GEODEZJA I KARTOGRAFIA
GEODESY AND CARTOGRAPHY
Vol. 55, No 3, 2006, pp. 147-167

Area neighbourhood models

Elzbieta Lewandowicz

Department of Surveying, University of Warmia and Mazury in Olsztyn
12 Jana Heweliusza St., 10-957 Olsztyn, Poland
e-mail: leela@uwm.edu.pl

Received: 22 February 2006/Accepted: 21 September 2006

Abstract: In GIS systems, the neighbourhood of areas within an analysed region is a term applied usually to raster-screened data. The aim of the study was to adapt this term to vector and descriptive data as well as to systemize models of so defined neighbourhood. The starting point was the assumption that the basic area neighbourhood model may be based on spatial data illustrated with a graph and described with a neighbourhood matrix. It provides the basis for building subsequent models, that are linked with the introduction of new neighbourhood measures, i.e. measures resulting from the characteristics of areas entered in tables of their attributes. Based on the proposed models, spatial analysis related to area neighbourhood can be performed and aggregate models, considered essential in multidimensional analysis of neighbourhood can be developed.

Keywords: Area neighbourhood, neighbourhood model, neighbourhood measure, neighbourhood matrix, spatial analysis

1. Introduction

A significant factor contributing to GIS development may be any new approach of spatial analysis. The paper focuses on the problem of analysis of the functions of areas making up a given geographical region.

In most GIS systems, an analysis that is based on vector data is included in a network analysis package. In such case the analysis of area neighbourhood consists in finding bordering on the network areas with given parameters. They are determined from the spatial bordering condition – contiguity (Ahuja et al., 1993; Zhan, 1998).

The neighbourhood can also be taken into account by selecting desired objects through search within a given range from a specified centre or by entering buffers within a given distance from the boundaries of an object (Autodesk, 2000; Moore, 2000; Bentley, 2002; ESRI, 2005). Numerous examples of such analysis are available in the literature (e.g. Caruso et al., 2005; Sadahiro, 2005).

It is difficult for GIS software programmers to formally describe a broadly understood neighbourhood of areas within a region. The formal description proposed in the paper may provide the basis for solving numerous analytical problems related to the location of new objects, land appraisal or the assessment of planned economic tasks.

The basis for the new approach is an area neighbourhood model presented in the form of a graph (Molenaar, 1998; Longley et al., 2001). That neighbourhood model can be described by many metrics and measures (Longley at al., 2001; Sullvan and Unwin, 2002; Lewandowicz, 2005), as well as modified (Fischer et al., 1996; Bera and Claramunt, 2003). The purpose of this study was to develop new forms of that model, resulting from the geometric and topological data of areas. The descriptive data characterizing areas, specified in tables of attributes, were used for a thematic separation of neighbourhood relations.

The paper provides the formal description of those issues, referring to the fundamentals of creating mathematical neighbourhood models. A new systematics of neighbourhood forms and respective neighbourhood models are presented. Examples are also provided to illustrate the use of the developed neighbourhood models.

Multivariate analysis, enabling to estimate the collective impact of various descriptive data, requires aggregate neighbourhood models. Building aggregate models must be preceded by the normalization of thematic neighbourhood models. The paper presents the appropriate procedure.

2. The concept of area neighbourhood

Assume that the region F is divided into n areas f_i, where $i = \{1, 2, \ldots, n\}$. The areas border on each other, and their total surface area equals the surface area of the region. In the literature (Molenaar, 1998; Longley et al., 2001), f_i and f_j areas are referred to as neighbouring areas if they have a common border e_{ij}. The bordering of two areas f_i, f_j is identified with 1, while no bordering is identified with 0. This can be presented as follows:

$$f_i \in F: \quad f_1 \cap f_2 \neq f_i \quad i = \{1, 2, 3, ..., n\} \tag{1}$$

$$[f_1, f_2 / e_{12}] = 1 \quad \text{when} \quad e_{12} = f_1 \cap f_2 \tag{2}$$

$$[f_1, f_2 / e_{12}] = 0 \quad \text{when} \quad f_1 \cap f_2 = \varnothing \tag{3}$$

Area neighbourhood may also be presented in the form of a geometric graph (Molenaar, 1998; Longley et al., 2001). That graph is easily generated by transforming geometric data specifying area boundaries on a digital map (Lewandowicz, 2004; Lewandowicz and Bałandynowicz, 2005). It is called a dual graph (Kulikowski, 1986; Wróblewski, 1997; Wilson, 2000). The neighbourhood matrix **S** for that graph corresponds the matrix form of spatial relations between areas f_i. Numbers $\{1, 2, \ldots, n\}$ of rows and columns of neighbourhood matrix **S** are also identifiers of areas of the examined region. Elements of the matrix **S** are denoted as s_{ij}. In the classic description of neighbourhood, elements s_{ij} take on two values: 1 – if areas f_i and f_j border on each other, and 0 if they do not.

Let us adopt the area neighbourhood graph in the region as the base model used to build derivative models. The starting point is the representation of this graph in the form of **S** matrix.

The essence of neighbourhood models is well illustrated by the example shown in Figure 1. The considered region was divided into areas so as to present many neighbourhood situations: diagonal neighbourhood (f_4, f_6), (f_3, f_5), so-called islands (f_7) and the isolation of areas (f_1).

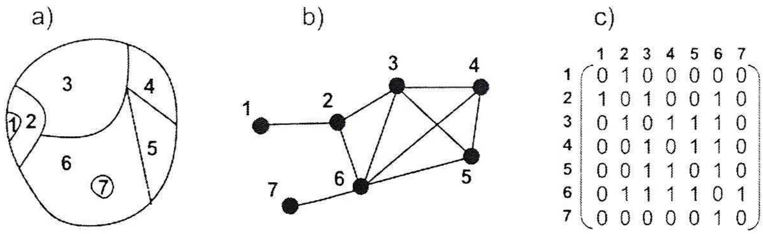


Fig. 1. Areas in the region a), area neighbourhood graph b), area neighbourhood matrix **S** c)

The classic definition of area neighbourhood may be developed by distinguishing the following forms of neighbourhood:

- direct,
    - direct dependent,
- indirect of degree q,
- indirect of degree not higher than q,
    - indirect of degree not higher than q with suppression,
- full with suppression.

The neighbourhood specified in literature (Molenaar, 1998; Longley et al., 2001) as well as presented above is a direct neighbourhood. The direct dependent neighbourhood takes into account the neighbours of neighbours in direct neighbourhood, while the indirect neighbourhood takes into consideration distant neighbours, not having a common border. Generally, neighbourhood of a specified degree as well as neighbourhood of degree not higher than q, which aggregates selected neighbours were also distinguished. Full neighbourhood describes the neighbourhood relations of each area with every other area.

After distinguishing such forms of neighbourhood, one can identify the following neighbours: direct, indirect and neighbours exhibiting neighbourhood of a given degree. A detailed description of the adopted neighbourhood classification is specified further on in the paper.

Each form of neighbourhood may be defined basing only on geometric relations, or considering also the thematic features of the neighbourhood. Taking into account thematic features in practice is possible only when the areas in GIS databases possess descriptive attributes t^{td}:

$$f_i \in F : f_i \rightarrow \left[t_i^1, t_i^2, \ldots, t_i^k\right] \tag{4}$$

Some of those attributes t^{td} ($td = \{1, 2, \ldots, k\}$), representing thematic descriptions, are significant from the point of view of the analysed problem for the region and should be applied by building thematic neighbourhood models.

Direct neighbourhood relations take place for areas that have a common border. The graph of area neighbourhood in the region (Fig. 1) represented in the form of **S** matrix, assesses the existence of direct neighbourhood with the use of values 0, 1. Matrix element s_{ij} is assigned 1, if areas f_i and f_j possess a common border $e_{i,j}$; 0 of the element s_{ij} indicates no neighbourhood between areas f_i and f_j. The values of s_{ij} describe the measures $|\mathbf{S}_{f_i f_j}|$ of the classic neighbourhood of areas f_i and f_j, which can be specified as

$$\bigcap_{i \in \{1,2,\ldots,n\}} \bigcap_{j \in \{1,2,\ldots,n\}} |\mathbf{S}_{f_i f_j}| \in \{0, 1\} \tag{5}$$

and presented in the form of a graph as illustrated in Figure 1. A quantum metric description of neighbourhood was adopted in this neighbourhood entry. It can be replaced with an Euclidean or weighted metric, as well as described with other thematic measures (Lewandowicz, 2005). Those measures may result from thematic descriptions of areas (4).

3. Direct neighbourhood taking into account thematic consequences of neighbourhood

The direct neighbourhood model described by the matrix **S**, may be transformed to various thematic models $\mathbf{S}^{td}$, $td = \{1, 2, \ldots, k\}$ of direct neighbourhood. Let us assume that the attribute t^{td} represents a significant feature in the description of the neighbourhood of the examined areas. The values of those attributes may be used to specify the impact of an attribute t_i^{td} of area f_i on neighbouring areas. By taking into account the values of that attribute in the neighbourhood model **S**, one acquires a thematic direct neighbourhood model $\mathbf{S}^{td}$, that can no longer be illustrated by a simple graph, since it is an asymmetric model. It can be presented in the form of a digraph, in this case, the digraph $\mathbf{S}^{td}$. The transformation

$$\mathbf{S} \rightarrow \mathbf{S}^{td} \tag{6}$$

can be specified in the form:

$$\mathbf{S}^{td} = \mathbf{T}^{td}\mathbf{S} \tag{7}$$

where **S** is the neighbourhood matrix, and $\mathbf{T}^{td}$ is the diagonal matrix, comprising on diagonal t_{ii}^{td} the values of an attribute t^{td} characterizing areas f_i, $i = \{1, 2, \ldots, n\}$. The indices of columns and rows of the matrix $\mathbf{T}^{td}$ correspond to those of the matrix **S**.

Product $\mathbf{S}^{td}$ (7) describes the thematic model of direct neighbourhood in the region. In this model, the measure of neighbourhood are the values t_i^{td} of the examined attribute. It is an asymmetric model, since

$$\left|\mathbf{S}^{td}_{f_i f_j}\right| \neq \left|\mathbf{S}^{td}_{f_j f_i}\right| \tag{8}$$

$$\left|\mathbf{S}^{td}_{f_i f_j}\right| = t^{td}_{ii} \text{ for } \bigcap_{i\in\{1,2,\ldots,n\}} \bigcap_{j\in\{1,2,\ldots,n\}} \left|\mathbf{S}_{f_i f_j}\right| = 1 \tag{9}$$

Let us notice (Fig. 2) that by presenting graphically $\mathbf{S}^{td}$ model with the use of a digraph, the directed edges are respectively weighted.

To illustrate such a model (Fig. 2), examples of numerical values of the descriptive attribute t_i^{td}, $i = \{1, 2, \ldots, n\}$ for the examined areas, were adopted for the region in Figure 1a, respectively: 2_1, 8_2, 18_3, 11_4, 10_5, 15_6, 4_7. They can describe, e.g. the value of land, number of students, schools, number of unemployed or the income of an administrative unit managing the area.

$$\begin{array}{c}\mathbf{T}^{td}\\ \begin{array}{ccccccc}1&2&3&4&5&6&7\end{array}\\ \begin{pmatrix}2&0&0&0&0&0&0\\0&8&0&0&0&0&0\\0&0&18&0&0&0&0\\0&0&0&11&0&0&0\\0&0&0&0&10&0&0\\0&0&0&0&0&15&0\\0&0&0&0&0&0&4\end{pmatrix}\end{array} * \begin{array}{c}\mathbf{S}\\ \begin{array}{ccccccc}1&2&3&4&5&6&7\end{array}\\ \begin{pmatrix}0&1&0&0&0&0&0\\1&0&1&0&0&1&0\\0&1&0&1&1&1&0\\0&0&1&0&1&1&0\\0&0&1&1&0&1&0\\0&1&1&1&1&0&1\\0&0&0&0&0&1&0\end{pmatrix}\end{array} = \begin{array}{c}\mathbf{S}^{td}\\ \begin{array}{ccccccc}1&2&3&4&5&6&7\end{array}\\ \begin{pmatrix}0&2&0&0&0&0&0\\8&0&8&0&0&8&0\\0&18&0&18&18&18&0\\0&0&11&0&11&11&0\\0&0&10&10&0&10&0\\0&15&15&15&15&0&15\\0&0&0&0&0&4&0\end{pmatrix}\end{array}$$

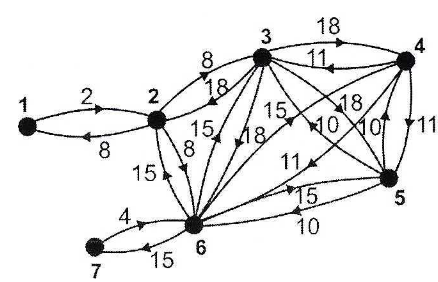


Fig. 2. Thematic model $\mathbf{S}^{td}$ of direct neighbourhood, in the matrix form and represented graphically

Model $\mathbf{S}^{td}$ is described with a matrix whose elements s_{ij}^{td} correspond to

$$s_{ij}^{td} = \begin{cases} s_{ij} * t_{ii}^{td} & \text{for } s_{ij} = 1 \\ 0 & \text{for } s_{ij} = 0 \end{cases} \tag{10}$$

Column j of the matrix $\mathbf{S}^{td}$ includes the values of the thematic impact of neighbouring areas on area f_j, resulting from attribute t^{td}. For every area f_j, a respective column in the form of the vector $\mathbf{S}^{td}_{f_j}$ can be distinguished

$$\bigcup_{td\in\{1,2,\dots,k\}} \bigcap_{j\in\{1,2,\dots,n\}} \mathbf{S}_{f_j}^{td} = [s_{1j}^{td}, s_{2j}^{td}, \dots, s_{nj}^{td}] \tag{11}$$

Based on the matrix $\mathbf{S}^{td}$, the thematic value of full direct neighbourhood of region areas can also be specified in the form of the following vector:

$$(\mathbf{SC})^{td} = [(\mathbf{SC})_{f_1}^{td}, (\mathbf{SC})_{f_2}^{td}, (\mathbf{SC})_{f_3}^{td}, \dots, (\mathbf{SC})_{f_n}^{td}] \tag{12}$$

where

$$\bigcup_{td\in\{1,2,\dots,k\}} \bigcap_{j\in\{1,2,\dots,n\}} (\mathbf{SC})_{f_j}^{td} = \sum_{i=\{1,2,3,\dots,n\}} s_{ij}^{td} \tag{13}$$

The vector $(\mathbf{SC})^{td}$ consists of the sums of values of elements in columns of the matrix $\mathbf{S}^{td}$.

The values of components of $\mathbf{S}_{f_j}^{td}$ and $(\mathbf{SC})^{td}$ vectors may easily be used to analyse neighbourhood in the region. In the example considered, the components of the vector $(\mathbf{SC})^{td}$ are as follows

$$(\mathbf{SC})^{td} = [8_1, 35_2, 44_3, 43_4, 44_5, 51_6, 15_7]$$

The largest value of the vector $(\mathbf{SC})^{td}$ component corresponds to area f_6, while the smallest one – to area f_1. Total value of a feature t^{td} of neighbouring areas is thus the largest for area f_6 and the smallest for area f_1. Assuming that t_i^{td} describes the values of areas f_i, the values of the components of the vector $(\mathbf{SC})^{td}$ exhibit the possible increase in value of area f_j if neighbouring areas were merged with it. Similarly, those values may indicate an increase in the number of schools in area f_i following a merging with neighbouring areas. When adopting t_i^{td} as the number of unemployed in area f_i, the values of the elements of the vector $(\mathbf{SC})^{td}$ indicate the value of available human workforce in the neighbourhood of each area.

3.1. Region model taking into account thematic features of areas and the impact of direct neighbourhood

The vectors $\mathbf{S}_{f_i}^{td}$, and $(\mathbf{SC})^{td}$ describe the thematic direct neighbourhood of areas. In tasks related to the neighbourhood analysis of a region, there is a need to sum the values of a selected attribute of a given area as well as neighbouring areas. In such cases, the neighbourhood model $\mathbf{S}^{td}$ described above should be modified. The new model was assigned the $(\mathbf{ND})^{td}$ symbol and presented as the transformation

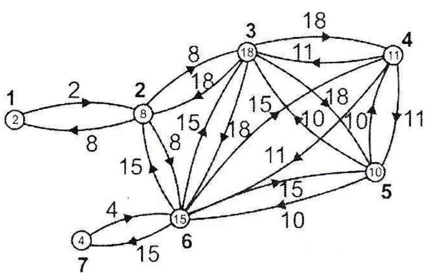
$$\mathbf{S}^{td} \rightarrow (\mathbf{ND})^{td} \tag{14}$$

expressed as follows

$$(\mathbf{ND})^{td} = \mathbf{S}^{td} \cup \mathbf{T}^{td} = (\mathbf{T}^{td}\mathbf{S}) \cup \mathbf{T}^{td} \tag{15}$$

where $(\mathbf{ND})^{td}$ is the thematic model supplementing the feature of each area with the impact of neighbours, $\mathbf{S}^{td}$ is the thematic neighbourhood model, $\mathbf{S}$ is the base neighbourhood model, and $\mathbf{T}^{td}$ is the matrix of the value of attribute t^{td} for the areas.

The $(\mathbf{ND})^{td}$ model may be represented graphically in the form of a weighted digraph and a matrix, as in Figure 3. In this case, the weights are assigned not only to directed edges but also the nodes.



$$(\mathbf{ND})^{td} = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 8 & 8 & 8 & 0 & 0 & 8 & 0 \\ 0 & 18 & 18 & 18 & 18 & 18 & 0 \\ 0 & 0 & 11 & 11 & 11 & 11 & 0 \\ 0 & 0 & 10 & 10 & 10 & 10 & 0 \\ 0 & 15 & 15 & 15 & 15 & 15 & 15 \\ 0 & 0 & 0 & 0 & 0 & 4 & 4 \end{pmatrix} \\ \begin{array}{cccccccc} \Sigma & 10 & 43 & 62 & 54 & 54 & 66 & 19 \end{array} \end{array}$$

Fig. 3. Thematic model of a region taking into account the impact of direct neighbourhood $(\mathbf{ND})^{td}$ in the graphic and matrix form

Vectors corresponding to columns of the matrix $(\mathbf{ND})^{td}$, include values of attribute t^{td} for area f_i as well as areas directly neighbouring f_i.

$$\bigcup_{td\in\{1,2,\ldots,k\}} \bigcap_{j\in\{1,2,\ldots,n\}} (\mathbf{ND})^{td}_{f_j} = [(nd)_{1j},(nd)_{2j},\ldots,(nd)_{nj}] \tag{16}$$

The vector $(\mathbf{NDC})^{td}$ includes the sums of attributes t^{td} of the areas in the region and their neighbours. It can also be acquired from $(\mathbf{SC})^{td}$

$$(\mathbf{NDC})^{td} = [(\mathbf{NDC})^{td}_{f_1},(\mathbf{NDC})^{td}_{f_2},(\mathbf{NDC})^{td}_{f_3},\ldots,(\mathbf{NDC})^{td}_{f_n}] = (\mathbf{SC})^{td} + diagi(\mathbf{T}^{td}) \tag{17}$$

where $diagi(\mathbf{T}^{td})$ is a vector of elements equal to the sums of elements in corresponding columns of the matrix $\mathbf{T}^{td}$, and

$$\bigcap_{j=\{1,2,\ldots,n\}} (\mathbf{NDC})^{td}_{f_j} = \sum_{i=\{1,2,\ldots,n\}} (nd)^{td}_{ij} \tag{18}$$

The vector $(\mathbf{NDC})^{td}$ in the example shown in Figure 3 is

$$(\mathbf{NDC})^{td} = [10_1, 43_2, 62_3, 54_4, 54_5, 66_6, 19_7]$$

The largest value of a component of $(\mathbf{NDC})^{td}$ points to area f_6. The total value of the examined feature of the area as well as the value of neighbouring areas is the highest for area f_6 and the lowest for area f_1.

Assuming in the example that the values of a feature t^{td} represent the number of unemployed residents in the area, then with the help of the above algorithm one can estimate the number of persons who will be offered work by locating a new investment in the area corresponding to a component of the vector $(\mathbf{NDC})^{td}$. That number is equal to the value of this component. According to the values of these components, the highest number of unemployed residents, as many as 66 in area f_6 and neighbouring areas, will be offered work by locating a new investment in that area.

In the neighbourhood models $(\mathbf{ND})^{td}$ and $(\mathbf{NDC})^{td}$ presented above, original values of area attributes t_n^{td} are taken into account, i.e. non-normalised values of neighbourhood. They are significant in most simple types of analysis.

4. Direct dependent neighbourhood

Direct neighbourhood was described above with a simple model $\mathbf{S}$. Thematic neighbourhood model $\mathbf{S}^{td}$ was the result of the introduction to $\mathbf{S}$ model of a selected descriptive attribute t^{td} for areas of the entire region. In practice, neighbourhood is often dependent not only on the thematic description but also on certain spatial factors. Let us try to define a neighbourhood model that will describe the sharing of features t^{td} of area f_i with its neighbours. The value of a feature depends on the factors related to the spatial characteristics of the neighbourhood, e.g. with the number of neighbours, length of the boundary line and surface area of neighbouring regions. Selected spatial features of neighbourhood affect the construction of the neighbourhood model and have a significant impact on the value of thematic neighbourhood descriptions.

Direct dependent neighbourhood takes into account the spatial characteristics of the neighbours of neighbours in direct neighbourhood. New direct dependent neighbourhood measures $\left|\mathbf{S}^{x}_{z_{f_i,f_j}}\right|$, may replace value 1 in $\mathbf{S}$ model. Let us assume that those measures are dependent on $x = \{nn, lbl, sur, \ldots\}$, i.e.

- number of neighbours (nn) – $\left|\mathbf{S}^{nn}_{z_{f_i,f_j}}\right|$
- length of boundary line (lbl) – $\left|\mathbf{S}^{lbl}_{z_{f_i,f_j}}\right|$
- surface area (sur) – $\left|\mathbf{S}^{sur}_{z_{f_i f_j}}\right|$
- other features.

Let us consider the case in which the area neighbourhood measure $\left|\mathbf{S}^{x}_{z_{f_i,f_j}}\right|$ is a number dependent on the number $(nn)_i$ of f_i area's direct neighbours. The $(nn)_i$ number of direct neighbours of every f_i area can be specified based on $\mathbf{S}$ model. Those numbers correspond to the sums of elements in rows (columns) of the matrix $\mathbf{S}$. They are equal to graph node degrees described by the matrix $\mathbf{S}$, and can be obtained with the use of the following formula

$$\bigcap_{i \in \{1,2,\ldots,n\}} (nn)_i = \sum_{j=\{1,2,\ldots,n\}} s_{ij} \tag{19}$$

The number of neighbours is important when examining the impact of each area on neighbouring areas. That impact will be larger if the number of neighbours is lower. The inverse of $(nn)_i$ can thus be adopted as a measure of impact of area f_i on direct neighbours. The measure of dependent neighbourhood of two areas $\left|\mathbf{S}^{nn}_{z_{f_i \cdot f_j}}\right|$ may be expressed by the following formula:

$$\left|\mathbf{S}^{nn}_{z_{f_i \cdot f_j}}\right| = \frac{1}{(nn)_i} \text{ for } \bigcap_{i \in \{1,2,\ldots,n\}} \; \bigcap_{j \in \{1,2,\ldots,n\}} \left|\mathbf{S}_{f_i f_j}\right| = 1 \tag{20}$$

Entered in **S** model, those measures transform it into $\mathbf{S}^{nn}_z$ model, i.e. neighbourhood dependent on the number of neighbours

$$\mathbf{S} \rightarrow \mathbf{S}^{nn}_z \tag{21}$$

The matrix $\mathbf{S}^{nn}_z$ may be generated with the following transformation

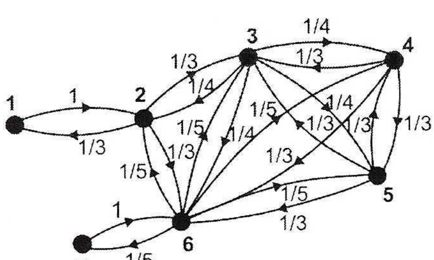
$$\mathbf{S}^{nn}_z = \mathbf{T}^{nn}\mathbf{S} \tag{22}$$

where values other than zero of diagonal matrix $\mathbf{T}^{nn}$ are respectively $t_{ii} = \dfrac{1}{(nn)_i}$.

The elements of the matrix $\mathbf{S}^{nn}_z$ will equal to:

$$(s^{nn}_z)_{ij} = \begin{cases} s_{ij}/(nn)_i & \text{for } s_{ij} = 1 \\ 0 & \text{for } s_{ij} = 0 \end{cases} \tag{23}$$

An example of neighbourhood model $\mathbf{S}^{nn}_z$ is based on base model **S** (Fig. 1); it is represented graphically with a digraph as well as in the matrix form in Figure 4.



$$\mathbf{S}^{nn}_z = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0 & 0.33 & 0 \\ 0 & 0.25 & 0 & 0.25 & 0.25 & 0.25 & 0 \\ 0 & 0 & 0.33 & 0 & 0.33 & 0.33 & 0 \\ 0 & 0 & 0.33 & 0.33 & 0 & 0.33 & 0 \\ 0 & 0.20 & 0.20 & 0.20 & 0.20 & 0 & 0.20 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Fig. 4. Direct dependent neighbourhood model $\mathbf{S}^{nn}_z$ taking into account the number of neighbouring areas (nn)

By multiplying the matrix $\mathbf{S}^{nn}_z$ by the matrix of thematic characteristics of examined areas $\mathbf{T}^{td}$ (7), a subsequent thematic neighbourhood model $(\mathbf{S}^{nn}_z)^{td}$ is generated. The vector $(\mathbf{S}_z^{nn}\mathbf{C})^{td}$, describing complete area neighbourhood (Fig. 5), is based on $(\mathbf{S}_z^{nn})^{td}$ model.

$$(\mathbf{S}_z^{nn})^{td} = \begin{array}{c} \\ \left(\begin{array}{ccccccc} 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2.7 & 0 & 2.7 & 0 & 0 & 2.7 & 0 \\ 0 & 4.5 & 0 & 4.5 & 4.5 & 4.5 & 0 \\ 0 & 0 & 3.7 & 0 & 3.7 & 3.7 & 0 \\ 0 & 0 & 3.3 & 3.3 & 0 & 3.3 & 0 \\ 0 & 3.0 & 3.0 & 3.0 & 3.0 & 0 & 3.0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 \end{array}\right) \\ \Sigma \;\; \begin{array}{ccccccc} 2.7 & 9.5 & 12.7 & 10.8 & 11.2 & 18.2 & 3.0 \end{array} \end{array} \begin{array}{c} \Sigma \\ 2 \\ 8 \\ 18 \\ 11 \\ 10 \\ 15 \\ 4 \\ \\ \end{array}$$

$$(\mathbf{ND}_z^{nn})^{td} = \begin{array}{c} \left(\begin{array}{ccccccc} 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2.7 & 8 & 2.7 & 0 & 0 & 2.7 & 0 \\ 0 & 4.5 & 18 & 4.5 & 4.5 & 4.5 & 0 \\ 0 & 0 & 3.7 & 11 & 3.7 & 3.7 & 0 \\ 0 & 0 & 3.3 & 3.3 & 10 & 3.3 & 0 \\ 0 & 3.0 & 3.0 & 3.0 & 3.0 & 15 & 3.0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 4 \end{array}\right) \\ \Sigma \;\; \begin{array}{ccccccc} 4.7 & 17.5 & 30.7 & 21.8 & 21.2 & 33.2 & 7.0 \end{array} \end{array}$$

Fig. 5. Matrices $(\mathbf{S}_z^{nn})^{td}$, $(\mathbf{ND}_z^{nn})^{td}$ as the basis for generating vectors $(\mathbf{S}_z^{nn}\mathbf{C})^{td}$, $(\mathbf{ND}_z^{nn}\mathbf{C})^{td}$

In the example shown in Figure 5, magnitudes of $(\mathbf{S}_z^{nn}\mathbf{C})^{td}$ vector components are smaller than those of the vector $(\mathbf{SC})^{td}$, provided in the example in Figure 3. This is due to the fact that the neighbourhood feature was divided by the number of the neighbour's neighbours. The magnitudes of $(\mathbf{S}_z^{nn}\mathbf{C})^{td}$ vector components, obtained basing on the matrix $(\mathbf{S}_z^{nn})^{td}$, after rounding up to integers, are respectively:

$$(\mathbf{S}_z^{nn}\mathbf{C})^{td} = [3_1,\ 10_2,\ 13_3,\ 11_4,\ 11_5,\ 18_6,\ 3_7].$$

Analogically, the magnitudes of vector $(\mathbf{ND}_z^{nn}\mathbf{C})^{td}$ components, specified based on $(\mathbf{ND}_z^{nn})^{td}$ amount to

$$(\mathbf{ND}_z^{nn}\mathbf{C})^{td} = [5_1,\ 18_2,\ 31_3,\ 22_4,\ 21_5,\ 33_6,\ 7_7]$$

Dependent neighbourhood model $\mathbf{S}_z^{lbl}$ may be built in a similar way, taking into account the length of boundary lines (lbl). The length of the boundary line between neighbouring areas is important when examining the impact of each area on neighbouring areas. The impact will be larger if the length of the common border is greater. In that case, the neighbourhood measure dependent on the length of the boundary line can be expressed as follows

$$\left|\mathbf{S}_{z_{f_i,f_j}}^{lbl}\right| = \frac{(lbl)_{i,j}}{(lbl)_i} \tag{24}$$

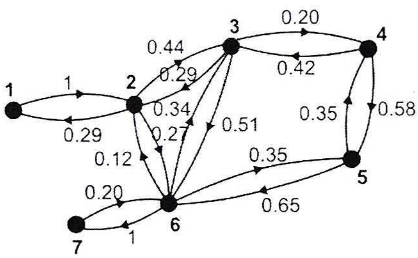
where $\left|\mathbf{S}_{z_{f_i,f_j}}^{lbl}\right|$ is the measure of f_i f_j area neighbourhood, determined based on the length of boundary lines; $(lbl)_{i,j}$ is the length of the boundary line between areas f_i,

f_j, $(lbl)_{i,j} = |e_{i,j}| = |f_i \cap f_j|$; $(lbl)_i$ is the length of the boundary line between area f_i and neighbouring areas in region F; $(lbl)_i = \sum_{j=1,2,\dots,n} e_{ij}$.

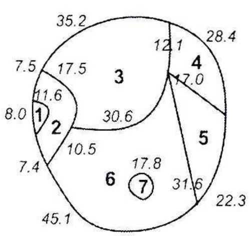
The matrix of dependent neighbourhood $\mathbf{S}_z^{lbl}$ will be composed of the following elements

$$(s_z)_{ij}^{lbl} = \begin{cases} s_{ij} * \left(\dfrac{(lbl)_{ij}}{(lbl)_i}\right) & \text{for } \; s_{ij} = 1 \\ 0 & \text{for } \; s_{ij} = 0 \end{cases} \tag{25}$$

According to that model, the neighbourhood of neighbouring areas f_3 and f_5 as well as f_4 and f_6 will be equal to 0. The boundary lines of those areas meet in one point only, so the length of common boundary lines $(lbl)_{3,5}$ and $(lbl)_{4,6}$ equal 0. In this case, the neighbourhood model (Fig. 6) is simplified.



$$\mathbf{S}_z^{lbl} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.29 & 0 & 0.44 & 0 & 0 & 0.27 & 0 \\ 0 & 0.29 & 0 & 0.20 & 0 & 0.51 & 0 \\ 0 & 0 & 0.42 & 0 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.35 & 0 & 0.65 & 0 \\ 0 & 0.12 & 0.34 & 0 & 0.35 & 0 & 0.20 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$



$$(\mathbf{S}_z^{lbl})^{td} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \Sigma \end{array}\begin{array}{c} \begin{pmatrix} 0 & 2.00 & 0 & 0 & 0 & 0 & 0 \\ 2.34 & 0 & 3.54 & 0 & 0 & 2.12 & 0 \\ 0 & 5.23 & 0 & 3.61 & 0 & 9.12 & 0 \\ 0 & 0 & 4.57 & 0 & 6.43 & 0 & 0 \\ 0 & 0 & 0 & 3.50 & 0 & 6.50 & 0 \\ 0 & 1.74 & 5.07 & 0 & 5.24 & 0 & 2.95 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 \end{pmatrix} \\ \begin{matrix} 2.34 & 8.97 & 13.18 & 7.11 & 11.67 & 21.74 & 2.95 \end{matrix} \end{array} \begin{array}{c} \Sigma \\ 2 \\ 8 \\ 18 \\ 11 \\ 10 \\ 15 \\ 4 \end{array}$$

Fig. 6. Neighbourhood models $\mathbf{S}_z^{lbl}$ and $(\mathbf{S}_z^{lbl})^{td}$ presented in the matrix form and represented graphically

One can move from neighbourhood model $\mathbf{S}_z^{lbl}$ to a thematic model $(\mathbf{S}_z^{nn})^{td}$ using the transformation (7)

$$(\mathbf{S}_z^{lbl})^{td} = \mathbf{T}^{td}\mathbf{S}_z^{lbl} \tag{26}$$

Then, based on the matrices $(\mathbf{S}_z^{lbl})^{td}$ and $(\mathbf{ND}_z^{lbl})^{td}$, the vectors $(\mathbf{S}_z^{lbl}\mathbf{C})^{td}$ and $(\mathbf{ND}_z^{lbl}\mathbf{C})^{td}$ can be determined. For the considered example, after rounding up to integers, the magnitudes of the corresponding vector components are as follows

$$(\mathbf{S}_z^{lbl}\mathbf{C})^{td} = [2_1,\ 9_2,\ 13_3,\ 7_4,\ 12_5,\ 22_6,\ 3_7],$$
$$(\mathbf{ND}_z^{lbl}\mathbf{C})^{td} = [4_1,\ 17_2,\ 31_3,\ 18_4,\ 22_5,\ 37_6,\ 7_7].$$

Upon assuming that the model depends on the neighbourhood weight, measured by the length of the boundary line between neighbouring areas and comparing vectors $(\mathbf{S}_z^{nn}\mathbf{C})^{td}$ with $(\mathbf{S}_z^{lbl}\mathbf{C})^{td}$, one can clearly see that the algorithm developed is more beneficial for area f_6 and less beneficial for areas f_4 and f_5.

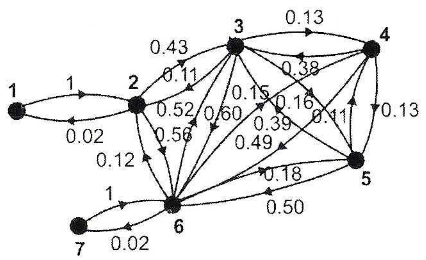
Analogically, a neighbourhood model dependent on the size of neighbouring areas can be generated. The neighbourhood measure will then be

$$\left|\mathbf{S}^{sur}_{z_{f_i,f_j}}\right| = \frac{(sur)_{f_j}}{\sum (sur)_{f_r}} \quad \text{for} \quad \bigcap_{r\in\{1,2,\ldots,n\}} \left|S_{f_i f_r}\right| = 1 \tag{27}$$

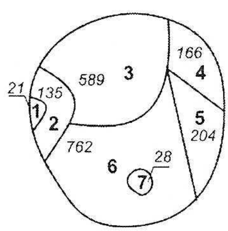
The matrix $\mathbf{S}_z^{sur}$ of dependent neighbourhood will be built according to the already known scheme:

$$(s_z)_{ij}^{sur} = \begin{cases} s_{ij} * \left|\mathbf{S}^{sur}_{z_{f_i,f_j}}\right| & \text{for} \quad s_{ij} = 1 \\ 0 & \text{for} \quad s_{ij} = 0 \end{cases} \tag{28}$$

Models $\mathbf{S}_z^{sur}$, $(\mathbf{S}_z^{sur}\mathbf{C})$ for the examined region are presented in Figure 7.



$$\mathbf{S}_z^{sur} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.02 & 0 & 0.43 & 0 & 0 & 0.56 & 0 \\ 0 & 0.11 & 0 & 0.13 & 0.16 & 0.60 & 0 \\ 0 & 0 & 0.38 & 0 & 0.13 & 0.49 & 0 \\ 0 & 0 & 0.39 & 0.11 & 0 & 0.50 & 0 \\ 0 & 0.12 & 0.52 & 0.15 & 0.18 & 0 & 0.02 \\ 0 & 0 & 0 & 0 & 0 & 1.00 & 0 \end{pmatrix}$$



$$\begin{array}{cc} (\mathbf{S}_z^{sur})^{td} = \begin{pmatrix} 0 & 2.0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 3.4 & 0 & 0 & 4.4 & 0 \\ 0 & 1.9 & 0 & 2.4 & 2.9 & 10.8 & 0 \\ 0 & 0 & 4.2 & 0 & 1.4 & 5.4 & 0 \\ 0 & 0 & 3.9 & 1.1 & 0 & 5.0 & 0 \\ 0 & 1.8 & 7.9 & 2.2 & 2.7 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 4.0 & 0 \end{pmatrix} & \begin{array}{c} \Sigma \\ 2 \\ 8 \\ 18 \\ 11 \\ 10 \\ 15 \\ 4 \end{array} \\ \Sigma \;\; 0.1 \quad 5.7 \quad 19.4 \quad 5.7 \quad 7.0 \quad 29.6 \quad 0.4 & \end{array}$$

Fig. 7. Neighbourhood models $\mathbf{S}_z^{sur}$ and $(\mathbf{S}_z^{sur})^{td}$ presented in the matrix form and represented graphically

For the case presented in Figure 7, the elements of $(\mathbf{S}_z^{sur}\mathbf{C})^{td}$ and $(\mathbf{ND}_z^{sur}\mathbf{C})^{td}$ vectors determined based on matrixes $(\mathbf{S}_z^{sur})^{td}$ and $(\mathbf{ND}_z^{sur})^{td}$ are presented as follows after rounding up to integers

$$(\mathbf{S}_z^{sur}\mathbf{C})^{td} = [0_1,\ 6_2,\ 19_3,\ 6_4,\ 7_5,\ 30_6,\ 0_7],$$
$$(\mathbf{ND}_z^{sur}\mathbf{C})^{td} = [\ 2_1,\ 14_2,\ 37_3,\ 17_4,\ 17_5,\ 46_6,\ 4_7].$$

Analysis of the results indicates that the value of neighbouhood strongly depends in this model on the size of the area; large areas (f_6) are assigned with large value of neighbourhood, while small areas (f_1 and f_7) with the small values.

It is easily noticeable that by designing the dependent neighbourhood model $\mathbf{S}_z^x$, the classic neighbourhood base model $\mathbf{S}$ is modified by changing only the magnitudes of those elements s_{ij} that were equal to 1. They are assigned new values resulting from adopted bases for neighbourhood assessment. After setting up models $\mathbf{S}_z^x$, $x = (nn, lbl, sur)$, the subsequent, derivative thematic models can be built using diagonal matrix $\mathbf{T}^{td}$ that reflects non-spatial descriptive attributes.

Let us notice a certain standardization of dependent neighbourhood models. The dependent neighbourhood models presented above are comparable. The sum of elements in rows of matrix $\mathbf{S}_z^x$ amounts to 1, while in $(\mathbf{S}_z^x)^{td}$ it is equal to the value of attribute t_i^{td} for area f_i.

Let us assume that in the thematic analysis, the value of descriptive attribute t^{td} is determined by the value of income of an administrative unit managing an area, assigned for integrated development. It is assumed that the funds for region integration should be shared with administrators of neighbouring areas. The allocation of funds for common investments in neighbouring areas can be carried out depending on the number of neighbouring areas, but one can take into account the impact of the length of the common boundary line or the size of neighbouring areas. In that case, the values of the elements of the vector $(\mathbf{S}_z^x\mathbf{C})^{td}$ specify the value of funds to be used in every area, originating from the funds of neighbouring administrators and divided in accordance with criterion x.

5. Indirect neighbourhood

Neighbourhood models $\mathbf{S}$, $\mathbf{S}^{td}$ concerned direct neighbours only. The neighbours of neighbours were taken into account in models $\mathbf{S}_z^x$ and $(\mathbf{S}_z^x)^{td}$. Distant neighbours were not taken into account in the presented models, despite often holding significance.

There are many distant neighbours and they should be classified. Let us assume that indirect neighbourhood will be described in the same manner as kinship with the use of various q degrees. The following classification principle will be applied: a direct neighbour will possess degree 1, while an indirect neighbour will be assigned subsequent q degrees, $q = \{2, 3, \dots\}$. Indirect neighbourhood of degree q will be described with neighbourhood matrix $\mathbf{S}^q$, which can be obtained with the following transformation

$$\mathbf{S} \to \mathbf{S}^q, \qquad q = \{\,2,\,3,...\} \tag{29}$$

expressed with the following formula

$$\bigcap_{q=\{2,3,...\}} (s^q)_{ij} = \begin{cases} 0 & \text{for} \quad \rho_{\min}(i,j) \neq q \\ 1 & \text{for} \quad \rho_{\min}(i,j) = q \end{cases} \tag{30}$$

where $\rho_{\min}(i, j)$ is the shortest path in graph **S** between nodes representing areas f_i, f_j. In the classic representation of graphs (Kulikowski, 1986; Wilson, 2000), the path in the graph is described with a quantum measure, which exhibits the minimum number of edges distinguished in the graph, which join together selected nodes in the form of a path. Thus, it adopts absolute values, which correspond to neighbourhood degrees q:

$$\bigcap_{i,j \in \{1,2,3,\ldots,n\}} (f_i f_j) \in \mathbf{S}^q \quad \Leftrightarrow \quad \rho_{\min}(i, j) = q \tag{31}$$

Indirect neighbourhood of degree q of area f_i will take into account only neighbours f_j, for whom minimum path $\rho_{\min}(i, j)$ in graph **S** between nodes i, j amounts to q.

Indirect neighbourhood models $\mathbf{S}^q$ for the examined region are represented by graphs in Figure 8.

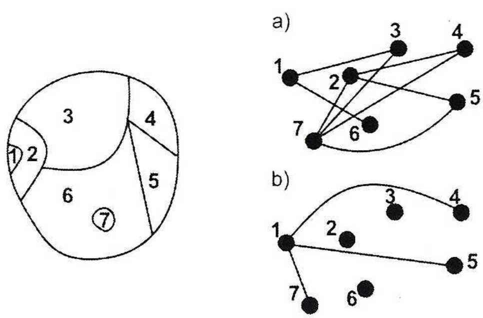


Fig. 8. Area indirect neighbourhood models: a) 2nd degree, b) 3rd degree

It seems that indirect neighbourhood of a specified degree would be rarely applied for analytical tasks. To change opinion, one only needs to notice that the maximum indirect neighbourhood degree informs of the dispersion of areas in the region.

6. Neighbours of degree not higher then q – $\mathbf{S}^{\leq q}$

For the purpose of solving analytical tasks one should search for a neighbourhood model that would aggregate, in a simple manner, direct and indirect neighbours in one simple entry. The neighbourhood degree was specified in the previous section and it was assumed that direct neighbourhood would correspond to neighbourhood of the first degree, while indirect neighbourhood would be assigned subsequent degrees. This can be generalized by simply introducing a neighbourhood degree. With this assumption, one can always refer to neighbours of a specified neighbourhood degree or even neighbourhood of degree not higher than q – $\mathbf{S}^{\leq q}$.

Upon referring to neighbourhood model $\mathbf{S}^{\leq q}$ and adopting $q = 1$, the model $\mathbf{S}^{\leq q}$ describing direct neighbourhood is generated, while if $q = 2$, $\mathbf{S}^{\leq 2}$ model aggregates direct neighbourhood as well as indirect neighbourhood of the 2nd degree. For $q = m$,

$\mathbf{S}^{\leq m}$ model aggregates direct neighbourhood model (**S**) and all indirect neighbourhood models $\mathbf{S}^q$ of degrees $q = \{2, 3, \ldots, m\}$.

The neighbourhood model of degree not higher than q – $\mathbf{S}^{\leq q}$ may be described with a simple transformation

$$\mathbf{S} \rightarrow \mathbf{S}^{\leqslant q} \tag{32}$$

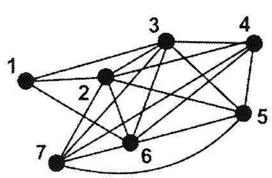
expressed with the following formula

$$\bigcap_{q=\{1,2,3,..\}} (s^{\leqslant q})_{ij} = \begin{cases} 0 & \text{for} \quad s_{ij} = 0, i = j \\ 1 & \text{for} \quad 0 < \rho_{\min}(i, j) \leqslant q \\ 0 & \text{for} \quad \rho_{\min}(i, j) > q \end{cases} \tag{33}$$

where $\rho_{\min}(i, j)$ is the shortest path in the graph **S** between nodes describing areas f_i and f_j. The same transformation may be expressed differently in a form stressing the aggregation process

$$\bigcap_{q\in(1,2,..m)} \mathbf{S}^{\leqslant q} = \mathbf{S} + \sum_{q=2}^{m} \mathbf{S}^q \tag{34}$$

The graphic illustration of neighbourhood models of degree not higher than 2, i.e. $\mathbf{S}^{\leqslant 2}$, is presented in Figure 9. The thematic model $(\mathbf{S}^{\leqslant 2})^{td}$ acquired through the multiplication of the matrix $\mathbf{S}^{\leqslant 2}$ with diagonal matrix $\mathbf{T}^{td}$ is presented in the matrix form in Figure 9. As for the graphic representation, it should be illustrated in the form of a digraph.



$$\mathbf{S}^{\leq 2} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \qquad (\mathbf{S}^{\leq 2})^{td} = \begin{pmatrix} 0 & 2 & 2 & 0 & 0 & 2 & 0 \\ 8 & 0 & 8 & 8 & 8 & 8 & 8 \\ 18 & 18 & 0 & 18 & 18 & 18 & 18 \\ 0 & 11 & 11 & 0 & 11 & 11 & 11 \\ 0 & 10 & 10 & 10 & 0 & 10 & 10 \\ 15 & 15 & 15 & 15 & 15 & 0 & 15 \\ 0 & 4 & 4 & 4 & 4 & 4 & 0 \end{pmatrix}$$

$$\Sigma \quad 41 \quad 60 \quad 50 \quad 55 \quad 56 \quad 53 \quad 62$$

Fig. 9. Neighbourhood models $\mathbf{S}^{\leqslant 2}$ and $(\mathbf{S}^{\leqslant 2})^{td}$ of degree not higher than 2

The elements of the vector $(\mathbf{S}^{\leqslant 2}\mathbf{C})^{td}$ corresponding to the sums of elements of columns or rows of the matrix $\mathbf{S}^{\leqslant 2}$, describe the number of neighbours of each area taking into account the neighbourhood degree $q \leqslant 2$. In thematic model $(\mathbf{S}^{\leqslant 2})^{td}$, the vector $(\mathbf{S}^{\leqslant 2}\mathbf{C})^{td}$ describes for each area the sum of features of neighbourhood areas of degree not higher than 2. The magnitudes of the elements of the vector $(\mathbf{S}^{\leqslant 2}\mathbf{C})^{td}$ are as follows

$$(\mathbf{S}^{\leqslant 2}\mathbf{C})^{td} = [41_1, 60_2, 50_3, 55_4, 56_5, 53_6, 62_7]$$

while for the vector $(\mathbf{ND}^{\leqslant 2}\mathbf{C})^{td}$ they are

$$(\mathbf{ND}^{\leqslant 2}\mathbf{C})^{td} = [43_1, 68_2, 68_3, 66_4, 66_5, 68_6, 66_7]$$

Let us assume that the values of the attribute t^{td} describe values of areas. Then components of the vector $(\mathbf{ND}^{\leqslant 2}\mathbf{C})^{td}$ indicate possibly new values of areas following their merge with neighbouring areas of neighbourhood degree not higher than 2.

6.1. Neighbours of degree not higher that q with suppression t – $S_t^{\leq q}$

In models $\mathbf{S}^{\leqslant q}$ aggregating neighbours of various neighbourhood degrees, one must often assume that neighbours of subsequent degrees have a lower and lower impact on the value of the neighbourhood. Upon building group models in such cases, a factor suppressing neighbourhood value – t, inversely proportional to the neighbourhood degree, e.g. $t = 1/q$ should be introduced. In that case, a neighbourhood model $\mathbf{S}_t^{\leqslant q}$ of degree not higher than q with suppression t may be presented in the form of the following transformation

$$\mathbf{S} \rightarrow \mathbf{S}_t^{\leqslant q} \tag{35}$$

described as follows

$$\bigcap_{q=\{1,2,3...\}} (s_t^{\leqslant q})_{ij} = \begin{cases} 0 & \text{for} & s_{ij} = 0, i = j \\ 1/\rho_{\min}(i,j) & \text{for} & 0 < \rho_{\min}(i,j) \leqslant q \\ 0 & \text{for} & \rho_{\min}(i,j) > q \end{cases} \tag{36}$$

where $\rho_{\min}(i, j)$ is the shortest path in graph $\mathbf{S}$ between nodes describing areas f_i and f_j. Analogically to (34)

$$\bigcap_{q=\{1,2,3,...,m\}} \mathbf{S}_t^{\leqslant q} = \mathbf{S} + \sum_{q=2}^{m} \frac{1}{q} \mathbf{S}_k^{q-1} \tag{37}$$

Thematic models $(\mathbf{S}_t^{\leqslant q})^{td}$ and respective vectors $(\mathbf{S}_t^{\leqslant q}\mathbf{C})^{td}$ can be derived from the model $\mathbf{S}_t^{\leqslant q}$. For $q = 2$, the elements of those vectors become as follows

$$(\mathbf{S}_t^{\leqslant 2}\mathbf{C})^{td} = [25_1, 48_2, 47_3, 49_4, 50_5, 52_6, 39_7]$$
$$(\mathbf{ND}_t^{\leqslant 2}\mathbf{C})^{td} = [27_1, 56_2, 65_3, 60_4, 60_5, 67_6, 43_7] \ .$$

Assume that the values of attribute t^{td} describe the number of unemployed residents in areas of region F and that all of the unemployed from the analysed areas and directly neighbouring areas are potential workforce that can be employed, and in addition that only 50% of them can be employed in areas with neighbourhood of degree 2. The values of the elements of the vector $(\mathbf{ND}_t^{\leqslant 2}\mathbf{C})^{td}$ then indicate that by locating a new plant in area f_6 one can expect around 67 unemployed residents to report for work in total from area f_6 and areas with neighbourhood of a degree not higher than 2. As a result, it was found that area f_6 was assessed as the most attractive with respect to the highest possible employment, i.e. favourable for new investments.

$$(\mathbf{S}_t^{\leqslant 2}) = \begin{pmatrix} 0 & 1 & 0.5 & 0 & 0 & 0.5 & 0 \\ 1 & 0 & 1 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 1 & 0 & 1 & 1 & 1 & 0.5 \\ 0 & 0.5 & 1 & 0 & 1 & 1 & 0.5 \\ 0 & 0.5 & 1 & 1 & 0 & 1 & 0.5 \\ 0.5 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 & 1 & 0 \end{pmatrix} \qquad (\mathbf{S}_t^{\leqslant 2})^{td} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \Sigma \end{array} \begin{array}{c} \begin{pmatrix} 0 & 2 & 1 & 0 & 0 & 1 & 0 \\ 8 & 0 & 8 & 4 & 4 & 8 & 4 \\ 9 & 18 & 0 & 18 & 18 & 18 & 9 \\ 0 & 5.5 & 11 & 0 & 11 & 11 & 5.5 \\ 0 & 5 & 10 & 10 & 0 & 10 & 5 \\ 7.5 & 15 & 15 & 15 & 15 & 0 & 15 \\ 0 & 2 & 2 & 2 & 2 & 4 & 0 \end{pmatrix} \\ \begin{matrix} 25 & 48 & 47 & 49 & 50 & 52 & 39 \end{matrix} \end{array}$$

Fig. 10. Neighbourhood models of degree not higher than 2 with suppression t: $\mathbf{S}_t^{\leqslant 2}$ and thematic model $(\mathbf{S}_t^{\leqslant 2})^{td}$ considering the values of attribute td

6.2. Full neighbourhood with suppression t

Neighbourhood model $\mathbf{S}_t^{\leqslant q}$ for the maximum neighbourhood degree in the region is a full neighbourhood model. It can be presented as:

$$\mathbf{S} \to \mathbf{S}_t^f \tag{38}$$

expressed with the following transformation:

$$(s_t^f)_{ij} = \begin{cases} 1 & \text{for} \quad s_{ij} = 1 \\ 1/\rho_{\min}(i,j) & \text{for} \quad s_{ij} = 0, i \neq j \\ 0 & \text{for} \quad s_{ij} = 0, i = j \end{cases} \tag{39}$$

where suppressing factor $\rho_{\min}(i, j)$ is the shortest path in graph $\mathbf{S}$ between nodes describing areas f_j, f_j.

Only diagonal elements of the matrix $\mathbf{S}_t^f$ of full neighbourhood equal to zero, while all others describe mutual relations between respective pairs of areas. The numerical quantities characterizing full neighbourhood are presented in the matrix $\mathbf{S}_t^f$ in Figure 11. The higher the measure the more important is the neighbourhood.

The elements of full thematic neighbourhood $(\mathbf{S}_t^f)^{td}$ matrix can be determined using e.g. the already known transformation (7)

$$(\mathbf{S}_t^f)^{td} = \mathbf{T}^{td}\mathbf{S}_t^f \tag{40}$$

Full neighbourhood $\mathbf{S}_t^f$ can be similarly defined also upon selecting suppression coefficient values other then in (36) and (39). The appropriate selection usually depends on the nature of the analysed problem. For example, for analysis related to transportation networks, it would be best to describe suppression with measures inversely proportional to the length of road connections between the centres of the examined areas, acquired from a digital map.

$$\mathbf{S}_t^f = \begin{pmatrix} 0 & 1 & 1/2 & 1/3 & 1/3 & 1/2 & 1/3 \\ 1 & 0 & 1 & 1/2 & 1/2 & 1 & 1/2 \\ 1/2 & 1 & 0 & 1 & 1 & 1 & 1/2 \\ 1/3 & 1/2 & 1 & 0 & 1 & 1 & 1/2 \\ 1/3 & 1/2 & 1 & 1 & 0 & 1 & 1/2 \\ 1/2 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1/3 & 1/2 & 1/2 & 1/2 & 1/2 & 1 & 0 \end{pmatrix} \qquad (\mathbf{S}_t^f)^{td} = \begin{array}{c} \begin{pmatrix} 0 & 2 & 1 & 0.7 & 0.7 & 1 & 0.7 \\ 8 & 0 & 8 & 4 & 4 & 8 & 4 \\ 9 & 18 & 0 & 18 & 18 & 18 & 9 \\ 3.7 & 5.5 & 11 & 0 & 11 & 11 & 5.5 \\ 3.3 & 5 & 10 & 10 & 0 & 10 & 5 \\ 7.5 & 15 & 15 & 15 & 15 & 0 & 15 \\ 1.3 & 2 & 2 & 2 & 2 & 4 & 0 \end{pmatrix} \\ \Sigma\; 32.8 \;\; 47.5 \;\; 47 \;\; 49.7 \;\; 50.7 \;\; 52 \;\; 39.2 \end{array}$$

Fig. 11. Full neighborhood model with suppression $t = 1/\rho_{\min}(i, j)$ described with matrix $\mathbf{S}_t^f$ and thematic model $(\mathbf{S}_t^f)^{td}$ of full neighbourhood

7. Neighbourhood models as aggregate models

The selection of the model depends on the purpose of the analysis. It is advisable to apply multidimensional analysis (Jajuga, 1993) when there is a need of taking into account more than one geometric or thematic feature $\mathbf{T}^{td}$, $td = \{1, 2, \ldots, k\}$ in the neighbourhood analysis. In order for such an analysis to be performable, neighbourhood models should be set up to be mutually comparable. It is usually done by data normalization (Strahl, 1998). Standardization is the most frequently applied normalization method. It should be performed for every model considered

$$(\mathbf{S}_c^x)^{td} \xrightarrow{standardization} (\mathbf{S}_c^x)_{st}^{td} \tag{41}$$

where $c \in (\varnothing, z, t)$, $x \in (\varnothing, nn, lbl, sur, q, \leq q, f)$, $td \in (\varnothing, 1, 2, 3, \ldots, k)$.

Standardization is performed for each model $(\mathbf{S}_c^x)^{td}$ separately, determining new values of neighbourhood matrix elements in accordance with the following principle:

$$\bigcup_{c,x,td} (s_c^x)_{st_{ij}}^{td} = \frac{(s_c^x)_{ij}^{td} - \frac{1}{m} \sum_{i=1}^{n} \sum_{j=1}^{n} (s_c^x)_{ij}^{td}}{\sqrt{\mathrm{var}(s_c^x)^{td}}} \tag{42}$$

where $m = n * n$ is the number of matrix elements in neighbourhood model $(\mathbf{S}_c^x)^{td}$, $\frac{1}{m} \sum_{i=1}^{n} \sum_{j=1}^{n} (s_c^x)_{ij}^{td}$ is the mean of elements of the matrix $(\mathbf{S}_c^x)^{td}$, $\sqrt{\mathrm{var}(s_c^x)^{td}}$ is the standard deviation of matrix element values.

Standardized models $(\mathbf{S}_c^x)_{st}^{td}$ are the basis for generating aggregate neighbourhood models. They can be developed with the assumption that measures from various models are equivalent or that certain models are dominating. In the first case, the measures S_{ag} of aggregate model are calculated as the arithmetic mean of measures of component models. In that case, the values of the elements of a matrix describing an aggregate model are mean values from respective elements of matrices taken into account in the model. In the other case, a weight w_m system should be introduced, in order to determine the importance of the model

$$\{w_1\mathbf{S}_{st}^1, w_2\mathbf{S}_{st}^2, w_3\mathbf{S}_{st}^3, \ldots., w_r\mathbf{S}_{st}^r\} \rightarrow \mathbf{S}_{agw} \tag{43}$$

Weighted aggregate measures $\mathbf{S}_{agw}$ are calculated using the following formula

$$s_{agw_{ij}} = \frac{\sum_{n=1}^{r} w_n s_{st_{ij}}^{n}}{r \sum_{n=1}^{r} w_n} \tag{44}$$

Aggregate models are used for multidimensional analysis, such as the selection of areas characterized by the best neighbourhood parameters, while taking into account many features. The results of similar analysis can be considered in the decision-making process related to the selection of a location for, e.g. a new investment.

Aggregate models are a function of standardized models. The selection of types of component models and weights to build an aggregate model depends on the user's assessment and the purpose of analysis (Jajuga, 1993; Strahl, 1998).

8. Summary

The presented approach to building area neighbourhood models is the result of applying graph theory, that allows to record neighbourhood models in a simple manner in the graphic and matrix form. Initial neighbourhood models are associated with a simple graph, while thematic models with a digraph. Graphs accounting for neighbourhood models can be automatically generated based on spatial data, acquired from digital maps.

The paper proposes a new systematics of neighbourhood types. The following terms were introduced: direct neighbourhood, direct dependent neighbourhood, indirect neighbourhood of degree q, neighbourhood of degree not higher than q and full neighbourhood. Models were developed for the listed types of neighbourhood. Neighbourhood models $\mathbf{S}$, $\mathbf{S}_z^x$, $\mathbf{S}^q$, $\mathbf{S}^{\leqslant q}$, $\mathbf{S}_t^{\leqslant q}$, $\mathbf{S}_t^f$ were described using quantum metric. If it were replaced with a different metric, e.g. Euclidean or weighted (Lewandowicz, 2005), then different results of analysis could be expected.

Simple examples were provided based on the presented models of thematic analysis taking into account selected area features described by the values of area attributes t^{td}. They are associated with the summing of features of neighbouring areas, while assuming that the values of features t^{td} of areas describe neighbourhood $\mathbf{S}^{td}$ fully or partially – depending on geometric features $(\mathbf{S}_z^x)^{td}$ and quantum distance, calculated between area centres $((\mathbf{S}_t^{\leqslant q})^{td}, (\mathbf{S}_t^f)^{td})$. In practice, the selection of a neighbourhood model should depend on the analytical task and user requirements.

The need to find a method enabling to consider simultaneously many neighbourhood features in analysis will be satisfied by aggregate neighbourhood models generated from models that take into account selected single features.

The processes of modelling various neighbourhood types described in the paper can be represented by one algorithm leading from the base model $\mathbf{S}$ to the model corresponding to a given type. Such a method of neighbourhood model building may provide the basis for formulating new algorithms increasing the scope of offered analytical GIS tools.

Acknowledgments

Works have been performed within the statutory research project No: 0309.0213 "Creation and updating digital map data bases for geodetic administration and teaching needs" supported by the University of Warmia and Mazury in Olsztyn.

References

Ahuja P.K., Magnanti T.L., Orlin J.B., (1993): *Network Flows, Theory, Algorithms and Applications*, Prentice Hall, Englewood Cliffs.

Autodesk, (2000): *User's Manual for AutoCad Map@-2000, Release 4* (in Polish).

Bentley, (2002): *Basics of MicroStation GeoGraphics*, Materials from basic course (in Polish).

Bera R., Claramunt Ch., (2003): *Topolgy-based proximities in spatial systems*, Journal of Geographical Systems, Springer-Verlag, 5, pp. 353-379.

Caruso G., Rounsevell M., Cojocaru G., (2005): *Exploring a spatio-dynamic neighbourhood – based model of residential behaviour in the Brussels periurban area*, International Journal of Geographical Information Science, Vol. 19, No 2, pp. 103-123.

ESRI, (2005): *Neighbourhood*, http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?ID=2811&TopicName=Neighbourhood%20filters&rand=588&pid=2808.

Fischer M., Scholen H.J., Unwin D., (1996): *Spatial Analytical Perspectives on GIS*, GIS DATA IV, Taylor & Francis, London.

Jajuga K., (1993): *Statistical multidimensional analysis* (in Polish), Polish Scientific Publishers PWN, Warsaw.

Kulikowski J.L., (1986): *Graph theory outline* (in Polish), Polish Scientific Publishers PWN, Warsaw.

Lewandowicz E., (2004): *Graphs as a tool for defining spatial relations between geographic data* (in Polish), Annals of Geomatics, Warsaw, Vol. II, book 2, pp. 160-171.

Lewandowicz E., Bałandynowicz J., (2005): *Some Ways of Formulation of Objective Functions for Chosen Space Analysis*, The 6th International Conference Faculty of Environmental Engineering, Vilnius Gediminas Technical University, Vol. 2, pp. 927-930.

Lewandowicz E., (2005): *Analysis of micro-region neighbourhood within a region based on spatial data recorded in the form of a geometric graph* (in Polish), Annals of Geomatics, Warsaw, Vol. III, book 1, pp. 73-82.

Longley P., Goodchild M., Maguire D., Rhind D., (2001): *Geographic Information Systems and Science*, John Wiley & Sons, LTD.

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

Moore K., (2000): *Resel filtering to aid visualisation within an exploratory data analysis system*, Journal of Geographical Systems, Springer-Verlag, 2, pp. 375-398.

Sadahiro Y., (2005): *Buffer Operation on Spatial Data with Limited Accuracy*, Transactions in GIS, Vol. 9, No 3, pp. 323-344.

Strahl D., (1998): *Taxonomy of structures in regional research* (in Polish), Wroclaw University of Economics.

Sullvan D., Unwin D., (2002): *Geographic Information Analysis*, John Wiley & Sons, INC.

Wilson R., (2000): *Introduction to graph theory* (in Polish), Polish Scientific Publishers PWN, Warsaw.

Wróblewski P., (1997): *Algorithms and structures of data and programming techniques* (in Polish), Helion, Gliwice.

Zhan F.B., (1998): *Representing Networks*, NCGIA Core Curriculum in GIScience, http//www.ncgia.ucsb.edu/giscc/units/u064/u064html, created 5 November 1998.

Modele sąsiedztwa obszarów

Elżbieta Lewandowicz

Katedra Geodezji Szczegółowej
Uniwersytet Warmińsko-Mazurski w Olsztynie
ul. Jana Heweliusza 12, 10-957 Olsztyn
leela@uwm.edu.pl

Streszczenie

W systemach GIS sąsiedztwo obszarów analizowanego regionu jest pojęciem stosowanym tylko dla danych rastrowych. Przeniesienie tego pojęcia na dane wektorowe i opisowe oraz usystematyzowanie modeli tak pojętego sąsiedztwa stanowi przedmiot tego opracowania. Punktem wyjścia jest założenie, że podstawowy model sąsiedztwa obszarów może być przygotowany na podstawie danych przestrzennych, zobrazowany za pomocą grafu i opisany macierzą sąsiedztwa. Stanowi on podstawę do budowy następnych modeli, które wiążą się z wprowadzeniem nowych miar sąsiedztwa – miar wynikających z charakterystyk obszarów zapisanych w tabelach ich atrybutów. W oparciu o proponowane modele można wykonywać potrzebne analizy przestrzenne związane z sąsiedztwem obszarów oraz konstruować modele agregatowe, niezbędne w wielowymiarowej analizie sąsiedztwa.