

A graph-based algorithm to define urban topology from unstructured geospatial data

J.-P. de Almeida^{a,b,d*}, J.G. Morley^c and I.J. Dowman^d

^aGeomatic Engineering Unit, Department of Mathematics, Faculty of Science & Technology, University of Coimbra (UC), Coimbra, Portugal; ^bInstitute for Systems Engineering and Computers at Coimbra (INESCC), Coimbra, Portugal; ^cCentre for Geospatial Science, School of Geography, University of Nottingham (UN), Nottingham, UK; ^dDepartment of Civil, Environmental & Geomatic Engineering, University College London (UCL), London, UK

Abstract

Interpretation and analysis of urban topology are particularly challenging tasks given the complex spatial pattern of the urban elements, and hence their automation is especially needed. In terms of the urban scene meaning, the starting point in this study is unstructured geospatial data, *i.e.* no prior knowledge of the geospatial entities is assumed. The aim of translating these data into more meaningful homogeneous regions can be achieved by detecting geographic features within the initial random collection of geospatial objects, and then by grouping them according to their spatial arrangement. The techniques applied to achieve this are those of graph theory applied to urban topology analysis within GIS environment. The role of graph theory in such a task and pre-processing of raw geospatial data were object of another previously published paper; this manuscript focuses primarily on the description of the procedure: “containment-first search” (CFS) – based on the well known breadth-first search algorithm for graph traversal, CFS analyzes and interprets the geospatial object arrangements in terms of the extension of the standard notion of the topological relation of adjacency to that of containment, what is defined within this research as “polygon-ring containment”. LiDAR data were used as an example scenario for development purposes; in order to test the system proposed as a whole, further tests will entail the use of real world LiDAR datasets with different technical characteristics as well as photogrammetric datasets.

Keywords: urban topology; graph theory; scene analysis; GIS.

1. Introduction

1.1 Motivation

Interpretation and analysis of urban topology are particularly challenging tasks given the complex spatial pattern of the urban elements, and thus their automation is especially needed. In this context, research has revealed the importance of the concepts from the mathematical areas of both topology and graph theory for interpreting the spatial arrangement of spatial entities. Graph theory in particular has been used in different applications of a wide range of fields for that purpose, however not many graph-theoretic approaches to analyse entities within the urban environment are available in the literature, hence the need for further investigation. Some examples should be mentioned though such as, Bafna (2003), Barr and Barnsley (2004), Bunn *et al.* (2000), Krüger (1999), Nardinocchi *et al.* (2003), and Steel *et al.* (2003), amongst others.

In particular, very little work has been devoted to the interpretation of initially unstructured (in terms of the urban scene meaning) geospatial datasets. In most applications – if not all – developed so far for the interpretation and analysis of spatial phenomena within the urban context, the starting point is to some extent a meaningful geodataset in terms of the urban scene. Starting at a level further back, before meaningful urban scene information is obtained, the interpretation and analysis of geospatial phenomena are more challenging tasks and require further investigation. LiDAR data were used in this research as an example scenario for algorithms development. LiDAR data are usually combined with other sources of data for geographic information production. Albeit our research was not carried out from the perspective of the remote sensing field, and hence LiDAR analysis was not its main

interest, further investigation still appears pertinent on the processing of LiDAR data as a single data source for the retrieval of higher-level geographic information

Moreover, the lack or even the non-existence of graph analysis functionalities in common GIS packages suggests the development of an interactive tool for the visual representation and analysis of such a spatial data structure. In fact, structured inquiries and spatial analysis are certainly integral to GIS, but they fail to emphasize the power of the human eye in detecting patterns and the role of subjectivity in GIS (Schuurman, 2004; Dykes *et al.*, 2005). As fact of matter, in most automated and semi-automated methods the user is not able to visualise any interim results until the end of the whole process. It is believed that the capability of visualising the results of interim steps is relevant and constitutes an extra value.

1.2 Aim and objectives

Further to the needs for more investigation stated above, two research questions emerged for the purpose of this work:

- To investigate whether it is possible to work with initially unstructured geospatial data – and hence no prior knowledge of the spatial entities is assumed in this case – in order to produce higher-level geographical information.
- Although geometric information is inevitably implicit in the whole process, the development of a purely topological approach was sought. The question is: how far it is possible to go just by looking at topological relations between the geospatial objects?

Given our topological approach, purely based on spatial relations, it was known beforehand that it would be awkward to obtain as final output low-level detail geographical information. Thus, the ultimate goal of the research hereby presented was

to find “higher-level geographical features” by delineating coherent, homogeneous features within the initial arbitrary collection of geospatial data. A “higher-level geographical feature” is defined in the context of this research as a set of geographical features like buildings, *e.g.* urban blocks of buildings, which may include trees as well. The aim of retrieving structured information from initial unstructured geospatial data, translated into more meaningful homogeneous regions, can be achieved by finding inherent relevant spatial relations within the initial random collection of geospatial objects and by inferring facts from that (Anders *et al.*, 1999). In these circumstances, not only the attributes of an object are relevant but most importantly both its spatial location and its spatial interaction with other objects (Barr and Barnsley, 1997; Anders *et al.*, 1999; Kim and Muller, 2002; Bauer and Steinnocher, 2001; Nardinocchi *et al.*, 2003; Forlani *et al.*, 2006).

In the light of the research questions above, some objectives have been identified as follows:

- The consideration of the problem as a general task of finding higher-level geographical features in an arbitrary collection of lower level details.
- The investigation of the topological relations between objects in the context of the whole spatial scene rather than within their individual neighbourhood (defined in this research as “higher-order analysis”).
- In order to meet the aims stated above, the design and implementation of a higher-order topological analysis method, based on both the investigation of the spatial relations beyond the first level of adjacency, and on the extension of the standard notion of adjacency to that of containment.
- The conception and implementation of the graph-theoretic approach above in a GIS framework.

- The development of the algorithm should be carried out as much independent as possible from the test environment.
- Finally, the extension of the range of the methods used in order to visualise the resulting graphs of adjacencies and display the urban scene topology on virtual maps. The objective was to link the higher-order topological analysis method back to the original environment, *i.e.* the digital map of urban spatial objects.

2. Background

As stated in section 1.1, it is believed that the task of understanding topological relations between spatial objects can be accomplished by both: applying graph theory and carrying out graph analysis. The purpose of this section is to give the reader an overview of how this mathematical framework has been used so far to address a wide variety of problems, not only in the GIS world – eventually our area of interest – but also in other research fields. As general examples, three graph-based applications for spatial system analysis available from the literature are briefly revisited below.

2.1 *Space syntax*

The space syntax method provides an efficient experimental approach to the understanding of spatial configuration. It has provided since the 1980s important computational support for the development of spatial morphological studies, in particular for the analysis of urban systems (Jiang *et al.*, 2000). Space syntax also attempts to explain human behaviour and social activities from a spatial configuration point of view (Jiang *et al.*, 2000; Bafna 2003). Thus, space syntax can also be described as “a research program that investigates the relation between human societies and the space from the perspective of a general theory of the structure of inhabited space in all its diverse forms: buildings, settlements, cities, or even landscapes”. Further to the

“social logic of space”, Bafna (2003) also stated that the ultimate aim of space syntax research is “to develop strategies of description for configured, inhabited spaces (of buildings, settlements, or built complexes) in such a way that their underlying social logic can be enunciated”.

The primary object of analysis within space syntax research is the configured space. Having said that a certain space is configured, space syntax theory implicitly assumes that the continuous space can be turned into a connected set of discrete units (large-scale vs. small-scale spaces view) (Jiang *et al.*, 2000; Bafna, 2003). There are many advantages in doing so because “different labels can be applied to its individual parts; these parts can then be assigned to different groups, people, or activities; different rules of behaviour and conventions can be associated with different parts of the space; and individual parts of space can be recognized as carrying a specific symbolic or cultural charge” (Bafna, 2003). The computational space syntax model that integrates the small-scale perspective is based on a two-step approach: first, the representation of the large-scale space as a finite number of small-scale spaces; second, to link these individual small-scale spaces to form a connectivity graph. Indeed, from the computational point of view, space syntax is based on a graph-oriented representation of the geographical space that adequately captures all the spatial configuration aspects (Jiang *et al.*, 2000; Bafna, 2003).

Most space syntax studies deal primarily with issues related to urban patterns. The spatial decomposition of the urban, vs. the building, environment into small-scale components concentrates on free spaces. This is because free spaces provide a unique view and hence are fundamental for understanding the configuration of an urban system (Jiang and Claramunt, 1999, cited in Jiang *et al.*, 2000). The distinction between the free spaces and spatial obstacles is generated by the existence of boundaries between

streets/rooms and the built environment, *i.e.* both are interdependent as they share a common physical boundary

Besides urban design and architecture, the method has proved to be relevant in a wider range of studies, such as: crime analysis, and accordingly space syntax theory can be used as a component for actions related to the prevention of crime in urban planning and building design (Jones and Fanek 1997, cited in Jiang *et al.*, 2000); in pedestrian modelling (Hillier *et al.*, 1993, cited in Jiang *et al.*, 2000); traffic pollution control (Penn and Croxford, 1997, cited in Jiang *et al.*, 2000); and way-finding processes (Peponis *et al.*, 1990, cited in Jiang *et al.*, 2000).

2.2 Built-form connectivity

Built-form connectivity method attempts to explain the spatial organisation of an urban system on the basis of its distribution of buildings. It focuses on both the intrinsic spatial relations of buildings between one another and also on their external relations with other urban features, such as the road network (Barr and Barnsley, 2004).

Traditionally, urban system connectivity research has concentrated overall on an analysis of the spatial topological organisation of built forms (Barr and Barnsley, 2004). A key example is the work of Krüger (1979a, 1979b, 1980, 1981a, 1981b), which is briefly reviewed in this section.

In Krüger's work, built forms are defined as "planar edge-connected set representing external walls and partitions (party walls) of buildings on the ground and can be represented by a particular combination of three types of mathematical graph". The three graph types represent the connectivity relations: amongst the built forms, between the built forms and the surrounding external environment, and between the built forms and the channel network (Krüger, 1979a).

Krüger's starting point is based on both the notions of homomorphism and isomorphism in the relation between buildings and built forms. In his model, the author defined built forms as "quasi-mathematical models" and used a graph-theoretic representation in order to express how buildings are connected and packed over an area of land: buildings are represented by points called built forms; external walls and partitions (party walls between buildings) by lines. The connected sub-graphs consisting of built forms and partitions are called arrays of built forms (Krüger, 1979a). According to the author, this constitutes a simplified view of the built-form subsystem that, together with the road channel, gives rise to a graph that models an urban system.

In Krüger's urban system model, a structural tree consisting of different levels of disaggregation stratifies the graph components. The whole urban graph system is at the level of greatest aggregation. The first level of disaggregation is divided into two different elements: the built-form galaxy and the channel network. The former represents all built forms within places, the latter relates to all kinds of links and vertices that represent some physical form of transportation between places.

The built-form galaxy is subdivided into built-form constellations (i.e. sets of built forms surrounded by part of the channel network). Built-form constellations are subdivided into built-form arrays (i.e. sets of independent connected components within a constellation of built forms). Last, the built-form arrays consist of one or more connected built form units.

In turn, the channel network is subdivided by type (into road, rail and river networks), and each of these is then subdivided into blocks (sub-graphs with a cycle structure) and cul-de-sacs (sub-graphs without a cycle structure).

In terms of graph representation, three different representations were proposed by Krüger (1979a): graphs of type 1 – the points represent the centres of the interiors of

the polygonal planar edge-connected sets (built forms), and the lines represent the common side walls between edge-connected sets; graphs of type 2 – the points represent the centres of the interiors of the built forms and the boundary neighbourhood centres adjacent to each external wall, and the lines represent the external walls; graphs of type 3 – the points represent the centres of the interiors of the built forms and the nearest points on the channel network, and the lines represent the access route to each built form from the channel network.

2.3 The eXtended Relational Attributed Graph (XRAG)

According to its authors, XRAG is defined as “a graph-based, structural pattern recognition system that might be used to infer second-order thematic information (broad categories of land-use) from first-order thematic information (normally, from high spatial resolution remotely-sensed images)” (Barr and Barnsley, 1997; Barnsley and Barr, 1998; Steel *et al.*, 2003). The input data is a land-cover map (*i.e.* a map of forms) to which is applied the XRAG to store, analyse and interpret the morphological properties of, and spatial relations between, discrete land-cover parcels; as a result, a land-use map (*i.e.* a map of functions) is obtained.

In conceptual terms, the system uses a graph (in the discrete mathematics sense of the world) to represent the structural relations between the land-cover parcels. Each one of these parcels is represented by a vertex in the graph, while the spatial relation between two such parcels is represented by an edge connecting them.

The system is able to represent and analyse the morphological properties of the individual parcels (such as areas, perimeter, and compactness), various structural relations between them (namely, adjacency and containment) and spatial relations between them (such as distance or cardinal directions) (Barr and Barnsley, 1997; Barnsley and Barr, 1998; Bauer and Steinnocher, 2002; Steel *et al.*, 2003; Barnsley,

2003). One way of representing those relations between land-cover parcels is the use of graphical pattern representation techniques (Schalkoff, 1989, 1992, cited in Barr and Barnsley, 1997). This sort of representation is based on graph theory for the representation of spatial relations.

3. Pre-processing of geospatial data

In order to provide the reader with some background, a general overview of the main preliminary steps dealing with preparation of raw geospatial data is given in this section. In fact, this was object of a previously published paper that the reader may refer to for further details (reference to be inserted by here).

To start with, LiDAR data were used as an example scenario. It is an unstructured dataset with no patterns pre-defined and meaningless in terms of urban scene; it has 3m point spacing and contains both ground points and object points reflected from trees, buildings and other small objects above ground level. The dataset refers to a 1470x1530m² urban area, comprising a total of 169,819 laser points, located in Kew, southwest London (UK).

In order to start structuring information and make it more explicit, some topological information was brought in by establishing a triangulated irregular network (TIN) through the given data set. Given that the TIN constitutes the maximal planar description of the initial point set's internal structure (Kirkpatrick and Radke, 1985), it expresses proximities and neighbourhoods between the input LiDAR points.

A classification was then applied to the TIN facets based on their gradient. The usage of a 45° gradient threshold was considered and an equal interval binary classification was performed. The underlying idea in considering two classes of gradients, *i.e.* “flat” and “steep” TIN facets, is that most of the important urban features (*e.g.* man-made structures and vegetation) are likely to be enclosed by rings of steep

facets. Several polygonal regions were then generated by aggregating TIN facets in accordance to the binary classification mentioned above, *i.e.* facets of the same class meeting on edges were merged; facets of the same class meeting at a node were preserved. A first attempt with a 60° gradient value showed that buildings clearly standing on their own were not well defined, for, in a few instances, the enclosing ring of steep facets was open. Albeit the great majority of TIN facets had a gradient less than 30°, another attempt using this threshold did not produce much better results for such a low gradient threshold inserts much more noise.

4. The “containment-first search” analytical method

4.1 The underpinning concept

Once the map of steep & flat regions is obtained, this constitutes the real input of the analytical analysis method (*vd.* Figure 1). Within such a map of polygonal regions, and for the purposes of this work, when two polygons share at least one arc, the spatial relation is called “adjacency”; if the two polygons happen to meet at a node, the spatial relation is distinguished from the previous one and is called “touching”.

INSERT FIGURE 1 BY HERE

Figure 1. Map of gradient regions used in our case as input of the urban topology analysis method (a 45° gradient threshold was used in this particular case).

As stated in the objectives above (*vd.* section 1.2), the graph-theoretic approach was designed based on the investigation of the topological relations between geospatial objects in the context of the whole geospatial scene, rather than within their own neighbourhood. In particular, the detection of containment relations across the whole map of polygonal regions is in fact of extreme relevance for where containment occurs within the “useful external border” (UEB) – the outer flat enclosing polygon, which

simulates the ground level, from where sequences of adjacencies/containments make most sense in terms of the urban scene – there is a high likelihood of an urban feature being present. Thus, as explained below, the analytical analysis method traverses graphs by looking for sequential relations of containment amongst sequences of adjacency, what we defined by “containment-first search” (CFS).

4.2 The network of connectivity across polygonal regions

Polygon and associate arc attributes were accessed in order to retrieve gradient-region adjacencies (ESRI, 1995; Rigaux *et al.*, 2002; ESRI, 2005). A network of connectivity throughout the map of flat & steep polygonal regions was then built up by applying graph theory, which resulted in a graph of adjacencies (*vd.* Figure 2): each node in the graph represents a gradient region; graph edges link up nodes corresponding to adjacent regions.

INSERT FIGURE 2 BY HERE

Figure 2. Establishing a network of connectivity between gradient polygonal regions.

The adjacencies graph can be processed either through depth-first (DFS) or breadth-first search (BFS) algorithms. Given the different ways each algorithm operates in traversing a graph, results obtained with BFS proved to be more meaningful in terms of the urban scene: the BFS tree branches are connected components of the original graph, and represent the shortest path between the root and their leaf (Sedgewick, 2002); it seemed they could be related to potential features on terrain. The design of the analytical method for urban topology analysis hereby proposed was then based upon BFS.

4.3 Polygon-ring containments

However, preliminary experiments revealed that CFS could not be developed simply based on BFS; it had to be extended in order to be able to detect the spatial relation of containment in a broader sense. This improvement enabled the derivation of particular cases of containment not explicit in the graph of adjacencies, *e.g.* when a ring of steep polygons meeting at nodes contains a single flat polygon - the so-called “polygon-ring containment” relation (the yellow circle in insert in Figure 2 above highlights such an instance). For the purpose, the spatial relation of touching between steep polygons is crucial and must be taken into consideration.

The information required for the retrieval of the steep-polygon touchings is not explicit in the database though. As for the retrieval of the polygon adjacencies, a complex task, entailing the manipulation and analysis of information spread over different INFO files, had to be carried out. After obtaining the spatial relations of touching between steep polygons, the respective graph of touchings was constructed.

4.4 Implementation

From the technological implementation point of view, the proposed method was coded in C foreseeing the advantages and potentialities of pointer structures in C for graph analysis (Kelley and Pohl, 1990).

The analysis method can be interpreted from the algorithmic perspective as follows. Considering the UEB flat polygon as the starting point of the search process, the original graph of adjacencies is traversed. When visiting the adjacent steep vertices of the root, CFS algorithm takes the first vertex appearing in the root’s linked list and, starting from this one, traverses the graph of steep-polygon touchings. Because the graph of touchings is a disconnected graph (recall that not all the steep polygons are in direct contact to one another by a shared node), the traversal process covers only the subgraph that the given steep vertex belongs to. While traversing this particular

subgraph, CFS tags all the steep vertices visited as belonging to the same connected unit. This process continues until the first level of adjacency of the graph of adjacencies is exhausted. When CFS comes across a root's adjacent vertex already tagged as belonging to a particular containment unit, this is skipped and the corresponding polygon remains intact as constituent part of the containment unit already identified.

To illustrate the concept implemented, let us take a simulated simple scene pictured in Figure 3. Let us suppose that steep polygons $3, \dots, 11$ (in dark green) are constituent parts of the steep-polygon rings enclosing flat polygons 12 and 13 (*vd.* Figure 3a); in other words, there is a sub-graph of the touchings graph that consists of vertices 3 to 11 .

INSERT FIGURE 3 BY HERE

Figure 3. The containment-first search process: a) before polygon-ring containments are detected; b) after polygon-ring containments are detected.

When vertex 2 is visited in the adjacencies graph, the algorithm takes the vertex at the top of 2 's adjacency list, vertex 10 , and the graph of steep-polygon touchings is traversed starting from 10 ; all the steep vertices belonging to the same sub-graph as that of 10 are tagged accordingly, indicating a potential containment unit. When vertex 10 is exhausted in the graph traversal, CFS moves on to visit vertex 9 ; this is now skipped since it was previously tagged as belonging to a containment unit already identified. And so on so forth until vertex 3 is visited, and the containment unit is complete. Visually, the translation of the facts above is accomplished by assigning the same colour to all steep (hashed pattern) and flat (solid colour) polygons within the same containment unit (*vd.* Figure 3b).

More formally, both detection of polygon-ring containments and CFS procedure are respectively presented below in terms of pseudo-code.

Polygon_ring_containment

Algorithm 1

Graph of steep-polygon touchings is traversed starting from vertex v (based on depth-first search):

Vertex v is visited;

```

For vertex  $t$  adjacent to  $v$  Do
  If  $t$  hasn't been visited yet Then
    Tag vertex  $t$  as belonging to the same polygon-ring as that of  $v$ ;
    Polygon_ring_containment( $t$ ); (take vertex  $t$  and proceed recursively)
  End if
End for
End

```

Containment_first_search

Algorithm 2

Graph of adjacencies is traversed (based on breadth-first search):

Start counting the nbr of containment units;

```

While Queue of vertices to be visited is not empty Do
  Visit vertex  $k$ ;
  Start counting the valence1 of  $k$ ;
  For each adjacent vertex  $v$  to  $k$  Do
    If  $k \neq$  root Then
      Count the valence of  $k$ ;
    End if
    If  $v$  hasn't been visited yet Then
      If  $k$  is the root Then
         $v$ 's parent = root;
         $v$ 's level of containment = 1;
        If  $v$  hasn't been tagged yet as belonging to a certain containment unit
          Then a new polygon-ring containment enclosing a new containment unit
            was detected:
            Count nbr of containment units;
             $v$ 's containment unit = nbr of units counted so far;
            Polygon_Ring_containment( $v$ ); //all steep polygons  $t$  touching  $v$ 
              are tagged as belonging to the same polygon-ring as
              that of  $v$  -> Algorithm 1
          End if
        Else
           $v$ 's parent =  $k$ ;
           $v$ 's containment unit =  $k$ 's containment unit;
           $v$ 's level of containment = level of containment of  $v$ 's parent+1;
          // $v$ 's level of adjacency represents a level of containment within the
            current containment unit
        End if
        Put  $v$  in the Queue to be visited thereafter;
      End if
    End for
  Root's "valence" = nr of containment units (recall that steep-polygon islands are discarded);
End while
Visual representation of graph analysis results;
End

```

A *rationalization* procedure was also implemented to be run after CFS. One of the aims of this algorithm is to detect and "clear out" single steep polygons enclosed by flat

¹ Within graph theory, "valence" of a node is the number of connecting arcs at that node, or in other words, the number of incident arcs into that node.

polygons. In most cases, steep-polygon islands are related to noise and do really not have any particular meaning in terms of urban scene; they can hence be ignored.

4.5 Preliminary results

As stated above, the analysis results of CFS are translated into different coloured patterns for visualisation purposes. The respective slope regions in the original map are dynamically mapped accordingly.

For the purpose above, a colouring function was implemented to set up a range of colours, and also an associate gradient of tones of each different colour. Different colours are meant to represent disjoint containment units. Solid colours correspond to flat polygons, and coloured hashed patterns correspond to steep polygons. The higher the level of adjacency/containment of polygons within a given unit, the darker the tone of the colour assigned to it. The idea of assigning a different colour to each containment unit detected during the analysis process is to help one in visually identifying the potential existence of a spatial feature. When applying the algorithm to geographical data, spatial features may correspond to disjoint buildings, sets of buildings, individual trees/bushes, or sets of buildings and trees/bushes.

The implemented method was then applied to the test geospatial dataset (Kew, London). Results obtained are depicted below in Figure 4 – just a detail is shown for the area highlighted by the yellow rectangle in Figure 1. It should be noted that owing to edge effects the steep polygons (and enclosed flat polygons) mapped in red in the northeast part of the case study area were wrongly tagged, and hence assigned a wrong colour. Those polygons constitute in fact upper levels of adjacency within the containment unit mapped in green, and should have been mapped with a darker tone of that colour.

INSERT FIGURE 4 BY HERE

Figure 4. Experiment applied to test urban data of Kew (southwest London, UK): the different units detected, potentially relating to high-level urban features, are mapped in different colours. (The letters on the map are polygon labels in insert).

In visual terms, the potential existence of urban geospatial features is mapped with a different colour; they represent each different containment unit detected by the algorithm. In the particular case illustrated in Figure 4, the arcs of the steep-polygon islands, within for instance the UEB mapped in white, are still mapped just to show the extent of the generalisation procedure. Recall that all steep-polygon islands are “merged” with their enclosing flat polygons during the rationalization process.

To summarise, a diagram is depicted in Figure 5 below to give an overview of the methodology described throughout section 4 for the analysis of urban spatial topology. It illustrates in particular where CFS – including polygon-ring containment detection – sits within the whole process.

INSERT FIGURE 5 BY HERE

Figure 5. An overview of the methodology proposed for the analysis of urban spatial topology.

5. Summary and conclusions

The study described in this paper has taken a step in the direction of the use of spatial topology for the analysis and understanding of complex urban scenes.

A graph-theoretic approach was proposed to analyse entities within the urban scene; its design and implementation was accomplished within GIS environment. More precisely, graph theory was applied in the interpretation of the urban spatial topology based on the extension of the standard notion of the spatial relation of adjacency to that of containment: the containment-first search algorithm. Although geometric information is to some extent implicit in the designed methodology, this study sought to take a

purely topological approach and investigate how far it would be possible to go just by looking at topological relations between spatial objects. In addition, this research attempted to work with initially unstructured geospatial data – no prior knowledge of the urban scene was thus assumed – in order to study whether it would be possible to produce at least higher-level geographical information.

As an example scenario, LiDAR data (referring to an urban area in Kew, London – UK) were considered for development purposes. The analysis of the local spatial topology was undertaken and conclusions were drawn in terms of the assertions made when designing the algorithms. Results obtained demonstrated that the algorithms do indeed make the urban spatial topology more explicit. In particular, results support the assumption that each BFS tree's branch does relate to a single containment unit within the initial map of gradient regions; moreover, sequences of containment relations do relate to higher-level urban scene features.

To conclude, a summary of the whole work developed so far is provided below: the diagram depicted in Figure 6 takes an overview of the main steps undertaken from initially unstructured data until higher-level geographical features are derived.

INSERT FIGURE 6 BY HERE

Figure 6. Illustrative summary of the urban topology analysis process, from start (unstructured data) to finish (detected containment units).

In terms of final recommendations, it must be noted that it would be beneficial to replicate the approach hereby outlined on more LiDAR datasets with different technical characteristics. Further experimentation on other sources of unstructured data, such as clouds of photogrammetric points, would also be interesting in order to test the versatility of the algorithms.

Furthermore, concepts drawn in the research hereby described and the algorithms implemented should serve as the basis for automatic analysis of spatial datasets, such as: analysis of image data; analysis of settlement structures; automation of land-use mapping for urban areas. Furthermore, albeit the aim of this work was not LiDAR analysis, preliminary experiments carried out with some real-world LiDAR data revealed that the methodology proposed has also promised as a tool that may well be extended towards the automatic segmentation of raw LiDAR data.

References

- ANDERS, K.-H., SESTER, M. and FRITSCH, D., 1999. Analysis of settlement structures by graph-based clustering. *Semantische Modellierung SMATI*, 99, 41-49. Munich (Germany).
- BAFNA, S., 2003. Space Syntax – A brief introduction to its logic and analytical techniques. *Environment and behaviour*, 35(1), 17-29.
- BARNSLEY, M.J., 2003. Mapping Land-Use in Urban Areas [online]. Available from <http://stress.swan.ac.uk/~mbarnsle/research/urban.htm> [Accessed 24 January 2003].
- BARNSLEY, M.J. and BARR, S.L., 1998. 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. and 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.
- BARR, S.L. and BARNSLEY, M.J., 2004. Characterising the structural form of an urban system using built-form connectivity model concepts. In KELLY, R.E.J., DRAKE, N.A. and BARR, S.L. eds. *Spatial Modelling of the Terrestrial Environment*. Chichester (England, UK): John Wiley, 201-226.

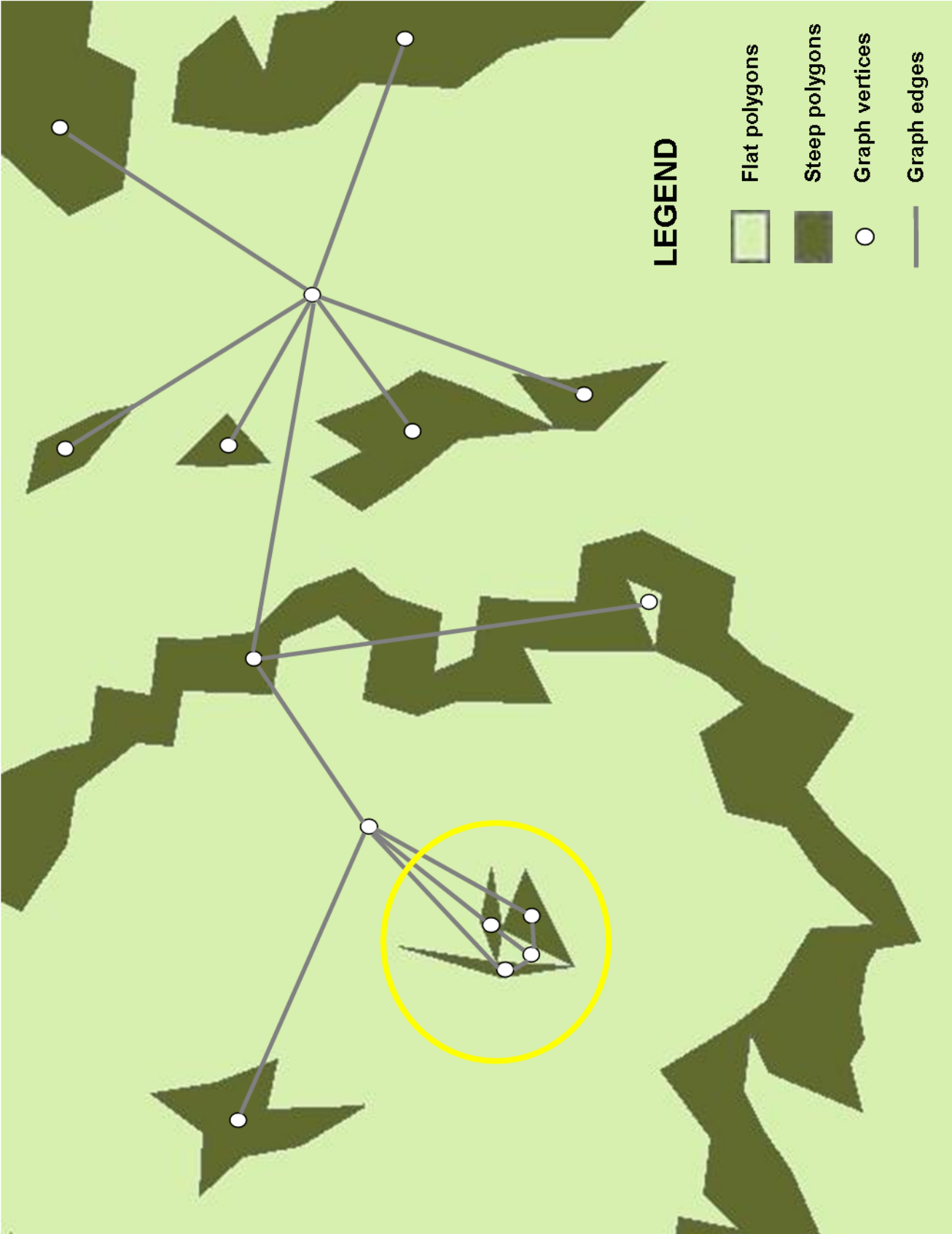
- BAUER, T. and STEINNOCHER K., 2001. Per-parcel land use classification in urban areas applying a rule-based technique. *GeoBIT/GIS* 6, 24-27.
- BUNN, A.G., URBAN, D.L. and KEITT, T.H., 2000. Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management*, 59, 265-275. Academic Press (USA)
- DYKES, J., MACEACHREN, A. M. and KRAAK, M.-J., 2005. Exploring geovisualization. In DYKES, J., MACEACHREN, A. M., KRAAK, M.-J., eds. *Exploring Geovisualization*, 1, 3-19.
- ESRI, 1995. *Understanding GIS, The ARC/INFO Method*, Lesson2. ESRI, Inc. Cambridge (England, UK).
- ESRI, 2005. *ArcGIS Topology* [online]. Available from <http://www.esri.com/products/support/ArcGIS%20UK%20EXT/Topology.asp> [Accessed 10 June 2005].
- FORLANI, G., NARDINOCCHI, C., SCAIONI, M. and ZINGARETTI, P., 2006. Complete classification of raw LiDAR data and 3D reconstruction of buildings. *Pattern Analysis and Applications*, 8(4), 357-374.
- HILLIER, B., PENN, A., HANSON, J., GRAJEWSKI, T. and XU, J., 1993. Natural movement: configuration and attraction in urban pedestrian movement. *Environment and Planning B*, 20, 29-66.
- JIANG, B., CLARAMUNT, C., LINK, L. and KLARQVIST, B., 2000. An integration of space syntax into GIS modelling urban spaces. *International Journal of Applied Earth Observation & Geoinformation*, 2(3/4), 161-171.
- KELLEY A. and POHL I., 1990. *A Book on C - Programming in C*, 2nd ed. Redwood City (CA, USA): Benjamin Cummings.
- KIRKPATRICK, D.G. and RADKE, J.D., 1985. A framework for computational morphology. In TOUSSAINT, G.T., ed. *Computational Geometry*. Amsterdam (The Netherlands): North-Holland, 217-248.
- KIM, J.R. and MULLER, J.-P., 2002. 3D Reconstruction from very high resolution satellite stereo and its application to object identification. *Symposium on Geospatial Theory, Processing and Applications*. Ottawa (Canada).

- KRÜGER, M., 1979a. An approach to built-form connectivity at an urban scale: system description and its representation. *Environment & Planning B*, 6, 67-88.
- KRÜGER, M., 1979b. An approach to built-form connectivity at an urban scale: variations of connectivity and adjacency measures amongst zones and other related topics. *Environment & Planning B*, 6, 305-320.
- KRÜGER, M., 1980. An approach to built-form connectivity at an urban scale: relations between built-form connectivity, adjacency measures, and urban spatial structure. *Environment & Planning B*, 7, 163-194.
- KRÜGER, M., 1981a. An approach to built-form connectivity at an urban scale: modelling the distribution of partitions and built-form arrays. *Environment & Planning B*, 8, 41-56.
- KRÜGER, M., 1981b. An approach to built-form connectivity at an urban scale: modelling the disaggregation of built forms types. *Environment & Planning B*, 8, 57-72.
- NARDINOCCHI, C., FORLANI, G. and ZINGARETTI, P., 2003. Classification and filtering of laser data. *ISPRS Workshop on 3D reconstruction from airborne laser scanner and InSAR data*. Dresden (Germany).
- PENN, A. and CROXFORD, B., 1997. Effects on street grid configuration on kerbside concentrations of vehicular emissions. In HILLIER B., ed. *Proceedings, First International Symposium on Space Syntax*, 16-18 April. University College London (England, UK).
- PEPONIS, J., ZIMRING, C. and CHOI, Y., 1990. Finding the building in way-finding. *Environment and behavior*, 22, 555-590.
- RIGAUX, J., SCHOLL, M. and VOISARD, A., 2002. *Spatial Databases with Applications to GIS*. San Francisco (CA, USA): Morgan Kaufmann, Chp 8.
- SCHALKOFF, R. J., 1989. *Digital Image Processing and Computer Vision*. New York (USA): John Wiley.
- SCHALKOFF, R. J., 1992. *Pattern Recognition: Statistical, Structural and Neural Approaches*. New York (USA): John Wiley.

SEDGEWICK, R., 2002. *Algorithms in C – Part 5*, 3rd ed. Reading (MA, USA): Addison-Wesley, 17, 18.

SCHUURMAN, N., 2004. *GIS – A short introduction*. Bodmin (England, UK): Blackwell Publishing Ltd, 4.



STEEL, A.M., BARNSLEY, M.J. and BARR, S.L., 2003. Inferring urban land use through analysis of the spatial composition of buildings identified in LiDAR and multispectral image data. In MESEV, T.V. ed. *Remotely-Sensed Cities*. London (England, UK): Taylor & Francis.

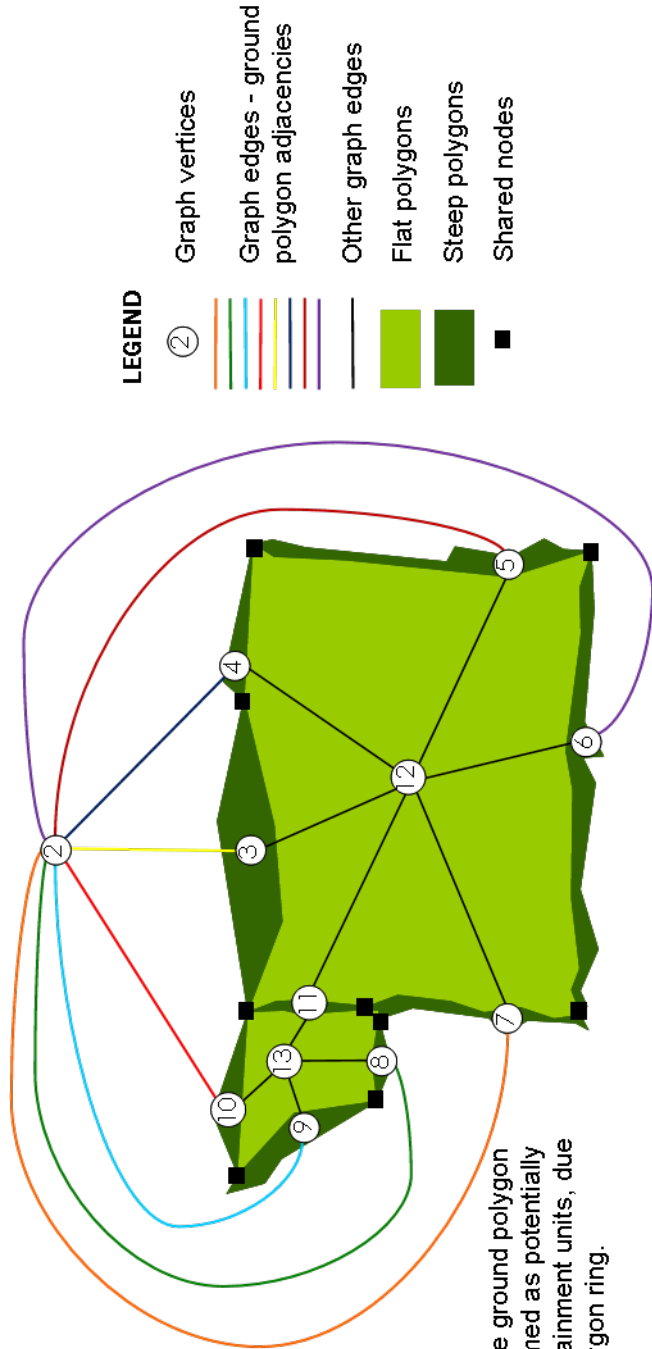




LEGEND

Gradients

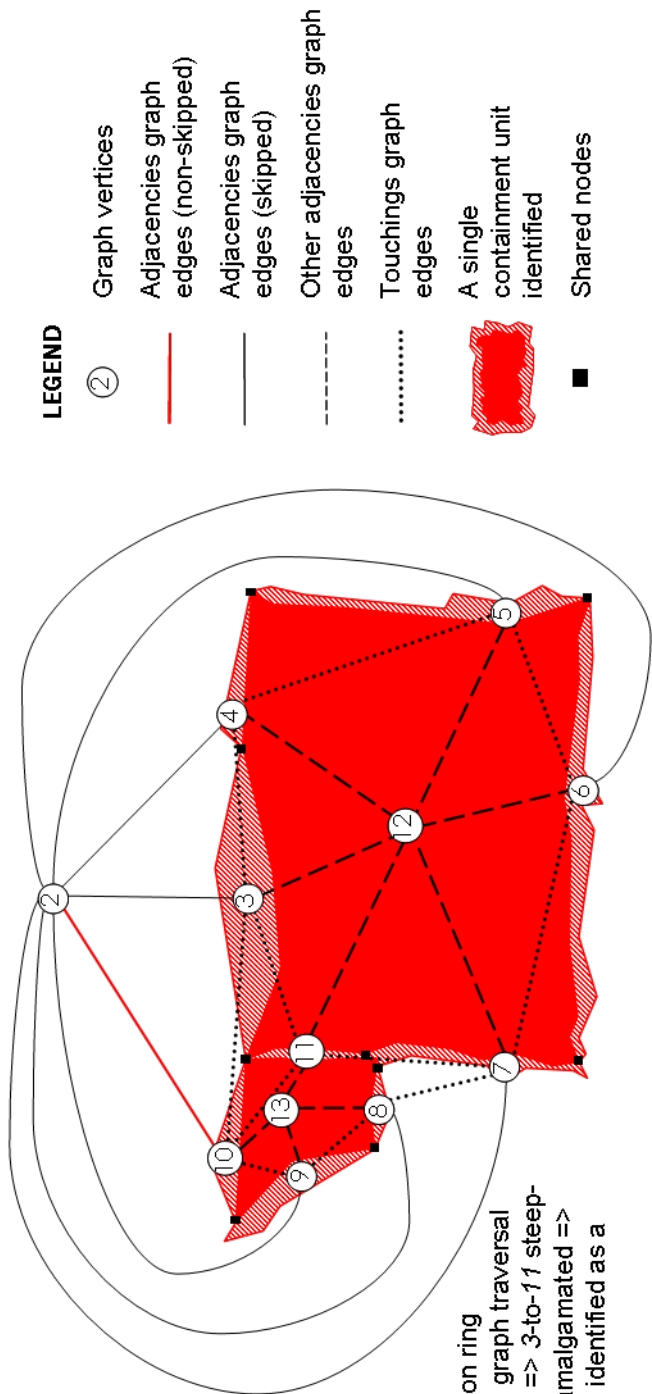
	0 - 45
	45 - 90



LEGEND

- ② Graph vertices
- Graph edges - ground polygon adjacencies
- Other graph edges
- Flat polygons
- Steep polygons
- Shared nodes

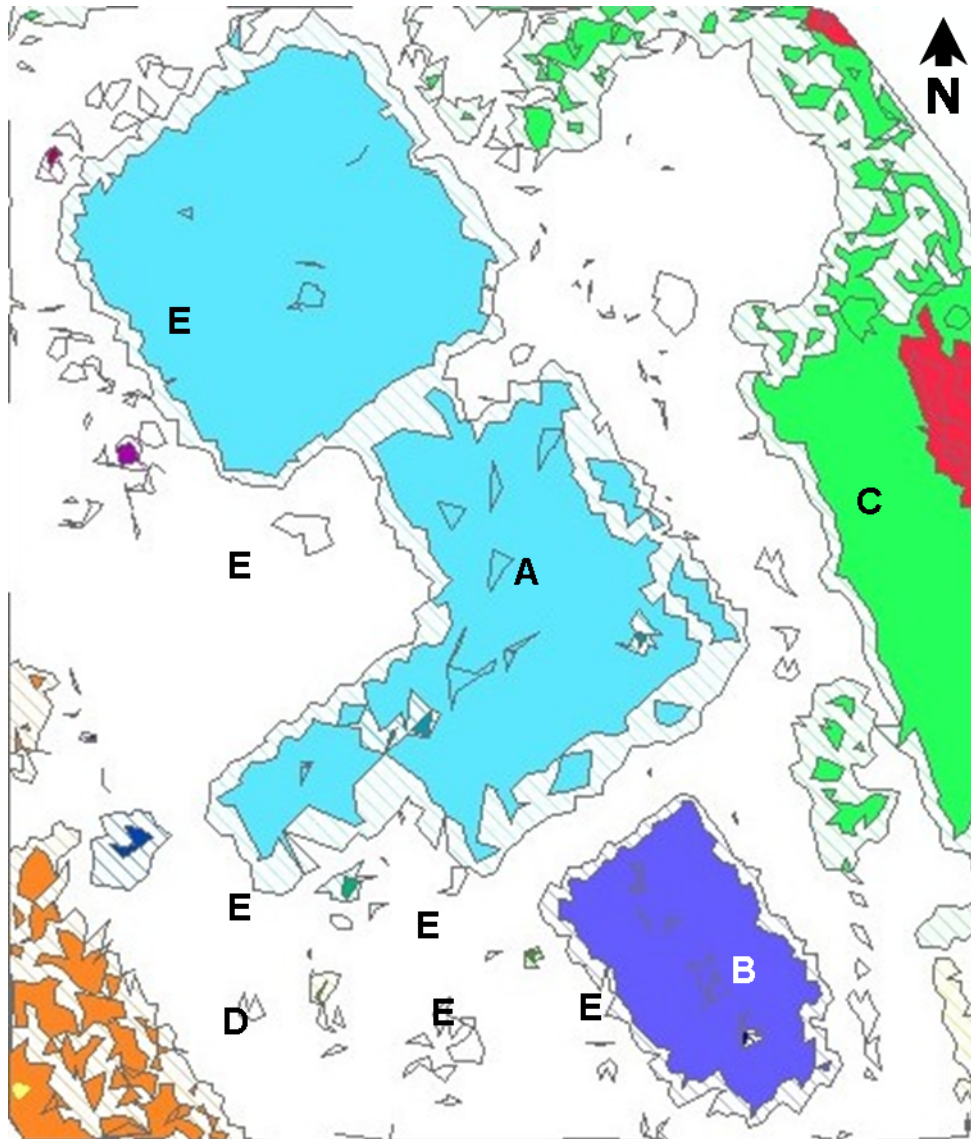
a) Adjacent polygons to the ground polygon (vertex 2) wrongly assumed as potentially enclosing separate containment units, due to the 3-to-10 steep-polygon ring.



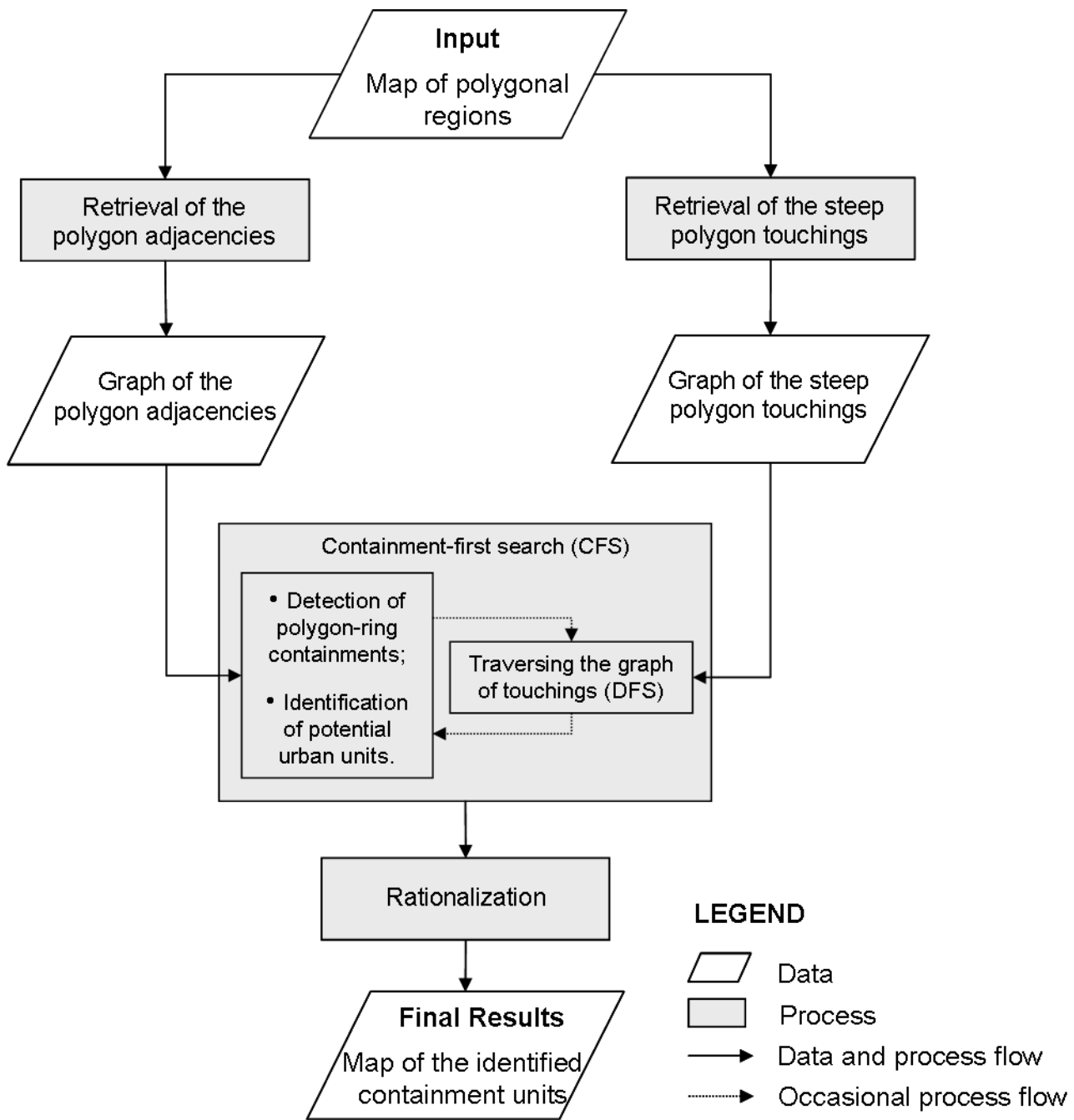
LEGEND




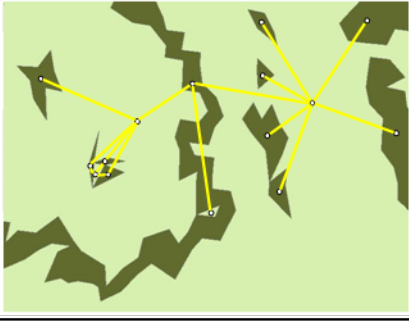
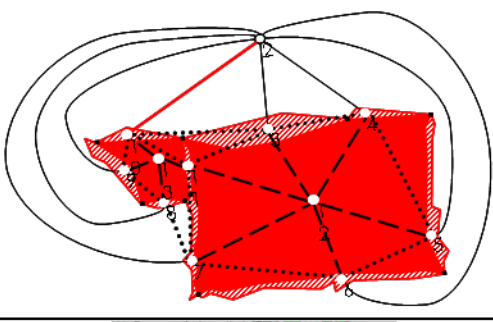
- ② Graph vertices
- Adjacencies graph edges (non-skipped)
- Adjacencies graph edges (skipped)
- Other adjacencies graph edges
- Touchings graph edges
- A single containment unit identified
- Shared nodes

b) Detection of steep-polygon ring containments: touchings graph traversal (starting with vertex 10) => 3-to-11 steep-polygon rings correctly amalgamated => structure in red correctly identified as a single containment unit.



- LEGEND**
- A – a building (National Archives)
 - B – a building
 - C – an urban unit (building & some vegetation)
 - D – an urban unit (residential buildings & vegetation)
 - E – some isolated trees/bushes



<p style="text-align: center;">Start</p> <p style="text-align: center;">Unstructured data: a cloud of 3D points</p>	
<p style="text-align: center;">Data preparation (i)</p> <p style="text-align: center;">Topology is brought in: construction of a TIN for the internal description of the original point pattern</p>	
<p style="text-align: center;">Data preparation (ii)</p> <p style="text-align: center;">Binary classification of the TIN facets based on their gradient</p>	
<p style="text-align: center;">Establishment of a network of connectivity between polygonal regions</p> <p style="text-align: center;">Construction of polygon adjacencies graph & steep-polygon touchings graph</p>	
<p style="text-align: center;">Containment-first search</p> <p style="text-align: center;">Analysis of the adjacencies graph structure & Detection of the polygon-ring containments by simultaneously analysing the touchings graph</p>	
<p style="text-align: center;">Finish</p> <p style="text-align: center;">Derivation of higher level spatial structures: containment units & Dynamic display of the containment units identified in the original map of polygonal regions</p>	