http://dx.doi.org/10.15576/GLL/2022.3.93

GLL

Geomatics, Landmanagement and Landscape No. 3 • 2022, 93–102

NETWORK ANALYSES WITH THE USE OF SPATIAL DATABASES

Szczepan Budkowski

Summary

An analysis is the process of browsing and searching for specific information from an entire dataset. The simplest analysis that can be performed on the data is visual analysis. However, it does not provide absolute certainty as to correctness and quality. A more advanced way of selecting required data is computer-based analysis. Analytical operations are performed on the data entered into the computer. The user defines the query, and the program performs calculations and displays the answer on the monitor screen. The aim of this publication is to conduct network analyses with the use of spatial databases. Besides focusing on the analysis as the leading research method, the paper also adopts this method to analyze the literature on the subject. In addition, the paper points to the complementary roles of the raster model and the vector model, emphasizing their coexistence.

The paper shows a variety of applications of GIS analyses, from simple buffers around selected areas, through selection, and the intersection of layers, to network analyses. The high degree of advancement of GIS tools allows to build advanced models in which analyses that go beyond the original application of the collected databases can be run.

Keywords

geographic information system • databases • network analysis • space management • raster data model • vector data model

1. Introduction – research issues

An analysis is the process of browsing and searching for specific information from an entire dataset. The simplest analysis that can be performed on the data is visual analysis. However, it does not provide absolute certainty as to correctness and quality [Fischer 2003]. A more advanced way of selecting required data is computer-based analysis. Analytical operations are performed on the data entered into the computer. The user defines the query, and the program performs calculations and displays the answer on the monitor screen [Dijkstra 1959].

However, analyses are applied to the input data that describe our reality. The development of both technology and the method of data presentation allowed for the development of specific spatial data models that represent objects in a specific way. The geometrical elements of a given model depend primarily on the criterion of the dimensions of these objects in space [De Smith et al. 2007]. Therefore, we can distinguish:

- point objects, e.g. geodetic network points,
- linear objects, e.g. fence lines,
- surface objects, e.g. parcels.

Objects that comprise the content of maps are presented in the form of raster and vector models. The raster model is used to collect and process data obtained from existing materials such as maps, remote sensing images, aerial photos, or satellite imagery. The real world data in this model is stored in the form of surface elements called pixels. Single pixels can have the shape of a square, less often a rectangle or a hexagon. The image formed by pixels is commonly called a raster [Fischer 2003].

In the vector model, the position of point objects is determined by the coordinates of the point locating a given object [Graser 2011]. For linear and surface objects defined by a larger number of points not only their coordinates are important, but also their appropriate arrangement, which allows to determine the shape of the object [Fischer 2003]. Having only a group of points without information on their arrangement, it is impossible to clearly define the shape of the object [Zhou et al. 2019]. The table (Table 1) compares the features of both models, which are crucial for the conducted analyses.

Table 1. Comparison of vector models according to their features

Characteristic	Vector model	Raster model
Data size	the data is saved very sparingly, takes up little space, there are only coordinates and identifiers of vector elements	very extensive and wasteful recording, image saved as a matrix, each pixel has its own value
Required hardware	low hardware requirements due to the simplicity of analysis and low memory requirement	high data storage requirements (capacious and fast disks); fast processors for raster analysis
Data acquisition	time-consuming and labor-intensive, each element must be properly processed	quick in comparison to the vector model
Acquisition cost	high; a complex process is required for each vector element, from data extraction to input and processing in a computer	low in relation to the cost of acquiring vector data
Data structures	complex	simple
Graphic depiction	quite good, even with large enlargement of elements it is possible to unambiguously identify	weak, dependent on the size of the field pixel, quality improvement has a negative impact on the size of the data, and this translates into the speed of analysis

Topology	simple	complicated
Folding data layers	poor	good
Generalization of maps	complex	simple
Geometric accuracy	high	short
Scale and resolution	Virtually any scale, and high resolution of the image	strongly limited by hardware capabilities (both at the time of data acquisition and processing)
Modeling of real objects	full; each vector element has an information base in the form of attribute tables, which can be enriched with the necessary data at any time	poor; an image is a collection of pixels in which the pixel does not represent a given object, so it is impossible to assign specific attributes to it

Source: Author's own study

Network Analysis is a special type of analysis. It requires a special data network, thanks to which it is possible, for example, to determine the travel time from one place to another, taking into account the time, date, and means of transport [Budkowski and Litwin 2019]. Network analyses are helpful in transport companies, although their application may be much wider and cover many issues related to public and social life [Curtin 2007], e.g.:

- travel time of public transport,
- time of arrival of ambulances to patients,
- availability of police stations in the area, etc.

This part of the work can be divided into three stages:

- data preparation,
- conducting analyses,
- presentation of the results.

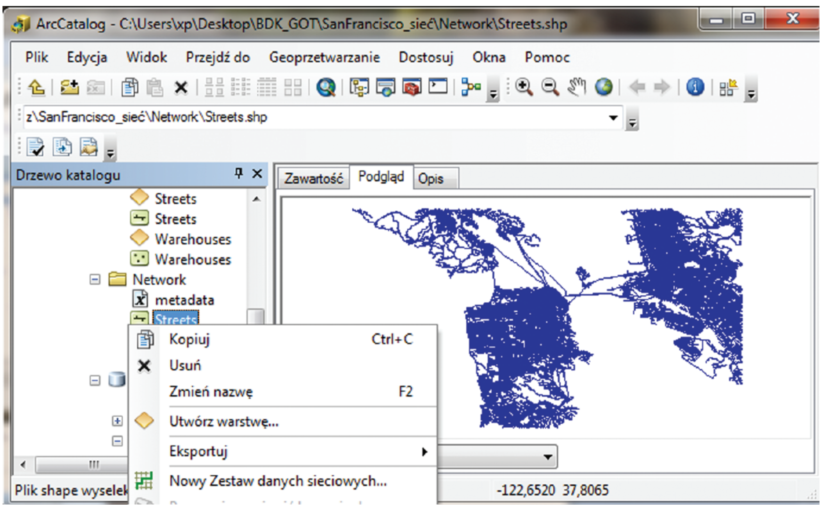
2. Conducted research

The research used the data available on the website of the manufacturer of the ArcGIS software. Such a selection of data was motivated by the availability of a large database containing detailed spatial information. The aim of the analysis was to find an area for the construction of a new warehouse in San Francisco. Data for network analyses was imported from GISTutorial [ESRI 1998], while data on land use and buildings from OpenStreetMap [Haklay 2008]. At the stage of data compilation, a height raster was also downloaded from the website.

Before starting the analysis, the following criteria were adopted:

- the area cannot be located in depression areas (below sea level),
- the area for development must be located at least 200 meters from the above-mentioned areas (completely outside the flood zones),
- must be at least 1000 meters from the parks,
- the property must be undeveloped,
- must be a maximum of 100 meters from roads,
- the area must be in the range of 100,000–200,000 m^2,
- the price of the localized area must be less than 100,000,
- additional assumptions:
  - the travel time from the new warehouse to the store should range from 2–7 minutes,
  - the new warehouse is to be located so as to serve at least 2 stores.

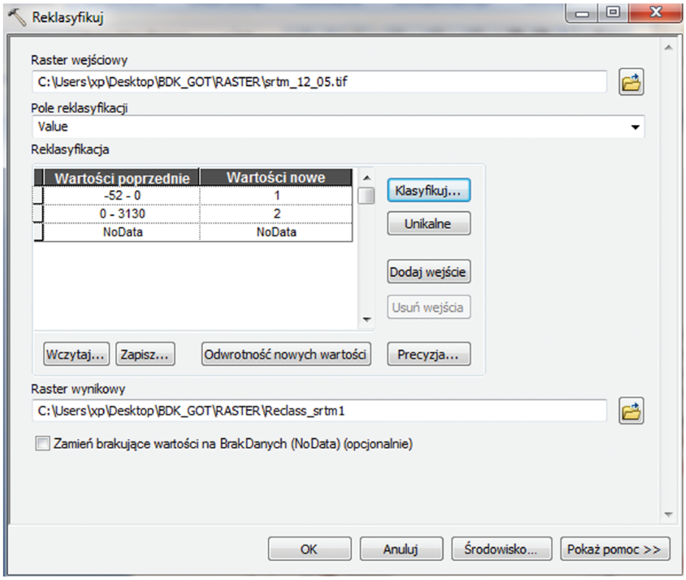
The first step was to create data that was then used for network analysis. Most of the optional solutions suggested by the program were adopted in the process of creating the network (Fig. 1).



Source: Author's own study

Fig. 1. Set of network data – preparation

The next step was to prepare the previously downloaded raster. For this purpose, the ArcToolbox tools were used, finally saving the new raster as a floating point and resampling with the cell size equal to 5. Then the raster was reclassified (Fig. 2). The purpose of this activity is to check the first criterion of the analysis. Then the reclassified raster was saved as a proving ground.



Source: Author's own study

Fig. 2. Raster reclassification

Data for three types of streets have been added to ArcMap 10.4: highways, major roads, and minor roads of local importance. Also added layers: parks and lakes. The ocean layer, as well as the layers with points for stores and wholesalers, were created by the user.

The first analysis that was carried out was to identify areas that are not suitable for the location of the warehouse. For this purpose, using the reclassified raster converted into a proving ground, the area below 0 meters above sea level was excluded, and a buffer of 200 meters was made from this area (Fig. 3).

Source: Author's own study

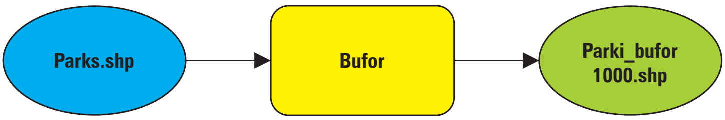
Fig. 3. Execution of a 200-meter buffer around depression areas

The next step was to select the areas that intersected with the buffer. This allowed to reject unsuitable areas, and by inverting the selection select areas for further analysis.

Based on the existing layers of parks and the vectorized park, a buffer was made from the fitted raster (Fig. 4, 5).

Source: Author's own study, raster with Google maps

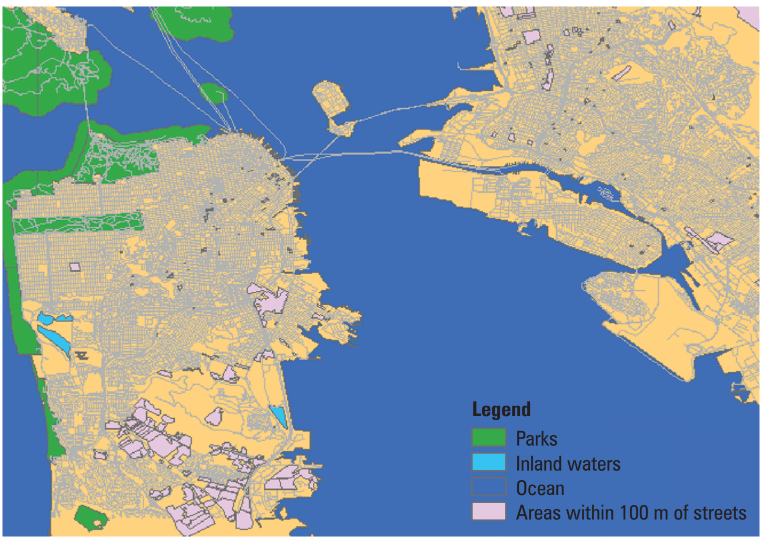
Fig. 4. Park from vectorization



Source: Author's own study

Fig. 5. Buffer from parks 1000 meters

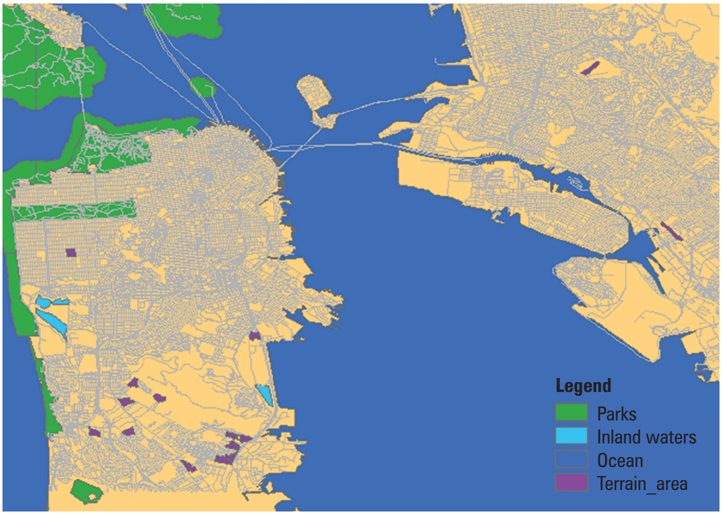
In the further part of the analysis, the existence of buildings in the analyzed areas was checked by making an intersection. All land that had any buildings on it was rejected. The others (Fig. 6) were subjected to another analysis.



Source: Author's own study

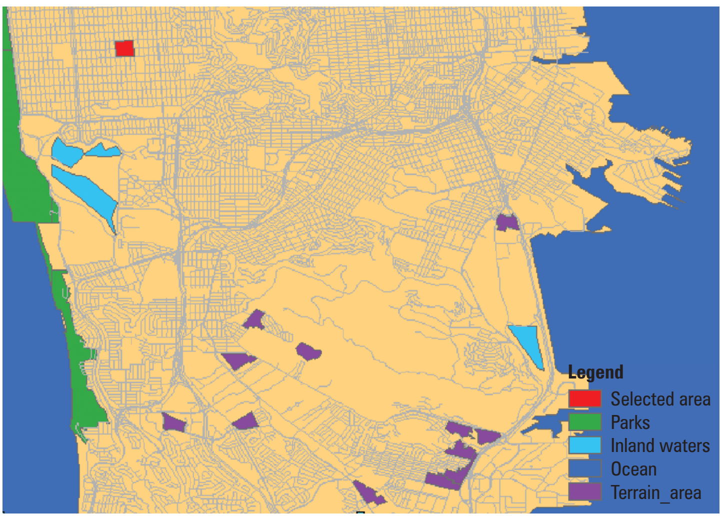
Fig. 6. Areas that meet the criterion of 100 meters from the streets, after applying all the other criteria

It turned out that each of the considered areas met the criterion of distance from roads. Then, the criterion of the size of the areas was checked by selecting the attributes. The Area column with the algorithm converting the area in square meters from the SHAPE_AREA field has been added to the table. The area criterion, i.e. the size of the area in the range of 100,000–200,000 m^2, was met by 19 areas (Fig. 7).



Source: Author's own study

Fig. 7. Areas that meet the area criterion, and the previous criteria

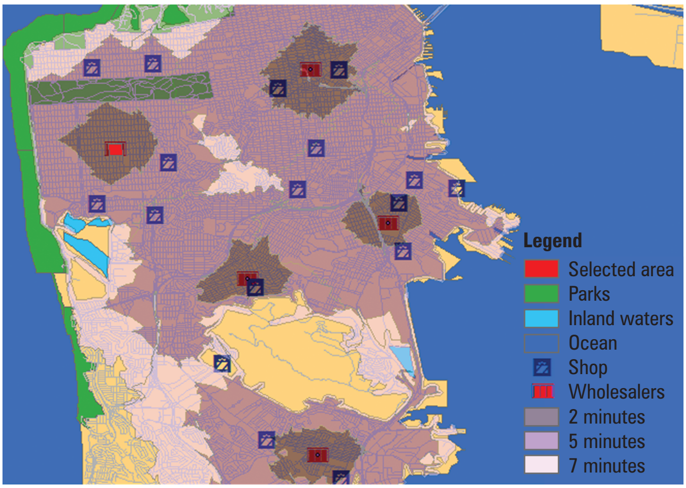


Source: Author's own study

Fig. 8. Area that meets the price criterion and all the previous assumptions

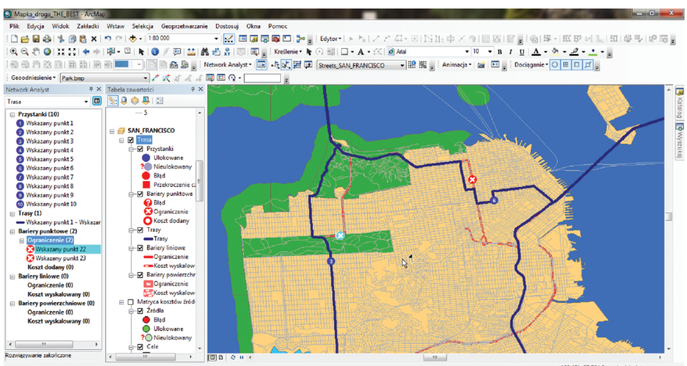
The next step was to add the price table to the attribute table and perform the appropriate analysis. The analysis indicated one area that meets all the selected conditions (Fig. 8).

Therefore, an additional criterion remained to be checked – the travel time in the 2, 5 and 7-minute intervals, as well as the criterion regarding the number of covered stores. For this purpose, network analyses were used, specifically relying on the allocation function (Fig. 9). In the analysis settings, the travel time was selected as 2, 5, or 7 minutes. As can be seen, then polygons were created, which made it possible to divide into specific areas. The area, therefore, meets the requirements for the number of covered stores, as well as the time of arrival.



Source: Author's own study

Fig. 9. Areas covered by wholesalers at a given time



Source: Author's own study

Fig. 10. Network analysis of access roads with a blockade

Additionally, network analyses were used again to simulate a failure on the originally selected route between wholesalers (Fig. 10). Thanks to the extensive database of travel times and directions in the streets, the program quickly found alternative solutions.

3. Summary and conclusions

The conducted analyses show explicitly that both models (raster and vector) are needed. Each of them has both advantages and disadvantages. And on this basis, it is important to make the right choices when it comes to achieving the intended goal. In the case of analyses carried out by the author, the raster is used to determine the height of the terrain. It is an excellent complementary function at the moment when the analysis condition appeared (buffer from parks). The vector model is particularly useful here, and without it the entire analysis would not have been possible. The following advantages are definitely present here:

a) the ability to save a large amount of data in the attribute table,
b) relatively small size of vector files used in the analysis,
c) accuracy and precision at close-ups of small streets.

The article shows the variety of applications of GIS analyses, from simple buffers around selected areas, through selection ant the intersections of layers, to network analyses. The high degree of advancement of GIS tools allows to build advanced models within which analyses that go beyond the original application of the collected databases can be conducted. Having accurate data on cubature objects, it is also possible to apply network analyses to the interiors of these objects. However, this requires further research in the field of efficiency in space management with the use of network analyses. The presented thematic proposal covers only a small range of all problems that can be solved with the use of Geographic Information Systems (GIS) and network analyses.

Funded by a subsidy of the Ministry of Education and Science for the Hugo Kołłątaj University of Agriculture in Krakow for 2022.

References

Budkowski S., Litwin U. 2019. Synergy in space. The use of GIS tools on the example of premises records in the LoD4 model. Acta Scientiarum Polonorum, Formatio Circumiectus, 18(1), 127–135.

Curtin K.M. 2007. Network Analysis in Geographic Information Science: Review, Assessment, and Projections. Cartography and Geographic Information Science, 34, 2.

De Smith M.J., Goodchild M.F., Longley P.A. 2007. Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. 2nd ed. Troubador Publishing.

Dijkstra E.W. 1959. A note on two problems in connexion with graphs. Numerische Mathematik, 1.

ESRI Shapefile Technical Description. 1998. An ESRI White Paper, Environmental Systems Research Institute, Redlands.

Fischer M.M. 2003. GIS and Network Analysis. ERSA Conference Papers. 43rd European Congress of the European Regional Science Association. Jyväskylä, Finland.

Graser A. 2011. A Beginner's Guide to pgRouting. Free and Open Source GIS Ramblings. http://underdark.wordpress.com/2011/02/07/a-beginners-guide-to-pgrouting/

Haklay M., Weber P. 2008. OpenStreetMap: User-Generated Street Maps. IEEE Pervasive Computing, 12–18.

Zhou G., Soufan O., Ewald J., Xia J. 2019. NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Research.

Dr inż. Szczepan Budkowski
University of Agriculture in Krakow
Department of Geodesy
30-198 Kraków, ul. Balicka 253a
e-mail: szbudkowski@o2.pl
ORCID: 000-0002-1806-1173