TOPOLOGICAL INFORMATION SYSTEMS FOR GEOGRAPHIC REPRESENTATION

Nicholas R. Chrisman
Harvard University

INTRODUCTION

The development of geographic information processing has occurred at such an explosive rate that we risk losing sight of our basic mission to understand spatial distributions and processes. Decisions originally made by adopting specific tactics in the organization of information now threaten to split our efforts into widely divergent streams. On one hand, grid systems may originally stem from programming expediency, but they are now reinforced by a supporting technology of satellites, raster scanners and special purpose computers.

The alternative to grid processing has also developed, but in radically different directions. Originally characterized as polygon or line and vertex systems, the new development has created a family of geographic information systems based on topology. The division between these two approaches transcends the technical arguments of efficiency and practicality. For, each approach imposes different attitudes with regard to the nature of space.

I cannot pretend to present both sides of a balanced debate between grid and topologic philosophies. Grid systems have many successful and perceptive proponents, plus large public financing for their hardware requirements. I wish to concentrate here entirely on the topological approach, presenting first the underlying ideas and an evaluation of the main weaknesses of existing attempts to implement them. Second, I will review the organizing principles and present status of the GEOGRAF project now under way at the Laboratory for Computer Graphics and Spatial Analysis.

I believe that topological systems will offer more effective geographic representation than that available through the grid approach. But I also hope that the gulf between the two philosophies does not become so wide that dialogue becomes impossible between their supporters. The final product of spatial analysis should take precedence over the transitory exigencies of data processing.

ASSUMPTIONS OF THE TOPOLOGICAL APPROACH

The first task of any project should be to define the eventual goals of inquiry. Developing a system to represent geographic entities requires the acceptance of some assumptions about the nature of space. Unless these assumptions are clearly stated, there will be a gap between expected and actual capabilities. Many systems fail not because they operate incorrectly, but because they do not provide as complete a spatial tool as required.

ENTITIES, CLASSES AND ATTRIBUTES

Dealing with space at the human level of experience, we are always dealing with aggregation and generalization. For different purposes we identify objects at different scales. A metropolitan area is a useful unit for analysis at the national scale. Each community must be distinguished inside the metropolis as analysis becomes finer. Eventually individual blocks or land parcels become the important spatial unit. The unifying factor in the progression parcel-community-metropolis is that some spatial aggregation is appropriate at any given scale. While the spatial unit has internal structure, our level of analysis allows it to be considered indivisible and relatively homogeneous in its attributes.

Thus for a given purpose, we naturally tend to distinguish a given set of discrete geographic entities. The criteria used to determine these entities usually form generic classes based upon specific attributes. For instance, the land use class consists of industrial, residential, agricultural, and other zones. By implication these uses are mutually exclusive, and so is their geographic expression. Similarly, linear objects such as roads can form a class which has network structure. In either case, the class is a set of geographic entities derived from a common set of attributes, thus sharing spatial character and structure.

Every entity has internal consistency at a particular level of analysis, and shares properties with similar, disjoint entities. But the entity need not have a specific size, shape, or orientation. It need not and should not be broken down to fit smaller cells, nor should it become mixed with others if it crosses some arbitrary gridding. The integrity of each geographic entity should be preserved.

NEIGHBORHOOD RELATIONSHIPS

Because they are located in a particular manner on the surface of the earth, all geographic entities are related to one another. The relative location of entities, in fact, describes the armature for spatial interaction and the overall structure of a class. Within the class of countries of the world, the contiguity, for example, of France and Germany is a significant fact. Within the class of land use, the proximity of residences to pollution generating facilities is similarly critical. Analysis of geographical information must take into account the basic connectedness and interdependence of all geographic entities. If an artificial unit such as a grid is used, the relationships are not those of the underlying geographic entities.

TOPOLOGY

A simple, concise means to maintain the integrity of entities, and to represent neighborhood relationships exists in the flexible science of topology and its dealings with the manipulation of networks. In our linguistic description of places we find the essential units of topological representation. In the case of France, we recognize that it is an area bounded by a set of frontiers with other countries and with the oceans. These frontiers are another sort of entity, dependent on the countries for definition, but serving a linear, not an areal function. The French-German frontier cuts across the Lorraine plateau, then follows the Rhine, terminating at Basel. This terminal point is also the terminal point of other frontiers. The relationships of areas, lines, and points in this example could be formalized by using a topological structure of polygons (cycles), boundaries (edges), and nodes.

Topological notation emphasizes the generic structure of geographic entities by stripping down to essentials. When dealing with complex problems where classes of zones overlap, the topological principles force us to recognize that only one network, the combined intersection set need be defined.

EXISTING SYSTEMS

The growing family of topological systems attests to the relevance of these concepts in a broad range of fields. Unquestionably, the original impetus can be found in the rapid creation of the DIME files for the 1970 Census. (Maxfield and Cooke, 1967). These concepts have been refined in the POLYVRT (Laboratory for Computer Graphics, 1974) and BZIS systems (Blumberg, 1974) among others. Also, the NRIS and LUDA systems demonstrate a parallel development without direct links to DIME.

Reviews of development and comparisons of relative efficiency of some of these systems have appeared and do not require repetition (Chrisman, 1974; Peucker and Chrisman, 1975). I wish instead to focus on the future of topological systems by describing first their main weaknesses and finally some attempts to solve them.

WEAKNESSES

Although topology offers a natural and effective means to formalize geographic information, there are a number of important needs that are poorly served at present. In general the weaknesses are not conceptual, they represent gaps between theory and reality to be filled by programming effort.

LEAST COMMON GEOGRAPHIC UNIT

The most important difficulty involves what has been called the "polygon overlay" problem. Most of the development of topological systems has centered on city street networks and administrative zones. In these applications the nature of the network frequently can be known from the start. The city block is a very reasonable choice of indivisible unit for the DIME files. A topological structure can be based entirely on hierarchical groupings of this unit with very little distortion. However, topological systems must seek applications with substantially different ground rules. In environmental situations, particularly, there is no unit which can be identified from the start as the common denominator. Attempts to do so are doomed to answer only limited questions (viz. James River Project).

Typically environmental data is composed of a series of classes of polygons which must be intersected and stored in an integrated system. To perform this task without recourse to the simplicity of a uniform grid is a great challenge, although the conceptual process is understandable. Taken together all the polygonal classes form a single network of zones which are the intersection set of all the classes. These zones may not bear direct relationship to any single class, so they need a new name: Least Common Geographic Units (LCGU's). Before the topological systems can claim generality of application, a procedure to effectively create the network of LCGU's from a complex set of polygonal and linear classes must be built. I do not think the existing, rather tradeworn answers to polygon intersection are capable of processing complex networks at a reasonable cost. The computational complexity must be lowered, and also the ability to recognize equivalent but slightly differently coded features must be enhanced.

DIRECTORIES

The second main weakness in existing approaches concerns the relative importance of information and methods to manipulate it. Too much emphasis is now put on data banks, and too little on software to use them. The relationship should be symbiotic; new software should allow refinement of information structure. At present, the DIME file is the main product of the DIME development. Beyond the maintenance modules and the address-matching capabilities, very little DIME software exists. I believe this lack is caused by the rigid external structure of the DIME file which seriously inhibits the design of an efficient internal data structure. There is only one type of record, the two point segment, in the DIME file. There are no directories that allow other information to be easily extracted. The topological relationships between geographic entities should not be merely theoretically possible, but also practically available. For instance, to use the DIME file to extract the outline of a zone, one should have an easier way than duplicating all records, sorting the whole file, and then linking segments. While the present DIME system allows for concise, compact external storage, there should also be a concern for internal processing requirements. It seems a waste to recreate directories from scratch for each program run.

GEOGRAF

Fulfilling the potential of the topological concepts is a challenge that has already attracted a number of researchers. Because most of the other projects are incomplete, the focus of this paper is the GEOGRAF project at the Laboratory for Computer Graphics. I will summarize the basic principles which underlie GEOGRAF and describe some of our early decisions in the system design. The audacity of our undertakings prompted a colleague to call the project MEGALOMANE I. But, in spite of high goals, I believe the project is reasonable and its goal attainable.

COMPLETE TOPOLOGY

The initial assumption behind GEOGRAF is an unqualified acceptance of topological analysis applied to geographic representation. Each class of geographic entities is taken to be a network of a known basic structure. A complex system of many classes is reduced to a single underlying intersection network. This lowest level has the structure of polygon, chain, and node developed for POLYVRT. The external data structure for POLYVRT which uses a multi-segment chain as the basic unit will provide the startling point. Above the lowest level, each entity is described in terms of components at the lowest level. A GEOGRAF representation should be a faithful translation of all spatial relationships into the data base.

STORE DIRECTORIES

Not only will all relationships be entered, they will be readily accessible through a software-generated internal data structure. Tables recording the chains incident at a node, the chains around a polygon, the lower level chains forming upper level boundaries, and more will all be constructed once and stored in the basic GEOGRAF data structure. The ease of access to this information should promote generality of application, flexibility of use, and efficiency of operation. The basic file structure required to operate these directories should be working by January 1976.

NEW PROCEDURES

In particular, many operations which now place large demands on computer resources can be restructured because of easy access to full topological information. If GEOGRAF is to work, the first priority must be the ability to combine diverse networks into the structure of LCGU's. In a previous paper (Chrisman, 1974) I have presented a method which, given access to full topological information on two networks, performs the work of intersection. The method presented operates at or near a linear order of complexity, because there is no need to compare every entity with every entity if the network connectivity is used. This procedure will require considerable experimentation, but the prototype should be working by October 1976.

A GEOGRAPHIC OPERATING SYSTEM

The mechanical programming problems of building a large and complete software system are monumental. The design of software for this project has been made subservient to a very demanding list of criteria. GEOGRAF should address the broadest range of applications and should be applicable to very large data files. The package should be built to foster transfer of geographic data and software without imposing arbitrary standards.

To implement these requirements GEOGRAF is being developed in a subset of FORTRAN amenable to transfer to all large and medium scale computer systems. Also the more restricted memory sizes of mini-computers have not been ruled out. In order to accomodate almost unlimited numbers of geographic entities and to operate with reasonable efficiency, a virtual memory system will allow constant interaction between software and storage. The resulting GEOGRAF package will consist of a set of subroutines that provide a basic geographic operating system. Requests will be made in a natural geographic manner without concern for the precise method of storage, internal or external. We plan to develop a number of complete programs using this environment, but the operating system should be useful for other projects. While such a system may expand indefinitely, the target for first release is in 1977.

SUMMARY

By offering a complete topological data management system in a flexible computer environment, new avenues for spatial analysis can be opened. The connected structure of geographical entities should foster the development of analytical procedures sensitive to spatial interdependence, rather than statistical independence. Spatial data collected for mapping can provide the base for more rigorous spatial understanding. GEOGRAF and similar systems should significantly enhance the ability to manipulate urban and environmental information.

REFERENCES

1. Blumberg, M., and L. van de Velde, "Developing a Computerized Impact Zoning System," Urban and Regional Information Systems: Resources and Results, Volume II, Papers from the Twelfth Annual Conference of the Urban and Regional Information Systems Association, August 1974, Montreal, PQ, Canada (Pomona, N.J.: Stockton State College, 1975), pp. 340-352.

2. Chrisman, N., "The Impact of Data Structure on Geographic Information Processing," as discussed in Panel on Cartographic Data Structures, Proceedings of International Conference on Automation in Cartography, Reston, Virginia, December 1974, (Washington, D.C.: American Congress on Surveying and Mapping, 1976), pp. 165-181.

3. Cooke, Donald F., and William Maxfield, "The Development of a Geographic Base File and Its Use in Mapping," Urban and Regional Information Systems for Social Programs, Papers from the Fifth Annual Conference of the Urban and Regional Information Systems Association, Garden City, N.Y., September 1967, (Kent, Ohio: Center for Urban Regionalism, Kent State University, n.d.) pp. 207-218.

4. Laboratory for Computer Graphics, POLYVRT, A Program to Convert Geographic Base Files, (Cambridge, Massachusetts: Harvard University, 1974).

5. Peucker, T. and N. Chrisman, "Cartographic Data Structures," The American Cartographer, volume 2, number 1 (April 1975), pp. 55-69.