[DM-02-020] Entity-based Models

Abstract

As we translate real world phenomena into data structures that we can store in a computer, we must determine the most appropriate spatial representation and how it relates to the characteristics of such a phenomenon. All spatial representations are derivatives of graph theory and should therefore be described in such terms. This then helps to understand the principles of low-level GIS operations. A constraint-driven approach allows the reader to evaluate implementations of the geo-relational principle in terms of the hierarchical level of mathematical space adopted.

Keywords: boundaries, fields, raster, tessellatoin, vector, Voronoi

Author & citation

Albrecht, J. (2022). Entity-based Models. The Geographic Information Science & Technology Body of Knowledge (2nd Quarter 2022 Edition). John P. Wilson (Ed.).
DOI: 10.22224/gistbok/2022.2.11.

This Topic is also available in the following editions:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Discrete entities. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Explanation

1. Context-setting Terminology
2. Hierarchy of Mathematical Spaces
3. The Vector Data Model
4. The Raster GIS Model
5. The Voronoi GIS Model
6. Entities as Combination of Spatial Footprints and Class Attributes

1. Context-setting Terminology

"Entity" is a term that the casual GIScience reader will likely only encounter in the context of entity-relationship diagrams in the database literature (see Conceptual Data Models). Here, it has been chosen because the ambiguity of more commonly used terms such as "object" or "feature" raises connotations we seek to avoid. In the eternal discussion of raster-vs.-vector (see Logical Data Models), entities represent the vector side: items defined by their boundaries – although this is overcome by representations such as Voronoi diagrams. As such, they are the data model counterpart to geospatial fields, phenomena that have no discernible boundaries, not even fuzzy ones (see section 6, below).

Regardless of the choice of data structure, we distinguish between the location of our object of interest and its characteristics, commonly encapsulated by the geo-relational

principle of spatial footprint and attribute. The spatial footprint can be a zero- to three-dimensional geometry, a Voronoi, or a raster cell (see Raster Data Model and Geometric Primitives and Algorithms). The attributes may be singular values, a table row, or a binary large object (BLOB). Inherent, although not necessarily stored explicitly, is a set of topological relationships between the spatial footprints.

The spatial footprint of entities can be expressed as point, line, or area geometry, as a set of adjacent, equal-valued raster cells, or as nodes, lines, or higher-dimensional simplexes in a graph. Before we discuss their respective implementations in the form of data structure, it is useful to lay a few mathematical foundations.

2. Hierarchy of Mathematical Spaces

Any discussion of the representation of spatial footprints touches upon a number of mathematical concepts that can be illustrated as a continuous application of constraints to mathematical spaces (Figure 1).

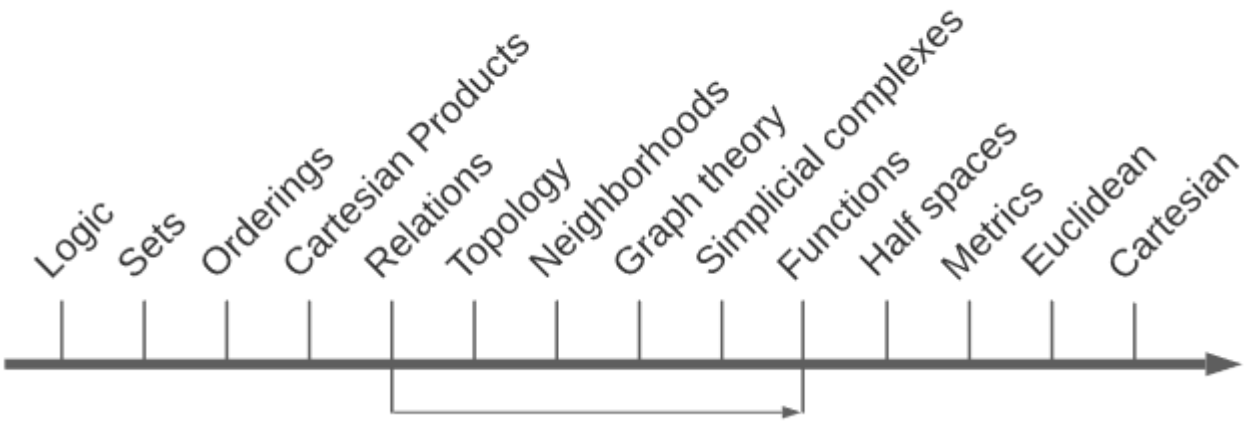


Figure 1. Hierarchy of mathematical spaces increasing specialization and adding constraints. Source: author.

At the lowest (most abstract) level, we have logic. We will use Boolean logic with its NOT, AND, and OR operators as a stand-in for a variety of logics but the interested reader is invited to have a look at Spencer-Brown's spatial logic (Spencer-Brown, 1969). As we move up (to the right) on our spatial progression, we come to sets, where we distinguish elements as belonging to one set or another and operations on these sets that remind us of vector GIS operations, namely union (logical OR), intersection (logical AND), difference or complement and disjunction (Figure 2).

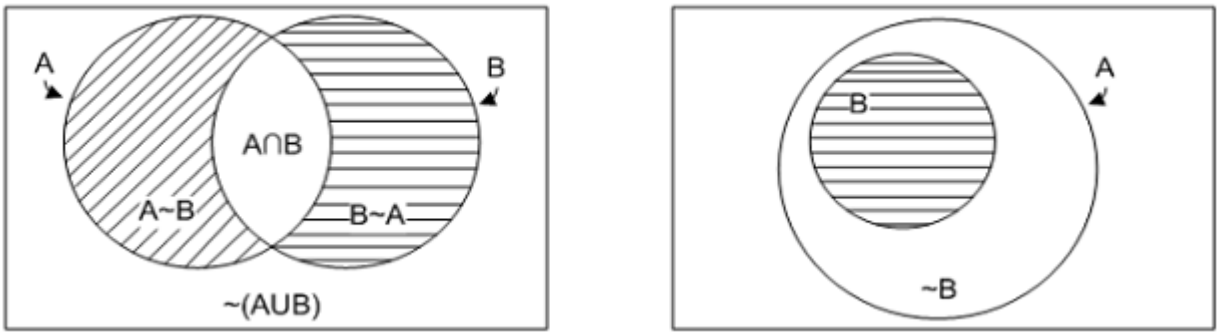


Figure 2. GIS-like operations on sets. Source: author.

In a set (see Set Theory), the order of elements does not matter; however, we may impose the constraint that it does matter, in which case we talk about the ordering of elements. The total sum of all ordered pairs between two sets is called a Cartesian product set and a subset of that is called a relation. Relations are defined either by the exhaustive listing of all ordered pairs or by specifying a property or rule. The topological rule of contiguity or the sharing of a boundary is an example of such a rule. A neighborhood of a point is a set of points containing that point where one can move some amount in any direction away from that point without leaving the set. The notion of leaving the set is defining a boundary, which is important for our definition of entities as well as for some of the quintessential overlay operations, which have been codified by Clementini et al.'s as the topological 9-intersection model (Figure 3).

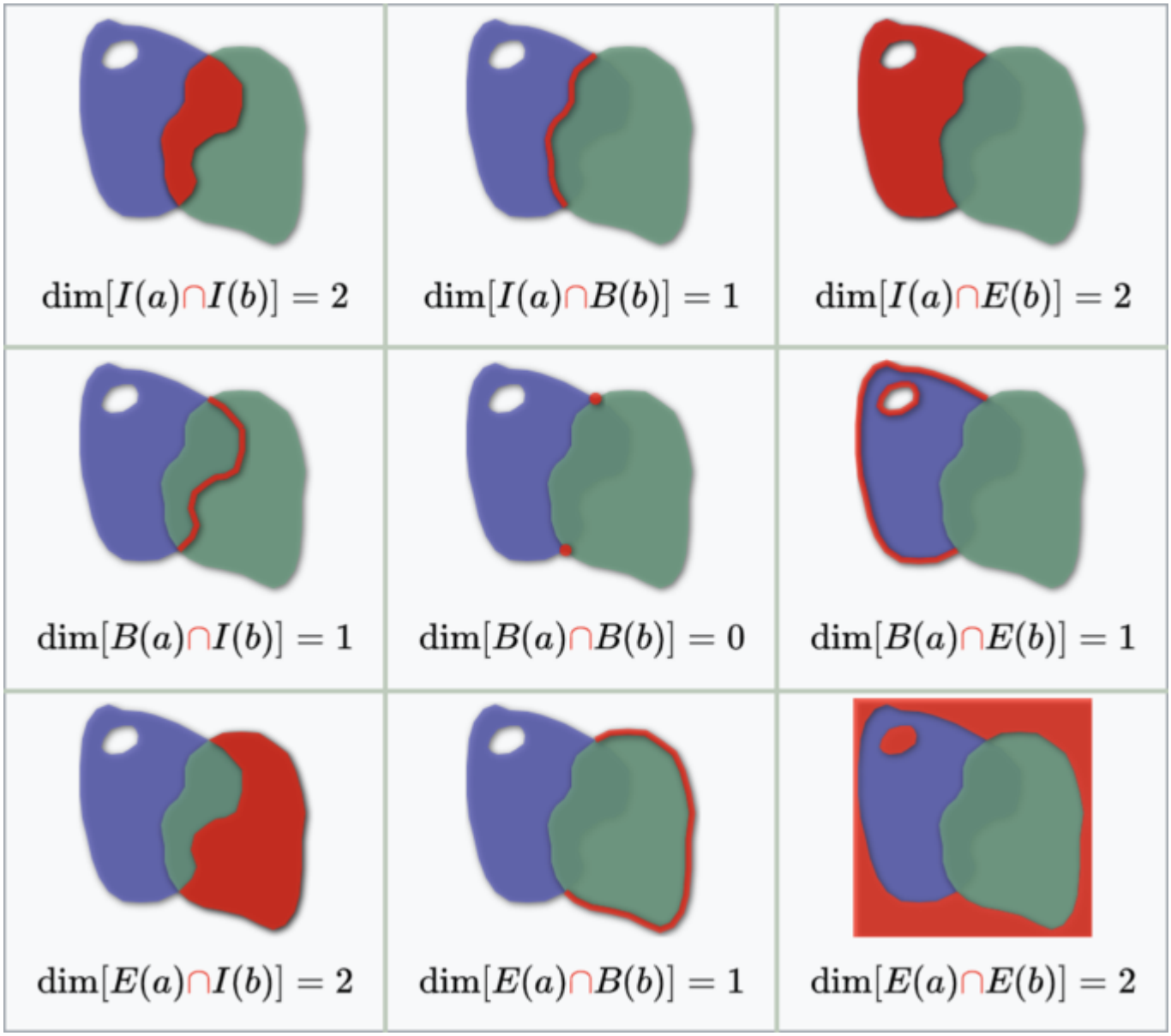


Figure 3. Topological relationships form the basis for (a) spatial predicates as well as (b) spatial footprint consistency checks in GIS. Source: Wikipedia.

Graph theory is the study of pairwise relations and their topological relationships. Graphs are n-1 dimensional structures with n being the number of nodes. A simplex is any triangle formed between three neighboring nodes of a graph. Neighboring simplexes (triangles) can then be combined to form simplicial complexes, which form the basis of Atkin's (1974) Q-analysis (see also Jiang & Omer 2007).

For the next step along the hierarchy of mathematical spaces, we are following a little shortcut bypassing topological spaces to simplicial complexes, to define functions as another form of relations. Functions are a special kind of relation where there is a correspondence between each element of the first set for each of the second, again forming ordered pairs. Parallel to the introduction of boundaries in topology, we now distinguish two spaces divided by the continuous line formed by a function. Each side of the two space is called a half space and we distinguish open from closed half spaces depending

on whether the boundary formed by the function is included or not.

So far, each set represents a dimension and all the relationships and their subsequent specializations as we follow our arrow of mathematical spaces to the right, have no metric associated with them other than the “distance” from one element of the set to the next one (in order). Therefore, in a graph, for instance, the length of the edge between two nodes does not matter (unless we create specializes, so-called labeled graphs). Similarly, there are no units of measurement in topological spaces or simplicial complexes; they are all qualitative in nature. The next step is hence to introduce a steady unit of measurement for the distance among the elements of a set. This is the idea behind a coordinate system, which initially is just providing a metric for each of the axes that span the dimensions of our sets. We may have as many axes (dimensions) as we want, they do not have to be perpendicular to each other and they may each have another unit of measurement. The multidimensional analysis of census variables would be an example of such a coordinate system, or the spherical coordinates of latitude and longitude. The most common example, however, is the Euclidean plane, which is simply the Cartesian product of two sets of real numbers. Yes, for GIS analyses, this is still not constrained enough. Here, we want all axes to have the same scale and to be perpendicular to each other. Only then do we have the luxury of applying Pythagoras’ theorem for simple distance measurements. And only then will high school math suffice to calculate quantitative relationships based on simple vector algebra.

3. The Vector Data Model

We needed the above stroll through mathematical spaces to follow the reasoning behind the formalism of entity-based data models. One more mathematical convention needs to be introduced here and that is the fact that any link between two coordinates that defines the boundary of a one- or higher-dimensional geometry must be a straight line (i.e., no curve). Following the terminology of graph theory, we identify three geometry types in a 2-dimensional plane (the formalism would apply to higher-dimensional graphs as well, but most GIS don’t implement them): nodes, edges and faces. They are used to represent point, line, and area entities. A consequence of the 2-dimensional plane is that as we project an n-dimensional graph onto a place, we are inserting nodes wherever to edges intersect. This is referred to as planar enforcement, and along the lines that constraints are good, it causes the vector data model to be a lot easier to implement than it would be without that enforcement (Figure 4).

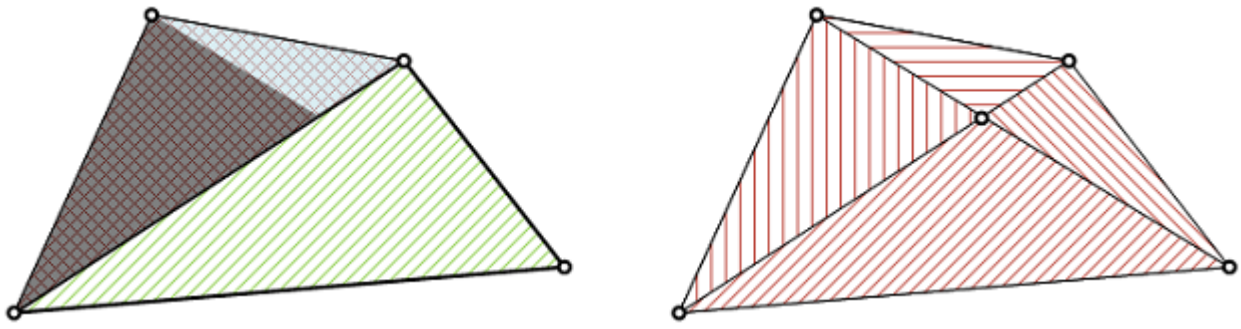


Figure 4. Planar enforcement. Source: author.

In section 2 above, we distinguished between labeled and unlabeled graphs. The latter allow us to impose directionality in the definition of edges and faces, which adds the benefit of uniquely identifying neighborhood relationships. The final piece is the agreement that the study area that holds all our spatial footprints is the entity that supports everything else: it is face zero or f0. With this, we can now exhaustively and uniquely describe the relative and absolute positions of anything that has a coordinate. In particular, we can establish the relationship of every geometry, be it point, line, or area to every other geometry in our database, and from that we can deduce all the relationship depicted in Figure 3, which is the basis for the vast majority of GIS functions. In earlier GIS systems, these relationships were explicitly stored (after being calculated) in a complex web of tables. With the advancement of computational power, the distance and topological relationships are now commonly calculated upon demand – and typically for the entities displayed on a screen rather than all entities of the database.

4. The Raster GIS Model

We jumped through a lot of mathematical hoops to describe a vector data model that could be a somewhat simpler if we weren't to express it in graph theoretical terms. We did this because on our arrow of a hierarchy of spaces, graphs are quite abstract and hence can spawn different complementary models of (geographic) space. The raster GIS model, typically applied to field-like phenomena can now be interpreted as defining tightly constrained faces with point entities represented as nodes, line entities as (a concatenation of) edges and area entities as one or more adjacent cells, where each cell is a face as defined for the vector model. The unique advantage of the cells in a raster GIS is that they all have the same size, shape, and orientation, which spares us the need to describe them repetitively. Distance and topological operations are now trivial. Let's keep in mind that although fields represent a continuous conceptualization of space, due to the discrete nature of our von Neuman based computers, we are forced to first discretize space into cells and then use the implicit topology to "sew the tiles back together" (Gold 2015, p. 16).

5. The Voronoi GIS Model

This Gold quote appropriately moves us into the realm of the Voronoi GIS model. More than anyone else in the realm of GIS, Gold advocated for the Voronoi GIS model as a bridge between the traditional chasm between the worlds of raster and vector, object versus field (Couclelis 1992). Voronoi diagrams (also known as Thiessen polygons or Dirichlet domains) are the dual of a planar graph (see Thiessen polygons for an elucidation of the concept). They can be applied to nodes, edges, and faces and define their respective neighborhood such that the inside of their faces is closer to their generating source geometry than to any other one (Figure 5). This has algorithmic benefits because all edit operations and their topological updates are now restricted to the neighborhood defined by this dual structure. If both the original planar graph and its Voronoi dual are stored, then these may serve as a recursive hierarchical index, which makes the suitable for very large data sets such as all property boundaries in a country.

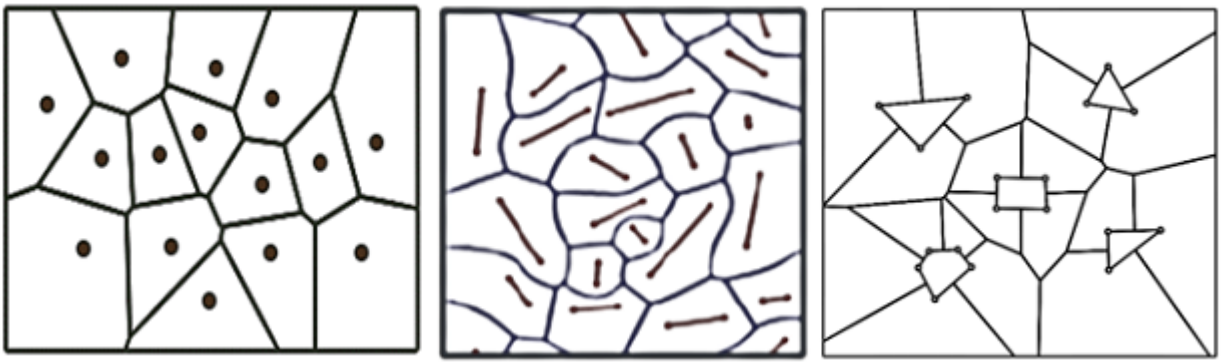


Figure 5. Voronoi diagrams of original nodes, edges, and faces. Source: author.

6. Entities as Combination of Spatial Footprints and Class Attributes

Returning to the definition of geospatial entities in section 1, our attention now turns to the characteristics of the entities whose spatial footprint occupied our attention in the previous three sections. Those characteristics or attributes can be anything we want them to be: quantitative or qualitative, structured or unstructured; as long as they support one research question or another. In order to be subject to query or analysis, those descriptors must be subject to some form of classification. i.e., we must be able to assign each attribute and subsequently the whole entity to one set or another (although fuzzy and hence overlapping memberships may be allowed in more advanced systems). The geo-relational principle uniquely and unambiguously links each combination of attributes with a particular geometric representation and vice versa. In other words, any change in a part of the spatial footprint requires the introduction of a boundary that delineates the partial change in the original entity. Our graph-based boundary notation, regardless of whether the entity is encoded as point/line/area geometry, as a graph, as a raster or as a Voronoi diagram, is not allowed to contain any changes among its descriptors within the bounds of its spatial footprint.

If class attributes are alphanumerical, then we may use set membership rules to reclassify entities with a likely effect on the boundaries of the spatial footprint (see Point, Line, and Area Generalization). Technically, this is addressed by the proverbial (set of) overlay operations(s) (see Overlay) but conceptually, this touches on unresolved problems of scale dependence (see Scale and Generalization and Data Properties) and ontology (see Ontology for Geospatial Semantic Interoperability and Problems of Scale and Zoning). An interesting twist on the relationship between geometries and attributes is the digital elevation model, where we store a single elevation attribute for each raster cell or vector geometry giving us a 2.5 dimensional representation as opposed to TINs/Voronoi diagrams, which are true three-dimensional.

Given the crucial role of higher-level mathematical spaces in all aspects of geospatial entities discussed here, it is worthwhile to return to set theory and look at the implications of not working with binary memberships but membership functions as in fuzzy set theory. Traditional approaches to GIS assume perfect knowledge about both the positional and the thematic characteristics of spatial data, a notion that may be unjustified (see Spatial Data Uncertainty). Burrough & McDonnell (1998) discuss lucidly the options provided by applying

membership functions to each aspect of an entity and the operations that then still allow for all the GIS functionality that we are used to.

References

Atkin, Ronald 1974. Mathematical Structure in Human Affairs. London: Heinemann.

Burrough, P. A. and McDonnell, R. (1998). Principles of Geographical Information Systems, 2nd Edition. Oxford: Oxford University Press.

Clementini, E., Sharma, J. and Egenhofer, M. J. (1994). Modelling topological spatial relations: Strategies for query processing, Computers & Graphics. 18 (6): 815–822.

Couclelis H. (1992) People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In: Frank A.U., Campari I., Formentini U. (eds) Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science, vol 639. Springer, Berlin, Heidelberg.

Egenhofer, M. J., & Franzosa, R. D. (1991). Point-set Topological Spatial Relations. International Journal of Geographical Information System, 5(2), 161-174.

Gold, C. (2016) Tessellations in GIS: Part I—putting it all together. Geo-spatial Information Science, 19:1, 9-25.

Jiang, B. and Omer, I. (2007). Spatial Topology and its Structural Analysis based on the Concept of Simplicial Complex. Transactions in GIS, 11(6): 943-960.

Molenaar, M. (1998). An Introduction to the Theory of Spatial Object Modelling. London: Taylor & Francis.

Rigaux, P., Scholl, M., and Voisard, A. 2002, Representation of Spatial Objects, in Spatial Databases with Application to GIS. San Francisco, CA: Morgan Kaufmann.

Spencer-Brown, G. (1969). Laws of Form. London: Allen & Unwin.