Spatial Topology and its Structural Analysis based on the Concept of Simplicial Complex

Bin Jiang[1*] and Itzhak Omer[2]

[1]Division of Geomatics, Dept. of Technology and Built Environment, University of Gävle, SE-801 76 Gävle, Sweden, Email: bin.jiang@hig.se

[2]Department of Geography and Human Environment, The Environmental Simulation Laboratory, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, 69978, Israel, Email: omery@post.tau.ac.il

ABSRACT: This paper introduces a model that identifies spatial relationships for a structural analysis based on the concept of simplicial complex. The spatial relationships are identified through overlapping two map layers, namely a primary layer and a context layer. The identified spatial relationships are represented as a simplical complex, in which simplices and vertices represent respectively two layers of objects. The model relies on simplical complex, a key concept in Q-analysis for overall structure analysis. To quantify structural properties of individual primary objects (or simplices), we define a set of measures by considering multidimensional chains of connectivity. With the model, the interaction and relationships with a geographic system are modeled both from local and global perspectives. The structural properties and modeling capabilities are illustrated with some examples.

1. INTRODUCTION

The first law of geography nicely describes a nature of geographic systems in which "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). The "things" can be understood from two different perspectives. The first is as established in existing GIS where *graphic primitives* are interconnected, and as such constitute different layers. For instance, a street network is considered a network of interconnected street segments though common nodes (or junctions); or different land use parcels constitute a land use layer with interconnected polygons. Behind the layers is the notion of topology that governs the relationship of various graphic primitives. The second understanding is from a geographic perspective that different *geographic objects* (in contrast to the graphic primitives) are interconnected and thus constitute a kind of network in terms of adjacency and connectivity at relative space, defined by the relationship between the objects themselves. For instance, a street network can be perceived from a purely topological perspective (Jiang and Claramunt 2004). The law appears to suggest an overall structure of any geographic system. However, actual structures may vary significantly from one system to another, which deserves in-depth study. This is what we intend to explore in this paper. More specifically, we intend to develop a model that can be used to explore structural properties of a geographic system as a whole, and the individual objects within the system.

Geographic systems, like many other natural and artificial systems, can be represented as a network in which nodes represent constituents and links represent possible relations between the nodes. The network representation can help illustrate various aspects of a geographic system, such as efficiency, robustness and stability. Distinct from previous studies mostly based on graph theory for network representation, this paper adopts the concept of simplicial complex, as defined in the theory of Q-analysis (Atkin 1974), for the representation and structural analysis of geographic systems. The

[*] Current address: Department of Land Surveying and Geo-informatics, Hong Kong Polytechnical University, Hung Hom, Kowloon, Hong Kong, Email: bin.jiang@polyu.edu.hk

structural analysis is fundamental to understanding flows of goods, information and people within a geographic system. Although Q-analysis shares with graph theory some similarity in terms of topology (Earl and Johnson 1981), simplicial complex representation is able to deal with structural properties from the perspective of multidimensional chains of connectivity.

In this paper, we direct our attention to an overall structure or topology of an entire geographic system, linking to how various parts (or objects) of the geographic system are interacted. In other words, our modeling effort is to take the interrelationship of spatial objects as a whole and explore its structure from both local and global levels. The study is also motivated by the emerging study of complex networks in a variety of disciplines to seek hidden structure of complex systems (Watts and Strogatz 1998, Dorogovtsev and Mendes 2003). Our model helps investigate structural properties and various structural analyses in terms of efficiency, dependency and stability of networks underlying the geographic systems.

The remainder of this paper is organized as follows. Section 2 defines the novel concept of spatial topology, a network representation of spatial relationships within a geographic system, illustrated with a simple example. Section 3 presents how Q-analysis in general and simplicial complex in particular can be used to characterize structural properties of a spatial topology. Section 4 introduces two new measures to characterize structural properties of individual objects within a geographic system. Finally section 5 concludes the paper and points out possible future work.

2. SPATIAL TOPOLOGY: DEFINITION AND ILLUSTRATION

Here we introduce a model to identify spatial relationships of geographic objects via common objects. To introduce the model, let's start with an example from social network. Two people, A and B, are not acquainted in a general social sense as colleagues or friends, but they can be considered as "adjacent", as both, for instance, belong to the same Geographic Information (GI) community. The community or organization people belong to is a context that keeps people "adjacent". Extending the case into a geographic context, we can say that, for instance, two houses are adjacent if they are both situated along a same street, or within a same district. We can further say that a pair of objects sharing more common objects is more "adjacent". Now let's turn to the formal model for identifying such spatial relationships. The model aims to represent spatial relationships of various objects as a simplicial complex. To this end, some formal definitions are also presented as well, but we are not implying that they are new.

2.1 Defining Spatial Topology

We start with a definition of map layer according to set theory. Map layer is defined as a set of spatial objects at a certain scale in a database or on a map. For example, $M = \{o_1, o_2, ..., o_n\}$, or $M = \{o_i \mid i = 1,2,...,n\}$ denotes a map layer (using a capital letter) that consists of multiple objects (using small letters). The objects can be put into four categories: point, line, area, and volume objects in terms of basic graphic primitives. Note that the definition of objects must be appropriate with respect to the modeling purpose. For instance, a street layer can be considered a set of interconnected street segments, or an interconnected named street depending on the modeling purpose (Jiang and Claramunt 2004); a city layer can be represented as a point or area object depending on the representation scale.

A spatial topology (*T*) is defined as a full set (note not subset) of the Cartesian product of two map layers, L and M, denoted by $T = L \times M$. To set up a spatial topology, we need to examine the relationship of every pair of objects from one map layer to another. As distinct to topological relationships based on possible intersections of internal, external and boundary of spatially extended objects (Egenhofor 1991), we simply take a binary relationship. That is, if an object λ is within, or intersects, another object *m*, we say there is a relationship $\lambda = (\lambda, m)$, otherwise no relationship, $\lambda = \varnothing$ [Note: the pair (λ, m) is ordered, and (m, λ) represents an inverse relation denoted by λ^{-1}]. The relationship can be simply expressed as "an object has a relationship to a context object". If a set of primary objects shares a common context object, we say the set of objects are adjacent or proximate. Thus two types of map layers can be distinguished: primary layer for the primary objects, with which a spatial topology is to be explored, and context layer, whose objects constitute a context for the primary objects. It is important to note that the context layer can be given in a rather abstract way with a set of features (instead of map objects). This way, the relationship from the primary to context objects can be expressed as "an object has certain features". For the sake of convenience and with notation $T = L \times M$, we refer to the first letter as the primary layer and the second the context layer.

The notion of spatial topology presents a network view as to how the primary objects become interconnected via the context objects. A spatial topology can be represented as a simplicial complex. Before examining the representation, we turn to the definition of simplicial complex (Atkin, 1977). A simplicial complex is the logical union of relevant simplices. Let's assume the elements of a set A form simplices (or polyhedra, denoted by σ^d where d is the dimension of the simplex); and the elements of a set *B* form vertices according to the binary relation λ, indicating that a pair of elements (a_i, b_j) from the two different sets *A* and *B*, $a_i \in A$ and $b_j \in B$, are related, the simplical complex can be denoted $K_A(B;\lambda)$. In general, each individual simplex is expressed as a *q*-dimensional geometrical figure with *q+1* vertices. The convergence or union of all the simplices forms the simplicial complex. For every relation λ it is feasible to consider the conjugate relation, λ^{-1}, by reversing the relations between two sets A and B by transposing the original incidence matrix. The conjugate structure is denoted as $K_B(A; \lambda^{-1})$.

A spatial topology can be represented as a simplicial complex where the simplices are primary objects, while vertices are context objects. Formally, the simplicial complex for the spatial topology $T = L \times M$ is denoted by $K_L(M;\lambda)$, where L represent the primary layer, and M the context layer, the relation between a primary object and context object $\lambda = (\lambda, m)$. A spatial topology can be represented as an incidence matrix Λ, where the columns represent objects with primary layer and the rows represent the objects with context layer. Formally it is represented as follows,

$$\Lambda_{ij} = \begin{cases} 1 : if \ \lambda = (i, j) \\ 0 : if \ \lambda = \varnothing \end{cases} \quad [1]$$

The incidence matrix is not symmetric, as $\lambda \neq \lambda^{-1}$ in general. However for two identical map layers i.e. $L = M$, the incidence matrix would be symmetric.

Spatial topology is not intended to replace the existing topology or spatial relationship representation, but rather to extend and enhance the existing ones for more advanced spatial analysis and modelling. The spatial topology can also be represented as a bipartite graph via a graph theoretical approach. However, the simplicial complex representation extends topological analysis in two aspects. First, spatial relationships can be explored both from local and global perspectives, while the existing ones are constrained with local relationships only. Secondly spatial topology is intended for structural analysis rather than for a simple query. In spite of the differences, spatial topology adopts existing topological settings for exploring structural properties of its own.

2.2 A Simple Example

In order to illustrate the concept of spatial topology, we shall take a look at a simple example. Let's assume with an environmental GIS, three pollution sources and their impact areas are identified, through a buffer operation, as a polygon layer. It is likely that the three polluted zones overlap each other. This constitutes a map layer, denoted by $Y = (y_j \mid j = 1,2,3)$. To assess the pollution impact on a set of locations with another map layer $X = (x_i \mid i = 1,2,...6)$, it is not sufficient to just examine which location is within which pollution zones. We should take a step further to put all locations within an interconnected context (a network view) using the concept of spatial topology. For instance, location x1 is out of pollution zone of y2, but it may get polluted through x2 and x5, assuming the kind of pollution is transmittable. Only under the network view are we able to investigate the pollution impact thoroughly.

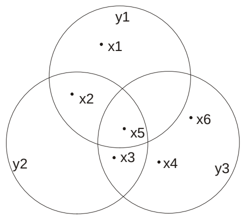

Figure 1: A simple example

In the example, the locations under pollution impact are of our primary interest. Using equation [1], the spatial topology can be represented as an incidence matrix as follows.

$$\Lambda_{36} = \begin{bmatrix} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ y_1 & 1 & 1 & 0 & 0 & 1 & 0 \\ y_2 & 0 & 1 & 1 & 0 & 1 & 0 \\ y_3 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

For the primary layer, the six locations in figure 1 and with respect to the columns of the matrix can be represented as six simplices as follows:

$$\sigma^0(x_1) =< y_1 >$$
$$\sigma^1(x_2) =< y_1, y_2 >$$
$$\sigma^1(x_3) =< y_2, y_3 >$$
$$\sigma^0(x_4) =< y_3 >$$
$$\sigma^2(x_5) =< y_1, y_2, y_3 >$$
$$\sigma^0(x_6) =< y_3 >$$

where the right part to the equation represents vertices that consist of a given simplex. The dimensionality of a simplex is represented by a superscript. For instance, $\sigma^2(y_5)$ denotes a 2-dimensional simplex or a 2-dimensinal face that consists of three vertices y_1, y_2, and y_3. The simplices and the related complex can be graphically represented as figure 2.

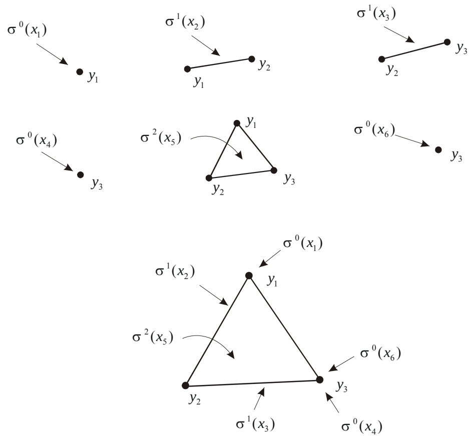

Figure 2: The six simplices and their union – the simplicial complex Kx(Y; λ)

Note that the primary and context object are relative and transversal (interchangeable) depending on different application interests. If for instance we take Y as the primary layer (whether it makes sense is another issue which we will not consider here), i.e. to transpose the incidence matrix defined in equation [1], then we would have three simplices as follows (see also figure 3),

$$\sigma^2(y_1) = < x_1, x_2, x_5 >$$
$$\sigma^2(y_2) = < x_2, x_3, x_5 >$$
$$\sigma^3(y_3) = < x_3, x_4, x_5, x_6 >$$

The union of the three simplices constitutes a three-dimensional complex (figure 3).

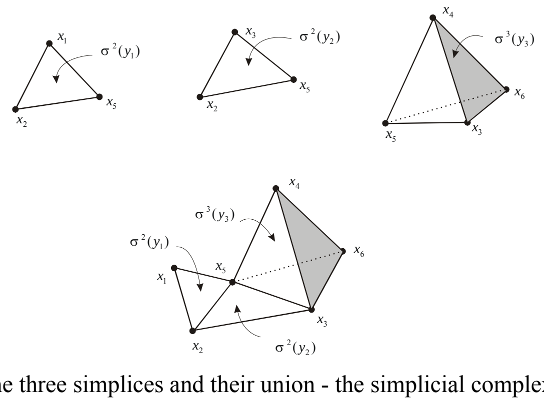

Figure 3: The three simplices and their union - the simplicial complex Ky(X; λ^{-1})

We have noted that a spatial topology can be represented as a geometric form (the simplicial complex). It is important to note the geometric representation makes little sense when the dimension of the simplicial complex exceed three because of human perceptual constraint. However, an algebraic representation to the spatial topology has no such constraint. The measures to characterize local properties to be introduced in the sections that follow are considered to enhance the algebraic method.

3. STRUCTURAL PROPERTIES OF A SPATIAL TOPOLOGY

A spatial topology can be represented as a connected simplicial complex for structural analysis based on Q-analysis theory. In this case, primary objects are represented as individual simplicies, while context objects as the vertices of simplices. For the primary layer, a pair of objects can be assessed to see if they share a common context object and how many common objects they share. If there are q+1 common context objects shared, we say the pair of primary objects is q-near. More specifically and for example, 0-nearness means that a pair of primary objects shares one context object, and 1-nearness means 2 context objects are shared. For the example in figure 3 the pair y_1-y_2 shares two objects, i.e. two vertices x_2 and x_5, so we say y_1 and y_2 are 1-near. If, on the other hand, there is no context object shared, i.e. –1-near, but each of the primary objects is q-near to one or several common primary objects, we say that the pair of objects is q-connected due to the transitive rule. For instance, the primary objects y_1 and y_3 have only one common context object (x_5), but they are 1-connected due to the primary object y_2 that serves as a transitional object; y_2 is 1-near with y_1 and with y_3. In a more general way, we say that two objects are q-connected if they are q-near or there is a chain of q-near simplices between them. That is, two objects can be connected through more than one transitional simplex (this is true of course for a case where there is more than 3 primary objects). As a general rule, two simplices are q-near, they would be also q-connected, but not vice versa.

The Q-analysis is based on the q-nearness and q-connectivity relations between the simplices of a given complex (or simplicial complex). A Q-analysis of a complex $K_L(M;\lambda)$ determines the number (#) of distinct equivalence classes, or q-connected components, for each level of dimension q ranging from 0 to q-1. The equivalence classes are decided by a rule as follows. If two simplices are q-connected (either q-near or q-connected), then they are in a same class. For example, the Q-analysis of the complex in figure 3 leads to the following equivalence classes at the different dimensional levels of q=0, q=1, q=2 and q=3. Each equivalence class is enclosed in the curly brackets.

q=3: $\{y_3\}$
q=2: $\{y_1\},\{y_2\},\{y_3\}$
q=1: $\{y_1,y_2,y_3\}$
q=0: $\{y_1,y_2,y_3\}$

The spatial topology can be analyzed by Q-analysis from a global perspective. That is, through counting the number of components at each q-level, we are able to get a structure vector Q. It is defined as follows:

$$Q = \{\overset{q-1}{\#_{q-1}} \ \overset{...}{...} \overset{2}{\#_2} \ \overset{1}{\#_1} \ \overset{0}{\#_0}\} \qquad [2]$$

where the superscripts of the structural vector denote the dimension q, i.e. 0, 1, 2, …,(q-1) from the most right to left. Large entries in Q indicate that the structure at a certain dimension is fragmented or disconnected, while small entries imply more integrity structure. For instance the structure vector for the complex in figure 3 $Q = \{\overset{3}{1} \ \overset{2}{3} \ \overset{1}{1} \ \overset{0}{1}\}$ indicates that there is one equivalence class (EC) at 0-dimensional level, one EC at 1-dimensional level, three ECs at 2- dimensional level, and one EC at 3-dimensional level. In other words, the complex is rather fragmented at the second dimension, and rather integrated at zero, first and third dimensions. Such an analysis can provide some critical insights into the structure of a spatial topology in terms of information flows, energy and goods exchanges.

From a local perspective we could assess the uniqueness of individual primary objects in the whole spatial topology by the eccentricity index. This index is defined by the relation between a dimension where an object is disconnected and another dimension where the object is integrated. It is formally expressed as follows.

$$Ecc(\sigma_i) = \frac{\hat{q} - \check{q}}{\check{q} + 1} \qquad [3]$$

where $\hat{q}$ or top-q denotes the dimensionality of the simplex (σ_i); $\check{q}$ or bottom-q denotes the lowest q level where the simplex gets connected to any other simplex. The eccentricity values of the primary objects in the complex presented in figure 3 are Ecc (y_1) = 0.5, Ecc (y_2) = 0.5, and Ecc (y_3) = 1.

According to equation [3], a primary object (as a simplex) gains a high value of eccentricity if it is different from every other. Thus it differentiates other simplices in the sense of uniqueness, rather than in the sense of importance or function in the flows and transmission within a complex. In other words, the eccentricity cannot precisely characterize the status of different simplices. This can be illustrated with the complex shown in figure 3, where both simplices y_1 and y_2 have the same eccentricity value (*0.5*). However, it is intuitive enough that y_2 tends to be more important structurally than y_1. Referring to the complex in figure 3, without y_2, the complex would be broken into two pieces at the dimension two, while y_2 and y_3 are still kept together with the loss of y_1. The reason for this stems from the fact that the status of y_2, as a transitional simplex in the equivalence classes of dimensions of 0 and 1, is not taken into consideration in the eccentricity computation. In other words, the computation of eccentricity does not consider nearness relations within the equivalence class, but such relations are critical in determining the status of a simplex.

4. NEW MEASURES FOR CHARACTERIZING STRUCTURAL PROPERTIES OF INDIVIDUAL SIMPLICES

The inability of eccentricity to describe the status of individual simplices within a complex enables us to seek alternative measures for it. Centrality measures initially proposed in the field of social network by Freeman (1979) aim to support the quantitative description of a node status within a graph. It consists of three separate measures of centrality: connectivity degree, closeness and betweenness. In the context of this paper, the latter two measures are of particular interest. Closeness quantifies how close a node is to every other node by computing the shortest distances from every node to every other; betweenness measures to what extent a node is located in between the paths that connect pairs of nodes, and as such it reflects directly the intermediary location of a node along indirect relationships linking other nodes.

4.1 Closeness measure for as simplex

The closeness centrality in a graph context aims to assess the status of a node within a graph. It examines how a node is integrated or segregated with a graph or network. It is interesting that the measure has different names in different contexts. For instance, it is called status in graph theory (Buckley and Harary 1990); In space syntax (Hillier and Hanson 1984) it is called integration, computed from the reciprocal value of path length, a key concept used in small world study (Watts and Strogatz 1998). Based on the idea of closeness centrality, we suggest here a similar measure to quantify how each simplex is close to every other. We define the measure for a simplex within a complex at the different dimensional level.

$$C(\sigma_i)_q = \frac{n-1}{\sum_{k=1}^{n} d(\sigma_i, \sigma_k)} \qquad [4]$$

where d is the shortest distance from a given simplex (σ_i) to every other simplex, n is the total number of simplices within a complex. Note the measure is given at a dimensional level q. Or put differently, the computation of the closeness considers multidimensional chains of connectivity. A simplex gains a closeness value at the different dimensional level if it gets connected to other simplices.

Logically the closeness measure can be accumulated along the different dimensional levels, to show an overall closeness of a simplex to every other with at all dimensional levels. The accumulated closeness can be formulated as follows.

$$C(\sigma_i) = \sum_{bottomq} \frac{n-1}{\sum_{k=1}^{n} d(\sigma_i, \sigma_k)} \quad [5]$$

4.2 Betweeness Measure for a Simplex

On the basis of the idea of betweenness centrality, we suggest here a similar measure for describing the linkage status of a simplex within a complex. Unlike the betweenness centrality in graph theory where the analysis focuses on zero-dimensional links using a term in q-analysis, we consider the multidimensional nature of the linkage. That is, to what extent a simplex is located in between the paths that connect pairs of simplices in the simplical complex at different dimensional levels and thereby determine the betweenness centrality of a given simplex at all the dimensions. Accordingly, we define betweenness centrality for a simplex (σ_i) in a given dimension q as follows:

$$B(\sigma_i)_q = \sum_i \sum_j \frac{P_{ikj}}{P_{ij}} \quad [6]$$

where P_{ij} denotes the number of shortest paths between simplices σ_i and σ_j, and P_{ikj} is the number of shortest paths from σ_i to σ_j that pass through transitional simplex σ_k, i.e. the simplices σ_i and σ_j are q-connected by the transitional simplex σ_k; P_{ikj} is a binary value that indicates if a simplex σ_k serve as a transitional simplex for σ_i and σ_j - if yes, $P_{ikj} = 1$, otherwise $P_{ikj} = 0$. Hence, the transitive weight of each transitional simplex σ_k for a given pair $\sigma_i \sigma_j$ equals to the proportion $1/P_{ij}$ i.e. if there is only one minimum path between i and j then σ_k gains the value 1; if there is two minimum paths σ_k gains the value ½.. Accordingly, the betweenness centrality of a simplex σ_k at a dimension level q equals to the cumulative transitive weight values of σ_k for all the pairs at that dimension. Note: the betweenness value does not exist when $P_{ij}=0$.

Applying the computation of $B(\sigma_i)_q$ for all the dimensions where the simplex is connected to other simplices defines the total betweenness centrality of the simplex within a complex K:

$$B(\sigma_i) = \sum_{bottomq} \sum_i \sum_j \frac{P_{ikj}}{P_{ij}} \quad [7]$$

The total centrality measure is a cumulative value obtained by applying the computation of $B(\sigma_i)_q$ from bottom-q (the highest dimension at which a simplex gets connected to other simplices) down to dimension 0. As its countpart defined in graph theory, this measure indicates to what extent a simplex is important for the linkage within a complex. If a simplex with higher value of betweenness gets removed, the whole complex may get broken into pieces.

5. CONCLUSION

This paper introduced a model based on Q-analysis for exploring spatial relationships through common context objects. Through overlapping two map layers (a primary layer and a context layer), a simplicial complex can be set up to explore various structural properties both from local and global perspectives. Major contribution of this paper is two-fold. First, the concept of spatial topology is suggested to extend the existing various topological concepts and to concentrate on the network view of a geographic system for more advanced spatial analysis in terms of flow of information and substance. Second, two centrality-based measures are introduced for characterizing structural properties of individual objects within a geographic system. Unlike the counterpart measures in graph theory, our measures consider multidimensional chains of connectivity, thus it is more appropriate to analyze the flows of information and substances within a system.

The model introduced is compatible to existing GIS model and thus is directly computable using layer-based geographic information. The model to a great extent is complementary to many graph theoretic approaches, and it provides new insights in terms of structural discovery, evolution and dynamics of a geographic system. We plan to extend the study by applying the model to some practical applications.

REFERENCES

Atkin R. H. (1974), *Mathematical structure in human affairs*, Heinemann Educational Books: London.

Atkin R.H. (1977), *Combinatorial Connectivities in Social Systems*, Birkhauser: Basel.

Beaumont J. R. and Gatrell A. C. (1982), *An introduction to Q-analysis*, in CATMOG book series on concepts and techniques in modern geography, Geo Abstracts, Norwich.

Buckley F., and Harary F. (1990), *Distance in graphs*, Addision-Wesley Publishing Company: Reading, MA.

Dorogovtsev S.N. and Mendes J F F (2003), *Evolution of networks: from biological nets to the Internet and WWW*, Oxford University Press: Oxford.

Egenhofor M. (1991), Reasoning about binary topological relations, *Second Symposium on Large Spatial Databases*, Zurich, Switzerland, O. Gunther and H.-J. Schek (eds.), *Lecture Notes in Computer Science*, Vol. 525, Springer-Verlag, pp. 143-160.

Earl C. F. and Johnson J. H. (1981), Graph theory and Q-analysis, *Environment and Planning B: Planning and Design*, 8, pp. 367 – 391.

Freeman L. C. (1979), Centrality in Social Networks: Conceptual Clarification, *Social Networks*, 1, pp. 215 – 239.

Hillier B. and Hanson J. (1984), *The Social Logic of Space*, Cambridge University Press: Cambridge.

Jiang B. and Claramunt C. (2004), Topological Analysis of Urban Street Networks, *Environment and Planning B: Planning and Design*, Pion Ltd, Vol. 31, pp. 151-162.

Tobler W. R. (1970), A Computer Movie Simulating Urban Growth in Detroit Region, *Economic Geography*, 46, pp. 234-240.

Watts D. J. and Strogatz S. H. (1998), Collective Dynamics of 'Small-World' Networks, *Nature*, 393, 440 – 442.