Spatial Evolution of Urban Fabrics Extracted from Spatiotemporal Topographic Database using Symbolic Dynamic Time Warping

Pierre Gançarski
*University of Strasbourg*, pierre.gancarski@unistra.fr

Anne Puissant
*University of Strasbourg*, anne.puissant@unistra.fr

François Petitjean
*Monash University*, francois.petitjean@monash.edu

Follow this and additional works at: [https://scholarsarchive.byu.edu/iemssconference](https://scholarsarchive.byu.edu/iemssconference)

Part of the Civil Engineering Commons, Data Storage Systems Commons, Environmental Engineering Commons, Hydraulic Engineering Commons, and the Other Civil and Environmental Engineering Commons


[https://scholarsarchive.byu.edu/iemssconference/2016/Stream-C/12](https://scholarsarchive.byu.edu/iemssconference/2016/Stream-C/12)

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Spatial Evolution of Urban Fabrics Extracted from Spatiotemporal Topographic Database using Symbolic Dynamic Time Warping

Pierre Gançarski(1), Anne Puissant(1), François Petitjean(2)

(1) University of Strasbourg - France
{pierre.gancarski, anne.puissant}@unistra.fr
(2) Monash University, Melbourne - Australia
francois.petitjean@monash.edu

Abstract: This article introduces a new methodology dedicated to extracting the evolution of urban blocks from spatiotemporal topographic databases where an urban block is defined as the smallest area that is surrounded by communication network. To achieve this analysis, we apply the ascendant hierarchical clustering to sequences of urban block states (i.e.; sequences of class labels to which the block belongs to on each date). The principal originality of this approach is to use a measure based on DTW (Dynamic Time Warping) which is able to apprehend temporal behaviours (mainly time lags in dates corresponding to a change of state) and which takes into account the semantic proximity between the different kinds of urban blocks. Several experiments have been carried out on areas in the city of Strasbourg (France). First results are relevant and highlight realistic urban dynamics.

Keywords: urban dynamics; historical vector topographic database; time series; clustering; dynamic time warping; edit-distance.

1 INTRODUCTION

The management of cities is at the heart of the major contemporary concerns. Monitoring urban sprawl and its consequences remains a major challenge for urban planning and management. Morphological analysis of urban fabric is an important precondition to better understand the urbanisation process of cities. The first assumption is that several urban evolutions correspond to a reconstruction and densification process. The densification is never random and depends on the type of urban blocks, where such object is defined as the smallest area that is surrounded by communication network (roads, railways …) and on the location of the blocks with respect to the city centre and the rural area. As one cannot assume that reference data (e.g., ground truth, training samples) are going to be available, methods that are able to process in an unsupervised way are needed.

This paper focuses on the aspect concerning the clustering of urban block evolutions. The objective is to explore and quantify evolution of urban blocks by taking into account period’s analysis.

Clustering is a well-known family of methods, which are able to process in an unsupervised way. This family is made up of well-known algorithms like the Kmeans algorithm or the Ascendant Hierarchical Clustering algorithm. Clustering aims at grouping similar data. In our case, it would consist of grouping urban block that have a similar evolution. The vast majority of the most used clustering algorithms are based (at least) on a similarity measure, which makes it possible to know how close two data points are. When handling evolution of urban blocks, i.e., sequences of states of evolution, defining a similarity measure is much harder than when handling numerical values in a vector space without natural ordering. There are actually four main problems: firstly, each state of evolution is symbolic, which means that the similarity between different states of evolution have to be defined by the expert. Secondly, the temporal structure of the data has to be taken into account: an urban block evolving from wasteland to individual housing, is very different than an urban block evolving from individual housing to wasteland. Thirdly, only few databases are available over the last fifty years. Therefore, only methods that deal with irregular temporal sampling will be able to fully exploit all the available databases. Finally, a single urban block (corresponding to a geographic area) can evolve into two blocks and conversely. For instance, in (Mizutani 2011), the author combines the land use information and the shapes of regions to characterize the transition patterns from one time point to the next (Mizutani 2009). He proposes six region states according to these aspects: stability, substitution (change of land use), division without land use change, and division with land use change, expansion...
and conversion. The goal of this work is the dual of ours. The author aims to study the land use transition processes themselves: statistics of transitions associated with each kind of region states for instance. In the same way, (Kulik 2011) defines 25 transition types to analyse vegetation time series.

Section 2 describes the databases used in this work and introduces how merges and splits of urban blocks can construct sequences of evolutionary states. In Section 3 the similarity measure DTW used to clusterise the states urban time series and the clustering algorithm are introduced. Experiments carried out on several databases are presented in Section 4. We conclude this paper and propose further work in Section 5.

2 MATERIAL
2.1 Geographic mono-date objects

The geographic objects contained within different mono-date topographic databases (produced by the French National Institute) on the urban area of Strasbourg are used as benchmark data. Historic spatiotemporal databases are then created for four test areas with the following six different dates due to the availability of historic data (aerial photographs and topographic maps): 1956, 1966, 1976, 1989, 2002 (benchmark data) and 2008. These four zones are chosen because they have been subject to a variety of typical urbanisation processes. For example, the area presented in Figure 1 corresponds to a typical expansion phenomenon called peri-urbanisation phenomenon. It is also subject to natural constraints (rise of ground water) in the centre. This area is localised in the North of Strasbourg.

![Figure 1: Area with a typical expansion phenomenon](image)

The proposed typology of urban blocks (Tab. 1) is compatible with existing land-cover/use nomenclature (e.g., Corine Land Cover). It is adapted to map the territory at the scale of 1:10,000. It does not take into account the functionality of urban fabrics because this information is impossible to create from historic maps.

<table>
<thead>
<tr>
<th>Tab 1: Urban block class names</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C1</strong></td>
</tr>
<tr>
<td><strong>C2</strong></td>
</tr>
<tr>
<td><strong>C3</strong></td>
</tr>
<tr>
<td><strong>C4</strong></td>
</tr>
<tr>
<td><strong>C5</strong></td>
</tr>
<tr>
<td><strong>C6</strong></td>
</tr>
<tr>
<td><strong>C7</strong></td>
</tr>
<tr>
<td><strong>C8</strong></td>
</tr>
<tr>
<td><strong>C9</strong></td>
</tr>
<tr>
<td><strong>C10</strong></td>
</tr>
<tr>
<td><strong>C11</strong></td>
</tr>
</tbody>
</table>
The urban blocks are built automatically by using linear objects (communication and hydrographic networks) available in the topographic database. For each temporal database, urban blocks have been labelled automatically by a supervised classification process (Lesbegueries et al., 2011, Puissant et al., 2011), which was only based on the morphological attributes and the spatial distribution of buildings and the open space within a block. Omission and commission errors in all the classifications have been manually corrected by an expert in urban planning.

2.2 Construction of sequences

Each state of evolutions of urban blocks corresponds to the class to which the block belongs on the considered date. However, several changes can appear over an area, like the building (resp. the removal) of a road. By assumption, roads form the boundaries of urban blocks (as urban blocks are surrounded by communication ways) and therefore building (resp. the removal) of a road can split (resp. merge) urban blocks. In fact, we are face to a graph of urban block evolutions as shown in Fig 2.

![Graph of urban blocks evolution](image)

**Figure 2:** Graph of urban blocks evolution (5 dates). \((b_i, C_k)^t\) denotes that the label associated to urban block \(b_i\) is \(C_k\) at date \(t\). Note that at date 2, \(b_1\) is split into two parts, while at date 5, two blocks are merged.

Each path in the oriented graph builds a sequence (e.g., \([b_1, C_1]^1, (b_2, C_3)^2, (b_3, C_3)^3, (b_3, C_3)^4, (b_7, C_3)^5\)). In the sequel of the paper, we will simplify notation into \(C_1 > C_3 > C_3 > C_3 > C_3 > C_1\) as only the class labels are taken into account in the classification process.

One can notice that a block can belong to several sequences and, as it will be shown in Section 4, it can belong to several kinds of evolution since the sequences (where it appears) can be classified into different clusters. Nevertheless, we assume that this is not a problem because we aim to highlight classes of evolution and not individual block behaviour.

3 CLUSTERING OF URBAN BLOCK EVOLUTIONS

3.1 Urban block evolution similarity

When studying the evolution of urban areas over time, the core of the process generally consists of comparing data in order to estimate (dis)similarity. The distance provides an estimation of this similarity. When the data is temporal, the choice of the distance is crucial since it completely defines the way of tackling the temporality of the data. This dissimilarity measure must exploit the temporal distortions and compare shifted or distorted evolution profiles. It must be able to deal with sequences whose time sampling is irregular. In this paper, we use the Dynamic Time Warping (DTW) based on the Edit-distance (Levenshtein 1965) and introduced in (Sakoe 1971, Sakoe 1978). It finds the optimal alignment (or coupling) between two sequences of numerical values, and captures flexible similarities by aligning the coordinates inside both sequences. The cost of the optimal alignment of two sequences \(A = \langle a_1, a_2, \ldots, a_N \rangle\) and \(B = \langle b_1, b_2, \ldots, b_P \rangle\) (where \(N\) can be different than \(P\)) can be recursively computed by:

\[
D(A_i, B_j) = \text{sim}(a_i, b_j) + \min \left\{ \begin{array}{l}
D(A_{i-1}, B_{j-1}) \\
D(A_i, B_{j-1}) \\
D(A_{i-1}, B_j)
\end{array} \right.
\]

with \(A_i = \langle a_1, a_2, \ldots, a_i \rangle, i \leq N\) and \(A_i = \langle a_1, a_2, \ldots, a_j \rangle, j \leq P\).

Thus to compare two time series we need to compare the items in pairs using a given similarity \(\text{sim}\). In the case of numeric values, the Euclidian distance is generally used. In our case, since values are labels of class, we must define a similarity between classes to which each block can belong. In this work, we propose to use a dissimilarity matrix (Tab. 2) manually built by an expert in urban planning and based on his/her knowledge of the composition and the spatial organization of buildings in a block.
and on the possibility to confuse them. Thus, it can be considered as representative of the “semantic distances” between the classes. A low value means that classes are close in terms of composition and that the probability to confuse them is very high. A high value means that classes have no similarity.

### Tab 2: Urban blocks dissimilarity matrix

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C4</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C7</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C8</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C9</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C10</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C11</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

The use of DTW as a similarity measure in clustering technique (Petitjean 2011) is relatively classical now. For instance, (Petitjean 2012) shows benefits to using such similarity measure in satellite image time series analysis. More recently, (Izakian et al. 2015) proposes some alternatives to fuzzy clustering methods to time series analysis based on the DTW distance and on the fuzzy C-means algorithm.

### 3.2 Ascendant Hierarchical Clustering

The Ascendant Hierarchical Clustering (AHC) algorithm makes it possible to cluster a set of data (Ward 1963). This algorithm performs in four steps:

1. Begin with K groups corresponding each to one sequence from the urban block evolutions.
2. Compute the distances between every group pair in a triangular similarity matrix.
3. Merge the two closest groups i.e., the groups which have the lowest dissimilarity value and modify the triangular similarity matrix.
4. If there are more groups than desired (generally, one group), go to step 3.

The distance between two clusters that contain single sequences is defined as the distance between these sequences. To compute the distances between two clusters that contain one or more sequences, we have chosen to use the Average Linkage criterion, which usually provides satisfactory results. By construction, we obtain a binary tree of clusters whose root is the cluster that contains all the sequences and whose a leaf contains only one sequence. Figure 3 shows the top of the cluster hierarchy obtained on our dataset with the ratio of objects in each cluster.

![Figure 3: Extract of the top urban block evolutions hierarchical clustering.](image-url)

### 1. RESULTS AND DISCUSSIONS

To assess the relevance of our method for the analysis (i.e., the identification) of the principal blocks evolutions, several experiments have been carried out on an area in the city of Strasbourg (France) corresponding to the one presented in Fig 1 (for the sake of readability, only three dates out five are...
Figure 4: Cluster E.1.1.1 mainly corresponding to evolution of low density mixed urban surfaces C5 and C7 into high density areas surface C6 (E.1.1.1.1) and into discontinuous urban fabric with individual houses C3 (E.1.1.2.2).

Figure 5: Cluster E.1.1.2 mainly corresponding to evolution of no or few building areas C9 into low density mixed areas C7 (E.1.1.2.1) and into discontinuous urban fabric with individual houses C3 (E.1.1.2.2).

Fig 6 (resp. Fig 7) illustrates the maps corresponding to the cluster E.1.1.1 (resp. E.1.1.2) with, in dark orange (resp. blue), the sub-cluster E.1.1.1.1 (resp. E.1.1.2.1) and in light orange the cluster E.1.1.2 (resp. E.1.2.2). In these figures, a block of a particular date is coloured if it belongs to at least one sequence of the considered cluster. For instance, all non-white blocks in Figure 6(a-c) correspond to urban blocks belonging to at least one sequence in cluster E.1.1.1. Finally, the hatched texture highlights blocks that belong to both subclusters E.1.1.1.1 and E.1.1.1.2 (resp. to E.1.1.2.1 and E.1.1.2.2).
Several distinctive evolutions can be extracted from the clustering of urban blocks evolutions. Thus, cluster E.1.1 seems to correspond to blocks with a densification of buildings and cluster E.1.2 to blocks with almost no evolution.

1. In Figure 6 the evolution corresponds to the transformation of "low density mixing/urban fabric and area" (C5 and C7). With the dark orange colour, we can observe the evolution of such areas into "high density mixed areas" (C6) while with the light orange colour, we can observe their evolution into "discontinuous urban with individual houses" (C3).

2. In Figure 7 the evolution corresponds to the transformation of areas with no or few buildings (C9). With the dark blue colour, we can observe the transformation of such areas into "low density of mixed areas" (C7) while in light blue, we can observe their transformation into "Discontinuous urban with individual houses individual houses area" (C3). Note that the difference to the previous evolution in light orange is the initial state of the blocks (C5 or C7 vs C9). The hatched dark and light orange block on the 1976 map belongs to the two clusters because, between 1976 and 1989 it has been split into seven blocks: the sub-block in the East has moved to C9 while the others have moved to C3.
3. In Figure 8 we can see blocks with low evolution, mainly roads and non-buildable areas shaded with green. The big block on the left appears in green in 1986 because it is the "father" of the small block corresponding to the roads on the left (2002 and 2008). The hatched dark and light blue block in the South has been split between 1976 and 1989 into a block belonging to C2 and a block belonging to C5. (The same for a block in the East and another in the North-West, both of which have been split between 1989 and 2002). The big block in the middle of the map (corresponding to the area subjected to the rise of the ground water) has been merged with its left neighbour between 2002 and 2008.

4. In these figures, one can see a very small urban block (localized at the far north, in the middle of the area). This block belongs to C4 in all dates. This is the only block with this behaviour: it is the only element of E.2.

According to these observations, we assume that two-thirds of the E.1 cluster corresponds to blocks, which have become denser between 1956 and 2008 and one-third of the cluster to blocks with low (or no) evolution. Whilst the E.2 cluster corresponds to the (very small) urban block (localized in the middle of the area corresponding to the area submitted to the rise of the ground water), which is only classed in C4 for each date (no change).

5. CONCLUSION

This article has presented a new methodology dedicated to extracting the evolution of urban blocks from spatiotemporal topographic databases. The principal originality of this approach is to use DTW distance measure which is able to apprehend temporal behaviours (mainly time lags in dates corresponding to a change of state) and which takes into account the semantic proximity between the different kinds of urban blocks. To validate this approach, we have applied it to the ascendant hierarchical clustering of sequences of block states (i.e., class labels). The class labels associated to the blocks on each date have been pre-calculated by applying a supervised algorithm to the database corresponding to the specific date. The results of the experiment have been studied by an expert and seem to correspond to the reality. This validates the relevance of the proposed methodology. Nevertheless, some additional experiments should be conducted to precisely quantify and identify the evolution patterns of one or more periods.

This work opens up several perspectives and different research directions. From a methodological point of view, we plan to study more formally (1) the definition of the blocks and of the sequences and (2) the quality of blocks and sequences built in order to evaluate their influences on the results.

Furthermore, it could be relevant to integrate an approach that enables the user to build the similarity matrix. Indeed, by asking the user for different constraint examples between the data (e.g., must-link or cannot-link constraints), semi-supervised clustering approaches could be used to learn/estimate the different values of the matrix.

From an applicative point of view, this methodology could be used for supervised classification (using the K-Nearest Neighbour algorithm for example). Although to define examples seems a difficult and time-consuming task that would require better theoretical definitions of the types of evolution.

ACKNOWLEDGMENTS

The authors would like to thank the French National Research Agency (ANR) for having supported the GeOpensim (ANR-08-RAN52-Geopensim, 2007-2011) and COCLICO projects (ANR-12-MN-001-COCLICO, 2012–2016).

REFERENCES

Izakian H., Pedrycz W., Jamal I., 2015, Fuzzy clustering of time series data using dynamic time warping distance. Engineering Applications of Artificial Intelligence, 39, 235–244.


