Supporting Data Analytics for Smart Cities: An Overview of Data Models and Topology

Patrick E. Bradley(✉)

Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
bradley@kit.edu

Abstract. An overview of data models suitable for smart cities is given. CityGML and G-maps implicitly model the underlying combinatorial structure, whereas topological databases make this structure explicit. This combinatorial structure is the basis for topological queries, and topological consistency of such data models allows for correct answers to topological queries. A precise definition of topological consistency in the two-dimensional case is given and an application to data models is discussed.

1 Introduction

Electronic government, planning systems, and citizen participation require data models suitable for performing such tasks. Geographic Information Systems (GIS) for smart cities are used for assessing transportation and mobility, urban risk management and GIS-assisted urban planning, including noise-mapping and solar energy issues [17].

In order to be useful for more than visualisation purposes, data models must be topologically consistent, meaning that the underlying combinatorial and geometric models must be compatible. The reason is that it is usually more efficient to use the combinatorial model for topological queries. The precise definition of topological consistency varies considerably in the literature. E.g. in [9], 3D-meshes via one-dimensional finite elements are considered. Their consistency rule is that line segments intersect only in boundary points. With this, they achieve self-correction of inconsistent meshes. In [3], topological-geometric consistency means that the interiors of 1-cells resp. 2-cells may not intersect. However, no statement about cells of different dimensions is made. The authors of [18] use only the geometric model for determining topological relations, and then make corrections of violations against topological integrity constraints. According to [1], a surface is consistent if and only if it is a valid 2D-manifold. In [8], the topological consistency of simplifications of line configurations is treated without defining topological consistency. The authors of [14] discuss arbitrary configurations of points, line segments and polygons in the plane, as well as their extrusions to 3D. Their consistency rules are that line segments may intersect only in their boundary points, and the interior of a polygon may not intersect any other object. Nothing is said about the intersection of a point with another

A. Gammerman et al. (Eds.): SLDS 2015, LNAI 9047, pp. 406–413, 2015.
DOI: 10.1007/978-3-319-17091-6_35

object, or how the boundaries of objects intersect. For [12], a surface in $\mathbb{R}^3$ is consistent if it is a so-called *2.8D map*. Gröger & Plümer extend this consistency rule by not allowing the interiors of points, lines, areas and solids to intersect [10,11]. An efficiently verifiable list of local consistency rules then requires a tesselation of $\mathbb{R}^3$. This excludes e.g. free or exterior walls, i.e. polygons which border at most one solid.

In the following section, we will discuss some data models for smart cities, namely CityGML, G-maps and topological databases. The next section gives some examples of queries involving topology in smart cities. This is followed by a section in which our point of view on topological consistency is made into a precise definition, and some consequences are discussed.

2 Data Models for Smart Cities

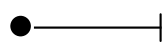
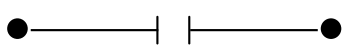
CityGML is an XML-based format for virtual 3D city models based on the Geographic Markup Language (GML). It is an international standard issued by the Open Geospatial Consortium (OGC) [13]. In essence, it models a configuration of points, line segments, polygons with holes, and solids together with a semantic structure by representing the thematic properties, taxonomies and aggregations. Solids and line segments are given by their boundary representations, and polygons with holes by their cycles of vertices under the tag `gml:LinearRing`. The combinatorial structure can be extracted from the xlink-topology together with the coordinate lists.

A *generalized map* (or short: G-map) is a topological model designed for representing subdivided objects in any dimension, and is an implicit boundary representation model. The definition of a G-map is given in [15] as

Definition 1. *An n-G-map is an $(n+2)$-tuple $G = (D, \alpha_0, \ldots, \alpha_n)$ such that*

- *D is a finite set of points called* darts
- *$\alpha_0, \ldots, \alpha_n$ are involutions on D, i.e. $\alpha_i \circ \alpha_i$ is the identity map on D*
- *$\alpha_i \circ \alpha_j$ is an involution if $i + 2 \leq j$ with $i, j \in \{0, \ldots, n\}$*

The combinatorial structure is given by i-cells which in turn are defined as G-maps whose underlying darts are an orbit of the group generated by the involutions other than α_i. G-maps can depicted in the following way:

- A dart is depicted as 
- An orbit under the group $\langle \alpha_0 \rangle$ generated by the involution α_0 is depicted as 
- An orbit under the group $\langle \alpha_i \rangle$ generated by α_i with $i > 0$ is depicted by connecting darts with i strokes. E.g. for $i = 1$: 

An example of a G-map with two 2-cells, five 1-cells and four 0-cells is given in Figure 1 (left). The 2-cells are the orbits of $\langle \alpha_0, \alpha_1 \rangle$ which are the upper and lower triangles of Figure 1 (right); the 1-cells are the orbits of $\langle \alpha_0, \alpha_2 \rangle$ which are

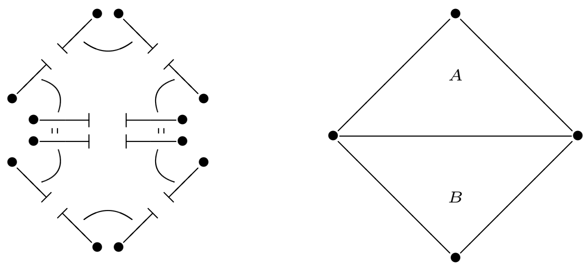


Fig. 1. Left: A G-map with two 2-cells, five 1-cells, and four 0-cells. Right: The cell complex associated with the G-map

the sides; and the 0-cells are the orbits of $\langle\alpha_1, \alpha_2\rangle$ which are the vertices of the figure. Notice that involution α_2 leaves some darts fixed, i.e. has fixed points.

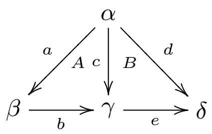
One problem with G-maps is their verbosity, especially in higher dimensions [6]. This means that it takes very many darts to describe a cell complex with relatively few cells. The reason is that a dart corresponds uniquely to a *cell tuple* which is a maximal sequence of cells $(c_n, \ldots, c_0)$ with c_i a face of c_{i+1}. The number of cell tuples is in general exponential in the number of cells for n large. The review article [7] contains some examples of urban models generated with G-maps.

A *topological database* [5,16] is mathematically given by a set of objects together with a binary relation on this set. As a relational database, it can be realised by two tables: one X for the objects, and one R for the relation. It is well known that a binary relation defines a topology on a set, and every finite topological space has a relation which defines the topology [2]. The table R can then be interpreted as the relation "bounded by" from which connectivity queries can be made.

It follows that in a topological database, the combinatorial structure is explicitly modelled. Additional orientation information yields an algebraic structure which turns R into a (partial) matrix, and under the condition $R^2 = 0$ this is a relational form of a chain complex with boundary operator R, important for some topological computations (e.g. Betti numbers). Figure 2 shows a cell complex with oriented cells and its relational boundary operator as a partial matrix. The orientation of the areas A, B is fixed here to be counter clockwise.

3 Topological Queries in Smart Cities

Once the combinatorial model underlying the data model has been established, topological queries can be made. A large class of topological queries are path queries: *Is there a path* $A \to B$*?* under some constraints. These constraints can be of geometric or of semantic nature. For example, the objects passed along a path must have a certain minimum size (geometric constraint) or the path must be inside a building (semantic constraint). Combinations of different types of constraints are also possible.



	A	B	a	b	c	d	e	α	β	γ	δ
A											
B											
a	+1										
b	+1										
c	−1	+1									
d		−1									
e		+1									
α			−1		−1	−1					
β			+1	−1							
γ				+1	+1		−1				
δ						+1	+1				

Fig. 2. Left: A cell complex with oriented cells. Right: Its relational boundary operator as a partial matrix

A particular type of path queries makes use of the *adjacency graph.* On a 2D-manifold, this is the *area adjacency graph*: the nodes are the polygons, and the edges are given by polygons adjacent along a common line segment. In 3D, it is the *volume adjacency graph.* Here, the nodes are the solids, and two solids are connected by an edge iff they share a common polygon at their boundaries. The advantage of the adjacency graph is that it does not use the complete dataset. However, it is quite common to have available only 2-dimensional data for city models. Then, in order to compute the volume adjacency graph of a complex building with many rooms, it is first necessary to fill all the shells surrounded by polygons with solids. This is a homology computation: the shells are the *second homology.* Homology computations are in essence algebraic computations, for which a chain complex derived from the data becomes useful. The number of shells is thus the *second Betti number.* The *first homology* is given by the loops, (window-like) openings, "inner courts", "tunnels" and "passages" of buildings. The *first Betti number* is their count. Consequently, the topological database enhanced with orientation information (i.e. the relational chain complex) has everything needed for homology queries.

4 Topological Consistency

As we have seen, topological queries exploit the combinatorial (or algebraic) model underlying the data model. This approach prefers the use of topological databases or relational chain complexes over data models which do not explicitly code the combinatorial structure. From CityGML or G-maps, a topological database can be extracted. In order for topological queries to yield correct answers, the geometric and combinatorial models must be compatible. This form of *topological consistency* means that both models must realise the same cell complex structure of the data. In other words, the data comprising of points, line segments and polygons must form a valid cell complex geometrically, and this is the combinatorial model underlying the data model. In Geographic Information

Systems, polygons may have holes. Such polygons are in general not cells, and we call these *quasi-cells*. A *quasi-cell complex* is obtained by sewing quasi-cells along their boundaries iteratively into the one-skeleton of the previously obtained quasi-cell complex. A one-dimensional quasi-cell complex is a cell complex. This generalises our notion of topological consistency:

Definition 2. *A finite set consisting of points, line segments, polygons with holes, together with the sides of each element is* topologically consistent *if its elements are closures of (quasi-)cells of a quasi-cellulation of the union of all elements in* $\mathbb{R}^3$.

A *quasi-cellulation* of a space is the quasi-cell complex obtained from a partition of the space into (quasi-)cells. The *combinatorial quasi-cell complex* associated with a configuration as in Definition 2 is the quasi-cell complex obtained by sewing each element along its boundary with each side as it occurs along the boundary. Such a configuration is topologically consistent if and only if the quasi-cellulation from Definition 2 coincides with the combinatorial quasi-cell complex associated with this configuration. Checking topological consistency amounts to computing pairwise intersections of elements. Such an intersection must then be a disjoint union of boundary elements plus possibly the interior of an element. Figure 3 shows an example of a consistent configuration. It consists

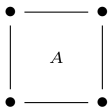


Fig. 3. A consistent configuration consisting of a square A, four line segments and four points

of a square, and the line segments and points at its boundary. These cells form a natural cellulation of the square, and coincide with the combinatorial cell complex associated with this configuration.

In [4], we prove the following intersection criterion:

Theorem 1. *A configuration as in Definition 2 is topologically consistent if and only if the intersection of any two distinct maximal elements* A *and* B *is the (possibly empty) disjoint union of boundary elements of* A *and* B.

An immediate consequence of Theorem 1 is that if an urban model is built up of polygons (with or without holes), topological consistency can be checked by computing the intersections of pairs of polygons. Line segments and points need not be intersected in this case. Figures 4–6 represent possible topological inconsistencies. Notice that in Fig. 5, the boundary vertices of polygon A are depicted as $\circ$ and $\odot$, whereas those of polygon B are $\bullet$ and $\odot$.

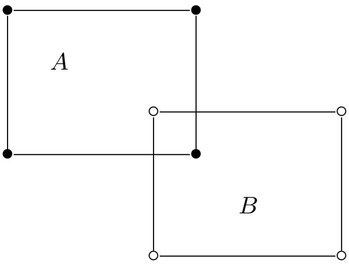


Fig. 4. A topologically inconsistent configuration where $A \cap B$ is not a union of objects

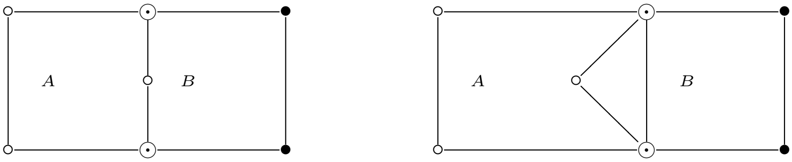


Fig. 5. Left: A topologically inconsistent configuration where $A \cap B$ is a non-disjoint union of objects. Right: The combinatorial cell complex associated with the configuration to the left contains a loop-hole.

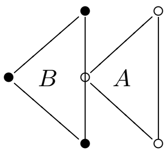


Fig. 6. A topologically inconsistent configuration where $A \cap B$ is a point which is not a boundary element of both A and B

5 Conclusions

We have discussed data models for smart cities: CityGML, G-maps and topological databases. The first two indirectly model the combinatorial structure underlying the data, whereas topological databases explicitly model it, and can be enhanced with orientation information on the individual building blocks. This combinatorial model can then be exploited to retrieve answers to topological queries which are correct if the model is topologically consistent, i.e. the underlying combinatorial and geometric models are compatible. Topological consistency can be checked through computing intersections of maximal building blocks which are polygons with holes or line segments in 3D. It is well known that this can be done in polynomial time. Applications of topological consistency are also beyond purely topological queries: for example, our results can lead to more efficient GIS-assisted urban planning systems, transportation and noise management, or improved management and evaluation of Smart Cities and public Open

Linked Data repositories, exploiting the advantages of topological consistency in data. Examples of possible applications could include airport terminal buildings, or commercial application of flying robots, e.g. the logistics and delivery drones that are used in some warehouses. A validator for CityGML data is work in progress. Future research will include the extension of the notion of topological consistency to higher dimensions and to the case where objects may be contained in higher-dimensional objects, like e.g. points or line segments inside a polygon. This is relevant for applications in Geographic Information Systems: for example a town in a region should be consistently represented as a point inside a polygon in certain levels of detail.

Acknowledgments. The author would like to thank Fionn Murtagh for suggesting to write about topological consistency in the context of smart cities, and also him and the anonymous referees for suggestions leading to a better exposition of this paper.

References

1. Akleman, E., Chen, J.: Guaranteeing the 2-manifold property for meshes with doubly linked face list. Int. J. Shape Model. **5**, 159–177 (1999)
2. Alexandroff, P.: Diskrete Räume. Matematiećeskij Sbornik **44**(2), 501–519 (1937)
3. Boudriault, G.: Topology in the TIGER file. In: Proceedings of the International Symposium on Computer-Assisted Cartography, pp. 258–263 (1987)
4. Bradley, P.: Topological consistency of polygon complexes in CityGML. Work in progress
5. Bradley, P., Paul, N.: Using the relational model to capture topological information of spaces. The Computer Journal **53**, 69–89 (2010)
6. Bradley, P., Paul, N.: Comparing G-maps with other topological data structures. Geoinformatica **18**(3), 595–620 (2014)
7. Breunig, M., Zlatanova, S.: 3D geo-database research: Retrospective and future directions. Computers & Geosciences **37**, 791–803 (2011)
8. Corcoran, P., Mooney, P., Bertolotto, M.: Line simplification in the presence of non-planar topological relationships. In: Gensel, J., et al. (eds.) Bridging the Geographic Information Sciences. Lecture Notes in Geoinformation and Cartography, pp. 25–42. Springer (2012)
9. Gattass, M., Paulino, G., Gortaire, J.: Geometrical and topological consistency in interactive graphical preprocessors of three-dimensional framed structures. Computers & Structures **46**(1), 99–124 (1993)
10. Gröger, G., Plümer, L.: How to achieve consistency for 3D city models. Geoinformatica **15**, 137–165 (2011)
11. Gröger, G., Plümer, L.: Provably correct and complete transaction rules for updating 3D city models. Geoinformatica **16**, 131–164 (2012)
12. Gröger, G., Plümer, L.: Transaction rules for updating surfaces in 3D GIS. ISPRS Journal of Photogrammetry and Remote Sensing **69**, 134–145 (2012)
13. Kolbe, T.: Representing and exchanging 3D city models with cityGML. In: Lee, J., Zlatanova, S. (eds.) Proc. 3rd Int. Workshop on 3D Geo-Information. Lecture Notes in Geoinformation & Cartography (2009)
14. Ledoux, H., Meijers, M.: Topologically consistent 3D city models obtained by extrusion. IJGIS **25**(4), 557–574 (2011)

15. Lienhardt, P.: N-dimensional generalized combinatorial maps and cellular quasi-manifolds. International Journal on Computational Geometry and Applications **4**(3), 275–324 (1994)
16. Paul, N., Bradley, P.: Eine Erweiterung des Relationalen Modells zur Repräsentation räumlichen Wissens. Datenbank-Spektrum, Online First (August (2014)
17. Tao, W.: Interdisciplinary urban GIS for smart cities: advancements and opportunities. Geo-spatial Information Science **16**(1), 25–34 (2013)
18. Ubeda, T., Egenhofer, M.: Topological error correcting in GIS. In: Scholl, Michel O., Voisard, Agnès (eds.) SSD 1997. LNCS, vol. 1262, pp. 281–297. Springer, Heidelberg (1997)