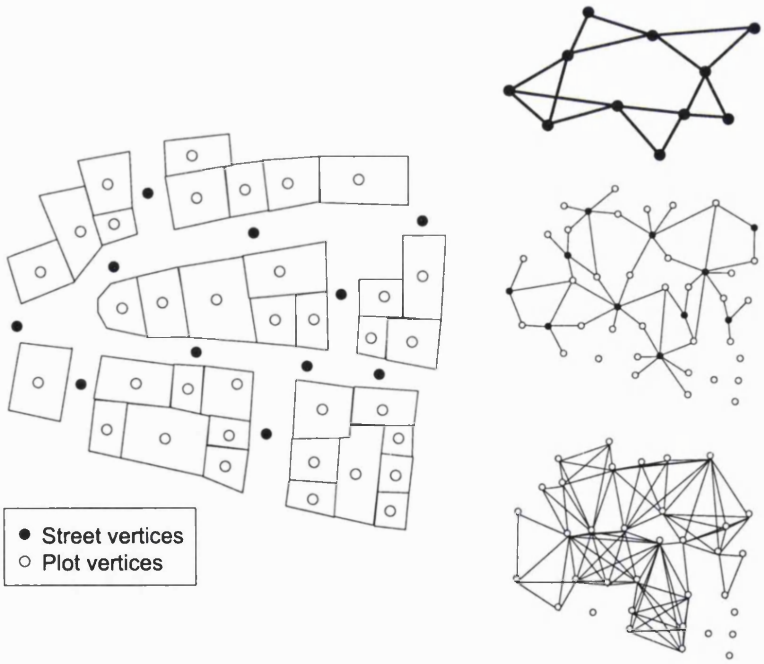


Figure 3 Proximal models of space based on urban morphological elements. Examples show a street network based model, a streets and plots based model, and one based on street frontages on the same street segment.

ing street segments which they meet at a junction. In the second example streets and buildings are included, but only the relation 'building on street' is represented. In the final example only buildings are represented, but each is taken to have a relation with all those other buildings with which it shares a street frontage. Thus, even this simple situation can give rise to a rich variety of proximal models. Note that there is already a whole series of decisions behind this representation as Krüger's (1979*a*, 1979*b*) work makes clear, since many other sets of spatial elements might be deemed relevant. More detail might be incorporated by considering rooms in buildings or floors; for a model covering a larger geographical area, city blocks might become the most relevant element. Having determined which elements are to be considered, the proximal model builder must determine which relations to include. All three examples shown use simple locality criteria to build the representation. Functional relations between elements might also be considered. The point is that many representations are possible, and that it is not really possible, without specifying our purpose, to privilege one representation over any other.

It is important at this stage to note that a proximal model of space entails a departure from the apparently self-evident assumption that space is uniform, since neighbourhoods are not necessarily defined according to a simple distance criterion or indeed, according to any universally applicable metric. For example, the approach might lead us to regard London and Tokyo as more meaningfully related to one another than London and Swansea. Stated in this way, it is perhaps obvious that it is the spatial unformity assumption which is actually unreasonable, in many socio-economic and human geography cases! There are clearly some processes, weather patterns and governance, for example, under which London and Swansea are closely related and London and Tokyo only distantly related, if at all (notwithstanding the flap of the famous butterfly's wings). But there are other contexts and processes, the movement of large scale capital flows, air-travel, under which the relative standings of the London–Swansea and London–Tokyo relations are very different. It is evident from the discussion earlier in this chapter that the disciplinary context (spatial analytic, cognitive, planning, economic, or whatever), as well as the subject under investigation, will influence which types of relations at what level are regarded as relevant.

It is this explicit invocation of the process of building a model which distinguishes a proximal model from a relational one. In a relational model, everything is related to everything else — which may in the end not tell us very much. In the proximal case, decisions about which relationships matter restrict the representation to just those relationships, thus structuring the model, and (hopefully) giving it more explanatory worth. The explicit recognition of the modelling process itself means that particular subsets of the possible elements and possible relations between those elements are selected in any particular case. In geographical models, it is likely that the resulting model will exhibit a certain 'spatial coherence', so that near elements are more likely to be related than distant ones, but this is strongly dependent on the scale (spatial and temporal) of the representation. It is also possible to imagine building a model in which there is a hierarchy of relations between and across different scales of investigation, an approach which seems particularly suited to Bhaskar's notion of the layering of reality, and to current geographical ideas about networks of world cities, trading blocs, nation states, economic regions and so on (Byrne 1998, pages 91–94).

In view of my emphasis on the model builder choosing a representation, and being conscious of the choices made, it might be (somewhat optimistically) thought that this approach stands up to the criticisms of postmodern relativist accounts of representation. But, of course, no representational approach can ever do that.[4] Once a particular representation is chosen, it is 'privileged' over the others which have not been chosen. To the extent that the representation is of a system of relationships and not of individual elements there may be some scope for a reconciliation with postmodern views (Cilliers 1998). Ultimately, however, in the best modernist tradition, the choice of representation will rely for its validity on the extent to which the resulting model can provide a plausible account (not necessarily a prediction) of what happens (past, present or future) in the world itself.

2.3 Conclusion

Figure 4 presents a possible way of thinking about the proximal models explored in more detail elsewhere in this thesis. There are many conceptions on which spatial models may be based from the mathematical or physical conceptions of science to

[4]To be cynical just once more, this seems to be the sole purpose of a great deal of postmodern theory — to criticise and deny the applicability of any account or representation — other, of course, than the postmodern account itself!

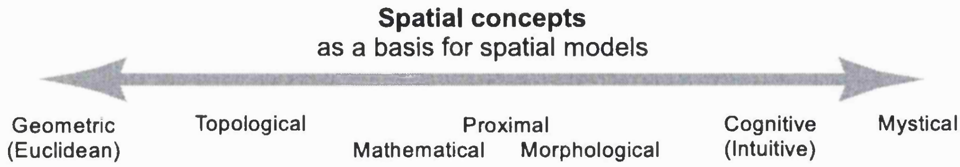


Figure 4 A possible spectrum of spatial concepts.

mystical concepts such as magic. Proximal models of space may result from spatial adjacencies which are entirely mathematical (the Voronoi proposal), but they may also be constructed on other bases such as urban morphology, where functional or perceptual relations are involved. The presentation of these ideas as a spectrum is appropriate, since there are no obvious dividing lines between the conceptions mentioned. Note that the location of proximal models on this spectrum is local (as it were) to this thesis, since there is no reason, in principle, why proximal models of cognitive–mystical space could not be constructed.

It should be clear from the foregoing that any typology of space and spatial concepts must necessarily be partial (in both senses of the word) and for something. It is introduced here as a basis for model building that seeks to examine the complexities of spatial phenomena in cities. Given the philosophical heritage of debates about space it would be arrogant to claim any more. This approach is informed by pragmatism, and an interest in tackling problems, not by the abstract ideas of philosophers, so that Gould (1997, page 128) captures the intent when he says:

> "we start with the idea that this strange no-thing [space] is structured by other things, which we relate in various ways to each other, and which we measure as various distances to each other as the fancy takes us according to our purpose of utility, curiosity, or ambition."

Chapter 3

Graphs and cellular automata

In pursuit of Gould's pragmatic programme we now turn to an examination of the two abstract mathematical tools on which the present investigation is founded. This chapter describes two widely used abstract mathematical representations for investigating systems of relations between elements — graphs and cellular automata. Both representations are applicable to a much wider range of problem domains than those where spatial relations are significant, which allows us to benefit from a great deal of earlier work across many disciplines. Attention is drawn to the particular aspects of problem domains which are emphasised by each approach. In the next chapter this emphasis is retained, when I examine some examples of the application of these representations to the urban morphology domain. The location of each representation in the conceptual scheme of chapter 2 is also briefly discussed.

The presentation in this chapter is necessarily fairly formal in order to avoid verbose descriptions. This definition of terms lays the necessary ground work for later material based on graph theoretic measures and cellular automata properties.

3.1 Graphs

The following sections are based on material in Beineke & Wilson (1997), Buckley & Harary (1990), Wilson (1996) and Wassermann & Faust (1994). Not all of the definitions in the sections which follow are used in subsequent chapters, but the concepts they relate to are useful and emphasise the generality of the graph representation.

3.1.1 Some definitions

A *graph* G consists of a finite non-empty set $V(G) = \{v_i\}$ of *vertices* and a finite non-empty set $E(G) = \{e_i\}$ of distinct unordered pairs of distinct elements in V, called *edges*. The number of elements in V, n, is the *degree* of G. The number of edges in G is denoted by m. The edge $\{v_i, v_j\}$ may be denoted by v_iv_j or e_{ij}. Figure 5 shows a typical small graph with $V(G) = \{a, b, c, d\}$, $E(G) = \{ac, bc, bd, cd\}$, $n = 4$, and $m = 4$.

Loops v_iv_i and multiple edges may be allowed if we drop the requirement that vertices in edges are distinct, or that edges be distinct. Directed graphs (*digraphs*) consist of a set of vertices $V(D)$, and a set of *arcs* $A(D)$ each of which consists of an ordered pair of vertices in V.

If $e = v_iv_j$ is an edge of G then e *joins* the vertices v_i and v_j, and these vertices are *adjacent*. We say that e is *incident* with v_i and v_j and that v_i is a *neighbour* of v_j. The *neighbourhood* $N(v_i)$ of v_i is the set of vertices in G adjacent to v_i. Two edges incident with the same vertex are said to be *adjacent* edges. In figure 5, $N(b) = \{c, d\}$, and edges bc and bd are adjacent.

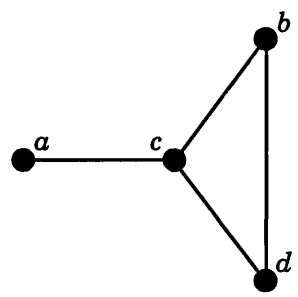


Figure 5 A typical graph.

Before going any further, it is worth noting the extreme generality of a graph representation. The vertex set $V(G)$ can clearly represent any set of entities — street junctions, houses, states, cities, trees; and the edge set $E(G)$ can similarly represent any relationship of interest between them — streets, intervisibility, trade volumes, railway lines, overlapping canopy cover. Another variant, *weighted graphs* where each edge has an associated value, which can denote the strength or the nature of the relationship between the vertices it joins, further extends the applicability. In terms of the discussion of spatial models presented in the preceding chapter, graph representations of spatial situations combine aspects of both absolute and relative spatial models. The vertices typically represent well defined objects located in absolute space, while edges may represent spatial relations of adjacency, nearness or whatever. Of course, the graph itself is aspatial and the relations may be completely non-spatial or some combination of non-spatial and spatial relations.

Vertex degree

The number of edges incident with a vertex is its degree denoted $\deg(v_i)$, and often denoted k_i. The maximum degree in G is sometimes denoted by Δ. A vertex of degree 0 is isolated, and a vertex of degree 1 is an *end-vertex*. The *degree list* of G is the set of degrees of the vertices of G often arranged in non-decreasing order. If all vertices in G have the same degree k, it is *regular* of degree k, or k*-regular*. The *density* of a graph is $m/\binom{n}{2}$, that is the fraction of all the possible edges which *could* exist between vertices, which *do* exist. In figure 5, $\deg(b) = 2$, and $\Delta = \deg(c) = 3$; a is an end-vertex, the degree list of the graph is $\{1, 2, 2, 3\}$ and its density is $\frac{2}{3}$.

Matrix representations

The *adjacency matrix* of G is the matrix $\mathbf{A}(G) = a_{ij}$ where $a_{ij} = 1$ if v_i and v_j are neighbours, and 0 if not. Thus the graph of figure 5 would be represented by the adjacency matrix

$$\mathbf{A}(G) = \begin{array}{c} (a) \\ (b) \\ (c) \\ (d) \end{array} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{3.1}$$

Note that the row order is meaningless, but that the row and column ordering must be the same. The *incidence matrix* of G is the $n \times m$ matrix $\mathbf{B}(G)$ where $b_{ij} = 1$ if e_j is incident with v_i, and 0 if not. The incidence matrix of the graph in figure 5 is

$$\mathbf{B}(G) = \begin{array}{c|cccc} & (ac) & (bc) & (bd) & (cd) \\ \hline (a) & 1 & 0 & 0 & 0 \\ (b) & 0 & 1 & 1 & 0 \\ (c) & 1 & 1 & 0 & 1 \\ (d) & 0 & 0 & 1 & 1 \end{array} \tag{3.2}$$

The matrix product $\mathbf{BB}^T = \mathbf{A}^*$, where $\mathbf{A}^*$ represents the adjacency matrix, modified such that the elements in the main diagonal are equal to the degree of the corresponding vertex.

Paths and cycles

A sequence of edges $v_0v_1, v_1v_2, \ldots, v_{r-1}v_r$ (or $v_0v_1 \ldots v_r$) is a *walk* of *length* r. If the edges are distinct it is a *trail*, if the vertices are distinct it is a *path*. If $v_0 = v_r$ then the walk is closed, and for $r > 0$ a closed walk in which the vertices are distinct is a *cycle*. If we denote the length of a particular path x from v_i to v_j by $l_x(v_i, v_j)$, then the *distance* $d(v_i, v_j)$ is the length of the shortest path in the set $\{l_x(v, w)\}$ of all possible paths from v_i to v_j. Thus $d(v_i, v_j) = \min\{l_x(v_i, v_j)\}$ is the distance from v_i to v_j. A shortest path between two vertices is called a *geodesic*. The largest distance between any two vertices in G is the *diameter* of G.

In a digraph walks may only proceed in the direction of the constituent arcs. In a weighted graph, the length of a path is normally calculated by summing the weight of its constituent edges, and the distance between two vertices is the length of the shortest path, as before. A graph is *connected* if there is a path joining each pair of vertices in G (alternatively if there is a walk including every vertex in G). A graph which is not connected is *disconnected* and any disconnected graph can be split into maximal connected sub-graphs called *components*.

Examples of graphs and sub-graphs

A graph in which every two vertices are adjacent is a complete graph denoted by K_n. K_n has $\binom{n}{2}$ edges and density 1. If edges ab and ad were added to the graph in figure 5 it would be K_4. In the same way, a graph which is a cycle of n vertices is denoted C_n, and one which is a path of length n by P_n. K_4, C_4 and P_4 are shown in figure 6.

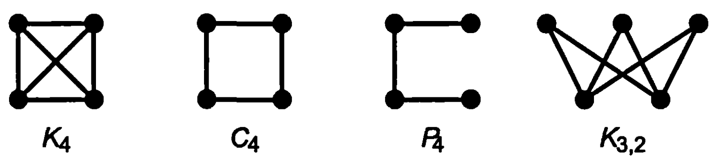


Figure 6 Examples of graphs (see text).

A *clique* in G is a complete sub-graph in G, and a maximum clique is a clique of maximum degree in G. The clique number $\omega(G)$ is the degree of a maximal clique.

A *bipartite* graph is a graph whose vertices can be partitioned into two sets, so that each edge joins a member of the first to a member of the second set. A complete bipartite graph is a bipartite graph where each member of the first set is adjacent to every member of the second set; if the sets have r and s members then $K_{r,s}$ denotes the complete bipartite graph. $K_{3,2}$ is shown in figure 6.

A connected graph with no cycles is a *tree.* Note that in a tree $n = m + 1$. We obtain another tree by removing the end-vertices of any tree T. A *spanning tree* of G is a connected sub-graph in G with the same degree as G, but containing no cycles.

Planar graphs

A *planar graph* is a graph that can be embedded in the plane so that no two edges intersect, except at a vertex to which they are both incident. The plane is partitioned into regions by such a graph, the regions being called *faces,* and the exterior region being called the *infinite region.* Note that any simple street network is likely to be readily represented by a planar graph, where junctions are the vertices and street sections between nodes are the edges. Obviously, complications arise when overpasses are included. Where there is a need to represent street directions then a directed graph will be required.

New graphs from old

The *line-graph* $L(G)$ of G is the graph whose vertices correspond to the edges of G, and where two vertices are joined when their corresponding edges are adjacent.[1] Figure 7 shows the derivation of the line graph from the graph in figure 5.

For a connected planar graph the *dual graph* is obtained from the adjacency relations of its faces. In the dual graph each face is represented by a vertex; each vertex has as neighbours the vertices representing faces with which it shares a common edge. This is the relationship between the commonly used dual the Voronoi diagram or tesselation and the Delaunay triangulation, and is shown in figure 8.

[1] The adjacency matrix $\mathbf{A}(L)$ is given by $\mathbf{B}^{\mathrm{T}}\mathbf{B} - 2\mathbf{I}_m$ where $\mathbf{B}$ is the incidence matrix of G as defined in equation 3.2 above, and $\mathbf{I}_m$ is the $m \times m$ identity matrix, where m is the number of edges in the original graph G.

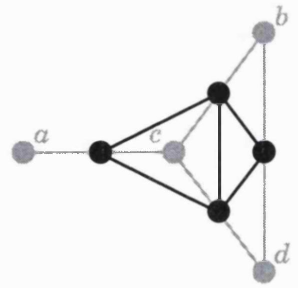


Figure 7 Derivation of the line graph for the graph in figure 5.

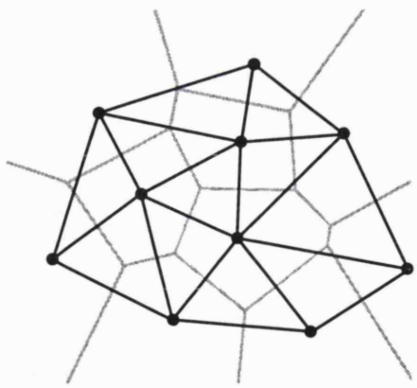


Figure 8 A Voronoi diagram or tesselation (grey) and its Delaunay triangulation (black).

3.2 Graph measures

One of the advantages of graph representations is the wide range of measures which have been developed in a number of disciplines to assist in understanding their structure. Some of these are explained and discussed in this section. The graph theory literature is extensive, and in many cases highly technical and mathematically obscure.[2] Just to illustrate: a great number of the celebrated Paul Erdös's 1500 or so publications were concerned with aspects of graph theory. Fortunately, one of the reasons for the extensive literature of graphs is their widespread applicability, and as a result there is also an extensive secondary literature examining applications of graphs in various fields. I have found social networks theory (Wassermann & Faust 1994) a particularly fertile field, and much of the material in this section is based on that literature, rather than network analysis in geography which tends to concentrate on the particular structures associated with transport and drainage networks (Haggett & Chorley 1969).

In the review which follows, three broad categories of structural measure are identified: *centrality*, the identification of *cohesive subgroups*, and the identification of *structural equivalence* classes. These ideas are quite general and seem to have possible relevance in urban morphology. Figure 9 illustrates the sorts of features which each of these approaches is likely to identify. From figure 9(ii) we see that 'centrality' identifies those vertices in a graph which are dominant in the structure. Figure 9(iii) illustrates cohesive subgroups as regions of a graph which are particularly strongly interconnected, and figure 9(iv) is intended to show that structural equivalence refers to the idea that various groups of vertices may be identified based on the similarity of their positions in the graph (end-vertices are an obvious example).

3.2.1 Centrality and centralisation

An obvious question to ask about any graph (especially one derived from some 'real-world' situation) is "which are the most important vertices?" One way of answering this question is by using one of a number of ways of measuring the *centrality* of the vertices. A vertex centrality measure is one which reflects the extent to which the vertex is involved in relations with other vertices. A related concept is *centralisation* which measures the extent to which the graph as a whole is centred on a limited

[2]Rather, it is 'obscure' in the same sense that most mathematics is obscure to non-mathematicians!

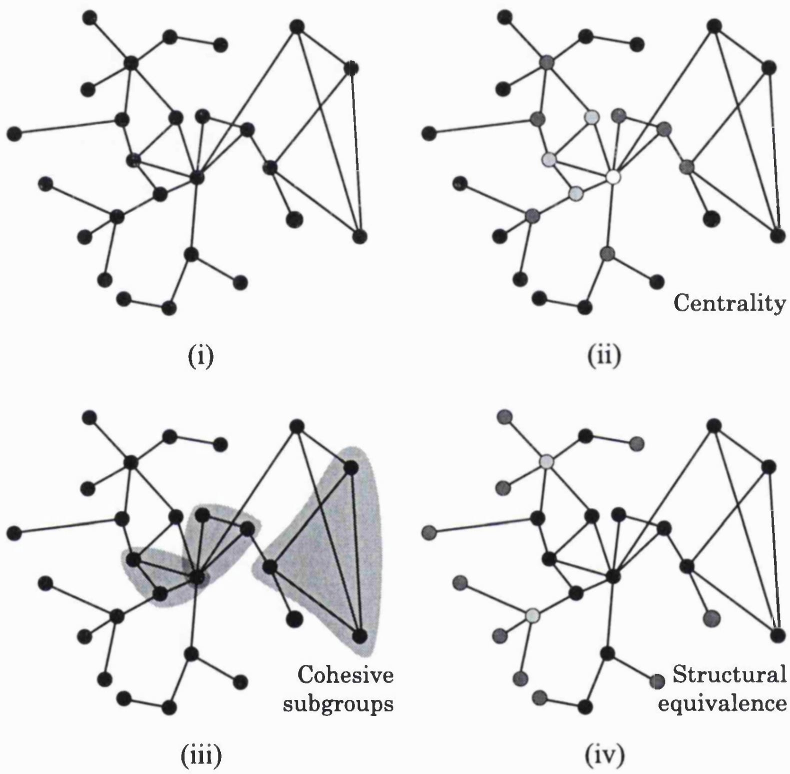


Figure 9 The three types of graph structural measure illustrated for graph (i): (ii) centrality, (iii) cohesive subgroups, and (iv) structural equivalence.

number of vertices — in the same sense that we describe a nation-state as centralised. Various measures have been proposed based on degree, distance, 'betweenness', information, and differential status/rank. The potential interest in the most (or least) central vertices, and in the centralisation, in graphs derived from urban systems, is obvious. In the following discussion the centrality of a single vertex is denoted by $C(v_i)$, and the graph centralisation is denoted by C. Normalised measures are denoted by $c(v_i)$ and c respectively. Subscripts denote the method used to derive the measure.

Degree centrality

Arguably, prominent vertices are those involved in relations with many other vertices — a central vertex is one with many neighbours. This definition is clearly particularly relevant in social network theory, but the idea is general enough. Freeman (1977, 1979), Freeman, Borgatti & White (1991), Nieminen (1974) and Wassermann & Faust (1994) all discuss these measures. A degree based measure in an urban context might highlight particularly significant nodes in a transport network, or prominent buildings in an graph of intervisibility relationships. A simple approach is to define degree centrality C_{deg} of a vertex as

$$C_{\text{deg}}(v_i) = \deg(v_i) \tag{3.3}$$

and an obvious normalisation is

$$c_{\text{deg}}(v_i) = \frac{\deg(v_i)}{(n-1)} \tag{3.4}$$

since $(n-1)$ is the largest vertex degree possible in a graph of n vertices.

For a centralisation measure based on degree, we could consider using the variance or standard deviation of the constituent vertex centrality measures, since these record how unequal are the centralities of vertices. However, such measures of central tendency may not be entirely appropriate, since centralisation as we are concerned with it here, has more to do with how *skewed* the centrality values measured at individual vertices are towards just a few, highly centralised cases. A skewness measure such as the third moment of the distribution of individual vertex centrality values

$$\sum_{i=1}^{n}(C_{\text{deg}}(v_i) - \overline{C}_{\text{deg}})^3 \tag{3.5}$$

could presumably be used but is not mentioned in the literature. Freeman (1979) advocates another approach. If we denote the maximum vertex centrality by $\widehat{C}_{\text{deg}}$, then the graph centralisation is given by

$$C_{\text{deg}} = \sum_{i=1}^{n} \left(\widehat{C}_{\text{deg}} - C_{\text{deg}}(v_i) \right) \tag{3.6}$$

which is a measure of how much more centralised than all other vertices is the most central one. Normalising centralisation measures is more complicated than for centrality measures. Freeman's approach is to divide the measure in equation 3.6 by the maximum possible value which could be obtained for a graph of the same degree n. This would occur in the star-like graph of degree n. Since the degree of the central vertex in such a graph is $(n-1)$, and the degree of all the $(n-1)$ other vertices is 1, the maximum possible value is $(n-1)(n-2)$. Thus a possible normalised centralisation measure is

$$c_{\text{deg}} = \frac{\sum_{i=1}^{n} \left(\widehat{C}_{\text{deg}} - C_{\text{deg}}(v_i) \right)}{(n-1)(n-2)} \tag{3.7}$$

This seems easy enough, but Donninger (1986), referring to Erdös & Renyi (1960), hints at the complex assumptions which are hidden by this sort of approach. Donninger's point is that any graph measure should ideally be related to the value the same measure might be expected to have in a randomly generated graph of the same degree, *with the same number of edges.*[3] He concludes that measures such as that in equation 3.7 are unstable for small graphs, and a correction factor (such as $\sqrt{n}$ or $\log n$) is recommended. The potentially intractable (and advanced) mathematics of large and random graphs is well beyond the scope of the current work. However, the general point that simplistic normalisation techniques may be subtly flawed is worth bearing in mind.

The general approach to centralisation presented here for degree-based measures is applicable to both distance and betweenness measures (i.e. either variance, skewness, or the 'Freeman' normalisation can be used), so, for brevity, I do not give details of these.

[3]Note that the models of random graphs which might be used are a rich subject area in themselves. See, for example, Bollobás (1985).

Distance or closeness centrality

An obvious alternative to vertex degree measures is the distance of a vertex from other vertices in the graph. Sabidussi (1966) proposes an index based on the distance of each vertex from every other:

$$C_{\text{dist}}(v_i) = \left[\sum_{j=1}^{n} d(v_i, v_j)\right]^{-1} \tag{3.8}$$

The use of the reciprocal is so that large values correspond to central vertices. Normalisation is achieved by multiplying through by $(n-1)$:

$$c_{\text{dist}}(v_i) = (n-1)C_{\text{dist}}(v_i) \tag{3.9}$$

This approach is similar to accessibility measures used in transport planning (Lee & Lee 1998, Pirie 1979, provide reviews), where weighted graphs are frequently used. The distance approach is also related to the method used in the space syntax approach to urban morphology (see section 4.3.3, pages 78ff. for more details), where a simple, unweighted graph is employed.

Betweenness centrality

These measures are based on the idea that vertices 'control' communication or movement between other vertices to the degree that they lie on (shortest) paths between them. Freeman (1977) argues that betweenness centrality is an intuitively useful measure for vertex centrality, especially in a social network where it may be interpreted as related to an individual's potential for control over information flows. He proposes a measure based on the geodesics between all vertices. If there are g_{jk} geodesic paths from v_j to v_k, then the probability of any one of them being used is $1/g_{jk}$. If $g_{jk}(v_i)$ is the number of geodesics on which v_i lies then the probability that a route including v_i will be used is $g_{jk}(v_i)/g_{jk}$. This leads straightforwardly to

$$C_{\text{between}}(v_i) = \sum_{k=1}^{n}\sum_{j=1}^{n} \frac{g_{jk}(v_i)}{g_{jk}} \,\forall\, i \neq j \neq k \tag{3.10}$$

The maximum value attainable, is clearly $(n-1)(n-2)/2$, since this is the number of pairs of other vertices not including v_i in a graph of degree n. This value is reached by the central vertex in a 'star' graph since all the geodesics between other vertices pass through the central one. A convenient normalisation is therefore

$$c_{\text{between}}(v_i) = \frac{C_{\text{between}}(v_i)}{[(n-1)(n-2)/2]} \tag{3.11}$$

This centrality measure has obvious application to urban morphology problems concerned with transport networks.

Information centrality

The most obvious weakness of the betweenness measure is the assumption that all geodesics are equally likely. Relaxing this assumption is difficult. Another assumption embedded in the betweenness measure is that shortest paths are the only paths likely to be used. Stephenson & Zelen (1989) point out this weakness, particularly in the context of their own interest in the transmission of infectious diseases. Their centrality measures weights the sum of 'control' values according to the length of the paths on which the vertex lies. Typically, geodesics are given weight unity, and other paths are weighted according to the information they contain. The information is defined as the inverse of the path length. This leads to an elaborate measure which is nevertheless easily calculated by manipulations of the adjacency matrix of the graph.

3.2.2 Cohesive sub-groups

Another set of graph measures thoroughly explored by social network theorists is intended to identify sub-graphs in a graph, which are in some sense more closely connected internally than they are to the rest of the graph. The sociological reason for an interest in such groups is that the more tightly individuals are tied to a network, the more likely it is that they will be affected by standards in that group. Two factors are likely to be of interest: the extent to which vertices of a sub-graph are related to one another, and the extent to which the sub-graph is isolated from other parts of the graph. The measures described below reflect these interests. In an urban context we might expect cohesive subgroups, if they exist, to correspond to 'neighbourhoods', or possibly to regions of similar land use, although there is likely to be some overlap between the two.

Subgroups based on complete mutuality

A clique in this context, is a maximal complete sub-graph of three or more vertices — a sub-graph in which every vertex is joined to every other vertex. Note that a vertex may be in several cliques, but that no clique can be a sub-graph of another since it would not then be maximal. A clique is a very strict definition of cohesive subgroup, and whilst it may be of interest in social network theory it seems unlikely to be of interest in urban morphology.

Subgroups based on comparing within to outside subgroup ties

An approach to cohesive subgroups based on comparing intra-set to inter-set ties seems more suited to the purposes of urban morphology, since the existence of subgroups is assessed against prevailing norms in the graph as a whole, rather than on the arbitrary requirement of compete mutuality. The starting point for these approaches is the *strong alliance* which is a complete component in a graph, that is a clique, with no connections to other vertices in the graph. The following definitions are based on various relaxations of this rather restrictive concept. We can envisage the subgroup definitions below as resulting from the progressive connection of such an isolated clique to other parts of the graph, as well as the progressive erosion of internal connections.

LS sets and lambda sets

The definition of an *LS set* (named after Luccio and Sami) is based on the greater frequency of ties within compared to outside the cohesive subgroup. Borgatti, Everett & Shirey (1990), Luccio & Sami (1969) and Seidman (1983) all offer definitions. For example: "a set of [vertices] S in a [graph] is an LS set if each of its proper subsets has more ties to its complement within S than to the outside of S" (Seidman 1983, page 98).[4] Borgatti et al. show that the LS sets in a graph form a nested hierarchy which is an interesting feature, suggestive, in the urban context, of regionalisation at different scales.

[4]This idea was originally developed in electronics to assist in the partitioning of devices amongst circuit boards, so as to minimise the number of (relatively unreliable) interconnections between circuit boards.

Borgatti et al. then extend the notion of an LS set, believing that it is important that cohesive subgroups should have some minimum connectivity. The edge-connectivity $\lambda(i, j)$ of a pair of vertices i and j, is the number of edge-disjoint paths (of any length) between them, that is paths which do not share any edges. A subgraph S is a *lambda set* in G, if any pair of vertices in S, has larger λ than any pair of vertices where one vertex is taken from S, and one from $(G \backslash S)$.

Lambda sets are more general than LS sets, and any given graph is more likely to contain lambda sets than LS sets. Successively larger values of λ give rise to a series of lambda sets, which do not overlap unless one is contained in another. This feature suggests close similarities with hierarchical clustering techniques and it may be that this is just another way of looking at such techniques — much as MacGill (1984) demonstrated that some of the mystery of Q-analysis (Atkin 1974c) disappeared when its close relationship to single link clustering was recognised.

Other cohesive subgroup definitions

A variety of other definitions for cohesive subgroups have been proposed such as *n-cliques, n-clans, n-clubs, k-plexes* and *k-cores*. The various definitions and properties of these are described by Wassermann & Faust (1994, pages 257–267). For a variety of reasons none of these seem likely to be of interest in urban morphology. For example, an n-clique is a maximal subset V_n of vertices in $V(G)$ such that the distance between any pair of vertices in V_n is not greater than n, that is

$$V_n = \{v_i : d(v_i, v_j) \leq n \; \forall \, v_j \in V_n \wedge d(v_i, v_j) > n \; \forall \, v_j \in V(G) \backslash V_n\} \tag{3.12}$$

This definition is such that an n-clique may be disconnected, which suggests that the n-cliques in a graph are unlikely to be of interest in urban morphology, certainly not where the search for cohesive subgroups is motivated by a desire to identify neighbourhoods. Networks of world cities, on the other hand seem likely to form n-cliques if we were to examine (say) the airline connections between cities.

Finally, in connection with cohesive subgroups it is worth mentioning that deterministic algorithms for finding them are slow, because of the vast numbers of subgraphs which must be tested. Most practical methods are based on permuting the rows and columns of the graph adjacency matrix, until dense clusters of 'ones' are found in particular regions. This sort of process is generally driven by heuristic methods, and is unlikely to produce unique solutions.

3.2.3 Structural equivalence

Another set of measures of graph structure is intended to allow groups of vertices which have similar sets of relations to the rest of the graph to be identified. A rather strict mathematical definition introduced by Lorrain & White (1971) and cited by Sailer (1978) defines as structurally equivalent any two vertices that are related in the same ways to the same other vertices. This sort of equivalence hardly ever holds in practice; at any rate, it only applies to rather obvious sets of vertices. For example, in a graph representing a family tree only sets of unmarried, childless siblings would be identified as structurally equivalent. A more fruitful concept (even if it has a more intimidating name) is *automorphic equivalence*. This is defined by Sparrow:

> "in a [graph G], a permutation of the [vertices] f is an automorphism if and only if it preserves adjacency: that is, given two [vertices] a and b, then $f(a)$ is linked to $f(b)$ if and only if a is linked to b. Two [vertices] are said to be role equivalent (or automorphically equivalent) if there is an automorphism which maps one on to the other. That is, a and b are automorphically equivalent if there exists an automorphism f such that $f(a) = b$." (Sparrow 1993, page 153)

To clarify, using the family tree example again, we would expect to find members of a single generation in automorphic equivalence classes, since they share similar relations to other parts of the graph. This seems like a concept which is likely to be of use in urban morphology since it might allow the identification of groups of urban form elements which exist in similar relations to the rest of the city; 'similar places' catches the approximate meaning. Various methods for detecting automorphic equivalence classes have been proposed (Borgatti & Everett 1989, Burt 1990, Sparrow 1993). Computationally, Sparrow's approach is the most tractable. His method involves the performance of a repeated numerical calculation for each vertex based on some property of the vertex's neighbours, and a transcendental number such as π. The calculation is repeated for every vertex in the graph as many times as the 'depth' to which vertex equivalences are required to be distinguished. If we denote the Sparrow equivalence value to depth d of some vertex v_i by $x_{i,d}$ then Sparrow suggests the calculation:

$$x_{i,d} = \prod_{j:v_j \in N(v_i)} \pi + x_{j,(d-1)} \tag{3.13}$$

where $x_{i,0} = \deg(v_i)$. This method can be refined to cope with directed graphs too. Its major drawback is that the numeric value associated with each vertex has no intrinsic meaning: it serves merely as a 'classifier' for the grouping of vertices into equivalence classes.[5]

3.2.4 Small world networks and their structure

More recent work on the properties of non-regular graphs by Watts & Strogatz (1998) and Watts (1999) has also produced two convenient measures. The interest in this work is to determine what properties a graph must have to exhibit the 'small world phenomenon'. This is the oft-remarked surprise we encounter when, on meeting a stranger, we soon find out that we share a mutual acquaintance ("small world, isn't it?!"). Such events, and our surprise at them, are indicative of how much more closely interrelated we all are than our intuitions would lead us to expect. Our unreliable intuitions are based on the perception that our own social networks are relatively tightly interconnected — locally dense, in graph terms. The concept has recently been popularised by the internet game 'six degrees of Kevin Bacon'.[6] Trying to develop suitable graph measures for understanding these graph properties, and after reviewing much of the above-cited literature, Watts (1999) concludes that (for his purposes) two measures are particularly useful: *characteristic path length* and *clustering coefficient*. Before discussing these measures, we define the characteristic path length of a graph G, $L(G)$ as

$$L(G) = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} d(v_i, v_j)}{[(n-1)(n-2)/2]} \tag{3.14}$$

$L(G)$ is simply the average shortest path length,[7] and thus closely related to the distance based centrality measure already defined in equation 3.8 on page 49. The clus-

[5]Malcolm Sparrow acknowledges this problem (personal communication) and has suggested that it can be partly circumvented by pre-classifying vertices into a small number of types. For example, in a social network, actor degree might be classified as very low (< 5), low (< 10), and so on. This greatly reduces the number of distinct classes which are identified at each depth in the calculation, which may assist in identifying and interpreting the classes .

[6]See `http://www.cs.virginia.edu/oracle/` for details. The Kevin Bacon game demonstrates the similarly surprising (or maybe not) interconnectedness of the graph constructed where actors are vertices, and edges represent their appearing in a film together. The game's title is based on the (false) allegation that no actor is more than six steps away from Kevin Bacon in this graph. Kevin Bacon is nevertheless 'well-connected' in this sense, with an average distance to other actors of only 2.83. For what it's worth, the 'best-connected' actor is Rod Steiger.

[7]Watts actually defines characteristic path length as the *median* of the means of the shortest path lengths connecting each vertex to all others. This is because a median is easier to estimate from a sample of the shortest paths, and he is interested in very large graphs, where determination of all shortest paths is impractically slow.

tering coefficient $\gamma(G)$ is defined as

$$\gamma(G) = \frac{1}{n} \sum_{i=1}^{n} \frac{|E(\Gamma_i)|}{\binom{k_i}{2}} \tag{3.15}$$

where Γ_i is a sub-graph in $G(V, E)$ made up of the neighbours of vertex v_i and all the edges which exist between them. Γ_i is considered to not include v_i itself, nor any loops $v_j v_j$. As before, k_i is the degree of v_i. The clustering coefficient is thus the mean *density* of a graph measured locally in each vertex neighbourhood. It is the average number of edges which *do* exist in each vertex neighbourhood Γ_i, expressed as a fraction of the number which *could possibly* exist. It therefore lies between 0 and 1. This measure, determined at each vertex (rather than averaged over the whole graph), seems likely to be closely related to cohesive subgroups, in that *ceteris paribus* we would expect vertices with higher γ to belong to larger cohesive subgroups.

Together, Watts suggests these average measures can provide a useful overall summary of the structure of a graph. He goes on to develop models for the expected values of L and γ in various random graphs. Small world graphs are then defined by the characteristic of having near-random characteristic path lengths combined with a high clustering coefficient. Watts does not spend much time on the *distribution* of these measures determined locally for all vertices (i.e. the distributions on which the means are based), but this also seems likely to be of interest. I return to the small world measures elsewhere in this thesis (see sections 7.1.1 pages 162ff., and 11.1 pages 253ff.), as conveniently calculated summary statistics, which, given the impact of Watts & Strogatz's (1998) work, seem likely to be widely used.

3.2.5 Conclusions on graphs

The above overview gives some idea of the range of structural measures available to the investigator, once a graph representation of some 'real-world' situation has been derived. Of course, the 'meaning' of any of these measures or partitions of the graph will depend on the situation, the representation (what do the vertices and edges represent?), and the interests of the investigator. But broadly speaking, it is the fixed structure of the relations between the objects of interest in any given situation which a graph allows us to examine. We will see this again in the next chapter where these sorts of measures are applied by some workers in the field to the description and understanding of urban systems. Which of the measures from those above should be used in a particular investigation is a moot point, and is likely to be determined partly by trial and error, partly by judgement of the situation's requirements, and partly by the tools available. The likelihood that the precise measures used will have a bearing on the overall patterns identified in a particular graph is low. For example, Bolland (1988) in a comparison of a number of centrality measures — not including information centrality — concludes that high correlations between vertices found to be central (or peripheral) to test networks "suggest considerable redundancy" (page 251). However, an awareness of the likelihood of certain configurations occurring is also necessary as Donninger's (1986) invocation of random graph theory makes clear. Thus, *some vertex* must be most central, and the fact that a particular one is, may not be especially significant, without consideration of the properties of random graphs, or, perhaps, 'similar' graphs (in some sense) of the same size.

3.3 Cellular automata

3.3.1 Definition

Stephen Wolfram, one of the foremost experts on cellular automata (CA) in the last twenty years defines them as:

> "[...] mathematical idealisations of physical systems in which space and time are discrete, and physical quantities take on a finite set of discrete values. A cellular automaton consists of a regular uniform lattice (or 'array'), usually infinite in extent, with a discrete variable at each site ('cell'). The state of a cellular automaton is completely specified by the values of the variables at each site. A cellular automaton evolves in discrete time steps, with the value of the variable at one site being affected by the values of variables at sites in its 'neighbourhood' on the previous time step. The neighbourhood of a site is typically taken to be the site itself and all immediately adjacent sites. The variables at each site are updated simultaneously ('synchronously'), based on the values of the variables in their neighbourhood at the preceding time step, and according to a definite set of 'local rules'." (Wolfram 1983, page 603)

This definition captures all the significant aspects of cellular automaton models: they are discrete in space, over time, and in their outcomes. Space is made discrete by adopting a uniform lattice of cells; time is made discrete by using the 'time step' as the least period in which events can occur; and outcomes are made discrete by forcing cells to exist in states from only a limited number of allowed states. The definition does not begin to hint at the rich variety of behaviour which may be exhibited by CA, and the consequent range of phenomena which they have proved useful for investigating. Wolfram's (1983) paper refers to around 50 other papers primarily concerned with the application of cellular automata rather than the theory of their behaviour.

The canonical example is Conway's 'Game of Life' (Levy 1992, Poundstone 1985). In the Life CA the cellular space is an infinite grid of cells. Each cell's neighbourhood consists of its eight immediate neighbours in the grid, that is, four orthogonal neighbours and four diagonal neighbours.[8] Each cell may be in one of two states: 'alive' or 'dead'. The rules governing the time evolution of any initial configuration are simple: If a cell is alive and has two or three live neighbours it 'survives' and remains alive in the next time step, otherwise it 'dies'; if a cell is dead and has exactly three live neighbours it 'gives birth' and is alive in the next time step, otherwise it remains dead. The simplicity of the rules disguises an almost incredible range of dynamic behaviour (Hensel 1996, Poundstone 1985). In fact, the Life CA has been demonstrated to be rich enough to support universal computation (Berlekamp, Conway & Guy 1982), which was what Conway was hoping to demonstrate when he devised it. The interesting feature is the apparent disjunction between a very simply specified cellular automaton and very complex behaviour. It is this feature more than any other which has led to the great interest in cellular automata models in the last two decades.

The complexity of the Life CA is a good example of the emergent complexity mentioned in section 2.2.1 (pages 32ff.). The idea that such complexity results in systems whose behaviour may be understood at more than one level is particularly well captured by the way in which it has proved possible to identify configurations in the Life CA whose behaviour is describable without reference to the underlying rules. The best known example of such a Life configuration is the 'glider'. A glider consists of 5 live cells arranged as shown in figure 10 at time step $t = 0$. The sequence of four time steps results in the changes shown, at the end of which the configuration has moved one cell up and to the right. The behaviour of a glider can therefore be described without any knowledge of the underlying rules, since, unimpeded it will progress diagonally across the cell space. Although the idea that a glider is emergent is evocative, some deep philosophical issues are involved here: principally a distinction between emergence arising from our inability to *analytically predict* certain CA

[8]This eight cell neighbourhood in a grid may be referred to as the 'Moore' neighbourhood. A neighbourhood of the four orthogonally adjacent cells only is the 'von Neumann' neighbourhood.

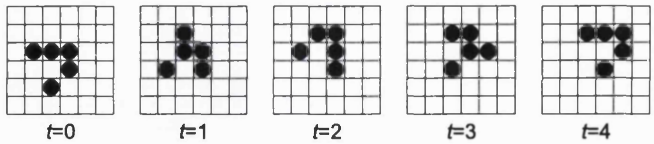


Figure 10 The Game of Life 'glider' configuration — an example of emergence. 'Live' cells contain black circles.

behaviour, and emergence in the sense touched on in section 2.2.1 where ontologically distinct new entities are involved (Faith 1998). It is important to remember that a 'glider' exists only in our perception of the Life CA, it is not an actual entity, altering the behaviour of its constituent cells. Nevertheless, the riches of the Life CA should be clear, and 'glider guns' — configurations which produce gliders at regular intervals — are key to the proof that Life is capable of universal computation, and therefore of arbitrarily complex behaviour. Poundstone (1985) provides details of many other configurations which also occur frequently, and have well understood properties.

3.3.2 Properties of cellular automata

The complexity exhibited by Life has stimulated others to search for underlying order in the behaviour of cellular automata. Current research is still dominated by the ground-breaking efforts of Wolfram (1983), who studied the behaviour of the simplest possible cellular automata. To clarify the discussion I will define the *cell space* to be the arrangement of cells and their neighbourhoods. In Wolfram's simple cellular automata the cell space consists of a linear array of cells, where each cell has as a neighbourhood itself and its two immediate neighbours in the linear array so that the neighbourhood size is three (conventionally, in a CA every cell is its own neighbour). Each cell may be in two states '0' or '1'. The cell space may be defined with reference to the set of allowed cell states $A = \{a_i\}$,[9] and the number of cells in each cell's neighbourhood k. There are $|A|^k$ possible neighbourhood configurations, in this case $2^3 = 8$.

The other aspect of the cellular automaton requiring definition is its *transition rule*. This can be thought of as a 'look up table' which defines for each possible neigh-

[9]Note that S is the more conventional notation for the allowed state set, but we will encounter another S later, so A is used here, and elsewhere in this thesis, for clarity.

bourhood configuration what the resulting state at the central cell will be. If we think of drawing up the look-up table representing the transition rule, since each of the $|A|^k$ rows in the table may be filled with any one of $|A|$ entries, there are clearly

$$|A|^{|A|^k} \tag{3.16}$$

possible cellular automaton rules associated with a given cell space. We can think of this as the size of the *rule space*. Wolfram's simple CAs have a rule space size of $2^{2^3} = 2^8 = 256$ possible rules. A two-state grid-based CA with the Moore neighbourhood has $2^{2^9} = 2^{512} (\approx 10^{152})$ possible rules. This is very, very large! The Life CA is only one possibility from among this number.

Introducing symmetry constraints to Wolfram's much simpler cell space (so that, for example 001 and 100 neighbourhoods both result in the same final state) results in only $2^5 = 32$ possible cellular automata. This is a manageable number and Wolfram proceeded to investigate the behaviour of all the possible cellular automaton rules of this very simple type, by simulation. His work set the pattern for recent cellular automata research, much of which has been concerned with exploring the rule space of given cell spaces (often linear), with a view to determining what criteria must be satisfied by a particular rule for interesting behaviour to be observed.

This begs the question of what constitutes interesting behaviour. Wolfram's (1984*b*) work was again ground-breaking, and is essentially an attempt to operationalise the intuitive ideas about complexity which we have already encountered in the Life CA. Using a variety of methods, Wolfram identified four distinct classes of behaviour:

Class 1 cellular automata evolve in a short period of time to a unique homogeneous state.

Class 2 cellular automata act as 'filters' so that the number of instances of the sequences they filter which are present in the initial system state is a good predictor of the density of their final stable state. These CA are statistically predictable. They may result in final states which exhibit periodicity over time.

Class 3 cellular automata evolve to aperiodic, chaotic behaviour, which although it is totally deterministic, is unpredictable, except with complete knowledge of the initial state. Any finite segment of an infinite automata in this class is not predictable, since its state will eventually depend on the initial states at arbitrarily remote sites. Typically, such cellular automata tend towards chaotic patterns whose average density of cells in any given state is stable, and similar for all initial states.

Class 4 cellular automata exhibit complex behaviour. Complexity is characterised by the evolution of structures in the cellular automaton which may propagate and interact with one another, and whose behaviour can be described and partially predicted without reference to the underlying rules (the Life CA is complex by this definition). Complex behaviour is characteristic of systems capable of computation, which can be 'programmed' by choosing suitable initial states, and are therefore capable of arbitrarily complicated behaviour.

Wolfram's classes are by no means unambiguous, and even in his own work he seems to be unsure as to whether certain 1-D linear CA fall into class 4 or not. However, ambiguously defined or not, class 3 and class 4 cellular automata exhibit behaviours which are highly suggestive of real-world phenomena, and this has stimulated a great deal of interest in determining what conditions must be satisfied by the transition rules for given cell spaces to operate in these domains. It is notable also that there is a close match between the Wolfram behavioural classes of these discrete dynamic systems, and the 'attractors' of continuous dynamics. Class 1 behaviour approximates to a limit point; class 2 to a limit cycle; class 3 to chaotic — or 'strange' — attractors; and class 4 behaviour seems to relate most closely to the dynamic behaviour of non-equilibrium dissipative structures in thermodynamics (Prigogine & Stengers 1984).

Langton (1990) devised a simple parameter, denoted λ, which appears to offer some possibility of predicting which class of behaviour a given rule will result in. λ is calculated simply as the proportion of neighbourhood states which result in some 'quiescent' state a_q in the next time period. Using the earlier terminology if r_q rules in the rule space lead to a_q then

$$\lambda = \frac{|A|^k - r_q}{|A|^k} \tag{3.17}$$

If all $|A|^k$ possible neighbourhood states lead to a_q, then $r_q = |A|^k$ and $\lambda = 0$. With $|A|$ cell states a randomly generated set of rules will have $r_q = \frac{1}{|A|}$. Langton therefore restricts his investigation to the region of the rule space where $0 < \lambda < 1 - \frac{1}{|A|}$, which is the portion of the rule space in which a_q is 'over-represented'. Langton investigates the behaviour of a simple linear cell space with $k = 5$ and $|A| = 4$. Since $|A| = 4$, λ is varied from 0 to 0.75. The interesting outcome is that at some critical value λ_c there seems to be a 'phase transition' from class 2 to class 3 behaviour. Langton speculates that class 4 behaviour is found near this phase transition (see figure 11). Intriguingly the value of λ for the Life CA is near Langton's critical value.

The evocative — although probably over-enthusiastic — further speculation is made that systems at the 'edge of chaos' (near λ_c) are 'alive' in some previously undefined sense. This notion has become one of the corner stones of North American research into artificial life. Others have challenged Langton's parameterisation of rule space as too simplistic and it is relatively easy to argue that it may be inadequate. Mitchell, Crutchfield & Hraber (1994) use the example of the two state automaton which would be required to transform a random array of cells into a homogeneous array in one state depending on which state was in the majority in the initial configuration (this is called the 'density classification problem'). Arguing from logic they show that this particular class 4 automata cannot have a λ parameter near to Langton's value. This argument shows some of the weaknesses in Wolfram's original classification scheme, since it is not absolutely clear that density classification is complex

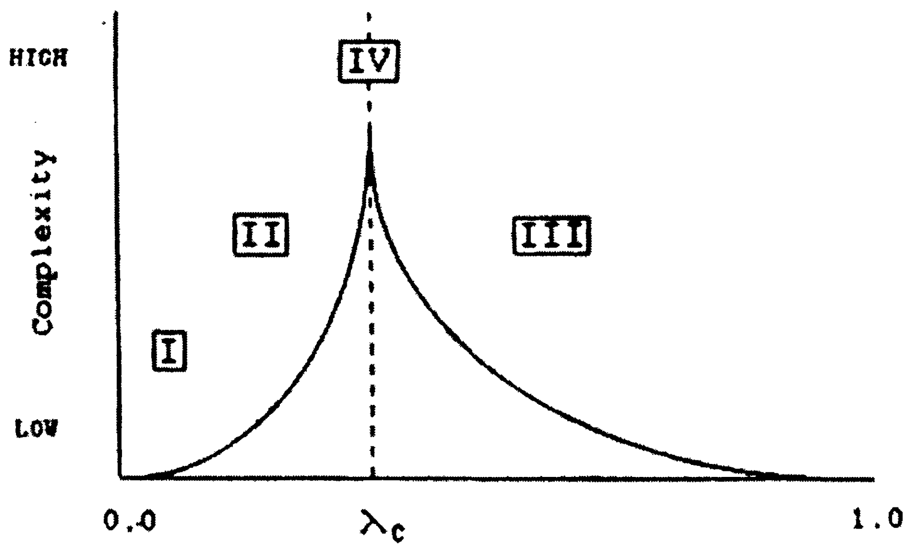


Figure 11 The phase transition in CA rule space from class 2 to class 3 behaviour as the λ parameter is varied (*source*: Langton 1990).

in the Wolfram sense — in fact, it has much in common with Wolfram's class 2.

There are other approaches to the investigation of cellular automata. More complex rule parameters have been described (Gutowitz 1989). The 'attractor-basin portrait' approach, which draws directly on dynamic systems theory, examines the trajectory of different CA in state space, that is the dynamics of how a particular cellular automaton moves over time between its various possible system states (Hanson & Crutchfield 1992, Wuensche 1994). Building again on earlier work by Wolfram (1984*a*), others have used formal language theory to model and predict CA behaviour using a new methodology called 'computational mechanics' (Hanson & Crutchfield 1997). This is related to investigations of the dynamics of CA 'particles' and their interactions (Hordijk, Crutchfield & Mitchell 1996). Particles are very much like gliders in the Life CA and occur when a pattern emerges in the evolution of a CA which moves through the cell space. In class 4 automata such particles are regular occurrences, and the CA behaviour can be characterised in terms of the relative frequencies of different particles, their speeds and their interactions. It is speculated that the development of useful CA (for complex processing tasks, for example), will depend on such particle-based CA, since particles can communicate information about local states between remote parts of the system, an essential function in complex processing tasks. Figure 12 shows a few such particles in a class 4, 1-D CA. Yet another perspective on CA behaviour uses ideas from information theory (Lindgren 1987, Lindgren & Nordahl 1988).

It is noteworthy that most of the mathematical, computer and physical science research on CA behaviour still focuses on linear CA. This is obviously an attempt to simplify matters, and does allow for the much more ready visualisation of CA outcomes over time, since a linear array of cells evolving over time can be conveniently presented as a two-dimensional image in its entirety (figure 12 is an example). However, linearity may also be a serious limitation. For example, the computational mechanical approach, based on the representation of CA rules as formal languages is entirely dependent on the natural ordering of cells which 1-D CA intrinsically possess. It is therefore difficult to see how this particular approach can be extended to two and higher dimensional cell spaces. This is an important observation in the context of applying CA models to two (or higher) dimensional problem domains, as is clearly required in geography and urban modelling. This does not make the applica-

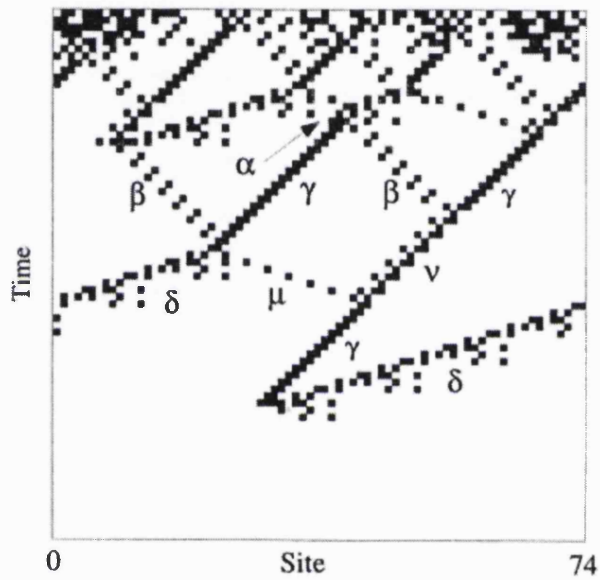


Figure 12 Particles and their interactions in a CA. Note the two-dimensional representation of the time evolution of a 1-D CA, with system states represented as rows across, and consecutive states stacked on top of one another down the page (*source*: Hordijk et al. 1996).

tion of CA models to these domains invalid, but it does make it important that any implications of the move to higher-dimensional CA be understood.

3.3.3 Cellular automata measures

For our current purpose, rather than examining any of the investigative approaches mentioned above in great detail (especially given the limitations of the 1-D case just mentioned), it is important to extract from them useful measures for characterising cellular automata behaviour. The key measure for this purpose is *entropy*. Entropy can be roughly defined as 'the amount of disorder'. The reason for its centrality to cellular automata behaviour studies is that it provides a way of detecting complex, class 4 behaviour. In classic equilibrium physics, which is concerned with the evolution of closed systems, the Second Law of Thermodynamics states that while energy is conserved in all processes, entropy is non-decreasing. No process in a closed system can reduce the entropy, that is, no closed system can spontaneously move from a disordered state to an ordered state. If a box full of ping pong balls coloured black and white is initially sorted into two distinct parts of the box and vigorously shaken it will tend to become disordered. No amount of further shaking is likely to cause the balls to sort themselves out again. What distinguishes complex systems is that they do spontaneously organise themselves — there is no clash with the second law because such systems are not closed systems: they are dissipative, which means that energy is used up to increase the order or structure (Prigogine & Stengers 1984). Such systems absorbing energy from their surroundings and use it to reduce their (internal) local entropy. Overall entropy in the system *plus* its surroundings will be conserved or rise, as waste products and heat are dissipated.[10] Hence, if the entropy local or internal to a system falls, it is becoming more ordered and it may be inferred that complex behaviour is occurring.

Wolfram (1984*b*) defines various entropies for characterising CA behaviour. *Spatial set entropy* is defined as

$$S^{(x)}(X) = -\frac{1}{X}\sum_{j=1}^{|A|^X} p_j^{(x)} \log_{|A|} p_j^{(x)} \tag{3.18}$$

[10]Note that we must distinguish *actual* CA from the Platonic or formal ideal of CA to make sense of the dissipative structure concept in this setting (Faith 2000).

where $p_j^{(x)}$ denotes the frequency of occurrence of each of the $|A|^X$ possible neighbourhood states in sequences of cells of length X. The (x) superscript in each term indicates that the measure is made at some specific location x. This measure is in the range 0 to 1, since it will be a maximum where each $p_j^{(x)} = |A|^{-X}$, when $\log_{|A|} p_j^{(x)} = -X$, so that the sum becomes $\sum_{j=1}^{|A|^X} |A|^{-X} = 1$. A value of 0 indicates a very ordered arrangement of cell states, where only one of the possible neighbourhood states is present, and 1 indicates a completely disorganised arrangement of states, where every possible neighbourhood state is equally represented. Note that the measure is applied over some sequence length of cells, so that a particular configuration of cell states may seem random at $X = 1$, but exhibit order at some other value of X.

Wuensche (1998, 1999) sets $X = k$ and evaluates the entropy using each cell's neighbourhood. He terms this the *input entropy* since it is calculated by examining the set of inputs to each cell at each time step. This seems a very natural way to use the measure, with the added advantage that the required statistics can be conveniently compiled at the same time as the determination of the next time step is being performed. Figure 13 shows respectively ordered, complex, and chaotic behaviour for simple linear CA along with (sideways on) plots of the time evolution of input entropy measured over cell neighbourhoods. The different behavioural classes can clearly be distinguished in this way. There are some simplifying assumptions in Wuensche's approach, which again relies in part on the inherent ordering of cells in a linear CA. In section 5.3.5 (pages 118ff.), I present an extension of Wuensche's input entropy which attempts to allow for its calculation in two- and higher-dimensional systems.

Note that an analogous temporal entropy can be defined for the time sequence of states at a particular site:

$$S^{(t)}(T) = -\frac{1}{T} \sum_{j=1}^{|A|^T} p_j^{(t)} \log_{|A|} p_j^{(t)} \tag{3.19}$$

where $p_j^{(t)}$ denotes the frequency of occurrence of each of the $|A|^T$ possible sequences of states in the time period T for which the measure is being evaluated.

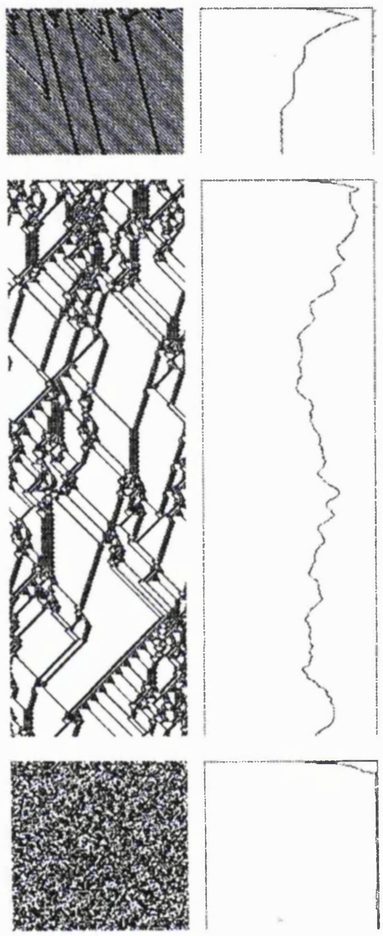


Figure 13 Use of input entropy to detect a CA's Wolfram class. The top example sees input entropy settle quickly to a stable value indicating class 1 or 2 behaviour. The middle case displays long-term variation below the maximum entropy which is indicative of organization and complex class 4 behaviour. The bottom example is class 3 with unpredictable high input entropy — or chaos (*source*: Wuensche 1998).

3.3.4 Conclusions on cellular automata

It is clear from the foregoing that a cellular automaton based model of a process is primarily concerned with the *dynamics* of the situation. The Wolfram classes provide a short-hand for broad classes of dynamic behaviour which may occur. Determination of the behavioural class of a CA model does not allow the modeller to make accurate predictions of outcomes at any particular location in the cell space, however, it does enable general predictions about the sorts of phenomena which are likely to occur to be made.

The key insight which CA models provide is the way in which globally complex and even apparently random (actually chaotic) configurations can arise out of simple behavioural relationships at a local level. By manipulating the rules in a CA and examining the resulting behaviour of the model, investigators can also explore the relationship between phenomena at different scales.

3.4 Conclusions

In terms of the discussion of spatial models presented in the preceding chapter, graph representations of spatial situations combine aspects of both absolute and relational spatial models. The vertices typically represent well defined objects located in absolute space, while edges may represent spatial relations of adjacency, nearness or whatever. However, since objects need not necessarily be well defined in absolute space (Campari 1995), a graph representation is essentially a relational model of space. It is the relations between elements which are represented, and it is the structure of these relations which measures of graph structure reveal. Of course, the graph itself is aspatial and the relations may be completely non-spatial or some combination of non-spatial and spatial relations. The addition of non-spatial relations to a graph representation makes the underlying spatial model more akin to a proximal conception of space, since the additional relations (edges) may be added to account for effects of functional influence in the system.

The model of space incorporated in CA models is different. The partitioning of space into a regular array of cells is effectively an absolute approach since it implicitly creates a co-ordinate system with respect to which the location of any individual cell can be specified. On the other hand CA might also be seen as relational, since only properties, not objects, are defined across the grid, and 'objects' only exist as patterns in the values of those properties. However, the centrality of the neighbourhood concept to the specification of rules in CA is really the origin of the proximal conception of space (Couclelis 1997, Takeyama & Couclelis 1997) and it is clear that we may also — perhaps best — understand CA models in this way. Which conception is more relevant in any particular model would seem to depend on the way in which that model is constructed. A grid-based model with a very large cell space, in which each cell represents only a very small portion of the 'world' seems particularly likely to embody an absolute conception of space. Where the cell space is small it is more likely that cells might represent objects, and a proximal or even relational conception of space is possible.

In Chapter 5, where a new modelling approach based on both of these tools is proposed, it is this connection of both to a proximal model of space that allows them to be combined in a straightforward way, without loss to the usefulness of the analytical concepts, measures and tools associated with each.

Chapter 4

An overview of urban morphology and micro-scale analysis

This chapter reviews a selection of work in the study of urban form at a micro-scale. My purpose is twofold. First, I am concerned to show that many approaches to urban morphology have tended to separate structure and process, often concentrating on one to the exclusion of the other. The modelling approach which I propose in chapter 5 attempts to address this weakness. Second, as discussed in chapter 2 (see especially figure 3 on page 35), I want to use urban morphology — predominantly at the scales identified by Moudon (1997) in the passage quoted in section 1.1 — as a source of ideas about spatial elements which can be used in the construction of proximal spatial models of cities.

As a result of these twin foci I do not spend time on any of the more broad brush stroke histories of the city (Mumford 1961, for example), nor on those approaches which fit 'the city' into much wider perspectives (Castells 1989, Harvey 1978, for example). Also, I am not concerned with models which may be broadly labelled operational land use–transportation models (Torrens 2000, reviews many of these). As a result, most of the work considered here originates in the geography, planning and architecture literature, at the scale of Montello's (1993) vistal or environmental spaces, rather than at the scale of regional models, although this focus does not mean that the modelling approach developed in the next chapter could not be applied at those larger scales.

4.1 An introduction to urban morphology

Urban morphology is a field which *does* need at least *some* introduction! There is an extensive urban morphology literature in architecture and planning theory. Prior to the 1960s this tradition was largely prescriptive and utopian, concerned with outlining 'ideal cities'. The *Garden Cities* of Howard (1898), Le Corbusier's (1929) *Ville Radieuse*, and Wright's (1945) *Broadacre* are the outstanding examples of this genre. The 1960s' critique of what became a planning orthodoxy loosely based on aspects of these manifestos, inspired many and sowed the seeds of more analytical approaches to urban form. The most prominent exponents of the critique were Lynch (1960), Jacobs (1961; 1994) and Alexander (1964, 1965) who argued for a more humane approach to city planning, based on observation of what actually works in existing cities. Crucially, Alexander's work introduced formal mathematical concepts into the debate for the first time.

A range of subsequent work in urban morphology attempted to grapple with what works, and what is possible in the urban design arena (March & Steadman 1971, Martin & March 1972, Steadman 1983, are typical). These are chiefly of interest here for the introduction of mathematical tools into the realm of urban morphology, in particular graph theory and set theory. An intriguing combination of these, *Q*-analysis (Atkin 1974*a*,*b*, 1975) was also influential. Work focusing on possible graph representations of urban form (Krüger 1979*a*,*b*) exemplifies the approach. More often applied in practice is a body of work known as 'space syntax' (Hillier & Hanson 1984, Hillier, Penn, Hanson, Grajewski & Xu 1993). This is also based on a graph representation of cities as systems of open space. Simple structural measures of the resulting graphs allow the structure of a city to be described.

Work in human geography in this field also has a long history. In Britain, it has its origins in the work of a German geographer M. R. G. Conzen. His work is in turn based on that of German geographers from the turn of the century onwards, such as Schlüter. Whitehand (1981, pages 1–24) provides a thorough overview of this history, and firmly grounds it in the German geographic tradition. In the hands of Conzen, it seems that the early German commitment to a thorough, almost geological, examination of historical evidence became a powerful tool for elucidating the development of towns and cities in Britain, especially in the north east. Conzen's (1960) *Alnwick, Northumberland: A Study in Town-Plan Analysis* is a landmark study. In it, Conzen develops two of his key ideas, the *burgage cycle*, and *fringe belts*. Both these phenomena (which are defined elsewhere in this chapter) are identified after a thorough study of the evolution of the town-plan based on historical map data. Whitehand (1967, 1972, 1987*a,b*, 1992) and Slater (1980, 1998) are pre-eminent amongst the researchers who have made contributions in this tradition more recently.

Two other European schools the 'Versailles school' and an Italian school originating with the work of Muratori and Rossi can be identified, but relatively little of this work is available in English (Rossi 1982, is an exception). Although the original motivation for this work is different, there are similarities in approach. The complex recent history of these groups is outlined by Moudon (1997). A similar American approach to urban morphology is to be found in Vance (1990), apparently with little reference to the large body of European work. There are more recent signs of communications channels opening up between some of these schools, especially in the recently launched journal *Urban Morphology*.

Some of the approaches touched on above are considered in more detail in the sections below. This is followed by a discussion of more dynamic micro-scale urban modelling approaches based on CA, which have a shorter but comparably rich history. Here, local rules for growth of cities are seen to produce city-like emergent forms. This leads to interesting conclusions about the degree to which cities can be planned at all, and also leads us back to a concern with urban form elements at the smallest scales. The theory of the 'fractal city' (Batty & Longley 1994, Batty & Xie 1996) provides a direct link between these models and more purely morphological concerns.

4.2 Jacobs's polemic and Alexander's analytic critiques of planning

One of the seminal works in the post-war development of urban planning was Jane Jacobs's (1961; 1994) *The Death and Life of Great American Cities*, a searing attack on the rationalist, modernist tradition of planning orthodoxy. She asserts that one of the consequences of apparently rational design approaches is the creation of lifeless urban places. Jacobs's recipe for the creation of lively cities is a call for recognition that

> "[u]nder the seeming disorder of the old city, wherever the old city is working successfully, is a marvellous order [...] It is a complex order. Its essence is its intricacy [...]" (Jacobs 1961; 1994, page 60 in the 1994 edition)

We may accept Jacobs's diagnosis, without learning very much about how to design urban places which promote and retain such marvellous order. Perhaps the earliest *analytical* attempt to describe the complex order desired, and thus how to *design* for it, was provided by Christopher Alexander in his ground-breaking paper, 'A city is not a tree'. The crux of his argument is that "I believe that a natural city has the organisation of a semi-lattice; but that when we organise a city artificially, we organise it as a tree" (Alexander 1965, page 58). This assertion is based on the morphological description of urban form using mathematical set concepts. Alexander's idea is that any concept of the city must necessarily involve breaking it down into units. These units, in turn, may be grouped into sets of related elements. A semi-lattice arrangement of such elements is formed when any intersection of two sets of elements is also a potentially meaningful set of elements. In contrast, a tree arrangement exists when only two possible relations may exist between sets of elements: either they share no elements (they are disjoint), or one is completely contained in the other (one is a subset of the other). The difference is most readily understood from the illustration in figure 14. The graph representations of the set relations in the lower half of this diagram are particularly interesting in the current context.

The essence of Alexander's argument is that urban designers tend to think in terms of more hierarchical and less complex tree-like structures, and not in terms of the more subtle semi-lattice structures which exist in organic cities. Planned and zoned cities are forced into tree-like structures because it seems like the rational sensible thing to do, and because that is the way that rational design methodology works. In a more detailed exposition, Alexander (1964) extends his argument to apply to the generality of design problems, and a case study design methodology which might help avoid such thinking is presented. Since this work, Alexander has spent many years working on similar ideas and trying to evolve a 'pattern language' — a detailed system of design rules — which could be used to design semi-lattice structures without simply creating chaos (Alexander, Ishikawa & Silverstein 1977). Opinions as to

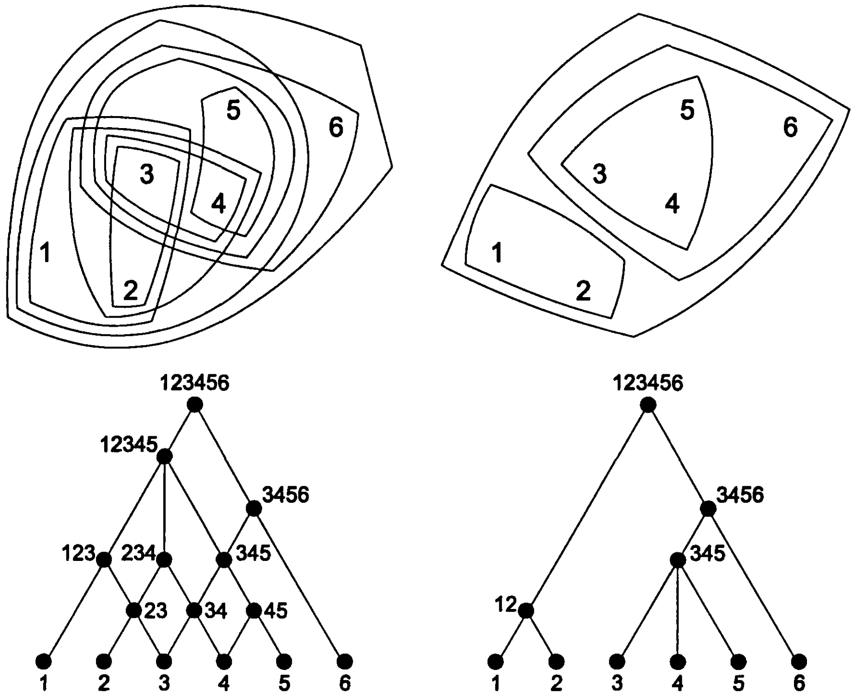


Figure 14 A semi-lattice and a tree, based on Alexander (1965).

Alexander's success differ.[1]

For our purposes, the most interesting development is the introduction to discussions of urban form and morphology, of abstract representations, in particular the graph representation of relationships between entities.

4.3 The analytical planning literature

Batty, Couclelis & Stiny (1994, page s3) comment that "[t]he notion that mathematics as models inside computers might be the vehicles to progress these ideas was so tantalising that few could resist its fascination."

Although Alexander's ideas about design were primarily a critique of architectural and urban design practice, his introduction of set, systems and graph theoretic ideas into the field was influential and a significant body of work on systematic design approaches started to emerge. Much of this work had more to do with the automatic

[1] Garreau (1992) is enthusiastic; others have questioned the original formulation or pointed to its inadequacy; see, for example, Harary & Rockey (1976). It is arguable that 'mixed use' as a core ideal of contemporary planning has its origins in these exchanges, and as such has become a new orthodoxy.

generation of urban and building plans than with the representation or understanding of urban systems as they are, and was extremely optimistic about the prospect of such approaches 'solving' the urban design problem. In retrospect this seems naïve, and the robust polemic of Jacobs has arguably had a more enduring impact than any number of formal design approaches.[2] However, our interest is primarily in such formal approaches to the representation of urban systems and some of those are reviewed in this section.

4.3.1 The 'Cambridge school'

Formed in 1967 at Cambridge University, the Centre for Land Use and Built Form Studies (CLUBFS) — renamed the Martin Centre for Architectural and Urban Studies in 1973 — was a pioneer in this field. Two publications in the early 1970s — March & Steadman's (1971) *The Geometry of Environment*, and Martin & March's (1972) edited collection *Urban Space and Structures* — exemplify those aspects of the research at the centre which are of interest here. March, Echenique & Dickens's (1971) call to arms in *Architectural Design* gives an idea of the ambitions of the group. A later collection of articles shows some further development of these ideas (Steadman 1983). These are drawn particularly from *Environment and Planning* (latterly *Environment and Planning B: Planning & Design*), the journal which emerged from this area, and continues to publish much of the research in the field. In the spirit of this review, I intend only to examine one aspect of this work which is of direct relevance to the current work in its use of graph representations of urban morphology.

Two papers by Krüger (1979*a*,*b*), derived from his doctorate at the Martin Centre, develop the idea of multiple graph representations of the urban system. He proposes that the urban graph system be subdivided into the *channel network* and the *built form galaxy*. The first of these graphs is based on the street network, and is familiar from transport geography and transport planning, where graph representations are both obvious and natural. The built form galaxy is less obvious and is based on abstract representations of *built form units*. Working from the most detailed level, *built forms* may be represented by a graph which shows the interconnections between rooms. *Built form arrays* can be represented by graphs showing the relations between their

[2] It is also ironic that a diagnosis of 'too much rationalism' should lead to liberal doses of a different brand of rationalism! — and perhaps indicates that rationalism is in the eye of the beholder...

constituent built forms. *Built form constellations* in turn are represented by graphs showing the relations between constituent built form arrays. Krüger's terminology is rather confusing but is clearly defined. Built-forms or *built form units* (BFUs) can be thought of as any building, or part of a building, with a single address, so that a single house in a terrace is a BFU. A built form array is a connected set of BFUs such as a terrace; and a built form constellation is a set of built form arrays in a single block. A typical example of the graphs produced by this approach is shown in figure 15.

On the basis of these definitions a range of possible graphs representing the morphology of the built up areas of settlements is proposed. A set of measures which might be used to characterise the resulting graphs is proposed and examined for redundancy and complementarity. The measures proposed are measures of overall graph structure and do not relate to particular vertices or individual elements in the graphs. Krüger settles on a subset of his original measures which could be used to distinguish different regions in a single settlement, or to specify the differing characteristics of different settlements.

In later papers (Krüger 1980, 1981*a*,*b*) these measures are further analysed and correlated with a variety of separately derived measures of 'urban structure'. The urban structure measures relate to aspects such as residential densities, employment rates, service availability and so on as used in earlier work at CLUBFS (Crowther &

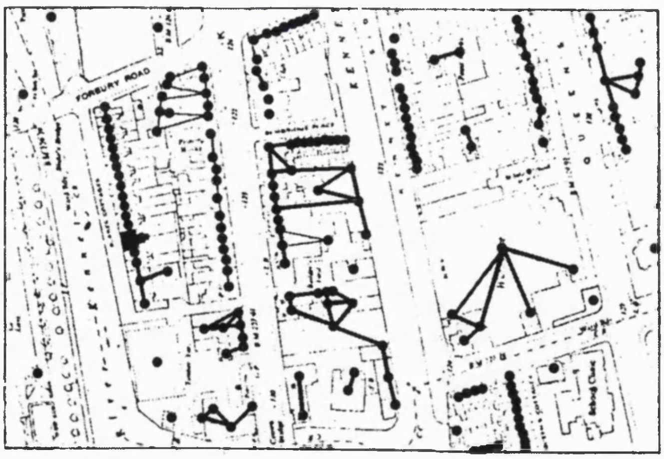


Figure 15 Typical urban built-form graphs (*source*: Krüger 1979*a*).

Echenique 1972, Echenique, Crowther & Lindsay 1972). Whether or not the measures have any predictive power is not considered; that they have, is taken as self-evident. As Echenique & Owers (1994, page 513) remark

> "[t]he initial work of the researchers at [the Martin] Centre was highly abstract and only remotely connected with the work of practitioners in the fields of architecture and urban planning."

The work of the Martin Centre has been central to the development of widely used land use–transport models, particularly Echenique's MEPLAN. For now, the important idea is the way in which abstract measures of the properties of graphs representing aspects of the built environment can be derived and related to other measures or aspects of urban development. Recent work on the automatic classification of built-up areas according to land cover as determined by remote-sensed imagery (Barnsley & Barr 1997, Barr & Barnsley 1997) shows the continuing relevance of this type of approach (see figure 16). Here, the graphs are based on the adjacency relations between homogeneous regions of ground cover, as detected in the remote-sensed data.

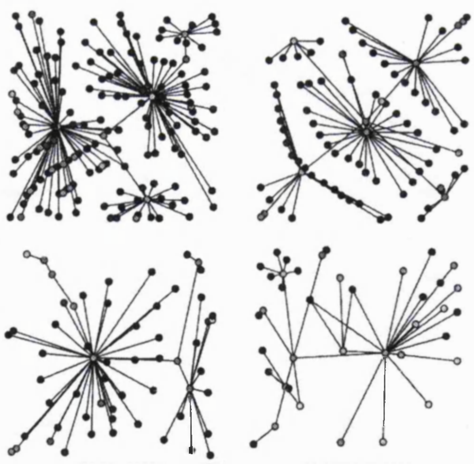


Figure 16 Typical graphs automatically generated from remote sensed imagery of different types of built environment — from left to right, top to bottom: a 1990s UK housing development, a late 1970s/early 1980s housing estate, 1930s semi-detached housing, and a hospital complex (*source*: Barnsley & Barr 1997).

4.3.2 *Q*-analysis

A methodology which makes a brief appearance in Krüger's work (1980, pages 190–193) is *Q-analysis*. Q-analysis is an analytical approach to binary relations between two sets of objects. In urban morphology, for example, it might be between streets and a particular land use. Beaumont & Gatrell (1982) provide a useful introduction which is easier to follow than some of the original papers of the developers of Q-analysis (Atkin 1974*a*,*b*, 1975, Atkin, Johnson & Mancini 1971).

In Q-analysis the relationships between two sets A and B are used to build a multi-dimensional structure — a *simplicial complex* — which can be described and analysed using some relatively simple geometric ideas. A and B must be inter-related sets for the method to have any application. In a worked example, Atkin (1975) uses a set of street segments in a town, and the set of land uses. The relations between $A = \{a_1, a_2, \dots, a_m\}$ and $B = \{b_1, b_2, \dots, b_n\}$ are first summarised by cross-tabulating the two sets in a matrix, say $\mathbf{Z} = [z_{ij}]$, where $z_{ij} = 1$ if a_i is related to b_j — in the streets and features case this would mean that feature type b_j is found in street segment a_i. The asymmetric matrix $\mathbf{Z}$ gives rise to two symmetric matrices $\mathbf{ZZ}^{\mathrm{T}}$ and $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ which can be represented as multi-dimensional structures.[3] These summarise the inter-relationships of the original two sets, and can be subject to further descriptive analysis and investigation.

The details of the methodology are not of direct concern here, and indeed its limitations (Couclelis 1983) and the close relationship of Q-analysis to clustering techniques (MacGill 1984) were eventually realised. The main interest in Q-analysis in the current context is that it provides another way of exploring the structure of spatial relations in urban environments.

It is also explicitly concerned with the relationship between *structural* aspects of an urban environment and *events* or *processes* which may occur in that environment. In common with many other aspects of the methodology this interest was expressed in rather idiosyncratic language. Atkin (1974*a*, page 57) introduced the idea of the "static backcloth of the urban structure", going on to say that the backcloth "offers a stage for the dynamic interplay of patterns" (Atkin 1974*a*, page 57). He further suggests that the geometry or topology of the backcloth in some sense permits certain

[3]Note the use of the inter-related matrix products $\mathbf{ZZ}^{\mathrm{T}}$ and $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$, and how this relates to the relationship between a graph and its line graph (see section 3.1.1 on page 43).

patterns and impedes others. In his explanation of this idea Atkin draws on ideas from physics about how the geometry of space affects the paths of particles and other physical phenomena (Atkin 1974*a*, pages 64–66). This recalls Nerlich's (1994) philosophical arguments about the 'shape of space'. Further ideas such as traffic, noise and even "a theory of surprises" (Atkin 1981) are introduced later. Not all of these concepts are particularly clear, in spite of later attempts to clarify them (Griffiths 1983, Johnson 1981, 1983). It is also striking that many of the contributions to the literature of the methodology use illustrative examples on a small scale, rather than extend it to potentially more interesting 'real-world' cases (although see Gould & Gatrell's, 1979 analysis of a football match!). However, the idea of relating structures to processes (backcloth to traffic) is an interesting one — albeit a difficult one for an essentially static methodology to grasp — and is a key idea in the current work which is developed later.

4.3.3 Space syntax

Another approach to representing urban structure which is rooted in a relational representation is *space syntax*, which I will discuss in more detail, because it has seen more actual application than either Krüger's graphs or Q-analysis, and because it raises some interesting issues.

Space syntax was originally developed in the Advanced Architectural Studies Unit at the Bartlett School of Architecture and Planning at University College London (UCL). It has latterly been developed by the 'Space Syntax Laboratory', a commercial venture which retains links with UCL. The ideas behind space syntax are outlined in Hillier & Hanson (1984). The basic spatial elements are relatively enclosed spaces and access routes between them. In an enclosed building the identification of these elements is a fairly obvious process, where enclosed spaces are identified with rooms and access routes are the connections between them.[4] This approach has its origins in the architectural provenance of this work, particularly in the graph analysis of complex building plans (March & Steadman 1971, Martin & March 1972, Steadman 1983).

When applied to the open public space of a city the definition of these elements is less obvious. The formulation which is adopted is to represent urban space as a

[4] Although modern, open plan buildings may present difficulties.

collection of *convex spaces* and *axial lines*. These spatial elements are loosely related to some theoretical ideas about the way in which space is occupied by moving people (often strangers), and stationary people (often residents). Axial lines relate to ways in which the overall structure of a system can be accessed by moving people — since they summarise some of the visual interrelationships in the system. Convex spaces on the other hand are distinguished by the fact that they promote awareness of mutual co-presence, since every location in a convex space is visible from every other. Thus an axial line and convex space representation captures some of the cognitive and social aspects of an urban space, and might even be understood as a (rough) formalisation of a scheme like Montello's (1993) (see section 2.1.3 on pages 28ff.).

The analytical approach is to first decompose the public space of the city under investigation into the smallest set of 'fattest' spaces which covers the entire space. These are the convex spaces. Having achieved this the smallest set of longest lines inside the public space, which together cover (i.e. intersect) all of the convex spaces at least once is constructed. This second construction is the *axial map* and is key to the most common further development of the analysis. The two representations are shown in figure 17. It is worth noting that the construction of the convex map is non-trivial, and similar problems in computational geometry such as decomposing a monotone polygon with holes into strong triangles are insoluble in polynomial time (de Berg, van Kreveld, Overmars & Schwarzkopf 1997). It has been argued that the convex decomposition and axial line construction are obvious in practice,[5], but this *is* a weakness in the methodology, and a clear obstacle to its widespread adoption. However, recent developments suggest that heuristic methods for generating the axial map can be devised (Peponis, Wineman, Bafna, Rashid & Hong Kim 1998). Tests of the sensitivity of the analysis to these steps in the process are required to resolve this issue.

At this point in the exposition, a wide range of possible measures, relating to both the convex space representation and the axial representation, are presented (Hillier & Hanson 1984, pages 90–123). Some measures describe the relationship between the two representations, and others also include relationships between the convex spaces or axial lines and building entrances. In later work many of these are not explored further, and analysis focuses on the axial map representation. The axial map is used

[5]Note that 'degenerate' cases are easy to invent (perfectly circular roundabouts, for example) but their practical relevance is difficult to argue.

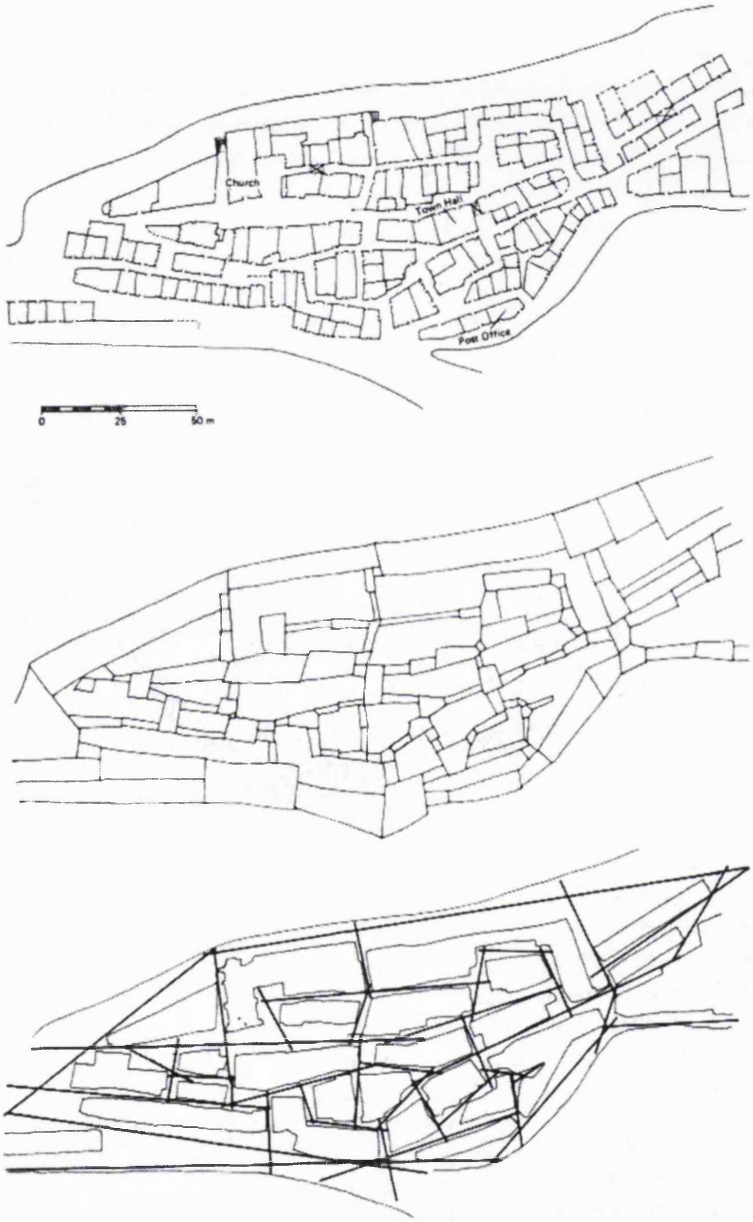


Figure 17 The convex and axial maps of space syntax. The top figure is a map of the study area, the middle image is the convex map, and the lower image is an axial map derived from it (*source*: Hillier & Hanson 1984).

to construct an *axial graph* wherein vertices represent axial lines, and an edge exists between two vertices if the corresponding axial lines intersect in the axial map. This construction is similar to the derivation of the line graph of a graph as illustrated in figure 7 on page 44.

This is a surprising representation, in that the axial lines which were initially conceived as in some sense connecting places (the convex spaces), are transformed into 'places' which are themselves connected by relations of mutual intersection. However, whatever its motivation, the axial graph can be used to represent and describe the urban structure, and this has been done by developing various graph structure measures. In contrast with Krüger's work, the graph measures used are not global, that is, measures are determined for the centrality of each vertex in the graph (for each axial line in the original map), not for the graph as a whole, so that the methodology relates individual elements to the whole system of elements in which they are embedded.

The principle measure in a space syntax study is *integration*, which is a distance based centrality measure (see section 3.2.1, pages 45ff.). Of note is the use of an unusual approach to normalising the distance sum on which the measure is based. Instead of averaging the total distance over the number of vertices included in the summation, the total is located relative to the maximum and minimum possible totals for a graph containing that number of vertices. Thus, whereas the 'conventional' normalisation for some $\sum d$ summed over n vertices would simply be $\sum d/n$, in space syntax analysis it is given by

$$RA = \frac{\sum d/n - (\sum d/n)_{\min}}{(\sum d/n)_{\max} - (\sum d/n)_{\min}} = \frac{\sum d - (\sum d)_{\min}}{(\sum d)_{\max} - (\sum d)_{\min}} \tag{4.1}$$

where the $_{\min}$ and $_{\max}$ subscripts represent the minimum and maximum possible values which the mean (or total) distance could have in a graph of this size. RA denotes *relative asymmetry* and the idea is to locate particular axial lines on a 'spectrum' from very integrated — the central node of a star, to very segregated — the end node of a sequence of lines — hence the (rather obscure) name. Integration values are given by $1/RA$ so that more central locations (low $\sum d$) have higher values (compare equation 3.8 on page 49). Since the extreme values in equation 4.1 are themselves dependent on n this is a more complex normalisation than is usual.[6]

[6]Although it is reminiscent of Freeman's (1977) approach: see equation 3.7 on page 48 and the accompanying discussion.

A further normalisation procedure is intended to allow comparison of integration values between different systems, producing a value known as *real relative asymmetry* (RRA). This calculation is based on values which would attain in regular 'diamond' graph systems of the same size (Hillier & Hanson 1984, provide details, although the explanation is obscure). It is not established that RRA does allow meaningful comparison between completely different urban systems. Some writers have suggested different normalisations which may be more stable (Krüger 1989, Teklenburg & Timmerman 1993), and the issue remains confused.

An interesting variation on distance based measures is also introduced — a radius-limited centrality measure. Instead of calculating $\sum_{j=1}^{n} d(v_i, v_j)$ over all other vertices in the graph, the distance sum is determined only for those vertices in the graph within some specified distance of the vertex of interest. Thus integration *radius-R* is based on the sum of distances from vertex v_i to all those vertices v_j in the graph such that $d(v_i, v_j) \leq R$. Given the use of integration values determined over graphs of widely varying sizes (since, for example a centrally located axial line may have 100 other lines within a radius of 3, whereas a peripheral line might have only 5) the normalisation issue discussed above is not a trivial one. It is arguable that the normalisation procedure used makes low radius integration values a more subtle measure than a simple 'average distance' and instead yields a value which in some sense reflects the structure of the graph on which the measure is based. This idea is rooted in the realisation that only a limited range of graphs are possible for low radius systems, and is similar to Donninger's (1986) remarks about Freeman's normalisations and random graphs. Further — probably mathematically involved — analysis is required to verify this idea. It is a relatively simple matter, for example, to examine the set of all possible radius R graphs of size n, and the distribution of integration values which these yield. Such research would shed light on the properties of these measures, and even preliminary investigation of these issues raises questions about the details of the calculations which are left unanswered in the space syntax literature. For example, is it sensible to use a value for $(\sum d)_{\max}$ based on a path graph P_n, when the n vertices under consideration are constrained by the radius R criterion? Questions like these arise naturally when we examine this work in light of the wider literature on graph analysis.

Whatever the precise mathematics of the various space syntax measures, they appear to have some empirical worth, in particular, correlations between pedestrian flow rates and the syntax centrality measures have been found. As a methodology which aims to provide useful tools to inform planned interventions in the urban fabric, space syntax can therefore be judged a relative success. It invokes an interesting new representation of urban space and analyses it as a graph. The relations between elements in the graph are the subject of the analysis, and the structure of these relations appears to have implications for processes which occur in the city. From the perspective presented in chapter 2, the spatial model used is interesting, since it decomposes space into elements which may be extremely varied in size and shape, but are treated similarly in the analysis. This is a significant departure from established metric geographic approaches to space and is interesting for that reason alone.

However, there are technical issues with the correlations on which the approach's claim to explanatory power largely rest. For example, there is the question of how a single flow rate can be attributed to a very long axial line — well over a mile long in some cases — but there also some more general issues, which are of more immediate interest. First, to find such correlations is less surprising than is sometimes claimed, since integration is obviously an accessibility measure, and, other things being equal, we would expect the highest flow rates in the most accessible locations. Indeed, it might be suggested that examination of cases where no — or only poor — correlations are found might be more revealing (see Batty & March 1976, for example).

But there is also a deeper problem, which is that of establishing causation (as opposed to correlation) in the static framework common to this and indeed all the analytic tools discussed so far. For example, Hillier et al. (1993, pages 30-31) argue that in considering the relations between spatial configuration, the location of 'attractors' (places of employment, shops etc.) and movement, only configuration is independent of the other two, and can therefore be held to be the causal factor. This argument has merit *only over time-scales where the configuration can be regarded as fixed.* Once all three factors become variable over the longer term, it seems clear that the urban configuration and the distribution of attractors are mutually *and dynamically* influential, with movement patterns largely an outcome of their interaction. Early criticism of space syntax that "the economics of location and land use are ignored" (Gatrell 1985, page 469), points to this difficulty, and recent efforts by others (Krafta 1994, 1996) who are beginning to incorporate land use aspects into a space syntax based framework, have barely scratched the surface of these issues.[7]

It should be emphasised that these problems are equally applicable to the other approaches we have discussed — even where there is an ostensible interest in the relationship between spatial structure and process dynamics, as in Q-analysis. At this point it may be worth noting part of the continuation of the passage from Jacobs quoted above, in section 4.2:

> "This [urban] order is all composed of movement and change, and although it is life, not art, we may fancifully call it the art form of the city and like it to the dance[...]" (Jacobs 1961; 1994, page 60 in the 1994 edition)

...and yet movement and change are conspicuously *absent* from all the approaches we have considered! This is a major failing, which the current research is intended to address, at least in part.

4.4 Urban morphology in geography

4.4.1 Conzenian urban morphogenetics

An intriguing aspect of the space syntax approach not discussed above is the idea that individual buildings can be thought of as particular *phenotypes* or instances of a more general *genotype* for that building type. In other words that buildings which have particular functions (residences, hospitals, prisons) have underlying rules governing their spatial organisation, and that individual buildings are particular realisations of those rules. This kind of analysis is more strongly emphasised in space syntax studies of buildings (Steadman 1983, chapter 12, is an example), and is largely lost when the analysis is transferred to the urban scale. The choice of words is suggestive, however, and perhaps unsurprisingly, appears in a completely separate strand or urban morphology research, this time in geography. According to Whitehand (1981, page 1)

> "The main line of post-war geographical research on the urban landscape — the Conzenian tradition — has its antecedence in the German-speaking countries. There a rich tradition of urban morphological research goes back to the turn of the century and it is there that we must look for the origins of the morpho*genetic* approach — the tracing of the evolution of the forms in terms of their underlying formative processes..." [emphasis added]

[7] This work is also apparently isolated from the accumulated years of experience in land use–transportation modelling which seems a pity.

The connection of such an approach to the space syntax concepts of phenotypes and genotype is clear but the two lines of research seem to have proceeded in virtual isolation from each other. The reasons for this are unclear, although I would suggest that there is a difference in emphasis in the two approaches which is interesting from the current perspective and which may help to explain the separate development of the ideas.

Space syntax analysis is primarily concerned with using the structural organisation of space, as revealed by its analytic tools, to shed light on underlying social and cultural processes which are thought to have led to these forms. This is particularly evident in the analysis of building forms. Some simulation experiments on the growth of artificial settlements (Hillier 1985, Hillier & Hanson 1984, pages 55–66) make this point particularly well, but thereafter the main focus of space syntax studies has been the analysis of urban and built space itself, and its *structure*. A similar focus on the description of the current urban structure is evident in other analytic planning approaches. This concern with structural description is dictated by a desire to intervene and to replicate those structures which appear to 'work' — so as to produce 'marvellous order', or whatever. In the urban morphogenetic tradition the emphasis is on tracing the historical evolution of urban form, through the map and other historical records, and on understanding and uncovering the *processes* which have led to this evolution.[8] Significantly, the concern is to understand the mutual interaction of the from and process, but the tools of analysis are largely interpretative (they might not even be recognised as 'tools' by the analytic planning fraternity!). This emphasis has a direct analogue in the morphology of physical geography — *geomorphology* — a point which is made by Whitehand (1981), in his survey.

The clearest example of the morphogenetic approach is Conzen's (1960) exhaustive study of Alnwick, Northumberland. In this and other work (Conzen 1938, 1949, 1958, Whitehand 1981) a methodological approach (*town plan analysis*) is developed, which involves the detailed analysis of large scale maps (town plans) of settlements over hundreds of years (see figure 18). The theoretical basis of the approach is the continuous adaptation of the human physical environment to meet current requirements.

> "Urban society, urban life, and townscape, therefore, form a unity in space despite, or rather because of, the tensions between society and landscape.

[8] Although see Hanson (1989) for a similar approach using space syntax.

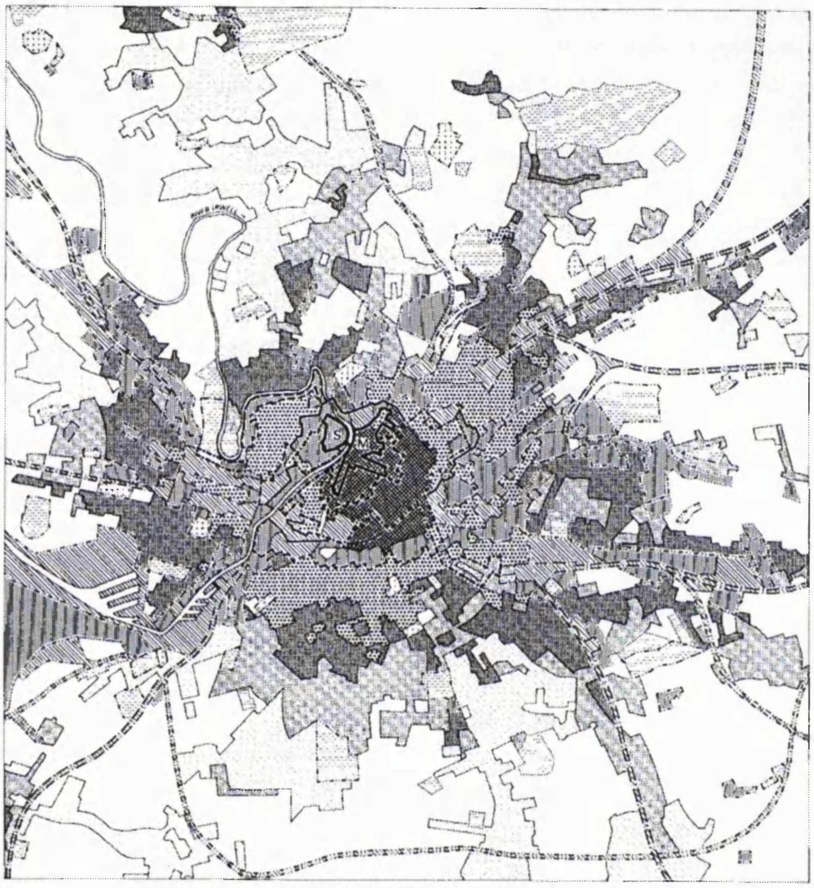


Figure 18 Urban fringe belts from a Conzenian town plan analysis of Manchester. The map shows successive boundaries of the urbanised area, using different weights and styles of lines. These are related to various (later) land uses in the city, which are indicated by different shading and hatching. No simple relationship between the two applies at any single time, but by tracing the historical development Conzen is able to demonstrate the inter-relationships (*source*: Conzen 1981).

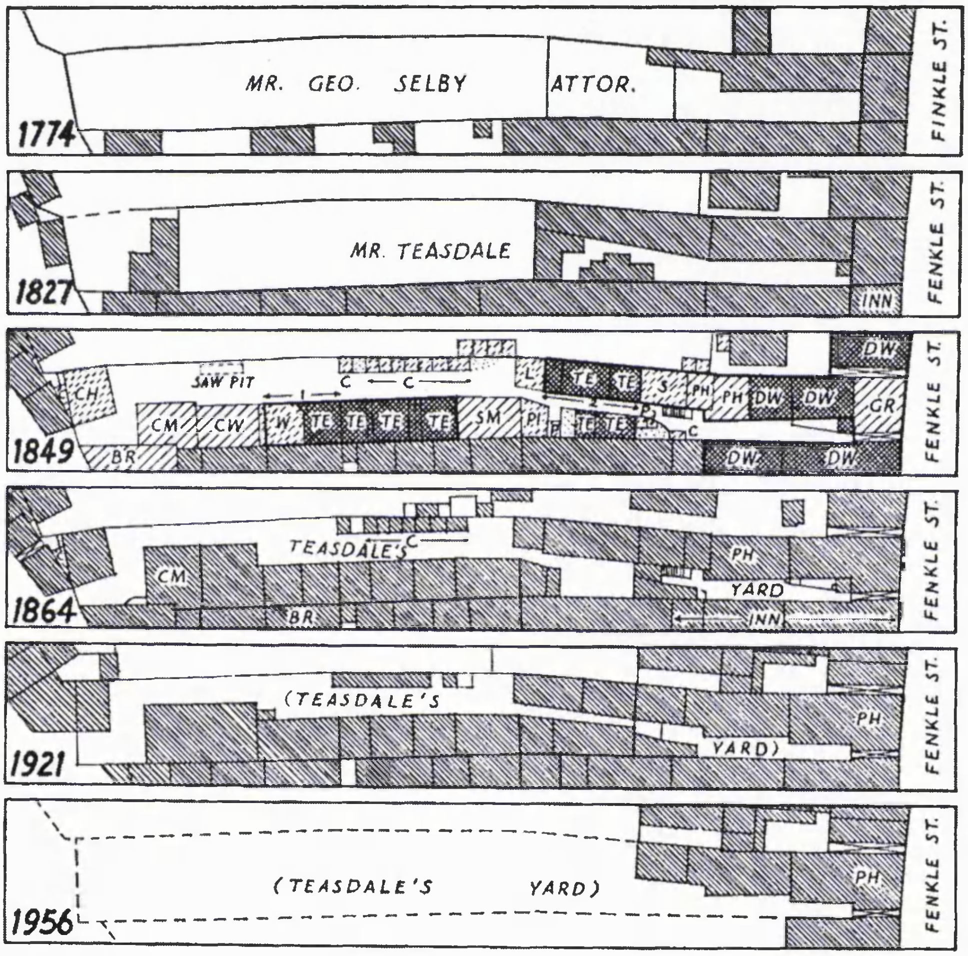


Figure 19 The burgage cycle. This map vividly demonstrates the detailed historical analysis on which Conzen based his approach (*source*: Conzen 1960).

> This tension results from the differential persistence of landscape matter in the face of the ever changing needs of society, thus creating in urban society a continual awareness and consequent re-evaluation of its spatial existence." (Conzen 1981, page 88)

The examination of the evolution of the built environment (townscape) over time will therefore shed light on the processes involved. In the course of his work Conzen hypothesises two common features of British urban development: *fringe belts* and the *burgage cycle*. According to Whitehand (1987*a*, page 254), fringe belts are

> "...the physical manifestations of periods of slow movement or actual standstill in the outward expansion of the built-up area [of a city] and in the initial stages of their development are made up of a variety of extensive uses of land, such as by various kinds of institution, public utilities and country houses, usually having below average accessibility requirements to the main part of the built-up area."

A typically detailed map, showing fringe belts is shown in figure 18.

The burgage cycle is

> "...the progressive filling in with buildings of the backland of burgages and terminates in the clearing of buildings and a period of 'urban fallow' [...] followed by a redevelopment cycle [...] the burgage cycle is [...] often associated with changed functional requirements, in a growing urban area." (Whitehand 1987*a*, pages 254–255)

Figure 19 demonstrates the detailed map analysis on which this concept is also based.

In more recent work, Whitehand (1987*b*), building on earlier efforts (Whitehand 1967, 1972) attempts to link these two concepts to well established ideas about building cycles in urban economics. This establishes a plausible explanation for the existence of both fringe belts and burgage cycles. Whitehand (1992) provides a wide ranging recent review of this whole body of working and its connections with the wider realm of urban geography.

4.5 Recent dynamic models of urban spatial processes

Finally, there is a great deal of work investigating urban form using more recently developed mathematical tools, particularly cellular automata and fractal geometry. Cellular automata models are intrinsically dynamic, and this is responsible for much of the excitement around them — the static nature of spatial interaction models being a major failing in that predominant earlier style of urban modelling (Batty 1976*b*). White & Engelen (1993) and White (1998) provide useful overviews of the more influential work in this area. An important point made in White & Engelen's paper is that urban models based on theories of economic equilibrium could only be matched to 'noisy' reality after complex spatial statistical processing. Thus

> "with complicated boundary conditions [such models] can [...] yield complex patterns. But, in this case, the complexity is essentially imposed on the system as an external condition, rather than being generated by the system itself." (White & Engelen 1993, page 1176)

The point is also made that it is not reasonable to characterise cities as existing in economic equilibrium when everyday experience tells us that they are dynamic, rapidly evolving places "undergoing continual growth, change, decline and restructuring — usually simultaneously." (White & Engelen 1993, page 1176). This is a telling criticism which can be applied equally to many of the structural models considered earlier in this review.

Given these inadequacies, CA models are an attractive approach, since they can be constructed without reference to equilibrium, and they are computationally efficient. Their fundamentals are easily grasped, and the point of entry for the modeller — which states can exist, and the definition of transition rules — is 'close' to theory. Of course, the CA approach also has the advantage of being easy to implement. Couclelis (1985, page 586) notes this happy accident in her consideration of the merits of CA-based models of urban systems. Couclelis (1985) and White & Engelen (1993) both note that the cellular space model in geography goes back some way, citing Tobler (1979), and, even earlier, Hägerstrand (1968), whose spatial diffusion model has been interpreted as a type of CA model. It is not surprising then, that a number of examples of CA-based urban models are to be found.

Xie (1994) developed a very general CA-based urban growth and land use modelling environment called *DUEM* (Dynamic Urban Evolution Model), which has since been extended (Xie 1996). *DUEM* produces fairly convincing results, including emergent effects such as the cyclical development of land use patterns. Batty, Couclelis & Eichen (1997) and Batty & Xie (1997) propose a number of variations on the CA theme for the construction of urban models. These models seem to be intended more as frameworks for the urban CA ideal, rather than as operational urban models, and there are definitely problems with the treatment of road networks, which grow in much the same way as the urban fabric itself — so that they do not actually 'go' anywhere. This is a symptom of the strictly local interaction invoked in this case.

In a series of papers over a number of years, Roger White and various colleagues have developed increasingly elaborate CA-based models of urban land use evolution (White & Engelen 1993, 1997, White, Engelen & Uljee 1997). These papers are particularly interesting in the way a simple basic framework is gradually developed, and in the systematic tackling of a number of technical issues in the application of CA models to urban processes. For example, more extensive cell neighbourhoods (up to a radius of 6 cells in a grid, 113 cells in all) are introduced early on (White & Engelen 1993, page 271). Later the edge-effect problem (that the effect of transition rules at the edges of a finite lattice is likely to be unpredictable) is handled by introducing a fixed-state 'buffer zone' around the edge of the system (White et al. 1997). Cells in the buffer zone may influence the development of neighbouring cells without themselves changing. Some typical model outputs are shown in figure 20. This paper also presents a modification of the basic CA formalism, which is common in urban CA: a non-homogeneous lattice where cells are not identical. Typically this is intended to cope with varying suitability of different regions for different land uses or for development itself. Thus, for example, steeply sloped cells in a model may not be developed, or parkland may not be converted to any other use.

In Clarke et al.'s (1997) convincing model of the urbanisation of the San Francisco Bay area, cell suitability for urbanisation is represented by proximity to the road network and land slope, and cell transition rules are linked to a simple economic model which may release steeper land for development if demand is sufficiently strong, but otherwise prevents development of those cells. The accuracy of their results is in large measure due to the exogenous updating of the regional transport network based on historical 'snapshot' data. Similar comments apply to a model of Brisbane (Ward, Murray & Phinn 1999). From the point of view of building operational, predictive models of urban growth the introduction of strongly determining or forcing variables in this way can be defended, since one of the key inputs to urban development over which policy-makers have almost complete control is the routing of new transport infrastructure. However, such multiplying modifications to the cellular automaton framework do muddy the waters for claims regarding their power as simple simula-

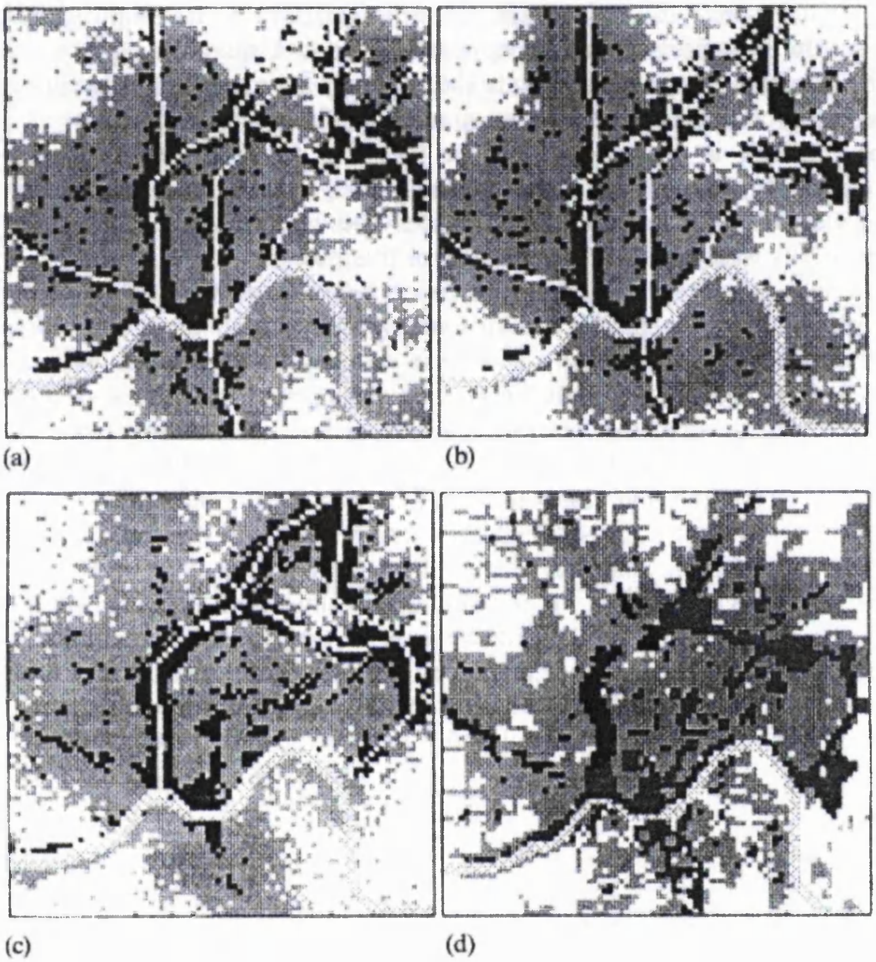


Figure 20 Land use dynamics in Cincinnati modelled using CA: (a) and (b) show two probabilistic variations of a single model run for the same starting conditions, (c) has been produced with a different representation of the transport network, and (d) is based on an actual land use map for Cincinnati in 1960, for comparison. In all cases darker cells are commercial and industrial areas, medium grey is residential. Linear features are parts of the transport network, while the curved pale grey feature in the lower half of the images is the river (*source*: White et al. 1997).

tion frameworks.

A model of urban land use development based on classical economic theory is presented by Webster and various co-workers (Webster & Wu 1999*a*,*b*). This represents a step forward from many previous models where rules can often be rather *ad hoc* and may be constructed with little reference to theory. An important aspect of this work is the explicit recognition that the transition rules in an operational land use model must be understood to represent some theory about various actors' behaviour in the urban land market. Thus rules must represent, in some sense, the way in which developers, communities, individuals, or regulatory authorities are likely to react to local situations as they pursue various goals of profit-seeking and (for communities) welfare-seeking. This conceptualisation of the meaning of cell transition rules is also evident in models presented by Portugali, Benenson & Omer (1997) and Benenson, Omer & Portugali (1999) where the residential preferences of different groups in a multicultural city are expressed at a local level, recalling Schelling's (1971, 1978) very simple models of residential segregation, but relocating them in increasingly realistic settings. These models also make use of theoretical measures of the isolation which may be felt by individuals living amongst neighbours of other groups. These are useful links to pre-existing concepts and theory which are only possible as a result of the relatively concrete nature of CA transition rules conceived as descriptions of actors' behaviour.

These models also make use of 'queues' of incoming and transitional households hoping to move around, out of, or in to the area represented in the model. Households evaluate their options in terms of the neighbourhoods present in the model and then move, leave or enter as appropriate. This introduces a distinctly non-local character to the rules. The same is true of transition potential approaches in White & Engelen's (1997) constrained cellular automata model. In such cases, all cell neighbourhoods are evaluated and assigned a transition potential based on the characteristics of their neighbourhoods. Different potentials are assigned for the possible land uses to which a cell might change in the next time step. Then, using an exogenous source of counts of how many cells *should* be in each state at each time step, each cell is converted to the state for which its potential is highest, until the number of cells required has been reached. At this point no further cells may be converted to the state in question during the time step. The CA therefore acts as an allocation mechanism for exogenously determined growth rates in the various land uses (or whatever) in the model. This introduces both non-local effects (since state transitions are based on the position of a cell's *local* state in the *global* system state distribution) and implicit asynchronous update of cells.[9] Clearly, whether or not models of this type should even be called cellular automata is open to question. On the credit side, the introduction of links to exogenous 'data streams' does open up the potential for embedding cellular models in larger multi-component models. The constraints on a cellular land-use model could clearly be provided by an economic model, which predicts activity in the relevant economic sectors, together with their regional locational preferences. This sort of integration of many different models is now underway (Phipps & Langlois 1997, White 1999, White & Engelen 1997).

Returning to a considerably more abstract model, Wu (1999) also invokes asynchronous update of cell states, whereby the best 'development niche' available at each time step is assigned a unit of investment. When no further sufficiently attractive investment niches are available, randomly generated changes are made to cells until a new niche appears. This in turn may trigger an avalanche of investments in nearby and surrounding cells. Wu links this model to the notion (also from complex dynamics) of 'self-organised criticality' by observing the frequency/size distributions of the avalanches of investment which result.

The link to self-organized criticality has also been made from fractal geometric views of urban morphology by Batty & Xie (1999). A direct connection between discretised cellular or grid-based simulations of urban growth processes and resultant fractal forms was established in earlier work by Batty & Longley (1986). The use of fractal geometry applied to urban morphology is thoroughly discussed by Batty & Longley (1994). The crucial insight here is the direct applicability of measures of fractal dimensionality to urban systems represented in various ways. Given the origin of fractal forms (see figure 21 for an urban example) in processes operating across a range of scales, but in generally local ways, this is a suggestive result, especially in relation to the kinds of models discussed in this section. At the very least, it seems to represent significant circumstantial evidence for the applicability of such models. The overlap between the fractal dimension of urban systems and these cellular models has

[9] Note that asynchronous CA have been explored to a limited extent in the natural sciences (Bersini & Detours 1994, Ingerson & Buvel 1984) and may exhibit significant differences in behaviour to otherwise similar, but synchronous examples.

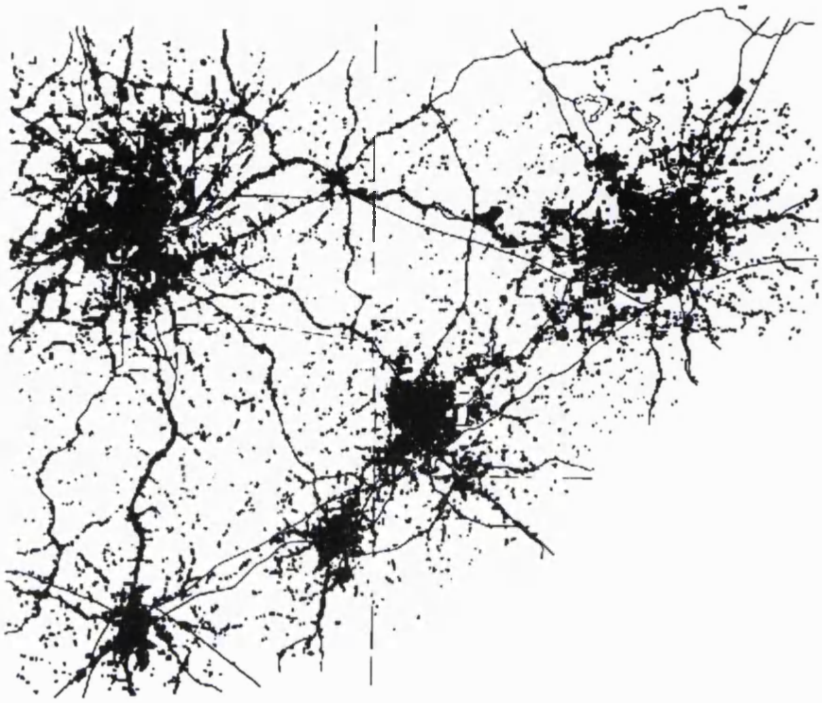


Figure 21 Urban fractals (*source*: Batty & Longley 1994, after Chapin & Weiss 1962).

been further recognised in the use of measurements of the fractal dimension of urban systems and complex models of the same, to validate the models (White & Engelen 1997). This has been criticised as focusing too much on the geometry of the emergent forms and not enough on an assessment of the validity of the transition rules (Macmillan 1999), and it is clear that the difficult issue of assessing the 'fit' of such complex models is vital to any chance of their widespread acceptance as operational models around which policy issues can be discussed.

4.5.1 CA models and complexity theory in urban dynamics

In spite of the tremendous expansion in the application of cellular approaches to the modelling of urban development processes at a micro- and meso-scale described above, there has been only relatively limited debate as to the applicability of such models. They belong to a wider reorientation of the urban modelling tradition towards dynamic models of urban processes, and away from the potentially dubious transfer of concepts borrowed from physics either directly (Newtonian gravitational attraction) or indirectly via economics (market equilibrium, perfect competition). This reorientation seem to have derived both from an acknowledgement of the inadequacies of these borrowed concepts (White 1977) and, ironically, from a further borrowing, in this case one which was initiated by physicists themselves (Allen & Sanglier 1979). This latter strand explicitly suggests that the physics of self-organising dissipative structures (Prigogine & Stengers 1984), first observed in non-equilibrium thermodynamics, is appropriate for the modelling of cities and urban systems. This argument has a strong intuitive appeal given the unpredictable and rich dynamics of systems of this kind. Notwithstanding Couclelis's (1984) contention that "intuitive appeal alone is not sufficient for establishing any model as an approximately correct description of the mechanism driving urban process" (page 472), such models have certainly been widely adopted, albeit not yet in operational form.[10]

What is clear is that the devastating critique of an earlier generation of urban models (Sayer 1976), as inappropriately focused on market equilibria, and unable to cope with dynamics, is not directly applicable to complex models, which can and do exhibit novelty and instability, and furthermore need not assume market equilibrium, nor be based on neo-classical economics. Indeed some work in this new tradition is very reminiscent of precisely the modelling approach recommended by Sayer — compare Sayer (1976, pages 218–226) and Portugali et al. (1997), for example. It is certainly a little disappointing that the political economy approach endorsed by Sayer is now largely isolated intellectually from ongoing research in modelling and analytical techniques, with the virtually unique exception of Sheppard & Barnes's (1990) *The Capitalist Space Economy*.

I return to these broader issues in chapter 12. For now, it is sufficient to note the importance of cellular automata, and by extension complexity theory, to current analytical approaches to the city. For many, the interest and justification is instructive or metaphorical. The way in which global structures are seen to emerge from purely local rules of interaction is certainly interesting, and sheds light on the otherwise random-seeming structure of many urban forms. At the scale of urban systems, a great deal of what actually occurs is contingent and dependent on individual hu-

[10] Although recent developments suggest that this is about to change as US cities gear up to implement the monumental TRANSIMS model (Nagel, Beckman & Barrett 1998).

man agency, which is only rarely modelled. Instead, such factors are usually included by introducing a random element into the rules. This strengthens the argument that these models can only be used as a way of exploring ideas rather than for proving theories or predicting particular outcomes. That is no small achievement, however, and the realisation that complexity can emerge in simple systems has been influential (Couclelis 1988, Stern 1992).

Whether or not such models can be used to *prove* anything is questionable. Many different models can presumably be constructed to produce convincing urban forms, so that a correspondence between a CA model outcome and reality does not necessarily imply that the rules employed in the model reflect the actual processes occurring in the 'real world'. However, in science *nothing* can be proven incontrovertibly — hypotheses may be *disproved* by experiment, but only *verified* for the time being. There is therefore no obvious reason why CA models can not be used to test theories of urban form and process. If the resulting model forms are not similar to real world forms in some appropriate sense (fractal dimension is frequently used, but other measures are required) then the theory so modelled may be brought into question.

4.6 Conclusions

This chapter has been only a partial review of the literature on what I have termed micro-scale urban analysis. An aspect of the work cited which is interesting is that in general urban spatial structure and urban spatial processes are separately studied. At a very general level, this may be a result of the missing 'spatiality' in critical theory, as Soja (1989) argues.

In much of the detailed and more structurally focused work (the Martin Centre models, Krüger's work, Q-analysis and space syntax) the emphasis is on understanding a *static urban structure* however defined. This static structural focus may be associated with an interest in short-term interventions on that structure. That is, technical planning and architectural perspectives are concerned with understanding the urban structure as it currently exists prior to intervening, whether it be to build, or re-zone, or otherwise alter that structure. This sort of work can only incorporate processes and change through 'comparative statics'. Snapshots of urban structure at various times can be compared in an attempt to understand underlying processes. If such snapshots are structurally analysed (using graphs or any other available tools) and related to other phenomena (economic activity, pedestrian movement) then it may be possible to deduce something about underlying processes. However, any conclusions cannot be tested easily in this way. Thus, for example in space syntax research, there is a theory of 'the movement economy' (Hillier et al. 1993) which loosely relates movement as correlated with integration values to changes in land use (retailing in particular), but given the comparative statics approach, the notion is difficult to test. The problem is that any single snapshot is the outcome of a historical process, the evidence for which is differentially present in any single historical moment.

In the geographic urban morphology literature (I include here morphogenetics and more recent CA models), as in the wider geographic theory literature, there is a more direct concern with understanding the *processes* which may lead to various urban spatial forms. There is a general understanding that the two are inter-related: urban structure influences the requirement for new forms and affects the social processes which occur; and the social processes occurring in a city in turn affect the forms which emerge. However, morphogenetics is hampered by the methods employed: these are not a great deal different from the various structural approaches, and they are difficult to generalise, because the descriptive tools needed to recognise similar outcomes in spatially distinct contexts are lacking.[11] Even so, detailed historical surveys do reveal something about the way in which structures have evolved.

Recent cellular automata and related non-linear dynamical models of city growth and evolution do incorporate processes more explicitly, although many of their developers are happier with using them as educational tools rather than as highly accurate models. In spite of reservations, these sorts of models seem to represent the best chance for deeper understanding to emerge — even if, as increasingly realistic *ad hoc* modifications to the underlying CA formalism are added, it becomes harder to draw any very *general* conclusions from the models. As Couclelis (1985, page 588) remarks, "all the simplifying assumptions of the basic cells space could be relaxed in principle: in practice, of course, the result would be forbiddingly confusing." — in fact, almost as confusing as the urban reality itself!

Whatever the merits of the various approaches discussed here, none of them attempts explicitly to relate urban spatial structure to urban social or economic processes, or to investigate the likely impact of changes in spatial structure on future

[11] The lack of exchange between the various strands of urban morphology research is frustrating in this respect.

dynamic outcomes in the urban system. This is generally held to be a relationship central to understanding cities (recall the quote from Harvey in the introduction to chapter 2), and thus seems to be a weakness of all of the work discussed. In the next part of this thesis a modelling approach is proposed and developed which may help to address this issue.

Part II

Model Development

Chapter 5

A spatial model combining graphs and cellular automata, and its implications

We are now in a position to bring together the ideas in the previous three chapters in a proposal for a spatial model suitable for the investigation of geographical systems — particularly urban systems. The proposed model is a proximal model in the sense discussed in chapter 2; it incorporates the methodological tools introduced in chapter 3 — graphs and cellular automata; and it may allow the user to come to some understanding of the relationships between urban structure and urban processes — an important, but often neglected aspect of urban morphology which was discussed in chapter 4. Although this model is presented and developed in the context of urban spatial problems, it is hoped that similar models may be useful more generally.

In the next section the proposed model — the *graph-CA* — is introduced in overview and formally specified. The succeeding section then reviews some possible lines of research arising from the model concept. The bulk of the chapter in section 5.3 then describes a methodology for pursuing one of these research directions. This research has been followed up, using tools developed for the purpose, based on the concepts described in this section. The tools themselves are described in chapter 6, and the research results reported in chapter 7. This chapter closes with a discussion of the concepts which have been presented.

5.1 The basic concept

At this stage the idea should already be obvious. It is simply that a cellular automaton system be constructed, based on a cell space which is irregular, and can therefore be represented as a graph. Since the operation of the CA is determined by rules based on cell neighbourhoods, the minimal requirement for a CA-like model is a set of cells with defined neighbourhoods. This relaxation of the regular cell-space of 'conventional' CA has already been proposed as an appropriate approach in regional and urban modelling generally (Couclelis 1997, Takeyama & Couclelis 1997). The extension of the idea that is proposed here is that the cell space be represented as a graph, which can therefore be investigated alongside the CA which is run on it. Thus, we may think of a CA in an irregular cell space as a *graph-based CA*.[1]

The advantage of this approach is that the model may be examined from a number of perspectives, using a range of already well-developed tools. At the simplest level, the graph representation of the cell-space allows the use of the measures described in section 3.2 (pages 45ff.), to examine the structure of the model; the GCA which runs on the graph, as it evolves will demonstrate the dynamic behaviour of the model. In addition, some of the tools of CA behavioural analysis described in section 3.3.3 (pages 64ff.) may also be brought to bear in understanding the model behaviour. Thus, using this type of model we may be able to begin to understand the relationships between structure and process in spatial systems. This broad perspective is shown in schematic form in figure 22.

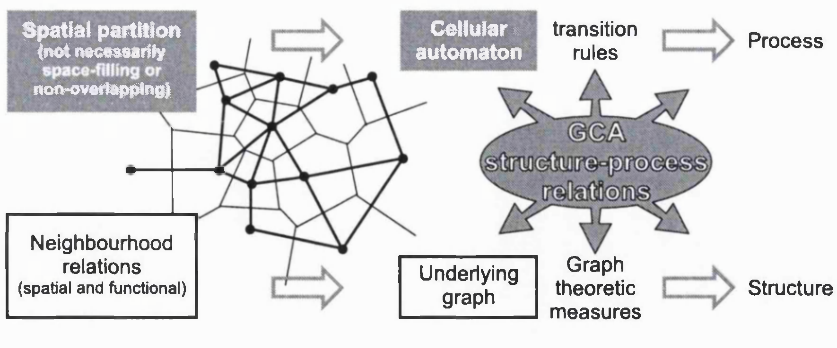


Figure 22 The proposed model structure, showing the links between structure and process which it provides.

[1]The terms *irregular CA*, *graph-CA*, and *GCA* are used interchangeably in the remainder of this thesis.

As can be seen in figure 22, a GCA model is based on some sort of spatial partition. The partition need not be space-filling and overlapping spatial elements might be used. The requirement is that a set of spatial elements be defined somehow. In the context of an urban model, the entities forming the basis of urban morphology (buildings, blocks, streets) are a natural set of spatial elements to use as the graph vertices. Thus any of Krüger's graphs might be used (see section 4.3.1, pages 74ff. for details). Alternatively, the graph might be constructed based on plots of land, census districts, administrative or planning authority zones, the axial lines of space syntax, or indeed any set of elements of interest. Edges in the graph represent some sort of relation between vertices, so that any relationships relevant to the model being constructed might be used. Where accessibility is under investigation the relations of interest will be associated with adjacency or distance relations between elements. Indeed, any model which is intended to investigate the spatial aspects of some problem domain is likely to make use of some sort of distance concept in the modelling of relations between entities. However, the graph representation is flexible enough to also allow relations of functional influence to be included.

The GCA is also based on a cellular automaton. In this context a CA requires that a set of cell states and transition rules be defined. Any more conventional CA could provide the basis for these. Some of the additional complications and difficulties of the transfer to an irregular cell space are discussed in section 5.3 after we formalise the model, and consider possible lines of research which it raises.

5.1.1 Formalism

We assume that a spatial system can be represented meaningfully as a graph G consisting of elements represented by the vertex set $V(G)$ and relations between those elements represented by the edge set $E(G)$. Each vertex v_i in $V(G)$ has a neighbourhood $N(v_i)$ which is a subset of $V(G)$ such that

$$N(v_i) = \{v_j : v_j v_i \in E(G)\} \tag{5.1}$$

Thus the neighbourhood of v_i consists of all the other vertices in $V(G)$ to which it is adjacent. Note that $v_i v_i$ may be a member of $E(G)$, so that $v_i \in N(v_i)$. In fact, to be completely general, it is preferable that we consider the graph to be a digraph, with directed edges or arcs so that $v_i v_j \not\equiv v_j v_i$. In this case we can consider the neighbourhood of v_i as divisible into two: an *in-neighbourhood* $N_{\text{in}}(v_i)$ and an *out-neighbourhood* $N_{\text{out}}(v_i)$, where

$$N_{\text{in}}(v_i) = \{v_j : v_j v_i \in E(G)\} \tag{5.2}$$

$$N_{\text{out}}(v_i) = \{v_j : v_i v_j \in E(G)\} \tag{5.3}$$

where $N_{\text{in}}(v_i)$ and $N_{\text{out}}(v_i)$ may overlap. In fact, in all the examples considered in the remainder of this thesis, the existence of an arc $v_i v_j$ implies the existence of the arc $v_j v_i$, so that although the graph is formally a digraph, it may be treated as a simple undirected graph with loops. Since this implies that $N(v_i) = N_{\text{in}}(v_i) = N_{\text{out}}(v_i)$ for all v_i, I use the simpler undirected graph notation in the remainder of the thesis, with no prejudice to the generality of the discussion (see O'Sullivan forthcoming, for a more detailed consideration of this point).

If each vertex v_i in G has associated with it a state $a_i^{(t)} \in A$ at time t then the cellular automata aspect of the model is incorporated such that

$$a_i^{(t+1)} = f\left(\mathbf{b}^{(t)}(N_i)\right) \tag{5.4}$$

where $\mathbf{b}^{(t)}(N_i)$ denotes a vector[2] representing the state of the members of $N(v_i)$ at time t, and A is the set of discrete allowed vertex (or cell) states. $\mathbf{b}^{(t)}(N_i)$ is itself some function of the states of the constituent cells of $N(v_i)$, that is

$$\mathbf{b}^{(t)}(N_i) = g\left(\left\{a_j^{(t)} : v_j \in N(v_i)\right\}\right) \tag{5.5}$$

In a strict digraph representation, we regard the in-neighbourhood as the set of influencing cells, so that $\mathbf{b}^{(t)}$ is based on the states of vertices in $N_{\text{in}}(v_i)$.

In a conventional CA the state a_i of a vertex in G is one of a finite set of discrete states. Thus in Wolfram's simple one-dimensional CA the set of allowed states $A = \{0, 1\}$. If $V(G) = \{v_0, v_1 \ldots v_n\}$ represents the linear array of cells numbered in sequence, then $E(G) = \{v_i v_j : |i - j| \leq 1\}$. In a CA intended to model say, the evolution of urban land uses the set of states will be more complex, say

$$A = \{\text{RESIDENTIAL}, \text{COMMERCIAL}, \text{INDUSTRIAL}, \text{DISUSED}\}.$$

$V(G)$ might consist of land plots, and $E(G)$ could represent the adjacency relations between them. In more complex models, the state a_i might be based on the values of

[2]The reason for adopting a vector representation of the cell neighbourhood state will become clearer below.

a number of variables, and $b(N_i)$ would be correspondingly more complex also. This does not fundamentally alter the mathematical description of the model presented above.

5.2 Two possible lines of research

At least two distinct research directions arising from this model can be identified. These are discussed in overview in this section, before a more thorough exploration of one of them is carried out in chapter 7 using tools developed for the purpose and described in chapter 6.

5.2.1 Exploring the structure-process relationship

The major theoretical motivation for introducing the GCA model is that it enables the exploration of relationships between the structure of spaces in which processes unfold and the sorts of processes which occur. This exploration is possible at two distinct levels — the global and the local. These are considered in turn below.

Global relations

As discussed in section 3.2 (pages 45ff.), it is possible to derive summary measures of the overall structure of a graph. One of the major example considered was graph centralisation, which allows the extent to which a graph is centralised to be quantified. A star-like graph is highly centralised, whereas a linear graph is not. The relationship between such measures of a graph's overall structure and the dynamic behaviour of a GCA running on it may be of interest. In the simplest case we might hope to find a relationship between some centralisation measure and the Wolfram class (Wolfram 1983) of the GCA, in much the way that Langton (1990) found a relationship between a parameterisation of CA rule space and dynamic behaviour, as described in section 3.3.2 (pages 58ff.). Research into the dynamic behaviour of random Boolean networks (Kauffman 1984) suggests that connections between the relational configuration of automata cells and their dynamic behaviour can be sensibly investigated. More recently, investigations of the behaviour of diffusion and other communications processes on small world networks (Watts 1999, Watts & Strogatz 1998) point to a great deal of riches in this general area. These other research efforts are described in more detail later in this chapter, and this broad theme occupies much of this part of the current work.

Local relations

Another aspect of the structure-process investigation of a GCA model is to examine whether there is a relationship between *locational measures of individual elements* in the underlying graph and their *individual dynamic behaviour*. Thus, we might seek to determine relationships between (say) the centrality of vertices in the graph and the time-sequence of states prevailing at those vertices. The various graph structure measures discussed in section 3.2 could all be investigated from this perspective. Thus we might find that members of different cohesive sub-groups in the graph behave differently, that vertices which belong to different automorphic equivalence classes behave differently, or that there is a relationship between vertex centrality and behaviour. Any of these localised relationships might also provide significant insight into the way in which the model behaves globally, and therefore into the way in which the real world behaves. These sorts of relationships are very similar to the usual concerns of spatial modellers to find relationships between locational variables and behaviour. Again, the current approach suggests new ways of thinking about these questions.

5.2.2 Extending the GCA model

Another possible research direction suggested by the GCA model framework is the exploration of various extensions of the model. Among others, the following are possibilities:

1. *Using structural measures of the underlying graph as input variables in the CA rule set*. Thus, we might want to use a centrality measure as a partial determinant of the behaviour of cells. Of course, such a formulation of the CA rules is a significant departure from any conventional CA model, where rules are held to apply globally. However, in the context of building working models of processes in the real world such an extension might be highly desirable. An obvious case is the centrality measure which in many situations can be thought of as a proxy for *global* accessibility, which it is generally agreed is a significant determinant of many urban processes, whether it be via the economics of the property mar-

ket, or the operation of the transport system. This sort of extension of a local rules-based model is difficult in a conventional regular CA where the location of all cells is effectively identical. Even if cell neighbourhoods are thought of as representative of a kind of accessibility, such accessibility is a purely local measure where $|N(v_i)| \ll |V(G)|$, which is usually the case.

2. *The introduction of multi-component GCA models.* This is a very generic notion, and in fact does not involve any extension of the model formalism, merely a change in the ways in which such models are conceived and constructed. It is also related to the constrained cellular automaton models, already discussed which have been developed to allow their embedding in multi-component models of regional development (Phipps & Langlois 1997, White 1999, White & Engelen 1997, White et al. 1997). The GCA approach may represent a particularly neat way of achieving this goal. The crux of the idea is that the graph vertices need not all represent elements of the same type and, by extension, graph edges can represent various different kinds of relations between elements of different types. For example, we could imagine elements representing buildings and planning authorities. The possible changes in building land uses would then be conditioned by which planning authorities they were related too. Not all elements in such a model would necessarily represent spatial elements.

   This sort of extension of the model also naturally lends itself to the creation of hierarchical geographical models so that elements at the bottom level belong by turns to (say) cities, regions, nations, super-regions and so on, with functional relations extending up and down the hierarchy. I have formalised this idea elsewhere (O'Sullivan forthcoming). Combining this extension with the other items in this list could allow the larger aggregates to evolve in diverse and unexpected ways.

3. *Allowing modification of graph structure as a result of local states of the system.* This would allow the generation of new graph edges and new vertices, or conversely their removal in response to local conditions in the model. The system would therefore modify its structure as a result of the processes occurring in it. Semboloni (2000, forthcoming) has carried out some initial explorations along these lines.

A thorough exploration of any of these possibilities would take us a long way from the current focus, and in any case, an understanding of the likely complex behaviours of such models would be dependent on understanding structure-process relations, which is the more immediate aim here.

These thoughts further illustrate the potential of the greater freedom provided by the irregular CA/graph-CA framework over the rigidity of conventional CA. One way of thinking about this is illustrated in figure 23. Models based on discrete elements (cells) may have spatially stationary or non-stationary structure in cell space. That is, they may be spatially homogeneous with similar neighbourhoods at all locations — or not. Strict CA have a spatially stationary cell space, whereas GCA do not. However, both CA and GCA are stationary in rule space — the transition rules apply in the same way at all locations in the system. This figure should be borne in mind in the discussion of section 5.3, when examples of other models in this 'space' are considered.

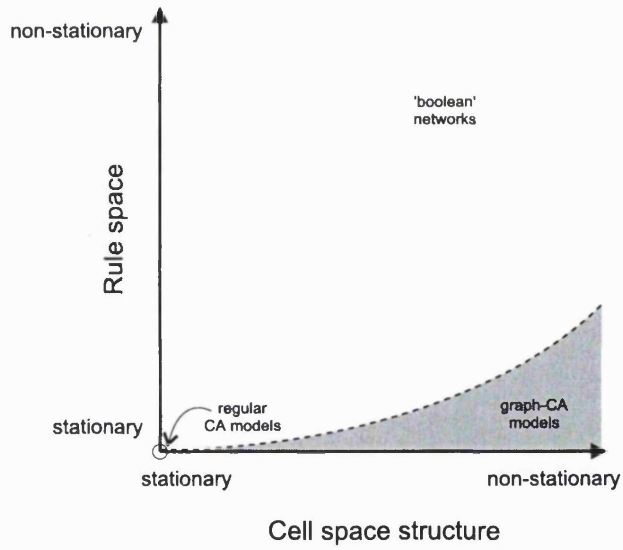


Figure 23 The conceptual space in which graph-CA models exist, relative to other discrete models.

5.3 Exploring global structure-process relations in graph-CA models

Given that the current work also explores the application of a GCA model to urban change, a thorough exploration of each of the above avenues of research is well beyond its scope. However, a limited investigation of the relations between structure and process in GCA models has been carried out. The next chapter describes the computer software developed for this investigation, and details of the results are presented in chapter 7. Here, we concentrate on the methodology of this investigation and the reasoning behind it.

5.3.1 The domain of all possible graph-CAs

A particular graph-CA model may be thought of as existing in a multi-dimensional domain which contains all possible such models. In a sense this domain is a subspace of the space of all discrete models — in effect, the shaded region in figure 23. This section now proposes a framework for thinking about this domain which leads naturally to a way of exploring it.

In setting up a graph-CA model we are free to vary it in two fundamental ways: we can vary the underlying graph; or we can vary the rules governing cell evolution. The outcome of these modelling decisions will presumably be different model behaviours. If we can describe each of these variables (graph structure or *cell space,* the rule set or *rule space,* and the *complexity* of the dynamic behaviour) in terms of a single variable then the domain of all possible graph-CAs can be thought of as a three-dimensional space (see figure 24). Of course, this is a simplified picture, since it should be clear that none of these three dimensions are likely to be fully describable in terms of a single one-dimensional variable. Note also that the cell space and rule space axes in figure 24 are not necessarily the same axes as those in figure 23, since the meaning of the axes is dependent on the measures adopted. In fact, the importance of the notion of *GCA-space* — represented in figure 24 — is as a conceptual device, which is useful in developing a methodological approach.

As we have already seen in section 3.3.2 (pages 58ff.), according to Wolfram (1983), Langton (1990) and others, there seems to be a correspondence between CA behaviour and CA rule space, and the first steps towards understanding the subtle

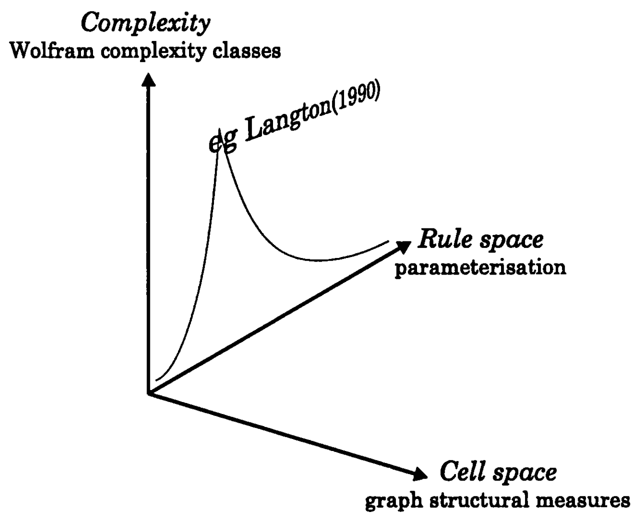


Figure 24 The domain of all possible GCA models — GCA state space.

effects of rules on global behaviour have been taken. The obvious question now becomes: *can relationships be found between the structure of cell space in a graph-CA and the emergent dynamic behaviour?* As the label on the cell-space axis of figure 24 suggests, the approach we adopt is to characterise cell space using graph structural measures, and to carry out similar work to Langton's, this time exploring variation in system behaviour as it moves through cell space rather than rule space. We can think of this line of research as an exploration of GCA cell space.

Before describing the methodology for this research in more detail, some recent work in the same general area in other disciplines is worthy of comment. None of this work is exactly like the current research, but taken together they offer some hope that the approach is worthwhile, *and* demonstrate that there is a 'niche' in this area for a specifically geographical research program.

5.3.2 Structure-process research using graph structures and discrete dynamics

Two previous research efforts in this area have been identified (note that these can be fruitfully understood in relation to figures 23 and 24 which should be borne in mind).

Kauffman's auto-catalytic networks

Stuart Kauffman's (1984) research investigates the dynamic properties of *random Boolean nets*. A (marginally) more accessible introduction is provided in Kauffman (1995). Kauffman's aim is to model the relationships of catalysis and inhibition which may exist among a set of organic chemical reactions. The model he proposes is a randomly generated graph with N vertices, in which each vertex has K randomly selected neighbours.[3] Such a graph he calls an N-K *network*. The second step in creating the model is to randomly assign a Boolean function to each vertex. The function assigned is like a CA transition rule except that each vertex may have a *different* transition rule. Thus, Kauffman's models belong in the upper right of figure 23 since they are spatially non-stationary in both rule space and cell space.

Typically, Kauffman runs 1000, 10 000 or 100 000 vertex N-K networks, and examines their dynamic behaviour. The striking finding is that for $K = 2$ such systems generally have very short transition times before they enter a relatively short cycle where a small number of the systems possible states are repeatedly generated. A 100 000 vertex system where each vertex may be 'on' or 'off', with $2^{100\,000} \approx 10^{30\,103}$ possible system states, actually settles rapidly into cycles with an average of only 317 ($\approx \sqrt{100\,000} = \sqrt{N}$) system states. Kauffman suggests that this means that *autocatalysis* of systems of reacting organic chemicals is much more likely than simple statistical considerations would lead us to expect. An autocatalytic chemical system is one which is self-sustaining given a ready supply of a small number of 'fuel' chemicals, and an environment into which the 'waste' products can be deposited. The finding that autocatalysis is less unlikely than simple statistical considerations would lead us to expect is a possible basis for an understanding of the emergence of life in the 'primeval soup'. Kauffman tries to explain this unexpected behaviour of his N-K networks with reference to 'forcing structures' — sets of connected vertices which tend to be stable and force the rest of the system into relatively few possible outcomes. Lynch (1994) has shown some of the weaknesses in Kauffman's argument without doing the general idea any serious damage, and the N-K network is now a well-established simple model for studies of some kinds of complex system.

Kauffman's work is important, but it is significantly different from the current

[3] I adopt Kauffman's own notation in this section for consistency with his work should the reader wish to refer to it. I will continue to use n and k for graph and neighbourhood size respectively, except when referring to Kauffman's work.

work because we are interested in very different system architectures, and moreover, in the current work the intention is to apply the same transition rules throughout the system as in conventional CA. Nevertheless, the example is interesting and establishes the possibility of making general but useful statements about the relationship between a system's structure as represented by a graph and its dynamic behaviour. It is also important to note that when K is only a little greater than 2, most systems observed by Kauffman exhibit chaotic behaviour. This puts the behaviour of systems like the Life CA ($K = 9$), in an interesting light. It suggests that the regular spatial structure of the Moore neighbourhood grid-based lattice is important to the behaviour of the Life CA, since the same rules on an irregular space would more than likely produce similar chaotic behaviour.

Watts and Strogatz's small world network dynamics

We have already encountered the small world phenomenon in section 3.2.4 (pages 54ff.). Watts & Strogatz (1998) show that small world networks can be readily constructed by random deformation of one-dimensional lattices. The deformation process used involves randomly breaking edges in the lattice with probability p and remaking them to randomly selected vertices elsewhere in the lattice. The small world measures L and γ can then be expressed as functions of p, $L(p)$ and $\gamma(p)$. Starting from a regular linear lattice with $L(0)$ and $\gamma(0)$, figure 25 shows the way in which only low values of p (around 0.01) are required to turn regular lattices into small worlds with low characteristic path lengths, where $L(p) \approx L(1)$, but high clustering, where $\gamma(p) \approx \gamma(0)$, so that most vertices are 'unaware' of the paths connecting them quickly to remote parts of the system.

Significantly, Watts & Strogatz go on to show that dynamic processes such as diffusion or disease transmission travel at significantly different rates on small world systems relative to regular lattices. This is a direct link between structure and process in systems. In more recent work along the same lines, examining the behaviour of cellular automata running on small world networks, with rules designed to solve the density classification problem (encountered in section 3.3.2, pages 58ff.), Watts (1999, page 187) comments that

> "[...] it is natural to ask whether or not high-performance CAs can be developed *by varying the coupling topology of the automata instead of their rules.*" [emphasis in the original]

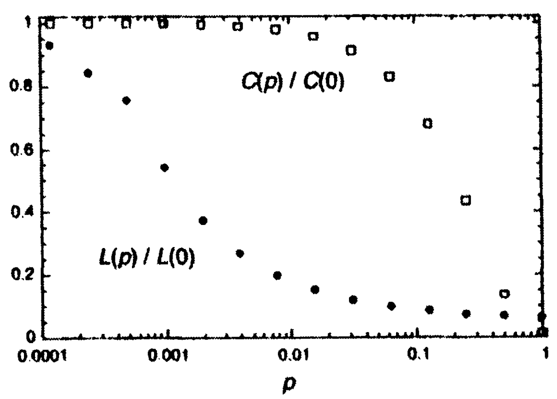


Figure 25 Watts & Strogatz's (1998) deformation of regular lattice by a random process — the effect on characteristic path length and clustering coefficient. Note that the symbol $C(p)$ has been used here for clustering coefficient, rather than $\gamma(p)$. The region with high clustering coefficient and low characteristic path length is where small world graphs are produced (*source*: Watts & Strogatz 1998).

The relationship between that question and the current one (*can relationships be found between the structure of cell space in a graph-CA and the emergent dynamic behaviour?*), is clear. Obviously, Watts's interest is expressed in terms more like an engineering problem, whereas the current work is more concerned with building an understanding of already existing geographical/urban systems. A more significant difference lies in the nature of the deformations to the graph which are allowed. In the current work, as we shall see, an attempt is made to limit deformations so as to preserve spatial proximity in the system. That is, near vertices tend to remain near vertices even after many deformations to the system. The resulting irregular lattices are not small worlds. As we shall (see figure 50 on page 162) this is a significant difference, with presumably different implications for the dynamic processes evolving on these systems

The important points about these two examples are (i) that relations between structure and process can be identified, and (ii) the models they examine are non-spatial — in the sense that the rules for their construction do not invoke spatial relations — so that there is room for specifically geographic research in this area.

5.3.3 Difficulties of exploring the cell-space of GCA models

Given that the current aim is an exploration of the cell space of the domain of all possible graph-CA models, we can identify a number of problems.

First, this domain is *unthinkably vast*. If we consider only graphs of 20 vertices, then there are around 10^{39} possible connected graphs of this size (Beineke & Wilson 1997, page 27). Considering only 2 state CA rules, each of these possible graphs has $2^{2^{20}} \approx 10^{315\,653}$ possible rule sets. Even considering only totalistic rules, where only the number of neighbours in a given state affects the next state, there are $2^{21} \approx 2 \times 10^6$ possible rule sets. $10^{39} \times 2 \times 10^6 = 2 \times 10^{45}$ is a very large number indeed.[4] *And* these calculations refer only to 20 cell 2-state models! We clearly need some way of exploring the space in an intelligent, directed way. The basic method adopted here is to start with well known 'interesting' regular CAs, and to progressively deform their cell space. Thus, we might start with the 'Game of Life' CA, and steadily alter the structure of its grid (its cell space), observing changes in the resulting behaviour. We can think of the deformation process as resulting in a *trajectory* in GCA-space.

Even limiting the approach in this way, difficulties remain. Apart from the fact that there is still a vast range of possible variation, perhaps the foremost difficulty is whether or not it is possible to alter the distribution of graph neighbourhood sizes and simultaneously hold the rule set 'constant' in a meaningful way. This effectively poses the question: is there a relationship between cell space and rule space? Put another way, can we choose parameters which make the cell space and rule space axes of figure 24 orthogonal? Consider a uniform grid of cells, where cell neighbourhoods include the cell itself and its immediately adjacent cells in all eight directions, so that all neighbourhoods consist of 9 cells. Assuming 2 allowable states, a particular rule set for this cell space need only specify $2^9 = 512$ outcomes. If deforming the cell space results in neighbourhoods of sizes other than 9, then how do we apply the original rule set to these neighbourhoods? There seems to be no easy answer. Possibilities include:

1. *Deform the cell space in ways which preserve the original distribution of neighbourhood sizes.*[5] This restricts the exploration to subsets of the cell space. It may also lead easily to unsatisfactory models. For example, the simplest way to

[4] There have been 'only' around 10^{18} seconds since the big bang!

[5] Preserving the graph's *degree list* would be slightly different. Here, the meaning is that each vertex's degree (or neighbourhood size) remain unchanged under the deformations.

deform a graph and retain the same distribution of neighbourhood sizes is to treat edges strictly as arcs, and randomly remove incoming arcs from vertices, replacing them with new incoming arcs from other vertices. Some vertex *out*-neighbourhoods will grow under such a process, but only *in*-neighbourhoods have a bearing on GCA behaviour in the current context. However, this deformation may quickly produce disconnected graphs, which does not seem entirely satisfactory. In any case, restricting changes to those which preserve neighbourhood sizes still does not make GCA-space *small*. For example, we can 'guesstimate' the number of k-regular graphs of size n as n times the number of ways of selecting $(k-1)$ neighbours from the remaining $(n-1)$ vertices. This gives us

$$n\binom{n-1}{k-1} = \frac{n!}{(n-k+1)!} \tag{5.6}$$

For $n = 400, k = 9$ this gives us something like 10^{16}.[6] Even restricting neighbour selection, so that connected graphs result, these are large numbers. In practice then, restricting ourselves to neighbourhood size preserving deformations may not be such a great restriction.

2. *Define rules for larger and smaller neighbourhoods from the outset.* This could work, but raises a different question. Can a 'reasonable' rule set contain rules which vary dramatically between similar neighbourhood sizes? As an extreme case, could a rule set specify that a cell with two neighbours one of which is in state 1 adopts state 0 in the next time step, whereas, with three neighbours one neighbour in state 1 leads it to adopt state 1 in the next time step? In effect, this consideration also limits the exploration, this time to rule sets which are well-behaved, according to some criterion, which must be specified.

3. *Specify rules in a way which is independent of neighbourhood size.* For example, we could express rules in terms of the percentage of neighbouring cells in various states. This also restricts the exploration to a subset of the rule space. This is also not a complete solution to the problem either. Say, for example, that we have a rule that any cell with 40% or more of its neighbours in state 1 changes to state 1 in the next time step. Consider the effect of this rule on neighbourhoods sizes k of 3, 4, 5 and 6. Table 2 lists the neighbourhoods of each size which

[6] Note that $n = 400$ and $k = 9$ are the values which apply to the systems examined in chapter 7.

will result in a final state of 1 with this rule. In effect, the critical percentage threshold of neighbours in state 1 jumps around from 67% via 50% and 40% to 50% as neighbourhood size increases from 3 to 6. Even a more subtle description of the cell transition rule will exhibit similar difficulties. This represents a sort of 'quantizing' distortion of the rule set. This is a result of the discretisation of space inherent in CA models and may be a fundamental limitation to the approach.

k	Neighbourhoods leading to state 1				Effective threshold
3	011	111			67%
4	0011	0111	1111		50%
5	00011	00111	01111	11111	40%
6	000111	001111	011111	111111	50%

Table 2 The effect of different neighbourhood sizes on a CA rule expressed in percentage terms. See text for discussion.

In the experiments described in chapter 7, method 1 above has been adopted, although the deformation process used is not the non-symmetric one described above, nor the random one of the small world case. The *neighbourhood-preserving edge pair swap* deformation used is described in section 5.3.4 below. All the experiments described in detail in chapter 7 start with regular grids in which all cells have the same neighbourhood size, so that a neighbourhood-preserving deformation process ensures that a rule set can be defined with one neighbourhood size in mind from the outset, thus side-stepping the difficulties of methods 2 and 3.

Whichever approach is adopted, the overall effect is that an experiment traces a trajectory through GCA space, where the position in rule space is approximately fixed. Clearly, the other two axes of figure 24 — cell space and complexity — must be calibrated or scaled in some way, for this trajectory to be fully described. The measures used for these calibrations are described in the next two sections.

5.3.4 Measures of cell space: graph structure

Numerous measures of graph structure, may be regarded as proxies for measures of cell space. The most obvious measures to use are those based on centrality, and centralisation, since they seem to be the most descriptive of the overall graph structure. The small worlds work suggests that a global measure such as characteristic path length, combined with a local one, such as the clustering coefficient, might work well together in describing graph structure.

However, for these initial investigations, a *relative* measure of graph structure is used, and that is explained here. This measure is based on how different from one another are consecutive graphs in a sequence along a trajectory through GCA-space. The easiest way to do this, is to deform the graph by the same amount in the same way at each step along the trajectory, so that trajectories start at some origin and each subsequent graph is 'equally spaced' along the cell space axis.

In keeping with the restriction mentioned in option 1 above, the deformation method used is a neighbourhood-preserving edge swap. See the illustration in figure 26. Four vertices v_0, v_1, v_2 and v_3 are randomly selected such that v_0v_1 and v_2v_3 are edges in the graph, and v_0v_2 and v_1v_3 are not. Edges v_0v_1 and v_2v_3 are then replaced by v_0v_2 and v_1v_3. Restrictions are placed on how remote from each other the four vertices may be. Thus, v_2 is chosen so that $d(v_0, v_2) = 2$, and v_3 is a randomly selected neighbour of v_2 which is not adjacent to v_1. Overall then, $d(v_0, v_1) = 1$, $d(v_0, v_2) = 2$, $d(v_2, v_3) = 1$ and $d(v_1, v_3) > 1$. Note that since we know that $d(v_0, v_2) = 2$ and $d(v_1, v_3) > 1$ and that these distances are equal to 1 after the deformation, then the characteristic path length seems likely to fall, although this is not certain since $d(v_0, v_1)$ and $d(v_2, v_3)$ may be increased by the deformation.

The distance between two graphs G_0 and G_1 in cell space can then be approximated by the number of edge swaps of this kind that are required to transform G_0 into G_1. Measures broadly of this type have been proposed in the mathematical literature on graphs (Chartrand, Kubicki & Schultz 1998, Goddard & Swart 1996, Hrnciar, Haviar, Monoszova & Bystrica 1996).

This deformation is significantly different from that used by Watts & Strogatz (1998) in their experiments, where edges are replaced by randomly connecting to any other vertex in the graph. The deformation considered here has the property of preserving what might be called spatial proximity in the graph. That is, if the initial

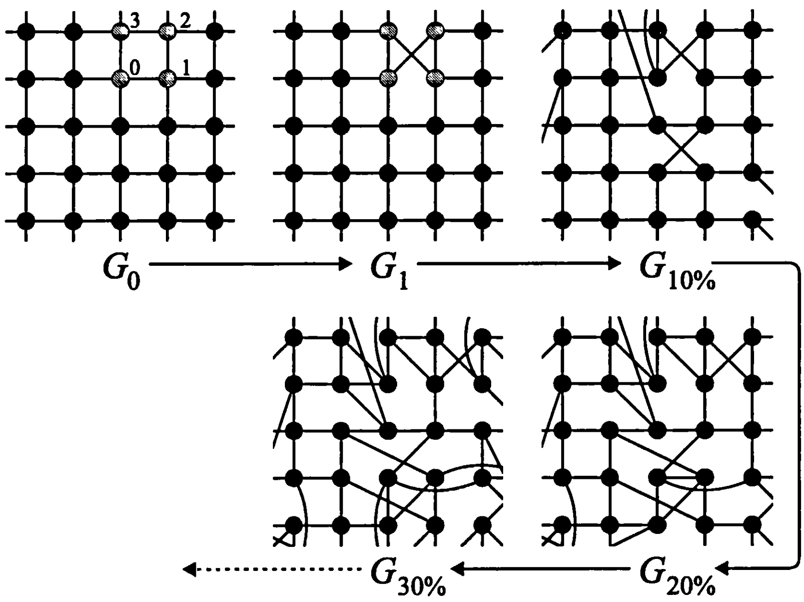


Figure 26 The neighbourhood-preserving edge pair swap. This figure shows the basic deformation $G_0 \to G_1$, and also the effect of subsequent swaps of increasing (indicated) percentages of the original edges for a fragment of a regular lattice.

graph is such that vertices are only connected to spatially near neighbours, then even after several deformations, *most* neighbourhood relations will still be between spatially near vertices. The difference between this sort of deformation and the 'small world' deformation is significant. However, it is equally clear that the difference is a matter of degree. The small world deformation has the effect of rapidly reducing the average distance in the graph while leaving local graph densities relatively unchanged; the neighbourhood-preserving edge swap also reduces average distances, but much more slowly, and local densities — clustering coefficients —are likely to fall at a similar rate (again, see figure 50 on page 162). Ultimately though, both deformations will reduce the graph, however ordered it is at the outset, to a random network.

5.3.5 Measures of dynamic behaviour: spatial set entropy and *spatial information*

We also need a measure of dynamic complexity to calibrate that axis of GCA-space. As we have seen in section 3.3.3 (page 64), Wolfram (1984*a*) introduced a measure — spatial set entropy — which attempts to assess the extent to which the evolution of a CA leads to statistically unlikely configurations. To recap, recalling equation 3.18, spatial set entropy S, in a system with $|A|$ allowable states, is given by

$$S^{(x)}(X) = -\frac{1}{X}\sum_{j=1}^{|A|^X} p_j^{(x)} \log_{|A|} p_j^{(x)}$$

where $p_j^{(x)}$ denotes the frequency of occurrence of each of the $|A|^X$ possible neighbourhood states in sequences of cells of length X. The (x) superscript indicates that the measure is determined over space (rather than time), and we drop it below, where all measures are calculated over space. This measure is in the range 0 to 1 where 0 indicates a very ordered arrangement of cell states and 1 indicates a completely disorganised arrangement of states. Wuensche (1998, 1999) uses this measure with X set such that the set of cells considered is the cell neighbourhood ($X = k$), and this is the approach adopted here. However, as previously discussed, Wuensche's work needs to be extended to apply to two (or higher) dimensional and non-regular graph-based CAs. The extension of the measure is described in this section.

It is interesting to compare this measure with spatial entropy measures in geography (Batty 1974*a*,*b*, 1976*a*, Morrill 1995, provides an application). The principle difference is that the Wolfram measure is summed over a *set of neighbourhoods* of spatial elements, rather than a simple set of spatial elements. To emphasise this point: in Wolfram's simple linear CA examples, a sequence of cells around each cell v_i, of length X, is examined, and is classified into one of the $|A|^X$ possible *neighbourhood* or sequence states. This classification yields the frequency distribution $P(X) = \{p_j\}$ on which the calculation of spatial set entropy in the equation above is based. A simple worked example is shown in table 3. The relatively high value of 0.79 is indicative of a high degree of randomness in the 10-bit string in this case.

System state: $\hookrightarrow$ 0101101001 $\hookleftarrow$

Neighbourhood	Count	Frequency
000	0	0.0
001	1	0.1
010	3	0.3
011	1	0.1
100	1	0.1
101	3	0.3
110	1	0.1
111	0	0.0

...which gives us, for $|A| = 2$:

$$S(X = 3) = -\tfrac{1}{3}[4 \times 0.1 \log_2 0.1 + 2 \times 0.3 \log_2 0.3] = 0.79032$$

Table 3 Sample calculation of Wolfram's spatial set entropy. Sequences of length $X = 3$ in the 10-bit system state string (which is understood to 'wrap' around at each end) are tabulated, and the resulting sum shown.

In Wolfram's and Wuensche's work the classification of neighbourhood states is regarded as unproblematic, because in a linear array neighbourhoods can be ordered in an obvious way — usually 'left-to-right' — so that an X cell neighbourhood, with $|A|$ allowed states has $|A|^X$ *equiprobable* possible states. In a graph-CA, — indeed any two or higher dimensional CA — classification of neighbourhoods in this way is problematic because *there is no natural ordering of cells in a neighbourhood.* A related *spatial information* measure has therefore been developed, which extends the spatial set entropy measure to two- and higher-dimensional systems.

We begin by re-writing the Wolfram formulation — ignoring his normalising $\frac{1}{X}$ factor, to begin with — as

$$S = -\sum_{j=1}^{|B|} p(\mathbf{b}_j) \log p(\mathbf{b}_j) \tag{5.7}$$

In this version, we take a *vector* $\mathbf{b}_j$ as representing neighbourhood state j, so that the set of all possible neighbourhood states is $B = \{\mathbf{b}_j\}$. Typically (and below) $\mathbf{b}_j$ is an ordered list of *cell* state counts $[b_1\, b_2, \ldots, b_{|A|}]$ where the vector components represent the number of cells in the neighbourhood in each of the allowed *cell* states. This is only one, relatively simple way of constructing a neighbourhood state vector from a set of cell states. Other approaches are possible, which might, for example, pay more attention to the spatial arrangement of cells. However, the state count vector method is very general, and alternatives do not affect the generality of the measure described below. Using this particular description of neighbourhood state, neighbourhoods of different sizes are not classifiable as similar,[7] so we must calculate a weighted sum of the entropies associated with the set of neighbourhoods of each size k:

$$S = \frac{1}{n} \sum_{k \in K} n_k S_k = -\frac{1}{n} \sum_{k \in K} n_k \sum_{j=1}^{|B_k|} p(\mathbf{b}_{j|k}) \log p(\mathbf{b}_{j|k}) \tag{5.8}$$

where S_k is the spatial set entropy associated with the set of neighbourhoods of size k, K is the set of neighbourhood sizes which exist in the system, n is the total number of cells in the system, n_k is the number of cells with a neighbourhood of size k, B_k is the set $\{\mathbf{b}_{j|k}\}$ of possible neighbourhood states of size k, and other symbols are as before. Then $p(\mathbf{b}_{j|k})$ is a member of a discrete probability distribution P_k which describes the observed frequency of occurrence of the various possible neighbourhood states $\mathbf{b}_{j|k}$ in neighbourhoods of size k. Note that $\sum_j p(\mathbf{b}_{j|k}) = 1$.

A further amendment to the Wolfram measure is also required in n-dimensional irregular systems. A state count vector approach to neighbourhood classification implies that *not all neighbourhood states are equally probable,* as is assumed in Wolfram's measure. To take this into account, *spatial information* is based on *both* the observed distributions of neighbourhood states, $\{P_k\}$, *and* a set of prior or expected probabilities of the same neighbourhood states, $\{Q_k\}$. This yields a spatial information

[7] Compare the quantization problem discussed above in section 5.3.3.

measure I

$$I = \frac{1}{n} \sum_{k \in K} n_k I_k = \frac{1}{n} \sum_{k \in K} n_k \sum_{j=1}^{|B_k|} p(\mathbf{b}_{j|k}) \log \frac{p(\mathbf{b}_{j|k})}{q(\mathbf{b}_{j|k})} \tag{5.9}$$

where $q(\mathbf{b}_{j|k})$ is the prior probability of occurrence of the k-sized neighbourhood state $\mathbf{b}_{j|k}$, which makes up the probability distribution $Q_k = \{q(\mathbf{b}_{j|k})\}$ and is defined for each neighbourhood size $k \in K$. Note that $\sum_j q(\mathbf{b}_{j|k}) = 1$ as for P_k. The log term in the inner summation is a *relative entropy* (Applebaum 1996, explains the relationship between relative entropy and information) which measures the extent to which a particular system state — described by the observed distributions of neighbourhood states $\{P_k\}$ — deviates from prior expectations of those states $\{Q_k\}$. The more 'unexpected' a particular arrangement of cell neighbourhood states is, the higher will be the spatial information. The inclusion of the set of prior probability distributions $\{Q_k\}$ is important because it means that the unexpectedness of a particular system state — the information which it contains — can be assessed with respect to the overall numbers of cells in each of the allowed cell states.

Note that the expressions in equations 5.8 and 5.9 do not specify a base for the logarithm in the inner summation. Re-examining Wolfram's measure, we note that the logarithm base is $|A|$. This is the key to the 'self-normalising' property of Wolfram's measure, since we can re-write his expression (again) as

$$S(X) = -\frac{1}{\log_{|A|}(|A|^X)} \sum_{j=1}^{|A|^X} p_j \log_{|A|} p_j \equiv -\frac{1}{\ln(|A|^X)} \sum_{j=1}^{|A|^X} p_j \ln p_j \tag{5.10}$$

This makes the relationship between the choice of logarithm base and the normalisation clearer. In fact, the logarithm base of the summation terms is unimportant, provided that the denominator in the scaling factor is the logarithm *to the same base*, of *the total number of possible neighbourhood states*. The appropriate denominator in the development of our new measure is then $\log_x |B_k|$, where the choice of x is dependent on the logarithm base in the inner summation. The introduction of the relative entropy term in equation 5.9 actually makes the choice of logarithm base arbitrary, since the ratio of the logarithms of two numbers is the same, regardless of the number base, so we may as well use natural logarithms. We can therefore finalise the spatial information I as

$$I = \frac{1}{n} \sum_{k \in K} \frac{n_k}{\ln |B_k|} \sum_{j=1}^{|B_k|} p(\mathbf{b}_{j|k}) \ln \frac{p(\mathbf{b}_{j|k})}{q(\mathbf{b}_{j|k})} \tag{5.11}$$

Note that this expression makes it important that the number of possible neighbourhood states of size k — that is $|B_k|$ — be calculable. The state count vector approach can be shown to admit $\binom{k+|A|-1}{|A|-1}$ possible neighbourhood states when it is used to describe k-sized neighbourhoods with $|A|$ allowed cell states.[8] In practice, in the examples below, where only one value of k applies throughout the system, the weighting of each I_k is unimportant, and we therefore ignore the $\ln|B_k|$ weighting factor, and use the expression in equation 5.9, with natural logarithms.

It would be nice to be able to normalise I so that direct comparisons could be made between systems of different sizes. A possible approach might be to estimate the maximum possible value of I_k for each neighbourhood size, and normalise by dividing each I_k by this estimated maximum. We can see from equation 5.11 that $\max I_k$ occurs when *all* k-sized neighbourhoods are in some least likely neighbourhood state $\mathbf{b}_{\min|k}$, so that the sum in equation 5.11 becomes the single term $\ln \frac{1}{q(\mathbf{b}_{\min|k})} = -\ln q(\mathbf{b}_{\min|k})$. However, there is a serious weakness in this approach, because the least likely *possible* system configuration depends not just on some aspatial combinatorial determination of Q_k, but *on the structure of the graph itself*, since this affects the combinations of neighbourhood states $\mathbf{b}_{j|k}$ which can occur. Inclusion of the interdependency effects due to the graph structure is forbiddingly complex, in general. It is unlikely that a configuration in which all neighbourhoods are in state $\mathbf{b}_{\min|k}$ is feasible. To see this, think about devising a string of bits for the example in table 3 which has any particular neighbourhood configuration, other than 000 or 111, in all locations. Since the main interest here is in the time evolution of spatial information for systems which do not change in size or in their neighbourhood size distribution, and not in comparisons between different systems, it therefore seems safer to use the un-normalised expressions in equations 5.9 or 5.11.

In fact, the criticism that the interdependency effects of the relational structure of these systems are ignored, could be levelled at almost any conceivable approach to developing a measure to determine how unexpected is a particular arrangement of the

[8] The number of state count vectors for k-sized neighbourhoods with $|A|$ allowed cell states is equivalent to the number of ways of summing $|A|$ whole numbers $\{0, 1, 2\ldots\}$ so that they total k. If we denote this by $X(k, |A|)$ it is clear that $X(k, |A|) = \sum_{n=0}^{k} X(k-n, |A|-1)$ since the expression on the right-hand side of this equation is the sum of the numbers of ways of extending the sum having chosen each of the allowed values $0\ldots k$ for the first term. Manipulation of this expression yields the recursive definition $X(k, |A|) = X(k, |A|-1) + X(k-1, |A|)$. Combining this with the observations that $X(k, 1) = 1$ and $X(0, |A|) = 1$ we arrive at the expression in the text, after constructing a table of values, which proves to be a transformed Pascal's triangle.

system. Thus, measures of spatial auto-correlation intended to assess the likelihood of some arrangement occurring by chance, by counting the number of 'joins' between elements in various states, are similarly prone to criticisms of circularity when the unlikelihood of some arrangement being random is forwarded as proof of a spatial process at work. Offering such 'proof' is potentially contradictory since it implies that the measure itself should be recalculated to take the existence of the posited processes into account — especially since there is no really satisfactory definition of randomness which does not rely on a notion of expectations. In fact, these difficulties go right to the heart of philosophical and methodological debates in statistics — and science more widely — about the nature of probability, evidence, and the scientific method. In the current work, we know that there is a process at work (since we will put it there!), so assessing it with respect to the likely outcome of a hypothetical, random process is arguably flawed. However, a notional random process seems to be the only datum available if we wish to compare the outcomes of a series of simulations to one another. In fact, as we shall see, actual values of I are not of interest in themselves. It is variation in I over time (which characterises the behavioural class of a system), and as system structure changes, which is interesting.

These are complex arguments, and they are ignored in much of what follows, for the purely pragmatic reason that spatial information I is only tentatively proposed as a way of beginning to explore this area, and is not intended (at least initially) to be any more widely applied than that. Therefore, the expression for I in equation 5.9 will be used to examine the complexity of the dynamic behaviour of models. The immediate neighbourhood of each cell is examined to determine the sets of neighbourhood state distributions $\{P_k\}$ and $\{Q_k\}$. It is worth noting that precisely the same measure could be used for other sub-sets of cells in the system. In the current context an obvious alternative would be the second and higher order neighbourhoods of each cell (compare the concept of 'lag' in spatial auto-correlation measures).[9] Whether other subsets might be of interest in other contexts is an open question. It is certainly possible to imagine a random sampling approach, or subsets based on walks on the graph of various lengths.

[9]Determination of spatial information for higher order neighbourhoods is computationally more intensive, however. Note that a whole spectrum of lagged spatial information values could be determined, which might well shed further light on the spatial patterns in a configuration — compare Dykes's (1994) use of differently lagged spatial autocorrelation statistics.

An example

To make these concepts more concrete,[10] it is useful to introduce an example, based on the system neighbourhood classification and prior probability calculation adopted for the experiments in chapter 7. The system is a 400 cell toroidal grid, where all cells have 9 cell neighbourhoods (the cell itself, and 8 other neighbours), so $n = n_{k=9} = 400$ and the set of neighbourhood sizes $K = \{9\}$ (see figure 49 on page 161 and the accompanying discussion for more details of the graph). Each cell may be in one of two states, hence $A = \{0, 1\}$ (say). Neighbourhoods are classified by the state count approach, so the set of neighbourhood states is

$$B_{k=9} = \{[0\ 9], [1\ 8], [2\ 7], \ldots, [9\ 0]\} \tag{5.12}$$

The prior probability distribution $Q_{k=9}$ of each of these $|B_{k=9}| = 10$ neighbourhood states is calculated as the probability of selecting 9 elements without replacement, from a population of $n = 400$, such that b_0 elements are in state 0, and b_1 in state 1, given that $n_{a=0}$ and $n_{a=1}$ are the respective numbers of cells in the whole system in each state. This is the *hypergeometric* distribution given by

$$q\left(\mathbf{b} = [b_0\ b_1]\right) = \binom{n_{a=0}}{b_0}\binom{n_{a=1}}{b_1} \Big/ \binom{n}{k} \tag{5.13}$$

Note that $k = b_0 + b_1$ and $n = n_{a=0} + n_{a=1}$. This calculation requires that we assume that the states of neighbourhoods are independent from one another. This is not strictly true. However in systems where neighbourhoods are small relative to the system size ($k \ll n$) it seems likely that we may be justified in adopting this simplifying assumption. A more correct calculation of Q_k would be sensitive to the order in which neighbourhoods in the system were considered, since the relationship between any particular neighbourhood's (unknown) state, and the states of any previously considered neighbours' (known) neighbourhood states is related to the graph structure itself. However, this approach is impractical in general, because it would require that the calculation of Q_k be based on all the $n_k!$ possible orderings of the n_k neighbourhoods of size k. This is another example of the forbidding complexity introduced into probability and statistics by the interdependency effects of a graph (or indeed of space).

[10] And also hopefully to demonstrate that the foregoing detail is not necessary for an appreciation of the basic idea!

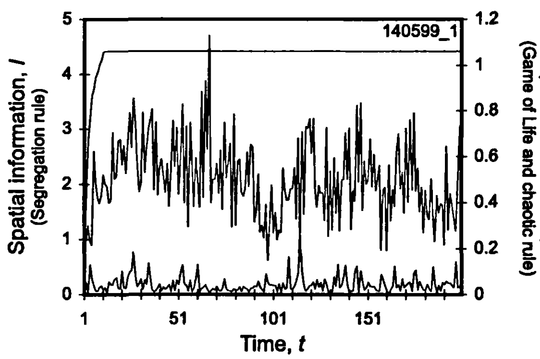
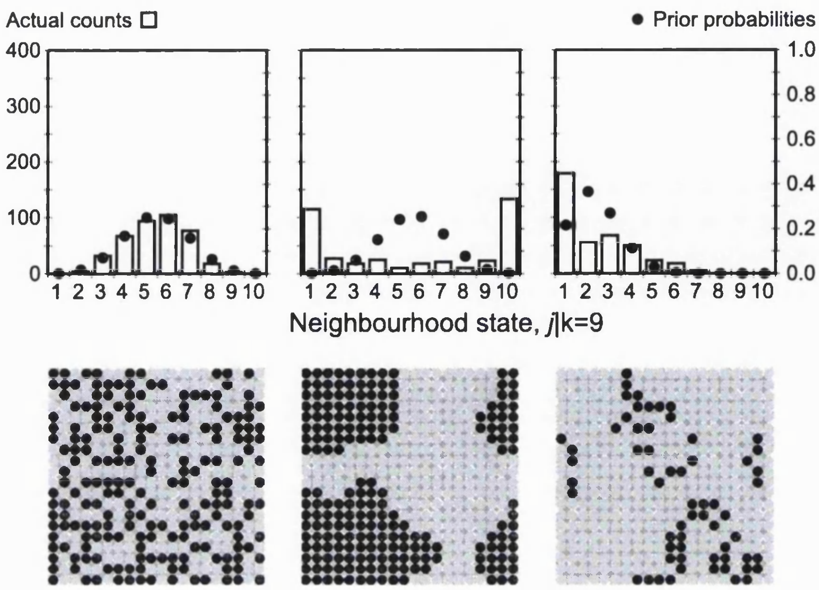


Figure 27 Typical spatial information time-series, for a segregation (class 2) CA, the Game of Life (class 4) CA and a chaotic (class 3) CA.

Noting these *caveats*, the time-series evolution for two well-known grid-based CA, and one other, are plotted in figure 27. The upper time-series represents the evolution of a segregation type CA rule, whereby each vertex (cell) adopts the same state as the majority of its neighbours. This sort of model is similar to Schelling's (1971) model of racial segregation phenomena in social systems. This rule exhibits a distinctive rapid increase in spatial information, as the system 'sorts' itself into distinct areas of connected vertices in the same state. Once segregation is complete the system's spatial information attains a fixed value. This is characteristic of class 2 system behaviour. The middle time-series shows the evolution on this measure of the 'Game of Life' CA. Although the actual value of spatial information attained is much lower than in the segregation CA case (note the right hand axis on figure 27) there is a distinct increase over the near-zero value of the random starting configuration, and this intermediate value is maintained over time — indicating complex class 4 behaviour. Finally the chaotic (class 3) behaviour of the third CA rule is evident in the apparently random near-zero fluctuations of the lowest time-series.

Understanding this measure is important, since it is extensively used elsewhere in this thesis. Figure 28 shows three examples of spatial distributions and the resulting measure to give a sense of the behaviour. In each case the distribution of neighbourhood states classified according to state counts is shown as a histogram. Superimposed on this is the prior probability distribution Q_k (the 'blobs' in each case). The measure's value in each case reflects the divergence of these two distributions. Thus,



(a) (b) (c)

Figure 28 The behaviour of spatial information for 3 different spatial patterns. Spatial information I has values of (a)0.023925, (b)4.420198, and (c)0.435288 in these cases.

in the first case, a randomly generated distribution of cell states, where the match between the expected and observed distribution of neighbourhood state is close, I is low, with a value of 0.024. The second and third cases show the patterns generated by 50 time steps of the segregation and Game of Life CA rules, starting from the random configuration of case a. In the segregation CA case the divergence of the arrangement of states from expectation is great since many neighbourhoods have either all cells in the same state — that is $\mathbf{b}_j = [0\ 9]$ or $\mathbf{b}_j = [9\ 0]$. This results in a large value of I of 4.420. For the Game of Life CA, although the effect is less clear-cut, a non-zero value of I (0.435) still results because of the over-representation of neighbourhoods with no dark cells, and the under-representation of cells with 1 or 2 dark cells. This over- and under-representation is indicative of the clustering of dark cells in the configuration.

This behaviour may be contrasted with spatial auto-correlation statistics (Cliff & Ord 1973) which count instances of adjacency of pairs of like or unlike states. Such measures are bipolar, producing either positive or negative values depending on whether adjacency tends to be associated with like or unlike states. Spatial information is not sensitive to these opposed tendencies but should pick up distributions which have many similar neighbourhoods, unless these are to be expected. Very little research has been done into the relationship between these two kinds of measure of spatial structure (Gatrell 1977, is an unusual exception), in spite of the widespread promotion of both autocorrelation and entropy statistics. Spatial information is preferred in the current context because it makes the connection with CA/complexity studies clear and because it is very directly aimed at assessing the way in which a particular process organises spatial pattern by forcing variation from random or disorganised outcomes. Phipps (1989) takes a similar line in exploring the spatial behaviour of a single CA model.

5.3.6 Summarising spatial information time-series data

The time-series shown in figure 27 do not in themselves provide a measure of complexity. This must still be determined by observing the 'typical' evolution of the spatial information measure over a range of cases. In general a Wolfram class 2 process may raise the information value attained but will stabilise or enter some (short) periodic cycle. A class 3 (chaotic) process will tend to show apparently random behaviour near $I = 0$, whereas class 4 (complex) processes will tend to show the sort of behaviour of the Game of Life CA with long transients with I definitely above zero. In fact, this classification is by no means clear cut, and any particular system is likely to require careful investigation of many different 'runs' before its behaviour can be characterised in this way.

To this end, it would be convenient to use summary statistics over a set of runs, from different starting configurations to characterise behaviour more easily. However, individual starting configurations show wide variation in the evolution of the measure. See figures 29 and 30. In these examples, the pattern of behaviour for the segregation rule is still clear, but that for the Game of Life is not so obvious. It is important to consider whether we can legitimately summarise the range of behaviour exhibited by these cases using simple statistics such as a mean spatial information calculated at each time step, for sets of different starting configurations such as these.

Applying the *Kolmogorov-Smirnov* (K-S) goodness-of-fit test (Kanji 1993) to see how well the 50 different spatial information values at each time step match a normal distribution (with mean and variance calculated from the 50 sample values), is one

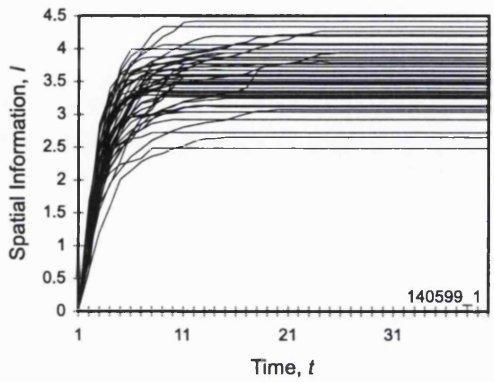


Figure 29 Spatial information for the same system run from 50 different starting configurations (Segregation rule). As already noted in figure 27 this type of rule leads rapidly to a high fixed value of the spatial information measure, and this behaviour is observed in all cases.

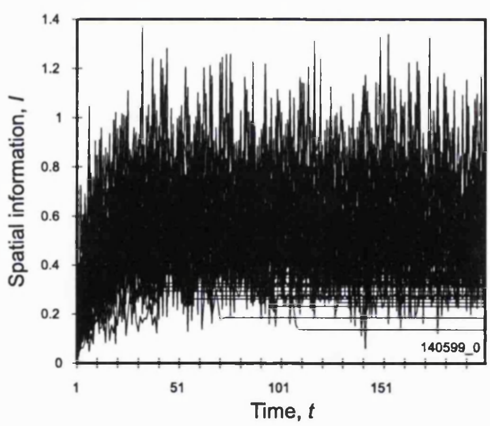


Figure 30 Spatial information for the same system run from 50 different starting configurations (Game of Life rule). Different configurations produce dramatically different sequences of values in detail, so that the overall behaviour is hard to summarise.

way to assess the validity of an averaging process. The K-S test involves finding the largest difference D_{ks} between the cumulative frequency distribution (CFD) of the empirical data, and the CFD of the distribution we wish to test against, in this case the normal distribution. The worst case example for the two sets of time-series data already presented is shown in figure 31. It is useful to see the worst case, so that the derivation of the statistic can be understood, and so that the generally better fit of most cases is appreciated.

For the 40 time-slice data sets from the segregation CA in figure 29, D_{ks} is never greater than the value which would be expected to occur randomly with $p = 0.01$. This indicates that it is reasonable to assume that the values of spatial information at any time step t are normally distributed, and that taking a mean value to produce a summary time-series is a reasonable procedure, *in the case of this CA*.

The Game of Life CA spatial information time-series are more problematic — as might be surmised from figure 30! The K-S test is conventionally applied to single sets of empirical data. The investigator consults tables of critical maximum values of D_{ks}. These list, for various sample sizes n, the maximum value of D_{ks} which can be

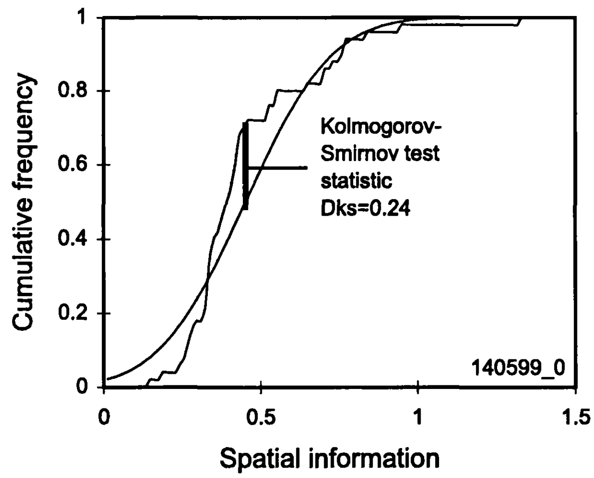


Figure 31 The Kolmogorov-Smirnov goodness-of-fit test. Cumulative frequency distributions for the normal distribution (smooth curve), and data for time step 173 of the Game of Life CA. The maximum vertical distance between these distributions is the Kolmogorov-Smirnov goodness-of-fit statistic D_{ks}.

expected to occur, with various probabilities, if the data matches the assumed distribution. With only one sample data set to assess, the conventional application of the test is to check whether the value of D_{ks} obtained is greater than the critical value at the probability of interest. This translates to a probability that the sample is normal. Thus, if D_{ks} is less than the critical value for all $p \leq 0.1$, we can say that there is a 90% likelihood that the distribution is normal. In assessing the normality of the 200 time-slice Life CA data sets derived from the 50 spatial information time-series in figure 30, it is instructive to assess the cumulative performance of the data sets against this test. That is, we count how many K-S tests are failed at various p levels (0.20, 0.15, 0.1, 0.05 and 0.01) by the data set at time t, *and all the preceding data sets*. This gives us a secondary test of the set of data sets up to time t, such that the actual frequency of occurrence of D_{ks} values above those to be expected at each probability p, $f_{\mathrm{FAIL}}(p)$, can be determined, and this can be compared to p. If $f_{\mathrm{FAIL}}(p) > p$, then the cumulative data sets up to that time period fail to support the assumption of normality with probability $(1 - p)$. Effectively, we are using the number of failures of the K-S test, as a gauge for how well the data fit the normal distribution. If we record the lowest p at which the cumulative data sets pass this multiple K-S test, we get an indication of the likelihood of normality of these data. The result is indicative only, since there are aspects of this procedure which may be arguable.

This information is plotted in figure 32 along with the fraction of time-series which show evidence of having attained long-run stability (or 'dropped out') at each time step. This shows that over the first 20 or so time steps, and after about 135 time steps, there are no grounds for assuming the normality of sets of Life CA spatial information time-series. Before $t = 20$ many of the spatial configurations observed are not typical of the Life CA — generally because the overall density of live cells has not yet settled to its usual level of around 30%. After $t = 135$, 4 or 5 of the 50 time-series have 'dropped out' by stabilising, often in low-density configurations with low values of I (a few of these are visible as horizontal lines in the lower right of figure 30). The implication of this result is that using simple averaged time-series to summarise the dynamics of the Life CA is unjustified before $t \approx 20$ and after $t \approx 135$. Between these times there may be reasonable grounds for assuming normality, and using a summary statistic.

This result demonstrates the difficulty of applying simple statistical methods to

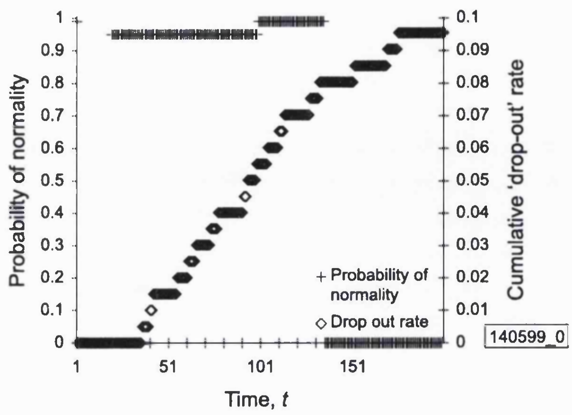


Figure 32 Assessment of Life CA time-series against normality. The cumulative failure rate on the K-S test of time-slice data sets for the CA has been compared to D_{ks} at various confidence levels. The most stringent confidence test passed by the data sets is plotted — demonstrating that before time step 20 and beyond time step 135 (approximately) the time-slice data sets for the Life CA cannot be regarded as normally distributed, so that taking a simple average value may be misleading. The increasing number of time-series which 'drop out' is also seen here and is a major reason for the difficulty of summarising behaviour using a simple average.

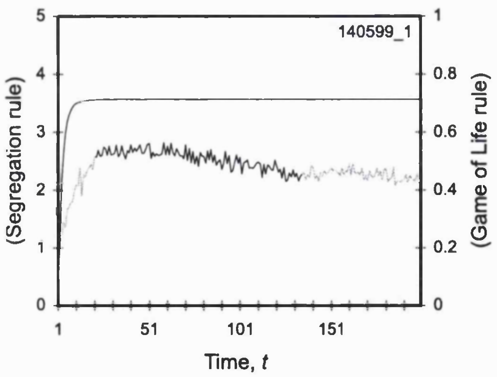


Figure 33 Mean spatial information time-series for the two CAs. The time period in the Life CA over which the averaging process is probably not valid is shown in grey.

complex systems. It is important to examine the various different kinds of behaviour exhibited by systems before attempting classification. In the Life case, the major problem is that in this finite 'world', many systems stabilise after 150+ time steps, skewing the distribution of system states such that no simple statistics can be applied.

Oblivious to these *caveats*, averaged time-series for 50 different starting configurations of the segregation and Life CA rules in a 400 vertex graph are plotted in figure 33. From these it is easy to classify the segregation CA as class 2. The Life time-series also tends to support its classification as class 4, although we must be wary of the simplification inherent in the averaging process, and a different approach is adopted in chapter 7.

These are relatively clear cut cases (chosen for that reason), but in general there is no easy way to convert a measure like spatial information into Wolfram's four-way classification of dynamic behaviour, and in practice each graph-CA's behaviour has to be observed in some detail before a manual classification into these categories can be made. The knowledge gained in this process is likely to suggest different appropriate summary statistics for each case. In chapter 7 simple interpretation of the time-series evolution of spatial information is used in investigating the structure-process relation in graph-CA models.

5.4 Discussion and conclusions

This chapter has briefly introduced the graph cellular automaton (GCA) model concept which is the main focus of the remainder of this work. An approach to the investigation of global structure-process relations in such models has also been developed, including the required measures of structure and process. It is evident from the exposition of these ideas that this is a fraught undertaking! The scale of the space to be explored and the difficulties of developing measures which are comparable across a series of GCA models with different structures are formidable problems. This may go some way to explaining why physicists and computational theorists have concentrated their efforts on the more tractable problems of how rule set variation influences the behaviour of 1-D cellular automata.

Nevertheless, the various methods and measures described *do* provide a basis for some sort of progress, as should become clear in chapter 7. Interestingly, the more difficult aspects of the methodology described draw attention to two similar issues. The question of what it means to hold a rule set constant while varying model structure is fundamental to the relationship between structure and process itself. Similarly, the spatial information measure derived in this chapter raises the issue of the relationship between patterns (generated by processes) and structure too — insofar as many of the difficulties encountered in developing and applying this measure relate to the interdependence of cell neighbourhood states due to the pattern of (spatial) relations between them.

Even so, if we bear the detailed arguments and *caveats* of this chapter in mind, there is still scope for the beginnings of an investigation into the relationship between spatial structure and spatial processes, using the graph-CA model. Before reviewing the results of such an investigation in chapter 7, the next chapter describes a computer implementation of the model, and various tools for manipulating, displaying, and analysing dynamic behaviour.

Chapter 6

Model implementation — the *graphca* program

This chapter describes the modelling environment — *graphca* — which has been continuously developed through the course of the research described in this thesis. After the rigours of the last chapter, and the impending rigours of the next, this chapter may come as light — albeit somewhat dull! — relief.

As will become clear, the *graphca* environment is not the 'finished article', but has been developed in response to the unfolding needs of the present research. This research is about urban spatial models in general, and graph-CA representations in particular, not about computer programming, or about a particular implementation of such models, so I do not intend to describe the internal workings of *graphca* in detail. However, this chapter should serve to indicate the scope of *graphca*, and its potential for further development. The intention then, is simply to demonstrate how the model proposal of the previous chapter has been developed, and how suitable it is for computer implementation.

6.1 The *graphca* program

Figure 34 shows a screenshot of the graphca program in use.

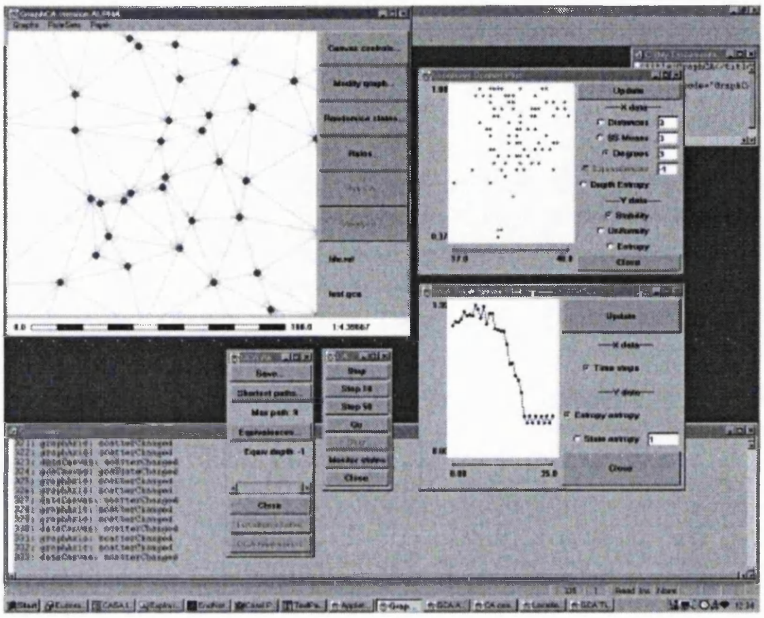


Figure 34 The *graphca* program.

6.1.1 Program functionality

We can use figure 34 to get a sense of the various capabilities provided by the *graphca* program. In the top left of the figure, the main program display window is visible, with a fragment of a spatially irregular graph-CA model displayed. The model is displayed as a graph, in a 2-D display area, with cells (vertices) represented by coloured circles, and edges as arrows joining them. Cell colours correspond to the different allowed discrete cell states. Standard 'pan-and-zoom' style navigation around the display area is possible so that the model structure can be seen in aggregate and in detail.

It is possible to click on cells to obtain more detailed information, and, in particular to see a list of neighbouring cells. Neighbouring cells may be removed from this list, and different cells added, so that the structure of the graph-CA is editable by hand. Individual cell states can also be changed by hand. Such manual adjustment of the model structure is rather tedious, however, and given the interest in carrying out experiments in the effects of structural change on system dynamics, automated tools are also provided for this purpose. A cell state randomiser allows approximate specified fractions of cells to be assigned states from among the allowed cell states in a random fashion. A tool which randomly adds or removes specified numbers of edges is also available. This 'graph editor' can also perform a specified number of random edge pair swaps, as described in the previous chapter, and perform Delaunay triangulation on the model cell locations.[1] Both these tools are controllable in a batch mode which was heavily used in compiling the results reported in the next chapter. These, and other program tools, are launched by clicking the large buttons on the right-hand side of the model display area. Tools appear as small windows like the two shown below the main display in figure 34.

The current state of a model — vertices, their locations and states, and edges — can be saved to a readable text-formatted file at any time. A more compact file format stores only the cell states for a particular model, which enables a number of model initial configurations to be stored for experimental work. No provision has been made for building a model interactively by hand. Initial grid-based models were generated using spread-sheet programs, since the regular spacing of a grid is relatively easily produced in this way. Limited graph generation capability is provided by a random vertex generator, which scatters 400 new vertices randomly around the display area. These may then be linked by hand, or triangulated. This method was used to generate the model discussed in section 7.4 (pages 175ff.). Examination of models based on real geographical settings is most easily achieved by building the model structure in a GIS, and exporting the resulting vertices and edges to a *graphca* readable text file. This method was adopted for the *Gentrification* extension of *graphca* described in part III. The required file formats are described in appendix A.

Actually running a graph-CA model requires transition rules to be defined. A simple rule editor tool is provided. The CA transition rules are themselves regarded as a graph — states are vertices, and the various possible transitions are edges. Using the rule editor tool up to 10 allowed cell states may be added to the rule set. A possible transition between any two states can be added, and is defined by specifying limits on a 'normalised state-count vector' which must be satisfied for that transition to occur. In a 2-state system the $0 \rightarrow 1$ transition might be specified as [0.25 0.5] [0.0 1.0], meaning that if between 25 and 50% of a cell's neighbours are in state 0 and any number of neighbours are in state 1, then the transition $0 \rightarrow 1$ will occur. This is supplemented by designating some transitions as 'not applicable' or 'default'. A transition which is

[1] The model fragment seen in figure 34 is based on a Delaunay triangulation.

not applicable is ignored; a default transition is the one which occurs if no other transition conditions are satisfied. The specification of rules is therefore flexible within limits — permitting many rules which Wolfram (1983) designates as 'totalistic', but making many more complex conditional IF-THEN-ELSE style rules rather difficult to implement. Such a decision-tree architecture could conceivably be implemented within the current program (also as a graph!) but the increased sophistication was not required for the research described in this thesis. Transition rules are also stored in a simple text format also described in appendix A.

With a graph and rule structure specified a graph-CA model may be run. Model behaviour is observable in the main program window, where cell states are dynamically updated as 'blinking lights', as cell colours are redisplayed in new colours to reflect changes in their state. Analysis tools are also provided, to enable more quantitative analyses of the model behaviour to be undertaken. Graph structure measures can be generated — distance based centrality measures (for the whole graph and radius-limited), the 'Sparrow' equivalence class method, and also the neighbourhood-size at various distances (that is the number of vertices within $1,2,\ldots,n$ edges of the vertex). These results can be stored with the model itself, for rapid retrieval, since they cannot be calculated quickly. The relevant formats are described in appendix A. Storing graph analysis in this way allows dynamic display of the behaviour of cells plotted against the graph structure measures to be produced, as in the upper right display in figure 34 (see also figure 46 on page 156). Provision is also made for a dynamically updating display of the spatial information value for the current system state — as shown in the lower right plot in figure 34, and also in figure 45 on page 155. The resulting time-series can be written to a simple spread-sheet readable text-formatted file.

The remainder of this chapter describes the implementation of the *graphca* program in outline. This description is not necessary for an appreciation of the results reported in chapter 7, or to a reading of part III, but is provided for completeness. It may be better, on first reading, to return to this chapter after reading more about the results of subsequent research.

6.1.2 Program implementation

Graphca is written in the *Java* programming language. The way in which *graphca* has been gradually developed has led to its structure and design being rather ungainly, so that any neat description of its details would be both misleading and irrelevant (the code is also extensive, with over 10000 lines, which puts detailed description well beyond the scope of this thesis). In effect, the development of *graphca* has been part of the research process itself. It is in this light that it is worth considering the advantages of the *programming* environment in such an open-ended problem domain. *Java*'s developers, SUN Microsystems[2], described it at its launch as

> "A simple, object-oriented, distributed, interpreted, robust, secure, architecture neutral, portable, high-performance, multi-threaded, and dynamic language." (Flanagan 1997, page 3)

Aside from demonstrating the commitment of the computer industry to an 'enhanced' version of English[3], many of these buzzwords refer to features of the *Java* language which have been useful to the development of *graphca*:

- *Simplicity* is obviously a subjective, non-technical aspect of the *Java* language. Nevertheless, its lineage (it bears many similarities to C, which in turn derives from *Pascal*) make it as readily understood as any modern, fully-featured programming language is likely to be. Much of the *Java* design philosophy seems to have stemmed from frustrations with the complexities of its older object-oriented cousin C++, and many of the more subtle complexities of that language have simply been removed from *Java*. Given the author's limited programming experience, this language feature has been a definite advantage.

- *Object-orientation* (OO) describes a language structure, which is also a way of thinking about software design problems. In OO languages, data and procedures are bundled together into software *objects*. The design task is to decide on an appropriate set of objects for representing the problem domain. Objects are specific instances of object classes, and a program consists of a set of *class* definitions. The class definition specifies both the constituent *instance variables* of any object in the class which describe its current state, and the *methods* which

[2] *Java* is available to download from `http://www.javasoft.com`.
[3] English 1.1, perhaps?

may be performed upon instance variables. This approach is not at all obvious at first, but generally leads to more easily understood programs. It also means that code is encapsulated into 'chunks' with well defined interfaces to the rest of the code, which aids in program maintenance and debugging. An advantage of *Java*'s object-orientation is that it is unavoidable: there is no way of writing programs in *Java* which are not object-oriented. This makes learning the OO paradigm much easier, although the learning curve is steep. The advantage of OO in *graphca* is clearest in the building of graph data-structures. Many of the *graphca* classes are described in more detail in the next section, where the significance of OO should be clearer.

- *Interpreted* languages while suffering some performance limitations relative to compiled languages have the significant benefit of much more rapid prototyping. Note that *Java* is not straight-forwardly interpreted. Source code is first compiled into byte-code `.class` files which are run in an interpreted manner by the native *Java Virtual Machine* (JVM). However, the debug-compile-run cycle is fast enough to be almost 'interactive' in feel and this is vital to a constantly changing program such as *graphca*. With the availability of faster alternatives to Sun's `javac` compiler (such as IBM's *Jikes!*[4]), and extensive compile-time debug information and on-line documentation, the development of *graphca* has proceeded relatively painlessly.

- *Architecture neutrality* refers to the fact that given an appropriate native JVM *Java* classes should be executable on any computer platform. Given the uncertainty at the outset of this work of the likely performance of any program, a certain amount of flexibility was desirable. In the event, a standard current x86 desktop PC has been adequate to the task.

- *Portability* refers to the ability to deliver *Java* programs (*applets* in this context) over the Internet. This feature was seen as desirable at the outset.[5]

- *High-performance* has always been the most controversial of Sun's many claims about *Java*. An interpreted language will always be slower than some of its

[4] Available at `http://www10.software.ibm.com/developerworks/opensource/jikes/`.

[5] A version of *graphca* is available at `http://www.casa.ucl.ac.uk/~david/webGraphCA.html`. This is a cut-down version, but gives a feel for the program. It is likely to remain available at this URL until mid-2002. It is unlikely that this implementation will remain current.

compiled rivals. In the *graphca* application such performance issues have not arisen, particularly since the advent of the much faster *Java 1.2* (latterly *Java* 2) In future, it is likely that *Java* processors — actual physical devices — will be developed so that any performance concerns will fade.

- *Multi-threading* (the ability to execute multiple sections of code simultaneously) is a core capability of *Java* which is paradoxically not found in 'bigger' languages such as C++. In *graphca* multi-threading is used but is not essential to the design. Arguably, its availability has been a distraction from the core design and in some cases has led to over-complication and difficulty in the programming of *graphca*.

In addition to these features, *Java*'s support of a range of *graphical user interface* (GUI) components, graphics features, and the `.zip` file compression format have been invaluable to the development of *graphca*.

6.2 The *graphca* packages

Java programs are organised into *packages* of related classes. The *graphca* program consists of nine different packages. These are described in more detail in the remainder of this chapter, but can be summarised as follows:

- `GraphCA` contains only one class, `graphca`, which is the main program itself, together with eight sub-packages which constitute the bulk of the program code.
- `GraphCA.superClasses` contains the base classes for the program.
- `GraphCA.gca` contains classes which can represent a graph-CA model. The `GCA` class in this package represents a graph-CA model and provides methods for update according to a CA rule, and also for modification by breaking, swapping, or adding edges.
- `GraphCA.ca` contains classes to represent a CA rule set.
- `GraphCA.gui` contains most of the graphical user interface (GUI) elements of the program.

- `GraphCA.analysis` contains classes for the analysis of graphs and graph-CA models. In particular classes are provided which can generate spatial information and graph structure measures.
- `GraphCA.experiment` provides classes for setting up multiple runs of the program and saving results to files on disk for subsequent analysis.

The above packages are described in more detail in the remainder of this chapter. Two other packages are not considered further, but for completeness these are:

- `GraphCA.gcaIO` contains classes for reading and writing instances of other classes in the program from and to files.
- `GraphCA.compGeom` contains classes to represent lines, edges, triangles and the like, and many methods for the manipulation of such geometrical entities. This package is based wholly on a package developed at the FernUniversität Hagen (Icking, Klein, Köllner & Ma 1997) for producing Delaunay triangulations and Voronoi diagrams.[6]

Note that the naming convention using the dot (`.`) separator indicates the hierarchical and nested organisation of packages.

Within and among packages of classes a more general set of OO class relationships and types apply. First, classes may *inherit* their functionality and structure from other classes. Subclasses inherit from *superclasses*, or, in *Java* terminology, they `extend` superclasses. The network of inheritance relationships among a set of classes is usually referred to as the *class hierarchy*. Inheritance is an important mechanism in the modelling of problem domains because it is a natural model for the relationship between generalised and specific examples ('shapes' as opposed to 'circles', for example).

Additionally, classes may be *concrete* or *abstract*. Concrete class definitions are those which define objects that may be instantiated at run-time. These are the classes which define program operations and functionality. Abstract classes may not be instantiated; rather they are generally used to refine the class hierarchy by bundling elements of the class functionality into manageable units, and by allowing sets of similar classes to provide different implementations of methods with the same names. Thus,

[6] Available at `http://wwwpi6.fernuni-hagen.de/Geometrie-Labor/VoroGlide/`.

an abstract class 'shape' might have subclasses 'quadrilateral' and 'ellipse'. These in turn might be abstract classes with subclasses 'square', 'rectangle', 'rhombus' and so on. This organisation of the class hierarchy can make it easier to manipulate sets of objects. So, for example each subclass could implement its own version of a method `getArea()` which returns the area of the shape, and no elaborate class checking would be required to perform this action on an array of shape objects.

A further refinement of the class hierarchy is provided in *Java* by *interfaces*. An interface is an abstract class definition which only defines a set of 'dummy' methods. Any class may be declared so that it `implements` an interface, in which case it must provide implementations for the dummy methods of the interface. This mechanism makes it possible to by-pass the rigidities of a strictly hierarchical organisation of class-definitions.

These abstractions should become clearer in the descriptions of *graphca* packages and classes which follow. Figure 35 shows the symbols used in the diagrams in the remainder of this chapter to describe the *graphca* class hierarchy. This formatting is

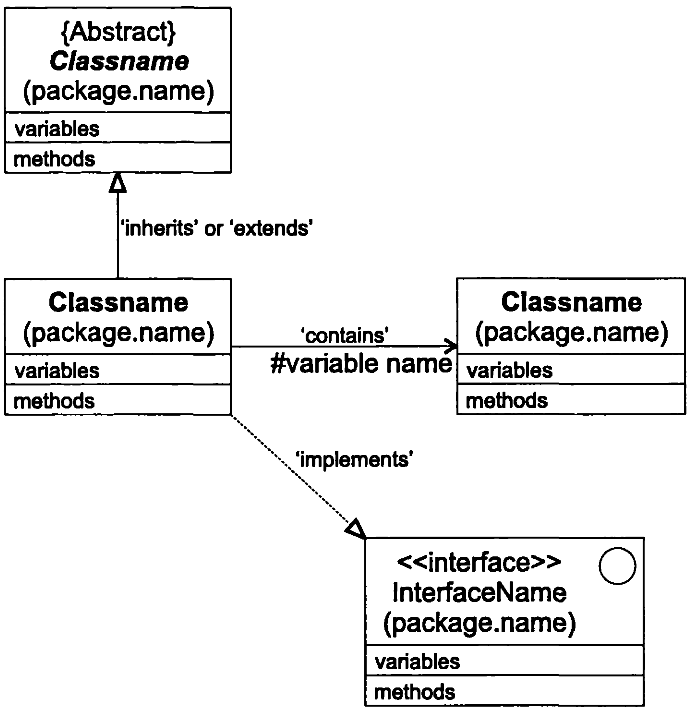


Figure 35 Key for the class hierarchy diagrams.

loosely based on the *Universal Modelling Language* (UML), which has emerged recently as the preferred systematic design methodology for object oriented systems (Alhir 1998, Anonymous 1999). UML is an extensive and complex descriptive language which I do not intend to explain in great detail or to adhere to in all its particulars for the limited purposes of the current software description. Strictly speaking, only the inheritance or extension relationship is properly included in a class hierarchy diagram. However the diagrams included here also show the *composition* or *containment* relationship when one class contains instances of another. Furthermore, although provision is made in the diagrams for presenting all the variables and methods of classes these are only shown where particular attention is drawn to them in the text. Prefixes −, + and # indicate the *privacy* status of class variables and methods. Those marked with a '−' are *private* and can only be accessed directly by the containing class itself. Those marked '#' are *package-protected* and can only be accessed by 'friendly' classes in the same package as the containing class. Finally those marked '+' are *public* and can be accessed by any other class in the program. The public variables (which are unusual) and public methods of a class define the ways in which other objects can interact with instances of that class and as such give the best idea of the functionality of a class since they show 'what can be done with it'. The principle behind these privacy levels is called *data-hiding* and is another aspect of the OO paradigm which contributes to its ease of maintenance and use.

Diagrams other than the class hierarchies are presented on an *ad hoc* basis. Note also that a `fixed-width font` is used in this chapter to refer to package, class, variable and method (function) names. Method names are followed by `parentheses()`.

6.2.1 Base classes — the `GraphCA.superClasses` package

Most of the root superclasses of the *graphca* program are found in this package. There are three main sets of classes in the `superClasses` package. The first of these is shown in figure 36. `Vertex`, `Edge` and `Graph` are the base classes for representing graphs and graph elements. Extensions of `Vertex` and `Edge` (`drawableVertex` and `drawableEdge`) add capabilities so that graph elements may be manipulated as graphical images. Most of the code associated with these classes is concerned with the creation, maintenance and analysis of graphs. It is worth noting that `Graph` is an extension of the *Java* `java.util.Observable` class which provides a simple mechanism for updating the

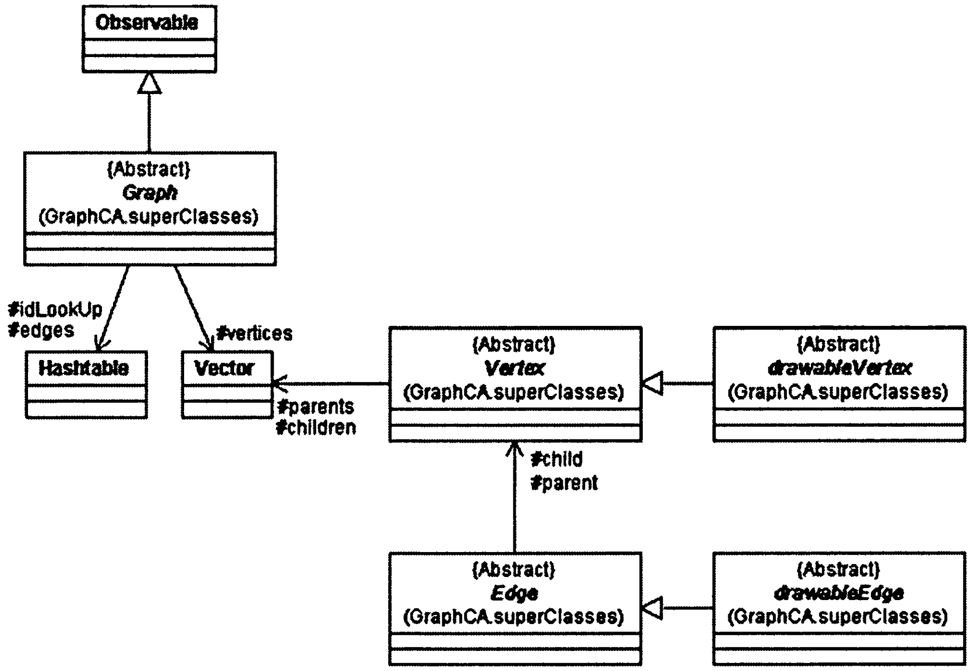


Figure 36 The graph classes in the `GraphCA.superClasses` package.

state of dependent objects (typically GUI components) when the `Observable` object changes.

The relationships between the three classes of object when *graphca* is running are shown in figure 37. When a `Graph` object is first created (from a file by classes in the `GraphCA.gcaIO` package) new `Vertex` objects are instantiated to represent each of the constituent vertices in the graph. These are instantiated with a sequentially allocated identification number, and a (not necessarily numeric) label, and added to the `Graph` object. Two indexes to the constituent `Vertex` objects are maintained by the `Graph`: a simple `Vector` (`vertices`) which stores `Vertex` objects in their sequence number order, and a random-access `Hashtable` (`idLookUp`) which can retrieve `Vertex` objects using their label.

Once all `Vertex` objects have been instantiated, `Edge`s between them are instantiated. `Edge` objects are all considered to be directed so that there is a 'from' `Vertex` (the 'parent') and a 'to' `Vertex` (the 'child'). `Edge` objects consist of two references one to the parent and one to the child `Vertex` which it joins. As `Edge` objects are added, the

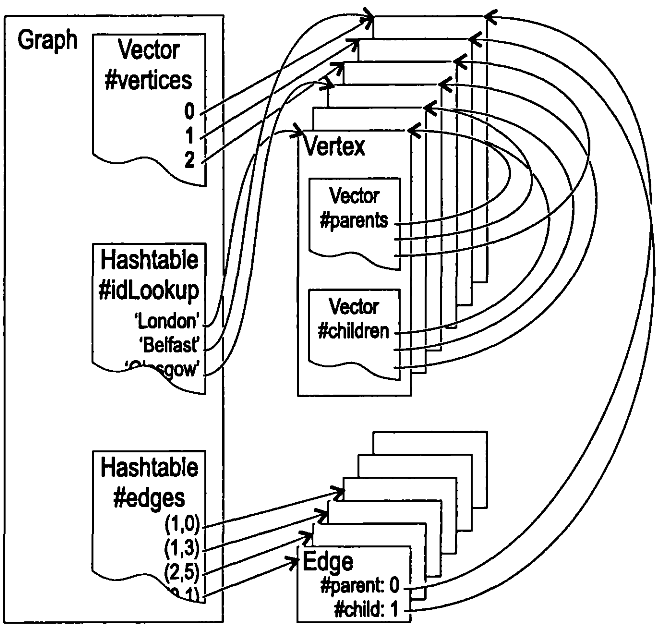


Figure 37 Relationships at run-time between the graph classes.

Vertex objects which they join also build a `Vector` list of references to their respective parent and child `Vertex` objects. The overall data structure is therefore multiply cross-referenced which makes it easy for other objects to access the graph objects in the most suitable way. Note that all of the relevant classes in the `superClasses` package are abstract, and so cannot be instantiated — it is various concrete extensions of these classes which are used in the *graphca* program. These classes are further explained in the context of their various packages.

The *graphca* program GUI is based on extensions of `gcaFrame` and `graphCanvas` (see figure 38). The `gcaFrame` class is a basic dialogue box based on the *Java* class `java.awt.Frame` which adds features such as maintaining a reference to its 'parent' `gcaFrame`, and a `Vector` list of all its 'child' dialogue boxes, so that these may be closed on exit. All of the GUI classes in the `GraphCA.gui` package are extensions of `gcaFrame`. Another basic class for the GUI is `graphCanvas` which can draw a `Graph` object sent to it. The `Graph` object itself maintains a `virtualCanvas` instance variable containing its bounding (geo-)co-ordinates to facilitate correct scaling by its containing `graphCanvas`.

Dynamic behaviour in *graphca* is implemented by extensions of the *Java* language

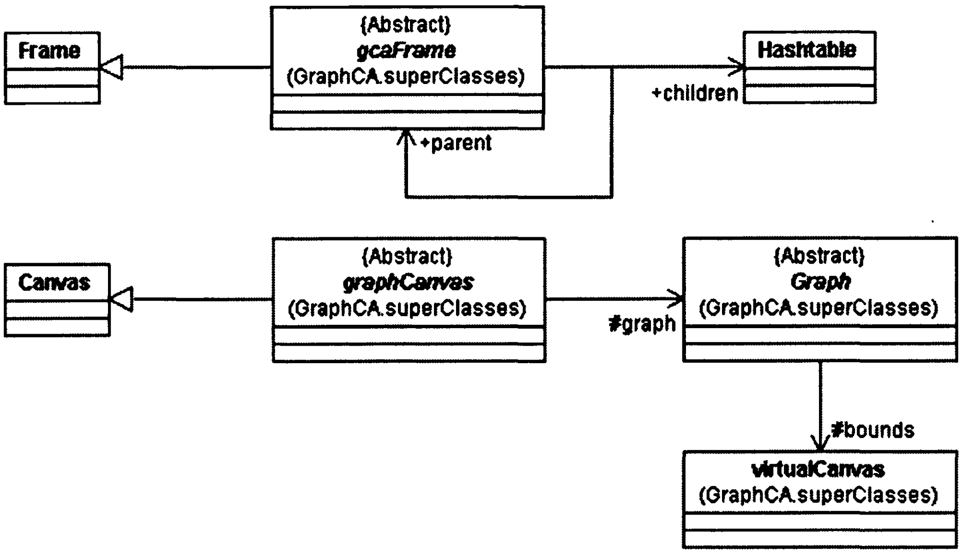


Figure 38 The graphca GUI classes in `GraphCA.superClasses`.

`Thread` class. A program *thread* is a 'thread of execution'. Older programming languages only support single-threaded operation. With the advent of programs with GUIs it has become increasingly important for programming languages to support multi-threaded operation which allows a program to do more than one thing at a time. In a single processor environment (as is still typical) multi-threaded operation is actually an illusion: the processor simply maintains a record of the various tasks it is working on, switching between them according to some priority scheduling process. Whether multiple threads actually operate simultaneously or not, the *Java* `Thread` class provides the means by which programs can be written with multiple logical threads of operation.

In *graphca* multiple threads have been used to ensure that as a graph-CA is updating according to its CA rules, any observing statistical measures or displays are updated before the next step of the CA. The relevant classes are shown in figure 39.

The operation of these classes is illustrated in figure 40. The important data structure here is a `LinkedThreadList` (actually a tree) which organises a set of `LinkedThread` objects into a sequence, such that each `LinkedThread` (except for the 'root') has a predecessor `LinkedThread`, and each also has a list (which may be empty) of successor `LinkedThread` objects.

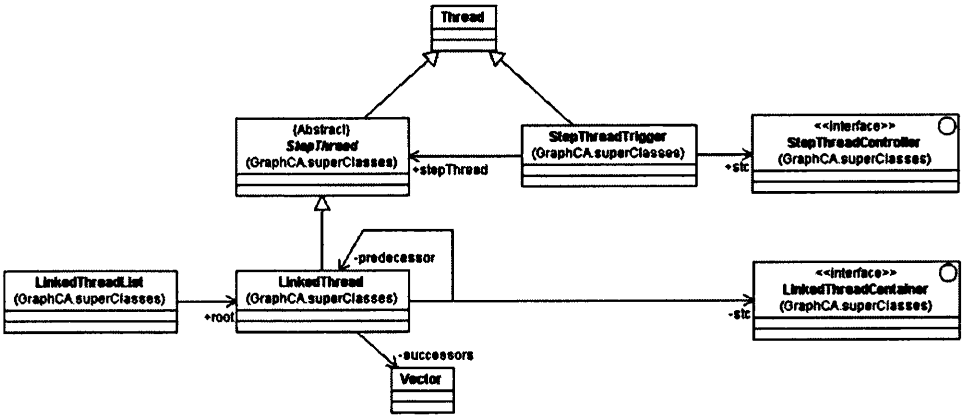


Figure 39 The `GraphCA.superClasses` thread classes.

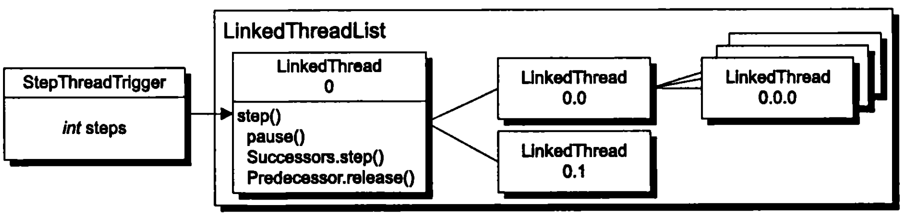


Figure 40 Dynamic behaviour of the *graphca* thread classes.

The root `LinkedThread` (0) has a `StepThreadTrigger` object which is instantiated with a variable `steps` set to the number of iterations required. Each iteration causes the root `LinkedThread` to execute its `step()` method once. The `step()` method executes one iteration of the thread's activity (say updating CA states according to the transition rule), followed by its `pause()` method, which suspends execution, while its successor LinkedThread objects (0.0, 0.1, 0.2. . .) execute their respective `step()` methods. Once all successor `LinkedThread` objects have completed their `step()` method, the `release()` method of the predecessor `LinkedThread` is called to 'unpause' it and allow execution to continue.

This means that the `LinkedThread` objects in the `LinkedThreadList` execute one iteration in 'depth-first' priority i.e. 0, 0.0, 0.0.0,. . . 0.0.k, 0.1.0,. . . 0.1.l, 0.m.0,...0.m.n, and so on. Many objects — in particular `Graph` objects — may contain `LinkedThread` objects linked together in this way. Although the overall effect is not multi-threaded operation (since only one `LinkedThread` is active at any given moment), the update actions associated with the dynamic behaviour of the graph-CA can be made dependent on one another in a controlled sequence using this mechanism.[7]

6.2.2 Graph-CA elements — the `GraphCA.gca` package

One of the most important packages in *graphca* is `GraphCA.gca` which contains the classes for representing a graph-CA model and its constituent locations (vertices) and relations (edges). These classes are `GCA`, `Location` and `Relation`. These are extensions of `Graph`, `drawableVertex`, and `drawableEdge` respectively in the `GraphCA.superClasses` package. These inter-relationships and many of the methods these classes provide are shown in figure 41.

Other classes also in this package support the implementation of a graph-CA model by `GCA`, `Location` and `Relation` objects. A `GCA` object maintains a single instance of `drawingDimensions` which stores parameters such as the diameter of the circles used in the graphic representation of `Location` objects. The `GCA.getDrawingDimensions()`, `Relation.setDimensions()` and `Location.setVertexRadius()` methods provide the required functions. The `gcaStateVector` class is

[7] In retrospect, it is not at all clear that the mechanism described is desirable, or even necessary. It was originally introduced to force GUI elements to update their view of the graph-CA model as quickly as the model itself, so that a clear view of model behaviour could be obtained. The mechanism implemented is not foolproof, and in the subsequent *Gentrification* model development (see part III) it was omitted.

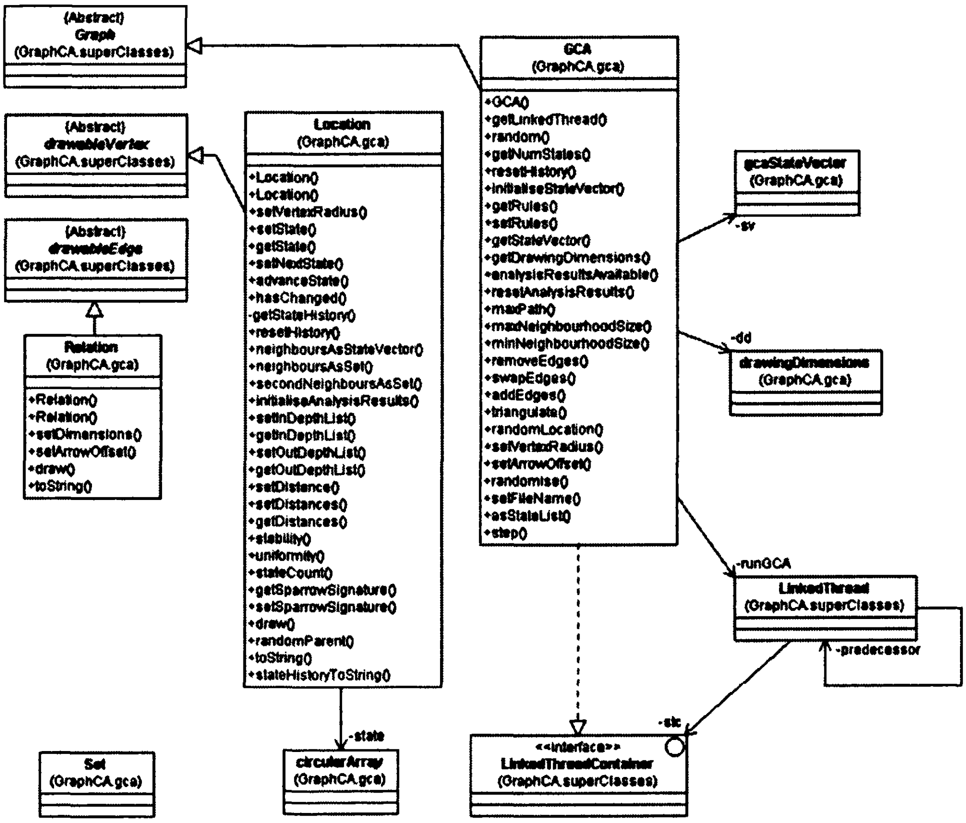


Figure 41 The `GraphCA.gca` package class hierarchy.

used by a `GCA` object as a summary of the current state of the model storing counts of the numbers of cells in different states.

Note that `GCA` implements `LinkedThreadContainer`. The `LinkedThread` contained in a `GCA` object uses the current CA rule set of the graph-CA to execute one time-step of the model. The `step()` method provides the necessary functionality. Other methods of the `GCA` class worthy of note are `randomise()`, `removeEdges()`, `addEdges()`, `swapEdges()` and `triangulate()` which provide many of the basic capabilities required for modifying graph-CA models for investigating structure-process relations.

The `Set` class provides another means of storing and manipulating information about the graph. It is an array of `boolean` objects (true/false flags), which is particularly suitable for performing vertex neighbourhood set union and intersection operations. This is used by the `swapEdges()` method of the `GCA` class, when deforming the graph by neighbourhood preserving edge swaps during structure-process experiments. The `swapEdges()` method also makes use of the `neighboursAsSet()` and `secondNeighboursAsSet()` methods provided by the `Location` class. The process used to select edges for swapping is described in section 5.3.4 on pages 116ff.

`Location` objects use the `circularArray` class to store their most recent states. In addition to making the examination of time series behaviour possible, this provides the option of (limited) processing of the time evolution characteristics of a model by methods such as `stability()`.

6.2.3 CA rule set elements — the `GraphCA.ca` package

The CA-like features of a graph-CA model can be represented by the classes in the `GraphCA.ca` package, shown in figure 42.

This package clearly demonstrates the advantages of re-use in object-oriented designs, since the core classes `RuleSet`, `State` and `Transition` are once again extensions of `Graph`, `drawableVertex`, and `drawableEdge` in `GraphCA.superClasses`. The other classes `criterion` and `StateVector` are used in defining transition rules. A `criterion` consists of a set of upper and lower limits on the fraction of cells in a neighbourhood which must be in each of the allowed states for the condition represented to be satisfied. Each set of limits (upper and lower) is represented by a `StateVector` object which is simply a list of floating point numbers between 0 and 1 — effectively a normalised state count vector. This class has methods (not shown)

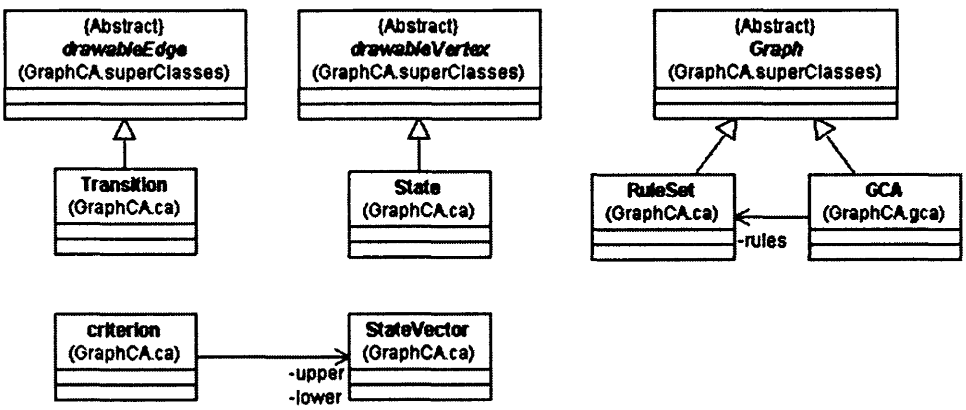


Figure 42 The `GraphCA.ca` package class hierarchy.

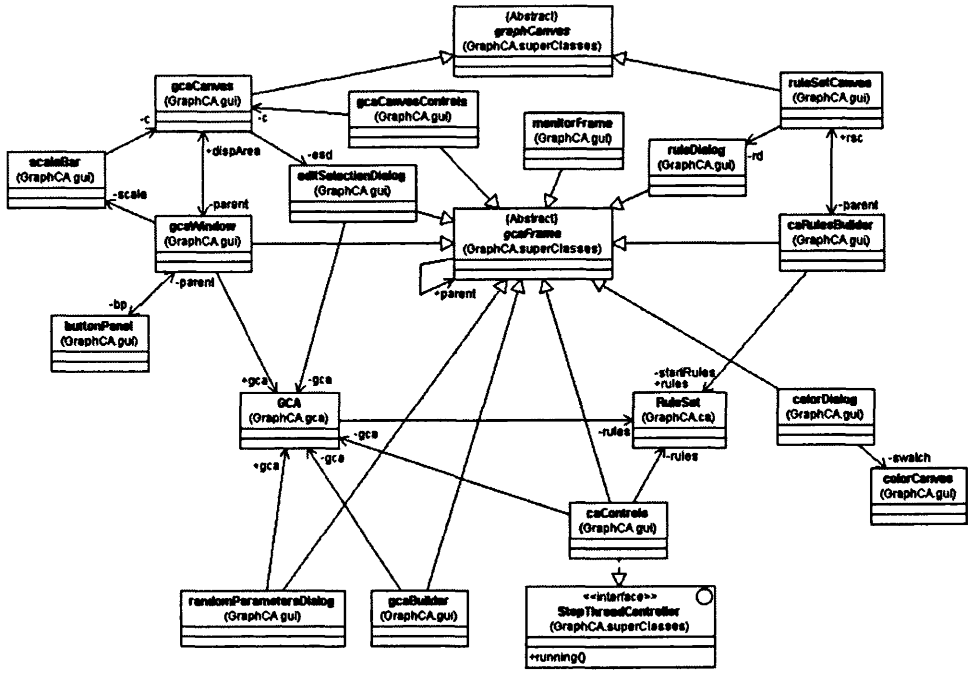


Figure 43 The `GraphCA.gui` package class hierarchy.

`greaterThanEqual()` and `lessThan()` which allow the state of a cell's neighbourhood to be compared to the `StateVector` objects in the current `RuleSet`. The comparison is facilitated by the `neighboursAsStateVector()` method of the `Location` class which converts the neighbourhood of a `Location` object to a `StateVector` in the required form.

Together the `gca` and `ca` packages provide the functionality required to implement a graph-CA model as described in the previous chapter. The remaining packages are concerned with displaying and storing the model, modifying it, analysing its behaviour, and running experiments to investigate its behaviour under modification.

6.2.4 Displaying the model — the `GraphCA.gui` package

This is the largest package in the program, but since it has little or no 'theoretical' content, instead being used mostly for the manipulation and display of graph-CA models in a graphical user interface we do not consider its operation in any great detail. Many of the classes are shown in figure 43.

Again, this package demonstrated the usefulness of OO design in developing reusable program code, since many classes inherit their functionality from the `gcaFrame` class. The most important class is `gcaWindow` which is the main display window where the user interacts with a graph-CA model. The `gcaWindow` object contains the graph-CA model (a `GCA` object), a `gcaCanvas` object on which it is displayed, a `scaleBar` object, and a `buttonPanel`. The latter is a set of buttons which launch various child dialogue boxes for performing various tasks which may be required. The various dialogues available are almost all in the `GraphCA.gui` package too, and include:

- `gcaCanvasControls` presents various controls for moving around the display area of the system, including a zoom facility, and adjustments for the size of vertex representations, which use the `drawingDimensions` object discussed earlier.

- `caRulesBuilder` provides a second `graphCanvas` for viewing and editing the graph-CA rule set together with some additional tools. These include `ruleDialog` and `colorDialog` which allow the rule set's `Transition` and `State` objects to be modified.

- `randomParametersDialog` presents controls for randomising the model's state by randomly allocating new states to all `Locations`. This dialogue uses the `GCA` object's `randomise()` method to perform randomisation.

- `gcaBuilder` is the control for applying the various methods of the model's `GCA` object to randomly add, remove or swap edges in the graph. The system can also have a Delaunay triangulation applied to it using this dialogue box. Triangulation is performed by the classes in the `GraphCA.compGeom` package.

- `caControls` is the dialogue used to start the model running using its current rule set. This class implements a `stepThreadController` so that it can start and stop the `LinkedThread` of the `GCA` which performs model state update according the transition rules.

A common feature of OO systems is illustrated here by the fact that the GUI controls do not themselves implement the processes which cause changes in the various objects representing data structures. Thus, it is the `GCA` class which implements `swapEdges()` and other behaviour, and not the related GUI component `gcaBuilder`.

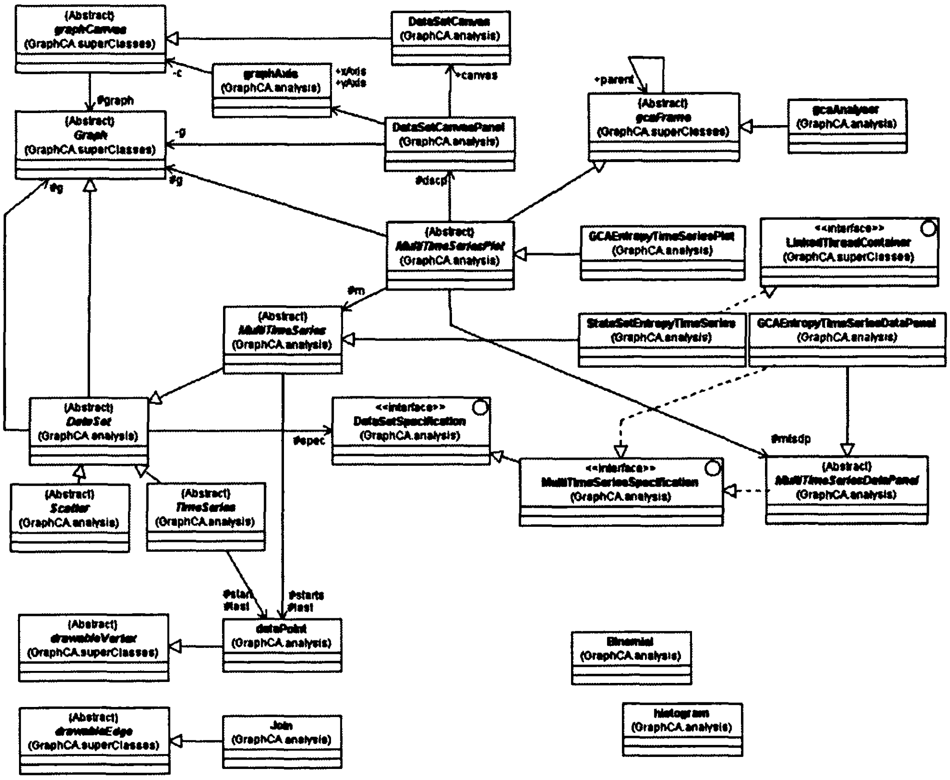


Figure 44 The `GraphCA.analysis` package class hierarchy.

This means that the GUI can be changed without the underlying behaviour of the model being affected. It also means (see below) that these behaviours may be initiated by other classes of object such as those in the `GraphCA.experiment` package which are used to run the program in batch mode. This organisation of the program code also results in substantial economies in coding time by avoiding unnecessary repetition.

6.2.5 Analysing the model — the `GraphCA.analysis` package

Once a model is up and running the next requirement is to analyse its operation. The derivation of various analytical measures has been discussed in section 5.3 (pages 108ff.). In this section the program structures required to provide these measures at run-time are described. The class hierarchy diagram in figure 44 shows many of the classes involved.

The most significant aspect of this package is that there is extensive re-use of classes from `GraphCA.superClasses`. Various dialogue box tools, particularly `gcaAnalyser`, are extensions of `gcaFrame`, whereas `DataSet`, `dataPoint` and `Join` classes used in the representation of various sorts of data plot are extensions of `Graph`, `drawableVertex` and `drawableEdge` respectively. In this case, the class interrelationships are rather more complex because various abstract extensions of the `DataSet` class are provided for different kinds of plot (`Scatter`, `TimeSeries`, and `MultiTimeSeries`). Each of these (only `MultiTimeSeries` is shown in detail, for clarity) can be extended to provide some measure of the state of the target `Graph`. The concrete extension of `DataSet` updates according to the programming of its `step()` method which it implements as a `LinkedThreadContainer`.

The display of the current state of a `DataSet` object is handled by extensions of `gcaFrame`, such as the `MultiTimeSeriesPlot` extension `GCAEntropyTimeSeriesPlot`. This class contains a `DataSetCanvasPanel`, and an appropriate `DataSetSpecification`. The `DataSetCanvasPanel` contains a `DataSetCanvas`, on which the data is actually displayed, and two `GraphAxis` objects for its x and y axes. The `DataSetSpecification` object supplies the `DataSet` object with any necessary parameters to define its behaviour. To make this clearer figure 45 shows these classes in use in a `GCAEntropyTimeSeriesPlot` object.[8]

This seemingly convoluted approach means that it is possible to quickly program additional components for the display of totally different measures of the current model state. The coding of certain components as interfaces means that it is also possible to generate these measures in a batch mode of operation as is done in the `GraphCA.experiment` package (see the next section). The overall structure of the analysis tools code is similar to the *Swarm* system[9] developed at the Santa Fe Institute (Minar, Burkhart, Langton & Askenazi 1996) for building multi-agent simulations, wherein all program objects are 'swarms' including 'observer swarms' which present data to the experimenter. In *graphca* many objects are `Graph` objects including the analysis tool `DataSet` objects.

Analysis tools provided by this package include:

- *Shortest path analysis of the graph.* The shortest path from every vertex to every other can be determined. Since calculation is relatively slow, results are stored

[8] Note that many of these class names reflect earlier thinking on the spatial information measure.
[9] Available for download from `http://www.swarm.org`.

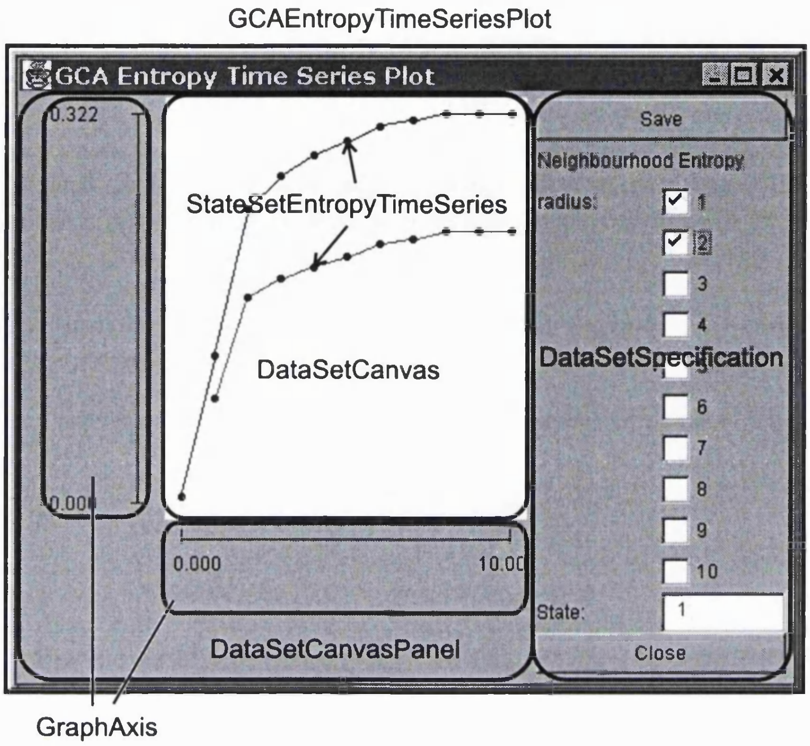


Figure 45 A typical plot in use: the `GCAEntropyTimeSeriesPlot`. This figure also shows the different components which make up a typical GUI display in the `analysis` package. This is an example of a `MultiTimeSeriesPlot` made up of the various elements shown.

in the `GCA` object itself as `histogram` objects. The `histogram` class stores a set of numbers as a frequency distribution. This allows rapid presentation of various statistics concerning shortest paths, such as mean shortest path, numbers of neighbours within some distance, and total distance in the graph. Space syntax normalisation of path lengths (see equation 4.1 on page 81 and the accompanying discussion) is also supported, together with radius-limited statistics.[10]

- *Sparrow's (1993) algorithm* for the determination of graph equivalence classes is also supported (see section 3.2.3 on page 53). Sparrow's method is particularly easily implemented since it involves repeated modification of the values at each vertex in the graph using values stored at neighbouring vertices.

[10]This proved instrumental to the raising of some perplexing questions about that the 'syntax-normalisation' touched on in the discussion in chapter 4.

- *Limited measures of the local behaviour of the state* at a particular location are supported. The two measures supported are (i) 'stability' defined as that fraction of previous time steps where the state at a location did not change, and (ii) 'uniformity' which is the fraction of neighbours of a central location currently in the same state as the central location.

Together these three types of measure are used to construct scatter plots showing the overall current state of the model and pointing to possible relationships between measures. Figure 46 shows a typical scatter plot which can be produced. Here the stability of locations is plotted against the 'degree radius-3' (the number of locations within 3 edges of a location). This feature of the program has not been used to any great extent but illustrates the potential for investigating these sorts of local relations in the graph-CA framework, as discussed in section 5.2.1 on page 105.

A further analytic measure provide by the `analysis` package is:

- The *spatial information* measure described in detail in chapter 5. Figure 45 illustrates this measure in use, and also shows that provision has been made for its calculation up to a radius (or 'lag') of 10 edges.

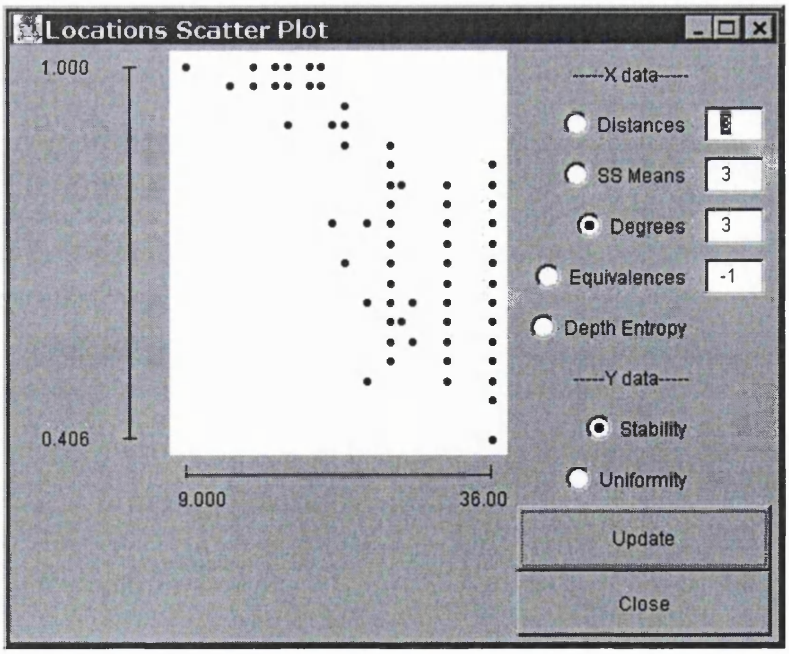


Figure 46 Typical scatter plot produced by the analysis tools.

6.2.6 Running a model many times — the `GraphCA.experiment` package

The `GraphCA.experiment` package (figure 47) supports the running of graph-CA models in a batch mode whereby multiple runs of the same model with different starting configurations can be set up, and the results stored to disk for subsequent analysis. This capability has been extensively used to generate the experimental results reported in chapter 7.

The easiest way to appreciate the `GraphCA.experiment` package's functionality is to examine the `ExperimentDialog` object shown in figure 48. Here the experimenter specifies the graph-CA to use ('`life.gca`' in the example), and any graph deformation to be performed between consecutive runs of the model. In the case shown, 10 edge-pair swaps are to be performed, and this is to be repeated 10 times, so that 11 different sets of results will be produced in total. The dialogue also allows the experimenter to either select previously stored starting configurations, or to create randomly generated new ones. Each of these will be run on each model for the number of time steps

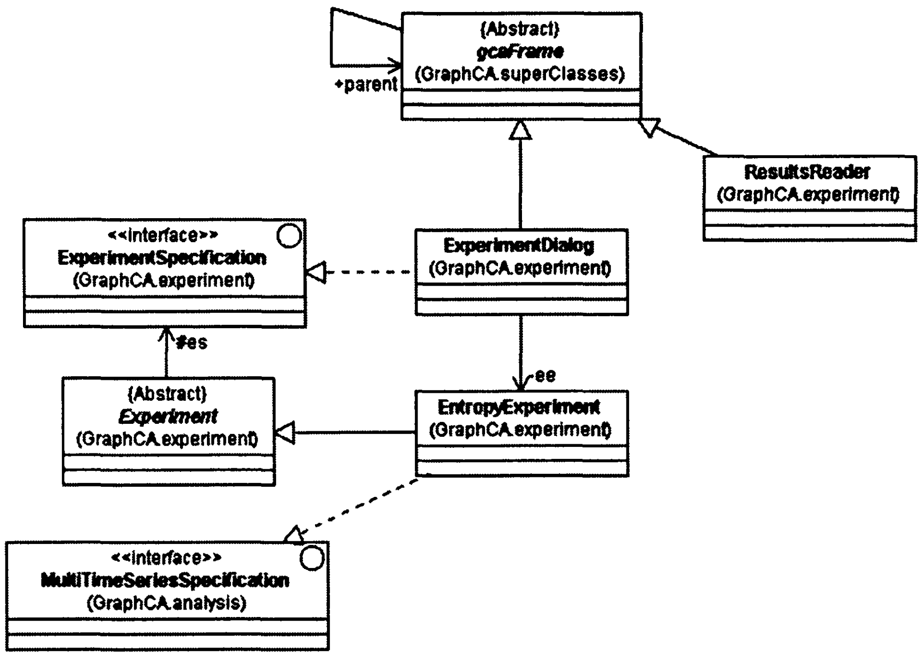


Figure 47 The `GraphCA.experiment` package class hierarchy.

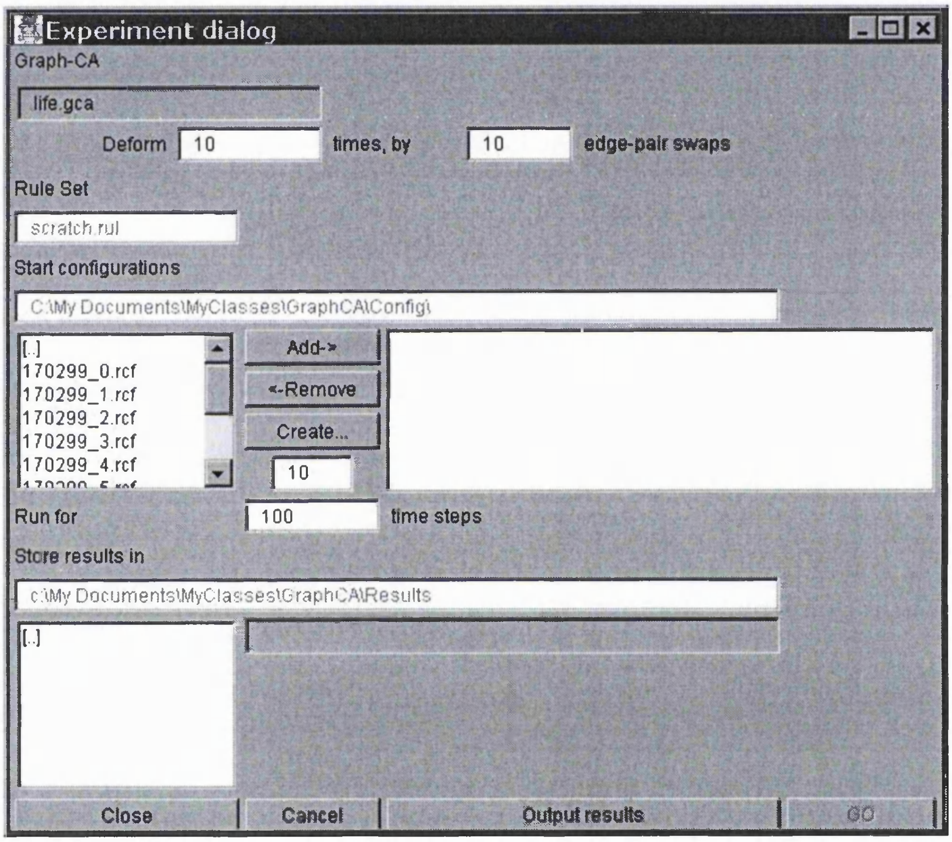


Figure 48 The `ExperimentDialog` class in use.

specified, and the results stored in files at the location also specified in the dialogue box. The resulting files can subsequently be read using a `ResultsReader` object which converts the results into a simple text format suitable for further analysis using other tools.

6.3 Conclusions

This chapter has reviewed the *graphca* program which has been used as the basis of the work reported in the remainder of this thesis. The suitability of the object-oriented programming environment to this sort of application has been emphasised. The generality of the graph representation has been further demonstrated by its re-occurrence throughout the software description, not only as the basic component of the graph-CA model itself, but for representing CA rule sets and statistical data sets.

As noted at the outset, *graphca* has undergone evolution throughout the course of this research, and enhancements and additions relevant to particular pieces of work are described alongside those results, in subsequent chapters.

Chapter 7

Exploring the structure-process relation using graph-CA models

Some investigations carried out along the lines proposed in chapter 5 using the *graphca* software described in the last chapter are now reported.

In section 7.1 some of the properties of the cell-space in which most of the experiments reported in this chapter were run are described. The effect of edge pair swap deformations on this cell space is also examined and described. This is an essential precursor to understanding the scope and meaning of changes in the dynamic behaviour of CA processes running in the cell-space.

An investigation of the effect of changing the spatial or relational structure of a grid on the dynamics of two graph-CA processes occupies most of the remainder of this chapter. The processes examined are a segregation rule and the Game of Life CA already encountered in chapter 5. These processes appear to respond differently to changes in structure and that becomes a major focus of the work reported here. The Some further brief experiments on very different structures are reported in section 7.4. The chapter concludes with a brief discussion of the major findings.

7.1 The cell-space

The main examples in this chapter are based on GCA-space trajectories originating with a 20 × 20 (i.e. 400) vertex 'toroidal grid' graph, in which each vertex has as its neighbours itself, and its four orthogonally and four diagonally nearest neighbours.

Vertices along the 'edges' of this grid are connected to vertices on the opposite edge so that all vertices are identical. Thus the structure is that of a Moore neighbourhood on a torus. The toroidal structure of the graph ensures that all vertices have the same number of neighbours and are therefore identical. This means that the transition rule specification is independent of neighbourhood size — since only one neighbourhood size is present.

This toroidal property also makes it difficult to appreciate the structure of the graph when it is presented as a 2-dimensional grid. Figure 49 is an illustration of the graph represented in three dimensions. The most important points to note about this structure are (i) its uniformity — every vertex is like every other — and (ii) when presented as a two-dimensional grid, apparently remote and separate clusters along 'opposite' edges or in 'opposite' corners are actually single clusters.

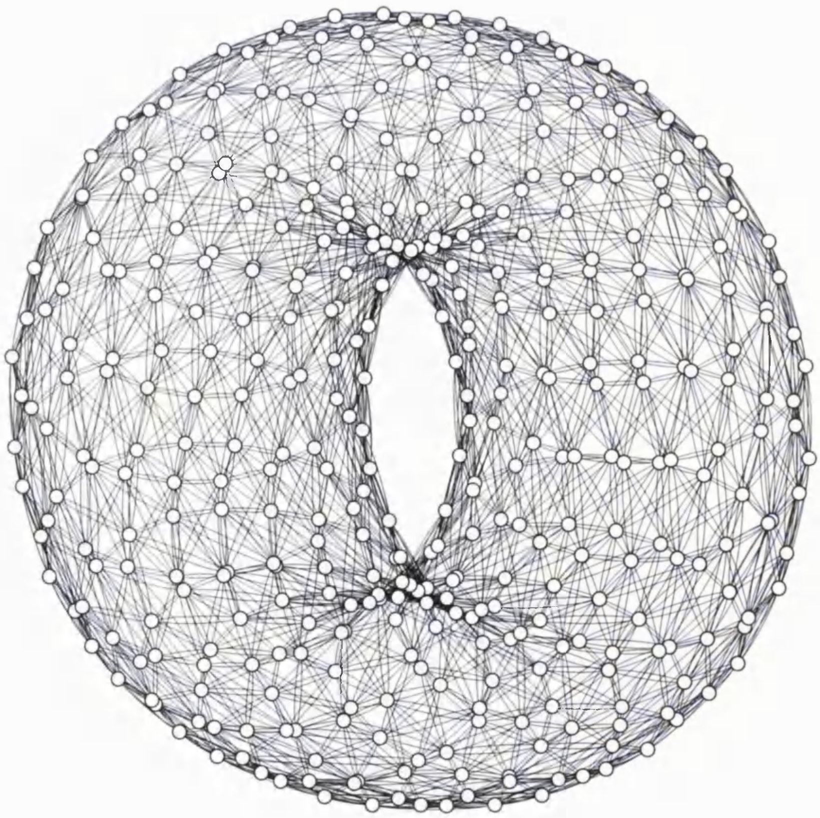


Figure 49 The toroidal grid graph represented in three-dimensional space.

7.1.1 The effects of edge pair swap deformation on graph structure

The structure-process experiments described in this chapter make use of the edge pair swap deformation described previously (see figure 26 on page 117 and the supporting text). As previously discussed the main motivation for using this deformation is to preserve the distribution of vertex neighbourhood sizes so that the dynamic process defined by the CA transition rules is strictly comparable between GCAs along any give trajectory. However, as also noted, the effect of such a restrictive definition of graph deformation is to restrict GCA-space trajectories to a limited subset of those graphs of any given size which are possible. Insight into this restriction is provided by tracking some of the graph's structural properties along a trajectory. This has been done in figure 50 for 10 GCA-space trajectories consisting of ten 10 edge pair swaps, followed by nine 100 edge pair swaps, up to a total of 1000 edge pair swaps. Note that results for *all* 10 trajectories are presented in this graph, but are so similar that they over-print. Graph structure is tracked using the small world measures characteristic path length and clustering coefficient described in section 3.2.4 on pages 54ff.

Two aspects are worth emphasising here. First, both of these measures change at very similar rates, and second, as a result, only a particular kind of graph is generated by this process. No particular vertex location — or small set of vertex locations — is highly distinct or dominant. The most central vertices are such, not by virtue of commanding strongly dominant locations in the graph structure, but by virtue of

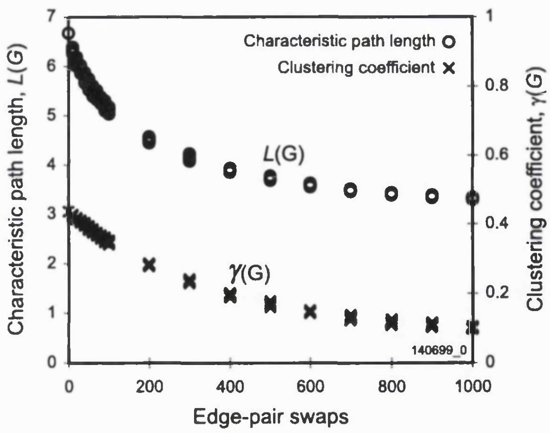


Figure 50 GCA-space trajectories in terms of characteristic path length and clustering coefficient.

the fact that their neighbourhoods are less cliquish, so that they tend to have larger numbers of vertices in their 'higher-order' neighbourhoods (i.e. at distances of 2, 3 or more edges). Thus, strongly strategic locations, such as may exist in a small world network do not occur in these graphs.

This is demonstrated in figure 51, where the development of structure in the toroidal grid graph is shown as edge pair deformations are introduced. Darker vertices in the left hand column are most central; the less 'cliquish' vertices are darker in the right hand column. There appears to be a weak relationship between the two factors, so that regions with low characteristic path lengths (the darker cells in the left hand cases), are often associated with less cliquish vertices (the darker cells in the right hand cases). Note that the shading in this figure is relative, not absolute: as the graph becomes more deformed overall, path lengths become shorter and cliquishness is reduced as already seen in figure 50.

Another aspect of the deformation worthy of note is that it has most impact on the graph structure over the initial deformations. Indeed, beyond 500 edge pair swaps relatively little further change occurs in the graph structure as described by these two measures, and after 1000 edge-pair swaps the graph is almost 'random' in its structure, so that the spatial adjacency of two vertices does not mean that they are highly likely to be joined.

We now examine the impact of such deformation of the structure of the toroidal grid on the spatial behaviour of two CA processes already familiar from earlier chapters — the segregation CA and the Game of Life CA.

7.2 Effects of deformations of the grid on the segregation graph-CA

7.2.1 Limited deformations of the grid

Recall that the segregation CA rule is simply described: each vertex adopts whichever state is in the majority in its neighbourhood at the next time step. This usually results in a system segregating into distinct regions of different states (a typical outcome is presented in the middle case in figure 28 on page 126). Occasionally, the system will evolve towards a homogeneous state where all cells are in the same state. However, this is a rare occurrence when the initial system configuration contains similar num-

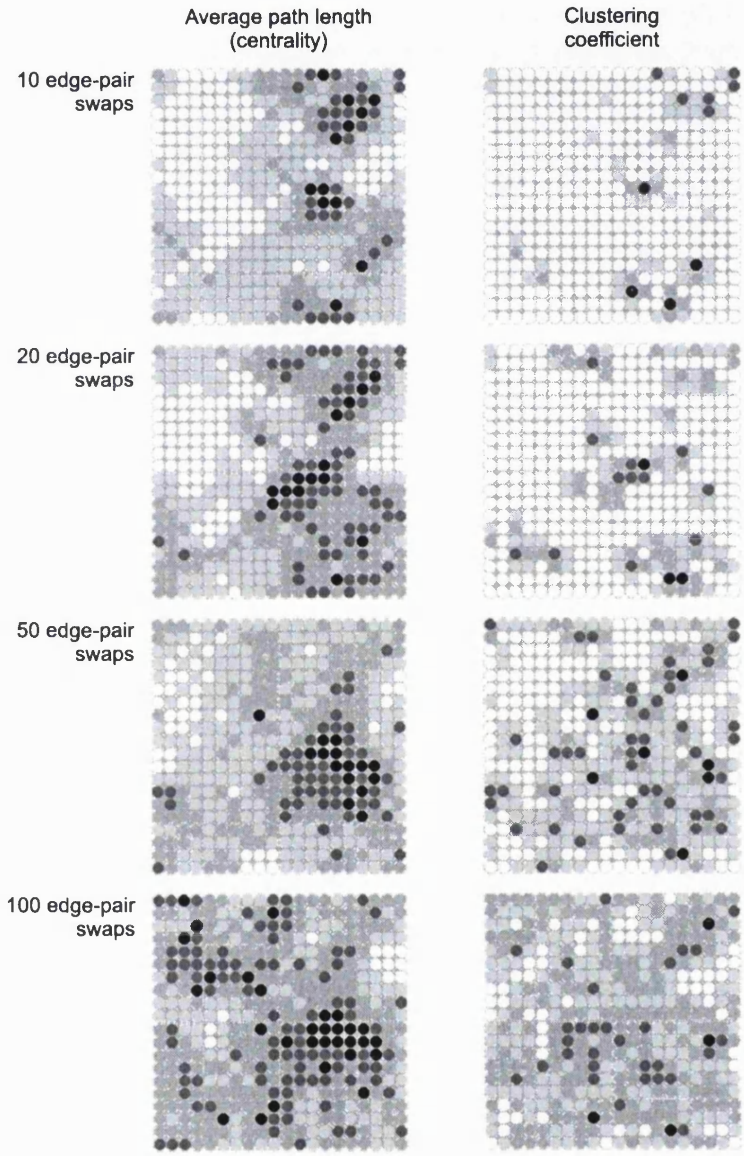


Figure 51 Changes in centrality and clustering coefficient under edge pair swap deformations.

bers of cells in each of the two allowed states, and the system usually segregates. The toroidal nature of the grid usually ensures that the resulting system configuration consists of just two 'super-regions' of cells in similar states, because small 'islands' of cells not in the same state as their surroundings are usually absorbed by those surroundings.

In figure 52 the effect of a series of 10 edge pair swap deformations on the segregation CA is shown. The 11 spatial information time-series shown are averages at each time step, based on a set of 20 different randomly generated starting configurations in which about 50% of vertices are in each of the two allowed states. In all cases, starting from the low spatial information associated with random configurations, there is a rapid increase in spatial information as the system segregates into distinct regions of vertices in one state only. The sort of dynamic behaviour observed, as described by the spatial information time-series, is similar, in spite of the deformation of the graph, although the final stable value of spatial information attained in each case is different. The evolution of the system in all cases is towards an atypical 'unexpected' configuration with high spatial information. This indicates that the general 'segregation' behaviour of this transition rule is *robust* under these sorts of deformation of the underlying graph.

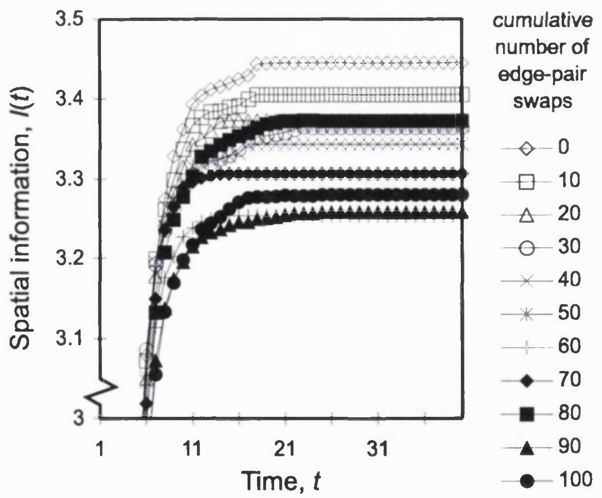


Figure 52 The effect of a series of 10 edge-swap deformations, up to 100 swaps, on the spatial information time-series for a segregation GCA.

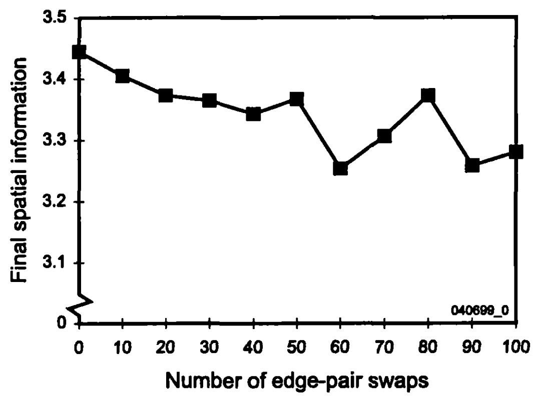


Figure 53 Summary diagram showing effect of deformation on the final value of spatial information attained by the segregation GCA.

It is useful to plot the trend in the final stable value of the spatial information attained, against the number of edge swaps made. This has been done in figure 53. There certainly appears to be a downward trend in the final value of spatial information attained by these GCA as the graph becomes more disordered due to edge pair swaps. However, the trend is not particularly strong — note particularly the false origin on the vertical axis of this figure. Over these 100 edge pair swap deformations, the fall in final spatial information is no more than about 5%. The trend is also not monotonic — in four cases of ten, (from 40 to 50, 60 to 70, 70 to 80 and 90 to 100 swaps) there is a reverse in the trend.

In figure 54 10 more GCA-space trajectories with the same characteristics are shown. Again, the 400 vertex toroidal graph has been deformed 10 times by 10 edge pair swaps, and these results have been compiled for the same 20 different starting configurations as before. There is a downward trend, but it is not particularly marked, and the first case considered above is fairly typical. Over all the trajectories shown here the fall in the final spatial information value attained is never more than about 10%. Furthermore, for deformations up to about 30 edge pair swaps, some trajectories actually result in final spatial information values a little above the initial case. It seems clear, therefore, that the behaviour of the segregation GCA is quite robust under this sort of deformation of the space it runs in. The easily observed fact of the system dynamics continuing to produce clearly segregated arrangements of cells

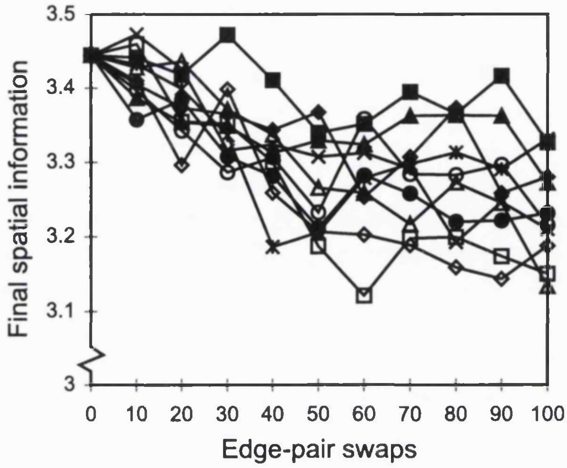


Figure 54 A number of GCA-space trajectories for the segregation transition rule.

in the two allowed states, as the lattice structure is deformed, is borne out by these spatial information time-series results.

This result is, perhaps, not surprising, since the segregation transition rule is not inherently spatial, in the sense that it is not sensitive to orientation or directionality, depending only on simple local properties of the system. The *local* properties of the toroidal grid are (largely) preserved up to this point in the trajectories produced by limited deformations.

7.2.2 Larger deformations of the grid

If we now deform the regular grid lattice for a segregation GCA considerably further than in the preceding section we can carry out analyses to see what happens to its dynamic behaviour. Results for continued deformation (up to 1000 edge pair swaps) of the same GCA-trajectories as in figure 54, are presented in figure 55. This shows that there is a continued downward trend in the final spatial information attained, a trend which accelerates significantly beyond about 300 edge pair swaps. This apparent change requires further investigation — after all, as we saw in figure 50, the most rapid change in the properties of the graph seems to occur well before this point.

The first point to note is that there is actually a change in the *qualitative* behaviour of the segregation GCA in these more severely deformed cases. This consists in an in-

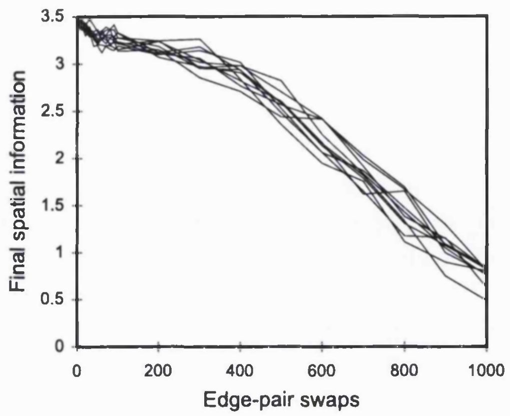


Figure 55 Further deformation of the segregation GCA cell space.

creasing number of starting configurations evolving to the homogeneous state, where all cells are in just one of the two allowed states. This occurs even with the initial system configurations considered here which have approximately equal numbers of cells in the two allowed states. A homogeneous system configuration has a spatial information value of 0, since there is nothing unexpected about the distribution of neighbourhood states (because only one distribution is possible). As an increasing proportion of starting configurations evolve to a homogeneous state, the average spatial information observed therefore falls.

Put another way, the system's dynamics no longer represent Wolfram *class 2* behaviour, but *class 1* behaviour. In fact, this draws attention to the weakness of Wolfram's classification scheme, because it shows that the observed behavioural class of a system is dependent on the starting configuration, and can thus only be described in statistical terms. That is, a particular GCA *tends* to be class 1 (or 2, 3, or 4), not a particular GCA *is* class 1 (or 2, 3, or 4).

The first case among these GCA trajectories where the tendency to produce homogeneity occurs is after 500 edge pair swaps. The spatial information time-series data for this graph and its 20 starting configurations are shown in figure 56. This should be compared to figure 29 on page 128. It is notable that the *range* of final spatial information values produced by the system is now considerably greater. However, the increase in range is all at the lower end — since the highest values of spatial information attained by any of the examples shown are very similar, at a around 4.5.

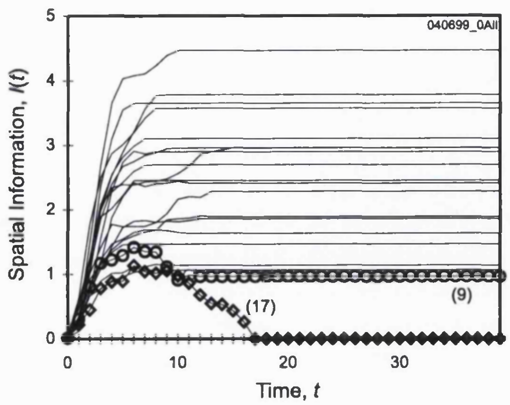


Figure 56 The 500 edge pair swapped graph time-series behaviour.

Some configurations now lead to considerably lower values, well below the minimum of around 2.5 in figure 29, and one configuration sees its spatial information time-series initially rise, but end up falling back to zero — indicating a homogeneous final state.

The development of the particular starting configuration (17, represented by ◇ markers in figure 56), which exhibits this class 1 behaviour can be seen in figure 57. Interestingly, another configuration (9, represented by the ○ markers in figure 56) shows signs of tending towards this behaviour, and this is also shown in figure 57. Configuration 9 stabilises with a single small isolated group of cells in the minority state (remember that the system is toroidal). In subsequent GCAs along this same trajectory (600, 700 edge pair swaps and so on), this small group becomes unstable and configuration 9's behaviour 'decays' to class 1.

It is difficult to generalise about the way in which the segregation GCA's behaviour changes as extensive deformation of the regular grid occurs. It seems that as deformation increases, it is unusual for a particular starting configuration to 'recover' from a decay to class 1 behaviour back to class 2 behaviour (the dominant behaviour on the regular grid). As a result, an increasing proportion of starting configurations lead to only a temporary increase in spatial information before 'decaying' to a state where all cells are in one or other of the two states, or to a state where only small clusters of the minority state are to be found. This is still 'segregation' behaviour of a sort, but it is distinctly different from the regular lattice where most random start-

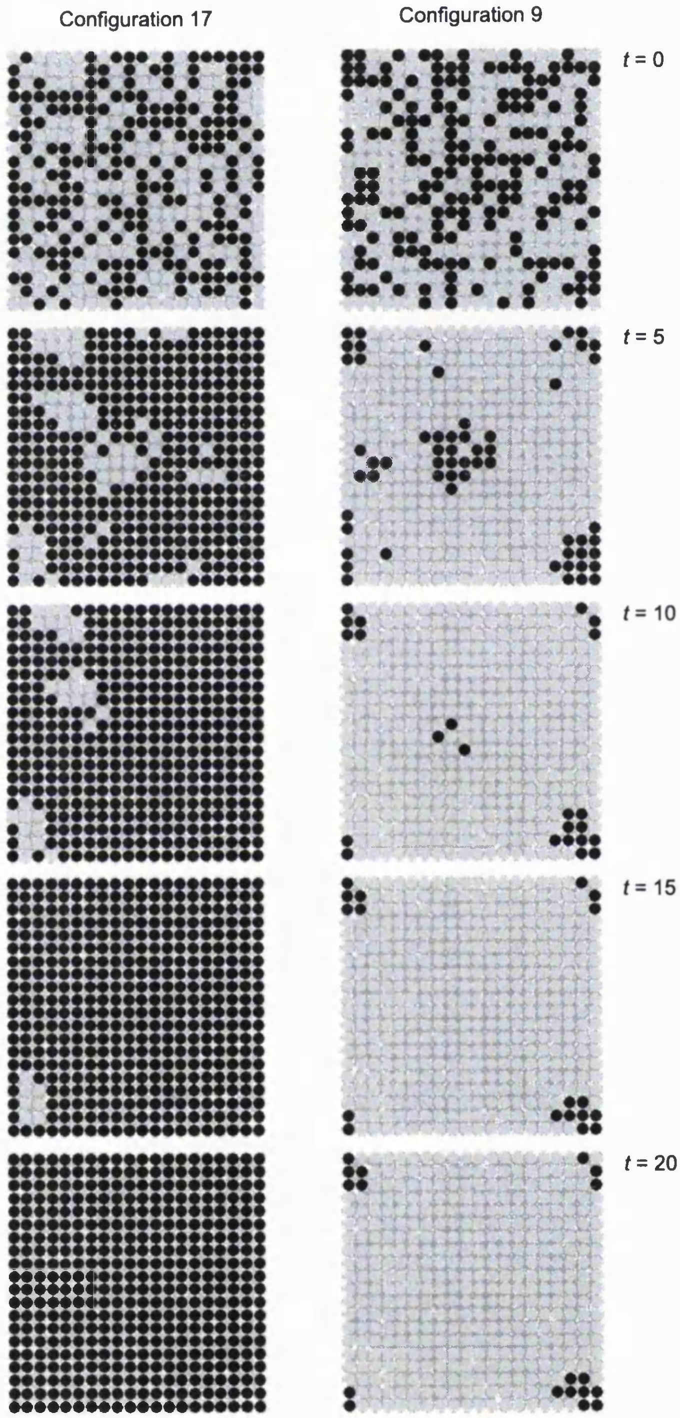


Figure 57 The evolution of configurations 17 and 9 on the 500 edge pair swapped graph.

ing configurations stabilise in a state with somewhere around half of the cells in each of the two allowed states. This suggests that the structure of deformed grid graphs imposes itself on the spatial outcomes which result, so that whichever state happens to be in the majority in 'dominant' regions of the structured graph ends up filling the whole system after a few time steps.

Alternatively, it suggests that only relatively large uniform regions are stable on the irregular structure of the deformed grid. These are unlikely to occur in the initial random configurations. More importantly, two large stable regions in *opposite* states are unlikely to occur in tandem, so that whichever cell state happens to be in the majority after a few time steps becomes more likely to rapidly occupy the whole system — since no stable regions of the other state are likely to exist. This contrasts strongly with behaviour in the regular grid, where any relatively compact homogeneous region with no sharp corners or elongated 'outgrowths' is stable. This means that it is possible for multiple distinct, stable regions to establish themselves in the regular grid, and to grow 'around one another', until large regions in each state are established with a smooth boundary (i.e. no corners) between them. Thus the overall change in the segregation GCA's behaviour can be considered as a result not of any obvious change in the local spatial properties of the system, but of its global properties — the properties of interconnectedness which the system's spatial structure establishes between regions in the system.

7.3 Effect of deformations of the grid on the Game of Life graph-CA

It is instructive to examine the effect of the same deformations on another GCA rule. The Game of Life GCA exhibits strong sensitivity to even limited deformation of its cell space.

As was discussed in section 5.3.6 (pages 127ff.), it is probably not valid to use a simple average of the system's spatial information to summarise the Life GCA's behaviour. Nevertheless, to establish that there is *something* going on, figure 58 shows this data for the same 20 starting configurations, for the same GCA-space trajectory as in figure 53. This suggests that there is a change in the dynamic behaviour of this GCA as the cell space is deformed. On the regular grid (the backmost time-series in

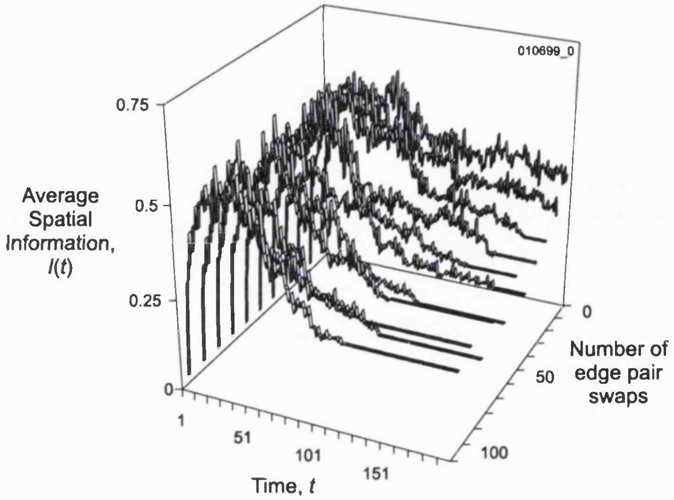


Figure 58 Average spatial information for the Life GCA under deformation.

the figure) there is evidence of sustained activity up to (and beyond) 200 time steps. After only 100 edge pair swaps (the front time-series) it appears that system activity continues for only 100 time steps or so.

This change can be characterised as a reduction in the *transient time* during which the Life CA rule raises spatial information above the level associated with initial random arrangements, before stabilising at some non-zero value. The meaning of this transient time is clearer in figure 59, where three distinctly different cases are shown with transient lengths of 200+ time steps, approximately 100 time steps, and approximately 50 time steps.

Now, if we compile frequency counts of the transient times for this GCA as the cell space deformation is increased, the effect of deformation on the Life CA is clearer. This is shown in figure 60. It is clear that in the case of the regular cell space (the front histogram) a wide range of different transient lengths is observed, with many greater than 200 time steps. As the cell space is progressively deformed (histograms nearer the back of the figure) the tendency is for the transient length to be reduced. There is one extreme case after 10 edge-pair swaps (with a transient length of 524 time steps), but the general trend towards shorter transient lengths is clear. A box-plot of average, median, and inter-quartile ranges for the distribution of transient times, against the number of edge pair swaps, for this set of 20 starting configurations (see figure 61)

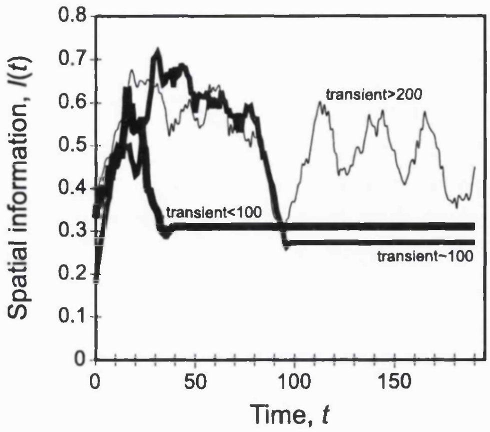


Figure 59 Spatial information time-series with three different transient times. The transient time is a measure of the 'vitality' of a system's behaviour since it records the number of time steps during which activity is still apparent in the spatial information time-series. In the figure transient times of over 200, around 100, and less than 50 time steps can be seen.

makes the trend clear. The change in transient times manifests itself as a reduction in the likelihood of configurations with very long transient times occurring. This is especially clear from the limited effect which deformation has on the recorded *median* transient time as opposed to the *mean*. The implication is that the overall dynamic complexity of this graph-CA process is reduced as the cell-space in which it runs becomes less ordered.

Again, we can present a tentative explanation of this observation. As has already been indicated, an important pattern which frequently occurs in the Life CA is the 'glider' (see figure 10 on page 58). Observation of the behaviour of the Life CA reveals that the glider is significant to the vitality of particular configurations. Often, the CA will settle into a relatively stable state where a number of isolated regions consist of 'bubbling' patterns of cells, but most of the system consists of uniform 'dead' regions. In these situations evolution usually proceeds in one of two ways: either

- the system stabilises to a fixed pattern, or
- one of the bubbling regions throws out a glider which 'crashes' into another bubbling region 'reactivating' it, and increasing the overall level of activity in

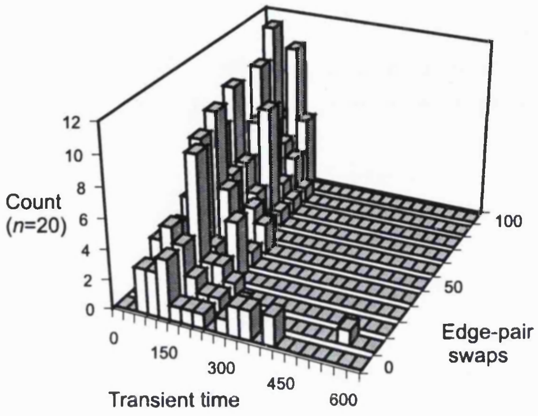


Figure 60 Effect of deformation on the distribution of transient times for the Life GCA.

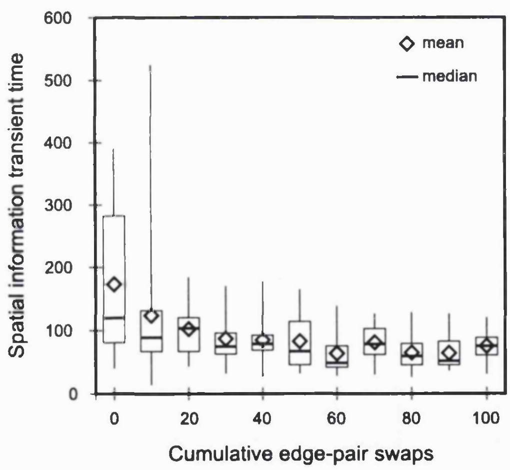


Figure 61 Summary statistics for changes in transient time as the Life GCA cell space is deformed.

the system.

The second of these alternatives makes a significant contribution to the overall vitality of any particular configuration — and thus to its transient time. If we consider how gliders are likely to be affected by deformations in the lattice in which they occur, it is clear that their regular diagonal progress is very likely to be disrupted by deformations of the lattice. Observation of isolated gliders on deformed lattices confirms that this is the case.

Although it is conceivable that with a small number of lattice irregularities some configurations will still exhibit long transients (because their particular evolution of the system does not 'run into' the irregularities), in general, as the lattice becomes more irregular, most configurations will stabilise more quickly. This is because the irregularities in the lattice prevent the consistent behaviour of structures such as gliders from sustaining the complex global behaviour over long time periods. It is possible that some very specific deformations of the lattice will cause gliders (say) to 'explode' into multiple fragments which actually contribute to the system vitality. However the required deformations are likely to be very specific, and unlikely to less common than those which simply stop gliders in their tracks.

7.4 Two further brief studies

In this section we briefly examine, in a qualitative way, the behaviour of the segregation CA transition rule on dramatically different spatial structures, rather than on progressively more deformed grid-based structures. Note that we have not transferred the Life CA to substantially different structures because it is non-trivial to do so. This is for reasons already touched on in section 5.3.3 (pages 113ff.), associated with defining an equivalent rule set for different neighbourhood size distributions. 'Similar' rules — birth and survival in medium density neighbourhoods, death otherwise — do not necessarily produce behaviour anything like as rich as the Game of Life. The segregation CA rule is easily transferred to other contexts, since a simple majority decision can still be applied. This foreshadows the work reported in part III, and chapter 11 in particular, where various spatial structures derived from a real urban fragment are examined. It also serves to emphasise the major point arising from the results already reported, that there may be dramatic differences in the sensitivity

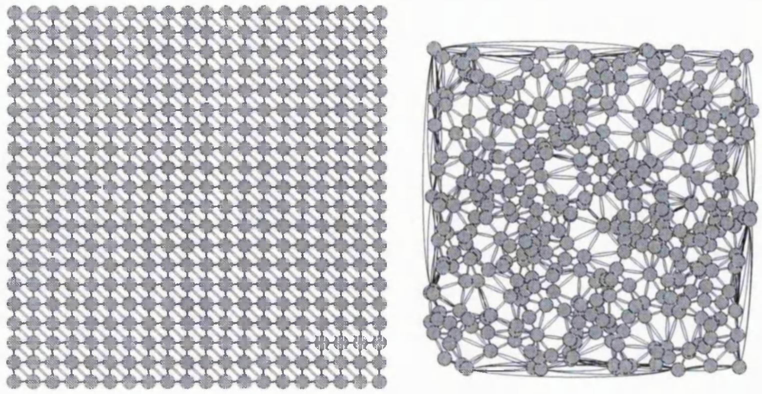


Figure 62 The two graphs examined in this section. The left hand example is a Delaunay triangulation of a grid-located set of 400 vertices. The right hand example is a Delaunay triangulation of 400 randomly located vertices.

of different process rules to spatial structure.

The observations in this section are based on two spatial structures. These are shown in figure 62.

First is a *Delaunay triangulation* of the 20×20 grid-located vertices of the previous sections. This graph is different from the toroidal grid in two inter-related and significant ways. First, it is not toroidal — so that it has 'edges'. Second, arising from this, vertices along the edges of the system and in the corners may have different numbers of neighbours from those in the centre. Vertices in the middle region of the system have 6 neighbours, those along edges only 4, and corner vertices may have either 2 or 3 neighbours.

Second is a *Delaunay triangulation of 400 randomly located vertices.* The range of variation in this structure is greater still, with some neighbourhoods as small as three other vertices, and others with perhaps 9 or 10. This structure is also not toroidal.

A typical outcome of the segregation rule from a random starting configuration on the Delaunay grid graph is shown in figure 63. Such outcomes also exhibit significantly higher spatial information than random configurations and the associated spatial information time-series are very similar to those we have already encountered

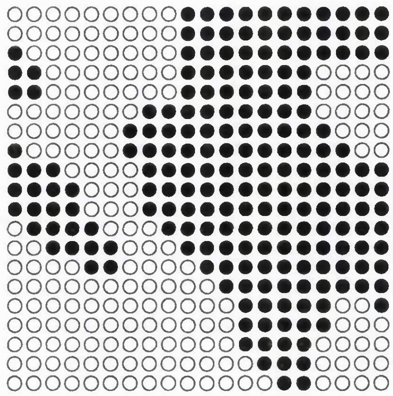


Figure 63 Typical segregation outcome on the Delaunay grid graph.

for the segregation GCA. This further demonstrates the robustness of the segregation process. The major difference evident here, apart from details in the shape of stable configurations, is that multiple distinct stable regions may emerge (as here) rather than almost always growing towards one another so that the whole system is partitioned into only two super-regions.

On the randomly distributed Delaunay structure (see figure 64) segregation also occurs. There is *perhaps* a suggestion in these 10 typical outcomes, that the spatial structure of the system subtly affects overall outcomes. Some regions of the system appear to be homogeneous more frequently than might be expected. However, this subjective impression — which is based on 'playing' with this system on many more cases than can be conveniently shown here — is difficult to confirm analytically. Derivation of cohesive subgroups in the graph coupled with a statistical study of their average homogeneity relative to system homogeneity, might allow the conjecture to be confirmed. An effect not evident from these single snapshot images is that almost all final configurations actually consist of short time period cyclic behaviour — each of the snapshots is a single one only, from the 2, 3, 4 or more system configurations in the cycle on which the system settles. However, no evidence has been found that this effect is strengthened by deformation of the Delaunay graph, which might ultimately lead us to expect chaotic behaviour from the system.

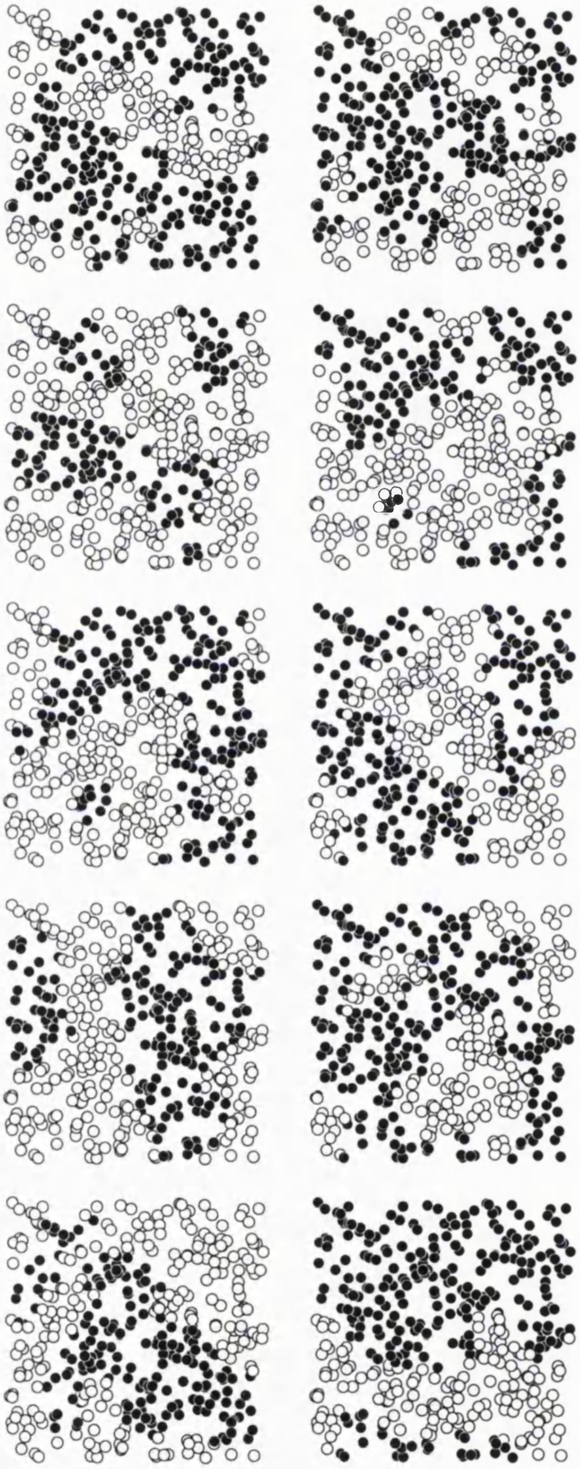


Figure 64 10 typical outcomes on the random Delaunay graph. Note that these are single snapshots from the final short periodic cycles in which the system usually settles.

7.5 Conclusions

The results presented in this chapter represent only a small proportion of the investigations which could be carried out into these questions. Nevertheless, interesting issues have been raised and some potentially interesting results obtained.

The difference between the Life GCA and the segregation GCA is marked. In both cases dynamic complexity is reduced — whether by a statistical change in the likelihood of evolution to homogeneity, or by the detrimental effects of deformation on particular, significant patterns. However, the effect occurs much sooner in the case of the Life CA, after relatively little deformation of the cell space. The segregation rule seems to be relatively unaffected by changes in its lattice, until deformation is so advanced that global structures (large homogeneous regions) become unstable. On the other hand, the Life rule is highly susceptible to changes in the local structure of its cell space, because of the significance of the behaviour of small patterns to its overall behaviour. The results reported therefore suggest that it is possible to distinguish between processes which are *robust* under spatial change, and those which are *fragile*. Further support for this claim is provided by the brief studies in the previous section, where segregation behaviour is still seen on much different graphs. 'Life-like' behaviour is possible on such graphs only by attending to the difficulties associated with varying neighbourhood sizes and constant rule sets already discussed in section 5.3.3 (pages 113ff.).

It is important to acknowledge that this conclusion may be highly dependent on the particular lenses through which we are viewing the behaviour of the systems under consideration. We have examined one particular general class of spatial process, which can be expressed in the rules of a CA model. We have also examined the dependence of such processes on structural change by a particular mechanism of spatial deformation — edge pair swapping — largely for the technical reason that it is preferable to use a process which preserves neighbourhood sizes. Finally, the spatial information measure and its time evolution, which has been used to summarise the dynamic behaviour of different system configurations, is only one possible approach among many.

Notwithstanding these considerations, and noting also the small sample of all possible graph-CA models on which this conclusion is based, it seems that the notion of robust and fragile spatial processes merits consideration. Transferring this observation to substantive geographic contexts is not simple. We might consider processes which are robust in this sense as aspatial, or, following Sayer (1992, pages 145–151), as cases where spatial arrangement is contingent. Correspondingly, fragile processes are sensitively dependent on spatial considerations, and may only be sustained by particular spatial structures, so that particular spatial arrangements are a necessary condition for their continuance. More pragmatically, these results suggest that geographic and urban modellers may need to pay more attention to the often unquestioned assumption that a regular grid lattice (or similar) is an adequate basis on which to construct cellular models in particular cases. This may send some back to earlier debates in geography about the nature of space — perhaps no bad thing in a period when the plentiful supply of digital, rasterised or pixelised data is a tempting place to start any investigation.

Finally, we should note also that these are 'awkward' findings, because they immediately raise questions about the original three-dimensional conceptualisation of GCA-space in figure 24 (see page 109). Spatial structure and spatial process are not properly separable or 'orthogonal' as that conceptualisation suggests. Instead they are intimately related; so much so that the nature of the spatial relational system in which a particular process 'prospers' is itself descriptive of what kind of process it is.

Part III

Application

Chapter 8

Theories of gentrification and a model

Having explored the usefulness of the graph-CA formalism in investigating some very general properties of dynamic spatial systems, the question remains open as to how useful the formalism is if we want to model a particular spatial phenomenon. Therefore, in this chapter and those which follow I attempt to apply the graph-CA model formalism to a particular urban process — gentrification. This chapter presents an overview of some of the relevant literature in this field, and proposes an abstract model of the process, based in part on ideas derived from this review. In subsequent chapters the resulting graph-CA model is explored and applied, and results and implications are examined.

8.1 The gentrification literature

Since the term's first appearance, *gentrification* (Glass 1964, pages xviii–xix) has occupied an important position in the urban geography literature (Hamnett 1991, Smith 1996, provide extensive bibliographies). In particular, after Neil Smith's introduction of the *rent gap hypothesis* (Smith 1979*b*) it became a hotly contended topic, with competing explanations vying for attention (Ley 1980, 1986, Rose 1984, Smith 1982, 1987*a*, Warde 1991, for example), and fairly frequent bad-tempered exchanges between the proponents of various approaches (Badcock 1991, Bourassa 1990, 1993, Clark 1992, Hamnett 1991, 1992, Ley 1987, Smith 1987*a*, 1992). More recently there appears to have been a general agreement to 'agree to disagree' and a more pluralistic approach has been evident (Bondi 1991, Bridge 1994, Redfern 1997*a*,*b*, Robson & Butler 1998, Rose 1984, Shaw 1998, Warde 1991). This section could not possibly thoroughly examine all of these explanations and debates in detail, but instead attempts to extract from this extensive literature the conceptual materials necessary for building a model of some aspects of gentrification processes.

From the outset it is important to acknowledge that gentrification is a rather complex concept. Glass's original description makes this clear:

> "Shabby, modest mews and cottages [...] have become elegant, expensive residences. [...] Once this process of 'gentrification' starts in a district, it goes on until all or most of the original working class occupiers are displaced, and the whole social character of the district is changed." (Glass 1964, page xviii)

Thus, gentrification involves physical as well as economic and social changes in a neighbourhood. A gentrifying neighbourhood is the focus of economic investment in its built environment, as well as an "invasion" (Glass 1964, page xix) by residents of higher social status. Many subsequent writers have insisted on this multifaceted view of the phenomenon: gentrification is "simultaneously a physical, economic, social and cultural phenomenon" (Hamnett 1984, page 284) and "it involves not only a social change, but also at the neighbourhood scale, a physical change in the housing stock, and an economic change in the land and housing market" (Smith 1987*a*, page 463). Ley (1996, page 34) is more equivocal, but defines gentrification to include "the effects of both renovation and redevelopment" which makes clear the physical aspect of the process.

This has led some commentators to label gentrification a 'chaotic conception' (Beauregard 1986, Rose 1984) in the sense of that term discussed by Sayer (1982). According to a later formulation, there is a difference between

> "a rational abstraction [...] which isolates a significant element of the world which has some unity and autonomous force [...]"

and a chaotic conception which

> "arbitrarily divides the indivisible and/or lumps together the irrelevant and the inessential, thereby 'carving up' the object of study with little or no regard for its structure and form [...]". (Sayer 1992, pages 138–140)

Given that the gentrification process encompasses economic, social and physical change, it is not surprising that the concept is vulnerable to such labelling. Smith (1987*a*, page 463) implicitly acknowledges the multiplicity of processes involved when he states "it is this combination of social, physical, and economic that distinguishes gentrification as an identifiable process or *set of processes.*" [emphasis added]

I am strongly inclined to agree with these accusations of conceptual chaos (Redfern 1997*a*, disagrees but does not elaborate). Rather than attempt to weave a single approach to the topic out of the resulting diverse strands, I intend to explore one of the more forcefully argued approaches — the *rent gap hypothesis* — and to use a graph-CA model to examine some particularly spatial aspects of the process which are not examined in the theory as presented by others.

8.1.1 The rent gap hypothesis

The essentials of the rent gap hypothesis are easily explained. The thrust of Smith's (1979*b*) paper is that whereas previous commentators had focused on 'demand-side' explanations of gentrification, "[a] broader theory [...] must take the role of producers as well as consumers into account." Furthermore "[...] when this is done, it appears that the needs of production — in particular the need to earn a profit — are a more decisive initiative behind gentrification than consumer preference." (page 540) In short:

> "Consumer sovereignty explanations took for granted the availability of areas ripe for gentrification when this was precisely what had to be explained." (pages 540–541)

In seeking to explain what amounts to the genesis of 'gentrifiable' areas, Smith turns to a description of a building cycle in urban neighbourhoods, whereby the gap between the potential rent if a site were developed for its 'highest and best use', and the actual rent realised increases beyond some threshold, at which point it is ripe for gentrification, because it has become a profitable investment opportunity. Smith presents his argument in relation to the diagram reproduced in figure 65. The price of a residential site is made up of two components: *house value* and *capitalised ground rent*. House value is regarded as the value of the raw materials and labour which have gone into construction, minus any subsequent depreciation due to wear and tear, and plus any improvements. Capitalised ground rent is somewhat confusingly the "actual quantity of ground rent that is appropriated by the landowner, given the present land use" (page 543). Since rent is a flow and not an amount, it seems likely that this value is actually the expected discounted cash flow from rents associated with the

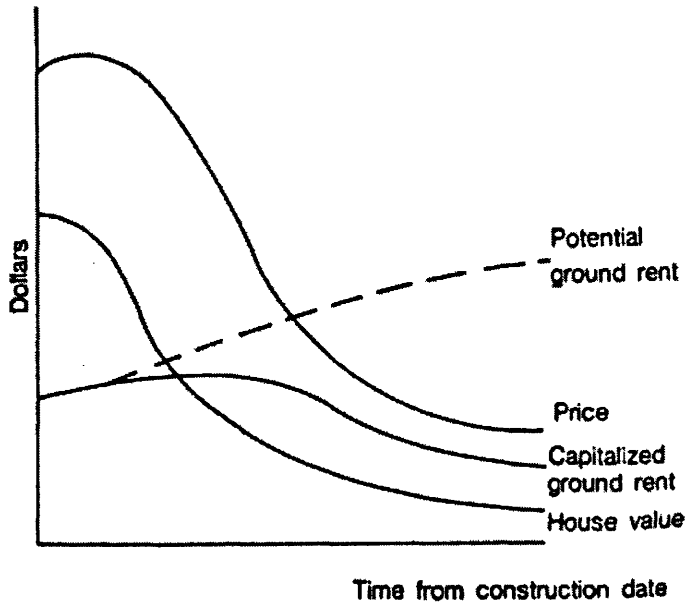


Figure 65 The rent gap hypothesis. Changes in various rents and values in an inner city neighbourhood (*source*: Smith 1979*b*, page 544).

site, assuming that the rent remains at or around the current level. Potential ground rent is the rent which it is expected could be realised, given the site location, under its "highest and best use" (page 543). This is how much a prospective developer or gentrifier would hope to be able to appropriate in rent after gentrification. Again, potential ground rent must be regarded as a discounted cash flow to make any sense of the argument. Note that we must think of the house as providing a flow of services (shelter, heat, light, amenity), to make sense of the rent concept in the owner-occupier case. The value of this flow is then equivalent to the rent in such a case. The fact that the sale price of a residential site is equal to the house value plus the capitalised ground rent means that an owner-occupier stands to benefit from any improvements which raise the capitalised ground rent of the property, since these will be recouped when the property is eventually sold.

Smith uses these concepts in a description of the life cycle of residential properties in 'gentrifiable' neighbourhoods. This is the cycle illustrated in figure 65. When a building is first constructed on an urban site, it will generally be well suited to that site, and in good condition, so that it allows the owner of the site to maximise the rental income. Thus, the capitalised ground rent matches the potential ground rent, and the rent gap is zero. Over time it is likely that house value may change. This is dependent on the extent to which regular maintenance tasks are carried out and also whether or not 'upgrades' in the form of extensions, re-plumbing, re-wiring and so forth, are performed to keep the property up to date. Smith argues for a Marxian labour theory of value, whereby the value of a property is dependent on the amount of raw materials and labour power involved in the initial construction, combined with any subsequent injections of raw materials and labour power. The main mechanism by which a property loses value is then continued technological progress in the construction industry which means that new houses can provide the same level of residential amenity, for a lower input of raw materials and labour power. In any case, the sale price is not related in any consistent way to house value, because that is dependent on the capitalised ground rent. This is a weak aspect of Smith's theory which others have focused on, to which I return below.

Notwithstanding the complexities of the relationship between house value and sale price, house depreciation[1] occurs in some areas, largely because more affluent households take up the better opportunities offered by the modern housing stock in the suburbs. This is more likely to occur in neighbourhoods where some deterioration in the housing stock has occurred. Such neighbourhoods are likely to start to move towards higher rates of tenancy. This is where Smith's focus on the institutional aspects of the neighbourhood life cycle comes to the fore. It is a rational economic response for landlords to under-maintain their properties in such neighbourhoods: since their income is primarily from rent, maintenance spending reduces their net profit, and any reduction in the value of their property is heavily discounted since it is in the medium to long term future. Such landlord behaviour will further reduce the value of the stock, accelerate the out-migration of more affluent households, and see the neighbourhood's decline accelerate further.

Alternative paths to this stage of the cycle are the particularly North American (US) phenomena of *blockbusting* and *blowout*. These may occur in initially less affluent skilled working class and lower middle class neighbourhoods, where deterioration of the stock is a gentler process. In these neighbourhoods

[1]Smith later adopts the different term "devalorization", see Smith (1982)

> "Real estate agents exploit racist sentiments in white neighbourhoods that are experiencing declining sale prices; they buy houses relatively cheaply, and then resell at a considerable markup to black families, many of whom are desperate to own their first home." (Smith 1979*b*, page 544)

Blowout describes the way in which such neighbourhoods in American cities are abandoned by white households as a result of the encroachment of (black) slums from inner areas. Here, many of the buyers are likely to be landlords, often at reduced prices. Again, they may sell to black households at inflated prices. Whichever process occurs, the result is either: new occupants who are ill equipped, partly because of the inflated prices they have paid, to maintain their newly acquired properties; or landlords whose interest in maintenance is minimal.

Another institutional practice may now further accelerate disinvestment. Financial institutions are likely to be reluctant to provide capital for owners (landlords or owner-occupiers) in a neighbourhood at this stage in the cycle. Whether a deliberate policy of *redlining* is pursued or not, the effect is further decline, which may prompt landlords to increase the subdivision of their properties to increase their return, since few other options are open to them. The final (catastrophic) stage of this process is abandonment — or even deliberate destruction through arson — by landlords who can no longer profitably fill their buildings.

8.1.2 Enter gentrification

The essence of Smith's hypothesis, is that the full or even partial passage of a neighbourhood through the various stages of this life cycle will manifest itself as a substantial rent gap between potential and capitalised ground rents. The stage is thus set for gentrification, because the rent gap represents an investment opportunity for owner-occupier and developer 'gentrifiers' alike.[2] Since both groups require finance capital to proceed, gentrification can be conceived as a movement back to the [inner] city by capital, not by people — to paraphrase the subtitle of Smith's paper. Smith claims that the rent gap is "the essential centrepiece to any theory of gentrification" (Smith 1987*b*, page 165).

[2] Note that it is not clear whether gentrification is to be expected when (or if) outer suburbs undergo similar cycles of decline, since the potential ground rent in outer locations may not be so high — in this sense the theory is more than a rehearsal of earlier neighbourhood life-cycle theories, which might suggest that an upturn *will* occur simply because it completes the cycle.

It is impossible to deny the importance of Smith's contribution to the literature. However, much of that importance derives from the at times heated debate which his contribution has provoked, and especially his own "[...] adversarial patrolling of [his] territory [...]" (Ley 1987, page 468) in the face of criticisms. There are certainly flaws in his approach. In no particular order these include:

A confused and confusing definition of rents. This has itself generated an interpretative literature. Clark (1987, 1988) goes some way towards sorting out the confusions, in particular the implication that the rent gap may widen indefinitely. In fact, since capitalised ground rent represents a market *estimation* of the likely future receipts from a site, the gap closes just before gentrification. This is perhaps a narrow technical issue, but it has led to confusion and difficulty in testing the theory. Bourassa (1993) has criticised the concepts of rent in the theory more severely, instead arguing that attention should be paid to the decisions made by developers with respect to a site. A site can be profitably developed if $V_n - C_n > V_c + D_c$ where V_n is the expected value of the site with a new or redeveloped building, C_n is the cost of redevelopment, V_c is the value of the current site, and D_c is the cost, if any, of demolishing the existing building. Since the valuations put on a site by a developer are related to discounted cash flows associated with the rent that various possible futures will yield, Bourassa's criticisms, seem overstated.

No empirical evidence for the rent gap exists. Considering its prominence in the literature, it is surprising how long it took for any serious empirical testing of the rent gap to be carried out. Ley (1987, page 466) draws attention to this when he remarks that

> "[...] almost ten years after its first presentation [the rent gap hypothesis] has not been made empirically accountable [...] [Smith] has no empirical results [...] to report."

This remark is made in the context of defending his own attempt to test the hypothesis (Ley 1986), which has been heavily criticised (Bourassa 1993, Clark 1988, Smith 1987*a*). However, since that time, several empirical investigations have appeared, which broadly support the hypothesis. Clark (1987) found some support for it in a painstaking examination of evidence from a small number of sites in Malmö in Sweden, and again later in Stockholm (Clark & Gullberg 1991). Badcock (1989) also finds limited support for the hypothesis in Adelaide. More recent work in Melbourne provides only qualified support for the theory (Yung & King 1998). All of these authors comment on the difficulty of empirically testing the concept. This has a lot to do with the confusions in Smith's definitions and with the lack of any readily available data which directly reflects the capitalised and potential land rent concepts.[3] This is a reflection of Bourassa's criticism that there is not really any distinction between capitalised and potential land rent, so that the categories are not readily measured nor the theory tested.

The hypothesis is not testable (and therefore not really a hypothesis!). This criticism is potentially more damaging than narrow definitional debates. Smith's initial argument can be simplified (or caricatured) as 'when the rent gap is wide enough gentrification will occur'. Critics reasonably ask "how wide is 'wide enough?' " (Hamnett 1984, page 308) or "When would capital switching occur?" (Ley 1987, page 467). Smith invites this question when he claims that "[w]hen [the rent] gap grows sufficiently large, rehabilitation [. . .] can begin to challenge the rates of return available elsewhere, and capital flows back [to the inner city]" (Smith 1979*b*, page 546). The problem is that if a test of the theory is attempted, and a neighbourhood fails to gentrify, then Smith can say 'not wide enough!' so that the theory survives. The problem here lies in what can be reasonably expected of theory in social science. Because social systems are open systems, no general laws can be expected to hold everywhere and at all times. Any theory can only describe a tendency inherent in particular sets of circumstances. In any particular instance, other factors or tendencies may act against the tendency described, so that the outcome is not the same. In fact, in later work, Smith concedes the point, and effectively denies the possibility of a deterministic theory of the type this caricature implies when he states (Smith 1987*b*, page 165)

> "It is the historical patterns of capital investment and disinvestment in the central and inner cities that establishes [*sic*] the *opportunity* (not the necessity) for [gentrification] in the first place." [emphasis in original]

[3]It is interesting to compare the difficulty Badcock (1992*a*,*b*) has operationalising Harvey's (1978) broad brushstroke theory of urban development. In fact, reliable, detailed data on local property markets is not as plentiful as might be expected, as will become clear in chapter 10.

Certainly, it seems unreasonable for critics (especially those committed to overtly cultural and contingent explanations of the process) to expect that the existence of a rent gap will mechanically lead to gentrification. Nevertheless, the criticism draws attention to the one-sided emphasis of Smith's theory on production or 'supply-side' matters. It is all very well to explain how the opportunity for reinvestment occurs; the rent gap fails to explain why such reinvestment can come to be profitable — what it is that changes the perceptions of a declining neighbourhood sufficiently for reinvestment to be seen as viable. After all, the neighbourhoods where gentrification occurs have not suddenly been relocated nearer the downtown — they have been there along! Hamnett (1991) effectively makes this point when he calls for recognition by the various schools of thought that they are theorising about different aspects of the process which are not inherently incompatible. Smith's (1992) response to this proposal seems unnecessarily vehement.

Conceding the inability of Smith's work to explain the timing of gentrification, it is worth digressing a little, to briefly consider some other explanations, before summarising the implications of all of these considerations for a model of the process.

8.1.3 Other explanations of gentrification

Broadly speaking other explanations of gentrification are rooted in consumer preference. Gentrification is taken as evidence of a shift in the preferences of middle class households from the previously assumed tendency of the affluent to happily trade commuting time for space which rendered 'accessibility' an 'inferior good' (Alonso 1964). This assumption was itself rooted in observations of the historically strong move by the more affluent to the suburbs. By the time gentrification had become widespread enough to be widely observed (say the mid-1960s to early-1970s) this assumption about preferences had assumed the mantle of hard fact, and it is this which accounts for surprise at the gentrification phenomenon, even though it can be readily explained by models of the Alonso kind simply by assuming a change in preferences. However, as Rose (1984, page 49, footnote 10) remarks

> "... the ability to modify the model to correctly predict the occurrence of gentrification does not mean that the neoclassical approach is valid."

This puts the neoclassical theory on similarly shaky foundations to rent gap theory — since if the world does not fit the theory, its assumptions can simply be changed!

Unsurprisingly then, a significant fraction of the gentrification literature is therefore concerned with explaining presumed changes in middle class tastes and preferences for housing. David Ley's work is the most prominent in this strand (Ley 1980, 1981, 1986, 1994, 1996). This work is at its most persuasive when it links changes in tastes to 'deeper', underlying shifts in employment patterns in general, and female employment patterns in particular. This might then account for variations in the tendency toward gentrification of cities with similar housing stock, because the phenomenon is associated most particularly with newer service industry patterns of employment. Sometimes Ley emphasises this aspect of his argument. It is his more 'soft focus' descriptions of the pioneering students, hippies and other counter-culture types who started the gentrification process in many places, which seems to particularly rile Smith, perhaps because it paints a rather rosy picture of the impact on previously resident communities.

The role of cultural changes in the process is also highlighted, especially by Zukin (1982). In her detailed study of New York's SoHo, the habitation of formerly light industrial lofts by artists and other 'alternative lifestyle' groups was a vital stage in the recognition of the potential of that neighbourhood as a high status residential district. This sort of approach highlights two factors ignored in Smith's more rhetorical defences of his theory.[4] First, the need for detailed contextual and historical research into the particulars of gentrification in any individual case; and second, how does the 'rent gap' come to be recognised and closed — what is it that changes the perceived value of particular locations at some particular time? In the case which Zukin studies, the cultural activity of artists not only caused a change in perceptions of particular neighbourhoods, it drew the attention of potential residents, property developers and city authorities to the residential possibilities of previously run-down areas. Similar processes have been identified in London's East End (Green 2000).

This again draws attention to what Bourassa's critique of Smith identifies as the "fundamental" question: what are "the sources of the changes in value that constitute gentrification."? (Bourassa 1993, page 1742) Two interesting recent responses to this question are provided by Robson & Butler (1998) who examine middle class strate-

[4] Although not in some of his other research: see Schaffer & Smith (1986), the first chapter of Smith's (1996) *The New Urban Frontier*, and even some of Smith's (1979*a*) earliest work.

gies for educating their children, and by Redfern (1997*a,b*) who looks at the role of domestic technologies. Redfern's contribution is particularly interesting.[5] He claims to focus on *how* gentrification occurs, rather than *why*, and in answering this question settles on the availability of relatively cheap domestic white goods — washing machines and refrigerators particularly — as the key ingredient, which made the middle class resettlement of areas like Islington, in North London, possible. Again, this is a complex, contextual argument, which focuses on the material aspects of the process. Redfern rejects Smith's gradual decline view of urban neighbourhood cycles, instead claiming that the 'servant problem' caused a rapid fall in the value of large family homes in many such areas. Thus, even only slightly less well off households were unable to live in such homes — because they could not afford to employ the requisite large numbers of domestic staff. As a result, these properties were bought by landlords and converted to multi-occupancy dwellings for the working class poor. It was only with the advent of the affordable washing machine and similar technologies that re-occupation by more affluent households became possible. Redfern's theory seems a little stretched, and certainly it appears to be limited to particular historical contexts. Nevertheless it has one other interesting aspect: he is able to place 'home improvements' on the same spectrum as gentrification, so that gentrification becomes a particular form of extensive home improvement 'with displacement' (of previous residents) rather than an entirely new and different phenomenon.

8.1.4 Discussion and conclusions: implications for modelling

Apart from conceding that even if the concept is *not*, then the gentrification literature itself certainly *is* chaotic, what are we to make of these extensive debates, and numerous perspectives and approaches?! It is not the intention here to unify all the disparate strands of this important aspect of urban geography. In any case, Clark (1994) suggests that such a synthesis may be impossible. The more excessive claims to explanatory primacy of the various schools of thought have been largely abandoned in recent years, which is a welcome development. As Clark points out it may be necessary to use some or all of the various theoretical approaches discussed above in attempting to understand and clarify the progress of gentrification in particular cases (Beauregard 1990, attempts to do just that).

[5] Not to say idiosyncratic!

On the face of it, this leaves model construction in a difficult position. In the absence of any widely agreed theory on which to base a model, but with a whole array of incommensurable theories which might explain various aspects of the process to varying degrees, there does not seem to be much hope of progress. However, I would argue that this is precisely the kind of situation in which building and 'playing with' relatively abstract models of reality is most likely to further knowledge. This is in keeping with methodology in social science, whereby conceptual models of some phenomenon are posited and interrogated in 'thought experiments' to determine which elements are necessary and which are contingent, the aim being to arrive at a view of the explanatory validity of the concepts embodied in the model (Sayer 1992, pages 212–220). In the case of a rather complex situation with many interacting elements, it does not seem unreasonable to carry out 'actual' experiments on our conceptual models using computers. That is one aspect of what is attempted in the remainder of this thesis.

It still remains to construct a process model of gentrification which attempts to pay due attention to the various theories discussed. The details of the process model are outlined in the next section. This model attempts to combine the following aspects of the various approaches described:

In order to unify the 'chaotic' gentrification concept, the model seeks to examine the more general process of investment in the built environment. Investment is simply regarded as expenditure on built infrastructure, whether geared towards residential upgrading or changes in land use. It may be undertaken by developers (large corporate interests down to locally operating landlords) or individual owner occupiers. In either case it is assumed that any substantial investment will have to meet some minimum criterion of financial wisdom, since most investment is funded by borrowing. On the other hand smaller scale investment ('home improvements') may be undertaken out of current income. This 'bottom end' of the process maintains the link with gentrification, as does the focus on an inner urban residential neighbourhood. Note that this means that the model is not strictly about residential reinvestment at all — it is a model of investment in the built environment.

A 'gap' between the current and potential rent/value of a location is regarded as a fundamental 'driver' of the investment decision making process. This is a consequence of the dependence of gentrifiers (whether corporate or individual) on the banking system as a source of finance.

The source of changes in value is a key issue. In this model it is regarded as dependent on actual or expected changes in the *neighbourhood* of a development site. The value of a potential development site is therefore not solely dependent on the current building on that site, but on the relational properties of its location within the model relative to other locations and their values. This is what makes property development and/or gentrification particularly appropriate for a proximal model of the graph-CA type.

Other factors (the change to a post-industrial economy, cultural preferences, for example) *are abstracted outside of the model* and appear as exogenous parameters, if at all.

It is clear that any model based only on these concepts will be limited in its predictive ability, but it is to be hoped that it can still provide some insight, and that the modelling process itself will be enlightening.

Before (finally) describing the model, it is worth remarking on its development. This has been an iterative, evolutionary process. The preceding discussion of the gentrification literature has raised a number of key aspects for any model to focus on — particularly the importance of perceived 'gaps' between the present and potential value of sites, and the relation of these to the profit/risk balance in the investment process. Thus 'gaps' — both in value and income — are important drivers of the model.[6] However, it would be dishonest to present the model described below as having sprung forth fully formed from a reading of the urban geography literature. Its current form was arrived at only after several rounds of trial and error experimentation with generally more detailed versions. A concern for clarity led to a reduction in the number of parameters to the current five. In the next chapter the behaviour of the current version of the model in some simple abstract spaces is described. Work very similar to that reported has been responsible for the current form of the model, insofar as it led to the conclusion that sufficiently interesting and varied behaviour was likely to be observed with the model in its current form.

[6] Wu's (1999) model works on a similar basis without specific reference to the wider literature.

8.2 Developing a graph-CA model of gentrification

8.2.1 An abstract model of property investment

As has been made clear, the purpose here is not to build a fully configured model of housing market behaviour and dynamics. Kain & Apgar (1985) give an idea of the scale of such a project. Recent work by Wu (1999) is closer in spirit to that attempted here. Benenson et al. (1999) and Portugali et al. (1997) also describe models in keeping with the current approach. The intention is to explore the specifically spatial aspects of a simple abstract model whose driving mechanism bears comparison with some of the mechanisms which operate in the housing market. The resulting model is intended as a tool for thinking about urban change, rather than as an operational model for policy guidance, which is well beyond the scope of the present thesis. The essential feature is an operationalisation of various 'gaps' in local property markets. These drive reinvestment and location decision processes, since they act as a draw (the potential profit from large rent or value gaps), as a deterrent (the risks associated with inappropriate locations), and also as a motive for the movement of individual households.[7] The most important drivers of investment decisions in any property market (commercial or residential) are potential profits and risk (Baum & Mackmin 1989, Brown 1991, Cadman & Topping 1995, Isaac 1996, for example), and these can be related to gap-based concepts. Profit potential is taken to be represented by the difference between a property's current value and those of its neighbours. 'Neighbours' must be somehow defined, and this is the 'proximal modelling' step in the process, as discussed in chapter 2. The possibilities in this regard are considered in the following chapters, especially chapter 11. Risk is based on an assessment of the income of the current occupants of a property since this determines the chances of an investor getting back their original investment. Precise definitions are laid out below.

8.2.2 Location state variables

These considerations lead us to a composite state variable at each location in a (graph-CA) model. Each location is taken to represent an individual building or property in some urban fragment. The state a_i of a location v_i is specified by the pair of variables

$$a_i = \langle V_i, Y_i \rangle \tag{8.1}$$

[7] Compare Schelling's (1971) residential segregation model, in this connection.

where V_i is the property's current 'intrinsic' value, and Y_i is the income of the current occupants. For convenience V_i and Y_i are restricted to the range 0.0 to 100.0.[8]

The property value V_i is a notional *intrinsic value* standardised for the property's physical attributes such as size and number of rooms, so that it can be thought of as a value 'per square foot'. The notion of an intrinsic value is an endorsement of the view that successive investments in the built fabric of a property result in increases in this quantity. In this view property is an effective store of value disregarding the effects of depreciation. The price of a property in the open market bears only a passing relationship to its intrinsic value (witness the dramatic, year-on-year changes in property prices commonly observed, particularly in world cities like London). A property's price is a 'virtual' phenomenon, only actualised at the moment of its sale and highly dependent on the buyer's assessment of the property's location and prospects. This is reflected in the model described below, insofar as a property's location is only indirectly related to the income of a new household. Household income Y_i is on a per (adult) capita basis, so that it captures the idea of 'place in the pecking order'.

These two abstractions together mean that we try to treat single 'yuppies' in studio flats as similar to established professional families with two children and a dual income in 3 and 4 bedroom houses. At this level of abstraction we can regard the model as an examination of the overall process of reinvestment in the built environment. The area modelled need not be solely residential, and no distinctions are made between residential and commercial activities in the model. As a result, the model examines processes which would not be regarded as 'gentrification' by many of the writers cited above. This reflects the fact of gentrification as a chaotic concept.

8.2.3 Cellular transition rules

As in any graph-CA model the state of a cell (graph vertex) at the next time step is dependent on the states of cells in its neighbourhood in the graph:

$$a_i^{(t+1)} = f(\{a_j^{(t)} : v_j \in N(v_i)\}) \tag{8.2}$$

Given that a is actually represented by two variables, it is easier to understand, develop, and describe the graph-CA process rules, as a two step process: first, house-

[8]It is a moot point whether V_i and Y_i should be discretised by restricting them to integer values. I am taking the pragmatic view that it is somewhat easier to model a range of *rates* of decline or resurgence with a continuously variable quantities.

hold incomes are modified by a predominantly stochastic process; second, given the current household income and the value of the property relative to neighbouring properties (the local rent gap), the household may invest in the value of the property.

Before outlining detailed rules describing these relationships, it is useful to define some notation and a framework for thinking about the formulation of rules in this context. Figure 66 shows how we can think of a location's state as existing in a two-dimensional state-space. A location's neighbourhood also has an overall state governed by the states of its constituent locations, and it is the relationship between this and the location's present state which determines the state at the next time step. Relevant 'gap' variables are shown in the diagram. Thus G_V^+ is the gap between the central location value, and the maximum value among the neighbouring locations. Similar notation yields the other gaps (G parameters) shown. Note that all the gaps shown are expressed as positive values when the central location state is 'inside' the envelope of its neighbours' states. If a location lies outside its neighbourhood envelope, then one or more of the G parameters may be negative, or exceed the value of the corresponding neighbourhood range (ΔV or ΔY).

It is convenient to 'normalise' the neighbourhood state (see figure 67), and use measures based on this normalisation to formulate graph-CA rules. This involves defining the central location relative to the neighbouring locations' states in a consistent way, independent of the absolute income and value figures involved. We define normalised gap parameters $v^{(-)}$, $v^{(+)}$, $y^{(-)}$ and $y^{(+)}$ as follows:

$$v^{(-)} = \frac{V - V_{\min}}{V_{\max} - V_{\min}} = \frac{G_V^-}{\Delta V} \tag{8.3}$$

$$v^{(+)} = \frac{V_{\max} - V}{V_{\max} - V_{\min}} = \frac{G_V^+}{\Delta V} \tag{8.4}$$

$$y^{(-)} = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} = \frac{G_Y^-}{\Delta Y} \tag{8.5}$$

$$y^{(+)} = \frac{Y_{\max} - Y}{Y_{\max} - Y_{\min}} = \frac{G_Y^+}{\Delta Y} \tag{8.6}$$

These range from 0 to 1 when a location is inside its neighbourhood envelope. They are less than 0 when a location state variable (value or income) is outside the neighbourhood envelope on the 'datum side' of the envelope. That is, $v^{(-)} < 0$ when $V < V_{\min}$, and $v^{(+)} < 0$ when $V > V_{\max}$ (similarly for y and Y). They are greater than 1 when location state variables are outside the neighbourhood envelope on the 'far

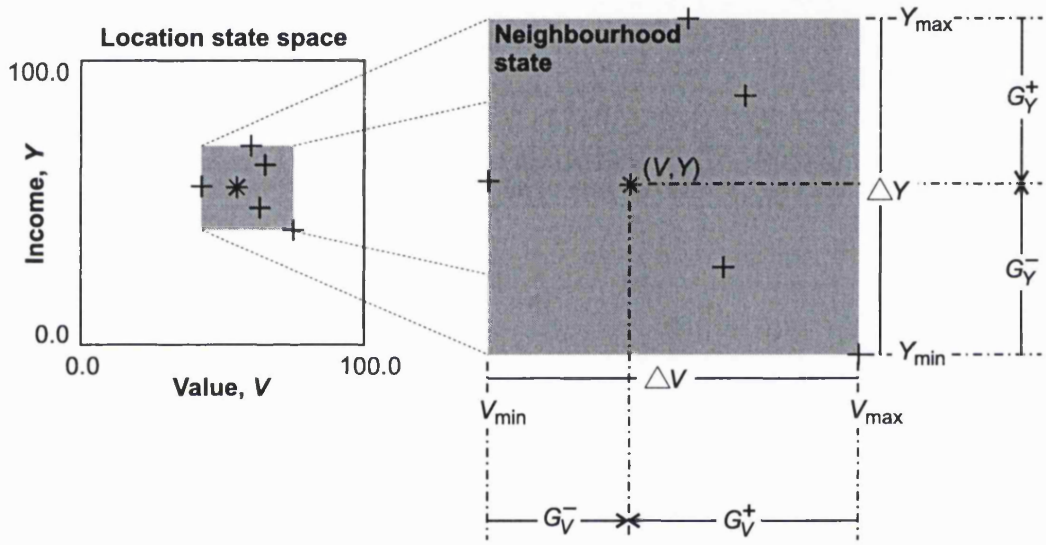


Figure 66 Location state space in the *Gentrification* model. The central location state is indicated by an asterisk, and neighbouring location states by crosses. Various indicated measures of the neighbourhood state are described in the text. Note that the central location state (V, Y) may not fall inside the limits of the neighbourhood state.

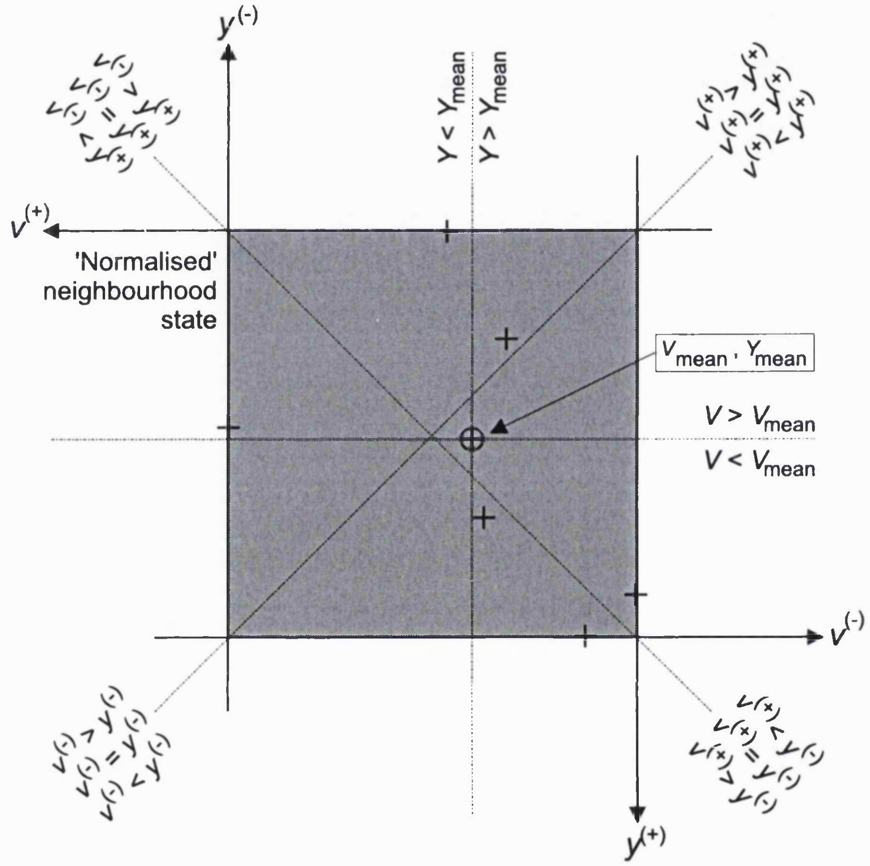


Figure 67 The normalised neighbourhood state and associated terminology. The $\oplus$ symbol indicates the mean value and income state of the neighbourhood. Various equalities and inequalities useful in defining transition rules are indicated.

side' of the envelope.

Importantly, as shown in the diagram, these relations allow regions of overall location state space to be defined relative to the neighbourhood state. It is also easy to use actual values of the gap parameters to define rules. A further possibility is to use the relationship between the central location state and the mean neighbourhood value and income. The graph-CA rules defined in the remainder of this description of the model make use of these inequalities and relationships.

As stated above, the model has been developed with a view to limiting the number of parameters which must be specified for its use. The five parameters which the modeller is required to choose are summarised in table 4. It should be noted that three of these parameters (p_0, r_D and r_M) are unlikely to be varied by very much — their main purpose is to make the model *do something*, and it is convenient in model exploration to have control over them. These are very generalised conceptual parameters and their meaning and likely effects should become clearer in the detailed description of the model below.

Three stages are involved in the change in state at each location in the model at each time step:

1. *Household income may change.* Two processes may be involved in income change. First households deciding to relocate, so that new occupants with a different income move in, and second, exogenous changes in household income, due to changes in employment status, for example. Considering the first of these, a household may decide to move home for any number of reasons, which are typically life-cycle dependent — new jobs, retirement, arrival of children, and so on. This is difficult to model in any simple manner without introducing a host of other household attributes and building a complex 'biographical' model of

Parameter	Description	Typical value
p_0	Willingness to move	0.05
k_A	Willingness to abandon	0.1
k_G	Willingness to gentrify	0.1
r_D	Depreciation rate	0.05
r_M	Income propensity to maintain	0.05

Table 4 The *Gentrification* model parameters.

household residential movements. Instead, households are tested at each time step, and decide to leave the area with probability p_L, where

$$p_L = p_0 + k_A \left[\max\left(0, y^{(-)} + v^{(+)}\right)\right] \tag{8.7}$$

Equation 8.7 states that there is some overall movement in the housing market (the p_0 factor), and that the richer a household is relative to its neighbours, and the less valuable the property is relative to neighbouring properties, then the more likely the household is to move out. The higher is the 'willingness to abandon' parameter k_A, the more likely it is that a household will leave.The use of only positive values of $y^{(-)} + v^{(+)}$ ensures that there is always a chance that *any* household may move. This rule introduces a basic pressure in the model towards residential segregation, by preferentially removing *richer* households [high $y^{(-)}$] from poorer neighbourhoods. It also preferentially makes *less valuable* property [high $v^{(+)}$] available to new households — and subject to either upgrading or decline. The overall impact of this rule is to base the relative probability of a household moving out on a modified 'Manhattan distance' metric in normalised neighbourhood state space. This is illustrated in figure 68.

A household which leaves a location is replaced by an incoming household. This is a key event in the model, since this is where the possibility of better or worse off neighbours arriving arises, potentially tipping the neighbourhood into gentrification or decline. An intermediate value for the new household's income is set according to

$$Y_{\text{temp}} = \begin{cases} Y^* & \text{if } v^{(-)} < k_G \\ Y^{(t)}_{\text{max}} & \text{if } v^{(+)} < k_G \\ Y^{(t)}_{\text{min}} & \text{otherwise} \end{cases} \tag{8.8}$$

where Y^* is a randomly generated income such that $y^{(-)}$ is in the range -0.1 to $+1.1$, that is, between a normalised 10% below the minimum local income and 10% above the maximum local income. The essence of equation 8.8 is that incoming buyers make a decision on the merits of a location based on its current value relative to local values, and that their income is in a similar range to that of their new neighbours. The effect of this rule (illustrated in figure 69) is that low value locations (anything less than a little above Y_{min}) attract incoming households whose income is unpredictable — they could be rich gentrifiers, a

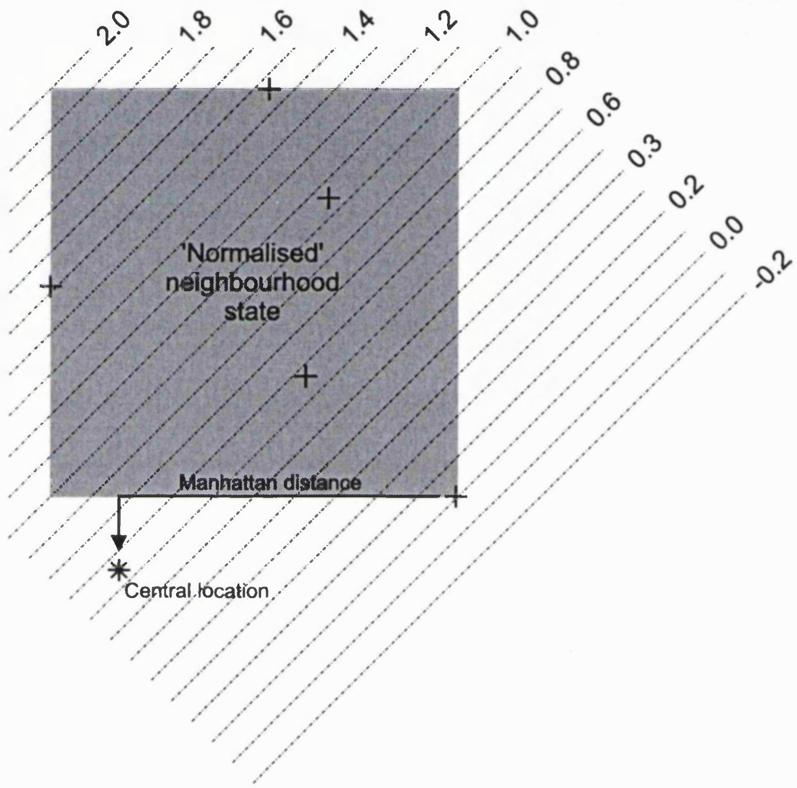


Figure 68 The Manhattan distance from the highest local value and lowest local income is used as to weight the relative probability for a household moving out. This is not a true Manhattan distance metric since $v^{(+)} + y^{(-)}$ may be negative. Negative values are treated as zero. The parallel lines in the diagram are isolines of the modified Manhattan distance measured on this basis.

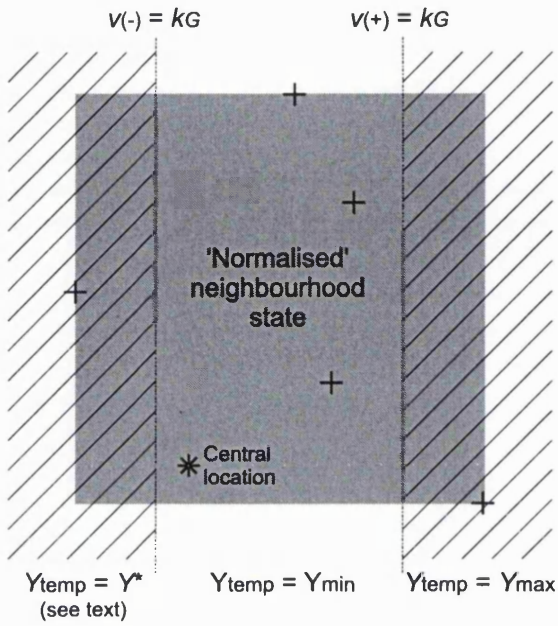


Figure 69 The income of a household moving into a vacated location is dependent on the value of the location relative to its neighbours.

poorer household, or anything in between (with even probability). High value locations (anything more than a little below Y_{min}) attract households with high incomes equal to the richest current locals. Finally, any other property draws low income occupants. This may be thought of representing a move into the local market by landlords who usually rent to less well off households.

2. *The intrinsic value of a property may change.* Once the current household income has been determined, changes in property value are determined. First, a decision is made as to whether mortgage lending funds are available for upgrade of the property. If a property's current value is less than the maximum value in its neighbourhood, then a loan is provided if it is not too risky. This judgement is made by comparing the potential profitability of a location with the potential risk. Both the profit rate and risk are operationalised using the normalised gap measures. Thus

$$\begin{aligned} &\text{IF } (v^{(+)} > y^{(+)}) \text{ AND } (y^{(-)} > v^{(+)}) \\ &\qquad \text{THEN } V_{\text{temp}} = V^{(t)} + x_{\text{LOAN}}(V_{\text{max}} - V^{(t)}) \end{aligned} \tag{8.9}$$

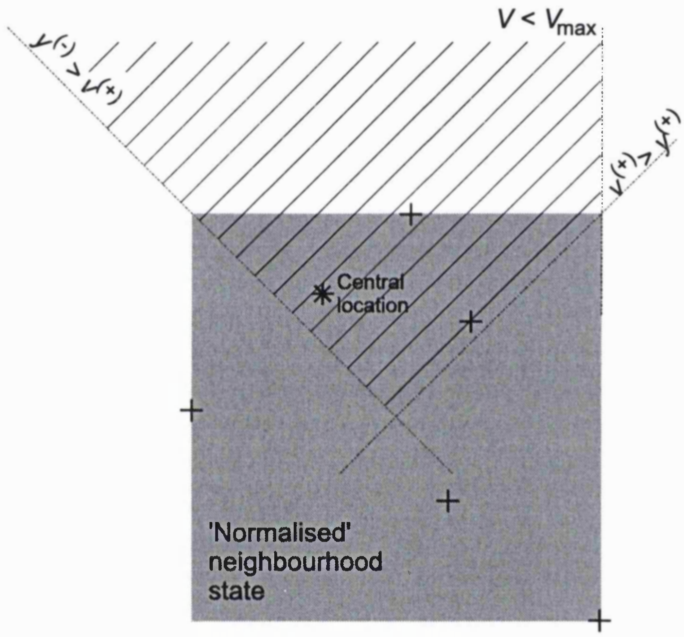


Figure 70 The region of normalised neighbourhood state space in which a location must fall for home improvement loans to be provided.

where x_{LOAN} is an evenly distributed random number in the range 0.9 to 1.1, which will bring the intermediate value of the location V_{temp} up to somewhere around the local maximum. If the above test is failed then $V_{\text{temp}} = V^{(t)}$. The region of normalised neighbourhood state space defined by these inequalities and the requirement that the current value be less than the local maximum is indicated in figure 70. The first inequality compares the potential for gain on the value of the location, against the risk attributable to a relatively low income household. The second inequality requires that the household's income be above some minimum limit relative to the local value gap, before funding will be provided by a lender. Note that both $y^{(+)}$ and $y^{(-)}$ are calculated based on the intermediate value Y_{temp} from equation 8.8.

3. *Whether or not a property changes hands,* or a home improvement loan is provided, the final values of income and value at a location are adjusted to account for exogenous changes in income, and property depreciation and maintenance. Exogenous income change is modelled by a simple stochastic process so that:

$$Y^{(t+1)} = Y_{\text{temp}} + \delta Y \tag{8.10}$$

where δY is an evenly distributed random variable between -0.5 and $+0.5$. Maintenance and depreciation are dependent on the final two model parameters r_D and r_M according to

$$V^{(t+1)} = V_{\text{temp}} + r_M Y^{(t+1)} - r_D V \tag{8.11}$$

which reflects the ability of better off households to carry out more routine maintenance, and the inevitability of some 'wear and tear' depreciation.

The program code which implements this process model is shown in figure 71. Although this code is in *Java,* it should be self-explanatory and implements the model in the order described above. The `Neighbourhood` objects which are created in lines 7 and 8 of this code, provide the method calls to determine the required normalised quantities $v^{(+)}$ and $v^{(-)}$. These method calls are `getRelativePositionFromMax` and `getRelativePositionFromMin` respectively, with the numerical value in question as a parameter. One point worth noting is that the implementation is structured so that the same number of calls to the random number generator occur regardless of the way that the system evolves. This makes direct comparison of slightly different scenarios

```java
Enumeration enum = getVertices();
while ( enum.hasMoreElements() ) { // for each Location

    Location L = (Location)enum.nextElement();
    float y = L.getIncome();
    float v = L.getValue();
    Neighbourhood values = L.getNeighbourhoodValues();
    Neighbourhood incomes = L.getNeighbourhoodIncomes();

    // four random numbers required every iteration
    // generate in this way to ensure similarity between
    // different runs with same random number seed
    float random0 = random.nextFloat();
    float random1 = random.nextFloat();
    float random2 = random.nextFloat();
    float random3 = random.nextFloat();

    // Test for household moving out
    if ( random0 < p.p0 + p.kA * ( Math.max( 0.0f, incomes.getRelativePositionFromMin( y )
                                    + values.getRelativePositionFromMax( v ) ) ) ) {
        if ( values.getRelativePositionFromMin( v ) < p.kG ) {
            y = incomes.getValueAtRelativePosition( -0.1f + 1.2f * random1 );
        }else if ( values.getRelativePositionFromMax( v ) < p.kG ) {
            y = incomes.getMax();
        }else {
            y = incomes.getMin();
        }
    }
    // Test for home-loan
    if ( v <= values.getMax() ) {// perceived need for a loan
        float profitRate = values.getRelativePositionFromMax( v );
        float risk = incomes.getRelativePositionFromMax( y );
        if ( profitRate > risk ) {// loan is worth making
            // test for household ability to cover it
            if ( incomes.getRelativePositionFromMin( y ) > values.getRelativePositionFromMax( v ) ) {
                v += ( values.getMax() - v ) * ( 0.9f + 0.2f * random2 );
            }
        }
    }
    y += random3 - 0.5f; // Exogenous changes
    v += ( p.rY * y - p.rD * v ); // depreciation and maintenance
    L.setNextIncome( y );
    L.setNextValue( v );
}
```

Figure 71 A fragment of the *Java* program code which implements the model described in the text (note that line numbers are for reference only) (initial conditions or parameters) easier, as it means that using the same random number seed in each case will result in the same 'roll of the dice' at each location on two different model runs.

8.2.4 Commentary on the model

As has already been noted, this is a very abstract model of some of the elements in a neighbourhood property market. Essential features which it embodies are:

1. *The exogenous nature of many important factors in the property market.* This is reflected particularly in equations 8.7 to 8.10, and in the probabilities (not certainties) attached to the decision to 'apply' for loans to upgrade property. In reality these decisions are taken by motivated actors in response to many factors in their life. Many of these factors are difficult to model in the cellular framework so a probabilistic approach has been adopted.

2. *Equation 8.8 enforces a somewhat arbitrary bifurcation in the likely income of new residents.* This is the 'engine' of the model, in the sense that changes in a neighbourhood could go either way at this point. If the first change in occupancy in a neighbourhood leads to a lower income household moving in, then subsequent removals seem more likely to lead to further lower income households moving in. The reverse is also true.

3. Such *positive feedback effects in the model are weak,* depending on a series of probabilistic decisions combining to reinforce a neighbourhood effect, with only weak reinforcement via household 'perceptions' of their neighbourhood. For example, it would be possible (and perhaps more convincing) to reinforce an upswing in a neighbourhood's fortunes by making each time step's maintenance dependent on recent changes in the values of neighbouring locations.[9] This could reflect the fact that as an area starts to 'come up' its residents see the advantage in looking after their residences and are more conscientious about maintenance. Equally, recent changes in the value of neighbouring locations could be an input to the decision rules for allotting home loans to households, reflecting lenders' preferences for neighbourhoods which are 'on the up'. This has been omitted

[9] According to Galster (1987) 'neighbourhood reinvestment' in the form of maintenance and property upgrading is strongly dependent on the perception that a neighbourhood is 'doing well'.

in the present version of the model largely to keep the process as simple as possible, in the hope that this will allow any effects of spatial variation in model structure to be seen more easily.

4. As has already been remarked, *the model is very abstract*. This is deliberate. Earlier versions of the model included more exogenous parameters (up to 9). A profusion of variables makes it difficult to assess the overall impact of varying different parameters without necessarily adding anything to the realism of the model, or its usefulness as a 'tool for thinking' about urban change.

5. *There are strong tendencies at work in the model which push neighbourhoods towards a norm of 'the higher the household income, the higher the property value'*, although this direct operation is clouded by time lags, particularly in the depreciation process when a lower income household moves into a previously high value location. The operation of these tendencies is examined in more detail in the next chapter.

8.3 Conclusions

In summary, the model can be interpreted as follows. Households (or businesses) for various reasons — mostly exogenous to the model — may decide to leave the area. Some factors relevant to such decisions are internal to the model, insofar as households who find themselves in relatively inferior properties, or whose income is well above that of their neighbours, are more likely to move out of the area. Incoming households are mostly of two types, near the bottom or top end of the local market. If the property value is locally high then the incoming household's income will be the same as the wealthiest neighbours. In most other cases it will be the same as the poorest neighbours, *unless* the property is very run down, in which case anybody might move in. Households whose property is in poor condition relative to their neighbours are assumed to have an interest in improving the property, but are dependent on lenders for loans to make improvements. Loans are preferentially provide to higher income residents, who may thereby significantly increase the value of their properties. Finally, depreciation, maintenance and exogenous changes in household income may also occur — generally at fairly low rates.

Implicit throughout the model is the concept of the neighbourhood of a location, which has important impacts on its prospects for redevelopment, and the changes likely to occur when the current residents of a location move out, and also on the decision of a household to move out. It is evident that the definition of the neighbourhood of a location will be a key determinant of the way the model behaves. This of course reflects the fact that the model has been implemented as a graph-CA. The seeming reasonableness of the above description hopefully shows that a plausible model of the graph-CA (or proximal) type, can be constructed, for the urban reinvestment process.

Although a full account of the mechanics of the model has been given, it is impossible (or at least very difficult) to anticipate the likely behaviour of a model based on a description of its transition rules alone. The only way to see the implications of this process model is to build it and see what happens! In the next chapter, the behaviour of the model described is reported on simple abstract spaces. This leads to a better understanding of the model dynamics, as an essential precursor to chapter 11 where the behaviour of the same rules is reported on a series of models representing possible neighbourhood morphologies of a real urban fragment in Hoxton, East London.

Chapter 9

Exploring and 'calibrating' the *Gentrification* model

We have just noted how difficult it is to anticipate the behaviour of a cellular model from a bald description of its cell state transition rules. In this chapter the gentrification model developed in the previous chapter is explored in an abstract, spatial setting, so that an appreciation of its dynamic behaviour is attained. This prepares the way for an examination of the model in more realistic settings in the next chapter.

Before describing the model's behaviour, some implementation details are described — effectively a variant of the *graphca* program has been developed: the *Gentrification* program. A set of GIS-based tools has also been developed to facilitate the creation of graph-CA model structures based on urban morphology, and these are also briefly described. These recall the concepts discussed in chapters 2 and 3, and in particular the ideas set out in figure 3 on page 35 and the accompanying explanation. The bulk of the chapter then goes on to explore the *Gentrification* model's properties, and this is used to inform the process of selecting suitable model parameter settings, for the real application of the next chapter. I refer to this process as model 'calibration' in inverted commas since the only criterion for judging the model's behaviour is a qualitative, subjective sense that the behaviour is interesting, and sufficiently rich to merit further exploration in a more concrete 'real' setting. The final stage of this calibration and exploration returns to the by now familiar theme of examining the impact of different underlying spatial structures on the process dynamics.

9.1 An extension of *graphca*: the *Gentrification* program

Before exploring the model of gentrification described in the previous chapter, further development of the *graphca* program was required. This has resulted in the development of a second program called *Gentrification*. Much of this development work consisted of stripping away functionality in *graphca* so that it could be further developed, without needless complication. In particular, the multi-threaded sequencing of the model and its 'in-line' analysis tools have been removed (see sections 6.2.1 pages 143ff. and 6.2.5 pages 153ff., for discussion of these). Furthermore, the concept of a highly flexible transition rule set is superseded by a much more 'hard-coded' implementation of the process model described in the previous chapter. This admits flexibility only in the setting of the model parameters. The core of the implementation of the transition rules has already been presented in figure 71 on page 206. Significant *new* functionality which has been added to *Gentrification* is described in the following sections.

9.1.1 Mapping capability

The most visible new capability of the program is the map-like display of graphs based on real geographical data and settings. This was found to be necessary both to ease interpretation of results, and as an aid in explaining the graph-CA concept to interested parties.[1] Rather than develop these capabilities from scratch a decision was made to use elements from other platforms. Fortunately, a *Java*-based mapping package — *GeoTools*[2] — has been developed at the Centre for Computational Geography at the University of Leeds (MacGill 2000). This provides a geographical mapping capability, and has made it possible to associate the graph-CA underlying a model with a geographical information system (GIS) polygon layer representing some portion of a real urban environment. A particularly useful feature is that *GeoTools* has the ability to read some standard GIS-formatted files. The end result is that the behaviour of a graph-CA gentrification model can be more easily viewed as changes in the state of

[1] As an aside, reaction to the 'blinking lights' of the abstract examples discussed in Part II has frequently been confused — I have found it surprising how few observers appeared to 'get' the concept that graph vertices could represent *any* geographical entities. Initial experience demonstrating the *Gentrification* model (with mapping) to interested parties has been very encouraging in this regard. It is immediately apparent that the intent is to model real processes in real settings, and not necessarily just to 'play' with wholly abstract models.

[2] *GeoTools* is available at `http://www.ccg.leeds.ac.uk/geotools/`.

particular buildings, locations and neighbourhoods in an urban environment. Figure 72 shows the appearance of a typical window in the *Gentrification* program. Graph vertices and edges are clearly visible, and related to the underlying urban morphology.

The coupling between *graphca* and *GeoTools* has deliberately been kept loose to minimise the software redesign required in both packages. Thus, the user first imports a GIS polygon layer. Then, a file containing graph vertices, complete with x-y (Easting-Northing) geo-coordinates is loaded (the `.gca` file format is described in appendix A). Each vertex location is then geometrically associated with a polygon in the GIS data via a point-in-polygon test. Once this link is established, subsequent changes in model location states can be displayed via the mapped polygon layer. The development of this mapping capability makes clear the potential for integration of the graph-CA modelling approach with standard GIS data layers. This possibility is

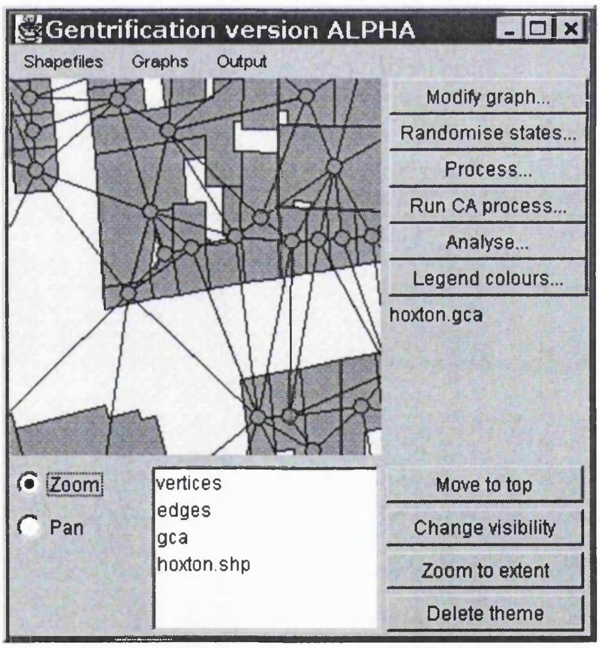


Figure 72 The *Gentrification* model development of the *graphca* program. This figure shows the mapping capability of the program provided by a (loose) integration of the program with the *GeoTools* package produced by the Centre for Computational Geography at the University of Leeds.

discussed further in chapter 12.

9.1.2 Generation of different model structures

The morphology shown in figure 72 is a simple Delaunay triangulation of vertex locations representing the buildings in the map fragment shown. This is just one of many possible bases on which the graph-CA model could be constructed. To complement the *Gentrification* model, an 'urban morphology' extension to a standard desktop GIS tool has been developed which enables the user to generate graph-CA compatible files on a number of different bases. This has been used to generate the graph-CA structures examined in this chapter and in chapter 11. In summary the graph generating processes available through the GIS extension are as follows:

- join with an edge those buildings whose centroids are within some specified distance of one another;
- join with an edge those buildings which are mutually visible; and
- join with an edge those buildings which are on the same straight street segment. This is achieved by determining the mutual visibility of building entrances.

Further, the *Gentrification* program retains *graphca*'s capability to create Delaunay triangulated graphs from building centroids. These various methods for the construction of relations among a set of buildings in an urban environment are the 'proximal guts' of the *Gentrification* model approach, and the GIS urban morphology tools are a key element. The generic spatial analytic capabilities of GIS software constitute a strong argument for the closer integration of any future implementation of the graph-CA concept with GIS.

Graph edges are loaded into the *Gentrification* program independently from the graph vertices. This means that the program allows the user to create logical combinations of the graph edges in two or more input files. Thus the edges in a final graph-CA structure could consist of all those relations between buildings which are less than 50m apart AND are mutually visible. Alternatively, using the same graph-CA input files a structure could be created based on all those buildings which are less than 50m apart OR are mutually visible. This can be thought of as a simple pair-wise logical test on the elements in the adjacency matrix of the underlying graphs. For example, if two graphs G_x and G_y, with adjacency matrices $[x_{ij}]$ and $[y_{ij}]$ have been constructed, then the ORed final graph G_z has adjacency matrix

$$\mathbf{A}(G_z) = [\max(x_{ij}, y_{ij})] \tag{9.1}$$

There is a great deal of flexibility in this scheme to create differently structured graph-CA models of the same urban settings, and to examine the effects of these different structures on the dynamic behaviour of the resulting models. This is in keeping with the experiments and discussion in earlier chapters of this thesis and will continue to be a major focus of this and the next two chapter — rather than, for example, considering the predictive accuracy of the models. The above listed graph construction methods are therefore discussed in more detail in section 11.1 (pages 253ff.), where consideration is given to the resulting overall graph structures in a real setting.

9.2 'Calibrating' complex dynamic models

As has already been implied in section 4.5.1 (pages 94ff.), the dynamic theory of complex systems has significant implications for the predictive role of models, and how we may interpret and use them (Allen 1997). In this chapter, we are faced with the task of calibrating the *Gentrification* model. Conventionally, this process would consist of running the model a number of times with a variety of parameters and choosing those parameter settings which produce the best fit with the real world. The immediately apparent problem is how to describe the dynamics of the real world in a way which allows us to develop a measure of how well model outcomes fit reality. This difficulty goes deeper than the data and the mathematical tools which would be required. A typical calibration process might involve starting the model based on empirical data for 1981, running it up to 1991, and assessing how well the model has post-dicted reality. This is unsatisfactory in light of the fact that complexity theory is precisely about the fact that the phenomenon of interest are not static, and therefore the route a system takes to a particular state is as important as what that state is. It is also fundamental to any complex system that it is underdetermined: many outcomes are possible at the outset and there is no way of knowing which will ensue. The only constraint this places on the choice of model parameters is that the model be seen to be capable of producing outcomes *somewhat like* the single outcome (historical reality) which actually occurred. Note that this is no basis for an adequately specified statistical test, especially (as here) where there is a probabilistic element in the model. If any *one* of the (effectively) infinite number of possible outcomes of our model is 'like' what actually happened, with some choice of parameter settings, then we could argue that those parameters are 'correct', although it would hardly be convincing. On the other hand, just how difficult must it be to find a set of parameters which produce a good fit, before we dismiss the model as 'wrong'?

Rather than attempt to tackle the difficult methodological issues herein, I propose a more exploratory approach. In effect, the remainder of this chapter constitutes a (clearly incomplete) description of the overall dynamics of the gentrification model, aimed at identifying an approximate region in the model parameter space (that is, rule space in the earlier terminology) where interesting and/or plausible model behaviour occurs. This is not a very satisfactory resolution of the issues raised by the unpredictable nature of complex system dynamics. In the end it is justified by the current purpose, the unavailability of any well-established approach to the calibration problem, and a view of complex models which suggests that their most significant application lies in just this sort of exploratory work, which aims at building our understanding of system dynamics rather than detailed prediction.

9.3 Behaviour of the model on a simple regular space

9.3.1 Introduction and intent

Implicit in many of the foregoing remarks is an assumption that the model discussed in the previous chapter will lead to a progress of distinct regions of different value and income housing across a model space — and that it will produce plausible kinds of dynamic behaviour. This is not at all obvious and will be dependent on the chosen values of the model's five free parameters. This section describes typical system behaviour for a range of these parameters on a simple model structure with a view to choosing values of the parameters, which at least produce interesting behaviour. The model space used is shown in figure 73.

This is a simple (non-toroidal) grid where each cell has a von Neumann neighbourhood consisting of the four orthogonally adjacent cells. Cells along the edges and in the corners obviously have only three and two neighbours respectively. The initial

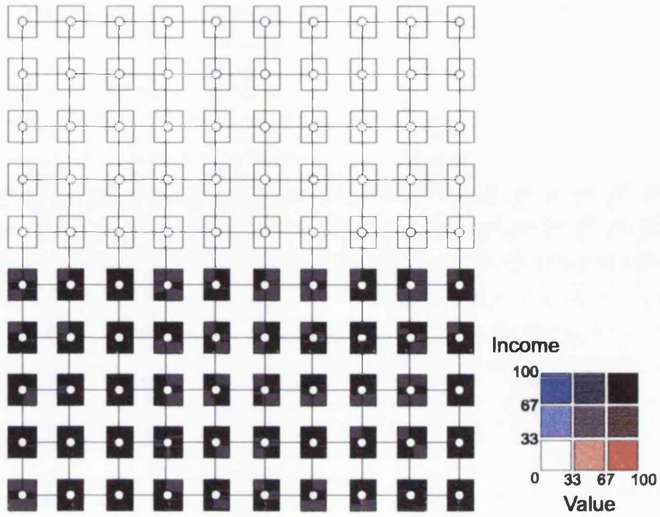


Figure 73 The test model space used to determine suitable model parameters.

state of this system is such that all the locations in the 'northern' half of the system have value and income of 30.0 and those in the 'southern' half have value and income of 70.0. The bivariate state at each location is displayed according to the colour scheme in the bottom right of the diagram, whereby red represents value and blue income, and various mixes of the two indicate corresponding values of the two state variables (Brewer 1994). Locations where value and income are broadly aligned will be various shades of purple/magenta. Locations of low value but high income (gentrifying areas?) tend to bluish colours, and locations of high value but low income (areas in decline?) tend to reddish colours. This scheme is not entirely satisfactory. A particular difficulty is the rather dramatic *perceptual* transition from white colouring to any of the immediately adjacent classifications, which tends to suggest that locations in white are badly run-down. Unfortunately, it is difficult to devise a consistent scheme in which more than a few colours can be readily distinguished.[3]

[3]These difficulties are quite apart from those associated with accurate colour reproduction on screen or in print.

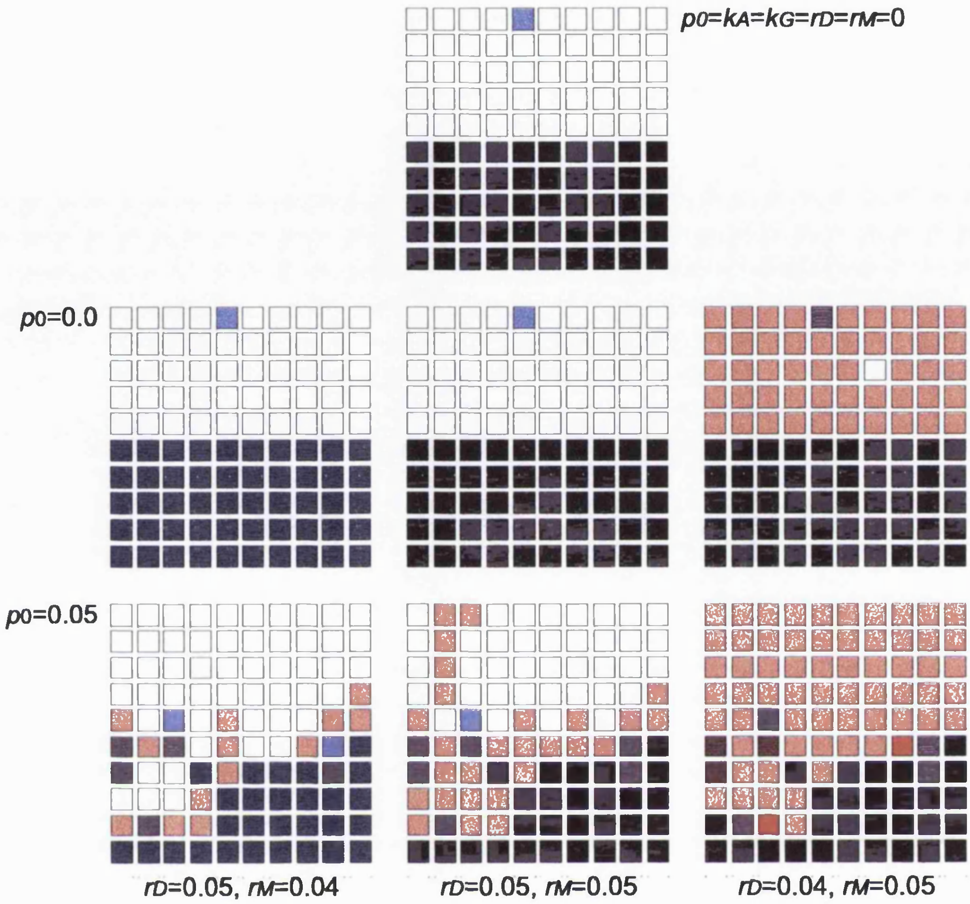


Figure 74 The resulting configurations at $t = 50$ for the gentrification model, for some simple settings of the parameters.

9.3.2 Simple initial parameter settings

We start by examining the model behaviour for some relatively predictable settings of the model parameters (which are listed in table 4 on page 200). Starting from the configuration described above, the model state after 50 time steps, for various parameter settings, is shown in figure 74.

Note that in all of these cases the various random numbers generated during the process have been generated from the same initial 'seed' value. The first point to note is that the overall behaviour is very much as we might expect. With all parameters set to 0 (the single case in the top row) not very much happens at all. The only change is due to stochastic variations in household income. Over 50 time steps this does not cause sufficient variation in income to blur the original boundary between the two regions. The case immediately below, where depreciation and maintenance are in balance ($r_D = r_M = 0.05$) and other parameters remain set to 0 produces a very similar outcome. To the left of this case, we see the effect of setting depreciation a little higher than maintenance ($r_D = 0.05, r_M = 0.04$), and to the right, the effect of the reverse settings ($r_D = 0.04, r_M = 0.05$). As might be expected, the former case sees a steady decline in all locations, and the latter a steady rise in values in all areas. However, there is still no blurring of the boundary line between the distinct regions in the initial configuration.

When the possibility that a household may move out of its neighbourhood is introduced in the cases in the bottom row ($p_0 = 0.05$), then more rapid and fundamental change in the system occurs, since some households begin to leave and be replaced by newcomers. Crucially, the boundary line between the regions does blur in these cases — in fact, in all three cases the lower income, lower value region is seen to 'advance' into the higher income, higher value region. This is because with k_G set to 0, there is a strong likelihood that the incoming household will have an income equal to the minimum income in the neighbourhood (refer back to equation 8.8 on page 201 to confirm this), so that an overall downward trend in income is observed. Note that the same location (top row, fifth from left) frequently stands out in these examples, demonstrating the repeatable use of the pseudo-random number sequence in the program implementation. The spatial pattern of advance of the lower status region is also very similar in each case, for the same reason. Different runs using other random number sequences produce outcomes which are different in detail, but similar overall.

It is difficult to draw any strong conclusions from these 'snapshots' of single system configurations at a particular time step. To improve our understanding of what is happening, we can examine the overall change in the system in location state space, and the evolution in state space of some specific locations. This view of the system behaviour is shown in figure 75, where the lower six cases of figure 74 are displayed in this way.

Each of these diagrams shows the evolution over 50 time steps of the average value and income of the whole system (heavy line) and of 10 individual locations in the system to give some sense of what particular local changes are contributing to the

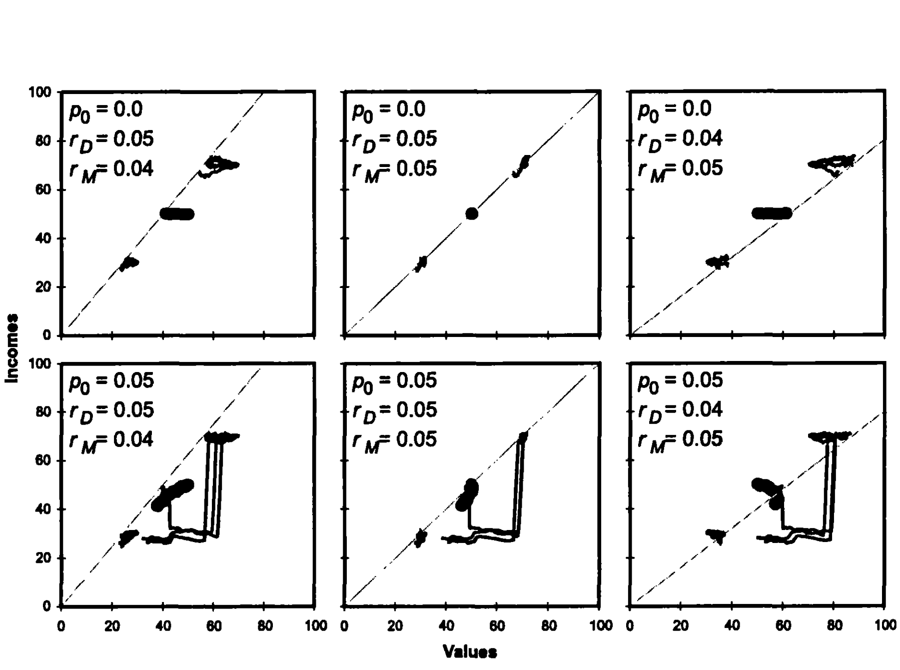


Figure 75 State space diagrams for the lower six cases of figure 74. The six diagrams are positioned in equivalent locations to those in figure 74.

overall system evolution. The ten individual locations are not necessarily the same in each plot, but have been selected to show something of the range of behaviours which occur in the model in each case. Note that interpretation of this and other similar diagrams is eased if one bears in mind that all locations start at either $(30, 30)$ or $(70, 70)$ and that the system average starts at $(50, 50)$. Also, dramatic (instantaneous) horizontal shifts in location state can only be associated with single time step large changes in value — home loans — whereas dramatic vertical shifts must be associated with new households of very different income moving into a location. Dramatic diagonal moves in state space can only be due to incoming new households immediately obtaining home loans and upgrading the property. The reverse transition is not possible — a location's value can only undergo relatively slow decline (dependent on the relative values of r_D and r_M).

The grey diagonals and near-diagonals in these plots are the line $Y = r_D V / r_M$ in each case. It is evident that this acts as an attractor for locations since along this line $r_D V = r_M Y$, and the net change in value of a location, in the absence of home loans, is zero (see equation 8.11 on page 205). In the upper row of this diagram the effect of varying the relationship between r_D and r_M can be clearly seen. Where $r_D > r_M$ the values at all locations drift down from their initial values towards the attractor. In the reverse case where $r_D < r_M$ values move up towards the attractor. Where $r_D = r_M$ the effect is still discernible, since random fluctuations in income mean that values constantly adjust towards the $Y = V$ attractor. If stochastic changes in income lead to $Y_i < V_i$ then the property value at v_i falls due to 'under-maintenance' restoring the equality; the converse is also true. This is an example of negative feedback dynamics, often associated with classic equilibrium economics.

The introduction of household movement out of the system ($p_O = 0.05$) drastically affects this equilibrium-seeking behaviour. This is most clearly seen in the individual dramatic falls in income visible in all three cases in the lower row of figure 75. These occur when households move out and are replaced by new households with incomes equal to the local minimum, and this is a common outcome because the willingness to gentrify $k_G = 0$. Even so, there is also one example of a richer household moving in. This case occurs when the value of the affected property has fallen below the local minimum, so that $v^{(-)}$ is negative and hence less than k_G, and a randomly selected new income is generated. The fact that the same new income of around 50 is generated in all three cases again forcefully demonstrates that the same sequence of random numbers has been used in all cases. In spite of these dramatic single instances of change in the system, it is also notable, that other locations proceed much as before. Routine depreciation and maintenance are not affected by the neighbourhood state, so that the underlying movements of those locations unaffected by households moving out are unchanged.

9.3.3 Introducing non-zero settings of k_A and k_G

Now we examine the behaviour of the model for various combinations of the abandonment and gentrification 'conditioning' parameters k_A and k_G which seem likely to influence the trajectory of the system over an extended run. Figure 76 shows outcomes at $t = 50$ for a range of values of these parameters for fixed values of $p_O = r_D = r_M = 0.05$. Looking at the progression up the $k_G = 0$ column (increasing k_A) there is a clear trend whereby the higher income area sees falling incomes and values, until with $k_A = 0.5$ it is completely 'abandoned'. By the time we reach the $k_G = 0.2$ column, this trend is reversed, and increasing values of k_A lead to a more

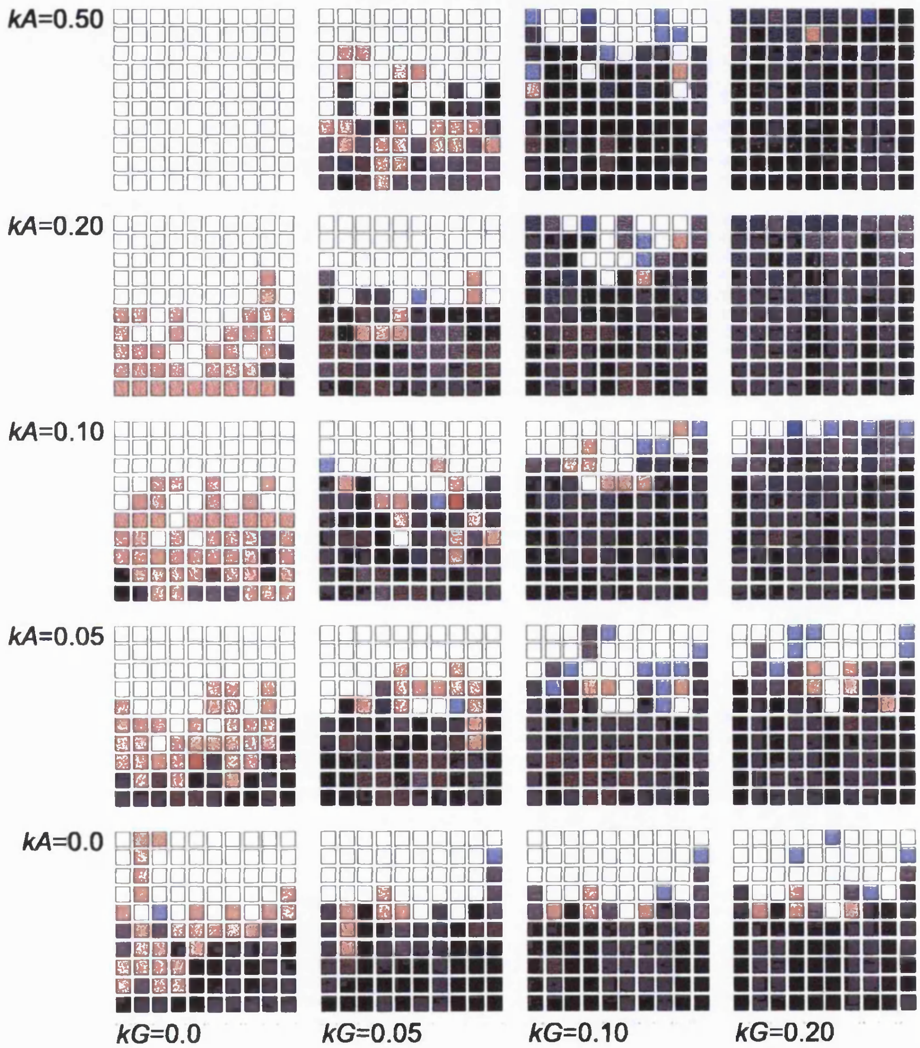


Figure 76 Outcomes at $t = 50$ with various values of k_A and k_G for $p_0 = r_D = r_M = 0.05$. Note that the increments in the values of the two parameters are not the same.

dramatic advancement of the higher status region into the lower status region.

Another way of looking at these results is across the rows. As k_G increases at some fixed value of k_A, there is less encroachment of the lower income area on the higher income area, so much so, that where $k_G \geq 0.1$ the higher income area advances into the lower income area, or an approximate balance is maintained. The effect of higher values of k_A on this trend is to emphasise it. At low values of k_A there is not much difference across the range of values of k_G, whereas the $k_A = 0.5$ case sees a dramatic shift in outcomes over the range of variation in k_G. This indicates that we may be able to think of the abandonment factor, k_A as a 'valve' controlling the freedom of the gentrification process (governed by k_G) to operate.

We can again examine this understanding further by considering the state space dynamics for these cases. These are shown in figures 77 and 78. In these diagrams, higher values of k_G consistently lead to increasing amounts of upgrading activity at individual locations. This is particularly evident when $k_A \geq 0.10$ where there are examples of locations seeing richer incoming residents who immediately obtain home improvement finance, resulting in dramatic, simultaneous increases in both value and income (seen as large diagonal movements in state space). Note that the large diagonal shift evident when $k_A = 0.10$ and $k_G = 0.05$ does not re-occur at higher values of k_A. This may be because the upward drift in incomes at neighbouring locations in those cases increases the assessed risk of the required loan, so that it does not occur.

It appears from these diagrams that more information might usefully be gathered about behaviour where $0.2 < k_A < 0.5$. For example, it appears that where $k_G = 0.2$, the change in k_A from 0.2 to 0.5 means that fewer rich incoming households stay long enough to receive home-loans and perform residential upgrading. Instead, since it is likely that rich incoming households will have high $v^{(+)} + y^{(-)}$ they are very likely to move out immediately, unless they receive a home loan (see equation 8.7 on page 201). All the non-zero cases of k_G where $k_A = 0.5$ show evidence of this effect. This suggests that this sort of setting of k_A may be too 'aggressive' for convincing model behaviour to be observed.

Such considerations together suggest a possible amendment to the model, whereby k_A is made dependent on some exogenous economic variables. This would result in k_A becoming a proxy for 'housing market activity' and perhaps regulating

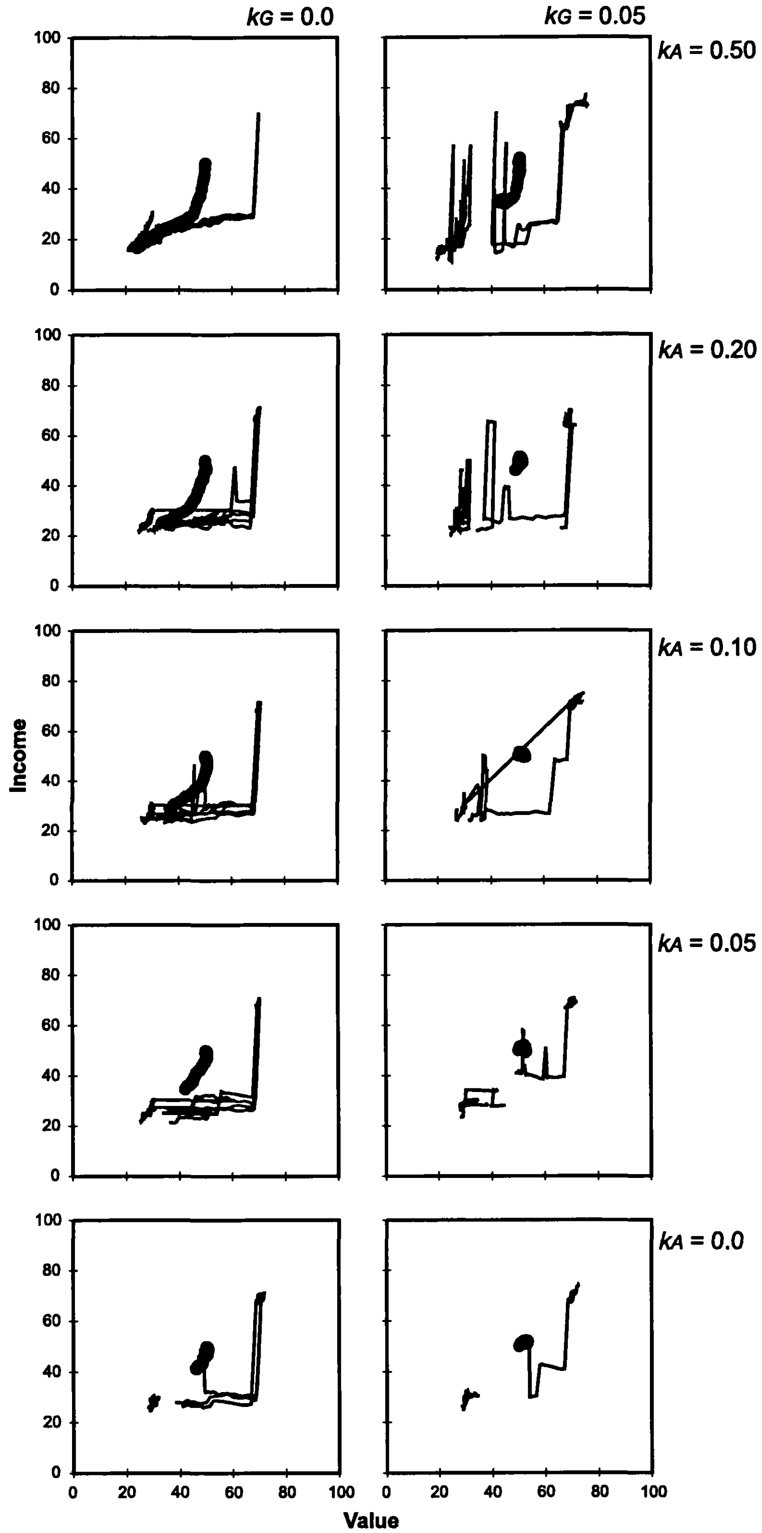


Figure 77 State space evolution over 50 time steps with various values of k_A and k_G for $p_0 = r_D = r_M = 0.05$ (part 1). Note that the increments in the values of the two parameters are not the same.

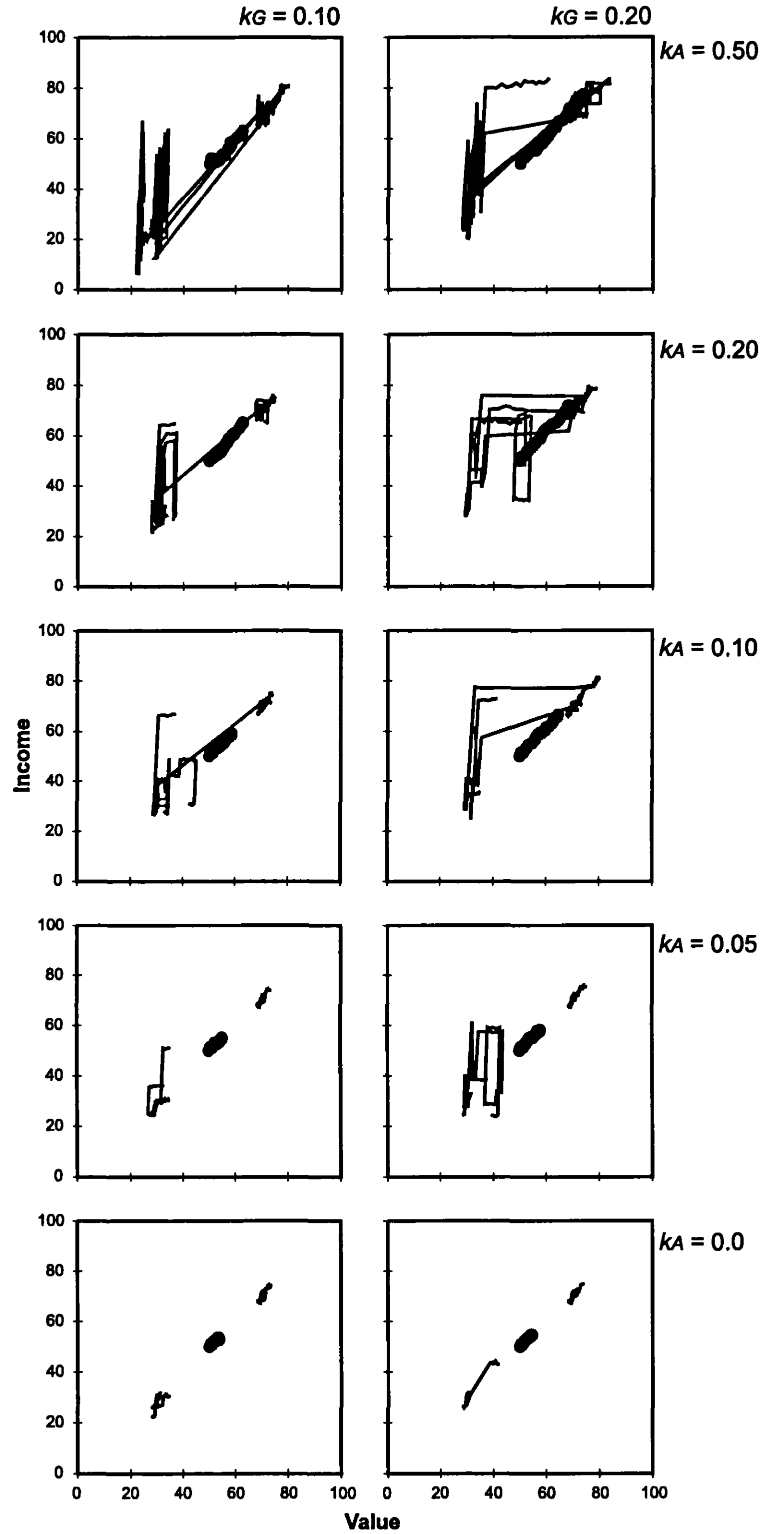


Figure 78 State space evolution over 50 time steps with various values of k_A and k_G for $p_0 = r_D = r_M = 0.05$ (part 2). Note that the increments in the values of the two parameters are not the same.

overall behaviour of the model. Such an approach fits well with the commonplace observation that the most rapid upward changes in neighbourhood status tend to occur at times of economic buoyancy, whereas the greatest downward mobility of neighbourhoods occurs in times of recession. An essential aspect of the gentrification process then is a ready supply of houses available on the market for intending incoming households or investors to upgrade. Linking k_A to an external economic model, or perhaps varying it cyclically to simulate the business/economic cycle, might result in interesting effects. Since the aim is only to demonstrate the feasibility of the graph-CA modelling approach, this suggestion is not pursued further here.

What can we conclude about suitable model parameters from all of the foregoing? It is difficult to draw any very firm conclusions. The cases presented so far do not represent a sufficient sample of the model's behaviour under variation in rule space to choose parameter settings. They do, however give a sense of the model's overall behaviour and scope, and show that interesting behaviour can be produced. Individual locations exhibit a range of trajectories, including gradual decline or upgrading, and occasional rapid upgrading on the basis of loans. Emergent neighbourhood effects are harder to verify, although with the particular model structure here, and the clearly segregated initial conditions, particular neighbourhood characteristics do appear to 'advance' or 'retreat' spatially across the model space.

9.4 Probabilistic variation in model behaviour

Before choosing final settings for the parameters, it is important that the effect of the probabilistic elements in the model be briefly examined. In this section various settings of the parameters are run for a number of different random number generator seeds. Following that, we settle on a single parameter selection, prior to examining the effect of varying the underlying graph-CA structure in the final section of this chapter.

We have already noted the repeatability of the random number generator in the model, with reference to similarities in outcomes at particular locations with different parameter settings. This is also obvious in figure 79, where each row represents the outcome after 50 model time steps for a different random number seed. The four cases in each row are for the different indicated settings of the model parameters. Similarities in the pattern of the outcome are evident for different parameter settings using

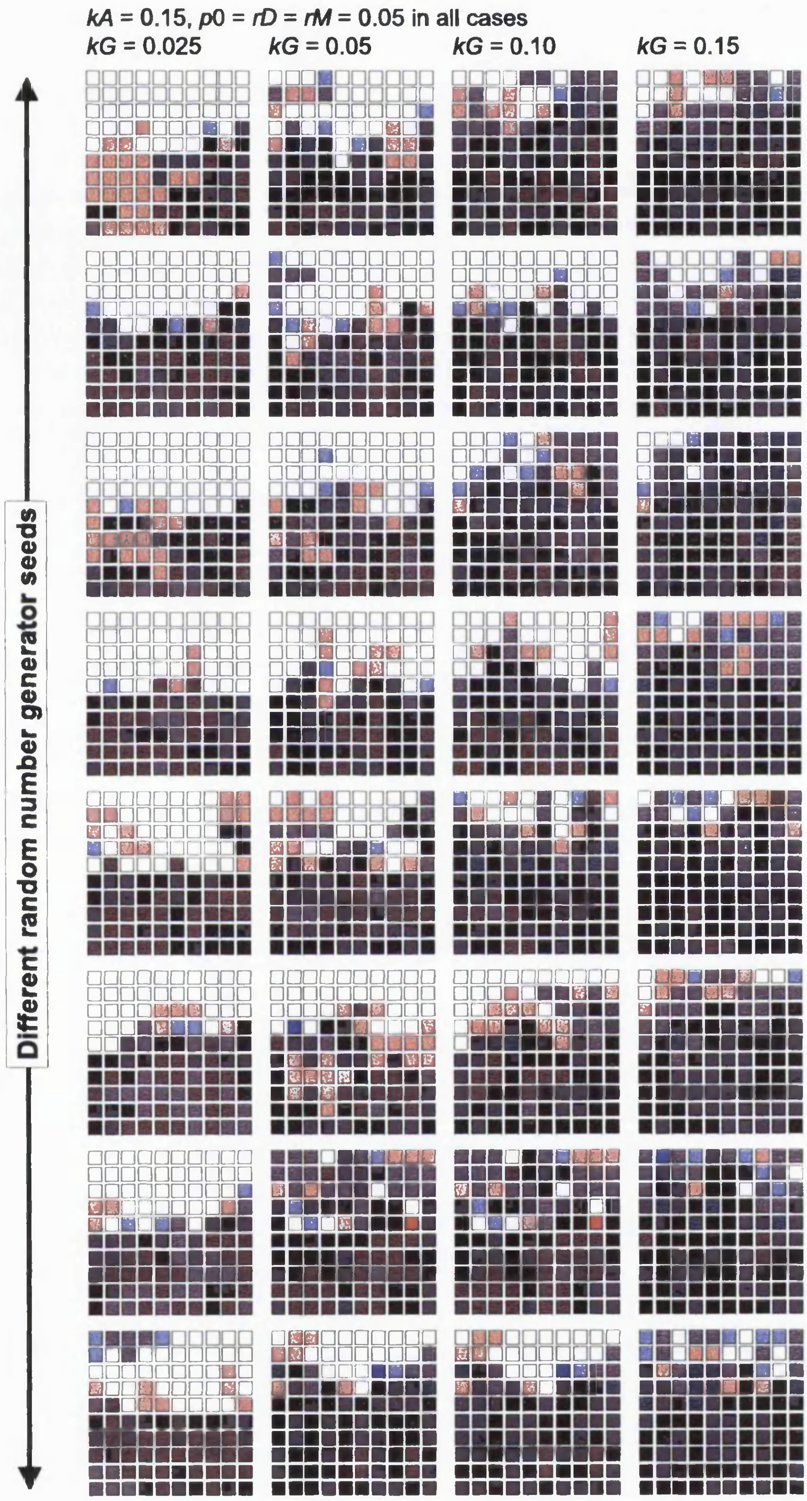


Figure 79 The effect of changing the random number seed over four further plausible combinations of the model parameters.

the same random number seed, especially in terms of roughly where the boundary between the two regions ends up. Conversely, each column in the diagram represents the outcomes for a single selection of parameter settings, for eight different settings of the random number seed. It is evident from variations in the outcomes in each column that the dominant effect of different random number sequences is in the detail of the model behaviour at different locations. The overall system state is broadly similar, in terms of the advance or retreat of the different status regions, for a fixed set of parameters as the sequence of random numbers is varied. This is probably what would be expected in the regular morphology of the present model: the spatial regularity of the system means that it really doesn't make very much difference which locations change initially, in terms of the interconnection of those locations to others in the system. Whether or not the same is true of non-regular spatial structures is an issue for exploration in chapter 11.

9.5 Behaviour of the model on other spatial structures

In the next two chapters the *Gentrification* model is applied to a real-world space with various irregular underlying graphs. Here, we anticipate possible effects of this by running the model on a series of different graphs with the same grid-based set of vertices, as used in the preceding section. The graph structures used are illustrated in figure 80. These have been generated using the GIS tools described above and discussed in more detail in chapter 11 (see section 11.1, pages 253ff.). The 'buildings' in this case are the regular blocks arranged in a rectangle, as in figure 73 on page 216, and have been omitted for clarity. Two points about the graphs themselves are worth noting. First, two of them are not symmetric: the Delaunay graph, and the street segment based graph. Second, the 'street segment' graph is not based only on street segment relations between elements, as this results in a graph with disconnected components. Instead, use is made of the logical graph combination feature already mentioned, and the graph is built by OR-ing its edge set with the edges for the Delaunay case. The same approach is adopted to avoid disconnected graphs in chapter 11.

The outcome states for each of these, after 50 time steps of model evolution from the same initial configuration as previously, for $p_0 = r_D = r_M = 0.05, k_A = 0.20$, and $k_G \in \{0.025, 0.05, 0.075\}$, are shown in figure 81. Considerable differences in the spatial outcomes are evident. The von Neumann and Delaunay graph cases show the

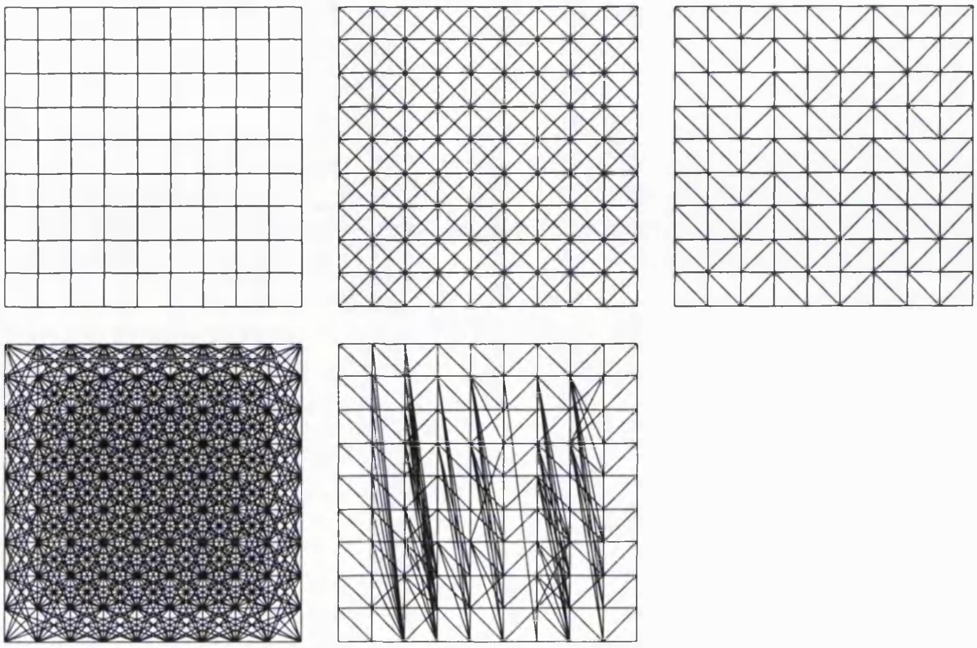


Figure 80 The five graph structures examined in this section. From left to right, top to bottom: a von Neumann grid, a Moore grid, a Delaunay triangulation, a visibility-based graph, and a street segment based graph. These have been constructed based on the blocks in figure 73 (page 216).

most spatial 'coherence' in that the initial distribution of low and high status areas in the system is more nearly preserved after 50 time steps in these cases. The Moore neighbourhood case seems to undergo change more rapidly than either of these — the overall system evolution (neighbourhood decline or regeneration) seen in the von Neumann and Delaunay cases still occurs, but is more advanced at the time step shown here. On the other hand, the visibility and street segment based structures have evolved towards very different distribution patterns, with little obvious relationship to the initial state. In fact, there seem to be distinct similarities in the arrangements produced in these cases. Since the structures themselves are not similar, this suggests that the probabilistic effect of the random number sequence used is the dominant influence on the outcome. In turn this suggests that spatial structure may have little impact on these cases. That is, the pattern which emerges is not dependent on the spatial structure, but on the sequence of random numbers drawn. The

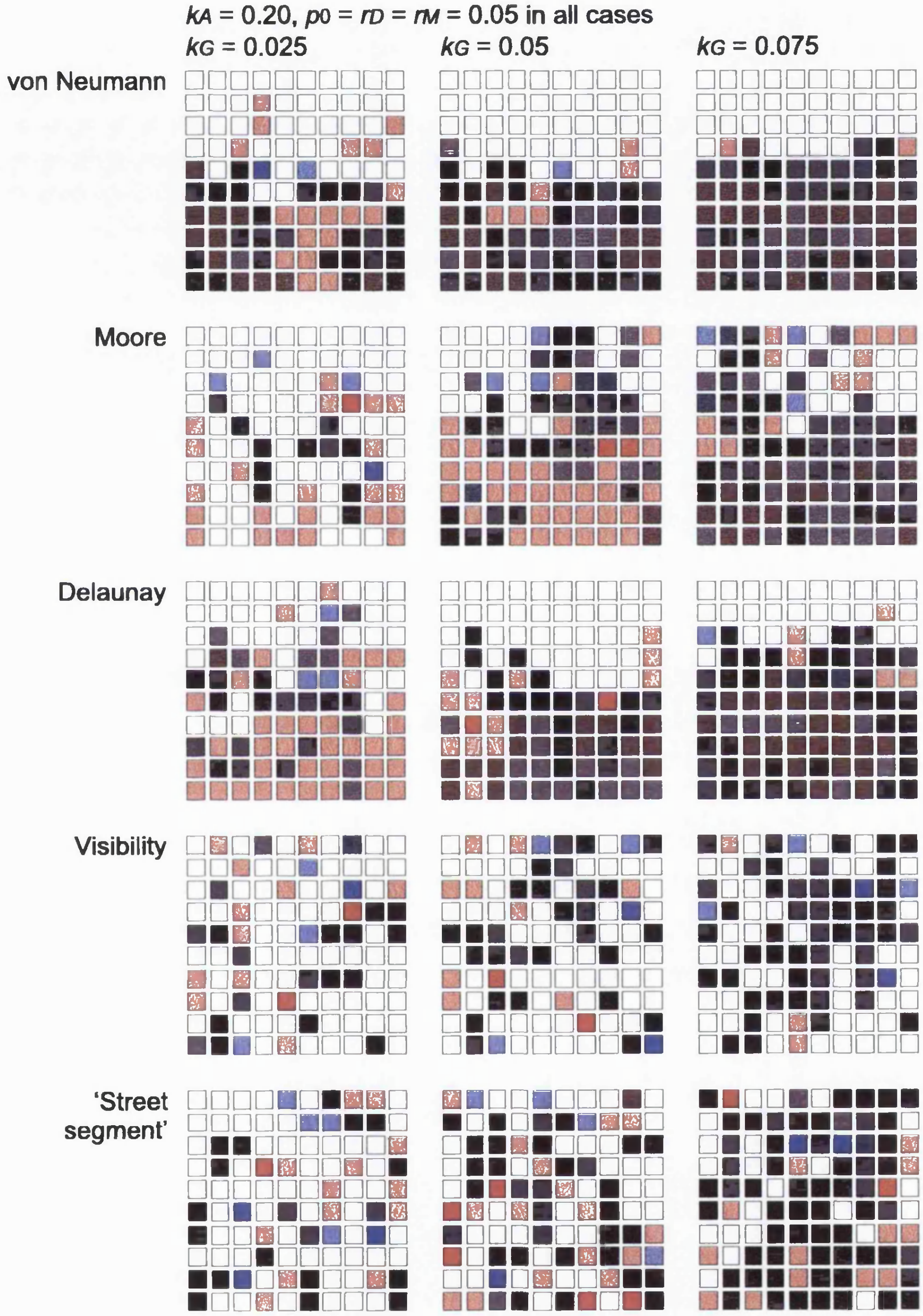


Figure 81 Outcome at $t = 50$ for the five different graph structures of figure 80.

most plausible explanation for this effect is that the large size of location neighbourhoods relative to the whole system means that individual location neighbourhoods are more similar to one another in these cases. This effect is probably more a reflection of the limited size of the examples here than of any more fundamental property of the model. Whether or not probabilistic effects are dominant in the model's evolution is again an issue for exploration in chapter 11, where the behaviour of irregular real-world-derived model structures is examined.

To try to understand these effects better, we can again examine the movement of the system in state space (see figure 82). The most striking effect here is the much wider range of variation in both value and income — particularly income — which is seen in both the visibility and street segment based structures. The common factor would seem to be that these two structures have larger location neighbourhood sizes. It is not obvious why this leads to the wider variations. Given the form of the model rules, it would appear that the difference must somehow lie in the likelihood of higher incoming household incomes being randomly selected according to equation 8.8 (page 201) which only occurs if $v^{(-)} < k_G$. This condition does not seem especially more likely to be satisfied with larger location neighbourhoods, although as soon as one or two high and low income locations appear, they are likely to affect subsequent events at many locations. Again, it may be not so much the absolute size of neighbourhoods which is important as their extent relative to the whole system of which they are a part. Thus in the visibility structure case, a single location neighbourhood may be up to nearly a quarter of the whole system size. Whatever the exact details of an explanation it seems clear from these plots that the effect is there. It seems likely that lower values for k_A are required for more stable outcomes with such relatively large location neighbourhoods. Another aspect of the different outcomes in these two cases (seen more clearly in the snapshots of figure 81) is a tendency for individual locations to be either low value and low income, or high value and high income and much less likely to fall anywhere in between. An explanation for this effect is elusive.

The clear implication of these findings, as in part II, is that the spatial structure of a dynamic system can have a significant impact on outcomes. In terms of the *Gentrification* model under discussion, the implication is that the way neighbourhoods are perceived by agents in the property market is likely to have a significant impact on spatial outcomes. By "way neighbourhoods are perceived" I do not mean the

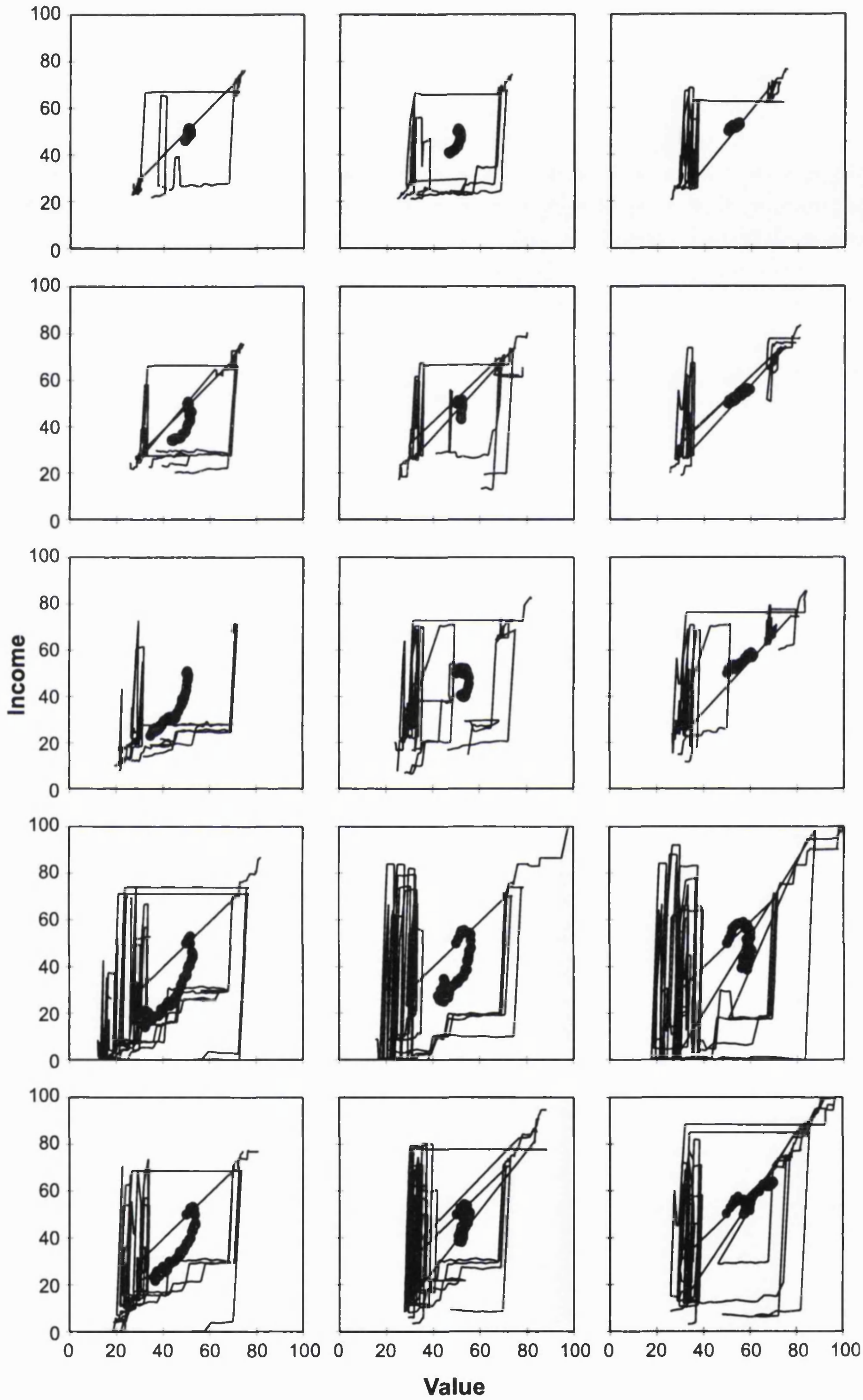


Figure 82 The behaviour in state space of the model for the five different graph structures of figure 80. Results are presented in the order of figure 81.

relatively obvious sense of whether or not an area is held to be doing well or not, but the less tangible notion of how interrelated different locations in a spatial system are perceived to be. The above reported results cannot easily be used to examine this sort of issue directly, but do perhaps suggest a model-based way of proceeding.

9.6 Conclusions

Returning to the main theme of this chapter of understanding the dynamics of the model and using that knowledge to select plausible parameters, prior to running it for more realistic data and spatial structures: what can we conclude? First, that the model as specified produces a range of system behaviours, which are composed, in most cases, of locations exhibiting the full range of behaviours made possible by the rules. That is, households may move out and be succeeded by richer or poorer incoming households. Richer incoming households may subsequently obtain home loans to carry out significant upgrading of a location, or they may gradually upgrade the location by virtue of their 'above-trend' income. Alternatively, they may fail to obtain finance and themselves move out after only a short stay. On the other hand when incoming households are less well off a location is likely to see prolonged slow decline in its value.

A range of settings of the parameters have been examined — primarily variations of the relationship between k_A and k_G, with low non-zero values of p_0, r_D and r_M. Apart from the observation that some upper bound on k_A is reasonable so that there is sufficient stability in the system for overall trends to emerge, no very firm conclusions can be reached. This is not unexpected: the model is after all not closely bound to any observable parameters (such as, say, interest rates) in the real world. However, it seems to be possible to understand k_A as a proxy for property market activity, and by setting a relatively low value of k_G (in the range 0 to about 0.1, say) to observe an interesting range of model behaviour. This suggests that parameter settings something like $p_0 = r_D = r_M = 0.05, k_A = 0.10$ and $k_G \in \{0.025, 0.05, 0.10\}$ are suitable for exploration of the model in real urban spaces. The further observation that the model's behaviour — in terms of both spatial outcomes, and location values and household incomes — may be affected by the underlying spatial structure suggests that caution is required before settling on any fixed set of parameters. In effect, this means that rule space and cell space are not independent, and that it may be important to continue to explore some of the potential for variation in the model as irregular system structure is introduced.

It also seems that it is important to check the degree to which differences in the stochastic events in the model affect overall outcomes. This is arguably the most important test for specifically *spatial* effects: very similar outcomes occurring spatially, regardless of stochastic effects, would indicate that the system's spatial structure was 'steering' or 'shaping' the outcome towards certain preferred system states. On the other hand, if stochastic events 'prevail' so that different outcomes, with no distinctive spatial structure emerge, it could be argued that there are no strong spatial effects resulting from the system structure. Running the model on the mainly homogeneous spatial structures of this chapter produces no obvious distinctly spatial effects, but the homogeneity of these spaces might lead us to expect that. The question of the emergence of distinct spatial clusters, or the like, will become more important in the next two chapters, where the model is transferred to a real urban system.

Chapter 10

Building a graph-CA model in a real urban setting

In this chapter and the next the graph-CA based model developed in chapter 8 and explored in chapter 9 is developed for a real urban area — Hoxton — in inner East London.

Two stages in building the model are covered in this chapter: preparing geospatial data and obtaining/synthesising attribute data. Difficulties in obtaining sufficiently detailed data (both spatial and attribute data) have resulted in an approach using synthetic data. This, in turn, suggested that the most appropriate application of the model in the context of this thesis, would again be to examine the effects of using different possible spatial structures for the model, and this is explored in the next chapter. Before proceeding to some of the more technical considerations of this chapter, a brief description of the study area is provided.[1]

10.1 The study area

Hoxton, in Hackney, East London, has been the subject of intense media and developer interest in recent years, as the latest 'hot area' in the inner London property market. Interest has coincided with the boom in London's international status as home to the 'sensation' of conceptual 'Brit-art' led by Damien Hirst, Tracey Emin, Rachel Whiteread, and Mark Wallinger amongst others. On closer inspection, the Hoxton

[1]Note that maps in this chapter and the next are based on the Ordnance Survey mapping data which is ©Crown Copyright.

real estate 'phenomenon' is less impressive than the associated media circus. Nevertheless, the notion that an artists' colony in the East End has been largely responsible for the reclamation of that part of the city is firmly established, and clearly borrows imagery and ideas from the earlier gentrification of New York's SoHo (Zukin 1982).

Like most neighbourhoods in London or any other city Hoxton itself is ill-defined. It can broadly be defined as on the 'city fringe' in South Hackney. The maps in figures 83 and 84 show the London Borough of Hackney, its neighbouring inner London boroughs, Islington and Tower Hamlets, and the City of London, in relation to the Greater London region. In spite of its proximity to the economic power house of the City of London, Hackney remains one of the poorest Local Authority areas in the United Kingdom, and has only shared patchily in Islington's famed 'yuppification' which has seen it rise to prominence as home to a sizeable proportion of the 1997 Labour Government's cabinet. To get a sense of this, see figure 85, which shows the Townsend deprivation index for this part of London (Townsend, Phillimore & Beattie 1988, see). Paler greys and whites are areas of higher relative deprivation (as measured by unemployment rates, overcrowding, non-home ownership and non-car ownership). Although the largest area of relative deprivation is in Tower Hamlets

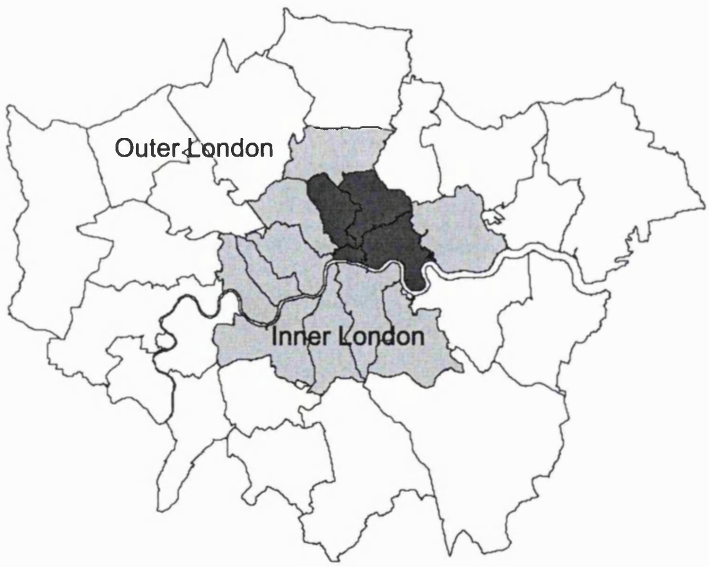


Figure 83 Hackney, Islington, Tower Hamlets and the City of London (dark grey) in relation to the Greater London region.

to the east, just north of the City in south Hackney deprivation is also severe. Note the large central expanse of Islington which is relatively well off — this area is centred around Islington's Upper Street and Highbury Fields, now thoroughly gentrified. Also note that the small resident population in the City itself prevents it from being classified on this index.

Broadly speaking, Hoxton is that part of Hackney immediately to the north of the City of London sandwiched between Finsbury/Clerkenwell in Islington to the west, and Shoreditch/Spitalfields in Tower Hamlets to the east. It is bounded to the north by the Regent's Canal. For the purposes of the current study this definition was restricted considerably, particularly in view of the high density of council owned social housing in the northern part of this area. As a result, the model study area is an approximate rectangle, bounded by the southern fringes of council owned housing to the north, Tabernacle Street to the west, Clere Street—Luke Street to the south and

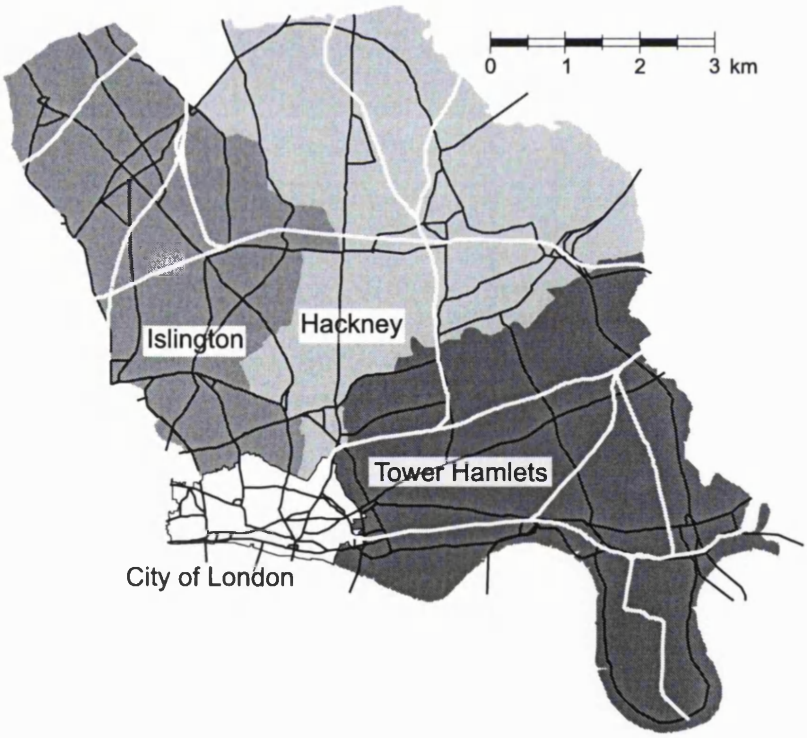


Figure 84 Hackney and its neighbours, with major roads (black) and rail lines (white).

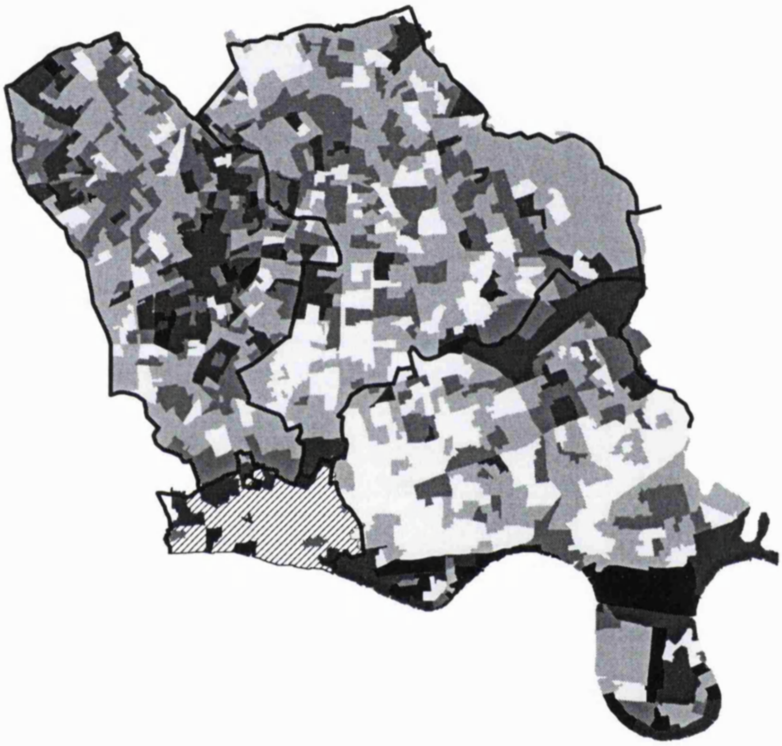


Figure 85 Relative Townsend deprivation scores for Hackney, Islington and Tower Hamlets (paler areas are more deprived).

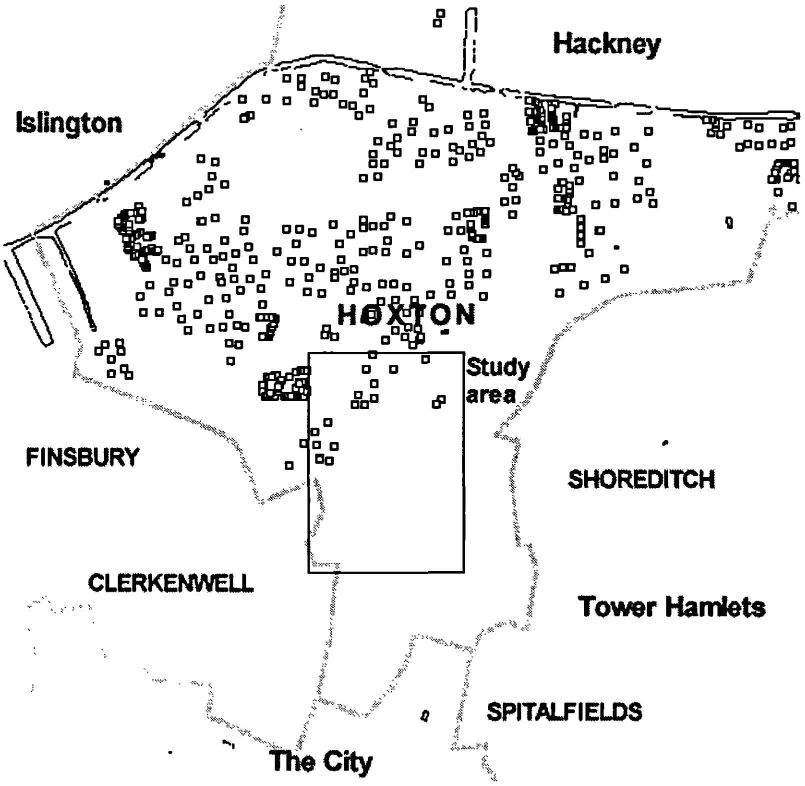


Figure 86 The study area in its context. Pale grey lines indicate local authority boundaries. The model study area is approximately contained within the marked rectangle.

a north-south line about halfway between Curtain Road—Hoxton Street and Shoreditch High Street—Kingsland Road to the east. The study area is indicated in figure 86 and, in detail, in figure 87. Figure 86 also shows the southern extent of the London Borough of Hackney's social housing stock[2] (the small black squares), south of the Regent's Canal.

It is worth noting that the model study area is more in keeping with current property industry notions of the whereabouts of Hoxton than longer standing local conceptions of the place, which certainly include the swathe of social housing to the north. Indeed, this kind of redefinition of place is arguably intrinsic to gentrification,

[2]This data was generously provided by Spencer Chainey, formerly the London Borough of Hackney's GIS manager. Hackney has well over 30 000 social housing units, around one-fifth of which are in the map area covered by figure 86. Many units are multi-occupancy and so may not appear so numerous on the map.

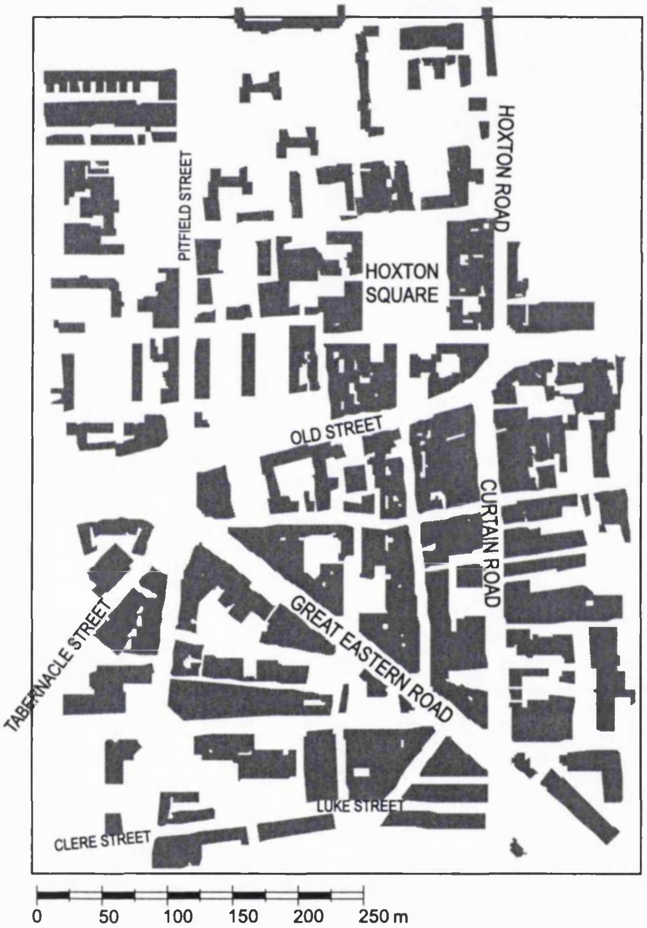


Figure 87 The model study area with major streets marked.

especially where the property industry itself plays a key role in the process as is happening in this case.[3] It is almost as though 'Hoxton' were a brand name intended to make the selling of property easier. Certainly this interpretation of the process is in keeping with the way in which the marketing of Hoxton has been closely linked to developments in the art world. This mobility of 'Hoxton' the place is remarked upon by Charles Jennings (2000, page 18), when he comments

> "A couple of years ago, Shoreditch north of Old Street was Hoxton. It was Shoreditch-Hoxton, piling up with designers and architects and beautiful people [...] And here they still are [...] Only now, south of Shoreditch is called Hoxton, and the geography has shifted around, so that Hoxton starts at Hoxton Square (so leafy, so Boho) and sprawls down below Old Street [...] all this is Hoxton: invest now! Get it while you can! A realtor's vision!
>
> "And the *other* Hoxton? The place on the map called Hoxton? Back up north? Hoxton Street? Another world: shops called Fags & Mags; Bits 'n' Bobs. Terse council estates; a guy picking through a heap of garbage left mid-pavement; kids patrolling the roads on chipped mountain bikes; the old Shoreditch Town Hall, bewildered and stained. But this place has lost its name to the concept down the road. You can't run your design studio up here, old Hoxton. You just have to sit tight, wait for the slow, incoming tide of money to wash up on your shore, think of a new name that sounds more right." [emphasis in original]

This quote also draws attention to the extremes of wealth and poverty which may often accompany gentrification and are a major element in more negative attitudes to the process.

Other significant factors in the choice of study area were a concern to ensure that the buildings in the selected area displayed a range of sizes and separations in order to ensure variety in the resulting graph-CA models, and that the number of buildings not much exceed 500. The most serious limitation on the model size is the ability of the urban morphology GIS component to calculate model structures in a reasonable time period with building counts much over 500. Were the model building processes to be implemented in native code, the performance of the *Gentrification* program would probably become the limiting factor.

[3] From personal conversations with Chris Hamnett at King's College London.

10.2 'Hoxton' on the ground

Green (2000), like Jennings, is sceptical about the media-led rush to invest in the artistic East End, and certainly seen on the ground it is less impressive than the hype. Very few artists' studios remain in the area and there is only limited evidence of their replacement by any concentrated development. According to Green only three studios remain in the area.[4] Only one of these is likely to survive impending rent reviews, and that is because StandPoint Studios happen to be the owners of their studio building in Coronet Street just off Hoxton Square. SPACE studios in Hoxton Road (see figures 88 and 89) is under threat from developments just up the road (see figure 90), and prices to buy and rent are clearly climbing (figure 91).

Although, there is still evidence of Hoxton's previous industrial incarnation as one of the centres of London's furniture industry (see figure 89), there is also plenty of evidence of commercial developments well underway, particularly those associated with the entertainment and creative industries. The Bean café (figure 92) and the Lux arts cinema (figure 93) are both now known throughout London, and the latter in particular has set the seal on the area's image — a self-consciously cool art world chic. But it is difficult to escape a feeling of edginess about the whole neighbourhood. Quite apart from its limited extent (only a few streets after all), and the evident deprivation in the council blocks which are never very far away, there are signs of resentment right in the heart of things (the photograph in figure 94 was taken on Charlotte Road, which is immediately parallel to Curtain Road). Even the supposed 'epicentre' of the Hoxton 'phenomenon' is not immune, as the boarded up café just a few doors up from the Lux (the photograph in figure 95 was taken in January 2000) bears witness.

In summary then, it now appears unlikely that the complete gentrification of (non-local authority owned) Hoxton, will ultimately 'fail', although it is at least *possible* that it could be left high and dry by any serious reverse in the booming London property market of the late-1990s and early-2000s. Further, these images and impressions are instructive in themselves. They certainly suggest that a realistic model, if stopped at any particular moment, would not necessarily show strong evidence of clear boundaries between gentrified and non-gentrified areas. Much of the time, an area in transition will be a confusing mess of old and new, developed, undevel-

[4]From personal conversations with Nick Green, to whom I am grateful for a guided walking tour of the 'artistic East End' in January 2000.

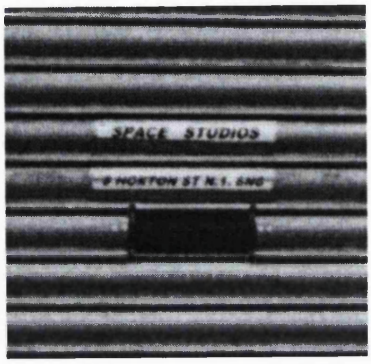


Figure 88 The entrance to SPACE studios in Hoxton Road.

Figure 89 SPACE studios (left) and the next door (former) cabinet making factory.

Figure 90 The upwardly mobile in Hoxton Road...

Figure 91 ...and the prices they're paying. Residential prices are around the £250,000 mark for about 1000 square feet at the time of writing, and are high, even by the standards of turn-of-the-(21st)-century London.

Figure 92 The Bean café — commercial life in Curtain Road.

Figure 93 The Lux cinema in Hoxton Square.

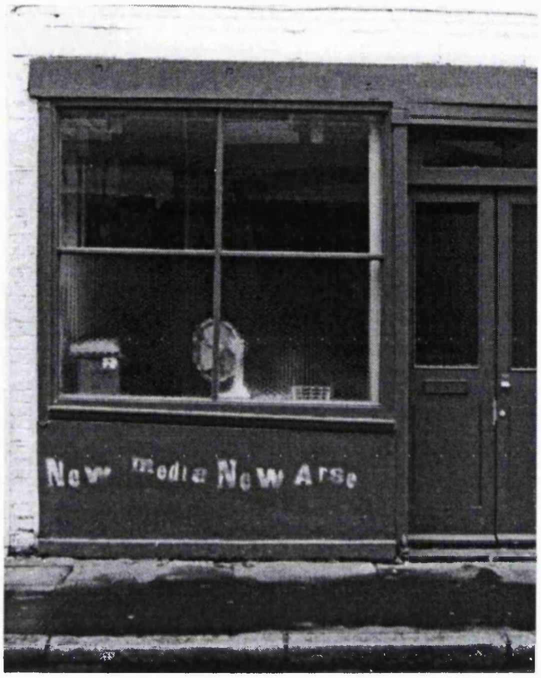


Figure 94 Signs of resentment (or irony?!) in Hoxton.

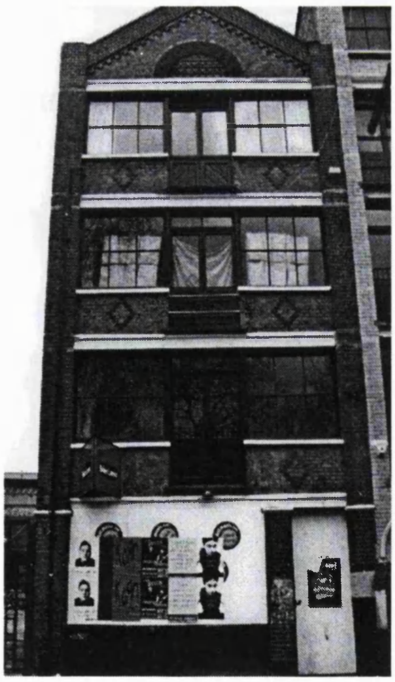


Figure 95 A boarded up café in Hoxton Square.

oped and yet-to-be-developed. This reality contrasts strongly with the rather clear-cut progress of segregation in a simple abstract model such as the majority rule graph-CA considered in Part II, and we would hope that the more complicated *Gentrification* model would show signs of this confusion and subtlety. The current state of affairs in Hackney more widely is also a challenge to Smith's (1996) idea of a clear cut gentrification 'frontier', although it might be argued that it is precisely the edginess of such confused areas that merits the description frontier at all. Experience suggests that however partial and confused the gentrification process may appear as it happens, ultimately, as Ruth Glass originally suggested, the process

> "goes on until all or most of the original working class occupiers are displaced, and the whole social character of the district is changed" (Glass 1964, page xviii)

This certainly seems likely to be the case in 'new Hoxton' south of the council estates.

10.3 Data sources for the model

Having settled on a real slice of urban space to model, it must be initialised with real world data. Two distinct types of data can be distinguished: spatial data describing and representing the urban fabric, and attribute data relating to locations in the fabric and representing our abstract variables, value and income. These are considered in the following two sections.

10.3.1 Spatial data

The UK Ordnance Survey's (OS) *Landline* digital data sets represent the best primary source for spatial data at the high resolution required for micro-scale simulation. Unfortunately, this data is primarily intended to support the production of map line work. This means that, for example, building footprints and plot boundaries are not usually stored as polygons, but as collections of line segments. Each line is associated with a numerical feature code so that it is possible to select all the lines which relate to buildings and plots. Even so, it is not trivial to further process these data to represent buildings or plots. Many blocks in the urban fragment are not subdivided, even where it is clear (from street number data also in the *Landline* data sets) that the block is not a single address.

As a result, a considerable amount of amendment and editing of the raw *Landline* data was undertaken to arrive at a GIS polygon layer which is a reasonable representation of the study area morphology. Given the abstraction of the *Gentrification* model's location state variables — which entails ignoring many of the important variables, such as floor space and land use — much of this editing was carried out on an *ad hoc* basis. Were more time available, a street by street survey could be undertaken to produce a more accurate representation of the study area. Note however, that it would not be possible to go very far down the route to accurate representation of the urban fabric before facing the issue of multi-occupancy buildings, and the relations internal to them. This leads immediately to the complexities of 3-dimensional representations of built forms, which are well beyond the scope of the present work. This also raises the issue of representing redevelopment in physical 3-dimensional detail, as for example, when a developer breaks a former factory up into small units, or when a new owner-occupier knocks former flats into a single family dwelling. This is a simulation task well beyond the current scope of the *Gentrification* model. Rather than claim absolute realism for the two-dimensional representation adopted, the pragmatic approach of building a representation at an adequate level of detail for the current purpose has been adopted. The final two-dimensional polygons used in the model can be seen in figure 96 on page 250.

10.3.2 Micro-scale property and household data

The detailed representation attempted in this model also faces formidable difficulties in its attribute data requirements, even with the relatively impoverished representation based on storing only 'value' and 'income' for each location.

The 'intrinsic value' notion of value which has been adopted is particularly problematic — although any other value measure which might be adopted would also face difficulties (recall the difficulties many have had testing the rent gap hypothesis cited in section 8.1.2 on pages 187ff.).

For example, prior to adopting Hoxton for study, the West End of Glasgow was also considered. Data on all property transactions in Scotland are readily available through the Sassines Register (unlike in England and Wales), and are being converted to digital forms by the Land Value Information Unit at the University of Paisley.[5] For

[5] I am grateful to both Nondas Pitticas at the Land Value Information Unit, for providing Glasgow

the Glasgow G12 postal district, there are around 11500 transactions recorded for the 11 year period from 1988-98 (inclusive). This was a particularly turbulent period in the UK property market, and it is immediately evident that considerable effort would be required to normalise these records, so that they could be regarded as directly comparable. The complexity involved is indicated by variations between different home-loan lenders' house price indices, and their interpretation (see Nicol 1996).

Further, as was remarked in chapter 8, the value of a property is a 'virtual' phenomenon, which is only realised when it changes hands at a price set in the market. However, even when market turnover is high it is rare for more than (say) 10% of the properties in *any* area to change hands in one year.[6] This introduces difficulties in developing an appropriate method for determining the value of the 90% or more of properties which did not change hands during the period for which data are available. Any such determination would require a method for normalising prices for variations in the characteristics of properties which *did* change hands. Only limited information about the numbers of rooms and their sizes, or the physical state of buildings is generally available. Since one of the aims of abstracting to a simple value figure is to reduce the data requirements, these complexities have simply been avoided by synthesising values from a proxy variable, as described in the next section.

The per capita household income is not so difficult to determine, at an *aggregate* level, given the detailed information available from the UK national census, for enumeration districts (EDs) of around 100–200 households. There is a considerable amount of work on the issue of disaggregating such data to produce realistic *micro-scale* data for simulation purposes (see Ballas, Clarke & Turton 1999 for a recent contribution, and Clarke & Holm 1987 for a review). Assigning synthetic households produced by such procedures to particular spatial locations — particular street addresses — is more difficult. A more difficult problem still, is the fact that the UK census does not provide actual *income* data for households in any case (Dorling 1999, makes some trenchant comments on this situation). Although some have suggested that the more detailed information provided by marketing surveys — which frequently do include an 'income question' — may represent a way out of this difficulty

G12 property transactions data, and to Iain Lake (see Lake, Lovett, Bateman & Langford, 1998) at the University of East Anglia for discussions as to its usefulness and applicability.

[6]The Glasgow G12 area, with its large student population, sees high turnover. According to the 1991 Census there are about 9750 properties (households) in G12 which could potentially change hands in the market (owner occupied or private landlords) so the 10% figure holds good in this case.

(Openshaw & Turton 1998), others are more wary (Harris 1998). Rather than develop what would amount to a complex sub-model to handle these problems, this issue has been fudged, as discussed in the next section.

10.3.3 Synthetic data sources

The two data required for each location in the model were generated as follows:

Value figures were synthesised by 'drilling down' into the Department of the Environment Transport and the Regions (DETR) *Index of Town Centredness* (ITC) surface data (Department of the Environment, Transport and the Regions 1998, Thurstain-Goodwin & Unwin forthcoming).[7] This is a composite data set developed on the basis of very detailed (and confidential) government data sets and intended for use in defining town centres for statistical monitoring purposes. The ITC takes into account parameters such as the amount of employment and turnover at different locations in urban areas. By 'drilling down' into a surface is meant a process whereby each location is assigned the value on the surface at which its centroid is located. For this particular area of London, values of the ITC surface are in the range 8.6 to 25.8. These numbers have been rescaled to a more appropriate range for *Gentrification* model purposes by multiplying all values by a factor of 3.0.

Income figures were set equal to the value figures summed with a normally distributed random offset ($\mu = 0, \sigma = 2.5$). This has the effect of making the model start from a situation where value and income are approximately matched. Alternative approaches based on disaggregating data from (say) the deprivation index, or another derived census variable were rejected as likely to be equally arbitrary, and also more likely to produce large initial disparities in the value and income at particular locations, which would be likely to strongly influence subsequent development of the model dynamics. The almost matching approach, on the other hand, ought to be fairly neutral in this respect, so that subsequent developments are more likely to be influenced by the spatial configuration of the model in question, which is the object of the current investigation.

[7] Town centres material was made available by kind permission of the Department of the Environment, Transport and the Regions.

The resulting overall location value distribution in the study area is shown in figure 96. Incomes closely match this spatial distribution, and the initial system configuration in state space is also shown.

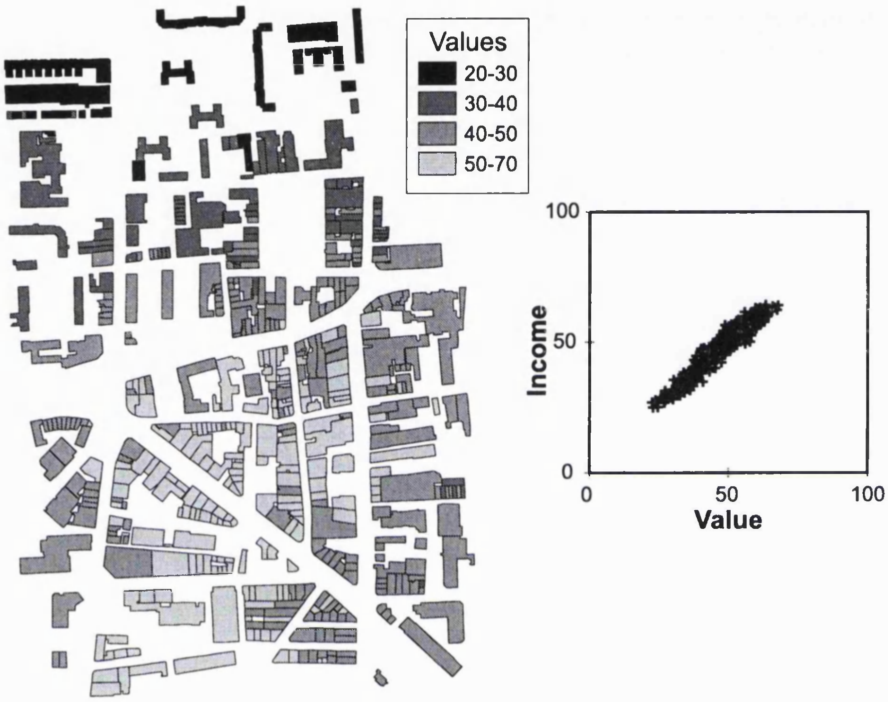


Figure 96 The distribution of values resulting from the methods described in the text, and the starting configuration for the Hoxton model in location state-space.

10.4 Conclusions

In this chapter we have seen that whilst it is certainly feasible to build a graph-CA model of an urban fragment at a high level of detail, the process is fraught with difficulties in terms of the availability and resolution of data. This applies to both spatial and attribute types of data. Problems with spatial data in this case are largely due to the type of spatial data considered in this case (buildings), and might well be less severe for other examples — the road network for example. Buildings present particular difficulties in the UK, given the map line-work form of the large-scale national digital data, although multi-occupancy buildings would present serious problems for any two-dimensional approach in a dense urban area. The attribute data required in this case are also particularly problematic in the UK, with no census 'income question' and (in England and Wales) no extant cadastral system. Once again, however, these data would present difficulties in any case because of the required household level resolution, and the 'virtuality' of house prices — indeed the general difficulty of the concept of value in the built environment.

In the current context of exploring the feasibility of graph-CA based micro-scale urban modelling these difficulties are not insurmountable. Given sufficient time and resources, detailed models of this kind could certainly be built, based on more convincing data sets than those used here, even if privacy laws are always likely to hinder micro-scale disaggregate modelling in much of the developed world.

However, as will become clear in the next chapter, where the model built in this chapter is actually run, the dynamic behaviour of complex models themselves is likely to be a more significant, and ultimately unavoidable, difficulty for micro-scale modelling, than the data problems faced in any particular case study.

Chapter 11

Running the *Gentrification* model of Hoxton

We can now apply the *Gentrification* model to the Hoxton study area. In keeping with the focus elsewhere in this thesis, and in light of the data collection problems discussed in the previous chapter, it was decided that an appropriate approach was to observe variations in model outcomes across a range of different proximal models based on the study area. These were built using the urban morphology tools already briefly discussed in section 9.1.2 (see pages 213ff.) and described in more detail in the first section of this chapter.

The main body of this chapter then consists of a series of observations made on the dynamic behaviour of each of these differently structured graph-CA *Gentrification* models. This discussion highlights a major problem facing all models of this type. Presenting the results of such dynamic, irregular spatial models is extremely difficult. In the thesis format there is not much choice except to present a series of snapshot images of typical model behaviour along with a verbal description.[1] The visualisation of dynamic outcomes is a major research issue which faces all complex modelling efforts (perhaps complex *spatial* modelling, especially), and an issue to which I return in the final chapter.

Before proceeding to a description of various runs of the Hoxton graph-CA model, one further modification to the *Gentrification* model was made to make provision for 'fixed-state' locations. These are locations whose current state may influence

[1] This is not as limiting as it sounds, however, since it is not so *very* different from the author's experience of just watching what the model does!

those around them by the normal CA rule process, but whose states do not themselves change as the model runs. Fixed-state locations were introduced to circumvent the dramatic edge effects which can occur in a model of this kind. This approach has been used for similar reasons on conventional urban CA models (White & Engelen 1997, page 327, for example).

11.1 Hoxton graph-CA model structures

As briefly discussed in section 9.1.2, a GIS extension for constructing various urban morphology based graphs has been written, and used to produce the graphs discussed in this section. It is first worth noting that all the graphs discussed, and the accompanying construction methods, use *polygon centroids* as point locations representing buildings. This simplifies the determination of the graph edges considerably, but leads to some mis-specification of graph edges — that is, the graphs may not precisely fit their loose verbal descriptions. Where relevant, this is considered in the discussion accompanying the graph description for each method below.

As a prelude to this discussion, the various graph structures have been characterised in terms of Watts's small world parameters (Watts 1999). Recall (see section 3.2.4 on pages 54ff.) that Watts suggests two measures for the overall characterisation of graph structure: characteristic path length, L, and average clustering coefficient, γ. Here we consider the distributions of these two measures, determined locally for each vertex, in the graphs. The distributions are shown in figure 97. These are based on 515 vertex graphs, fragments of which are shown in the figures in the next four sub-sections.

Since determination of these measures is problematic for disconnected graphs, all graphs are based on a 'Delaunay substrate'. This means that all the graphs considered include all the edges from the Delaunay triangulation graph, together with additional edges constructed according to the various methods. Since all locations in a point set have neighbours in a Delaunay triangulation, regardless of the remoteness of 'nearby' points, this ensures that all the graphs considered are connected, with no isolated components. This characteristic is also important in allowing comparison of different model behavioural outcomes, since disconnected subgraphs would necessarily behave differently in different cases, and so confuse the overall pattern of results.

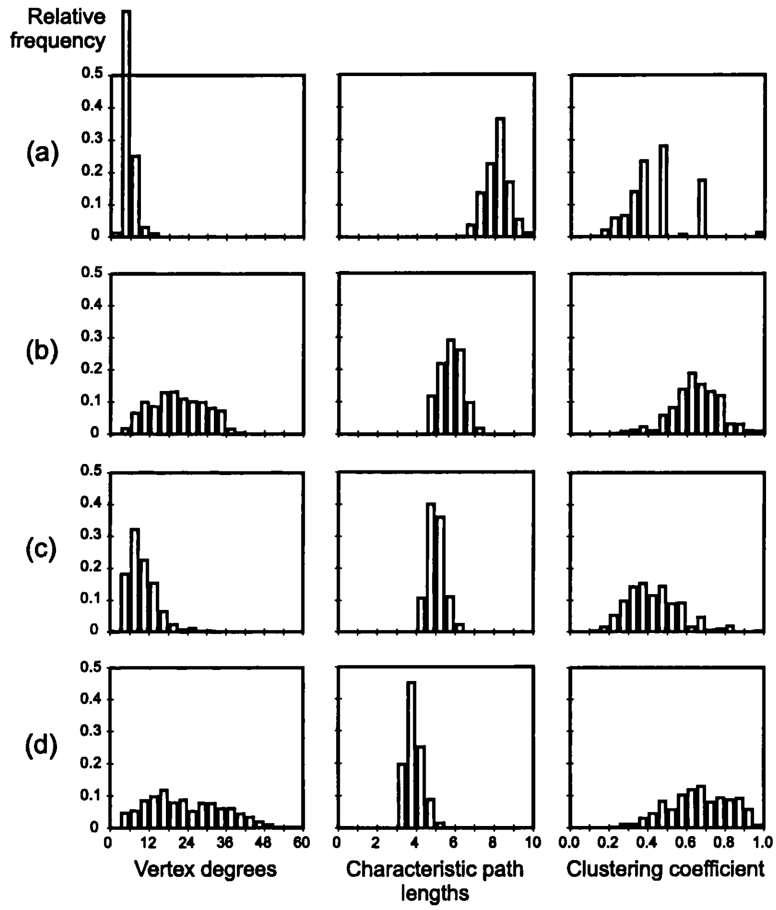


Figure 97 Distributions ($n = 515$) of the small world measures for (a) the Delaunay graph, and the Delaunay graph logical OR-ed with (b) the distance graph, (c) the visibility graph, and (d) the street segment based graph. Fragments of the graph on which these measures are based are shown in figures 98, 99, 100 and 101.

11.1.1 Delaunay triangulation based graphs

The Delaunay triangulation is well known in spatial analysis and other fields. It is the dual of the Voronoi (or Thiessen) tesselation, whereby each point in a set is associated with a proximity polygon consisting of all those locations which are nearer to that point than any other in the set. Many Delaunay triangulation algorithms rely on this duality, and these — and others — are thoroughly reviewed by Okabe, Boots & Sugihara (1992). The characteristic structure of a Delaunay triangulation graph is seen in figure 98. Such graphs are planar, and this usually restricts the neighbourhood of any vertex to relatively low numbers of vertices (in this case a maximum of 15). The upper bound on the average degree of vertices in a Delaunay triangulation is 6, so typical vertices have 5 or 6 neighbours. This results in long characteristic path lengths in the graph, which will tend to grow with the square root of the number of vertices in the graph, $\sqrt{n}$, since very few 'short-cuts' are present between spatially remote parts of the graph.

The clustering coefficients of vertices in a Delaunay graph are fairly low. The high values evident in figure 97 are attributable to small neighbourhoods with $3 \leq k \leq 5$, when it is not unusual for the vertex neighbourhood to be isomorphic to the cycle graph C_k, which results in $\gamma = k/\binom{k}{2} = 2/(k-1)$. For example, a vertex with

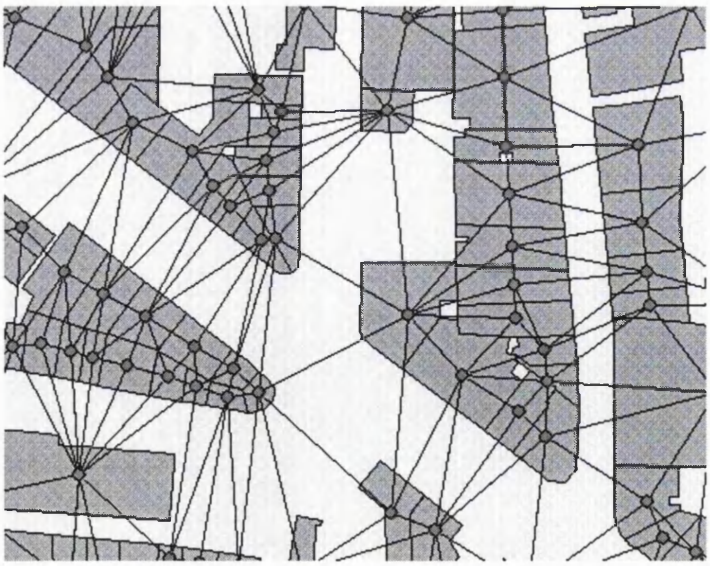


Figure 98 The Delaunay triangulated graph structure.

three neighbours will frequently have all three connected to each other, so that the neighbourhood sub-graph is $C_3 \equiv K_3$, with a clustering coefficient of 1.0.

11.1.2 Maximum distance based graphs

Graphs can be constructed on the basis of some user-specified maximum distance criterion, whereby vertices located at the centroid of building polygons are joined by an edge if they are less than some specified distance apart. A fragment of a typical graph of this type (maximum distance 50m) is shown in figure 99. This is a much denser graph than the pure Delaunay graph above. This density results in a wider range of vertex neighbourhood sizes and much shorter characteristic path lengths. Relatively high clustering coefficients are also evident because neighbouring vertices' neighbourhoods tend to overlap due to the spatial construction.

We could estimate the expected clustering coefficient using Clark & Evans's (1954) results for the separation distance of n randomly distributed points. At areal density ρ, the mean separation distance is $\sqrt{\rho}/2$, with standard deviation $0.261/\sqrt{n\rho}$. Two equal circles of radius R, with centres separated by distance d, have an intersection area, easily derived from the lens area formula (Weißtein 1996), which is a

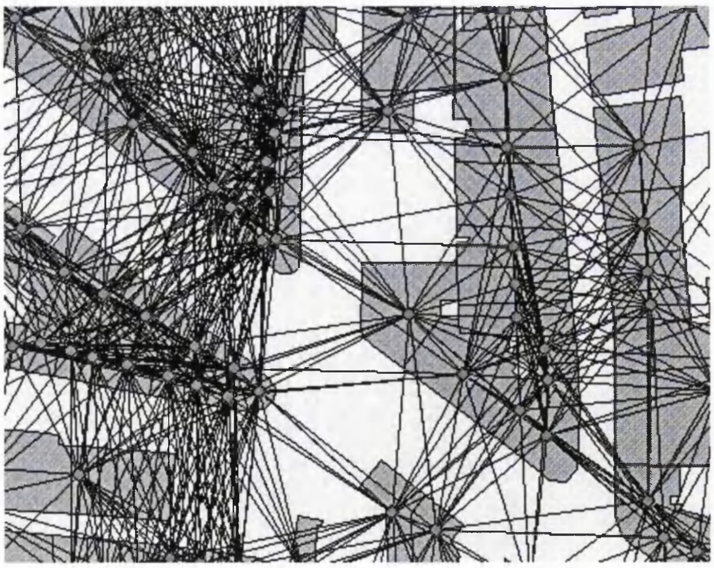


Figure 99 The graph based on a maximum distance criterion and ORed with the Delaunay graph of figure 98.

fraction a of each circle area, given by

$$a(x) = \frac{2}{\pi}\left[\arccos\left(\frac{x}{2}\right) - \frac{x}{4}\sqrt{4 - x^2}\right] \tag{11.1}$$

where $x = d/R$ and the inverse cosine is expressed in radians. Substituting an appropriate distribution of values of x from Clark & Evans's results into equation 11.1 ought to allow determination of an estimate for the clustering coefficient in a spatial graph, with the distance criterion is set to R.

11.1.3 Mutual visibility based graphs

A graph based on the mutual visibility of buildings is shown in figure 100. Here two buildings are joined by an edge if they can be seen from one another. Because the graph shown has been constructed depending on whether the building *centroids* can be joined by a straight line which does not intersect any other building polygon, this is only an approximation to a mutual visibility graph. However, it has the virtue of being well defined. More complex definitions would require that building heights and street elevation also be considered. They would also require that consideration be given to how many lines-of-sight between buildings are necessary (at

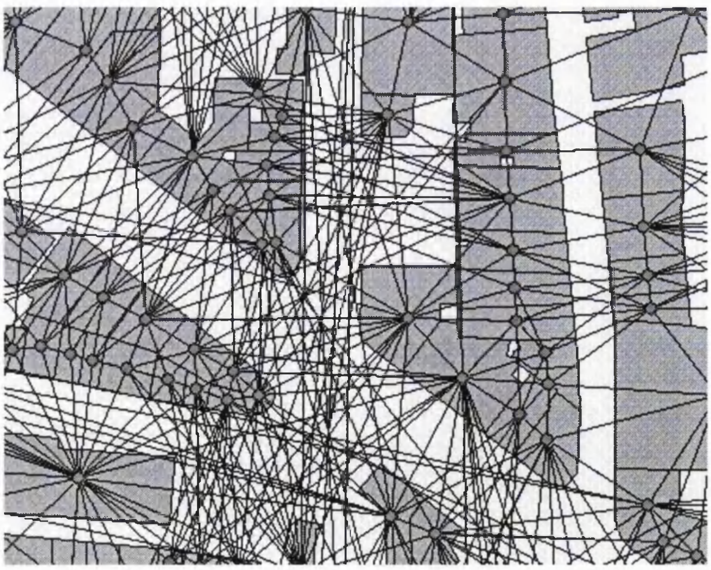


Figure 100 The graph based on a mutual visibility criterion and ORed with the Delaunay graph of figure 98.

some resolution) for them to be considered mutually visible. This leads immediately to 'data-hungry' considerations such as the number of windows, and their position in the building, and so on. Practically speaking it is much easier to opt for an approximation to a 'true' mutual visibility graph, than to build such a complete version. Furthermore, such a true mutual visibility graph would probably have to be a weighted graph to reflect the complexity of these issues, and no provision has been made for this in the *Gentrification* model.

This graph also exhibits a wider range of neighbourhood sizes than the base Delaunay graph, but less so than the distance graph. Nevertheless, the characteristic path lengths in this graph are shorter than in the distance based case. The lower clustering coefficient evident in the statistics accounts for this difference, insofar as it implies that more new options are available in constructing a shortest path at each 'step' along a shortest path, so that more remote vertices are reached relatively quickly. Of course, the simple fact that edges in this graph have no restriction on their lengths also leads directly to this conclusion.

11.1.4 Street segment based graphs

An even more dramatically 'foreshortened' graph is based on street segments (figure 101). Here two buildings are joined by an edge if their 'entrances' are mutually visible. This example has been constructed on the basis of artificially generated entrance points, although detailed surveys could also obviously be used to construct an accurate graph of this type. A simplistic rule for entrance placement was used based on the proximity of street centre-lines and the longest outside face of a building. This rule normally produces entrance locations in reasonable locations, particularly for buildings organised in terraces with only one facade on a street, which is the case for many buildings in the study area. The major difference between computing mutual visibility based on 'entrances' and simple mutual visibility is that it biases the lines of sight so that they tend to follow the street network.

Spatially then, this graph has edges which strongly follow the street pattern — as we would expect. This is the densest graph considered here, as can be seen from the distribution of vertex neighbourhood sizes. It also has the shortest path lengths — but relatively high clustering coefficient values (higher than the building visibility graph, and comparable to the distance based graph). As such there may be grounds

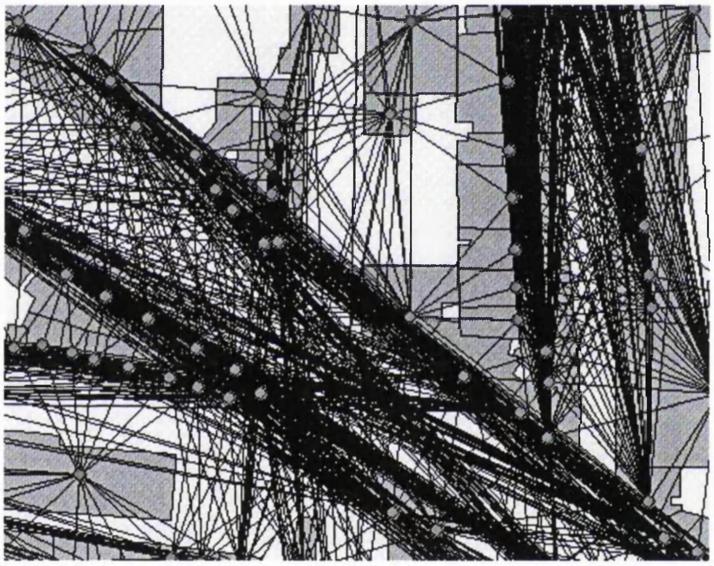


Figure 101 The graph based on street segments and OR-ed with the Delaunay graph of figure 98.

for considering this graph to be a small world, although the number of edges would also have to be considered in the analysis to confirm this suggestion, since we expect high vertex clustering coefficients in a dense graph.

11.1.5 Implications of structural properties for dynamic behaviour

In the current context the important question to address is how, if at all, are the graph properties likely to affect the dynamic behaviour of a graph-CA process model — and specifically the *Gentrification* model of chapter 8 — which runs on them? It is worth pausing to consider what we might expect to happen, in the light of the remarks on graph structure above, before looking at actual behaviour.

Various dimensions of variation in the graph structures are evident:

- There are differences in the *size and range of variation in vertex neighbourhoods.* These will interact with those parts of the process model where gaps between maximum, minimum and central vertex values affect outcomes. Other things being equal, larger neighbourhoods seem likely to exhibit wider gaps and therefore the potential for more unstable or chaotic behaviour. Graphs with wider variation in neighbourhood sizes are likely to witness greater variation in locational stability as a result of this. To the extent that vertices with larger neighbourhoods are clustered together or not, such effects taken together might be expected to result in distinct 'regional' differences across a system.

- Variations in the *distribution and length of shortest path lengths* have also been noted. These are likely to relate to the speed at which any state changes propagate across a system (this is the import of Watts & Strogatz's, 1998 findings on disease transmission on small world networks). Thus we would expect neighbourhood change in a Delaunay graph system to propagate more slowly across the system than across the street segment system.

- Closely related to this supposition, *variations in clustering coefficient* seem likely to affect the 'steadiness' of the progress of a change across a system. In a system with highly clustered vertex neighbourhoods, adjacent vertices are likely to have similar neighbourhood states, and thus to undergo similar sequences of locational state changes at the same time. In a less clustered case, a state change in one location might turn out to be quite transient, since surrounding neighbourhoods are potentially very different in character, and not subject to the same pressures for change. This ought to lead to less certain progress of changes across a system.

All these observations would apply absolutely in a completely deterministic process model, and presumably could represent strong tendencies in a probabilistic model. These ideas are difficult to confirm, and an *analytic* investigation (including the development of measures and tests for the tentative hypotheses above) is beyond the scope of the current investigation. In the next section, therefore, I concentrate on qualitatively describing the evolution of the *Gentrification* model on the model structures described, starting from the initial synthetic configuration which has already been presented (figure 96, page 250). However, the foregoing discussion does illustrate the strength of the graph-structure/CA process-dynamics split (central to this thesis), for generating hypotheses, intuitions and speculations on these fundamental issues. These ideas have also influenced the discussion of the model results throughout this chapter.

11.2 Description of model outcomes

In this section, the dynamic evolution of the *Gentrification* model is presented for the model structures described. Initially only one set of model parameters ($p_0 = r_D = r_M = 0.05, k_A = 0.10, k_G = 0.05$) is presented for all four model structures. Other parameter settings are briefly considered later in the chapter.

11.2.1 The Delaunay triangulation based graph

In figures 102 and 103 the evolution of the *Gentrification* model on the Delaunay triangulated graph structure is shown over 300 time steps of evolution from the initial conditions shown in figure 96 on page 250.

It is immediately evident how difficult it is to describe the dynamic behaviour based on 'snapshots' like these! Over the whole sequence there is a distinct polarisation in the system, reminiscent of the rather unexpected outcomes when less local structure was introduced to the regular arrangement of cells in section 9.5 (pages 227ff.). Also, there is a distinct tendency for parts of the model to move together. This

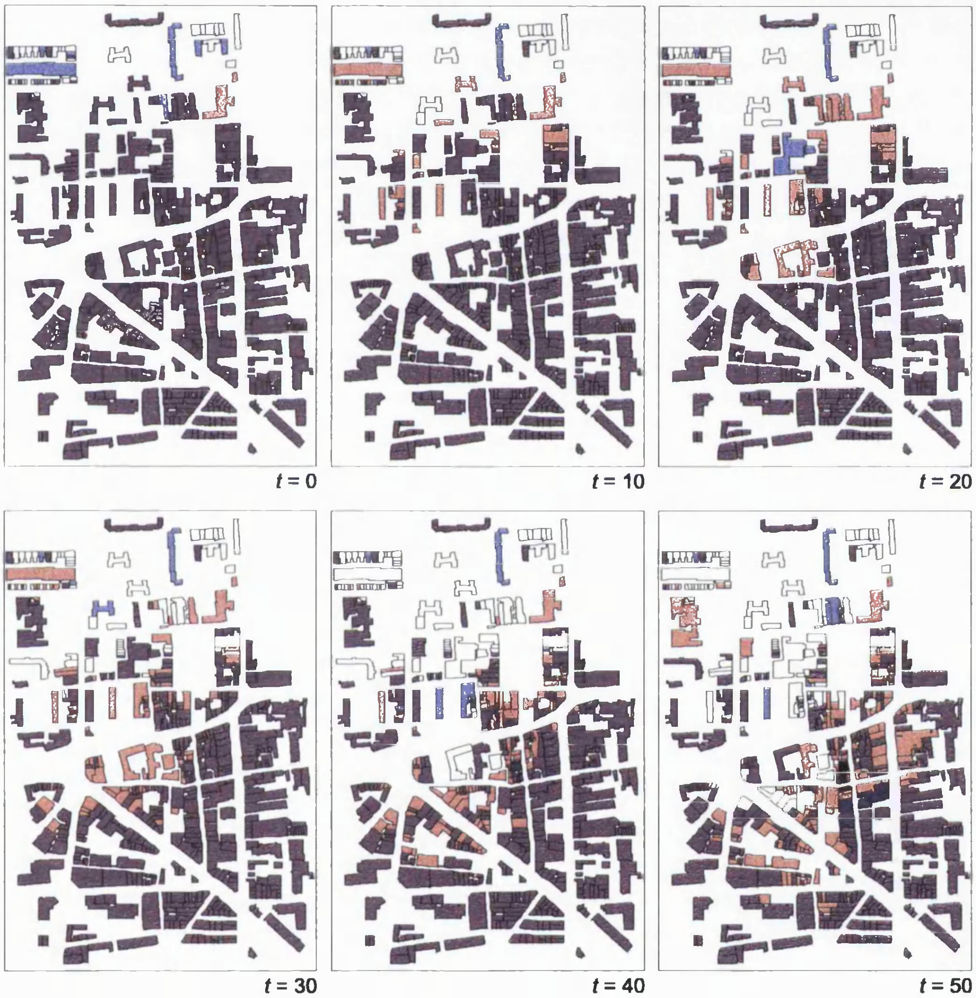


Figure 102 The evolution of the *Gentrification* model from $t = 0$ to $t = 50$ on the base Delaunay triangulated graph structure.

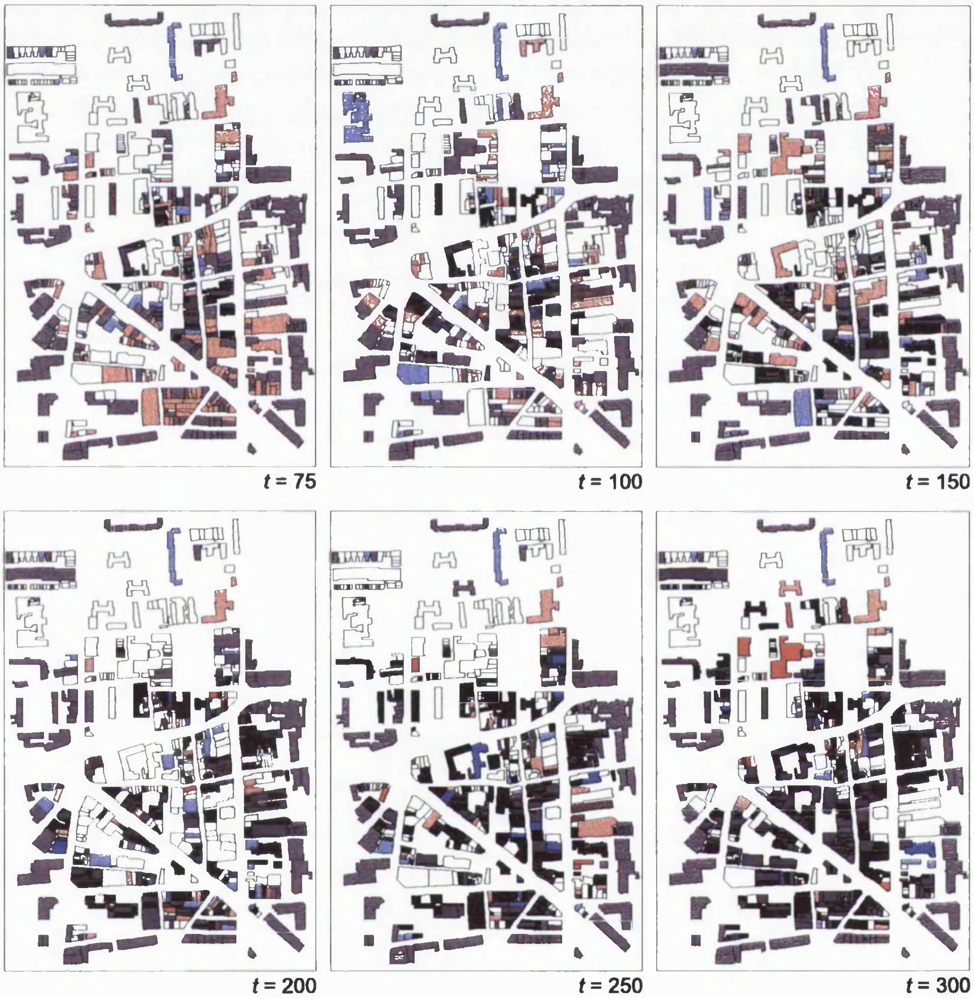


Figure 103 The evolution of the *Gentrification* model from $t = 75$ to $t = 300$ on the base Delaunay triangulated graph structure.

is particularly clear over the sequence from $t = 75$ to $t = 150$, when the 'gentrification' of the system seems to really take hold in the region of the model just south of centre, on the north side of Great Eastern Street, and along Charlotte Road in particular. The progress of the system to the point where most locations are 'up-market' by $t = 300$ is by no means assured, however, as can be seen from the snapshot at $t = 200$ where there has been some retreat, together with a northeast-ward shift in the centre of activity. It seems likely that edge effects are significant in the model here, as the anchoring of the gentrified locations to the eastern, fixed-value edge of the system around $t = 200$, seems to make the further gentrification of the remainder of the system 'more secure' over succeeding time steps. This unsteady progress of gentrification in this model may be a result of the relatively low clustering coefficients in this graph structure as conjectured above.

11.2.2 The maximum distance based graph

Results over 300 time steps at 50 time step intervals are shown in figure 104 for the maximum distance based graph, where the distance criterion is set to 50 metres. The system in this case still ends up predominantly gentrified, but much more quickly than before, a direct result of the shorter characteristic path lengths already noted. This is clearest when the snapshots at $t = 150$ are compared for the two cases. There is also no obvious interim reversal in the 'upwards' trend, except in the initial formation of 'niches' as many locations all across the system drift down in value and income. It is this downward drift which creates potential attractive locations for incoming households of higher income.

The locations which are gentrified first are in different locations in this case, towards the south and eastern edges of the system, up until about $t = 150$, after which gentrification 'advances into' the central and northern areas of the system. Thus, although the end result is similar, the spatial progression of the system towards full gentrification, is different in this case. This suggests that sufficiently varied spatial structures may 'shape' the model gentrification process, so that the stochastic effects at particular locations have a less pronounced impact on the overall outcome, although this is impossible to confirm from just one case. Of course, this result could also be seen as supporting the view that the most important input overall is the model parameter settings, which govern the overall trajectory of the system towards either

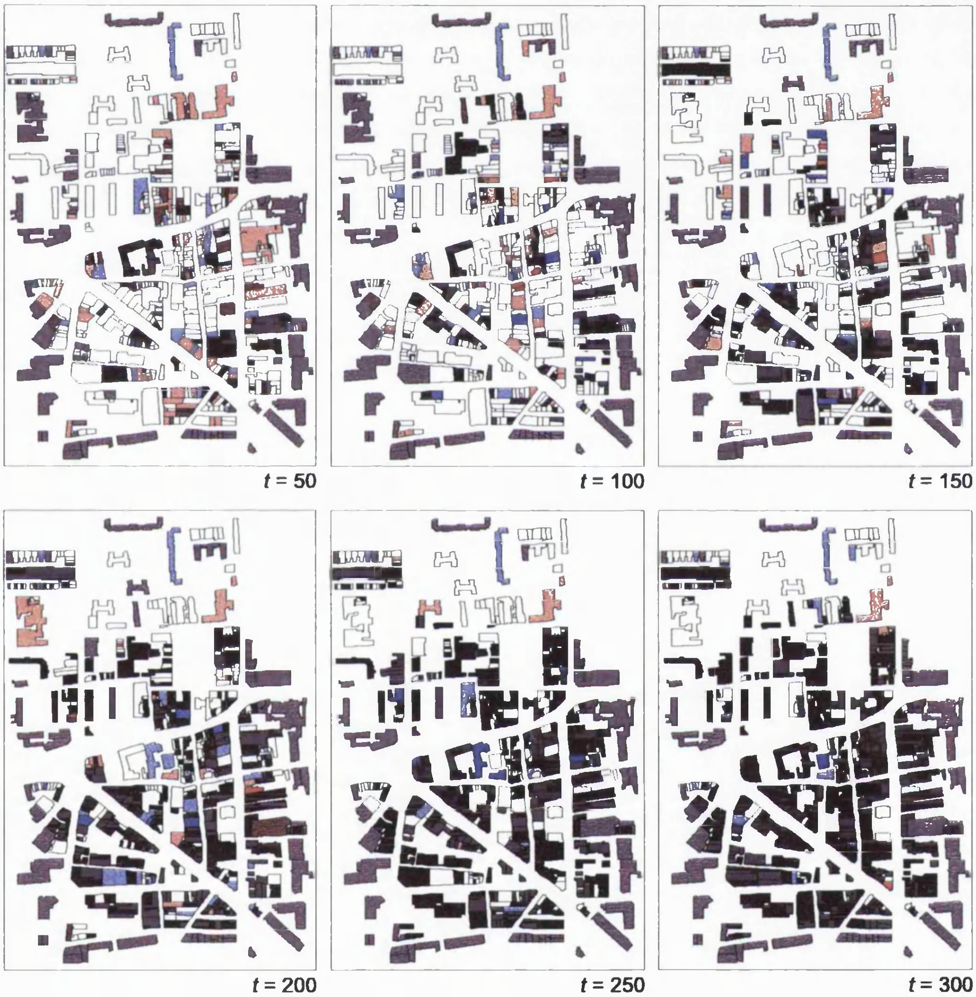


Figure 104 The evolution of the *Gentrification* model from $t = 50$ to $t = 300$ on the maximum distance (50m) graph structure.

'regeneration' or 'decline'. Thus, although the distance based graph results in more rapid progress to full gentrification, the eventual outcome in both these cases is similar.

On the basis of this and the previous case, it is more evident that the dynamics of the model depend on a very particular mechanism. The system only gentrifies *after declining* to the point where many properties are low-value and low-income. In both cases this decline is most pronounced in the northern half of the system, which is unsurprising given that the fixed-value boundary locations to the north are rather less well off than those to the south. Having seen this decline, numerous investment niches appear between the better and less well off regions, and this is where the resurgence, and eventual dominance of high-value, high-income locations has its genesis. This aspect of the system mechanics was hidden in the earlier consideration of abstract grid-based cases in chapter 9, because of the initial two-region distribution considered in all cases, which effectively had many investment niches in the system from the start. Although such relative decline is almost a definitional prerequisite for gentrification to occur, its occurrence in such a localised manner in the model seems a little unconvincing, and largely a product of the rule summarised in equation 8.8 on page 201 which will *usually* lead to decline in a location with low values of k_G, but will also occasionally result in significantly richer households moving into low-value properties. These are the important events in the model in terms of sparking a series of upgrading events, because richer incomers then find it easier to obtain finance for improvements. This advances the gentrification 'frontier' locally and moves the next investment niche a little further forward into the less affluent regions of the model.

11.2.3 The mutual visibility based graph

Similar results for the mutual visibility based graph-CA model are shown in figure 105. The process here proceeds at a similar rate (perhaps even a little more quickly) to the maximum distance graph, and certainly more quickly than on the simple Delaunay graph. It is difficult to discern any strong differences in the spatial pattern of change between this and the maximum distance case. At $t = 50$ there is less gentrification in evidence in Hoxton Square, and more in Great Eastern Road. However, by $t = 100$, the overall pattern is broadly similar, and from this point on there is not much to distinguish the two cases.

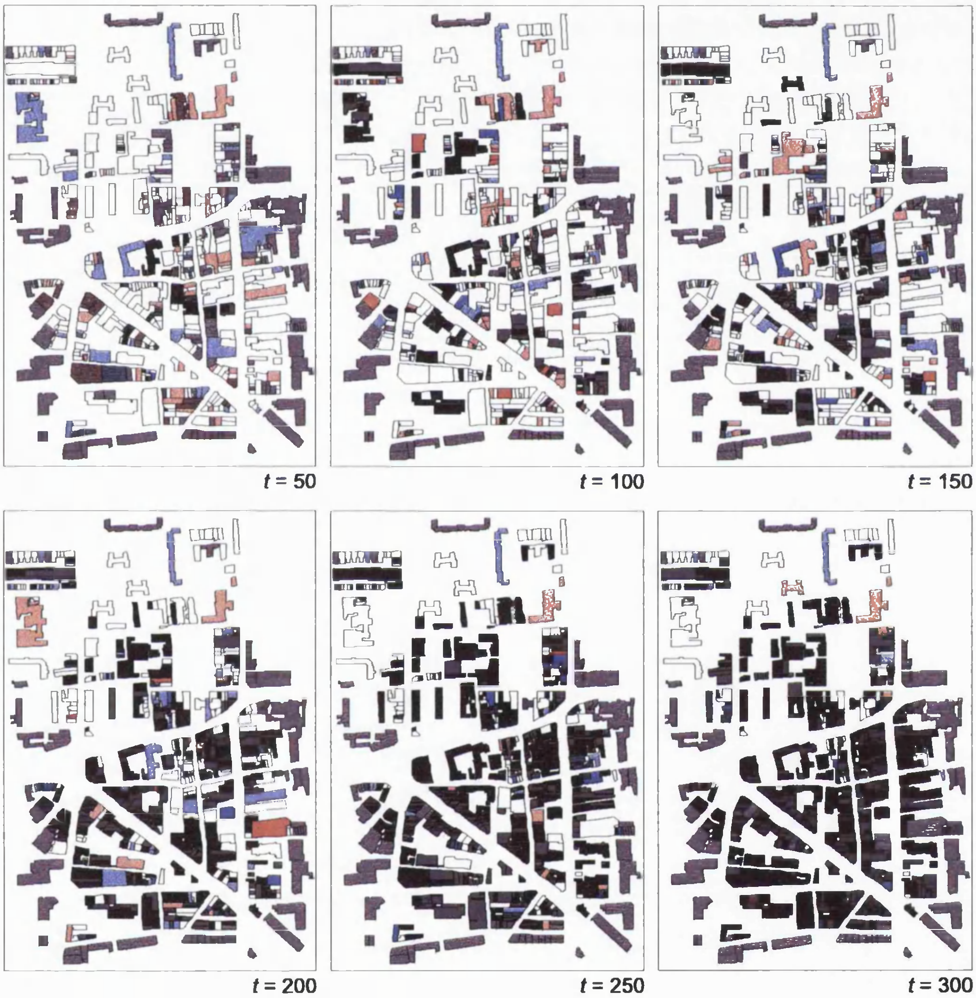


Figure 105 The evolution of the *Gentrification* model from $t = 50$ to $t = 300$ on the mutual visibility graph structure.

11.2.4 The street segment based graph

Finally, the progress of the model over the first 300 time steps is summarised in figure 106 for the street segment based structure. Again the model behaviour is accelerated from the simple Delaunay triangulated case, and broadly similar to the pattern in the other two cases. Model development again proceeds to a point where most of the non-fixed locations in the model are high-value and high-income.

In all three of the denser graph structures, this stage is not a final equilibrium, with no further change occurring, because of the continuing possibility of lower income households moving in, as current occupants probabilistically leave. However the general pattern at this point is that only small numbers of households move out, and these are generally replaced by similar income households. Around the edges of the system, lower income households do occasionally move in, but are often replaced soon after by high income households. The net effect is of a fairly stable final state being reached, with a small amount of visible activity 'around the edges'.

It is a *little* surprising, in the 'street segment' case, that gentrification does not proceed along roads and streets in any obvious way. This is probably because vertex neighbourhoods are often very large in relation to the overall model size, so that, although progress is street-by-street, it is not necessarily *building-by-building* along any particular street. The overall effect, actually appears almost random, as might be surmised from the fairly general 'spotty' gentrification visible between time steps 100 and 150 in figure 106. In the context of a much more extensive model, where the large neighbourhoods in this structure would be *relatively* smaller, it is likely that gentrification would not give the appearance of 'jumping around' quite so dramatically as here, but would instead be more local in character.

11.2.5 Discussion

These cases suggest that the single most important effect of the model structure on the operation of this particular spatial process is the characteristic path lengths (refer back to figure 97 on page 254). Because the overall process is very much like a diffusion process (albeit, in a disguised form), the shorter characteristic path lengths from each location to every other in the latter 3 cases promote rapid diffusion of lower status attributes in the initial phase of the model's operation, and also of higher status attributes, when the consequent appearance of 'investment niches' on the boundary

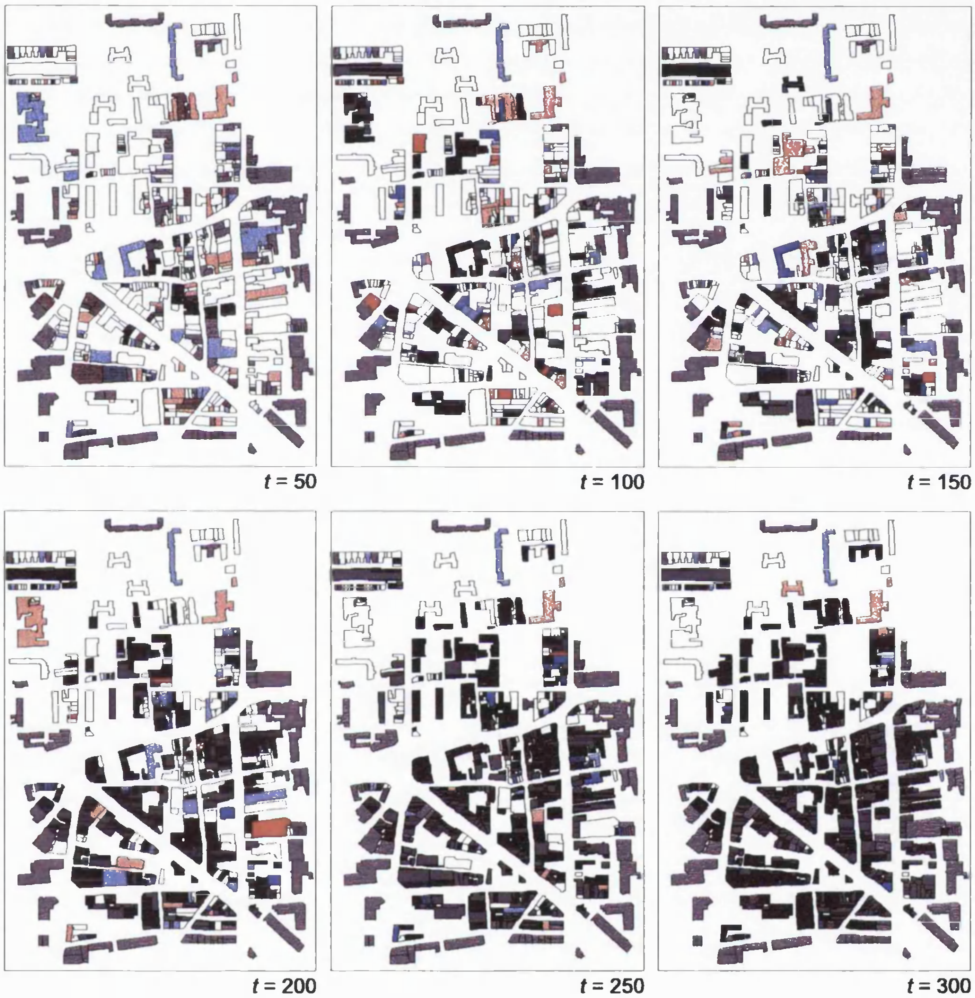


Figure 106 The evolution of the *Gentrification* model from $t = 50$ to $t = 300$ on the street segment structure.

between higher and lower status areas results in gentrification.

There is also a suggestion in the uncertain progress of gentrification in the Delaunay case, that low vertex clustering coefficients may produce such an effect on a diffusion-type process, for reasons already discussed in section 11.1.5 — that is the likely similarity (difference) in adjacent vertex neighbourhoods arising from high (low) clustering coefficients.

There are no strongly distinctive spatial patterns apparent in the differences between the various model structures, and by the same token, the overall trajectory of the system does not vary greatly between the various structures. We can confirm that the spatial distribution of average path lengths in each structure is indeed different by inspecting figure 107, where this graph structural variable is mapped. Lighter colours represent the shortest average path lengths associated with the most central (accessible) locations in each case. The most striking pattern here is the strong concentric pattern of centrality in the maximum distance graph case. The patterns are less globally ordered in the other three cases but clearly different from one another. There is no obvious mapping between these differences and differences in the evolution of the gentrification process in the various examples we have considered. On this basis, there is no evident 'steering' or 'shaping' of the model by the underlying spatial structure, and apparently no distinctly spatial effects. This suggests that the most important determinants of global system behaviour are the model's global parameters — k_A, k_G and so on — with 'space' playing relatively little part. It also suggests that the impact of stochastic events (individual household decisions) in this model is relatively limited, since the overall tendency of the system, with these particular global parameter settings, is more or less inevitable.

Variation in model outcomes for different global parameter settings is shown in figure 108. Only two of the structures are included here, since we have already observed that behaviour can be broadly divided between the 'slow' Delaunay structure and the 'faster' other structures, here represented by the mutual visibility graph case. Only a single snapshot is shown at $t = 150$, for $k_G = 0.025$ and 0.10 respectively for the two examples. Reference back to the same snapshot in figures 103 (page 263) and 105 (page 267) reveals that the difference is just as we might expect: gentrification is more advanced with $k_G = 0.10$, and much less so with $k_G = 0.025$. Again, no strong spatial effect is evident.

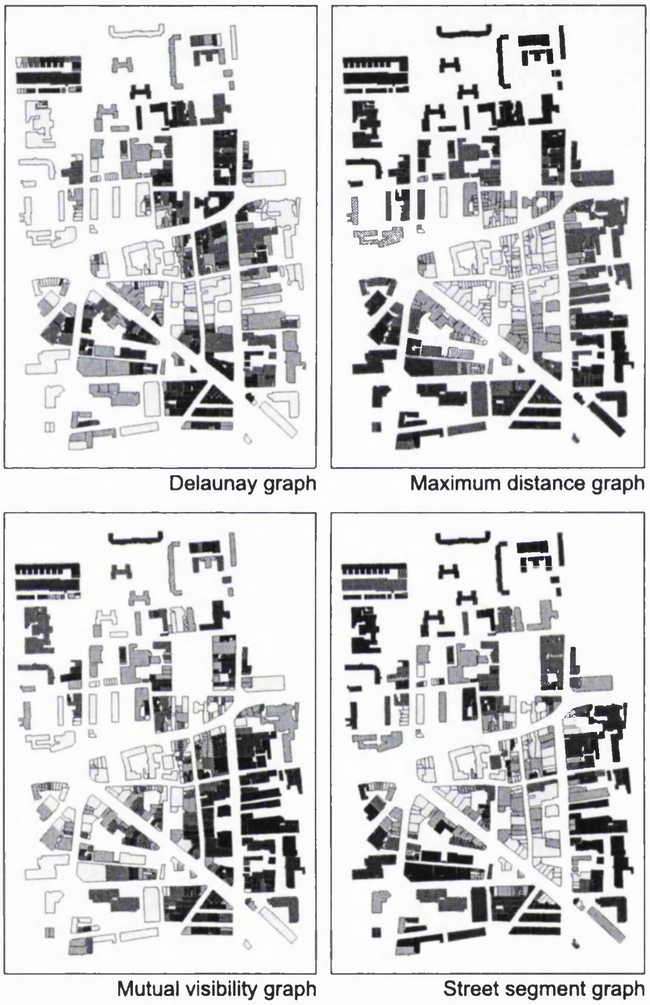


Figure 107 The distribution of characteristic path lengths for the four different cases. Light greys represent short characteristic path lengths (more central).

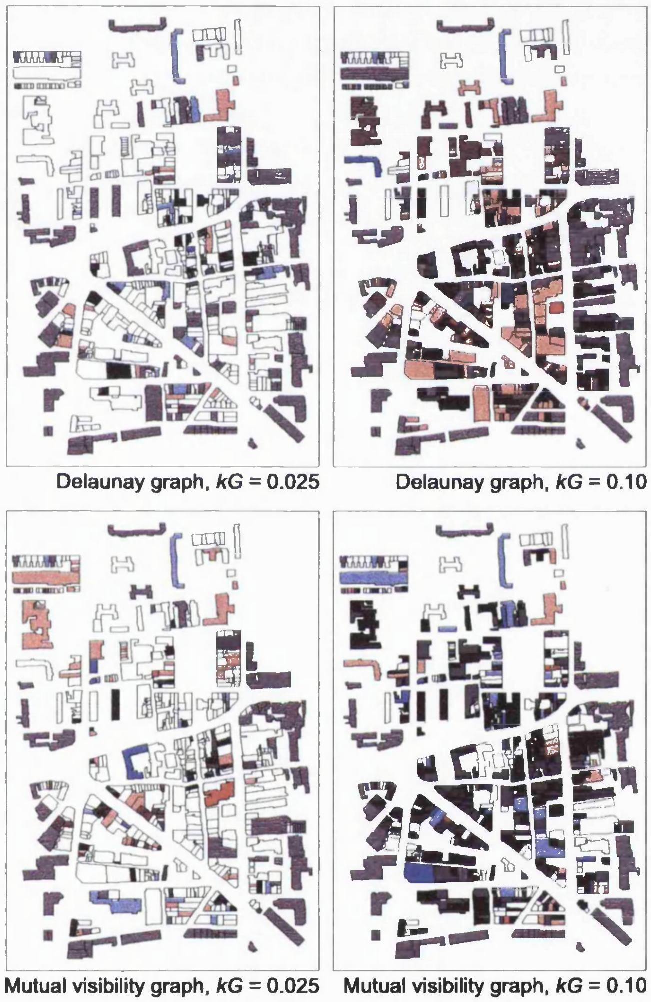


Figure 108 The effect of varying model parameters. Snapshots at t=150, for the Delaunay and visibility graph structures, with $k_G = 0.025$, and $k_G = 0.10$.

Overall these findings are — if not disappointing — then certainly slightly damaging to the whole notion of building an irregular or graph-CA type of model, in this case. How we conceptualise spatial effects in the model, seems to make very little difference to the overall outcomes which result, except in the rather obvious effect on the speed at which location state changes 'diffuse' through the system. We might have anticipated this finding, based on the eventual results of part II, where fragile and robust processes were distinguished, and, it was suggested that segregation type processes are robust. The gentrification model under investigation shares some of the characteristics of the much simpler segregation model, and appears to be similarly robust under variation in the spatial structure in which it is applied. There are differences in the detailed outcomes in each case, but the overall system trajectory is largely unaffected by spatial structure.

11.3 Variation of stochastic events in the model

We should be able to partly confirm this tentative finding by examining the effect of varying the stochastic elements of the model — in effect, the sequence of individual decisions made by households — for the same spatial structures as before. If spatial structure is a strongly controlling factor in system evolution, then it ought to produce very similar results even when the stochastic elements in the model specification are varied. For the examples in this section, the same setting of k_G ($= 0.05$) has been used, to allow comparison of results with those in the previous sections. For brevity, only the Delaunay and visibility graph structures are examined.

In figure 109, system a single snapshot at $t = 150$ from various system 'histories' is shown for the Delaunay triangulated case. Here, the model has been run with nine different sequences of random numbers drawn to determine which households move out, the income of new households, and stochastic changes in household income. The 'case number' under each panel is the random number generator seed value which was used. Case 0 is the same history as illustrated previously in figure 103. The overall extent of gentrification by time $t = 150$ appears similar in each case. The spatial distribution of upgraded locations is different in detail amongst these examples, and no clear overall similarities in pattern are evident. There is perhaps the suggestion that developments usually occurs more rapidly towards the southern half of the model, particularly south of Old Street, but the effect is not marked, and cases

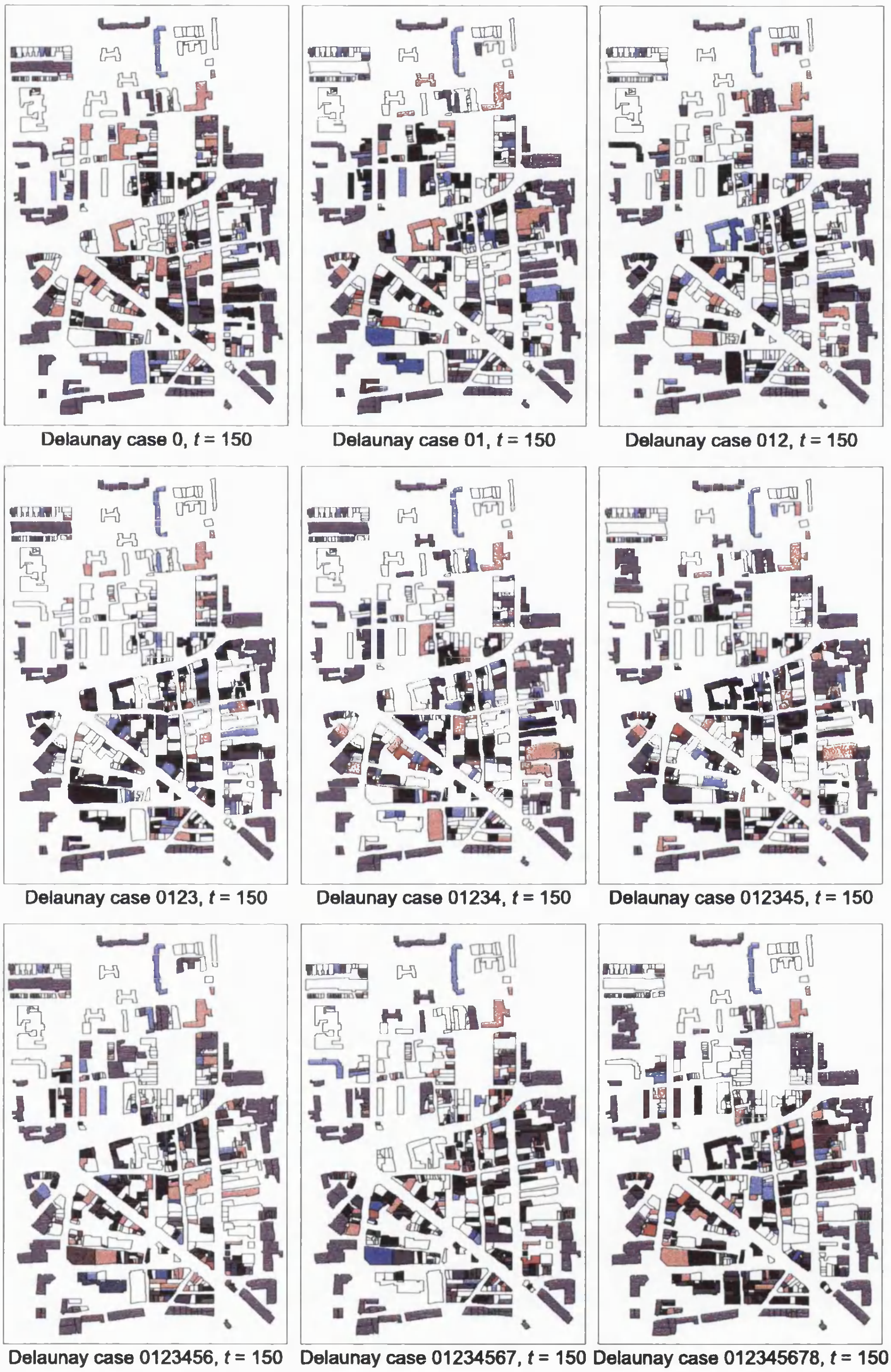


Figure 109 The effect of varying model histories for the Delaunay graph structure. Snapshots at $t = 150$.

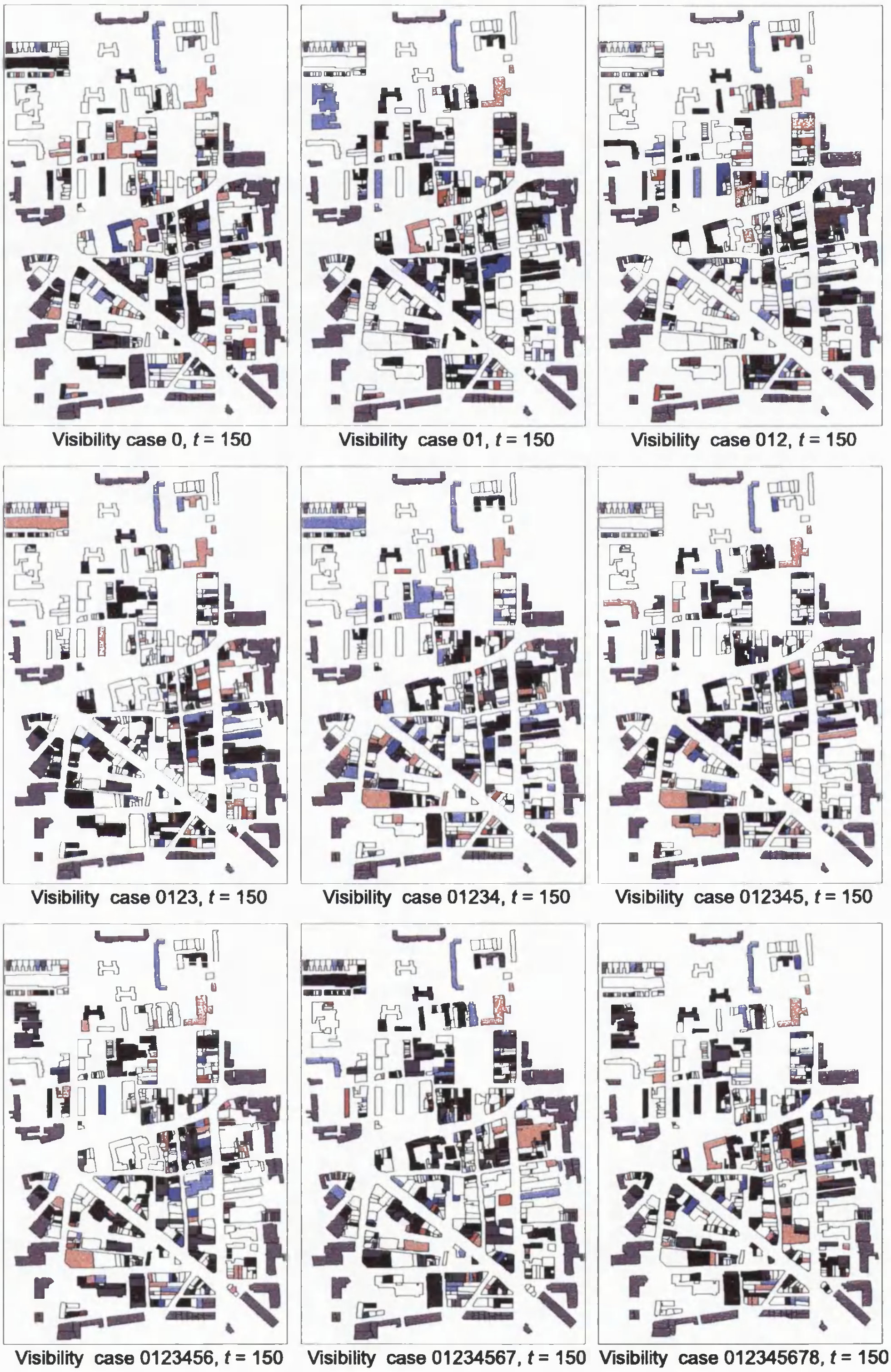


Figure 110 The effect of varying model histories for the mutual visibility graph structure. Snapshots at $t = 150$.

0, 01, 012, 0123456 and 012345678 all see some gentrification in the north. In all cases there is a definite tendency for similarly up- or down-market locations to be clustered together, although not in single large groups, but in a number of areas dotted around the system.

Equivalent data for the visibility graph structure are presented in figure 110. The results in this case appear less uniform. There may well be larger differences in the extent of gentrification by $t = 150$ in the eight cases (with case 012 appearing to lag behind the other cases a little). However, again, there are no obvious similarities in the distribution of gentrified locations for these different histories, which would suggest a strong spatial steering effect. There is less clustering of like locations on this structure compared to the Delaunay system. This is presumably because the visibility-based structure includes edges connecting more distant locations so that any clustering (on the graph structure) would not necessarily appear as clustering in (metric) space.

Taken together, these results tend to confirm the suggestion that the major impact of the model's spatial structure is in the detail of locational change across the system. This is itself much more influenced by the model 'history' — represented by different sequences of random number draws or 'household decisions' — than it is by the spatial structure itself. Of course, the possibility that more subtle structural effects are present can not be ruled out. For example, partitioning the graph structures into equivalence classes or cohesive subgroups might give us a different view of events — and demonstrate that the structure was significant in less obvious ways than can be discerned from simple snapshot image observations. Developing appropriate ways to investigate the complex patterns which *may* exist in a complex model such as this remains a major challenge for contemporary visualisation and spatial analysis research.

11.4 Discussion and conclusions

Although the results of this chapter are inconclusive, taken together they seem to confirm that the spatial evolution of the *Gentrification* model is not strongly influenced by the structure on which it runs. The exception to this observation is that the speed at which neighbourhood change occurs *is* dependent on the underlying graph structure of the model insofar as denser, more interconnected graph structures, promote more rapid diffusion of change across the model. This is not a surprising result. No more subtle spatial effects are readily discernible from the examples presented[2] and there is no support for a hypothesis to the effect that underlying spatial structure can steer or shape the outcome of this type of model.

Of course, this is only a particular example of the sort of micro-scale spatial model which might be constructed using the graph-CA approach, and its basic mechanism — a segregation-type behaviour — has already been seen in chapter 7 to be fairly robust under spatial variation. This should probably lead us to be unsurprised by the outcomes in these examples.

More constructively, it is evident from the results presented here, that perhaps the most forbidding obstacle to progress in this type of detailed, irregular, spatial modelling of phenomena, is not the (formidable) problems of data collection and verification, but the provision of tools which assist in the interpretation, summary, and comparison of model outcomes. It is difficult even with the few cases presented here to grasp any fundamental properties of the model from this relatively complete presentation of the model's behaviour. Consideration of more cases — an urgent requirement if such a model were ever to be applied 'in the real world' — would certainly demand that interpretative and visualisation tools be provided to assist the user/observer in homing in on significant patterns and regularities in the model's behaviour.

[2]Nor were any such effects noted by the author in observing many more cases than those presented here.

Chapter 12

Discussion and conclusions

We can now attempt to draw together the research reported in all of the previous chapters and to examine some of its implications from the perspective of geography generally, complex systems generally, more specifically in relation to geographic processes and models, and finally gentrification and related urban transformation processes in particular.

12.1 On 'the difference that space makes'

Many very general theories of urban processes (social, economic, cultural) which attempt to incorporate space (Harvey 1978, for example) end up having no necessary, testable spatial implications, and it thus becomes difficult to see what difference space makes (Sayer 2000, chapter 5). One of the explicit aims of the graph-CA model proposed in chapter 5 was to link spatial structure to process dynamics. The graph-CA model formalism is more a contribution to ways of thinking about dynamic spatial systems, than it is a novel type of spatial model. This is especially the case given that proximal model has also been formalised using Geo-Algebra (Couclelis 1997, Takeyama & Couclelis 1997).[1] As a dynamic process model, described by CA transition rules, which lends itself naturally to analytic spatial description, through graph structural measures, the graph-CA model is a good vehicle for the exploration of specifically spatial aspects of dynamic processes.

The results reported in chapter 7 suggested that it is possible to distinguish be-

[1]The relationship between the graph-CA and Geo-Algebra approaches is discussed in O'Sullivan (forthcoming).

tween processes which are robust under spatial change, and those which are fragile. Some remarks about the difficulty of making this finding especially meaningful or useful in substantive contexts have already been made in section 7.5 (pages 179ff.). The difficulty of drawing any strong conclusions from the more 'concrete' application of the model described in chapters 10 and 11 has emphasised this point. This only serves to strengthen Sayer's contention that in many (perhaps most?) socio-economic contexts space is not a *necessary* aspect of the situation, but rather a *contingent* one, so that many spatial realisations of particular social institutions or economic phenomena are possible, and, in general, the socio-economic relations in some situation are likely to carry more explanatory weight. Every situation is always simultaneously social *and* spatial, and the spatial structure of the phenomenon will more than likely *add* to our understanding. However, it is unlikely that it will ever *wholly explain* it. It is worthy of note that Allen (1997) reaches similar conclusions after running a simple model of regional development, leading him to suggest that, "a single set of functional requirements, acted upon by 'fairly rational' actors, can be realized in many different ways [...] the equations of change can give rise to a whole range of different possible cities" (pages 22–23). This conclusion lends considerable support to the numerous 'grounded' research programs in geography which concern themselves with detailed, local, archive and interview based investigation of the unfolding of processes like gentrification, on the ground (Foster 1999, is a good recent example).

Put another way, perhaps less pessimistically for some of the more technical spatial disciplines, consideration of the spatial structure of a particular area of enquiry usefully focuses attention on the processes under investigation. This brings us to an important point: space and process are not really separable at all. Of course, this means that the conceptual framework sketched in chapter 4 and developed in detail in the context of graph-CA models in chapter 5 does serious violence to the way the world is. The problem may lie in part with the notion of proximal space. There *is* something a little *odd* about a 'model of space' which can accommodate spatial and allegedly a-spatial effects equally — recall that in proximal space a neighbourhood is defined by relations of nearness between spatial elements, and nearness in turn depends on both (spatial) adjacency and (functional) influence. In fact, proximal space is more like an intellectually accessible formalisation of the interaction structure of any discrete system (spatial or not). As such, it seems to have its origins in earlier formulations by Couclelis, particularly the autonomous sequential machine, which she suggests is a generalisation of the cellular automaton formalism (Couclelis 1985). Thus, proximal space implies *from the outset* that process and structure are inseparable, since interaction between elements implies but does not require that elements be spatially adjacent, and spatial proximity implies, but does not require interaction.

In the end, though, proximal space remains a useful vehicle for understanding and conceptualising the behaviour of the infinite varieties of model which could be constructed (recall the earlier remarks about the pragmatism of the proximal framework — see especially the concluding quote from Gould in chapter 2). GCA-space is a similar conceptual — and pragmatic, framework — which hopefully delivers some of the same conceptual benefits. The surprise, given the lineage of the proximal model, is that the analytical apparatus of graph theory or other discrete structures (Q-analysis, perhaps) has not been invoked before in approaching this issue. The insight that such application is possible and opens up new areas for exploration and research is one of the contributions of this thesis.

12.2 Researching complex systems

From the outset the relationship between the current work and the 'sciences of complexity' has been made clear (see sections 2.2 pages 31ff., 3.3.2 pages 58ff. and 4.5.1 pages 94ff.). Cellular automata are a paradigmatic example of the very notion of complexity, exhibiting as they do diverse and rich behaviour based on only a few simple rules. Graph-CA models are also exemplary complex systems. It is useful to distinguish two types of complexity which face us in investigating such systems:

- First is the *combinatorial 'complexity' of this class of models.* As has been made clear in a number of places (see section 5.3.3 on pages 113ff. especially), the number of models which might be constructed in either of these classes — CA or graph-CA — is, for all intents and purposes, infinite. The numbers involved are so large that no foreseeable improvements in the processing power of computers are likely to make manageable the comprehensive investigation of such models.

  Unfortunately, there really seems to be nothing for it, but to admit that making findings concerning these models statistically robust is very difficult. The combinatorial immensity means that it is very difficult to build a statistically meaningful sample of the possible variations. Nevertheless, it has been possible to show that there is still potential for interesting geographical research in this area, drawing on related research in other fields (Kauffman 1984, 1995, in particular, attempts something similar). The research reported in part II represents one way of addressing this research agenda, and demonstrates that the spatial structure of complex models can affect their behaviour. It also suggests that the effects of structure may be most marked for more complex modes of behaviour (such as exhibited by the Life CA).[2]

- A second major difficulty concerns *the limits of what we can learn from complex models of complex systems*. A mysterious lacuna in much recent work on non-linear dynamics in social systems is an acknowledgement of some of the very fundamental properties of such dynamics, and in particular, their *inherent unpredictability*. Although we can hope to understand some of the general characteristics of such systems — the research agenda of the previous item aims at advancing such understanding — they are simply unpredictable in the traditional scientific sense. This is abundantly clear in part III of this thesis, where even the relatively predictable global behaviour of the *Gentrification* model is composed of unpredictable local behaviour. Built up into a much larger model covering a more representative region of a whole city, such local unpredictability is likely to translate into almost wholly inaccurate prediction — or at any rate prediction at local sites to which it is difficult to attach any sort of confidence.

  This gloomy outlook may be qualified with a little optimism. The most dramatic failures in prediction will usually arise from the failure of models to 'innovate' new mechanisms and behaviours. Between periods of innovation more stability can be expected in events and in system behaviour, and to the extent that a model captures the most pertinent mechanisms and structures it may provide us with a useful tool for examining the implications of limited interventions in urban systems. This is the implicit understanding on which any application of the 'static-structural' analytic planning tools discussed in section 4.3 (pages 73ff.), such as Krüger's graphs or space syntax, must be based. It is also consistent with the uses of models in urban planning suggested by Allen (1997).

[2]This may also be seen as a partial confirmation of Kauffman's findings about the strong tendencies towards chaotic dynamics inherent in k-regular random graphs, where $k > 2$.

Also, to the extent that studies of complex system dynamics of the kind discussed above allow us to develop 'intuitive' understandings of urban systems, they may help us to anticipate the impacts of public policy more sensitively. This is not to be underestimated, since it is precisely the type of systemic, holistic understanding which might have allowed policy-makers to anticipate the impact of (for example) out-of-town development and enthusiastic road-building programmes, on the shape of the late twentieth and early twenty-first century metropolis. With luck, it is also the sort of understanding which may help us to find imaginative ways out of that particular developmental *cul de sac*.[3] Gentrification itself may be seen as an example of a complex system dynamic effect — the capacity for innovation and novelty — which has been commandeered by some policy-makers in a conscious attempt to nudge the development of complex city neighbourhoods in desirable directions. Recent developments in the London Borough of Southwark around the new Tate Gallery at Bankside are a good example of this, where the cultural *cachet* of a major new art gallery is being harnessed as an attractor for new, more affluent residents.

Two further, more technical issues also arising from the study of complex spatial models, deserve further consideration. These are:

- *The need for measures of complexity.* In part II the spatial information time series, an extension of measures used in the physics literature on CA dynamics, was developed and proposed as a possible approach to measuring the dynamic behaviour of the graph-CA type of model. In the abstract context of chapter 7 this measure proved adequate to the task. However, although the spatial information measure used in characterising dynamic behaviour in part II is capable of general application, it must be carefully applied in each case. Thus, we can happily summarise segregation graph-CA behaviour with a single value of final spatial information (see section 7.2.1, pages 163ff.), but this method is meaningless for the Life graph-CA (see section 5.3.6, pages 127ff.). Equally, when segregation behaviour starts to break down, so that many configurations evolve to states in which all cells are in the same state (see section 7.2.2, pages 167ff.) then a different description of behaviour is required. In the case of the Life GCA,

[3] Although the utopian, *Garden Cities*-esque proclamations of the UK's Urban Task Force report (1999), apparently unencumbered as they are by any attempt to address the socio-economic, as well as the *solely* infrastructural aspects of the sustainable city, do not bode well in this regard.

transient time is a candidate measure, whereas the deformed segregation GCA may require the probability that a starting configuration will evolve to the homogeneous state to be determined — although determination of that parameter would itself require a large number of cases to be examined, so that we are immediately faced with the combinatorial problem again.

In the more complicated — though still very abstract — model of part III there are immediately evident difficulties with applying this or any other measure of developing spatial structure. One difficulty with spatial information is in attaching meaning to any particular value of the measure. More practically (and relevant to part III) is the computational intensity of determining its value in a multivariate, irregular setting. Nevertheless, it is likely that measures developed to assist in understanding the behaviours of complex systems will share characteristics with the spatial information measure — particularly the attention to the ideas of order and disorder, and therefore of entropy. This is because such concepts are fundamental to the understanding of complex behaviour. A major theme in the research agenda for further exploration of complex spatial systems must therefore be the development of this or (more likely) other, related measures. The present work contributes to this development in the suggested extension of Wolfram's (1984*a*) spatial information measure to two- and higher-dimensional systems, although much remains to be done.

- *The need for visualisation techniques and tools.* Closely related to the development of measures of complex behaviour must be the development of methods for identifying and understanding it, and for generating hypotheses and ideas for further exploration. The most likely perspective from which this need will be addressed is that of the scientific visualisation research community. The presentation of model outcomes in part III has dramatised this requirement, because we have seen how difficult it is to understand the behaviour of a system from a series of snapshot images of its development. Such difficulties are not much alleviated by the more complete, 'real-time' and animated view of the system dynamics enjoyed by the author!

  A possible contribution to scientific visualisation efforts arising naturally from the work presented here, might be to view models of this type in three-dimensions[4] with *graph measures* on the third (height) axis. Thus the graph vertices underlying a discrete relational model might appear as points on a 'surface', with graph centrality or other characteristics serving to locate cells 'vertically'. This might allow us to see subtle effects in the evolution of the model, with states appearing to 'flow' up- or down-hill (for example, Wood, Fisher, Dykes, Unwin & Stynes 1999, show how such landscape-metaphorical understanding of data patterns can inform our understanding). The graph structural properties might also be used to redraw the system structure in more immediately comprehensible arrangements so that its effects may be easily understood (Tollis, Di Battista, Eades & Tamassia 1999). Again, such efforts would benefit from presentation in 3-D interactive form. This can be seen as an alternative take on multi-dimensional scaling methods in earlier research.

Thus, complex systems present themselves as not just complex behaviourally, but complex to research! However, such complexity, while limiting what can be directly predicted with models, also raises many interesting and challenging research questions, which seem likely to further research in a number of areas in the quantitative geography tradition.

12.3 Geographical theory and geographical models

Couclelis (1984), in wide ranging consideration of the applicability of modelling to urban — or, more generally, any geographical — areas probes the philosophical question of why models should work at all. After all, we know that many models offer only gross simplifications of what happens in urban settings. Particularly mysterious is the question of why mathematical structures transferred from completely different problem domains (planetary motion, thermodynamics, evolutionary theory) should shed any light at all on urban development. Couclelis is critical of the adoption of these concepts on the basis of intuitive, analogical arguments alone:

> "[I]t will be argued that the existence of substantive analogy between thermodynamics and urban systems is highly unlikely, and that formal analogy in the usual sense of accidental similarity of structure between substantively unrelated processes, provides insufficient grounds for adopting the model." (page 325)

[4] Probably actually 2.5D, see MacEachran (1995).

Seen from the philosophical perspective of section 2.2.1 (pages 32ff.) this objection is hard to sustain, provided that models share the same or similar structures to those in other fields, *as a result of reasonable abstractions of the behaviour of their constituent elements.* One of the motivations for the graph-CA model has been to make the construction of cellular geographical models more 'concrete' in exactly this way, by enabling a more convincing level of representation in which cells correspond to more readily identified and possibly less contentious geographic units than the grid-squares of conventional CA.

The graph-CA is also intended as a single well-defined variation of the CA formalism, which admits exploration and permits understanding in a controlled manner. This is in contrast to the numerous *ad hoc* modifications to the formalism made by many others (see the review in section 4.5, pages 88ff.). To be fair to those efforts, they are largely aimed at making urban CA models more realistic. However, the problem of any understanding yielded by our models disappearing like the "smile of the Cheshire cat" as suggested by Couclelis (1985, page 588), as models become less formally elegant, is very real. A model which faithfully represents every known likely contributory factor to some geographical phenomenon may be adequate for a time, but it is unlikely to aid understanding, since its emulation of reality will be unsurprising and unrevealing, as well as full of all the contradictions and confusions of reality itself. It is also likely to prove (paradoxically) confining, since the impact of innovations in the real world may be difficult to incorporate. As a result, such detailed models may prove positively unhelpful when the complex system modelled, does what complex systems do, and adapts or alters its behaviour. Then it becomes likely that the model's predictions will be more than 'a little bit off', because of calibration-related problems, and actually entirely misleading because new causal mechanisms operating in the real world are missing entirely.[5]

On the other hand, the graph-CA model has itself proved difficult to understand, even in the relatively simple model of gentrification developed in part III. This implies that a 'bifurcation' in researchers responses to the possibility of complex models is appropriate. The use of simplistic formal models to explore general properties of different classes of phenomena, and to elucidate some general principles, such as the

[5]For example, the TRANSIMS model (Nagel et al. 1998) may exemplify this problem, since it has difficulties with representing non-automobile modes of transport, which seems likely to limit its usefulness to policy-makers severely.

possibility of the emergence of global structure from local action is one route we can follow. It is equally sensible however, to build more realistic, less directly comprehensible models as decision support systems, provided developers and practitioners remain aware of the limitations that complex dynamics impose. Each route forward serves to illuminate different aspects of the complex geographical world. The graph-CA formalism belongs squarely in the former camp, although as part III makes clear it clearly has potential as a basis for detailed micro-simulation models in the latter camp also.

Micro-simulation models present other difficulties in terms of data collection, since they are data-hungry beasts. This is evident from the material in part III where a not-particularly-detailed model has been attempted. This suggests to me that the exploration of the general dynamic properties of complex spatial systems may ultimately prove more practical and enlightening than the development of detailed and expensive micro-simulation models. However, such a preference *is only a preference*, arising more from my own interest in using models as vehicles for furthering our *understanding* of cities, rather than as decision support tools.

Finally, in this connection it is worth mentioning that a fairly simple route to integration of the graph-CA formalism into current GIS seems perfectly feasible. Graphs are simple relational data structures, easily implemented with linked lists and pointers in any computing environment. Their implementation in existing GIS therefore seems eminently practicable. The major limitation is likely to be that of memory, since even relatively small sets of spatial objects (say 1000 elements) may have very many inter-relations (edges in the GCA). Depending on what additional information is attached to these edges, large data structures — in storage terms — will quickly result. The advantages of such an integration from the GCA modelling perspective are: first, the increased analytic and visualisation capabilities offered by GIS, as is clear from the immediacy of the dynamic maps presented by the loose coupling between the *Gentrification* model and the *GeoTools* package; and, second the spatial, and geometric analysis capabilities provided by GIS enable the construction of various alternative proximal model structures from the same spatial elements, as demonstrated in chapter 11, where various structures were generated based on a variety of 'neighbourhood inclusion' rules. From the GIS perspective, the implementation of dynamic, relational objects in a graph-CA or similar framework might offer one possible way forward on the perennial knotty problem of time and dynamics in GIS — although this would require that the issue of synchronous and asynchronous behaviour be more thoroughly explored than hitherto.

12.4 On gentrification

The principal conclusion from part III must surely be, as has already been suggested above, that no matter what advances in computing and simulation technology lie ahead, that there will remain a role for detailed empirical, archival and historical research in the investigation of urban phenomena. Gentrification may be a *particularly* contentious example, but it is clear that any sufficiently generalised concept is likely to be amenable to numerous competing explanations.

It was suggested in section 8.1.4 (pages 192ff.) that such contested arenas may even be the ones which can be most usefully modelled, since models may represent one way to distinguish adequate from inadequate explanations. It is not hard to guess how this suggestion might be received beyond the confines of the quantitative tradition, in the wider geographical research community. A negative response is not *entirely* unreasonable however, since it must be acknowledged that it is impossible to actually prove anything using models, especially when it is impossible or very difficult to explore the whole model rule space. Nevertheless, there ought to be room for considered and constructive engagement between advocacy of recent modelling approaches and non-quantitative social theory (O'Sullivan & Haklay forthcoming, is an attempt to engage in such debate).

In this light, what has been learned about gentrification, from the graph-CA model presented? Certainly, that it is possible to work the rent gap hypothesis up into a model, although formidable problems are encountered at the resolution/scale attempted here in terms of data requirements. The model exhibits some of the behaviour we would hope to see in a model of gentrification. It is likely that the model's rules would have to be complicated somewhat for more convincing behaviour to emerge. In the particular study area examined, some conceptualisation of different land uses is desirable. It also seems that an agent-based approach might be more appropriate for the highly detailed level of simulation which has been attempted, although this would be even more demanding in data terms.

In spite of the probabilistic elements in the *Gentrification* model, there is a seeming inevitability about the onward march of gentrification into the area modelled. This may be as much to do with the artificial closure of the model enforced by time and computing constraints as with any limitation in the model's applicability. The graph-CA framework may itself offer a relatively painless way out of the difficulty, since there is the potential for a hierarchical graph-CA model in which other neighbourhoods are represented alongside Hoxton. These could be modelled in less detail, so that the overall size of the model is not much increased, but so that developments in Hoxton are more realistically constrained by developments elsewhere. Ultimately though, it is hard to see how constraints exogenous to any graph-CA model — the state of the economy, interest rates, government policy — can be avoided. Of course, this is perfectly reasonable, since the world is not closed either, at least not at the level of this model.

However, none of the observations in chapter 11 rule against the development of another model based on one or more of the *other* theories of gentrification discussed in section 8.1.3 (pages 190ff.). Assessing the explanatory adequacy of various gentrification models *against one another* would be an obvious next step. The suspicion must remain that such comparative assessments will establish nothing definitive. Partly this is because of the difficulty of developing measures to assess how well models fit observable complex reality. In the case of gentrification, it seems likely also to result from the fact that it is a multi-faceted phenomenon, not explicable in terms of any single theory. As has been noted, the chaos in the gentrification literature, reflects the chaos in the concept, and the fact that it is a complex outcome of many interacting mechanisms, combining in various combinations in various historical and geographical settings. However, that this is the case for gentrification does not rule against the notion of subjecting theory in the social sciences to model-based examination and interpretation. Indeed, such research seems likely to become more prevalent in years to come. The present work is intended as a contribution to such efforts.

A more specific point arising from the *Gentrification* model also merits attention. We might have hoped to draw some conclusions from the reported results regarding which representation of the relational structural is most appropriate in the model context. In fact, any assessment of that issue would require that more attention be paid to the interaction between spatial and temporal scale in the model. The temporal scale of the *Gentrification* model is implied by the types of transformation which can occur in one CA time-step. This is itself partly governed by the choice of parameter settings, since a different value of (say) p_0 would imply different levels of movement in the housing market. In fact the whole issue of the representation of time in CA-based models is a complex one, raising all sorts of questions about the approach — and coincidentally about the representation of time which would be implied if graph-CA were implemented in GIS as suggested in the previous section. Returning to the specific issue of which of the different relational structures employed in chapter 11 is most appropriate, it is evident that the implication of the different structures and resulting neighbourhood definitions is related to the way in which individual's perceive and assess place, and act on those perceptions and assessments. The interaction between the graph-CA model rule definitions and temporal scale make it difficult to draw even tentative conclusions, and, in fact, suggest that another modelling approach — probably agent-based micro-simulation — would yield more insight on this point.

12.5 And finally...

It is difficult to summarise the findings of the previous sections beyond suggesting that the major finding of the present work is that the irregular of graph-based CA model formalism which has been introduced, seems sufficiently interesting to be examined further. It has proved to be an interesting vehicle for the investigation of both very general spatial dynamical properties, and for the development of more concrete representative models of real spatial systems. The reported research has raised a number of issues relating particularly to the systematic investigation of complex systems of which the graph-CA is an example, and these have been considered above.

Where a proximal conception of space is appropriate, if its structure is best represented and understood as a graph — which seems incontrovertible — then the fact that any particular graph can be mapped to a limitless number of actual spatial realisations suggests that 'the difference that space makes' can only ever be contingent: processes are what matter, they are always realised in and through space, but the relationship is always likely to be *one* process to *many* possible spatial realisations. Any social theory which puts space at the heart of its concerns must confront this difficulty. One short-sighted outcome of such a view is precisely that of the pre-quantitative era in geography — a regional geography deeply concerned with uncovering the specifics of different places with little reference to the higher level of generality of causal mechanisms and structures. Such an approach is short-sighted, because it is only by attempting to understand those mechanisms and structures that we can ever hope to explain the specifics of regions and places. In this context, theoretical geographical models, carefully applied and critically examined, are a useful complement to more concrete approaches based on empirical observation — as the present work has attempted to show.

Bibliography

Alexander, C. (1964), *Notes on the Synthesis of Form*, Harvard University Press, Cambridge, MA.

Alexander, C. (1965), 'A city is not a tree', *Architectural Forum* **122**, 58–62.

Alexander, C., Ishikawa, S. & Silverstein, M. (1977), *A Pattern Language: Towns, Buildings, Construction*, Oxford University Press, New York. With M. Jacobson, I. Fiksdahl-King & S. Angel.

Alhir, S. S. (1998), *UML in a Nutshell*, O'Reilly & Associates Inc., Sebastopol, CA.

Allen, P. M. (1997), *Cities and Regions as Self-Organizing Systems: Models of Complexity*, Vol. 1 of *Environmental Problems and Social Dynamics*, Gordon Breach Science Publishers, Amsterdam.

Allen, P. M. & Sanglier, M. (1979), 'A dynamic model of growth in a central place system', *Geographical Analysis* **11**(3), 256–272.

Alonso, W. (1964), *Location and Land Use: Toward a General Theory of Land Rent*, Harvard University Press, Cambridge, MA.

Anonymous (1999), 'UML Resource Center: Unified Modeling Language, Standard Software Notation'. Materials available on-line at `http://www.rational.com/uml/index.jtmpl`.

Applebaum, D. (1996), *Probability and Information Theory: An Integrated Approach*, Cambridge University Press, Cambridge, England.

Atkin, R. H. (1974*a*), 'An approach to structure in architectural and urban design 1: Introduction and mathematical theory', *Environment and Planning B: Planning & Design* **1**, 51–67.

Atkin, R. H. (1974*b*), 'An approach to structure in architectural and urban design 2: Algebraic representation and local structure', *Environment and Planning B: Planning & Design* **1**, 173–191.

Atkin, R. H. (1974*c*), *Mathematical Structure in Human Affairs*, Heinemann Educational, London.

Atkin, R. H. (1975), 'An approach to structure in architectural and urban design 3: Illustrative examples', *Environment and Planning B: Planning & Design* **2**, 21–57.

Atkin, R. H. (1981), 'A theory of surprises', *Environment and Planning B: Planning & Design* **8**(4), 359–365.

Atkin, R. H., Johnson, J. H. & Mancini, V. (1971), 'An analysis of urban structure using concepts of algebraic topology', *Urban Studies* **8**, 221–242.

Badcock, B. (1989), 'An Australian view of the rent gap hypothesis', *Annals of the Association of American Geographers* **79**(1), 125–145.

Badcock, B. (1991), 'On the nonexistence of the rent gap, a reply', *Annals of the Association of American Geographers* **80**, 459–461.

Badcock, B. (1992*a*), 'Adelaide's heart transplant: 1. Creation, transfer, and capture of 'value' within the built environment', *Environment and Planning A* **24**, 215–241.

Badcock, B. (1992*b*), 'Adelaide's heart transplant: 2. The 'transfer' of value within the housing market', *Environment and Planning A* **24**, 323–339.

Ballas, D., Clarke, G. & Turton, I. (1999), Exploring microsimulation methodologies for the estimation of household attributes, *in* '4th International Conference on GeoComputation', 25–28 July, Mary Washington College, Fredericksburg, VA. On-line at `http://www.geovista.psu.edu/geocomp/geocomp99/index.htm`.

Barnsley, M. J. & Barr, S. L. (1997), 'Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system', *Computers Environment and Urban Systems* **21**(3/4), 209–225.

Barr, S. L. & Barnsley, M. J. (1997), 'A region-based graph-theoretic data model for the inference of second-order thematic information from remotely-sensed images', *International Journal of Geographical Information Science* **11**(6), 555–576.

Batty, M. (1974*a*), 'Spatial entropy', *Geographical Analysis* **6**, 1–31.

Batty, M. (1974*b*), 'Urban density and entropy functions', *Journal of Cybernetics* **4**(2), 41–55.

Batty, M. (1976*a*), 'Entropy in spatial aggregation', *Geographical Analysis* **8**, 1–21.

Batty, M. (1976*b*), *Urban Modelling: Algorithms, Calibrations, Predictions*, Vol. 3 of *Cambridge Urban & Architectural Studies*, Cambridge University Press, Cambridge, England.

Batty, M. & Longley, P. A. (1986), 'The fractal simulation of urban structure', *Environment and Planning A* **18**, 1143–1179.

Batty, M. & Longley, P. A. (1994), *Fractal Cities: A Geometry of Form and Function*, Academic Press, London.

Batty, M. & March, L. (1976), 'The method of residues in urban modelling', *Environment and Planning A* **8**, 189–214.

Batty, M. & Xie, Y. (1996), 'Preliminary evidence for a theory of the fractal city', *Environment and Planning A* **28**, 1745–1762.

Batty, M. & Xie, Y. (1997), 'Possible urban automata', *Environment and Planning B: Planning & Design* **24**(2), 175–192.

Batty, M. & Xie, Y. (1999), 'Self-organized criticality and urban development', *Discrete Dynamics in Nature and Society* **3**, 109–124.

Batty, M., Couclelis, H. & Eichen, M., eds (1997), *Special Issue — Urban Systems as Cellular Automata*, Vol. 24(2) of *Environment and Planning B: Planning & Design*, Pion Ltd, London.

Batty, M., Couclelis, H. & Stiny, G. (1994), 'Lionel March — three sketches', *Environment and Planning B: Planning & Design* **21**(Special Issue), s1–s6.

Baum, A. & Mackmin, D. (1989), *The Income Approach to Property Evaluation*, 3rd edn, Routledge, London.

Beaumont, J. R. & Gatrell, A. C. (1982), *An Introduction to Q-analysis*, Vol. 34 of *Concepts and Techniques in Modern Geography (CATMOG)*, Geo Abstracts, Norwich, England.

Beauregard, R. A. (1986), The chaos and complexity of gentrification, *in* N. Smith & P. Williams, eds, 'Gentrification of the City', Allen & Unwin, Boston, MA.

Beauregard, R. A. (1990), 'Trajectories of neighbourhood change: the case of gentrification', *Environment and Planning A* **22**, 855–874.

Beineke, L. W. & Wilson, R. J. (1997), *Graph Connections: Relationships between Graph Theory and Other Areas of Mathematics*, Vol. 5 of *Oxford Lecture Series in Mathematics and its Applications*, Clarendon Press, Oxford, England.

Benenson, I., Omer, I. & Portugali, J. (1999), An agent-based model of residential mobility in the Tel-Aviv metropolitan area, *in* '4th International Conference on GeoComputation', 25–28 July, Mary Washington College, Fredericksburg VA. On-line at `http://www.geovista.psu.edu/geocomp/geocomp99/index.htm`.

Benko, G. & Strohmayer, U. (1997), *Space and Social Theory: Interpreting Modernity and Postmodernity*, Blackwell, Oxford, England.

Berlekamp, E. R., Conway, J. H. & Guy, R. K. (1982), *Winning Ways for Your Mathematical Plays*, Academic Press, New York.

Bersini, H. & Detours, V. (1994), Asynchrony introduces stability in cellular automata based models, *in* R. A. Brooks & P. Maes, eds, 'Artificial Life IV Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems', The MIT Press, Cambridge, MA and London, pp. 382–387.

Bhaskar, R. (1998), *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*, Critical Realism: Interventions, 3rd edn, Routledge, London and New York.

Bolland, J. M. (1988), 'Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks', *Social Networks* **10**, 233–253.

Bollobás, B. (1985), *Random Graphs*, Academic Press, London.

Bondi, L. (1991), 'Gender divisions and gentrification: a critique', *Transactions of the Institute of British Geographers* **NS16**, 190–198.

Borgatti, S. P. & Everett, M. G. (1989), 'The class of all regular equivalences: algebraic structure and computation', *Social Networks* **12**, 337–358.

Borgatti, S. P., Everett, M. G. & Shirey, P. R. (1990), '*LS* sets, lambda sets, and other cohesive subsets', *Social Networks* **11**, 65–88.

Bourassa, S. C. (1990), 'On 'An Australian view of the rent gap hypothesis' by Badcock', *Annals of the Association of American Geographers* **80**, 458–459.

Bourassa, S. C. (1993), 'The rent gap debunked', *Urban Studies* **10**, 1731–1744.

Brewer, C. A. (1994), Color use guidelines for mapping and visualization, *in* A. M. MacEachren & D. R. F. Taylor, eds, 'Visualization in Modern Cartography', Elsevier Science, Tarrytown, NY, pp. 123–147.

Bridge, G. (1994), 'Gentrification: a reappraisal', *Environment and Planning D: Society and Space* **12**, 31–51.

Brown, G. R. (1991), *Property Investment and the Capital Markets*, E&FN Spon, London.

Buckley, F. & Harary, F. (1990), *Distance in Graphs*, Addison-Wesley, Redwood City, CA.

Burt, R. S. (1990), 'Detecting role equivalence', *Social Networks* **12**, 83–97.

Byrne, D. (1998), *Complexity Theory and the Social Sciences: An Introduction*, Routledge, London.

Cadman, D. & Topping, R. (1995), *Property Development*, 4th edn, E&FN Spon, London.

Campari, I. (1995), Uncertain boundaries in urban space, *in* P. Burroughs & A. U. Frank, eds, 'Geographic Objects with Indeterminate Boundaries', Vol. 2 of *GISDATA*, Taylor & Francis, London.

Castells, M. (1989), *The Informational City*, Blackwell, Oxford, England.

Chapin, F. S. & Weiss, S. F., eds (1962), *Urban Growth Dynamics in a Regional Cluster of Cities*, John Wiley & Sons, New York.

Chartrand, G., Kubicki, G. & Schultz, M. (1998), 'Graph similarity and distance in graphs', *Aequationes Mathematicae* **55**, 129–145.

Chorley, R. J. & Haggett, P., eds (1967), *Models in Geography*, Methuen, London.

Cilliers, P. (1998), *Complexity and Postmodernism*, Routledge, London.

Clark, E. (1987), *The Rent Gap and Urban Change: Case studies in Malmö 1860–1985*, Vol. 101 of *Meddelanden från Lunds Universitets Geografiska Institutioner*, Lund University Press, Lund, Sweden.

Clark, E. (1988), 'The rent gap and transformation of the built environment: case studies in Malmö 1860–1985', *Geografiska Annaler B* **70**(2), 241–254.

Clark, E. (1992), 'On blindness, centrepieces and complementarity in gentrification theory', *Transactions of the Institute of British Geographers* **NS17**, 358–362.

Clark, E. (1994), 'Towards a Copenhagen interpretation of gentrification', *Urban Studies* **31**(7), 1033–1042.

Clark, E. & Gullberg, A. (1991), 'Long swings, rent gaps and structures of building provision — the postwar transformation of Stockholm's inner city', *International Journal of Urban and Regional Research* **15**(4), 492–504.

Clark, P. J. & Evans, F. C. (1954), 'Distance to nearest neighbour as a measure of spatial relationships in populations', *Ecology* **35**, 445–453.

Clarke, K. C., Hoppen, S. & Gaydos, L. (1997), 'A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area', *Environment and Planning B: Planning & Design* **24**(2), 247–262.

Clarke, M. & Holm, E. (1987), 'Micro-simulation methods in human geography: a review and further extensions', *Geografiska Annaler B* **69**, 145–164.

Cliff, A. & Ord, J. K. (1973), *Spatial Autocorrelation*, Pion, London.

Collier, A. (1994), *Critical Realism: An Introduction to the Philosophy of Roy Bhaskar*, Verso, London.

Conzen, M. R. G. (1938), 'Towards a systematic approach in planning science: geoproscopy', *Town Planning Review* **18**, 1–26.

Conzen, M. R. G. (1949), 'The Scandinavian approach to urban geography', *Norsk Geografisk Tidsskrift* **12**, 86–91.

Conzen, M. R. G. (1958), The growth and character of Whitby, *in* G. H. J. Daysh, ed., 'A survey of Whitby and the surrounding area', Eton, England, pp. 49–89.

Conzen, M. R. G. (1960), *Alnwick, Northumberland: A Study in Town-Plan Analysis*, Vol. 27 of *Institute of British Geographers Publications*, George-Philip, London.

Conzen, M. R. G. (1981), The morphology of towns in Britain during the industrial era, *in* J. W. R. Whitehand, ed., 'The Urban Landscape: Historical Development and Management Papers by M. R. G. Conzen', Academic Press, London, pp. 87–126.

Couclelis, H. (1983), 'On some problems in defining sets for Q-analysis', *Environment and Planning B: Planning & Design* **10**, 423–438.

Couclelis, H. (1984), 'The notion of prior structure in urban modelling', *Environment and Planning A* **16**, 319–338.

Couclelis, H. (1985), 'Cellular worlds: a framework for modelling micro-macro dynamics', *Environment and Planning A* **17**, 585–596.

Couclelis, H. (1988), 'Of mice and men: what rodent populations can teach us about complex spatial dynamics', *Environment and Planning A* **20**(1), 99–109.

Couclelis, H. (1992*a*), Location, place, region and space, *in* R. F. Abler, M. G. Marcus & J. M. Olson, eds, 'Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography', Rutgers University Press, New Brunswick, NJ, pp. 215–233.

Couclelis, H. (1992*b*), People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS, *in* A. U. Frank, I. Campari & U. Formentini, eds, 'Theories and Methods of Spatio-Temporal Reasoning in Geographic Space', Vol. 639 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 65–77.

Couclelis, H. (1997), 'From cellular automata to urban models: new principles for model development and implementation', *Environment and Planning B: Planning & Design* **24**, 165–174.

Crowther, D. & Echenique, M. (1972), Development of a model of urban spatial structure, *in* Martin, ed., 'Urban Space and Structures', Cambridge University Press, Cambridge, pp. 175–218.

de Berg, M., van Kreveld, M., Overmars, M. & Schwarzkopf, O. (1997), *Computational Geometry: Algorithms and Applications*, Springer, Berlin and New York.

Department of the Environment, Transport and the Regions (1998), *Town Centres: Defining Boundaries for Statistical Monitoring: Feasibility Study*, The Stationery Office, London.

Donninger, C. (1986), 'The distribution of centrality in social networks', *Social Networks* **8**, 191–203.

Dorling, D. (1999), 'Who's afraid of income equality?', *Environment and Planning A* **31**(4), 571–574.

Dykes, J. (1994), Area-value data: new visual emphases and representations, *in* H. M. Hearnshaw & D. Unwin, eds, 'Visualization in Geographical Information Systems', John Wiley & Sons, Chichester, England, pp. 103–114.

Echenique, M. & Owers, J. (1994), 'Research into practice: the work of the Martin Centre in urban and regional modelling', *Environment and Planning B: Planning & Design* **21**(5), 513–514.

Echenique, M., Crowther, D. & Lindsay, W. (1972), A structural comparison of three generations of New towns, *in* L. Martin & L. March, eds, 'Urban Space and Structures', Cambridge University Press, Cambridge, England, pp. 219–259.

Edwards, G. (1993), The Voronoi model and cultural space: applications to the social sciences and humanities, *in* A. U. Frank & I. Campari, eds, 'Spatial Information Theory: A Theoretical Basis for GIS', Vol. 716 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 202–214.

Egenhofer, M. J. & Mark, D. M. (1995), Naïve geography, *in* A. U. Frank & W. Kuhn, eds, 'Spatial Information Theory: A Theoretical Basis for GIS', Vol. 988 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 1–15.

Einstein, A. (1960), *Relativity: the Special and the General Theory*, Methuen, London.

Entrikin, J. N. (1991), *The Betweenness of Place: Towards a Geography of Modernity*, Johns Hopkins University Press, Baltimore, MD.

Erdös, P. & Renyi, A. (1960), 'On the evolution of random graphs', *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17–61.

Faith, J. E. (1998), Why gliders don't exist: anti-reductionism and emergence, *in* C. Adami, R. Belew, H. Kitano & C. Taylor, eds, 'Proceedings of the Sixth International Conference on Artifical Life', The MIT Press, Cambridge, MA.

Faith, J. E. (2000), 'Emergent Representations: Dialectical Materialism and the Philosophy of Mind'. University of Surrey, unpublished D.Phil. thesis.

Flanagan, D. (1997), *Java in a Nutshell*, O'Reilly & Associates Inc., Sebastopol, CA.

Foster, J. (1999), *Docklands: Cultures in Conflict, Worlds in Collision*, UCL Press, London.

Freeman, L. C. (1977), 'A set of measures of centrality based on betweenness', *Sociometry* **40**(1), 35–41.

Freeman, L. C. (1979), 'Centrality in social networks: conceptual clarification', *Social Networks* **1**, 215–239.

Freeman, L. C., Borgatti, S. P. & White, D. R. (1991), 'Centrality in valued graphs: a measure of betweenness based on network flow', *Social Networks* **13**, 141–154.

Galster, G. C. (1987), *Homeowners and Neighbourhood Reinvestment*, Duke Press Policy Studies, Duke University Press, Durham, NC and London.

Garreau, J. (1992), *Edge City: Life on the New Frontier*, Anchor, New York.

Gatrell, A. C. (1977), 'Complexity and redundancy in binary maps'', *Geographical Analysis* **9**(1), 29–41.

Gatrell, A. C. (1985), 'Review of *The Social Logic of Space*', *Progress in Human Geography* **9**(3), 468–469.

Glass, R. (1964), Introduction: Aspects of change, *in* Centre for Urban Studies, ed., 'London: Aspects of Change', MacGibbon & Kee, London, pp. xiii–xlii.

Goddard, W. & Swart, H. C. (1996), 'Distances between graphs under edge operations', *Discrete Mathematics* **161**, 121–132.

Gold, C. M. (1992), The meaning of neighbour, *in* A. U. Frank, I. Campari & U. Formentini, eds, 'Theories and Methods of Spatio-Temporal Reasoning in Geographic Space', Vol. 639 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 220–235.

Golledge, R. G. & Stimson, R. J. (1997), *Spatial Behavior: A Geographic Perspective*, The Guilford Press, New York.

Gould, P. (1997), 'The structure of space(s)', *Geografiska Annaler B* **79**(3), 125–140.

Gould, P. & Gatrell, A. (1979), 'A structural analysis of a game: The Liverpool v. Manchester United Cup Final of 1977', *Social Networks* **9**, 277–282.

Gray, J. (1989), *Ideas of Space Euclidean, Non-Euclidean and Relativistic*, 2nd edn, Clarendon Press, Oxford, England.

Green, C. N. (2000), From Factories to Fine Art — History and Analysis of the Visual Arts Networks in London's East End, 1968–1998, unpublished Ph.D. thesis, University College London.

Griffiths, H. B. (1983), 'Using mathematics to simplify Q-analysis', *Environment and Planning B: Planning & Design* **10**(4), 403–422.

Grunbaum, A. (1970), *Philosophical Problems of Space and Time*, 2nd edn, Routledge & Kegan Paul, London.

Gutowitz, H. A. (1989), 'A hierarchical classification of cellular automata', *Physica D* **45**, 136–156.

Hägerstrand, T. (1968), *Innovation Diffusion as a Spatial Process*, University of Chicago Press, Chicago.

Haggett, P. (1979), *Geography: A Modern Synthesis*, 3rd edn, Harper & Row, New York.

Haggett, P. & Chorley, R. J. (1969), *Network Analysis in Geography*, Edward Arnold Ltd, London.

Hamnett, C. (1984), Gentrification and residential location theory: a review and assessment, *in* D. T. Herbert & R. J. Johnson, eds, 'Geography and the Urban Environment', Vol. VI of *Progress in Research and Applications*, Wiley, London.

Hamnett, C. (1991), 'The blind men and the elephant: the explanation of gentrification', *Transactions of the Institute of British Geographers* **NS16**, 173–189.

Hamnett, C. (1992), 'Gentrifiers or lemmings? a response to Neil Smith', *Transactions of the Institute of British Geographers* **NS17**, 116–119.

Hanson, J. (1989), 'Order and structure in urban design: The plans for the rebuilding of London after the Great Fire of 1666', *Ekistics* **334–335**, 22–42.

Hanson, J. E. & Crutchfield, J. P. (1992), 'The attractor-basin portrait of a cellular automaton', *Journal of Statistical Physics* **66**(5/6), 1415–1462.

Hanson, J. E. & Crutchfield, J. P. (1997), 'Computational mechanics of cellular automata: an example', *Physica D* **103**, 169–189.

Harary, F. & Rockey, J. (1976), 'A city is not a semilattice either', *Environment and Planning A* **8**, 375–384.

Harris, R. (1998), Considering (mis-)representation in geodemographics and lifestyles, *in* '3rd International Conference on Geocomputation', 17–19 September, University of Bristol, School of Geographical Sciences.

Hartshorne, R. (1961), *The Nature of Geography*, Association of American Geographers, Chicago.

Harvey, D. (1969), *Explanation in Geography*, Edward Arnold Ltd, London.

Harvey, D. (1973), *Social Justice and the City*, Edward Arnold Ltd, London.

Harvey, D. (1978), 'The urban process under capitalism: a framework for analysis', *International Journal of Urban and Regional Research* **2**, 101–131.

Harvey, D. (1990), *The Condition of Postmodernity*, Blackwell, Cambridge, MA.

Harvey, D. L. & Reed, M. (1996), Social science as the study of complex systems, *in* L. D. Kiel & E. Elliott, eds, 'Chaos Theory in the Social Sciences: Foundations and Applications', University of Michigan Press, Ann Arbor, MI, pp. 295–323.

Hensel, A. (1996), 'Conway's game of life'. On-line at `http://www.mindspring.com/~alanh/life/`.

Hillier, B. (1985), 'The nature of the artificial — the contingent and the necessary in spatial form in architecture', *Geoforum* **16**(2), 163–178.

Hillier, B. (1996), *Space is the Machine*, Cambridge University Press, Cambridge.

Hillier, B. & Hanson, J. (1984), *The Social Logic of Space*, Cambridge University Press, Cambridge, England.

Hillier, B., Penn, A., Hanson, J., Grajewski, T. & Xu, J. (1993), 'Natural movement: configuration and attraction in urban pedestrian movement', *Environment and Planning B: Planning & Design* **20**(1), 29–66.

Hordijk, W., Crutchfield, J. P. & Mitchell, M. (1996), Embedded-particle computation in evolved cellular automata, Working Paper 96-09-073, Santa Fe Institute. On-line at `http://www.santafe.edu/sfi/publications/96wplist.html`.

Howard, E. (1898), *To-morrow: A Peaceful Path to Real Reform*, Swann Sonnenschein, London.

Hrnciar, P., Haviar, A., Monoszova, G. & Bystrica, B. (1996), 'Some characteristics of the edge distance between graphs', *Czechoslovak Mathematical Journal* **46**(121), 665–675.

Icking, C., Klein, R., Köllner, P. & Ma, L. (1997), 'VoroGlide, interaktive Voronoi-Diagramme'. On-line at `http://wwwpi6.fernuni-hagen.de/Geometrie-Labor/VoroGlide/`.

Ingerson, T. E. & Buvel, R. L. (1984), 'Structure in asynchronous cellular automata', *Physica D* **10**(1–2), 59–68.

Isaac, D. (1996), *Property Development: Appraisal and Finance*, London, Macmillan.

Jacobs, J. (1961; 1994), *The Death and Life of Great American Cities*, Random House; Penguin Books, London.

Jennings, C. (2000), 'Lost and Found'. Pages 18–19, *Space* supplement to *The Guardian* newspaper, 20 January.

Johnson, J. H. (1981), 'Q-discrimination analysis', *Environment and Planning B: Planning & Design* **8**(4), 419–434.

Johnson, J. H. (1983), 'Q-analysis: a theory of stars', *Environment and Planning B: Planning & Design* **10**(4), 457–470.

Kain, J. F. & Apgar, W. C. (1985), *Housing and Neighbourhood Dynamics: A Simulation Study*, Harvard University Press, Cambridge, MA and London.

Kanji, G. K. (1993), *100 Statistical Tests*, Sage Publications, London.

Kauffman, S. A. (1984), 'Emergent properties in random complex automata', *Physica D* **10**, 145–156.

Kauffman, S. A. (1995), *At Home in the Universe: The Search for Laws of Complexity*, Penguin, London.

Keiller, P. (1999), *Robinson in Space*, Vol. 7 of *Topographics*, Reaktion Books, London.

Krafta, R. (1994), 'Modelling intraurban configurational development', *Environment and Planning B: Planning & Design* **21**, 67–82.

Krafta, R. (1996), 'Urban convergence: morphology and attraction', *Environment and Planning B: Planning & Design* **23**, 37–48.

Krüger, M. J. T. (1979*a*), 'An approach to built form connectivity at an urban scale: system description and its representation', *Environment and Planning B: Planning & Design* **6**, 67–88.

Krüger, M. J. T. (1979*b*), 'An approach to built form connectivity at an urban scale: variations of connectivity and adjacency measures amongst zones and other related topics', *Environment and Planning B: Planning & Design* **6**, 305–320.

Krüger, M. J. T. (1980), 'An approach to built-form connectivity at an urban scale: relationships between built-form connectivity, adjacency measures, and urban spatial structure', *Environment and Planning B: Planning & Design* 7(2), 163–194.

Krüger, M. J. T. (1981*a*), 'An approach to built-form connectivity at an urban scale: modelling the distribution of partitions and built-form arrays', *Environment and Planning B: Planning & Design* **8**(1), 41–56.

Krüger, M. J. T. (1981*b*), 'An approach to built-form connectivity at an urban scale: modelling the disaggregation of built forms by types', *Environment and Planning B: Planning & Design* **8**(1), 57–72.

Krüger, M. J. T. (1989), On node and axial grid maps: distance measures and related topics, Unpublished paper, UCL Bartlett School of Architecture and Planning, London.

Kuipers, B. J. (1982), 'The 'map in the head' metaphor', *Environment and Behaviour* **14**, 202–220.

Lake, I. R., Lovett, A. A., Bateman, I. J. & Langford, I. H. (1998), Modelling environmental influences on property prices in an urban environment, *in* '3rd International Conference on Geocomputation', 17–19 September, University of Bristol, School of Geographical Sciences.

Langton, C. G. (1990), 'Computation at the edge of chaos: phase transitions and emergent computation', *Physica D* **42**, 12–37.

Le Corbusier (1929), *The City of To-morrow and Its Planning*, John Rodher, London.

Lee, K. & Lee, H. (1998), 'A new algorithm for graph-theoretic nodal accessibility measurement', *Geographical Analysis* **30**(1), 1–14.

Lefebvre, H. (1991), *The Production of Space*, Blackwell, Oxford, England.

Levy, S. (1992), *Artificial Life: The Quest for a New Creation*, Pantheon Books, New York.

Ley, D. (1980), 'Liberal ideology and the postindustrial city', *Annals of the Association of American Geographers* **70**(2), 238–258.

Ley, D. (1981), 'Inner-city revitalization in Canada: a Vancouver case study', *Canadian Geographer* **XXV**(2), 124–148.

Ley, D. (1986), 'Alternative explanations for inner-city gentrification: a Canadian assessment', *Annals of the Association of American Geographers* **76**(4), 521–535.

Ley, D. (1987), 'Reply: the rent gap revisited', *Annals of the Association of American Geographers* **77**(3), 465–468.

Ley, D. (1994), 'Gentrification and the politics of the new middle class', *Environment and Planning D: Society and Space* **12**, 53–74.

Ley, D. (1996), *The New Middle Class and the Remaking of the Central City*, Oxford Geographical and Environmental Studies, Oxford University Press, Oxford, England and New York.

Lindgren, K. (1987), 'Correlations and information in random cellular automata', *Complex Systems* **1**, 529–543.

Lindgren, K. & Nordahl, M. G. (1988), 'Complexity measures and cellular automata', *Complex Systems* **2**, 409–440.

Lorrain, F. & White, H. C. (1971), 'Structural equivalence of individuals in social networks', *Journal of Mathematical Sociology* **1**, 49–80.

Luccio, F. & Sami, M. (1969), 'On the decomposition of networks in minimally interconnected subnetworks', *IEEE transactions* **CT-16**, 184–188.

Lynch, J. F. (1994), A phase transition in random boolean networks, *in* R. A. Brooks & P. Maes, eds, 'Artificial Life IV Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems', The MIT Press, Cambridge, MA and London, pp. 236–245.

Lynch, K. (1960), *The Image of the City*, MIT Press, Cambridge MA.

MacEachran, A. M. (1995), *How Maps Work: Representation, Visualization and Design*, The Guilford Press, New York.

MacGill, J. (2000), 'Introduction to GeoTools'. On-line at `http://www.ccg.leeds.ac.uk/geotools/`.

MacGill, S. M. (1984), 'Cluster analysis and Q-analysis', *International Journal of Man-Machine Studies* **20**, 595–604.

Macmillan, W. D. (1999), Cellular strategies for the simulation of human spatial systems, *in* '25-28 July, 4th International Conference on GeoComputation', 25-28 July, Mary Washington College, Fredericksburg, VA.

March, L. & Steadman, P. (1971), *The Geometry of Environment: An Introduction to Spatial Organization in Design*, RIBA Publications, London.

March, L., Echenique, M. & Dickens, P. (1971), 'Models of environment: Polemic for a structural revolution', *Architectural Design* **XLI**, 275.

Martin, L. & March, L., eds (1972), *Urban Space and Structures*, Cambridge University Press, Cambridge, England.

Minar, N., Burkhart, R., Langton, C. G. & Askenazi, M. (1996), 'The swarm simulation system: a toolkit for building multi-agent simulations'. On-line at `http://www.swarm.org/archive/swarmdocs.ps`.

Mitchell, M., Crutchfield, J. P. & Hraber, P. T. (1994), Dynamics, computation, and the "Edge of Chaos": A re-examination, *in* G. Cowan, D. Pines & D. Meltzer, eds, 'Complexity: Metaphors, Models and Reality', Vol. XIX of *Santa Fe Institute Studies in the Sciences of Complexity*, Addison-Wesley, Reading, MA, pp. 497–513.

Montello, D. R. (1993), Scale and multiple psychologies of space, *in* A. U. Frank & I. Campari, eds, 'Spatial Information Theory: A Theoretical Basis for GIS', Vol. 716 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 312–321.

Morrill, R. (1995), 'Racial segregation and class in a liberal metropolis', *Geographical Analysis* **27**(1), 22–41.

Moudon, A. V. (1997), 'Urban morphology as an emerging interdisciplinary field', *Urban Morphology* **1**(1), 3–10.

Mumford, L. (1961), *The City in History: Its Origins, Its Transformations and Its Prospects*, Harcourt Brace, New York.

Nagel, K., Beckman, R. J. & Barrett, C. L. (1998), TRANSIMS for transportation planning, Technical Report LA-UR 98–4389, Los Alamos National Laboratory. On-line at `http://www-transims.tsasa.lanl.gov/PDF_Files/LAUR98-4389.pdf`.

Nerlich, G. (1994), *The Shape of Space*, 2nd edn, Cambridge University Press, Cambridge, England.

Nicol, C. (1996), 'Interpretation and compatibility of house-price series', *Environment and Planning A* **18**, 119–133.

Nieminen, J. (1974), 'On the centrality in a graph', *Scandinavian Journal of Psychology* **15**, 322–336.

Okabe, A., Boots, B. & Sugihara, K. (1992), *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, Wiley, Chichester.

Okabe, A., Boots, B. & Sugihara, K. (1994), 'Nearest neighbourhood operations with generalized Voronoi diagrams: a review', *International Journal of Geographical Information Systems* **8**, 43–71.

Openshaw, S. (1991), 'A view of the GIS crisis in geography, or, using GIS to put Humpty-Dumpty back together again', *Environment and Planning A* **23**(5), 621–628.

Openshaw, S. (1992), 'Further thoughts on geography and GIS: a reply', *Environment and Planning A* **24**(4), 463–466.

Openshaw, S. & Turton, I. (1998), Geographical research using lifestyles databases, *in* 'RGS-IBG annual conference', 5–8 January, Kingston University, Guilford, England.

O'Sullivan, D. (forthcoming), 'Graph cellular automata: a generalised urban and regional model', *Environment and Planning B: Planning & Design*.

O'Sullivan, D. & Haklay, M. (forthcoming), 'Agent-based models and individualism: Is the world agent-based?', *Environment and Planning A*.

Peponis, J., Wineman, J., Bafna, S., Rashid, M. & Hong Kim, S. (1998), 'On the generation of linear representations of spatial configuration', *Environment and Planning B: Planning & Design* **25**(4), 559–576.

Philo, C. (1998), 'Reconsidering quantitative geography: the things that count', *Environment and Planning A* **30**(2), 191–201.

Phipps, M. (1989), 'Dynamical behaviour of cellular automata under the constraint of neighborhood coherence', *Geographical Analysis* **21**(3), 197–215.

Phipps, M. & Langlois, A. (1997), 'Spatial dynamics, cellular automata, and parallel processing computers', *Environment and Planning B: Planning & Design* **24**(2), 193–204.

Pirie, G. H. (1979), 'Measuring accessibility: a review and proposal', *Environment and Planning A* **11**, 299–312.

Portugali, J., Benenson, I. & Omer, I. (1997), 'Spatial cognitive dissonance and sociospatial emergence in a self-organizing city', *Environment and Planning B: Planning & Design* **24**(2), 263–285.

Poundstone, W. (1985), *The Recursive Universe*, Morrow, New York.

Prigogine, I. & Stengers, I. (1984), *Order out of Chaos: Man's New Dialogue with Nature*, Bantam Books, Toronto and New York.

Redfern, P. A.. (1997*a*), 'A new look at gentrification: 1. Gentrificatian and domestic technologies', *Environment and Planning A* **29**(7), 1275–1296.

Redfern, P. A.. (1997*b*), 'A new look at gentrification: 2. A model of gentrification', *Environment and Planning A* **29**(8), 1335–1354.

Reichenbach, H. (1958), *The Philosophy of Space and Time*, Dover, New York.

Robson, G. & Butler, T. (1998), Plotting the middle classes in London, *in* 'Cities at the Millennium Conference', 17–19 December, Royal Institute of British Architects, London.

Rose, D. (1984), 'Rethinking gentrification: beyond the uneven development of marxist urban theory', *Environment and Planning D: Society and Space* **1**(1), 47–74.

Rossi, A. (1982), *The Architecture of the City*, MIT Press, Cambridge, MA.

Sabidussi, G. (1966), 'The centrality index of a graph', *Psychometrika* **31**, 581–603.

Sack, R. D. (1980), *Conceptions of Space in Social Thought: A Geographic Perspective*, Critical Human Geography, Macmillan Press Ltd, London.

Sack, R. D. (1997), *Homo Geographicus: A Framework for Action, Awareness and Moral Concern*, Johns Hopkins University Press, Baltimore, MD and London.

Sadalla, E. K. & Magel, S. (1980), 'The perception of traversed distance', *Environment and Behaviour* **12**, 65–79.

Sadalla, E. K. & Staplin, L. J. (1980), 'The perception of traversed distance: intersections', *Environment and Behaviour* **12**, 167–182.

Sailer, L. D. (1978), 'Structural equivalence: meaning and definition, computation and application', *Social Networks* **1**, 73–90.

Sayer, A. (1976), 'A critique of urban modelling: from regional science to urban and regional political economy', *Progress in Planning* **6**(3), 187–254.

Sayer, A. (1982), 'Explanation in economic geography: abstraction versus generalization', *Progress in Human Geography* **6**, 68–88.

Sayer, A. (1992), *Method in Social Science: A Realist Approach*, 2nd edn, Routledge, London.

Sayer, A. (2000), *Realism in Social Science*, Sage, London.

Schaffer, R. & Smith, N. (1986), 'The gentrification of Harlem?', *Annals of the Association of American Geographers* **76**(3), 347–365.

Schelling, T. C. (1971), 'Dynamic models of segregation', *Journal of Mathematical Sociology* **1**, 143–186.

Schelling, T. C. (1978), *Micromotives and Macrobehaviour*, Norton, New York.

Seidman, S. B. (1983), 'Internal cohesion of LS sets in graphs', *Social Networks* **5**, 97–107.

Semboloni, F. (2000, forthcoming), 'The growth of an urban cluster into a dynamic self-modifying spatial pattern', *Environment and Planning B: Planning & Design*.

Shariff, A. R. B. M., Egenhofer, M. J. & Mark, D. M. (1998), 'Natural-language spatial relations between linear and areal objects: the topology and metric of English language terms', *International Journal of Geographical Information Science* **12**(3), 212–245.

Shaw, W. (1998), Constructing cultures of acceptability, *in* 'Cities at the Millennium Conference', 17–19 December, Royal Institute of British Architects, London.

Sheppard, E. & Barnes, T. J. (1990), *The Capitalist Space Economy: Geographical Analysis after Ricardo, Marx and Sraffa*, Unwin Hyman, London.

Sklar, L. (1974), *Space, Time and Spacetime*, University of California Press, Berkeley, CA.

Slater, T. R. (1980), The analysis of burgages in medieval towns, Working Paper 4, Department of Geography, University of Birmingham, England.

Slater, T. R. (1998), *Creating Spaces: Model Settlements, Garden Cities and New Towns*, Paul Chapman, London.

Smith, N. (1979*a*), 'Gentrification and capital: practice and ideology in Society Hill', *Antipode* **11**(3), 24–35.

Smith, N. (1979*b*), 'Toward a theory of gentrification: a back to the city movement by capital not people', *Journal of the American Planning Association* **45**(October), 538–548.

Smith, N. (1982), 'Gentrification and uneven development', *Economic Geography* **58**, 139–155.

Smith, N. (1987*a*), 'Gentrification and the rent gap', *Annals of the Association of American Geographers* **77**(3), 462–465.

Smith, N. (1987*b*), 'Of yuppies and housing: gentrification, social restructuring, and the urban dream', *Environment and Planning D: Society and Space* **5**, 151–172.

Smith, N. (1992), 'Blind man's buff, or Hamnett's philosophical individualism in search of gentrification', *Transactions of the Institute of British Geographers* **NS17**, 110–115.

Smith, N. (1996), *The New Urban Frontier: Gentrification and the Revanchist City*, Routledge, London and New York.

Soja, E. W. (1989), *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*, Verso, London and New York.

Sparrow, M. K. (1993), 'A linear algorithm for computing automorphic equivalence classes: the numerical signatures approach', *Social Networks* **15**, 151–170.

Steadman, P. (1983), *Architectural Morphology: An Introduction to the Geometry of Building Plans*, Pion Ltd, London.

Stephenson, K. & Zelen, M. (1989), 'Rethinking centrality: methods and examples', *Social Networks* **11**, 1–37.

Stern, D. I. (1992), 'Do regions exist? Implications of synergetics for regional geography', *Environment and Planning A* **24**(10), 1431–1448.

Takeyama, M. (1996), Geo-Algebra: A Mathematical Approach to Integrate Spatial Modelling and GIS, unpublished Ph.D. thesis, UCSB Santa Barbara, CA.

Takeyama, M. & Couclelis, H. (1997), 'Map dynamics: integrating cellular automata and GIS through Geo-Algebra', *International Journal of Geographical Information Science* **11**, 73–91.

Taylor, P. J. & Overton, M. (1991), 'Further thoughts on geography and GIS', *Environment and Planning A* **23**(8), 1087–1090.

Teklenburg, J. A. F. & Timmerman, H. J. P. (1993), 'Space syntax: standardised integration measures and some simulations', *Environment and Planning B: Planning & Design* **20**, 347–357.

Thurstain-Goodwin, M. & Unwin, D. (forthcoming), 'Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations', *Transactions in GIS*.

Tobler, W. R. (1979), Cellular geography, *in* S. Gale & G. Olsson, eds, 'Philosophy in Geography', D. Reidel Publishing Company, Dordrecht, The Netherlands, pp. 379–386.

Tollis, I. G., Di Battista, G., Eades, P. & Tamassia, R. (1999), *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice Hall, Englewood Cliffs, NJ.

Torrens, P. (2000), How land-use-transportation models work, Working Paper 20, Centre for Advanced Spatial Analysis, University College London. On-line at `http://www.casa.ucl.ac.uk/working_papers.htm`.

Townsend, P., Phillimore, P. & Beattie, A. (1988), *Health and Deprivation: Inequality and the North*, Croom Helm, London.

Tuan, Y.-f. (1977), *Space and Place: The Perspective of Experience*, University of Minnesota Press, Minneapolis, MI.

Unwin, T. (2000), 'A waste of space? Towards a critique of the social production of space...', *Transactions of the Institute of British Geographers* **25**, 11–29.

Urban Task Force (1999), *Towards an Urban Renaissance*, Department of the Environment, Transport and the Regions, London.

Vance, J. E. (1990), *The Continuing City: Urban Morphology in Western Civilization*, Johns Hopkins University Press, Baltimore, MD.

Voltaire (1738; 1967), *The Elements of Sir Isaac Newton's Philosophy*, Originally published by Stephen Austen, reproduction by Frank Cass & Co. Ltd, London.

Waddington, C. H. (1977), *Tools for Thought*, Jonathan Cape, London.

Ward, D. P., Murray, A. T. & Phinn, S. R. (1999), An optimized cellular automata approach for sustainable urban development in rapidly urbanizing regions, *in* '4th International Conference on GeoComputation', 25–28 July, Mary Washington College, Fredericksburg VA. On-line at `http://www.geovista.psu.edu/geocomp/geocomp99/index.htm`.

Warde, A. (1991), 'Gentrification as consumption: issues of class and gender', *Environment and Planning D: Society and Space* **9**, 223–232.

Wassermann, S. & Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, England.

Watts, D. J. (1999), *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton Studies in Complexity, Princeton University Press, Princeton, NJ.

Watts, D. J. & Strogatz, S. H. (1998), 'Collective dynamics of 'small-world' networks', *Nature* **393**(4 June), 440–442.

Webster, C. J. & Wu, F. (1999*a*), 'Regulation, land use mix and urban performance. Part 1 Performance', *Environment and Planning A* **31**(8), 1529–1545.

Webster, C. J. & Wu, F. (1999*b*), 'Regulation, land use mix and urban performance. Part 2 Theory', *Environment and Planning A* **31**(7), 1433–1442.

Weißtein, E. W. (1996), 'Eric weißtein's world of mathematics'. On-line at `http://mathworld.wolfram.com/`.

White, R. (1977), 'Dynamic central place theory: results of a simulation approach', *Geographical Analysis* **9**, 226–243.

White, R. (1998), 'Cities and cellular automata', *Discrete Dynamics in Nature and Society* **2**, 111–125.

White, R. (1999), High resolution integrated modelling of the spatial dynamics of urban and regional systems, *in* '4th International Conference on GeoComputation', 25–28 July, Mary Washington College, Fredericksburg VA.

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

White, R. & Engelen, G. (1997), 'Cellular automata as the basis of integrated dynamic regional modelling', *Environment and Planning B: Planning & Design* **24**(2), 235–246.

White, R., Engelen, G. & Uljee, I. (1997), 'The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics', *Environment and Planning B: Planning & Design* **24**(3), 323–343.

Whitehand, J. W. R. (1967), 'Fringe belts: a neglected aspect of urban geography', *Transactions of the Institute of British Geographers* **41**, 223–233.

Whitehand, J. W. R. (1972), 'Building cycles and the spatial pattern of urban growth', *Transactions of the Institute of British Geographers* **56**, 39–55.

Whitehand, J. W. R. (1987*a*), *The Changing Face of Cities: A Study of Development Cycles and Urban Growth*, Basil and Blackwell Ltd, Oxford, England.

Whitehand, J. W. R. (1987*b*), Urban morphology, *in* M. Pacione, ed., 'Historical Geography Progress and Prospect', Croom Helm, London.

Whitehand, J. W. R. (1992), 'Recent advances in urban morphology', *Urban Studies* **29**(3/4), 619–636.

Whitehand, J. W. R., ed. (1981), *The Urban Landscape: Historical Development and Management: Papers by M. R. G. Conzen*, Vol. 13 of *Institute of British Geographers Special Publications*, Academic Press, London.

Wilson, R. J. (1996), *Introduction to Graph Theory*, 4th edn, Longman, Harlow, England.

Wolfram, S. (1983), 'Statistical mechanics of cellular automata', *Review of Modern Physics* **55**, 601–643.

Wolfram, S. (1984*a*), 'Computation theory of cellular automata', *Communications in Mathematical Physics* **96**, 15–57.

Wolfram, S. (1984*b*), 'Universality and complexity in cellular automata', *Physica D* **10**, 1–35.

Wood, J. D., Fisher, P. F., Dykes, J. A., Unwin, D. J. & Stynes, K. (1999), 'The use of the landscape metaphor in understanding population data', *Environment and Planning B: Planning & Design* **26**(2), 281–295.

Wright, F. L. (1945), *When Democracy Builds*, University of Chicago Press, Chicago.

Wu, F. (1999), A simulation approach to urban changes: experiments and observations on fluctuations in cellular automata, *in* P. Rizzi, ed., 'Sixth International Conference on Computers in Urban Planning and Urban Management', 7–10 Spetember, Venice, Italy. On-line at `http://brezza.iuav.it/stratema/cupum/`.

Wuensche, A. (1994), Complexity in one-D cellular automata: gliders, basins of attraction and the Z parameter, Working Paper 94-04-025, Santa Fe Institute. Available on request, abstract on-line at `http://www.santafe.edu/sfi/publications/94wplist.html`.

Wuensche, A. (1998), Classifying cellular automata automatically, Working Paper 98-02-018, Santa Fe Institute. On-line at `http://www.santafe.edu/sfi/publications/98wplist.html`.

Wuensche, A. (1999), 'Classifying cellular automata automatically: finding gliders, filtering, and relating space-time patterns, attractor basins, and the Z parameter', *Complexity* **4**, 47–66.

Xie, Y. (1994), Analytical Models and Algorithms for Cellular Urban Dynamics, unpublished Ph.D. thesis, SUNY Buffalo, NY.

Xie, Y. (1996), 'A generalized model for cellular urban dynamics', *Geographical Analysis* **28**(4), 350–373.

Yung, C.-F. & King, R. J. (1998), 'Some tests for the rent gap theory', *Environment and Planning A* **30**, 523–542.

Zukin, S. (1982), *Loft Living: culture and capital in urban change*, Johns Hopkins University Press, Baltimore, MD.

Appendix A

Program file formats

A.1 The *graphca* and *Gentrification* .gca file formats

There are two versions of the .gca file format used to store details of a graph-CA model: version 2.0 and 3.0. Both are readable by the *graphca* program. Only version 3.0 is readable by the *Gentrification* program. In version 3.0 vertices and edges are stored in two separate files; a .gca file of vertices, and a .edg file of edges. The *graphca* program can only read version 3.0 .gca files, and *not* .edg files. The *Gentrification* program can read both files.

The *graphca* and *Gentrification* program file readers are rather fragile, and particularly sensitive to the correct spacing of input text, therefore in all the definitions in this appendix the symbol ␣ indicates a single white space. The correct number of spaces, as shown, should be included in files, unless otherwise stated.

A.1.1 The version 2.0 .gca file format

In this format, graph data is recorded in three sections, vertices, analysis data, and edges. The first line of every file must be

`GraphCA␣GCA␣File␣v2.0`

This string is checked by the program before it reads the file, and if it is incorrect reading will terminate. A list of vertex definitions follows.

A full vertex record is formatted as follows

```
[vertex]
␣␣␣␣␣␣␣␣[label]␣␣␣␣␣␣␣0
␣␣␣␣␣␣␣␣[x]␣␣␣␣␣␣␣␣␣␣␣10
␣␣␣␣␣␣␣␣[y]␣␣␣␣␣␣␣␣␣␣␣100
␣␣␣␣␣␣␣␣[state]␣␣␣␣␣␣␣0
␣␣␣␣␣␣␣␣[in-depths]␣␣␣36,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,1
␣␣␣␣␣␣␣␣[out-depths]␣␣36,1,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,1
[end]
```

where `[label]` may be any string, and `[x]` and `[y]` are floating-point Easting-Northing coordinates for the vertex. Other elements are optional. The `[end]` flag indicates the end of one vertex definition and the `[vertex]` flag the start of another one. The program assigns sequential internal numeric IDs to vertices so the labels specified may be any useful textual information.

The optional `[state]` is an integer between 0 and 9, specifying the cell state. The two optional elements `[in-depths]` and `[out-depths]` are comma-separated lists of the number of vertices in the graph, followed by the number of vertices at successive distances 0, 1, 2... from the vertex. The `[in-depths]` list records the numbers of vertices moving in reverse along arcs in the graph, whereas the `[out-depths]` list records the numbers of vertices encountered at each distance moving along arcs in the correct direction. In an undirected simple graph, both lists will be the same, and are not both required, although this will produce unpredictable results. This format closely mirrors the internal storage of analysis results in the *graphca* program, and enables rapid calculation of derived analysis results, since it summarises the whole graph structure from the perspective of every vertex.

After the vertices, is an *optional* verbose set of graph structural analysis data. A full vertex record with analysis data looks like this:

```
[short-paths]
␣␣␣␣␣␣␣␣[id]␣␣␣␣␣␣␣␣␣␣25
```

```
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣18,7
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣17,6,8
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣16,5,9
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣15,4,10
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣14,3,11
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣13,2,12
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣12,1,13
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣11,0,14
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣10,15,35
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣9,16,34
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣8,17,33
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣7,18,32
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣6,19,31
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣5,20,30
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣4,21,29
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣3,22,28
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣2,23,27
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣1,24,26
␣␣␣␣␣␣␣␣[set]␣␣␣␣␣␣␣␣␣0,25
[end]
```

All the indented elements are optional. `[short-paths]` flags the start of a set of analysis data; `[id]` is the sequential integer ID of the vertex to which the data applies, where 0 corresponds to the first vertex in the vertex list section of the file, and increments sequentially up to the final vertex (a 100 vertex graph will have `[id]`s from 0 to 99); and `[end]` flags the end of the record for this vertex. the `[id]` *must come first* in the `[short-paths]` record or an error will result.

The `[set]`s of vertex IDs record the IDs of vertices at the distance from the specified vertex given by the first number in the numerical list. Thus the first `[set]` above indicates that vertex 7 is at a distance of 18 away from this one (vertex 25). Vertices 6 and 8 are at a distance 17 from this one, and so on. This data is only required for the calculation of 'lagged' spatial information values in the analysis package and will only rarely be required.

The final (also *optional*) section contains edge data. All edges are regarded as directed, so a simple undirected graph will include edges twice, once in each direction. An edge record simply consists of

```
[edge]
        [child]       8
        [parent]      7
[end]
```

The `[edge]` flag indicates the start of a record, and the `[end]` tag the end. The `[child]` and `[parent]` tags indicate respectively, the 'to' and 'from' vertices of the edge, again using the internally assigned ID numbers, which correspond to the order in which vertices appear in the vertex list. The `[child]` and `[parent]` elements may appear in either order. If either is missing the program will assign vertex 0 by default for the missing element. If both are missing no edge will be created.

Finally the file is terminated with an optional

```
[eof]
```

flag. In fact, any line other than `[vertex]`, `[short-paths]` or `[edge]` will terminate file reading.

A.1.2 The version 3.0 file formats

As noted above in the version 3.0 file format, vertices and edges are stored in two separate files, with file extensions `.gca` and `.edg` respectively.

The version 3.0 `.gca` vertex file format

The `.gca` file contains both the vertex definitions and the detailed graph analysis results. The first line of a file must be

```
GraphCA GCA File v3.0
```

The vertex definitions then follow. A full vertex record looks like this:

```
[vertex]
        [label]       9496
        [x]           533117.031
        [y]           182304.648
        [value]       57.677
        [income]      54.230
```

```
␣␣␣␣␣␣␣␣␣[fixed]␣␣␣␣␣␣␣true
[end]
```

This is similar to the version 2.0 format and similarly tagged lines have the same meanings as before. The new `[value]` and `[income]` records are floating-point numbers between 0.0 and 100.0 for the two state variables used in the *Gentrification* model. The `[fixed]` record is either `true` or `false` and indicates whether or not the state of this cell (vertex) may be altered by the model operation. All three new records are optional. If they are missing the *Gentrification* program will assign default values 0.0, 0.0 and `false` respectively. The *graphca* will ignore these records if they appear.

The records `[state]`, `[in-depths]` and `[out-depths]` may be included in a version 3.0 file. The `[state]` will be ignored by the *Gentrification* program, but it will read the `[in-depths]` and `[out-depths]` normally.

Vertex definitions may be followed by detailed graph `[short-paths]` analysis results as in version 2.0. These will be read by both programs, as described above.

Edge definitions are not normally included in a version 3.0 `.gca` file. However, the (optional) detailed analysis results section, may optionally be followed immediately by a set of `[adjacencies]` (see below). These will be correctly read by both programs.

The `.gca` file should terminate, as before, with a `[eof]` flag.

The version 3.0 .edg edge file format

The `.edg` file records graph edges in an adjacency matrix format. This format is more compact than the version 2.0 format, which is a considerable advantage with relatively dense large graphs.

The first line of the `.edg` file is as for the `.gca` file:

```
GraphCA␣GCA␣File␣v3.0
```

There then follows an adjacency matrix in the following format

```
[adjacencies]
0100000001
101
0101
00101
000101
0000101
00000101
000000101
0000000101
[end]
```

Each row records the existence (1) or non-existence (0) of a directed edge from the vertex corresponding to the line in the text file, to the vertex indicated by position in the string. Trailing 0s may be omitted as shown here. The row and column order corresponds to the order in which vertices have been read in from the `.gca` file. The file shown describes the 10 vertex cycle graph C_{10}, where each vertex is connected to its immediate neighbours.

The `.edg` file should terminate with `[eof]` as usual.

A.2 *graphca* `.rul` transition rule files

A `.rul` file specifies the graph-CA transition rules for use in the *graphca* program. A typical file looks like this:

```
GraphCA␣RuleSet␣File␣v1.1
Dead,0,96,224
Alive,255,128,0
transitions
Dead,Alive,n,X,X,0.28,0.38
Alive,Alive,n,X,X,0.28,0.50
Alive,Dead,d,X,X,X,X
Dead,Dead,d,X,X,X,X
ends
```

The first line is required for the file to be recognised by the *graphca* program. Lines up to the `transitions` declaration each define an allowed cell state. The first term is the state name, and the three comma-separated integers specify the red, green and blue colour components for display of the state on screen (these may subsequently be changed by the user). Note that as with vertices states are assigned an internal sequential ID number; in this case Dead will have internal ID 0 and Alive will be 1. Up to ten states, from 0 to 9 may be declared.

The state declarations are followed by definitions of the various transitions. The first two elements in each line are the 'from' and 'two' states for the transition. The next item, which may be one of `n`, `d` or `na`, indicates that a rule is 'normal', 'default' or 'not applicable', respectively.

The remainder of each line is a sequence of pairs of lower and upper bounds on the fraction of neighbouring cells required in a neighbourhood for that transition to occur. The order of appearance of pairs in this list corresponds to the definition of states in the first part of the file. An `X` in any position indicates that the lower or upper bound may be ignored. Thus the first rule here may be read as "the transition from Dead to Alive will occur by normal application of this rule, if between 0.28 and 0.38 of neighbouring cells are in the Alive state". The file shown describes the Game of Life CA where each cell has nine neighbours (including itself. It approximately describes the Game of Life for other lattice structures — bearing in mind comments in this thesis about the difficulty of specifying rules in a neighbourhood-size independent manner.

A.3 *graphca* `.cfg` configuration files

Finally, a `.cfg` file specifies a complete system configuration in terms of cell states, for the *graphca* program. A typical file looks like this:

```
GraphCA␣State␣File␣v1.0
10000101011100000001110101100110000
```

As usual, the first line is required for the *graphca* to be able to identify the file. The remaining line simply lists the states, by internal ID (from 0 to 9), for the graph-CA vertices in sequence number order — that is corresponding to their order of appearance in the `.gca` file vertex definition list.