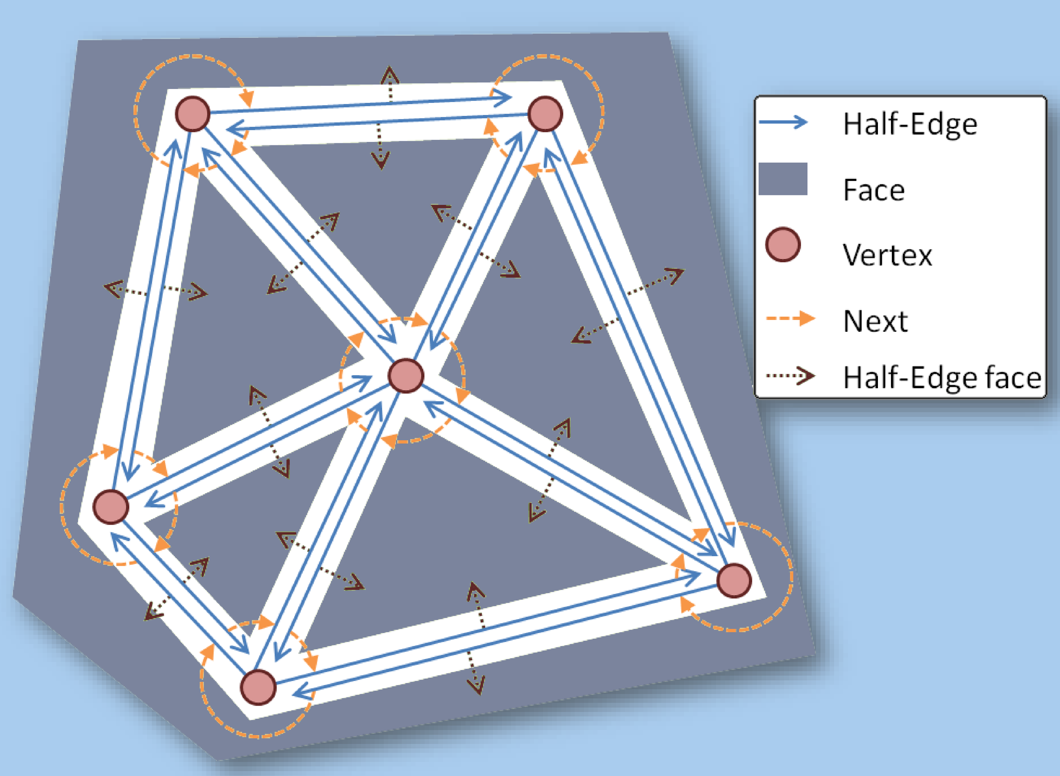
High Abstraction Level Topological Data Structures

Máté Cserép
Department of Cartography and Geoinformatics
Eötvös Loránd University

Introduction

Topology is a base concept of vector data models used for spatial reasoning and visualization. Both play a vital role not only in *Geographic Information Systems* (GIS), but also in computer graphics applications and more widely in mathematics.
The most expansively used spatial data formats – like the Shapefile, the Well-Known Text (WKT) or the Relational Data Format (RDF) – do not contain any topological information, rather they treat geospatial objects merely as shapes. Therefore the determination of spatial relations and overlays between vector objects and object layers are a common, although a complex and time-consuming task [1], while data imported from such formats are abound with joint errors on the boundaries and superfluous vertex and edge data.

Figure 2. The model of the half-edge topological data structure.



High level model

My research aims to specify a topological data structure over the half-edge model which is capable to answer higher level spatial relational queries without and inadequate memory overhead.

The classic half-edge data structure ensures to solve any spatial query with a linear $\Theta(|V|+|H|+|F|)$ algorithmic complexity both as an average and as a worst case scenario. In order to perform spatial queries on the original input geometries an increased memory storage addition of minimum $O(|V|+|H|+|F|)$ is also required for metadata. While it is a considerable amount compared to the really compact representation of the data structure itself, it is also remarkable that required storage space can be optimized already at this level and effortlessly refined to $O(|V|)$ or $O(|H|)$, as the association to the input geometries are easily reconstructible when defined either for the vertices or the halfedges only.

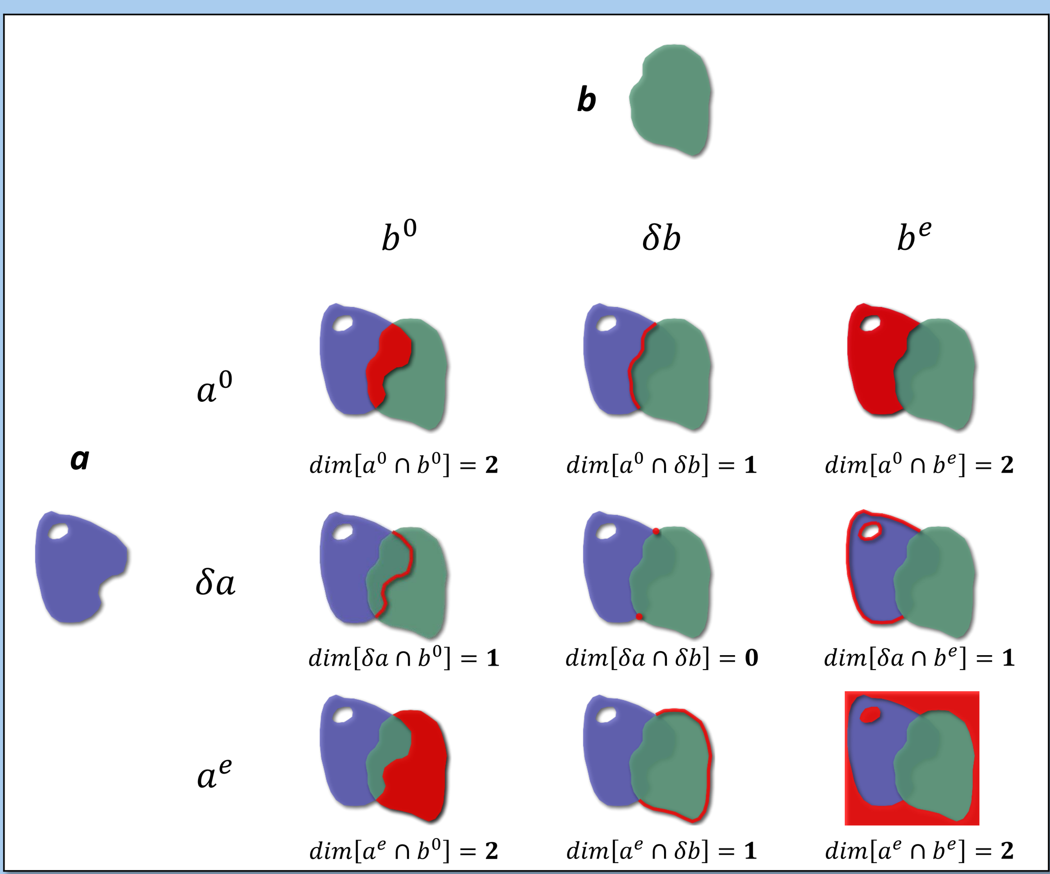
By indexing the vertices of a half-edge graph with the references of the represented original input geometry set *G*, the linear computational efficiency can be improved to logarithmic $\Theta(log(|G|))$, where $|G|{\sim}|F|$. This index structure does not necessarily require significant extra memory storage, since the metadata about the geometries attached to the components of the half-edge graph are no longer needed and can be omitted in exchange.

Spatial relations

The *Dimensionally Extended Nine-Intersection Model* (DE-9IM) [2] is a mathematical approach that defines the pair-wise spatial relationship between geometries of different types and dimensions by intersections of their interior (a^0), boundary (δa), and exterior (a^e) with consideration for the dimension of the resulting intersections. The intersections of respectively the interior, the boundary and the exterior of two subject geometries produce a 3-by-3 matrix, where each intersection can result in geometries of different dimensions. For example in a two dimensional space the dimension of an intersection can be -1, 0, 1, or 2, where the negative value corresponds to the null set that is returned when no intersection was found.

The DE-9IM matrix provides an approach for classifying geometry relations. There are 10 classification schemes (relations) standardized by the Open Geospatial Consortium (OGC) that have a name which reflects their semantics, namely: *equals*, *intersects*, *disjoint*, *touches*, *contains*, *within*, *covers*, *coveredby*, *crosses* and *overlaps*. [3]

Figure 1. Example of DE-9IM for two overlapping polygons.



Source of images: PostGIS Manual

Topological data structures

In many cases geospatial data sets are static or rarely altering, therefore the repeated recalculation of spatial relationships impacts the performance of geospatial operations superfluously. In these scenarios a topological representation created and stored in advance could significantly improve the computational efficiency of topological spatial queries. The usage of topological data structures like the *half-edge model* or the *winged-edge model* are widespread for the purpose of persisting the spatial relational information among the parts of the dataset, thus boosting the performance of later relational queries, meanwhile also eliminating the redundancy in the data set.

Most topological data structures store low abstraction level topological information, like the half-edge model [4] contains adjacency data through pair of halfedges. Deducing higher abstraction level spatial relations – defined by e.g. the DE-9IM – is a non-trivial task on such an abstraction level. The topological model requires the input geometries to be preprocessed into non-overlapping segments named *faces*, therefore the original geometries could be only stored as attached metadata and operations cannot be executed directly on them. As an example, the *overlaps* relation between geometry *a* and *b* can be defined as follows for the set of vertices (*V*), halfedge pairs (*H*) and faces (*F*):

$$dim(a) = \dim(b) \wedge$$
$$\exists v_1 \in V: v_1 \in a \wedge v_1 \notin b \wedge$$
$$\exists v_2 \in V: v_2 \notin a \wedge v_2 \in b \wedge$$
$$\exists v_3 \in V: v_3 \in a \wedge v_3 \in b \wedge$$
$$dim(a) = 1 \Longrightarrow \exists (h_1, h_2) \in H: h_1 \in a \wedge h_2 \in b \wedge$$
$$dim(a) = 2 \Longrightarrow \exists f \in F: f \in a \wedge f \in b$$

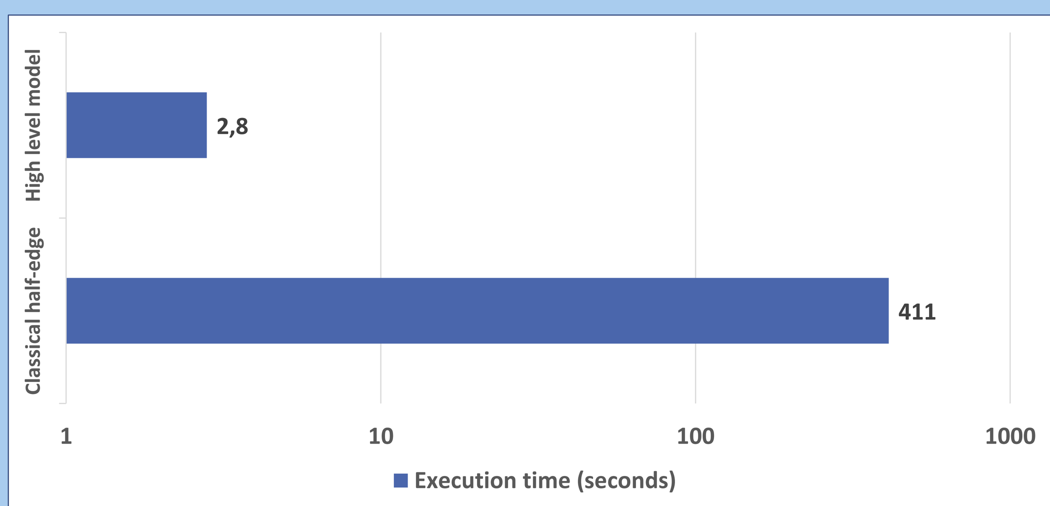
As the above formula demonstrates, the main efficiency issue is that all vertices, edges and faces must be observed to determine the spatial relation between any two input geometries (or layers), inflicting a significant performance drawback.

Results

To prove the usability of the described model an implementation was carried out with the intense usage and as part of the *AEGIS geospatial framework* [5]. The AEGIS system was initially developed for education and research goals at the Eötvös Loránd University, and is currently used both as a learning tool for computer science students and as a back-end engine for prototype implementations in GIS researches. It is based on the *.NET/Mono Frameworks* and implements well-known standards using state of the art programming methodologies with adaptability and extensibility in mind.

The accumulated test execution time of a few thousand high level spatial queries are displayed on Figure 3 respectively with the classic half-edge structure and the indexed representation.

Figure 3. Performance comparison on spatial query execution.



Contact

Author: Máté Cserép
Email: mcserep@inf.elte.hu
Date: January 15, 2016

References

1. E. Clementini, P. Di Felice, P. Oosterom. A Small Set of Formal Topological Relationships Suitable for End-User Interaction. *In Proceedings of the Third International Symposium on Advances in Spatial Databases (SSD '93)*, pp. 277-295 , Springer-Verlag, London, UK, 1993.
2. E. Clementini, J. Sharma, M. J. Egenhofer. Modelling topological spatial relations: Strategies for query processing. *Computers & Graphics*, 18 (6), pp. 815-822, 1994.
3. OpenGIS Implementation Standard for Geographic information - Simple feature access - Part 2: SQL option. Ed. J. R. Herring, Open Geospatial Consortium Inc., version 1.2.1, 2010.
4. M. Mäntylä. An Introduction to Solid Modeling. Computer Science Press Inc., New York, NY, USA, 1987
5. R. Giachetta. AEGIS – A state-of-the-art spatio-temporal framework for education and research. *In OSGeo Journal*, 13, pp. 68–77, OSGeo Foundation, Beaverton, Oregon, USA (2014).