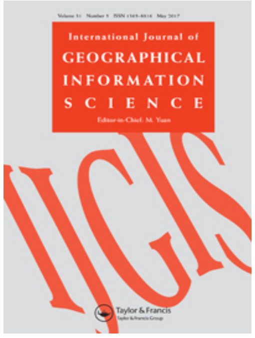


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

The Network-Max-P-Regions model

Bing She, Juan C. Duque & Xinyue Ye

To cite this article: Bing She, Juan C. Duque & Xinyue Ye (2017) The Network-Max-P-Regions model, International Journal of Geographical Information Science, 31:5, 962-981, DOI: 10.1080/13658816.2016.1252987

To link to this article: https://doi.org/10.1080/13658816.2016.1252987

Published online: 04 Nov 2016.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2017
VOL. 31, NO. 5, 962–981
http://dx.doi.org/10.1080/13658816.2016.1252987

The Network-Max-P-Regions model

Bing She[a,b], Juan C. Duque[c] and Xinyue Ye[d]

[a]LIESMARS, Wuhan University, Wuhan, China; [b]China Data Center, University of Michigan, Ann Arbor, MI, USA; [c]Research in Spatial Economics (RiSE), School of Economics and Finance, Universidad EAFIT, Medellin, Colombia; [d]Department of Geography, Kent State University, Kent, OH, USA

ABSTRACT

This paper introduces a new p-regions model called the Network-Max-P-Regions (NMPR) model. The NMPR is a regionalization model that aims to aggregate *n* areas into the maximum number of regions (max-p) that satisfy a threshold constraint and to minimize the heterogeneity while taking into account the influence of a street network. The exact formulation of the NMPR is presented, and a heuristic solution is proposed to effectively compute the near-optimized partitions in several simulation datasets and a case study in Wuhan, China.

ARTICLE HISTORY
Received 23 February 2016
Accepted 16 October 2016

KEYWORDS
Network-Max-P-Regions; heuristics; spatial contiguity; streets; regionalization

1. Introduction

With the introduction of the p-regions problem (Duque *et al*. 2011b), the use of mixed integer programming (MIP) for regionalization has increased in the max-p-regions model (Duque *et al*. 2012), the p-functional regions model (Kim *et al*. 2013), and the p-compact regions model (Li *et al*. 2014b). However, most models ignore transportation networks (Borgatti *et al*. 2009). In addition, streets constrain human activities (Wheeler 1973, Whitehand and Larkham 1992, Jiang and Claramunt 2004).

This paper extends the max-p-regions model (Duque *et al*. 2012) to the Network-Max-P-Regions (NMPR) model. This is a regionalization model that aims to aggregate *n* areas into a maximum number of regions (max-p) such that (1) the resulting regions minimize a measure of intraregional heterogeneity, in terms of a set of areal attributes (e.g., slope, types of soil, types of land use, price); (2) the areas along streets can be grouped into 'corridor regions,' and other areas are grouped into 'max-p regions;' and (3) each region must satisfy the condition that the value of a spatially extensive areal attribute (e.g., population, households, size) exceeds a predefined threshold value.

Physical constraints imposed by transportation networks shape the way in which human dynamics evolve (Jiang and Jia 2011). Therefore, the presence of a road might modify the attributes of the areas along it (e.g., percentage of vegetation cover, income, population density, commercial activity, land prices, percentage of impervious surfaces). This influence can be asymmetrical or symmetrical. The asymmetrical influence leads to the different attributes of the areas on either side of the road. For example, ring roads usually create a clear division between rural and urban areas (e.g., the northern edge of

CONTACT Juan C. Duque jduquec1@eafit.edu.co

the A10 in Amsterdam, Netherlands), and streets separate neighborhoods with different socioeconomic characteristics (e.g., the 8 Mile Road in Detroit, USA). In those cases, the road acts as a regional border. The symmetrical influence might generate areas with similar attributes along both sides of the road (e.g., shopping strips, edge cities, or fishbone deforestation along roads). In this case, the regionalization process tends to generate corridor regions along the roads.

However, due to the nature of the phenomenon (e.g., wildlife trails), or because of its short time period (e.g., recently constructed roads), the presence of the road may not have modified the attributes of the areas along it. In these cases, if the road is not explicitly included in the regionalization model, the infrastructure will be 'invisible' to the model. The NMPR is useful in such cases in which the influence of the roads cannot be captured through the area attributes, but the practitioner wants the resulting regions to be affected by those roads. To the best of our knowledge, the NMPR is the first regionalization model that allows for two types of regions: corridor regions (elongated regions along roads) and max-p-regions (shape-free regions located either close to or far from roads).

Well-defined corridor regions are important in spatial planning. With proper land use planning along the roads connecting cities, the benefits of a road can spread to the intermediate areas that it crosses. Habitat attributes and information about migratory pathways can be integrated within the NMPR to delineate wildlife corridors used to protect species. In intra-urban applications, the NMPR can use a street network and block-wise commercial activity to design a zoning plan that includes both elongated shopping strips and compact business districts.

The paper is structured as follows: Section 2 provides a literature review. The problem formation is established in Section 3. Section 4 explains the heuristic framework used to solve the problem. Section 5 illustrates the proposed approach using several simulated datasets and a case study in Wuhan City, China. The last section discusses the strengths and limitations of the proposed method and suggests future work.

2. Literature review

Regionalization addresses the aggregation of spatial units (hereafter areas) into contiguous regions. It has been widely studied since the early 1960s, when electoral districting became a matter of quantitative modeling (Vickrey 1961, Nagel 1965, Bunge 1966, Aboolian *et al*. 2009). A few years later, the concept of spatial contiguity was applied to school districting (Clarke and Surkis 1968), sales territory alignment (Hess and Samuels 1971), health care districting (Thomas 1979), zone design (Openshaw 1977), and police patrol areas (Curtin *et al*. 2010), among others (see Duque *et al*. 2007 for a literature review of regionalization). Since the appearance of the p-regions problem as an MIP formulation of the regionalization problem (Duque *et al*. 2011b), research in this area has focused on proposing new exact formulations (Duque *et al*. 2012, Kim *et al*. 2013, Li *et al*. 2014b) and exploring new possibilities for solving large instances of this NP-hard problem using parallel computing (Laura *et al*. 2015). These regionalization algorithms are implemented in spatial analytical libraries, such as PySAL (https://pysal.readthedocs.io), and are integrated into GIS platforms (such as ArcGIS and QGIS) to assist researchers and practitioners in solving real-world spatial problems.

The max-p-regions model (Duque *et al.* 2012) is the first MIP model to endogenize the number of regions. This is achieved by maximizing the number of homogeneous regions such that the value of a spatially extensive regional attribute (e.g., area, population, presence of a given type of facility) is above a predefined threshold value. Because intraregional homogeneity is an important characteristic in regionalization, maximizing the number of regions is a good condition to keep heterogeneity low (minimum aggregation bias).

The use of road networks in regionalization serve two purposes: first, it defines neighboring relationships between areas (Zoltners and Sinha 1983). In this case, two areas are neighbors if there is a road that connects them. This representation allows the inclusion of natural obstacles, such as mountains, lakes, and rivers. Second, it quantifies the attractiveness between areas by measuring flows of people, vehicles or products. This method is useful in designing functional or compact regions (Kim *et al.* 2013).

Our NMPR model proposes a novel way to incorporate road networks into regionalization problems. In our model, the road network exerts an influence on the shape of the regions by considering the position of the areas relative to the road network. Thus, areas closer to roads tend to group along the roads, forming what we call 'corridor regions.' A challenging condition in the design of this model is that corridor regions are a possibility rather than a requirement. Corridor regions occur when they do not degrade the intra-regional homogeneity. This implies that corridor regions will not appear everywhere along the road network but only where they are needed.

The NMPR is an NP-hard problem, implying that the computational complexity increases rapidly with the number of areas. This characteristic precludes the use of exact optimization techniques in large instances of the problem and requires the use of heuristic methods. Because of its capacity to escape from local optimal solutions, the Tabu search algorithm (Bozkaya *et al.* 2003) has proved to be the best option when addressing regionalization problems. It has been applied in political districting by Bozkaya *et al.* (2003) and Ricca and Simeone (2008), in zone design by Openshaw and Rao (1995), in home care districting by Blais *et al.* (2003), and more recently for the max-p-regions problem (Duque *et al.* 2012) and the p-compact-regions (Li *et al.* 2014a).

3. Problem formulation

The NMPR model is formulated as a MIP model. It builds upon the max-p-regions model devised by Duque *et al.* (2012), which implies that (1) the number of final regions is endogenously defined, (2) the intraregional heterogeneity is minimized, and (3) each region is spatially contiguous.

Parameters:

I = set of areas, $I = \{1, \ldots, n\}$;

i, j = indices used to represent areas, with $i, j \in I$;

k = index of potential regions, $k = \{1, \ldots, n\}$;

c = index of order for areas, $c = \{0, \ldots, q\}$, with $q = (n - 1)$;

$$w_{ij} = \begin{cases} 1, \text{if areas } i \text{ and } j \text{ share a border, with } i, j \in I \text{ and } i \neq j, \\ 0, \text{otherwise} \end{cases}$$

$N_i = \{j | w_{ij} = 1\}$, the set of areas that are adjacent to area i;

d_{ij} = dissimilarity of the relationships between areas $i$and j, with $i, j, \in I$ and $i < j$; it is left to the practitioner to decide which dissimilarity function to use. In this paper, we use the squared Euclidean distance.

$h = 1 + \log\left(\sum_i \sum_{j|j\rangle i} d_{ij}\right)$, which is the number of digits of the floor function of $\sum_i \sum_{j|j\rangle i} d_{ij}$, with $i, j \in I$;

l_i = spatially extensive attribute value of area i, with $i \in I$;

threshold = $m$inimum value for attribute l at a regional scale;

E, e = set and index of edges in an undirected network, $E = \{1, \ldots, m\}$;

$v_i = \begin{cases} 1, & \text{if area } i \in I \text{intersects an edge } e \in E, \\ 0, & \text{otherwise} \end{cases}$

$O_i = \{e | e \in E$ and it is the nearest edge to area $i \in I\}$; for each area i, there is only one nearest edge e;

s_{i,O_i} = spatial distance between area $i \in I$ and its closest edge e, O_i;

Decision variables:

$t_{ij} = \begin{cases} 1, & \text{if areas } i \text{ and } j \text{ belong to the same region } k, \text{with } i < j, \\ 0, & \text{otherwise} \end{cases}$

$x_i^{kc} = \begin{cases} 1, & \text{if area } i \text{is assigned to region } k \text{ in order } c, \\ 0, & \text{otherwise} \end{cases}$

$c^k = \begin{cases} 1, & \text{if region } k \text{ is a corridor region}, \\ 0, & \text{otherwise} \end{cases}$

$b_{ij}^k = \begin{cases} 1, & \text{ifareas} i \text{and} j \text{belongtothesamecorridorregion} k, \text{with} i < j, \\ 0, & \text{otherwise} \end{cases}$

Minimize:

$$z = \left(-\sum_{k=1}^{n} \sum_{i=1}^{n} x_i^{k0}\right) * 10^h + \sum_i \sum_{j|j\rangle i} d_{ij} t_{ij} - \sum_i \sum_{j|j\rangle i} d_{ij} f\left(s_{i,O_i}, s_{j,O_j}\right) \sum_{k=0}^{n} b_{ij}^k \tag{1}$$

Subject to:

$$\sum_{i=1}^{n} x_i^{k0} \leq 1, \forall k = 1, \ldots, n; \tag{2}$$

$$\sum_{k=1}^{n} \sum_{c=0}^{q} x_i^{kc} = 1, \forall i = 1, \ldots, n; \tag{3}$$

$$x_i^{kc} \leq \sum_{j \in N_i} x_j^{k(c-1)}, \forall i = 1, \ldots, n; \forall k = 1, \ldots, n; \forall c = 1, \ldots, q; \tag{4}$$

$$\sum_{i=1}^{n} \sum_{c=0}^{q} x_i^{kc} l_i \geq threshold * \sum_{i=1}^{n} x_i^{k0}, \forall k = 1, \ldots, n; \tag{5}$$

$$t_{ij} \geq \sum_{c=0}^{q} x_i^{kc} + \sum_{c=0}^{q} x_j^{kc} - 1, \forall i,j = 1,\ldots,n|i<j; \forall k = 1,\ldots,n; \tag{6}$$

$$x_i^{kc} \in \{0,1\}, \forall i = 1,\ldots,n; \forall k = 1,\ldots,n; \forall c = 0,\ldots,q; \tag{7}$$

$$t_{ij} \in \{0,1\}, \forall i,j = 1,\ldots,n|i<j; \tag{8}$$

$$\sum_{i=1}^{n} \sum_{c=0}^{q} x_i^{kc} v_i \geq c^k, \forall k = 1,\ldots,n; \tag{9}$$

$$3b_{ij}^k \leq \sum_{c=0}^{q} x_i^{kc} + \sum_{c=0}^{q} x_j^{kc} + c^k, \forall i,j = 1,\ldots,n|i<j; \forall k = 1,\ldots,n; \tag{10}$$

$$b_{ij}^k \geq \sum_{c=0}^{q} x_i^{kc} + \sum_{c=0}^{q} x_j^{kc} + c^k - 2, \forall i,j = 1,\ldots,n|i<j; \forall k = 1,\ldots,n; \tag{11}$$

$$c^k \in \{0,1\}, \forall k = 1,\ldots,n; \tag{12}$$

$$b_{ij}^k \in \{0,1\}, \forall i,j = 1,\ldots,n|i<j; \forall k = 1,\ldots,n. \tag{13}$$

Objective function (1) has three terms combined in such a way that term I dominates terms II and III. This is achieved by multiplying term I by a scaling factor 10^h, with $h = 1 + \log\left(\sum_i \sum_{j|j\rangle i} d_{ij}\right)$. When terms I, II, and III are added as a single value (-I+II-III), this scaling factor moves the term I to the left so that it is not combined with terms II and III. This strategy was first applied by Duque *et al.* (2012) in the max-p-regions model.

Term I represents the number of regions calculated as the sum of root areas (i.e., the area assigned at order equal to zero, x_i^{k0}). To maximize the number of regions, term I is multiplied by −1 so that the greater the number of regions, the lower the objective function. Term II adds dissimilarity between areas assigned to the same region (i.e., $t_{ij} = 1$). For this, practitioners can use either distance or dissimilarity functions in a univariate or multivariate context. Finally, term III represents an incentive to create corridor regions by subtracting a portion of the intraregional heterogeneity of the corridor region from term II. The portion to be subtracted is determined by a factor f with the following structure:

$$f\left(s_{i,O_i}, s_{j,O_j}\right) = MaxDisc - \frac{MaxDisc}{MaxDist}\left(s_{i,O_i+}s_{j,O_j}\right), \tag{14}$$

where *MaxDisc* is the maximum discount to be applied to d_{ij}. The maximum discount occurs when both areas *i* and *j* are very close to a road (i.e., $s_{i,O_i} + s_{j,O_j} \to 0$). This *MaxDisc* decreases as the distance between *i* or *j* and the road increases (i.e., $s_{i,O_i} + s_{j,O_j} \to \infty$). *MaxDist* indicates the threshold distance beyond which the discount becomes negative,

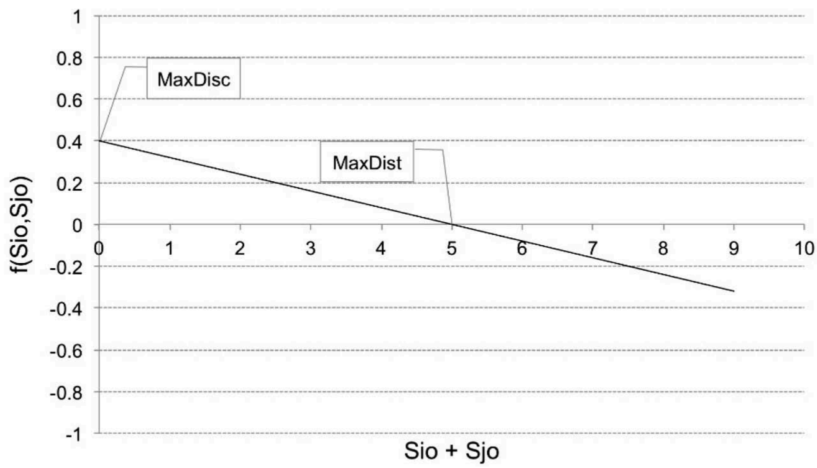


Figure 1. Example of a discount function.

which discourages the appearance of corridor regions far from roads. Figure 1 presents an example of a discount function with $MaxDisc = 0.4$ (40%) and $MaxDist = 5$.

According to Figure 1, the discount function is:

$$f\left(s_{i,O_i}, s_{j,O_j}\right) = 0.4 - \frac{0.4}{5}\left(s_{i,O_i} + s_{j,O_j}\right). \tag{15}$$

Note, that areas closer to the roads have a greater incentive to become corridor areas because the discount is higher. The discount function can have other functional forms depending on the research question.

Solutions with a higher number of regions are always preferred, and for a given number of regions, solutions with lower intraregional heterogeneity (calculated with terms II and III) are preferred over solutions with higher intraregional heterogeneity.

Constraints (2)–(8) correspond to the max-p-regions. Constraints (2)–(4) guaranty the spatial contiguity of the regions by using an extension of the ordered-area assignment conditions proposed by Cova and Church (2000). Constraint (2) guarantees that each region k has no more than one root area. Constraint (3) requires that each area i is assigned to only one region k and only one order c. Constraint (4) ensures spatial contiguity by allowing an area i to be assigned to a given region k in order c if and only if there exists an area j in the neighborhood of i that is assigned to the same region k in the order c-1. Constraint (5) ensures that each region satisfies the minimum threshold value imposed over a spatially extensive attribute. Constraint (6) identifies the pairwise relationships involving areas assigned to the same region. Constraints (7) and (8) impose variable integrity.

The NMPR adds the following constraints to the original max-p-regions model. Constraint (9) allows a region k to become a corridor region if at least one of its areas intersects an edge of the network. A link b_{ij}^k indicates that both areas i and j belong to a corridor region k. However, because of the form of the discount function, the activation of link b_{ij}^k can imply both a decrease in the objective function (when the areas are close to a road, i.e., a positive discount function), or an increase in the objective function (when the areas are far from a road, i.e., a negative discount function). Because all the b_{ij}^k links between areas belonging to a corridor region are activated, a region is declared a corridor region when the total effects of the discount function lead to a positive value of the third term in the objective function. Constraints (10) and (11) activate all the b_{ij}^k links between areas belonging to the region, but they act in different situations: constraint (10) controls the appearance of b_{ij}^k links with a positive discount value by requiring that b_{ij}^k links are only accepted when both areas belong to a corridor region. In contrast, constraint (11) forces the opening of those b_{ij}^k links with a negative discount value; it mandates that these b_{ij}^k links be opened when areas i and j are assigned to the same corridor region. Finally, constraints (12) and (13) impose variable integrity.

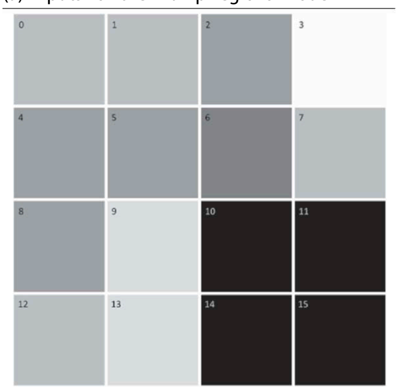
To illustrate how the NMPR model works, Table 1 presents an example of the max-p-regions (a) and the NMPR (b) models.

The solutions are shown in Table 2. Note the differences in the configuration of the regions; the presence of the road in the NMPR model has a direct effect on the shape of the regions. The NMPR model places a corridor region along two segments of the road network in areas 0, 4, 5, and 6. The other three regions are max-p-regions.

Table 3 summarizes the results using the exact formulation to solve instances with different sizes and thresholds. These are the same instances used by Duque *et al.* (2012), but in this case, we included a vertical road in the middle of the lattice. In 4 h, only two instances were solved to optimality, and feasible solutions were obtained for seven

Table 1. Examples of the (a) the max-p-regions and (b) the Network-Max-P-Regions model.

(a) Inputs for the max-p-regions model



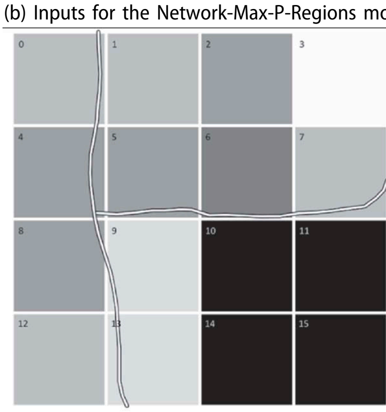
a_i : $\{3.0, 2.6, 3.1, 1.5, 3.5, 3.4, 3.6, 3.0, 3.3, 2.4, 5.7, 5.4, 2.9, 2.1, 5.3, 5.6\}$

$d_{ij} = (a_i - a_j)^2$

li :$\{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$

threshold $= 4$

(b) Inputs for the Network-Max-P-Regions model



a_i : $\{3.0, 2.6, 3.1, 1.5, 3.5, 3.4, 3.6, 3.0, 3.3, 2.4, 5.7, 5.4, 2.9, 2.1, 5.3, 5.6\}$

$d_{ij} = (a_i - a_j)^2$

li :$\{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$

threshold $= 4$

s_{i,O_i} : $\{395, 520, 1613, 1306, 333, 454, 507, 436, 565, 498, 592, 659, 714, 398, 1501, 1763\}$

Note : thedistancesaretheshortest Euclideandistancesfromtheareacentroidto theroad.

$f(s_{i,O_i}, s_{j,O_j}) = 0.5 - \frac{0.5}{912}(s_{i,O_i} + s_{j,O_j})$

$v_i = \{1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0\}$

Table 2. Solutions for (a) the max-p-regions and (b) the Network-Max-P-Regions model.

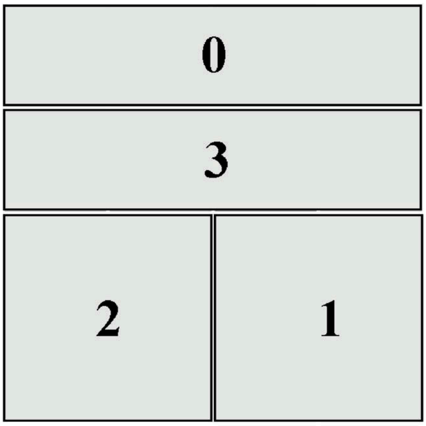
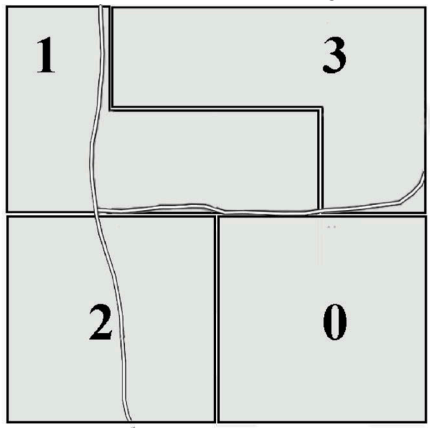
(a) Solution for the max-p-regions model	(b) Solution for the Network-Max-P-Regions model
	
Best objective -3.988940000000e+03, Non-zero decision variables: t_{ij} : t_[0, 1], t_[0, 2], t_[0, 3], t_[1, 2], t_[1, 3], t_[2, 3], t_[4, 5], t_[4, 6]t_[4, 7], t_[5, 6], t_[5, 7], t_[6, 7], t_[8, 9], t_[8, 12], t_[8, 13], t_[9, 12], t_[9, 13], t_[10, 11], t_[10, 14], t_[10, 15], t_[11, 14], t_[11, 15], t_[12, 13] t_[14, 15]. x_i^{kc} : x_[0, 0, 2], x_[1, 0, 1], x_[2, 0, 0], x_[3, 0, 1], x_[4, 3, 2], x_[5, 3, 1], x_[6, 3, 0], x_[7, 3, 1], x_[8, 2, 2], x_[9, 2, 1], x_[10, 1, 1]x_[11, 1, 0], x_[12, 2, 1], x_[13, 2, 0], x_[14, 1, 2], x_[15, 1, 1].	Best objective -3.988971501500e+03 Non-zero decision variables: t_{ij} : t_[0, 4], t_[0, 5], t_[0, 6], t_[1, 2], t_[1, 3], t_[1, 7], t_[2, 3], t_[2, 7], t_[3, 7], t_[4, 5], t_[4, 6], t_[5, 6], t_[8, 9], t_[8, 12], t_[8, 13], t_[9, 12], t_[9, 13], t_[10, 11], t_[10, 14], t_[10, 15], t_[11, 14], t_[11, 15], t_[12, 13], t_[14, 15]. x_i^{kc} : x_[0, 1, 2], x_[1, 3, 1], x_[2, 3, 0], x_[3, 3, 1], x_[4, 1, 1], x_[5, 1, 0], x_[6, 1, 1], x_[7, 3, 2], x_[8, 2, 1], x_[9, 2, 0], x_[10, 0, 1], x_[11, 0, 2], x_[12, 2, 2], x_[13, 2, 1], x_[14, 0, 0], x_[15, 0, 1]. c^k : c_[1]. b_{ij}^k : b_[0, 4, 1], b_[0, 5, 1], b_[0, 6, 1], b_[4, 5, 1], b_[4, 6, 1], b_[5, 6, 1].

Table 3. Computational experiment with GUROBI.

Problem	*n*	Threshold	MaxDist	Solution	Time (sec.)	# max-p regions	# corridor regions
1	9	28	2	$-2.947 \cdot 10^2$	15.30	0	3
2	9	38	1	$-1.851 \cdot 10^2$	21.21	1	1
3	16	51	1	$-2.996 \cdot 10^4$	Ŧ	2	1
4	16	68	2	$-1.993 \cdot 10^4$	Ŧ	1	1
5	25	52	3	$-4.916 \cdot 10^3$	Ŧ	3	2
6	25	79	3	$-2.842 \cdot 10^3$	Ŧ	2	1
7	25	105	3	$-1.771 \cdot 10^3$	Ŧ	1	1
8	36	53	2	–	Ŧ	–	–
9	36	68	3	$-4.966 \cdot 10^4$	Ŧ	5	0
10	36	120	3	$-5.241 \cdot 10^3$	Ŧ	3	0
11	49	54	3	–	Ŧ	–	–
12	49	65	3	–	Ŧ	–	–
13	49	82	3	–	Ŧ	–	–
14	49	109	3	–	Ŧ	–	–

Ŧ Run stopped after 4 h.
– No solution found.

instances. These results illustrate the complexity of this NP-hard problem and the convenience of using heuristic approaches to solve large instances.

4. Heuristic solution

In this section, we present a heuristic solution for solving the NMPR model. We followed a classical heuristic approach for solving regionalization problems, extending the algorithm to solve the Max-p problem (Duque *et al*. 2012). First, we generated a large amount of initial feasible solutions, and second, we performed a local search by modifying a subset of initial solutions with the goal of improving the objective function.

4.1. *Initial feasible solution*

This stage generates a large number of initial feasible solutions at random with different numbers of corridor regions and max-p regions. Pseudocode 1 presents the steps for generating initial feasible solutions. It consists of four steps: area filtering, adaptive region formation, partition ranking and filtering, and enclaving.

Area filtering

In this phase, the areas become regions by themselves when $l_i \geq$ *threshold*. If an area intersects the road network, it is categorized as a corridor region; otherwise, it becomes a max-p region.

Adaptive region formation

This phase builds upon the seeded regions strategy first proposed by Vickrey (1961), which is widely used in the context of regionalization (Thoreson and Liittschwager 1967, Openshaw 1977, Rossiter and Johnston 1981, Duque *et al*. 2012). This method selects an area at random and grows a region around it. The algorithm starts with areas that intersect the road network to generate corridor regions. During the growing process, a region can change from a corridor to a max-p region if this benefits the objective function. This phase is completed when it is not possible to grow new feasible regions. The remaining unassigned areas are called 'enclaves' (Duque *et al*. 2012).

Partition ranking and filtering

At this point, the method has yielded *n* partial solutions. Each partial solution generates a given number of regions. This stage creates a set **P** of partial solutions with the maximum number of regions (the other partial solutions are discarded).

Enclaving

This stage takes each partial solution in **P** and transforms it to a feasible solution by assigning the enclave (unassigned) areas to an existing region. This process is based on the greedy-based assignation process, *AssighEnclaves*, devised by Duque *et al*. (2012).

As with the max-p-regions model, the random component in this stage (specifically in the *Adaptive region formation* step) is a key feature to guarantee proper exploration of the solution space. Because the number of regions (*p*) is determined in this stage, it is important to generate a large number of initial feasible solutions to increase the possibility of finding a large number of regions.

4.2. *Local search*

The local search algorithm iteratively modifies an initial feasible solution by moving areas from one region to a neighboring region. This type of move was first applied by Nagel (1965) and is widely utilized in the literature. Duque *et al.* (2007) summarized other types of moves.

To guarantee a good exploration of the solution space, we provide the local search with the following features: first, the candidate areas that move between regions are selected at random. Second, we use a local Tabu search that allows temporary worsening of the objective function as a strategy to escape from local optimal solutions.

When an area is swapped from one region to another, the area–region relationship is altered; a corridor region could become a max-p region or vice versa. Finally, a unique feature of the Tabu search within the context of the NMPR problem regards local movements that modify a region intersecting the road network; these force the algorithm to evaluate the objective function under two scenarios, corridor and max-p regions, for both the donor and the recipient regions.

Pseudocode 1. Generate initial feasible solutions.

```
P = {} Candidate partition set
for i = 1 to n:
  P = Initialize a partition P with areas G.
  S_U = emptylist; unsigned areas
  S_E = emptylist; enclaved areas
Area filtering:
  for each g_i in G do:
    if a_ti ≥ threshold:
        e_k = the nearest edge to l_i
        R_k = create a corridor region if e_k intersects with g_i,
          if not, create a max-p region
        add g_i to R_k
      else:
        add g_i to S_U
Adaptive region formation:
  while S_U ≠ ∅:
    g = randomly select an area from S_U,
      starting from those areas that intersect with the network N
    R_i = Initialize a new region for P, with the seed area g
    add g to R_i, remove g from S_U;
    while l_R < threshold:
      if R_i don't have neighbor areas:
        break;
      g' = R_i.popMinNeighborArea();
      add g' to R_i, remove g' from S_U;
    if a_R < threshold:
      remove R_i from P
      add all areas in R_i to S_E
  P.setEncalvedAreas(S_E)
  P.add(P)
Partition ranking and filtering:
max = P.maxNumRegions()
P = P.filterByNumRegions(max)
Enclaving:
for each P in P do:
  enclave(P, P.getEnclavedAreas())
```

Pseudocode 2 presents the Tabu search procedure. S_b and S_c represent the best solution and the current solution. The candidate moves contain all the moves that will not break the spatial contiguity constraints. These moves are computed using the inner border areas across all regions in the partition. S_d is the temporary best solution from applying a move, and it is chosen if it is not in the Tabu list and if it is better than S_c. This is the design of Tabu search to avoid local entrapments as much as possible. convTabu and tabuLength are the two Tabu search parameters used to control the stop condition and the length of the Tabu list.

Pseudocode 2. Tabu search procedure.

```
S_b, S_c = the current solution
int c = 0
P = {} Tabu solution set
while c≤convTabu
  S_d = ∅
  while true:
    move = select a random move
    if move = ∅:
      break
    S_n = apply the move to S_c, and remove the move from candidate moves
    if S_n = ∅:
      continue
    if tabuList.contains(S_n):
      continue
    if S_n.betterThan(S_c):
      S_d= S_n
      break;
  if S_d = ∅
    break
  if S_d.betterThan(S_b):
    S_b = S_d
  S_c = S_d
  S_c.updateCandidateMoves()
  tabuList.add(S_d)
  if |tabuList|> tabuLength
    remove the first element from tabuList
return S_b
```

5. Experiments

This section analyzes the two network-related parameters, *MaxDisc* and *MaxDist*. The solution quality is also validated through randomized experiments. A real-world example is then given to show the model's scalability. The algorithms were implemented in Java and tested on a machine with an i7-4710HQ Intel CPU and 16G DDR3 memory.

5.1. *Parameter analysis*

The simulated areas were retrieved from sample data in the ClusterPy library for regionalization research (Duque *et al.* 2011a). The data include three datasets with different numbers of areas (n = 529, n = 1024, and n = 2025). The attribute value for each area is generated using a spatial autoregressive process with ρ=0.9. The attribute for the threshold constraint follows a uniform distribution of [10, 15] (https://code.google.com/p/clusterpy/). A network dataset with 16 edges was designed over the

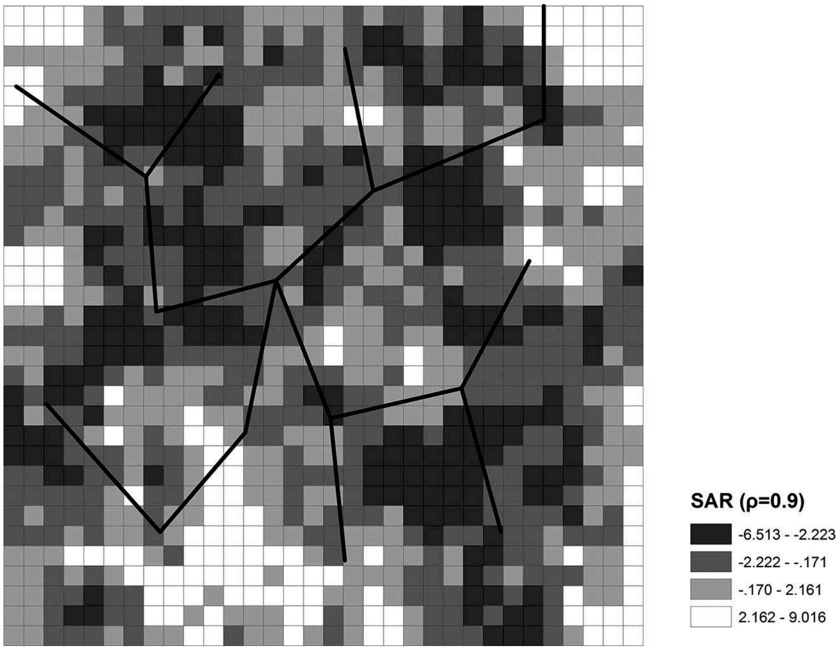


Figure 2. Spatial distribution of one of the simulation datasets (n = 1024).

grid areas. Figure 2 displays the spatial distribution of the heterogeneity attribute and the road network for one of the datasets with n = 1024.

Figure 3 shows one heuristic solution for this dataset. The parameters were set as follows: threshold = 500, number of initial solutions = 10,000, TABU length = 85, *MaxDisc* = 0.3 and *MaxDist* = 10. Corridor regions are marked in dark grey.

To quantitatively assess how MAX_DIST and MAX_DISC affect the results, two measures are proposed: the number of corridor regions, N_c, and the average corridor region compactness, $\bar{C}$. These are defined as follows:

$$\bar{C} = \sum_{R \in \mathbf{R_c}} \frac{c_R}{N_c}, \tag{16}$$

where $\mathbf{R_c}$ is the set of corridor regions of the partition and $c_R = \left(\sum_{j \in R} s_{i,O_i}/N_R - s_{min}\right)/(s_{max} - s_{min})$ is the compactness of a particular corridor region. c_R is a standard measure with the range [0, 1]. s_{min} and s_{max} are the minimum and maximum values of all the area-closest-edge distances.

A series of experiments was performed on a dataset with n = 529 to evaluate the impacts of MAX_DIST and MAX_DISC. The other parameters were set as follows: threshold = 110, number of initial solutions = 10,000, and TABU length = 85. Figures 4 and 6 illustrate the results of this comparison and emphasize that MAX_DIST has major effects. For a given value of MAX_DISC, both $\bar{C}$ and N_c increase with MAX_DIST. In Figure 4(a), as MAX_DIST grows, the discount function has a positive sign over a greater portion of the map, which implies a reduction in the

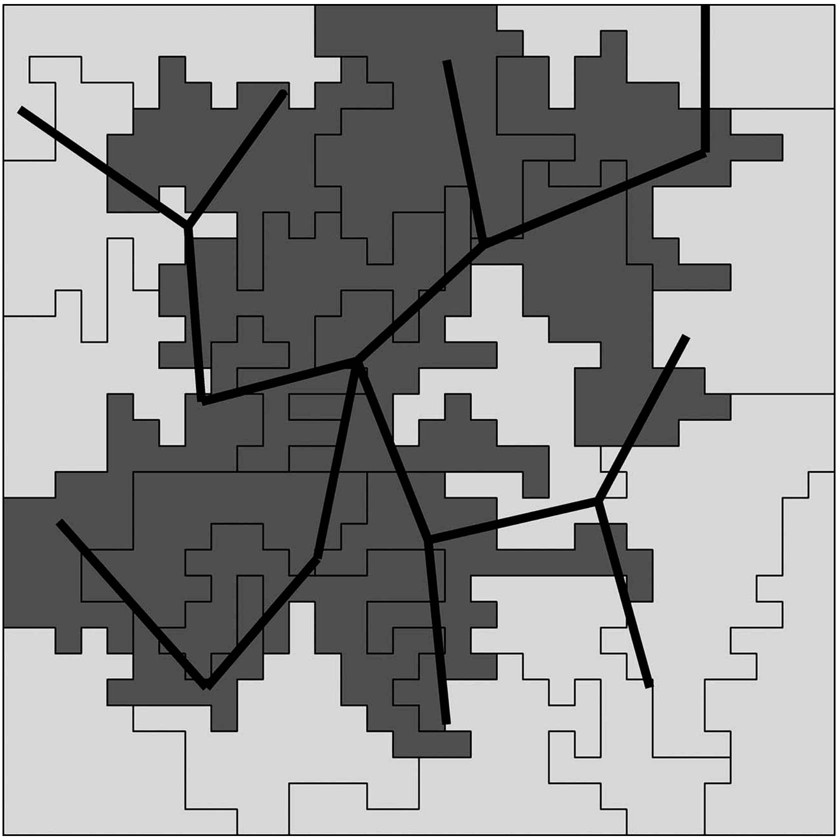


Figure 3. Regionalization result visualization.

objective function via the third term. Thus, the capacity for constructing corridor regions is higher. The speed of growth slows as MAX_DIST approaches 20 and then stabilizes. This indicates that the measure has reached a threshold where no further benefit can be gained through the third term with larger values of MAX_DIST. The same rationale applies to the compactness measure in Figure 4(b): small MAX_DIST values force the corridor regions to be very close to the road (i.e., they become more elongated). Conversely, larger MAX_DIST values force areas far from the road to become part of a corridor region. Again, there is a certain point after which MAX_DIST produces no more benefits in terms of the OF value.

As shown in Figure 5(a,b), MAX_DISC does not differ significantly for a given value of MAX_DIST but remains relatively stable. This suggests that in real applications, what matters is how far from the road we extend the premium for corridor regions (i.e., MAX_DIST). Given the linearity of the discount function, it is straightforward for practitioners to relate MAX_DIST to the parameters in domain-specific problems.

The following experiment is conducted to illustrate the solution quality of the heuristic algorithm. For the same parameters of the problem shown in Figure 3, 1000 random solutions are generated (not improved by the heuristic algorithm). Then, 20 heuristic solutions are generated and compared with those random solutions. Because

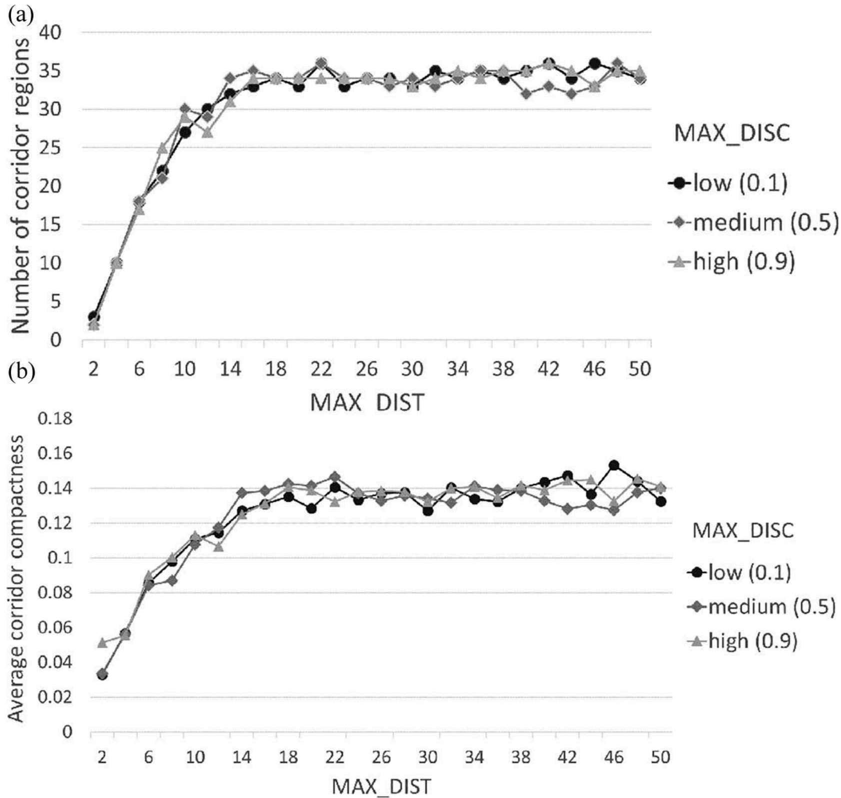


Figure 4. Parameter analysis of MAX_DIST with three different MAX_DISC values: low (0.1), medium (0.5), and high (0.9). (a) Number of corridor regions. (b) Average corridor region compactness.

the number of regions dominates other terms in the objective value function, the random solutions and the heuristic solutions are generated for a fixed number of regions k. This is achieved by generating as many random solutions as possible until 1000 random solutions are produced with the number of solutions equal to k, and the same process is repeated to produce the 20 heuristic solutions.

The result is shown in Figure 6 (with the number of regions = 20). The histogram represents the distribution of the objective values for the 1000 random solutions, and the vertical lines indicate where the objective values of the heuristic solutions are located. The vertical lines are clear outliers of low values against the histogram, which shows that the quality of the heuristic is reasonable.

5.2. *Case study: Wuhan (China)*

Wuhan is a rapidly growing city and is the largest city in central China. Its modernization process is reflected by the expansion of the street network. By 2014, the total road length had reached 14,520 kilometers, with a road density of 180.85 kilometers per 100 square kilometers (http://www.whtj.gov.cn/details.aspx?id=2513). By integrating these networks into the regionalization process, the NMPR formulation has practical implications because the development of the urban street network in

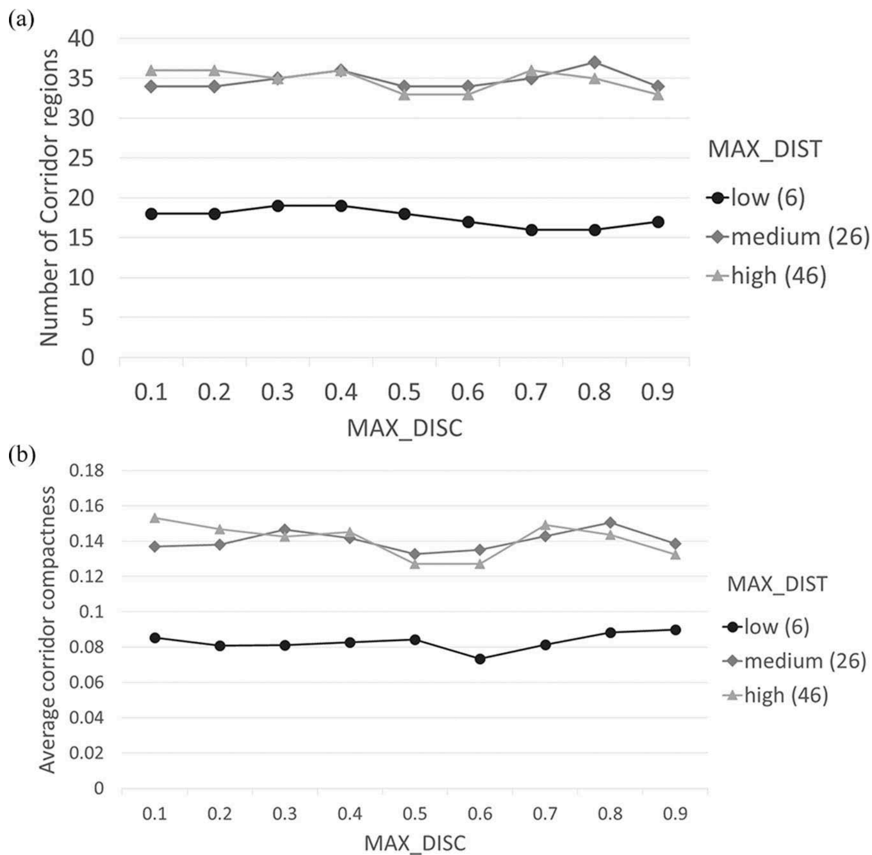


Figure 5. Parameter analysis of MAX_DISC with three different MAX_DIST values: low (6), medium (26), and high (46). (a) Number of corridor regions. (b) Average corridor region compactness.

Wuhan has led to greater accessibility among areas and rapid development along the streets.

We use the Wuhan grid-level data with its main street network. There are 7766 one-kilometer grids and 386 network edges. The population from the 2000 census was used as the aggregation variable, and the threshold was set to 100,000 people. The urban land use area was chosen as the heterogeneity variable. As shown in Figure 7, urban land use is largely clustered in the center and scattered along the roads far from the center. Some roads act as regional borders, such as the ring roads around the central area. Areas around the roads that expand outwards largely show similar attributes, regardless of their distance from the roads. The core of the city has a dense street network occupied by many urbanized areas.

We ran both the max-p-regions and the NMPR models. The number of initial partitions was set to 200, the threshold value was 100,000 people, and the Tabu length was 85. For the NMPR formulation, MAX_DISC was set to a moderate value of 0.5, and MAX_DIST was set to 2 kilometers. The actual values of the parameters, including MAX_DISC, MAX_DIST and the threshold, should be based on physically meaningful values determined by practitioners.

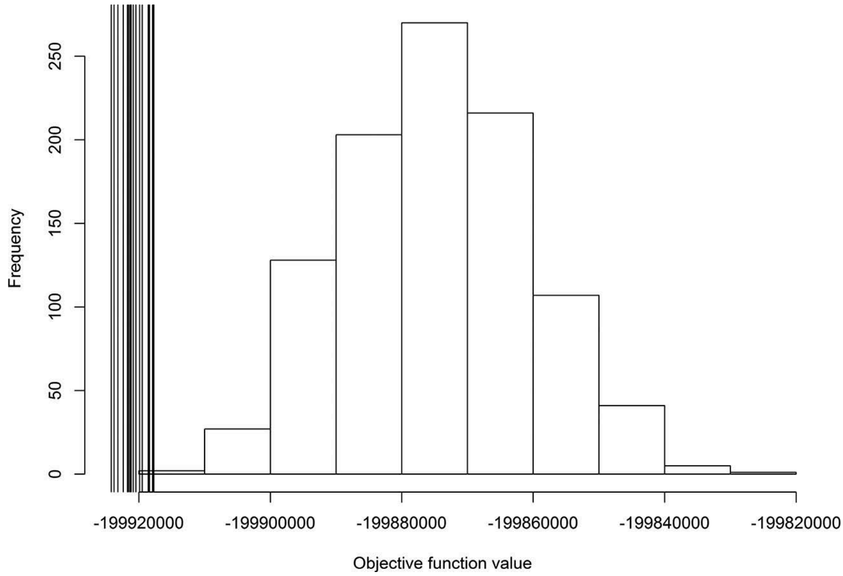


Figure 6. Solution quality analysis for the heuristic algorithm.

Figure 8 compares the resulting partitions of the max-p-regions and NMPR models. The max-p-regions procedure took 495.3 s to finish, while the NMPR procedure took 935.5 s. This means that incorporating the network can still lead to reasonably good performance for a dataset of moderate size. Both partitions generated 66 regions, but the NMPR partition included 43 corridor regions (dark grey). The result shows a clear pattern of corridor regions in the NMPR formulation. The corridor regions are largely aggregated in the urban center, where the street networks are also clustered. These areas are more likely to benefit from the formation of corridor regions, producing lower objective function values due to a higher value of the third term.

6. Discussion and conclusions

The NMPR model is a novel approach to incorporate the street network structure into a regionalization problem. This is the first regionalization model that allows both corridor regions and max-p-regions. The model would be useful in a wide array of applications, including land use planning, districting, and other areas within the domain of geographical information science. The NMPR model allows researchers and practitioners to explicitly take the structure of the street network into account, balancing the formation of corridor regions and regional heterogeneity.

The MAX_DIST parameter in the objective function affects the formation and features of the corridor regions in a near-logarithmic fashion. MAX_DIST influences the partitions through the objective function in a linear way. Compared to more complex measures, such as the tortuosity of network segments, it is a simple parameter; however, it performs surprisingly well. Parameter analysis reveals the bounded

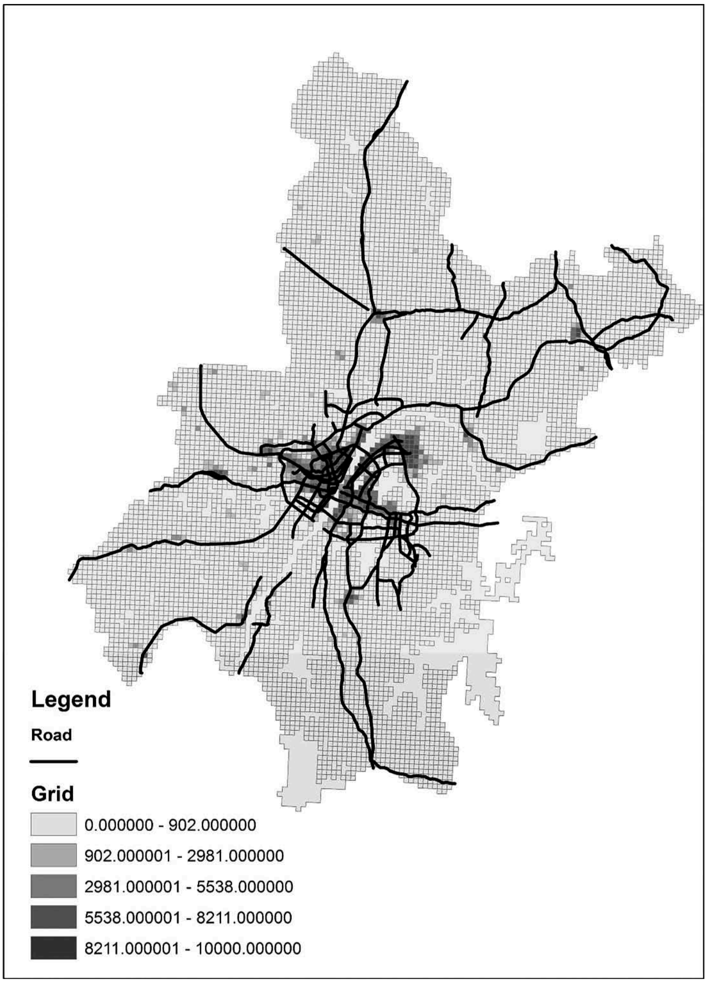


Figure 7. Spatial pattern of urban land use at the grid level of Wuhan overlaid with the main roads.

characteristics of this parameter, which could serve as guidelines for practitioners to modify it.

This study assumes a homogeneous network and only considers geometric network properties. Future work could investigate stricter formulations for the network. For example, an additional restriction could be that edges in a corridor region must be a connected subnetwork. In addition, the street structure contains rich semantic information, including rank, capacity, and traffic load. These attributes can be integrated into the objective function.

This paper establishes a framework for network regionalization problems. The network structure is incorporated into the exact formulation without sacrificing its rigor, while the heuristic algorithms provide an efficient method for regional growth testing and solution optimization. Both the formulation and the heuristic can be transferred to other optimization problems. This would greatly facilitate the regionalization applications.

Figure 8. Regionalization result visualizations for the Wuhan grid data. (a) Max-p-regions formulation. (b) NMPR formulation.

Acknowledgments

The authors are thankful to the editor and anonymous referees for their comments and suggestions that improved this paper. The usual disclaimer applies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation: [Grant Numbers 1416509 and 1637242] and the National Natural Science Foundation of China: [Grant Number 41329001].

References

Aboolian, R., Sun, Y., and Koehler, G.J., 2009. A location–allocation problem for a web services provider in a competitive market. *European Journal of Operational Research*, 194 (1), 64–77. doi:10.1016/j.ejor.2007.11.057

Blais, M., Lapierre, S.D., and Laporte, G., 2003. Solving a home-care districting problem in an urban setting. *Journal of the Operational Research Society*, 54 (11), 1141–1147. doi:10.1057/palgrave.jors.2601625

Borgatti, S.P., *et al*., 2009. Network analysis in the social sciences. *Science*, 323 (5916), 892–895. doi:10.1126/science.1165821

Bozkaya, B., Erkut, E., and Laporte, G., 2003. A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, 144 (1), 12–26. doi:10.1016/S0377-2217(01)00380-0

Bunge, W., 1966. Gerrymandering, geography, and grouping. *Geographical Review*, 56, 256–263. doi:10.2307/212882

Clarke, S. and Surkis, J., 1968. An operations research approach to racial desegregation of school systems. *Socio-Economic Planning Sciences*, 1 (3), 259–272. doi:10.1016/0038-0121(68)90014-1

Cova, T.J. and Church, R.L., 2000. Contiguity constraints for single-region site search problems. *Geographical Analysis*, 32 (4), 306–329. doi:10.1111/j.1538-4632.2000.tb00430.x

Curtin, K., Hayslett-Mccall, K., and Qiu, F., 2010. Determining optimal police patrol areas with maximal covering and backup covering location models. *Networks and Spatial Economics*, 10 (1), 125–145. doi:10.1007/s11067-007-9035-6

Duque, J.C., *et al*., 2011a. *ClusterPy: Library of spatially constrained clustering algorithms, Version 0.9.9* [online]. RiSE-group (Research in Spatial Economics). EAFIT University. Available from: http://www.rise-group.org

Duque, J.C., Anselin, L., and Rey, S.J., 2012. The max-p-regions problem. *Journal of Regional Science*, 52 (3), 397–419. doi:10.1111/jors.2012.52.issue-3

Duque, J.C., Church, R.L., and Middleton, R.S., 2011b. The p-regions problem. *Geographical Analysis*, 43 (1), 104–126. doi:10.1111/j.1538-4632.2010.00810.x

Duque, J.C., Ramos, R., and Surinach, J., 2007. Supervised regionalization methods: a survey. *International Regional Science Review*, 30 (3), 195–220. doi:10.1177/0160017607301605

Hess, S.W. and Samuels, S.A., 1971. Experiences with a sales districting model: criteria and implementation. *Management Science*, 18 (4–part–ii), P-41–P-54. doi:10.1287/mnsc.18.4.P41

Jiang, B. and Claramunt, C., 2004. Topological analysis of urban street networks. *Environment and Planning B: Planning and Design*, 31 (1), 151–162. doi:10.1068/b306

Jiang, B. and Jia, T., 2011. Agent-based simulation of human movement shaped by the underlying street structure. *International Journal of Geographical Information Science*, 25 (1), 51–64. doi:10.1080/13658811003712864

Kim, H., Chun, Y., and Kim, K., 2013. Delimitation of functional regions using a p-regions problem approach. *International Regional Science Review*, 38 (3), 235–263. doi:10.1177/0160017613484929

Laura, J., *et al*., 2015. Parallelization of a regionalization heuristic in distributed computing platforms–a case study of parallel-p-compact-regions problem. *International Journal of Geographical Information Science*, 29 (4), 536–555. doi:10.1080/13658816.2014.987287

Li, W., Church, R.L., and Goodchild, M.F., 2014a. An extendable heuristic framework to solve the p-compact-regions problem for urban economic modeling. *Computers, Environment and Urban Systems*, 43 (0), 1–13. doi:10.1016/j.compenvurbsys.2013.10.002

Li, W., Church, R.L., and Goodchild, M.F., 2014b. The p-compact-regions problem. *Geographical Analysis*, 46 (3), 250–273. doi:10.1111/gean.2014.46.issue-3

Nagel, S.S., 1965. Simplified bipartisan computer redistricting. *Stanford Law Review*, 17, 863–899. doi:10.2307/1226994

Openshaw, S., 1977. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers*, 2, 459–472. doi:10.2307/622300

Openshaw, S. and Rao, L., 1995. Algorithms for reengineering 1991 census geography. *Environment and Planning A*, 27 (3), 425–446. doi:10.1068/a270425

Ricca, F. and Simeone, B., 2008. Local search algorithms for political districting. *European Journal of Operational Research*, 189 (3), 1409–1426. doi:10.1016/j.ejor.2006.08.065

Rossiter, D.J. and Johnston, R.J., 1981. Program GROUP: the identification of all possible solutions to a constituency-delimitation problem. *Environment and Planning A*, 13 (2), 231–238. doi:10.1068/a130231

Thomas, J.W., 1979. Techniques for defining geographic boundaries for health regions. *Socio-Economic Planning Sciences*, 13 (6), 321–326. doi:10.1016/0038-0121(79)90013-2

Thoreson, J.D. and Liittschwager, J.M., 1967. Computers in behavioral science. Legislative districting by computer simulation. *Behavioral Science*, 12 (3), 237–247. doi:10.1002/(ISSN)1099-1743

Vickrey, W., 1961. On the prevention of gerrymandering. *Political Science Quarterly*, 76, 105–110. doi:10.2307/2145973

Wheeler, J.O., 1973. Social interaction and urban space. *In*: M. Albaum, ed. *Geography and contemporary issues: studies of relevant problems*. New York: John Wiley & Sons, 303–308 (reprinted from Journal of Geography).

Whitehand, J.W.R. and Larkham, P.J., 1992. *Urban landscapes: international perspectives*. London: Routledge Chapman & Hall.

Zoltners, A.A. and Sinha, P., 1983. Sales territory alignment: a review and model. *Management Science*, 29 (11), 1237–1256. doi:10.1287/mnsc.29.11.1237