Journal of Geovisualization and Spatial Analysis (2018) 2:4
https://doi.org/10.1007/s41651-018-0013-y

Exploring Regionalization in the Network Urban Space

Xinyue Ye [1,2] · **Bing She** [3,4] · **Samuel Benya** [2]

Abstract
Isotropic homogeneity does not hold in urban areas. Street networks exert a great influence on human mobility. As a result, city structure is largely shaped by this network, especially the streets that carry a higher volume of traffic. In practice, small areas along network edges often need to be grouped into regions for management purposes. This work formalizes the extension to the *P*-regions problem that takes the network as the underlying constraint and proposes a heuristic-based approach to solve the problem to near optimality. The network is subdivided into aggregator edges that attract regions and separator regions that divide areas apart. Two types of regions emerge in the region formation process: regions that grow along a certain network edge (network regions) and regions that grow from areas that are far away from all the network edges (planar regions). The heuristic solution effectively uses pre-computed spatial contiguity and distance matrices. The global objective function consists of the original heterogeneity factor and the discounted network proximity factor. This approach is elaborated with both a simulated and a real-world dataset. The regionalization results help design, study, and service regions that explicitly consider the network configuration with flexible parameter controls.

Keywords *P* regions · Network constraints · Streets · Regionalization · Heuristics · China

Introduction

Network science has become an increasingly popular area of research (Borgatti et al. 2009). Streets, which are one of the most common types of physical networks, constrain various human activities and are influenced by social and economic forces (Wheeler 1973; Whitehand and Larkham 1992; Jiang and Claramunt 2004; Parthasarathi et al. 2014; Huang et al. 2015, Zhao et al. 2017). The network space imposes physical constraints on how people act and interact (Jiang and Jia 2011; Porta et al. 2014). Many planar spatial analysis methods have been introduced into the network space over the last decade (Okabe and Yamada 2001; Yamada and Thill 2007; Okabe et al. 2008; Xie and Yan 2008; Okabe et al. 2009; Yamada and Thill 2010, Eckley and Curtin 2013; She et al. 2015). Research shows that planar spatial analysis methods can produce false alarms regarding the clustering of points distributed along a network (Yamada and Thill 2004). Regionalization is the process of segmenting the planar areal units into several spatially adjacent areas, given a set of constraints. Duque et al. (2011b) proposed the *P*-regions problem, which aims to aggregate small areas into *P* spatially contiguous regions while optimizing certain criteria. Based on this work, recent efforts in regionalization research include the max-*P*-regions problem (Duque et al. 2012), the *P*-compact-regions problem (Li et al. 2014b), and the network-max-*P*-regions model (She et al. 2017). The max-*P*-regions problem finds the maximal number of *P* regions, each satisfying a threshold variable, while simultaneously minimizing the heterogeneity. The *P*-compact-regions problem explicitly takes into account the shape compactness for each input region. The network-max-*P*-regions model is a regionalization model that aims to aggregate *n* areas into the maximum number of regions (max-*P*) that satisfy a threshold constraint and also minimizes the heterogeneity while taking into account the influence of a street network.

✉ Xinyue Ye
xye5@kent.edu

Bing She
bingshe@umich.edu

Samuel Benya
sbenya1@kent.edu

1 Department of Geography, Kent State University, Kent, USA

2 Computational Social Science Lab, Kent State University, Kent, USA

3 State Key Lab of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan, China

4 China Data Centre, University of Michigan, Ann Arbor, USA

Published online: 24 January 2018

Regionalization serves the managing practice and research needs in various disciplines. These include school districting, political redistricting, and police patrol-area partitioning (Lemberg and Church 2000; Barkan et al. 2006; Curtin et al. 2010). Furthermore, the underlying socioeconomic processes are greatly influenced by streets, especially in the urban space. Therefore, it would be theoretically interesting, as well as practically relevant, to investigate street distribution integration in the regionalization procedure. For example, in some Chinese cities, the original administrative levels are organized by streets. Therefore, the regionalization results would be better if the regions also fit the administrative hierarchies. In such a process, areas are grouped around a certain street segment and grow into a network region. Areas far away from the streets form a planar region. The final partitions consist of a set of network regions and a set of planar regions. In addition, there would be edges that separate areas of its two sides, such as ring roads. A proper network-constrained regionalization process could help researchers define regions that focus on the street-related phenomenon. This also helps practitioners design regions that are more naturally suited to underlying human activities. In practice, the street network contains either a large quantity of short roads inside regions or carries relatively little traffic. For this reason, only main roads are chosen for the partition. Thus, the network is much sparser compared to the original road network. While a region may belong to multiple edges, this work considers the case that a region only has one edge. This is common because practitioners would normally use major roads to partition management regions.

She et al. (2017) undertook a study to extend the max-*P*-regions problem into the network space. The following are the main differences between that study and the current research: (1) the *P*-region problem is used in this research instead of the max-*P*-regions problem; (2) in this research, the network edges are categorized into aggregators and separators. Aggregator edges attract areas around it to form network regions, while separators act as barriers to the areas around it. The areas intersected with separators are barrier areas that are excluded from the partition process; (3) this paper only allows one edge per network region; and (4) the heuristic algorithms are much different because P is fixed in the *P*-regions problem (this paper), but varies in the max-*P*-regions problem in She et al. (2017). Due to the differences in model assumptions and formulations, these two models, as well as other extensions such as the max-*P*-region problem and *P*-compact region problem, are not directly comparable in terms of efficiency or effectiveness. The main rationale for this work is to consider both aggregator and separator edges in the network regionalization process. Also, the edges in this work are assumed to be main roads; therefore, a region is only allowed to have one edge.

A challenge in network-constrained regionalization is the issue of how to effectively generate better initial partitions. To solve this, this paper adopted a heuristics-based approach in the optimization process by first generating a set of initial solutions relying on network constraints. After generating these initial solutions, the next step was to improve these solutions by using a local search algorithm. The TABU search algorithm is commonly used in regionalization research due to its capacity to lower the chances of getting stuck in local optimums (Duque et al. 2012; She et al. 2017). Therefore, this work also used the TABU search algorithm in the local search stage. The global objective function for evaluating solutions considers the number of regions, a heterogeneity factor, and a network proximity factor. To efficiently solve the problem, the spatial matrix structure that holds data for both area-area and network-area contingency and distance matrices was pre-computed. In the local search stage, a status update algorithm was designed to efficiently update valid candidate moves for a partition.

This work presents an instance of network-constrained regionalization that extends the *P*-regions problem. The network-constrained regionalization can assist researchers and practitioners in designing homogeneous regions (Fischer 1980) that explicitly consider the network configuration with flexible parameter controls. A "Literature Review" is given, while the "Methodological Framework" is established. The section "Heuristic Solution Implementation" describes the heuristic framework in detail. The section "Applications" illustrates this solution through a simulated dataset and real-world data for Wuhan, China. The final section discusses the characteristics and future works of the proposed methods.

Literature Review

Regionalization has special kinds of clustering problems that are constrained by spatial contiguity (Gordon 1996; Hansen et al. 2003). Such contiguity generally makes the problem harder to abstract and model (Tong and Murray 2012). Sometimes, locally defined constraints may produce regions of varying shapes which have less shape compactness. As a spatial optimization problem, attributes of regions, such as population or economic indicators, are considered in the regionalization process. Depending on the problem specification, these attributes and the shape compactness can all be integrated into the global objective function. In practice, researchers often do not know how many regions are necessary. Instead, they have domain knowledge about the local constraints on regions. Duque et al. (2012) made a specific contribution in this regard that makes the region growing process adaptive and produces a maximum number of regions, given that the constraints are satisfied. The assumption held in this paper is that researchers also have difficulties integrating the physical street network into the region building process. Intuitively, researchers often want the regions to follow the street edge whenever possible, given that other constraints satisfied. This trade-off can be reflected in a multi-objective function design.

The street network can affect a spatial problem in multiple ways. The most common perspective is to replace the Euclidean distance measures with network distances (Miller and Wentz 2003). Spatial optimization models often treat network distances between locations either as a quantitative measure such as in a point location model (Church and Murray 2008b) or as a virtual link network that is transformed from relationships between neighboring objects. This is useful in cases such as siting a corridor using flow constraints (Church and Murray 2008a). Network flow attributes are also considered in research during the regionalization process (Whiteaker et al. 2007). This work takes a different perspective by explicitly taking the geometric properties of a network into the objective function of the regionalization process. This specific component of the objective function can be seen as minimizing the degree of spatial adjacency (Tong and Murray 2012).

Street structure is often related to connectivity. As business activities often occur on the street, areas can be grouped based on the attraction from street edges. From this point of view, a street edge can be seen as connecting a set of areas into a network region. The connectivity of a network region is less developed in regionalization research but appears more often in research in transportation modeling (Li et al. 2014a). Shape factors are considered extensively in the literature and are often related to a specific application setting (Shirabe 2005; Williams et al. 2004). Li et al. (2013) introduced the compactness measure of a shape based on the moment of inertia and presented the mathematical formalization. Shape compactness affects the regionalization process by taking into account the intrinsic geometric properties of a single shape. In contrast, the network constraint poses a local constraint on the resulted shape of a region.

Regionalization problems can often be formed as integer programming problems. For example, Duque et al. (2011b) presented three model formulations, a tree-based model, an order-based model, and a network-flow-based model. These models explicitly embed the spatial contiguity constraints into the integer programming framework. The mixed-integer-programming formulation can be difficult to construct and are computationally intensive (Duque et al. 2012). One advantage of using heuristic procedures is that the heuristic algorithms themselves are relatively independent of constraints. This allows contiguity testing procedures to be more flexible in their construction and computation, therefore making them relatively easy to integrate. Thus, this work avoids considering the order of assignment or flow in our model formulation. Rather, the contiguity is ensured by separate procedures invoked by the heuristic algorithms.

Heuristic solutions for regionalization problems typically embed an existing heuristics algorithm for improving the partitions. Frequently used ones include simulated annealing (D'Amico et al. 2002; Bergey et al. 2003; Ricca and Simeone 2008), Greedy Randomized Adaptive Search Procedure (Feo and Resende 1995; Brás et al. 2013), and TABU (Tung and Chou 2002; Bozkaya et al. 2003). Between the generation of initial partitions and improving them, some heuristic solutions choose to store information in an external data structure, which helps accelerate the local search algorithms. For example, the MERGE algorithm in Li et al. (2014b) memorizes the candidate plans for a given region and therefore allows the region-swapping procedure to be more targeted and avoids repeated computation in the objective value function. This could be categorized as a specialized example of spatial memorization (Hardisty and Klippel 2010).

Methodological Framework

The network-constrained P-regions problem is based on the P-regions problem from the planar space. The original P-regions problem aims to minimize regional heterogeneity in the final solution. The network-constrained P-regions problem aims to integrate the distance factor into the heterogeneity computations. In other words, the shapes of the created regions will be influenced not only by the heterogeneity of its areas but also the closeness to a certain aggregator edge. Moreover, the final solution is constrained by the separator edges, i.e., the created regions will be separated by these particular edges.

Basic Concepts

(1) Areas

Areas are the basic unit which consists of a set of attributes that are grouped into regions in the final partition. This work dealt with grids in the experimental section; thus, a set of grid areas is used, denoted as $G = \{g_1, g_2, \ldots, g_n\}$.

(2) Network

The network in this work refers to an undirected and planar network $N = (V, E)$, formed by a set of nodes V and edges E. The edges E consist of two parts: the aggregator edges E_a and the separator edges E_s. The network regions are grown around aggregator edges, while the separator edges will separate regions away. This work models separator edges as physical barriers between regions. The areas crossing separator edges are called separator areas and are removed from the final partition process.

(3) Regions

Regions are aggregated by areas. Since the urban space has a comparatively sparse network, two types of regions naturally emerge in terms of growth: regions that grow along a certain network aggregator edge (network regions R_{N}) and regions that grow from areas that are far away from all the network aggregator edges (planar regions R_{P}). The type of a region is determined in the initialization stage. A condition necessary for a network region is that it must intersect with at least one network edge. An edge e is selected as the root edge of the network region. The root edge might get changed during the optimization process when regions are randomly swapping areas.

(4) Attributes

The attributes used in the study consist of a set of attributes $A_h = \{ a_{h1}, a_{h1}, \ldots, a_{hm}\}$ that are used in the heterogeneity calculation. All of these attributes are spatially extensive.

The dissimilarity d_{ij} between two areas i and j is used in the heuristic procedures when regions greedily choose candidate areas or regions start to swap areas. To simplify the case, this work only uses one attribute a_h for the heterogeneity calculation in the experimental section and chooses the absolute difference $|a_{hi} - a_{hj}|$ to represent d_{ij}.

The heterogeneity of a region ht(R) is reflected in the dissimilarity of its areas. With the absolute difference representing dissimilarity, it is written as follows:

$$\mathrm{ht(R)} = \sum_{i,j:g_i\in R, g_j\in R, i<j} \left|a_{hi} - a_{hj}\right| \tag{1}$$

(5) Spatial matrices

The spatial matrices in the network-constrained P-regions problem store information about area-area and area-edge relationships. There are two types of matrices: spatial contiguity and spatial distance. The spatial contiguity matrices are composed of two binary matrices **C** and **L**. **C** is constructed so that it records area-area neighboring relationships; it is used to ensure that all areas are connected in a region. The separator areas are removed from **C**. **L** records the edge-area intersection information, whereas each element l_{ij} in **L** records whether an edge e_i interests with an area g_j. This paper uses only one distance matrix **D** to store the area-edge distances. Each element d_{ij} in **D** records the nearest distance of an edge e_i to an area g_j. Since such information will stay unchanged during the optimization procedures, pre-computations of this information are beneficial in regard to execution speed. Both **C** and **L** could be stored as sparse matrices. Although **D** is a dense matrix, it is still of moderate size. This is because the number of edges is considerably smaller than the number of areas.

(6) Network proximity

The network regions grow along a certain aggregator edge. However, alongness itself is a vague spatial relationship for R_N. Therefore, a more appropriate interpretation would be that areas in a network region are more proximal to its root edge relative to neighboring regions, satisfying other constraints.

This work models the network proximity as a discounting factor to the heterogeneity factor for network regions. In other words, network regions get rewarded when they are growing along a network edge. If the areas added are too far away from the edge, the rewards may turn into penalties. The proximity function is defined for a pair of areas with regard to a certain aggregator edge:

$$d\left(g_i, g_j, R\right) = scale*\left(1-e^{D_{i,j,R}-extent}\right)$$
$$D_{i,j,R} = \left(d_{iR} + d_{jR} + d_{ijR}\right)/3$$

The function $d(g_i, g_j, R)$ represents the discounting factor of areas g_i and g_j in region R. The *scale* and *extent* are two parameters that control the influence of network proximity. As shown in Fig. 1, the extent factor represents the threshold when the function turns from reward into penalty. The scale factor controls the magnitude of influence. When $D_{i,\ j,\ R}$ is smaller than *extent*, the reward is bigger with a larger *scale* value. When $D_{i,\ j,\ R}$ is larger than *extent*, the penalty is also bigger with a larger *scale* value.

The distance measure $D_{i,\ j,\ R}$ is the average of three distance sub-measures that aims to quantify the alongness of areas to a network edge. d_{iR} and d_{jR} are the closest distances between areas g_i and g_j to the root edge of region R. d_{ijR} represents the closest distance between a virtual edge e_{ij}, which links the centroids of areas i and j to the root edge of region R. The representation of $D_{i,\ j,\ R}$ is simple and easy to compute.

Model Objectives

A valid partition P is a set of valid regions that cover all the input areas. Regions should be non-overlapped in terms of the assigned areas. The goal is to find a valid P with a minimal objective function value. The objective function O refines the original objection function to minimize heterogeneity across all regions that are discounted by the network proximity factor in network regions.

$$O = \sum_i \sum_{j|j>i} d_{ij} t_{ij} - \sum_i \sum_{j|j>i} d_{ij}\ d\left(g_i, g_j, R\right) \sum_{R=0}^{n} n_{ij}^{R} \tag{2}$$

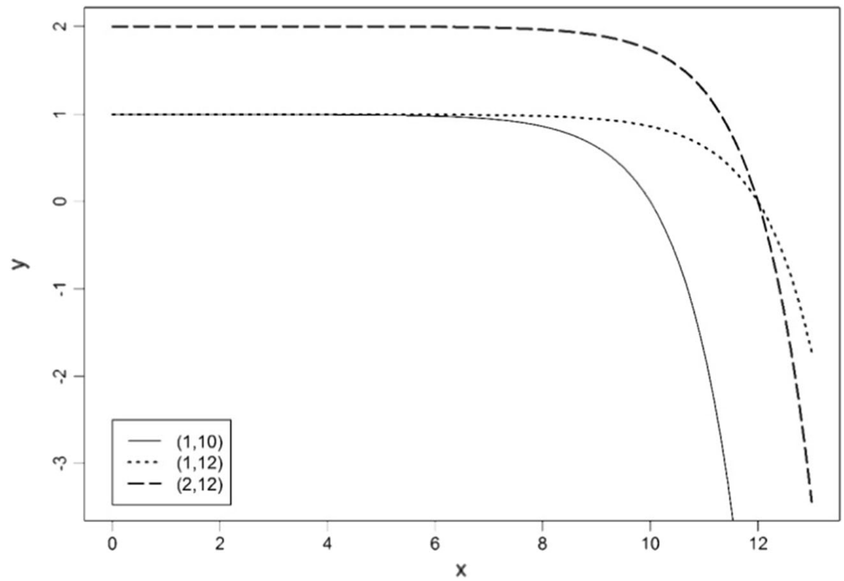

Fig. 1 Examples of different parameter combinations: (scale, extent)

t_{ij} is the decision variable that decides whether areas i and j belong to the same region R. To integrate the network proximity factor, two new decision variables o^R and n_{ij}^R are constructed as follows:

$$o^R = \begin{cases} 1, \textit{if region } R \textit{ is a network region}, \\ 0, \textit{otherwise} \end{cases}$$
$$n_{ij}^R = \begin{cases} 1, \textit{if areas } i \textit{ and } j \textit{ belong to the same network region } k, \textit{with } i < j, \\ 0, \textit{otherwise} \end{cases}$$

The function is subject to the same set of constraints of the original P-regions problem. This work adds three more conditions:

(1) A region R may become a valid network region if at least one of its areas intersects an edge of the network. In other words:

$$\sum_{i \in R} \sum_{e \in E_a} l_{i,e} > 0$$

$l_{i,\,e}$ is an element in the edge-area intersection matrix $\mathbf{L}$.

(2) For a valid network region R, the root edge of R is the edge that interacts with the most number of areas in R, among all edges that intersect with R.

$$e_R = \underset{e}{argmin} \sum_{i \in R} l_{i,e}$$

(3) A valid network region R is counted as a true network region only if the network proximity factor is positive. Otherwise, it will remain as a planar region.

$$\sum_i \sum_{j|j>i} d_{ij}\, d\left(g_i, g_j, R\right) \sum_{R=0}^{n} n_{ij}^R > 0$$

The problem is clearly highly nonlinear and would need a heuristic algorithm to solve it in a reasonable amount of time. The following section introduces the implementation of the heuristic solution.

Heuristic Solution Implementation

Solving a regionalization problem based on the heuristic solution generally consists of three steps:

(1) Data input and transformation: this stage involves constructing a proper data structure to hold the attribute and spatial data. Typically, a set of area objects are initialized. The spatial contiguity matrix is static; thus, it is often precomputed at this stage to avoid repeated computation at a later phase. Model developers often choose a specific contiguity type to construct the contiguity matrix. The two most common contiguity types are queen and rook. Queen defines an area's neighbor with either shared borders or vertices, while rook defines contiguity as an area's neighbor with only shared borders.

(2) Initial partition generation: this stage constructs a large number of initial partitions with the randomized greedy strategy. Li et al. (2014b) described this process as *dealing* since it resembles the strategy of a card dealer assigning cards. The initial partitions largely constrain the final partition, given that local search algorithms still have limits in terms of speed and power. Thus, better and more diversified initial partitions will likely produce a final partition with lower objective values, speeding up the local search phase.

(3) Local search: as a general algorithmic technique, a local search algorithm is fairly independent of the

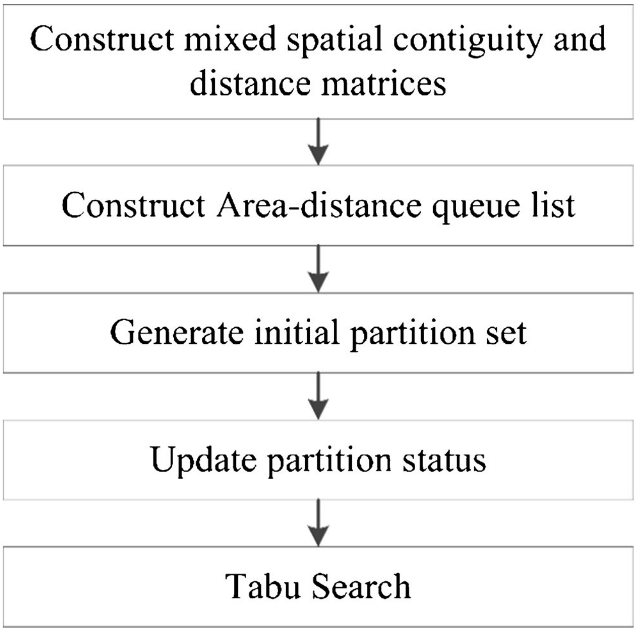

Fig. 2 Flowchart of solving the network-constrained *P*-regions problem

regionalization algorithm that invokes it. Three procedures are provided from specific regionalization algorithms to the local search algorithm: candidate move selection, partition validation and update, and objective value calculation.

In a sense, an object-oriented paradigm is adopted in the optimization procedure design. Areas and spatial contiguity matrices are immutable. A region R_i is mutable in its containing areas GR_i and neighboring areas NGR_i for candidate local search moves. Whenever an area is added to R_i, NGR_i is also updated by taking this area's neighbors in the spatial contiguity matrix **C**. A partition P_i is mutable in its containing regions RP_i and area-region pairs ARP_i. Each pair in ARP_i consists of an area and a neighboring region to it that represent a valid move. These moves are then randomly selected at the local search phases. A new partition is then formed through object cloning and status updates from the old partition.

Specifically, the adapted three-stage process of solving the network-constrained *P*-regions problem is illustrated in Fig. 2. In the data reading stage, additional information about the network structure is put in a similar structure to accelerate later computations. The spatial contiguity and distance matrices discussed in the section "Basic Concepts" are used both in the initial partition generation and local search. In terms of internal data updates, this work chooses to store decomposed parts of the objective function and neighboring areas in individual regions. This memorization serves the purpose of accelerating the local search stage. After generating a set of initial partitions, each partition will pre-compute the status regarding area-region pairs for valid moves. This work chose the TABU search as the local search algorithm that limited the

Table 1 Initial partition generation procedure

```
P = initialize a partition P with areas G
S_U = all areas in G; unsigned areas
E = all aggregator edges
Separator areas filtering:
Remove all separator areas from S_U
Region seeding:
For i from 1 to p:
    e = randomly choose and remove an edge from E
    g=chooseAreaFromMatrixL(edge)
    If g ! = ∅ then:
        R= Initialize a new network region in P with e as its root edge;
        Add g to R, remove g from S_U;
    Else:
        g=randomly choose an area from S_U
        R= Initialize a new planar region in P
        Add g to R; remove g from S_U;
Region formation:
While S_U ≠ ∅ :
    For each R_i in P regions:
        If R_i do not have neighbor areas:
            Break;
        g = R_i.popMinNeighborCandidate();
        Add g to R_i, remove g from S_U;
```

moves to area swapping between regions. The moves are generated using a randomized approach.

Initial Partition Generation

The initial partition generation stage is illustrated in Table 1, consisting of three steps.

(1) Separator areas filtering: this stage removes all areas that intersect with the aggregator edges *E*.

(2) Region seeding: this stage randomly generates *P* regions. Because the aggregator network *E* is sparsely distributed the space, the seeding process will favor regions that start as a network region. In detail, for each edge *e* in *E*, the *chooseAreaFromMatrixL* will randomly choose an unassigned area *g* from the edge-area intersection matrix **L**. *g*, with smaller distance to the centroid of *e*, is selected first. Then, a network region will be created with *e* as its root edge. If there are no more edges to select from or no more candidate areas that intersected with the remaining edges in E, *g* will be randomly selected from S_U and a planar region will be created from it.

(3) Region formation: the region then grows by taking its neighboring areas *NGR* through a greedy function *popMinNeighborArea*. This function will greedily choose a neighboring area with a minimum combined score of heterogeneity and distance to this newly constructed region. This function will use a greedy strategy that protects network regions from growing too far away from their root edge. For a network region, the *popMinNeighborArea* will favor areas in *NGR* that intersect with the root edge of *R* first. If there are no such areas, the whole *NGR* set is used in the greedy selection process. For a planar region, a similar process is done, but the *popMinNeighborArea* will favor areas in *NGR* that do not intersect with any aggregator edges.

Table 2 Partition status update procedure

```
ARP = empty list;
MAR = empty map;
G_IB = empty set;
For each R in P do:
  G_R = areas in R;
    For each g in G_R do:
      Update MAR with R and g;
      NG_g = neighboring areas of g from M_AA;
        NG_gR = NG_g ∩ G_R;
        If |NG_gR| = |NG_g|:
          Continue;
        Connected = True; default to True
        For each g' in NG_gR do:
            NG_g' = Neighboring areas of g' from M_AA;
            NG_g' = NG_g' ∩ G_R;
            Remove(g);
            If |NG_g'| = 0 :
              Connected = False; break;
            InnerConnected = False;
            For each g'' in NG_gR and g'' ≠ g do:
                NG_g'' = neighboring areas of g' from M_AA;
                NG_g'' = NG_g'' ∩ NG_g' ∩ G_R;
                NG_g''.remove(g);
            If |NG_g''| > 0:
                  InnerConnected = True; break;
            If innerConnected = False:
                Connected = False; break;
        If connected:
            Add g to G_IB
For each R in P do:
    NGR = neighboring areas in R;
    For each g in NGR do:
      If g in G_IB:
          Add the pair of (g, R) to ARP;
updateObjectiveFunctionValue();
```

Local Search

The local search stage follows the TABU search process, similar to the one used in solving the max-*P*-regions problem (Duque et al. 2012). A partition is considered to be improved if it has a smaller objective function value than a previous partition. The status of a partition contains the objective function value, the area-region pairs (*ARP*), and a mapping between all areas and regions (*MAR*) used in the region swapping process. Specifically, *ARP* contains all moves that will not break the spatial contiguity of either of the regions when they are swapping areas.

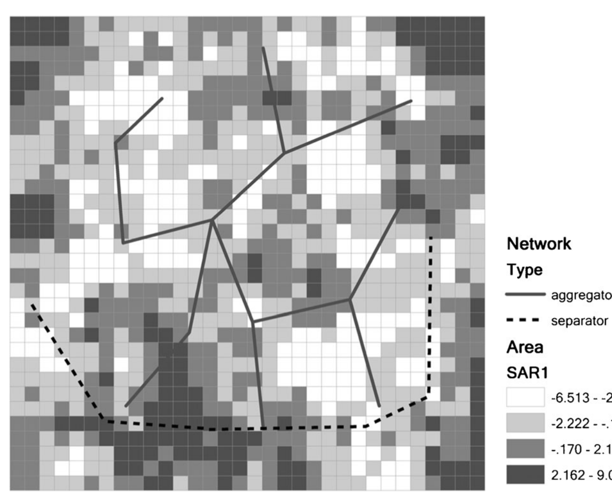

Fig. 3 Simulated dataset with areas ($n = 1024$) and network

Table 3 Comparing different weight combinations

	O	H	PR	t_{all}	t_{dp}	t_{i}	t_{ls}
(0.5, 8)	26,262.04	29,010.02	2747.98	80.888	1.438	8.931	70.519
(1.0, 8)	23,917.05	31,991.35	8074.30	78.411	1.446	7.705	69.26
(1.5, 8)	19,272.93	33,139.19	13,866.26	90.626	0.251	10.251	80.124
(0.5, 12)	22,286.18	27,265.10	4978.92	87.236	0.128	8.681	78.427
(1, 12)	17,607.31	31,081.68	13,474.37	89.303	0.096	7.481	81.726
(1.5, 12)	7915.58	33,453.24	25,537.66	97.85	0.105	8.483	89.262
(0.5, 15)	21,928.65	27,832.18	5903.53	103.325	0.146	10.694	92.485
(1, 15)	16,651.25	33,776.74	17,125.49	115.6	0.209	7.268	108.123
(1.5, 15)	341.5	42,907.32	42,565.82	136.837	0.103	9.487	127.247

When a move is applied, two regions will swap areas with each other. Because of the third condition described in the section "Model Objectives," a region might change types because the objective value of its current type might become bigger than the value of it being the other type. In other words, a planar region may become a network region, and a network region can revert back to a planar region. When a planar region becomes a network region, the root edge will be selected based on condition 2 in the section "Model Objectives." The swapping takes place by cloning regions and constructing a new partition $P^{'}$ from P. The status of $P^{'}$ is then updated and returned to the local search algorithm.

The status update process is shown in Table 2, which is invoked explicitly after the initial partitions are generated, as well as when new partitions are transformed from old ones in the local search stage. This process uses the pre-computed spatial contiguity matrices to efficiently construct all inner border areas G_{IB} across all regions in the partition, and then uses G_{IB} to update ARP. Essentially, G_{IB} is updated by checking whether neighbors of an area inside a region are all adjacent to each other or not.

Applications

This section studies the effectiveness and performance of the proposed method through a simulated dataset and a case study. First, a simulation experiment is given to show the partitions generated from different parameter combinations which shows the usage of the *extent* and *scale* factors in the model. The case study demonstrates the feasibility of the proposed model for real-world data. To quantify the performance, the time consumption (in seconds) of different stages was measured, including the overall running time t_{all}, time spent on the data preparation stage t_{dp}, initialization stage t_{i}, and local search stage t_{ls}. The decompositions of the objective function were also considered, including the final objective value O, and the heterogeneity factor H, and discounted network proximity factor PR. Afterwards, a practical experiment is conducted to show its scalability. In this section, the maximum number of initial solutions was chosen to be 100, and the TABU length was set to 85. The algorithm was implemented in Java and tested on an i7-4710HQ Intel CPU, with 8GB DDR3 memory, and a Windows 8.1 64-bit operating system.

Simulations

The simulated dataset is from the ClusterPy library (Duque et al. 2011a) and consists of 1024 areas. The heterogeneity attribute is generated under a spatial autoregressive process where $\rho = 0.9$.[1] A network dataset of 16 edges is constructed which is distributed randomly over the grid areas. Figure 3 displays the network overplayed on the dataset where $n = 1024$. The thick solid lines represent the aggregator edges, while the dashed line represents the separator edges. Areas intersected with the separator edges are removed from the computations and final visualizations.

The partitions are compared through three combinations of parameters (*scale, extent*) on the dataset. The number of regions is set to 30. Table 3 gives the numerical results, and Fig. 4 visualizes the partitions generated from these three parameter combinations. The dark grey regions mark the network regions, while the light grey regions are planar regions. The visualization clearly shows how the degree of attachment varies for network regions to edges. The time elapsed between different stages is relatively stable across different parameter combinations. The local search phase takes the most time, and it is proportional to both the *extent* and *scale* factors. The *extent* factor clearly influences the formation of network regions, with a bigger extent value producing more and larger network regions. The *scale* factor is more about controlling the compactness of the regions rather the number of network regions. A larger scale value means regions will grow more freely while maintaining the benefits of being a network region. Thus, the network regions with a smaller scale factor are more compact.

[1] https://code.google.com/p/clusterpy/

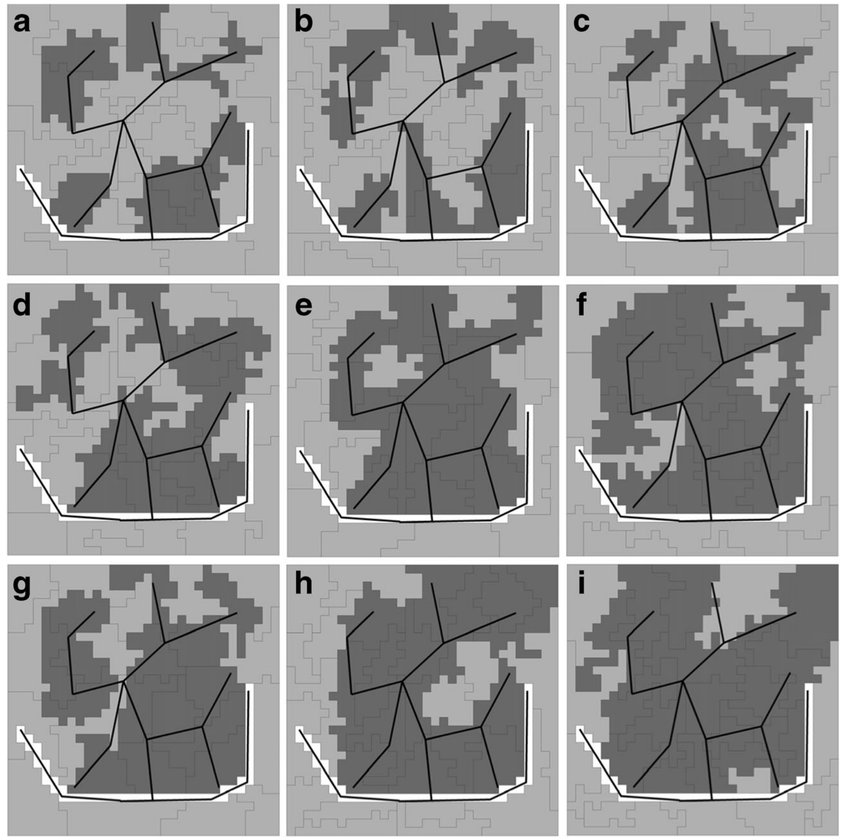

Fig. 4 Result of regionalization comparing the different parameter combinations of *scale* and *extent*. **a** (0.5, 8), **b** (1.0, 8), **c** (1.5, 8), **d** (0.5, 12), **e** (1.0, 12), **f** (1.5, 12), **g** (0.5, 15), **h** (1.0, 15), **i** (1.5, 15)

A Practical Experiment on Wuhan, China

The moderation process has been fast in China in recent decades. Wuhan is a fast-growing, large city in central China. The network-constrained regionalization process described in this work would help government agencies and researchers better understand and model the spatial layout under the constraints of a street network (Wang et al. 2011). The seven central districts are selected as the experiment areas. The area of urban land use is selected as the heterogeneity variable. The network is constructed from the major roads in Wuhan. Parts of the 3rd ring road are modeled as the separator edges in this experiment. There are 889 grids and 49 aggregator edges. The parameters are set to *scale* = 1, *extent* = 8000, and the number of regions = 30. The results are displayed in Table 4 and Fig. 5. The local search stage takes a lot more time than the initialization time. It also shows that the algorithm is fairly stable for a moderate-size dataset. As shown in Fig. 5, the network regions are formed based on the balance of the heterogeneity factor and the discounted network proximity factor. The separator border indeed separates most of the areas apart, but it also shows that regions could grow by extending from the end of a separator edge. Other ways of manipulating separator edge forms could be incorporated to deal with constraints in practice. It is important to recognize that urban dynamics is a complex process. This case study only serves as a demonstration to test the proposed model in this work. The detailed analysis of the created regions deserves a more thorough analysis and will be the focus of our future directions.

Discussion and Conclusion

There has been a growing interest in implementing planar spatial analysis methods into the network context. This work aims to partition the planar areas that are constrained by the network space. This work contributes to the extension of regionalization problems into the network space.

A city is a large and complex system (Ye and He 2016). The presence of streets and roads could have different types of influence over the nearby areas; a ring road would produce a clear cut between rural and urban areas, while an urban street

Table 4 Regionalization result and performance measures for the Wuhan grid data

O	H	PR	t_{all}	t_{dp}	t_{i}	t_{ls}
1.525566E7	4.382687E7	2.857122E7	4631.997	3.366	1827.174	2801.457

Fig. 5 Regionalization result for the grid data in Wuhan, China

would attract areas on both sides. This work formalizes the network as two sets of edges, the aggregator edges and the separator edges, and integrates them into the regionalization problem. This gives practitioners flexible control over how regions grow around network edges.

The discounted network proximity factor quantifies the edge-area relationships in the objective function. The two parameters *scale* and *extent* have clear physical meanings. The measures are surprisingly simple but perform reasonably well. More formally, the network compactness criteria could be further developed to rigorously define the compactness of an edge-surrounding shape. This work chose grid areas in the problem formulation. In practice, areas could have varying shapes. When dealing with such issues, a rigorous shape compactness measure of network regions becomes more important.

The construction of spatial contiguity and distance matrices, as well as the partition status, accelerates both initial partition generation and local search stages. These data structures can also be stored and used in later executions with different parameter settings. The network structure in this work is considered to be homogeneous and relatively sparse. In practice, researchers might want to reverse this assumption when the network is dense. A region would belong to a set of connected edges. In addition, the street hierarchy could affect the proximity factor. For example, a small road may attract fewer areas, and the big road may attract more areas. Such rules can be integrated into the region generation process and thus foster an adaptive way of generating regions.

In conclusion, this paper proposes an efficient way for the regionalization problem extended into the network space. The regionalization process results can help researchers and practitioners in designing network-constrained service regions with a fixed number of regions. In addition, it explicitly integrates the network structure into the modeling process in a flexible way. Other types of optimization problems can also be expanded into the network space following such a heuristic framework.

Acknowledgments Portions of this research were supported by National Science Foundation Grants 1535031 and 1637242. We have benefitted from the suggestions and comments of Juan C. Duque, the anonymous reviewers, and the editors.

Funding This study was funded by the National Science Foundation (1535031 and 1637242).

Compliance with Ethical Standards This article does not contain any studies with human or animal subjects.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This paper does not contain any studies with human participants performed by any of the authors.

Informed Consent This paper does not involve any individual participants.

References

Barkan JD, Densham PJ, Rushton G (2006) Space matters: designing better electoral systems for emerging democracies. Am J Polit Sci 50(4):926–939. https://doi.org/10.1111/j.1540-5907.2006.00224.x

Bergey P, Ragsdale C, Hoskote M (2003) A simulated annealing genetic algorithm for the electrical power districting problem. Ann Oper Res 121(1–4):33–55. https://doi.org/10.1023/A:1023347000978

Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. Science 323(5916):892–895. https://doi.org/10.1126/science.1165821

Bozkaya B, Erkut E, Laporte G (2003) A tabu search heuristic and adaptive memory procedure for political districting. Eur J Oper Res 144(1):12–26. https://doi.org/10.1016/S0377-2217(01)00380-0

Brás R, Cerdeira JO, Alagador D, Araújo MB (2013) Linking habitats for multiple species. Environ Model Softw 40:336–339. https://doi.org/10.1016/j.envsoft.2012.08.001

Church RL, Murray AT (2008a) Line-based location. Business site selection, location analysis and GIS. John Wiley & Sons, Hoboken, pp 159–186

Church RL, Murray AT (2008b) Point-based location. Business site selection, location analysis and GIS. John Wiley & Sons, Hoboken, pp 135–158

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

D'amico SJ et al (2002) A simulated annealing approach to police district design. Comput Oper Res 29(6):667–684. https://doi.org/10.1016/S0305-0548(01)00056-9

Duque J, *et al.* (2011a) ClusterPy: library of spatially constrained clustering algorithms. Version 0.9, 9

Duque JC, Church RL, Middleton RS (2011b) The p-regions problem. Geogr Anal 43(1):104–126. https://doi.org/10.1111/j.1538-4632.2010.00810.x

Duque JC, Anselin L, Rey SJ (2012) The max-p-regions problem*. J Reg Sci 52(3):397–419. https://doi.org/10.1111/j.1467-9787.2011.00743.x

Eckley DC, Curtin KM (2013) Evaluating the spatiotemporal clustering of traffic incidents. Comput Environ Urban Syst 37:70–81. https://doi.org/10.1016/j.compenvurbsys.2012.06.004

Feo T, Resende MC (1995) Greedy randomized adaptive search procedures. J Glob Optim 6(2):109–133. https://doi.org/10.1007/BF01096763

Fischer MM (1980) Regional taxonomy: a comparison of some hierarchic and non-hierarchic strategies. Reg Sci Urban Econ 10(4):503–537. https://doi.org/10.1016/0166-0462(80)90015-0

Gordon AD (1996) A survey of constrained classification. Comput Stat Data Anal 21(1):17–29. https://doi.org/10.1016/0167-9473(95)00005-4

Hansen P, Jaumard B, Meyer C, Simeone B, Doring V (2003) Maximum split clustering under connectivity constraints. J Classif 20(2):143–180. https://doi.org/10.1007/s00357-003-0011-7

Hardisty F, Klippel A (2010) Analysing spatio-temporal autocorrelation with LISTA-Viz. Int J Geogr Inf Sci 24(10):1515–1526. https://doi.org/10.1080/13658816.2010.511717

Huang L, Zhu X, Ye X, Guo W, Wang J (2015) Characterizing street hierarchies through network analysis and large-scale taxi traffic flow: a case study of Wuhan, China. Environ Plann B: Plann Des 43(2):276–296. https://doi.org/10.1177/0265813515614456

Jiang B, Claramunt C (2004) Topological analysis of urban street networks. Environ Plann B 31(1):151–162. https://doi.org/10.1068/b306

Jiang B, Jia T (2011) Agent-based simulation of human movement shaped by the underlying street structure. Int J Geogr Inf Sci 25(1):51–64. https://doi.org/10.1080/13658811003712864

Lemberg DS, Church RL (2000) The school boundary stability problem over time. Socio Econ Plan Sci 34(3):159–176. https://doi.org/10.1016/S0038-0121(99)00019-1

Li W, Goodchild MF, Church R (2013) An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. Int J Geogr Inf Sci 27(6):1227–1250. https://doi.org/10.1080/13658816.2012.752093

Li W, Church RL, Goodchild MF (2014a) An extendable heuristic framework to solve the p-compact-regions problem for urban economic modeling. Comput Environ Urban Syst 43:1–13. https://doi.org/10.1016/j.compenvurbsys.2013.10.002

Li W, Church RL, Goodchild MF (2014b) The p-compact-regions problem. Geogr Anal 46(3):250–273. https://doi.org/10.1111/gean.12038

Miller HJ, Wentz EA (2003) Representation and spatial analysis in geographic information systems. Ann Assoc Am Geogr 93(3):574–594. https://doi.org/10.1111/1467-8306.9303004

Okabe A, Satoh T, Furuta T, Suzuki A, Okano K (2008) Generalized network Voronoi diagrams: concepts, computational methods, and applications. Int J Geogr Inf Sci 22(9):965–994. https://doi.org/10.1080/13658810701587891

Okabe A, Yamada I (2001) The K-function method on a network and its computational implementation. Geogr Anal 33(3):271–290

Okabe A, Satoh T, Sugihara K (2009) A kernel density estimation method for networks, its computational method and a GIS-based tool. Int J Geogr Inf Sci 23(1):7–32. https://doi.org/10.1080/13658810802475491

Parthasarathi P, Hochmair H, Levinson D (2014) Street network structure and household activity spaces. Urban Studies 0042098014537956

Porta S, Romice O, Maxwell JA, Russell P, Baird D (2014) Alterations in scale: patterns of change in main street networks across time and space. Urban Stud 51(16):3383–3400. https://doi.org/10.1177/0042098013519833

Ricca F, Simeone B (2008) Local search algorithms for political districting. Eur J Oper Res 189(3):1409–1426. https://doi.org/10.1016/j.ejor.2006.08.065

She B, Zhu X, Ye X, Guo W, Su K, Lee J (2015) Weighted network Voronoi diagrams for local spatial analysis. Comput Environ Urban Syst 52:70–80. https://doi.org/10.1016/j.compenvurbsys.2015.03.005

She B, Duque J, Ye X (2017) The network-max-P-regions model. Int J Geogr Inf Sci 31(5):962–981. https://doi.org/10.1080/13658816.2016.1252987

Shirabe T (2005) A model of contiguity for spatial unit allocation. Geogr Anal 37(1):2–16. https://doi.org/10.1111/j.1538-4632.2005.00605.x

Tong D, Murray AT (2012) Spatial optimization in geography. Ann Assoc Am Geogr 102(6):1290–1309. https://doi.org/10.1080/00045608.2012.685044

Tung C-P, Chou C-A (2002) Application of tabu search to ground water parameter zonation1. JAWRA J Am Water Resour Assoc 38(4):1115–1125. https://doi.org/10.1111/j.1752-1688.2002.tb05550.x

Wang F, Antipova A, Porta S (2011) Street centrality and land use intensity in Baton Rouge, Louisiana. J Transp Geogr 19(2):285–293. https://doi.org/10.1016/j.jtrangeo.2010.01.004

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

Whiteaker TL, Maidment DR, Gopalan H, Patino C, Mckinney DC (2007) Raster-network regionalization for watershed data processing. Int J Geogr Inf Sci 21(3):341–353. https://doi.org/10.1080/13658810600965255

Whitehand JWR, Larkham PJ (1992) Urban landscapes: international perspectives. ROUTLEDGE CHAPMAN & HALL, London

Williams JC, Revelle CS, Levin SA (2004) Using mathematical optimization models to design nature reserves. Front Ecol Environ 2(2):98–105.

Xie Z, Yan J (2008) Kernel density estimation of traffic accidents in a network space. Comput Environ Urban Syst 32(5):396–406. https://doi.org/10.1016/j.compenvurbsys.2008.05.001

Yamada I, Thill J-C (2004) Comparison of planar and network K-functions in traffic accident analysis. J Transp Geogr 12(2):149–158. https://doi.org/10.1016/j.jtrangeo.2003.10.006

Yamada I, Thill J-C (2007) Local indicators of network-constrained clusters in spatial point patterns. Geogr Anal 39(3):268–292. https://doi.org/10.1111/j.1538-4632.2007.00704.x

Yamada I, Thill J-C (2010) Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. Ann Assoc Am Geogr 100(2):269–285. https://doi.org/10.1080/00045600903550337

Ye X, He C (2016) The new data landscape for regional and urban analysis. GeoJournal 81(6):811–815. https://doi.org/10.1007/s10708-016-9737-8

Zhao P, Qin K, Ye X, Wang Y, Chen Y (2017) A trajectory clustering approach based on decision graph and data field for detecting hotspots. Int J Geogr Inf Sci 31(6):1101–1127