BUILDING A HYPERGRAPH-BASED DATA STRUCTURE

The Examples of Census Geography and the Road System

ROBERT D. RUGG, *Department of Urban Studies and Planning, Virginia Commonwealth University, Richmond, Virginia, USA*

INTRODUCTORY CONCEPTS

HYPERGRAPH-BASED Data Structures are built on four fundamental concepts, which taken together are sufficient to describe any real world phenomenon: object, class, attribute, and relation. The concepts combine to form the six basic 'abstract data types' of the HBDS model: object, class, attribute of object, attribute of class, relation between objects, relation between classes. Figure 1 illustrates the symbolic representation of these data types as developed by Bouillé. 'Objects' of 'Classes' correspond to 'elements' of 'sets' in set theory; an 'Object' may also be viewed as a vertex in graph theory; an 'Attribute' is a property of an object (viewed as an element or as a vertex); 'Relations' correspond to links (either 'arcs' or 'edges') in graph theory. A hypergraph (Berge 1970) is a graph in which more than two vertices are linked by the same edge, hence allowing for the manipulation of 'Classes' as well as 'Objects'. The theory of hypergraphs thus provides a conceptual basis for combining set theory and graph theory in the same data structure model. (The complete model is presented in Bouillé 1977).

CENSUS GEOGRAPHIC UNITS

Census geographic units include (among others) states, cities, counties, towns, minor civil divisions, tracts, block groups and blocks. These fit more or less well into a hierarchy of areal units, as blocks nest into block groups, block groups into tracts, tracts into cities and counties, and the latter into states. The relationships could also be diagrammed as a rooted tree, with the State as the root. However, the structure of Census geography is already more complex than a hierarchy in the ordinary sense. At least two hierarchies are present: in Virginia for example, cities are geographically and juridically independent of their adjacent counties. The relationships between the two 'trees' also pose several complications. Both trees have the same root, the State; in Virginia, cities are 'county equivalents' for some purposes but not others (e.g., counties are subdivided into minor civil divisions whereas cities are not); portions of some counties are

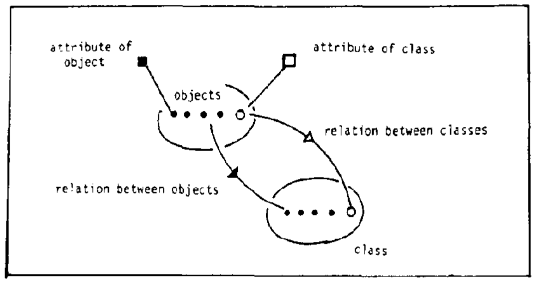


FIGURE 1. *Six abstract data types of the HBDS model.*

tracted while others are not, and the tracts and block groups in the portions of counties which are tracted may be split by, rather than nesting into minor civil divisions. Finally, again using the example of Virginia, the Census Bureau recognizes three kinds of 'places': independent cities, incorporated towns, and 'census designated places' or unincorporated towns. Places, also, may or may not nest within a given minor civil division. (See Figure 2a).

Considering the full range of relationships indicated in Figure 2a, it is clear that the hierarchical model is inadequate to fully describe the relationships that actually exist. A next step in HBDS methodology is to find commonalities in some of these relationships and, in so doing, to simplify the structure to present a clear, more generalized view of the phenomenon. In Virginia, where cities are independent of counties, counties and cities have many equivalent relationships and may thus be treated, for some purposes, as a 'hyperclass' – a set of classes whose objects may each verify the same relationships with objects of other classes in the system. Counties and cities share the same relationships with the state as a whole and with tracts and places (viewing the independent city as both a 'place' and as the county equivalent it nests within, in accordance with Census Bureau usage). Counties and cities thus form a 'hyperclass' and the relationships of inclusion between them and certain higher and lower order geographic units are 'hyperlinks'.

The twin relationships 'split by' or 'nests in' apply to tracts, block groups and places in relation to minor civil divisions. Tracts and places thus form a hyperclass in their relationships to cities and counties on one hand, and minor civil divisions on the other hand. The relationships 'split by' or 'nests in' can be subsumed under the HBDS concept of 'multilink', thereby further simplifying the diagram. The multilink between minor civil divisions and tracts and places, the latter being a hyperclass, would be termed a 'hypermultilink'. The power of the HBDS model can begin to be understood when one realizes that, with relatively little effort, the 16 relationships indicated in Figure 2a have now been reduced to only 8, as shown in Figure 2b. While the diagram has been simplified, none of the information concerning the relationships among the various geographic units has been lost in the process.

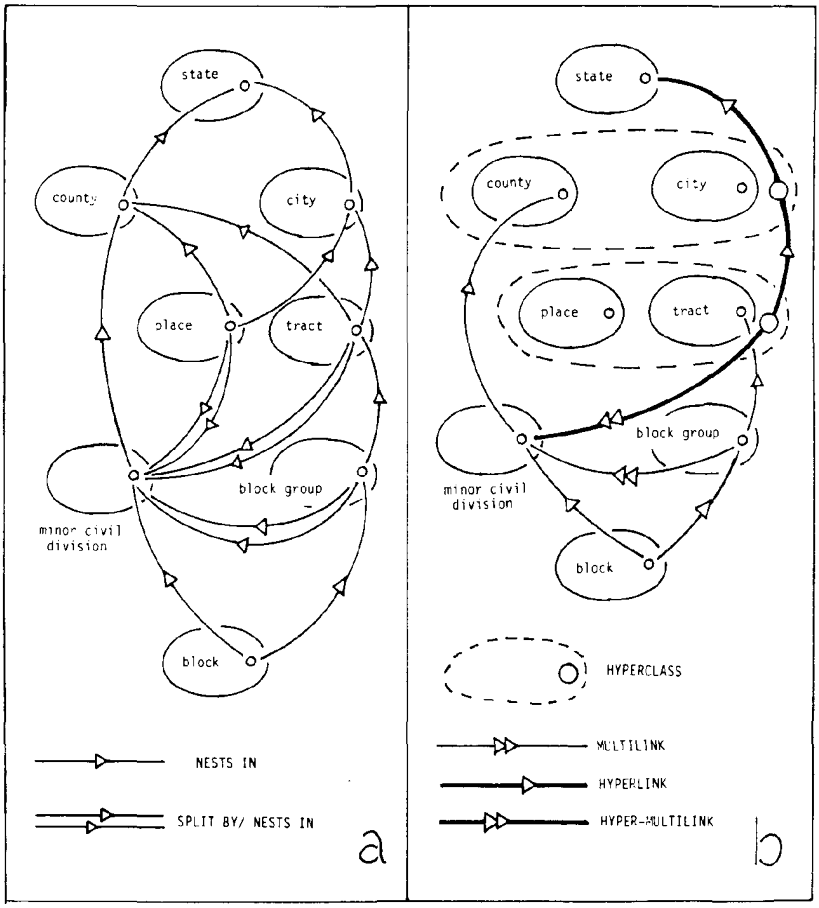
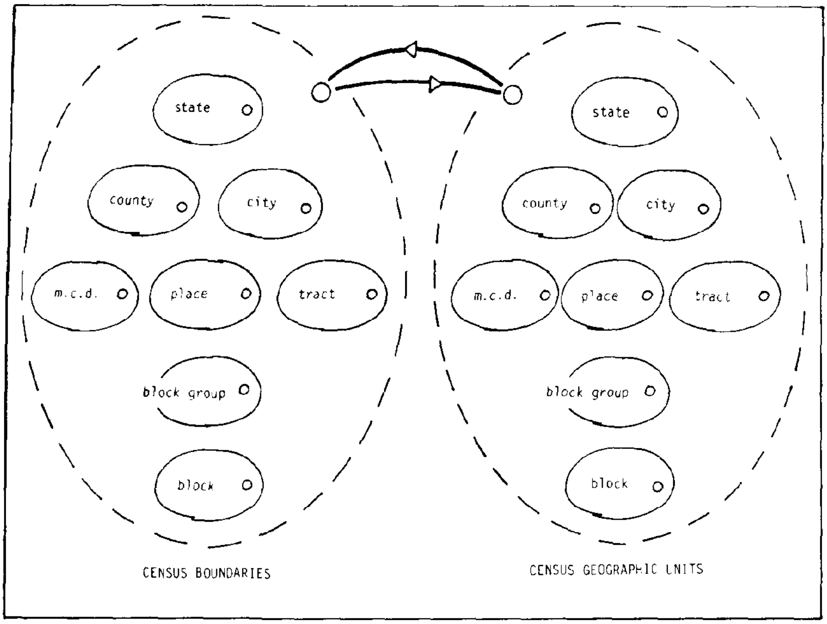


FIGURE 2. *Census area geographic structure:* a *relationships among census geographic areas;* b *structure using hyperclasses and multilinks.*

RELATIONSHIP TO THE ROAD NETWORK

The road network relates to Census geography in that it forms the boundaries of Census units. However, not all Census areas are bounded by roads: rivers, railroad lines, utility rights-of-way, invisible administrative boundaries, and other features are also used. Bouillé (1980: Figures 9–11) has already solved the analogous problem of relating administrative entities to administrative boundaries. Following his initial model, all of the Census geographic units shown in Figure 2 can be considered a hyperclass, related to a second hyperclass of Census area boundaries. (These constitute hyperclasses because, for example, a census block boundary may also be part of a county boundary, and in general all types of boundary lines potentially help define the limits of all types of Census geographic units. (See Figure 3a).



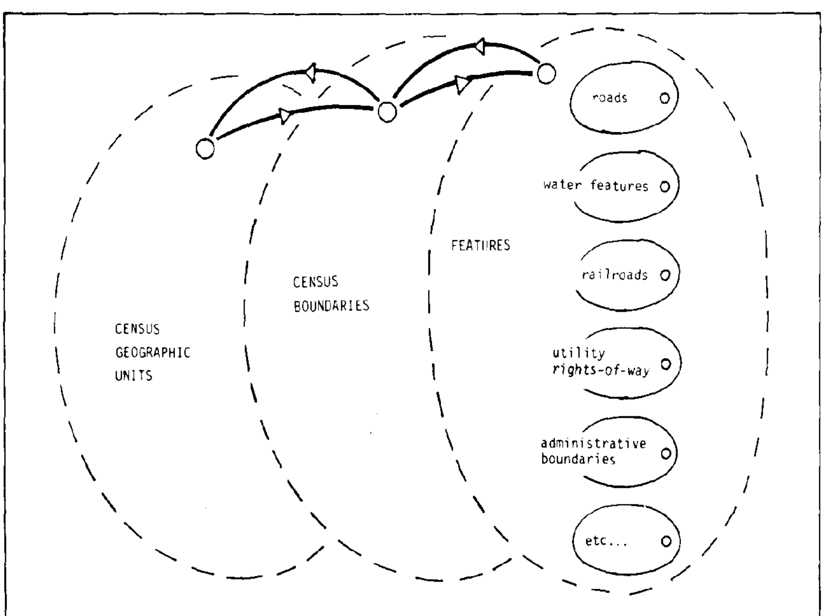


FIGURE 3. a *Census geographic areas and boundaries (after Bouillé);* b *Census areas, boundaries and boundary features.*

Considering that the road network is but one element of the various features used to define Census boundaries, the problem of relating Census geography to the road network can easily be solved by introducing a third hyperclass of features from which the various kinds of Census boundary lines are selected, as shown in Figure 3b. However, this view does not take into account the structure of the road system itself.

THE ROAD SYSTEM

The 'road system' consists of various types of roads (interstate highways, U.S. routes, state highways, county roads, etc.). Although it is possible to conceive of each road type as a separate class, in general it is sufficient to consider the type of road as an attribute of the class 'road'. The class road is illustrated in Figure 4a as the class 'whole roads', in order to emphasize the distinction between whole roads as features and road segments as elements of a conventional transportation network, described below.

In a classical network model, emphasis is given to the intersections between whole road (nodes, or vertices, of the road network), and to the resulting road segments connecting each pair of intersections (links, or in graph terminology, 'arcs' or 'edges' of the road network graph).

A complete view of the road system incorporates both perspectives: roads as features and roads as the building blocks of a transportation network composed of nodes and links. Roads, intersecting with each other, produce road segments which, together with the intersections, form the graph of the road network. Further, the road segments are related to the whole roads of which they are a part, sharing certain of their attributes including the type of road. The road system can thus be fully described in terms of the relationships among three classes – whole roads, intersections, and road segments – together with the attributes of each class (see Figure 4b). Of these three, it is the class 'road segments' that forms portions of the boundaries of Census areas. As it turns out, the result presented in Figure 4b is the same for any network, as has been demonstrated in LeForestier (1979) and LeGraverend (née LeForestier) (1980). It depends on transforming one abstract data type (intersection, which begins as a loop on the class 'whole roads'), into another (the class intersections, which constitute vertices in the graph of the road network, and divide whole roads into segments, the latter constituting the arcs of the road network graph).

CENSUS GEOGRAPHY AND THE ROAD SYSTEM

If all the Census geographic areas were bounded by roads, the problem of relating Census geography to the road network could be solved at this point by a simple extension. Regardless of the higher order structure of Census geographic areas, the fundamental unit is the Census Block, which nests into all higher level areas and is never split by any of them. Beginning with the graph of the road network (say, graph G), one can define the dual (G^*) as the graph of areas bounded by roads and intersections (each such area represented as a vertex of G^*), having links of adjacency defined as the boundaries between the areas (the arcs of graph G become arcs of G^*). Each Census block would then be composed

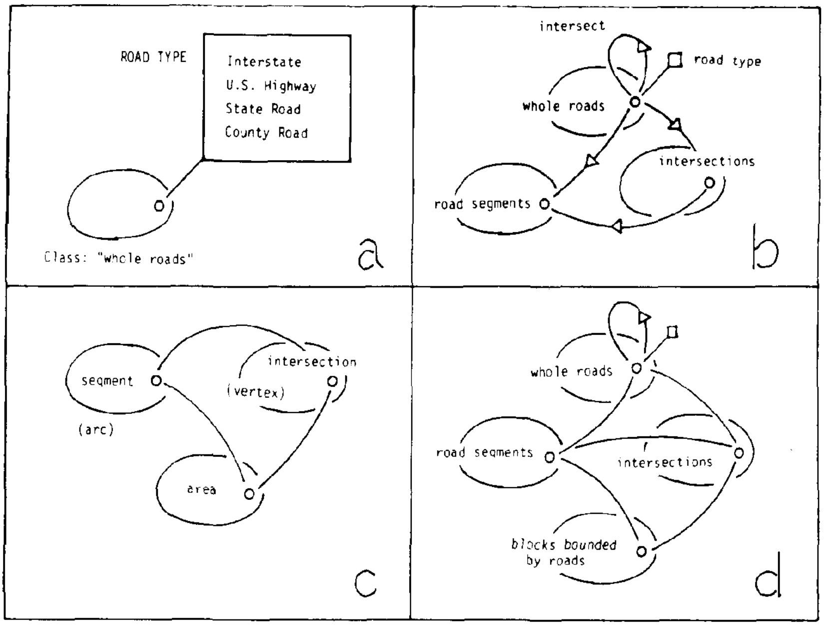


FIGURE 4. *Structure of the road system:* a *whole roads and road types;* b *road network (after Le Graverend);* c *generalized network model (after Bouillé);* d *synthesis of* b *and* c.

of one or more of the areas defined in G*. The simplest case is illustrated in Figure 5. Census Block 101 is bounded on the north by Prospect Avenue, on the east by Second Street, on the south by Glendale Avenue, and on the west by First Street. The physical block defined by the road network as block (1) is the same as Census Block 101.

The definition of areas bounded by the road network results in a new class 'areas' and the extended general solution to networks of all kinds, including the areas bounded by network segments, presented by Bouillé (1983), as shown in Figure 4c. Unfortunately, the reality of Census boundaries is more complex, involving not only roads but many other kinds of features that potentially serve as the boundaries of Census Blocks. Consider hypothetical Census block 102 (Figure 5). It is bounded on the north by Prospect Avenue, on the east by a Pennsylvania Power and Light Company power line, on the south by Glendale Avenue, and on the west by Second Street. It is intersected by Harry's Brook, which is not a census boundary here, but is used as a census boundary elsewhere along its course. To define the area of Block 102, it is necessary to consider all the intersections between the road network, the electric power distribution network, and the stream network. Each of these networks, of course, has its own structure, in general composed of classes, analogous to those of 'whole roads', 'road segments', 'intersections', and 'areas bounded by roads'. However, for our immediate purpose, the structure of each network considered separately is not

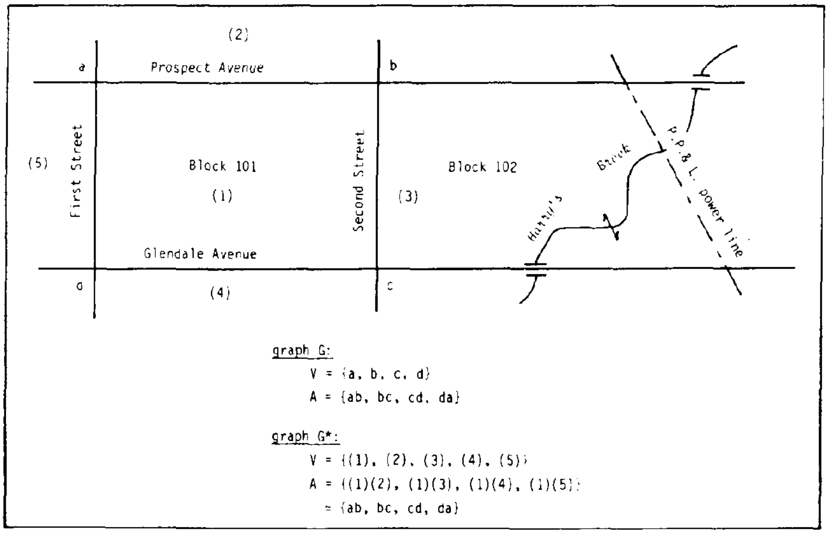


FIGURE 5. *Two hypothetical census blocks: (101) block bounded by roads; (102) block bounded by roads and other features.*

important; rather, we are interested in a different network, composed of all the segments and intersections formed by all the features that potentially serve as Census area boundaries.

The more general 'feature network' consisting of the segments formed by the intersection of all such features is illustrated in the upper portion of Figure 6. The segments defined by the intersections of all features become the building blocks of Census area boundaries and geographic units, as shown in the lower portion of Figure 6.

CONCLUSION

In describing the structure of the road system, obviously some information about the system is lost if it is modelled simply as a network. Conceiving of 'whole roads' as entities within a hierarchical model does not fill the gap, either. The notion of hierarchy, while tempting to apply to Census geographic units, fails to describe all of the potential relationships among elements of the latter system. Without using HBDS methodology, it would be all but impossible to include such divergent phenomena as a road network and the quasi-hierarchy of Census geographic areas within the same data structure model. Using HBDS, not only is this task made possible, but also, through such higher level constructs as 'hypergraph', 'hyperlink', and 'multilink', a generalized overview of the structure is provided, without loss of any of the detailed information concerning specific relationships of individual objects within the larger system.

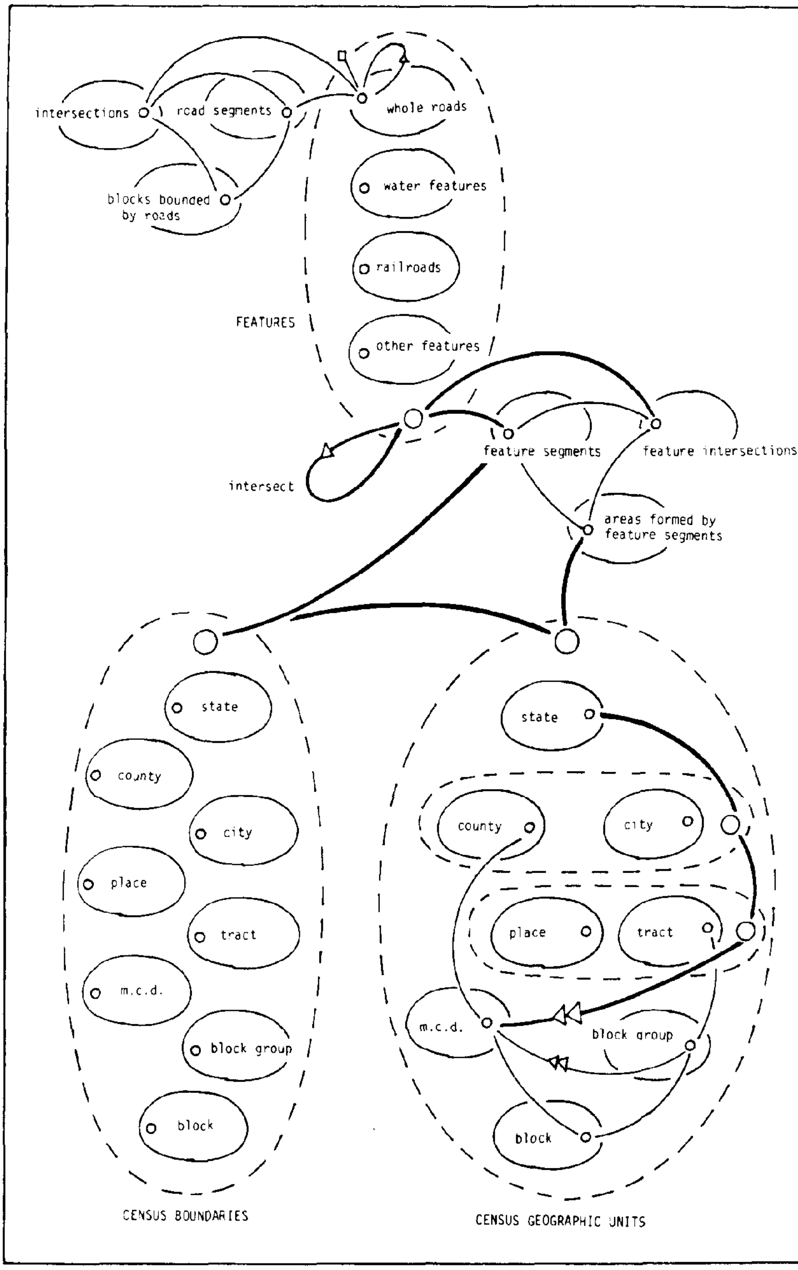


FIGURE 6. *Census geography and the road system.*

ACKNOWLEDGEMENTS

The research was supported by a Faculty Development Grant, School of Community and Public Affairs, and the Department of Urban Studies and Planning, Virginia Commonwealth University, 1982.

REFERENCES

BERGE, C. 1970. *Graphes et hypergraphes*. Paris: Dunod. English version: *Graphs and hypergraphs*, translated by E. Minieka. New York: North Holland Publishing Company, 1973.

BOUILLÉ, F. 1977. *Un modèle universel de banque de données, simultanément partageable, portable et répartie*. Thèse de Doctorat d'Etat ès Sciences Mathématiques (mention Informatique), Université Pierre et Marie Curie.

——— 1980. Contributions of graphs and hypergraphs to cartographic information theory. Translated by R.D. Rugg from 'Apport de graphes et hypergraphes en informatique cartographique,' *Bulletin du Comité Français de Cartographie*, March 1977. Richmond: Virginia Commonwealth University, Center for Public Affairs.

——— 1983. Lectures presented in a week-long series at *HBDS Tutorial '83*, Richmond Virginia.

LEFORESTIER, F. 1979. Les structures d'une banque de données urbaines. *Proceedings of the First International HBDS Seminar*, Lisbon, Portugal.

LEGRAVEREND, F. (née LeForestier) 1980. Cartography and urban planning: a tool for decision-makers. *Communications Presented at the Tenth International Conference of the International Cartographic Association*, Tokyo, Japan.

US DEPARTMENT OF COMMERCE, Bureau of the Census 1981. *Census of population and housing, 1980: summary tape File 1. Technical documentation*, Washington, D.C. File structure and Geographic coverage, pp. 9–16.