Institut für Verkehrswirtschaft
Lehrstuhl für Betriebswirtschaftslehre, insb. Verkehr
Schwerpunkt: Operations & Supply Chain Management

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

A District Design Approach to Construct Pilgrim Camp Layouts

Matthes Koch*

May 27, 2019

Report HH-ITE 001-2018
Institute for Transport Economics
www.bwl.uni-hamburg.de/vw.html

*Institute for Transport Economics, Hamburg Business School, Universität Hamburg, 20148 Hamburg, Germany, matthes.koch@uni-hamburg.de

Abstract

This paper focuses on a districting problem that arises in the context of Hajj pilgrim camp site design. The goal of this design approach is to construct a camp layout plan for up to 500,000 pilgrims given a set of basic tent units. We will formulate a bi-objective mathematical planning model for the problem. It incorporates constraints to ensure that each camp provides access to a sanitary facility, an entry point to the road network and sufficient space to accommodate the pilgrim group. The model's first objective is to spatially distribute the camps in accordance with a prescribed pilgrim departure schedule. The second objective is to generate regular, simple and contiguous camp shapes. To attain results for large instances, we develop a two stage solution procedure. In the first stage, we use a mixed-integer programming approach to partition the solution space. This allows us to solve multiple districting subproblems in the second stage independently. Our districting model establishes a contiguous core of layout units for each camp. Additionally, outer layout units may be partially assigned to adjacent camps. We illustrate this novel approach by applying it to the real dataset of Hajj season 2018 with more than 395,000 pilgrims in 282 camps and 3,058 spatial units. We construct a feasible starting solution in under 12 minutes computation time. Using parallel processing of the subproblems, we generate partial solutions in 32 cpu seconds on average during the improvement phase.

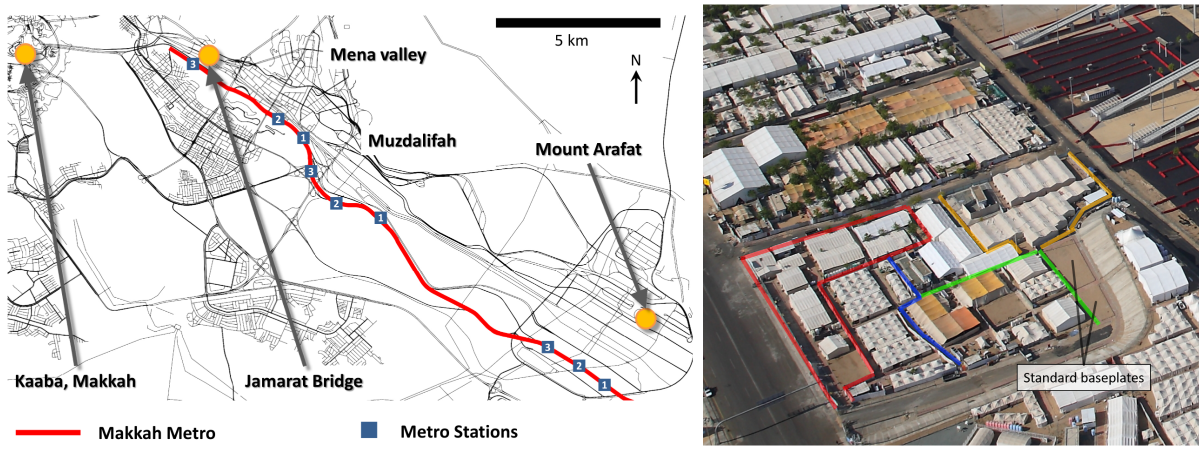
1. Introduction

The Hajj – the great Islamic pilgrimage towards Mecca – is one of the largest mass gatherings in the world (Müller, 2015). As part of the Hajj rituals, pilgrims gather in the plain of Arafat on the 9th day of Dhul-Al hija (Figure 1a). Within only a few hours, a modern high-capacity metro system currently transfers between 350,000 and 500,000 pilgrims from a permanent tent city in the Mena valley to the target site in Arafat. Hajj administration organizes the pilgrims in groups of different sizes and nationalities. For safety reasons, each group is dispatched towards one of three metro stations according to a prescribed schedule (Haase et al., 2016). As a result of the coordinated dispatching procedure, pilgrim group arrivals follow the scheduled sequence in Arafat. Pilgrim organizations, or *service offices*, then accommodate their groups in separate, temporary camps around the target stations.

Given the set of pilgrim groups, their pilgrim count, space requirement, their designated arrival station, arrival platform side, and sequence of arrival we seek to generate a tent layout plan for the target area. Each camp in the tent layout plan must provide sufficient space for its pilgrim group. It must grant access to at least one of the fixed sanitary facilities. Moreover, each camp must adjoin at least one of the streets so that pilgrim groups can access the planned camp.

Only days before the Hajj, more than 220 service offices independently build up their tents for often thousands of pilgrims. They usually conduct this operation under considerable time pressure when the authorities release the final layout plan. To facilitate the implementation of a layout plan our first objective is to generate simple, contiguous and block-like shapes. Our second objective is to distribute the camps in the allocation zone in accordance with their arrival sequence. To avoid that groups block each other when they walk from the station to their camp in the target area, early arriving pilgrim groups are to be located farther away from the station than later groups. Thus the pilgrim groups fill the tent space by following their scheduled arrival sequence.

Figure 1b highlights the borders of five camps near one of the three metro stations at the target site. Most tents are non-standardized, temporary accommodations only used on the Arafat-day of Hajj. Since pilgrim group sizes and locations are subject to change until only a few days before the event, new layout solutions need to be provided quickly.



(a) Tent City and Arafat Area (b) Camp allocation around Station 1

Figure 1: Camps at the target area

2. Related Research

The goal of districting or *regionalization* problems is to group a set of basic spatial units into larger geographic sub-regions such that the considered area features a desired structure (Li et al., 2013). Districting problems arise in diverse application areas such as political districting, service districting, sales territory alignment, natural reserve planning, and many others. In a comprehensive literature review, Kalcsics (2015) found that most research articles in this area report on specific, practical applications. Only a few authors attempt to formulate a generalized version of the districting problem.

Most approaches share general districting criteria of compactness, contiguity, balance, and exclusiveness, but differ in their actual specification (Kalcsics et al., 2005; Ricca and Simeone, 2008; Kalcsics, 2015).

Li et al. (2013) provide an overview of different measures of *compactness* in spatial optimization, which they categorize into four groups: area-perimeter measurement, reference shape, geometric pixel properties, and dispersion of elements of the area. Each application requires an adequate method to measure compactness of the constructed regions.

In commercial applications, the objective is to maximize the accessibility to all areas of the sub-regions and thus to minimize operational costs. In sales territory alignment problems, for example, compact territories are designed to minimize sales agents' travel costs (Drexl and Haase, 1999). In natural reserve design problems, habitat management costs such as fencing costs and land negotiation with neighbors can be minimized by finding compact habitat areas (Fischer and Church, 2003; Jafari and Hearne, 2013).

In political applications, like electoral districting, compact shapes are desired to avoid the impression of *gerrymandering* districts, that is the drawing of electoral districts to gain demographic advantages for a specific party (Webster, 2013).

Shirabe (2009) differentiates between economic and geometrical compactness. Optimizing for geometrical compactness, i.e., minimizing the total distances between regional centers and their assigned spatial units may promote circular configurations (Cova and Church, 2000; Onal et al., 2015). In the camp layout problem, cost minimization is irrelevant, because the land is provided free of charge. Instead, planning authorities' requirements for the camp layout plans describe simple, block-like subregions that are easy to memorize by field operators such as construction staff, crowd managers, and pilgrim group dispatchers. Circular and irregular shapes should be avoided. Moreover, camps may fill remaining space. Contrary to commercial land-acquisition models (Williams, 2002), there is no cost penalty for the assignment of layout cells.

The growth of each camp is restricted by the its neighbors' minimum space requirements. Similar minimum area constraints are known in preservation planning problems (Williams et al., 2005), in which each natural habitat must support a minimum population of the protected species.

In many districting problems, balancing criteria are either incorporated in the constraint set or in the objective function (Hess et al., 1965; Fleischmann and Paraschis, 1988; Blais et al., 2003). In sales territory alignment problems, for example, sales potential and agents workloads need to be balanced (Zoltners and Sinha, 2005) while political districts are planned to balance demographic figures (Webster, 2013). In camp layout planning, we use soft constraints to assign the camp areas equitably. Constructing camps greater than the minimum size is feasible. Yet, we apply a balancing criterion for camps above the minimum size to avoid rejection of a camp layout plan for fairness reasons.

In a recent paper, Wang et al. (2018) investigated different modeling approaches to enforce spatial *contiguity*. Contiguity constraints have been modeled using dual-graph based integer programming (Williams, 2002), network flow models (Shirabe, 2005, 2009) and path building constraints (Zoltners and Sinha, 1983; Cova and Church, 2000). We tested network flow and path building constraints to construct contiguous camp zones. In the context of natural reserve planning, computational tests indicate good performance for the path building approach (Wang et al., 2018), which is in line with the findings that network flow models suffer from computational difficulties as their scale grows (Shirabe, 2009).

The path building strategy was first introduced for the case of fixed center units, around which contiguous zones were to be constructed (*rooted* regionalization problem, Zoltners and Sinha (1983); Cova and Church (2000)). In their review, Kalcsics (2015) found that in most applications, the specific location of district centers is not an optimization objective in itself. However, compactness and contiguity criteria are usually calculated using a spatial unit declared as the center.

In the camp layout problem, we simultaneously select camp centers from a set of potential centers along with their assigned satellite units (*unrooted* regionalization problem). The reasons for this simultaneous approach are twofold. First, our objective is to align the camps spatial distribution with their prescribed arrival sequence in the target area. We use the camp centers to measure the distance between a camp and its target station. The selection of centers thus determines the quality of our solution and cannot be defined exogenously. Second, we define the set of potential center units such that each unit has direct access to a sanitary facility. Since each camp must start layout construction from one of the units in the set of potential centers, access to a sanitary facility is implicitly guaranteed for each pilgrim group.

In contrast to most integer programming approaches in the literature, we relax the requirement of assigning each spatial unit exclusively to one central unit and the requirement of complete assignment (Kalcsics, 2015). Our formulation allows sharing units along camp zone borders without compromising the effectiveness of the contiguity constraints. This relaxation is beneficial in the partitioning approach used to solve the problem for the large dataset of the real application.

3. Model Formulation

Figure 2: Layout cells

The photograph in Figure 1b shows two standardized solid tent baseplates (unoccupied). The plates cover the plains around the Arafat stations and are the basic spatial units in our model. We define the set $\mathcal{U}$ to contain all base plates, which we also refer to as layout cells. Standard cells, cells with street access, and toilet cells are highlighted on the area map in Figure 2. Each non-sanitary cell provides accommodation space denoted as capacity parameter r_u.

To ensure that each camp has access to a sanitary facility, we assign all cells adjacent to such units to the set of potential start cells for camp construction $\tilde{\mathcal{U}}$. Start cells in $\tilde{\mathcal{U}}$ selected for layout construction represent the center of the respective camp (*district*). For each start cell $u \in \tilde{\mathcal{U}}$, we define one subset $\mathcal{K}_u \subset \mathcal{U}$ containing all candidate cells for camp construction. The candidate cells for a start cell are typically in the same street block[1]. However, they can also be restricted by a range distance to allow camps to stretch over multiple street blocks.

To model the grouping of cells around a start cell, we define a layout construction network as illustrated in Figure 3. In the network, we connect cells $u \in \mathcal{U}$ with its

[1]Streets and walkways naturally partition the catchment area into blocks (see Appendix B, Figure 8).

adjacent cells $v \in \mathcal{N}_u$ at a distance of $e_{u,v}$. We avoid diagonal connections between cells to favor rectangular layouts in the optimization.

Figure 3: Excerpt of the layout network

Each pilgrim group requires a separate camp. We define the set of camps $\mathcal{C}$ with one element $c \in \mathcal{C}$ for every pilgrim group. Each camp demands accommodation space, denoted q_c, dependent on its pilgrim count and its service commitment (area reserved per pilgrim).

The metro service at Arafat operates multiple train convoys per station. It is optimized to provide large-scale passenger capacity in the space of a few peak hours (Haase et al., 2016). To ensure an efficient metro operation, a predetermined, capacity-balanced schedule prescribes the station and platform for each pilgrim group. The group uses the station and platform upon arrival and departure from the target area. It cannot switch stations[2]. Hence, each pilgrim group is assigned to exactly one of six spatially overlapping partitions $p \in \mathcal{P}$ (see Appendix B, Figure 7), defined in line with past years' layout partitions. Each partition p defines a subset of assigned camps $\mathcal{C}_p \subset \mathcal{C}$. Only the layout cells contained in a camp's assigned partition are valid for layout construction. $\tilde{\mathcal{U}}_c$ denotes the set of valid start cells for layout construction of camp c. By

[2]Metro-users must carry a colored wristband equipped with an RFID tag. The color indicates the planned station. Security forces control access to the stations

the integer parameter $s_c\ \forall c \in \mathcal{C}$ we describe the arrival sequence in each partition. The destination point of each partition is the corresponding metro station. The shortest-path distance between the destination point of partition p and the layout start cells inside of p is defined as parameter $d_{u,p}\ \forall p \in P, u \in \tilde{U}_p$, where $\tilde{U}_p$ denotes the subset of start cells within the catchment of partition p (see Figure 7).

3.1. Objectives

We use the binary variable $Y_{c,u}$ to indicate if camp c is assigned to start cell u in the solution. To achieve a spatial alignment according to the dispatching sequences, we define the following objective function:

$$\text{minimize } L_1 = \sum_{p \in \mathcal{P}} \sum_{c \in \mathcal{C}_p} \sum_{u \in \tilde{\mathcal{U}}_c} \frac{d_{u,p}}{s_c} \cdot Y_{c,u} \tag{1}$$

According to (1), we prefer startpositions close to the destination points for all camps. Additionally, camps are weighted by sequence parameter s_c to facilitate the allocation of camps with the lowest values of s_c to the closest start cells.

We use binary variable $Z_{u,v,w}$ to indicate if two adjacent cells (v, w) are assigned to the same start cell u in the solution. To facilitate the generation of block-like rectangular layouts, we maximize the number of such neighbor connections within the camps. Figure 4 illustrates this principle with a stylized example. In Figure 4a three cells are occupied by Camp C1. It began layout construction in cell 4. To provide enough space for its pilgrims, it requires one additional cell. If we added cell 6 to the layout, one additional neighbor connection (5,6) would be realized (Figure 4b). If instead we added cell 2 to the layout, two additional neighbor connections (1,2),(2,5) would be part of C1 (Figure 4c). We maximize the weighted sum of the neighbor connections realized for each camp to achieve the desired layout. Thus, we prefer the layout in Figure (4c) over the solution shown in Figure (4b). As a diagonal connection between cell 5 and 3 (Figure 4d) would result in a non-rectangular shape, we exclude such links from the adjacency list of the layout graph.

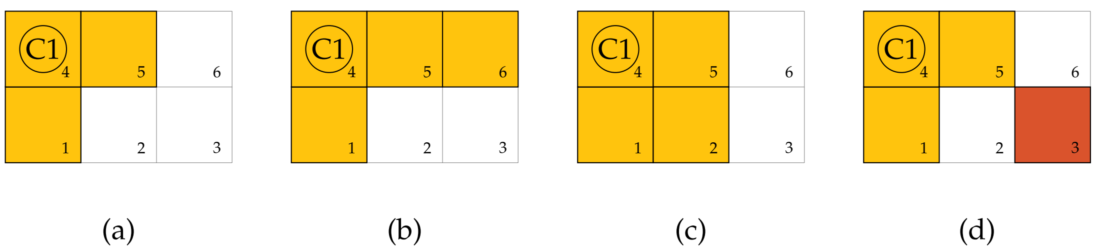


Figure 4: Non-optimal, optimal, and invalid expansions of a camp zone

Maximization of the camp-internal neighbor connections is expressed in the second objective function:

$$\text{maximize } L_2 = \sum_{u \in \tilde{\mathcal{U}}} \sum_{(v,w) \in \Omega_u} \beta_{u,v,w} \cdot Z_{u,v,w} \tag{2}$$

The subsets Ω_u contain all pairs of neighboring layout units that can be assigned to a source cell u, i.e., $\Omega_u = \{\ (v,w) \mid v \in \mathcal{K}_u, w \in \mathcal{K}_u \cap \mathcal{N}_v, v < w\} \quad \forall u \in \tilde{U}$.

The weight parameter $\beta_{u,v,w}$ is calculated as follows:

$$\beta_{u,v,w} = \frac{1}{\max\{e_{u,v}, e_{u,w}\}} \quad \forall u \in \tilde{U}, (v,w) \in \Omega_u$$

It decreases with increasing shortest-path-distance between the pair (v,w) and its potential source cell u. Each addition of a pair of cells to a camp-layout raises the objective function value. Thus, camp areas tend to expand beyond their minimum space requirements if abundant space is available.

Figure 5 illustrates the application of three alternative formulations for the objective function. On a 10x10 grid, we allocate one large-sized, one medium-sized and one small camp. Jointly, the camps require fewer cells than the available total.

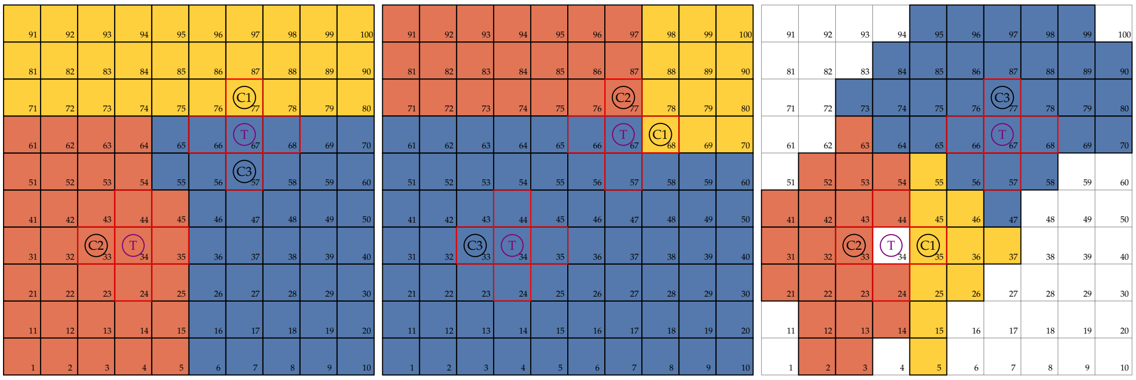
Implementing the coefficients as defined in (3.1) yields the solution shown in Figure 5a. The camps fill the remaining space, but the decreasing gain in objective value yields a well-balanced allocation of space to the camps. In contrast, Figure 5b shows a solution with all weights defined uniformly, i.e., $\beta_{u,v,w} = 1\ \forall u \in \tilde{\mathcal{U}}, (v,w) \in \Omega_u$. Similar to 5a, the camps occupy the entire space in this solution. The large camp must consume more cells than the smaller camps to satisfy its space demand. Because of its greater surface area, it is beneficial to expand the largest camp most. The extent of the smaller camps almost reduces to the minimum requirement.

An alternative approach is shown in Figure 5c, where we minimized the total distance between the assigned cells and their respective centers. While that approach is often used to generate compact layouts in other application contexts, the resulting circular shapes of the districts do not resemble past camp layout planning results very well. Additionally, such plans do not fill the remaining space.

3.2. Camp Layout Model

The introduction of the following two additional variables and related sets allows us to formulate the necessary constraints that ensure contiguity and space demand satisfaction of the camps. First, binary variable $X_{u,v}$ indicates a complete assignment of a cell $v \in \mathcal{K}_u$ to a source cell $u \in \tilde{\mathcal{U}}$.

Second, cells of a camp adjacent to cells of another camp may be shared. In the model,



(a) 0.81 CPU seconds (b) 173.31 CPU seconds (c) 1.53 CPU seconds

Figure 5: Results of different objective functions and computation times

non-negative variable $\tilde{X}_{u,v}$ represents the share of the cell $v \in \mathcal{K}_u$ occupied by the camp with start cell $u \in \tilde{\mathcal{U}}$.

Subset $\mathcal{U}^{\text{street}} \subset \mathcal{U}$ contains all cells with direct access to the pedestrian walkways. For every start cell u, we define a subset $\mathcal{U}_u^{\text{fix}}$, which contains all cells that must be part of the layout if u is a start cell for a camp. These subsets include at least the respective start cell itself. However, they can also be used to pre-define certain areas for a start cell.

The contiguity constraints rely on the definition of the following subsets:

$$\tilde{\mathcal{N}}_{u,v} = \{w \in \mathcal{K}_u \cap \mathcal{N}_v \mid e_{u,w} < e_{u,v}\} \qquad \forall u \in \tilde{\mathcal{U}}, v \in \mathcal{K}_u$$

For a given combination of (u, v) with $u \in \tilde{\mathcal{U}}, v \in \mathcal{K}_u$, the subset $\tilde{\mathcal{N}}_{u,v}$ contains cells adjacent to v which are also closer to u than v itself. Contiguity of a camp rooted in u is established by forcing at least one element of $\tilde{\mathcal{N}}_{u,v}$ into the camp if v is to be assigned to u.

Having established the necessary notation, we define the camp layout planning model **P1**:

Maximize

$$L_0 = \quad a_1 \left(\sum_{p \in \mathcal{P}} \sum_{u \in \mathcal{U}_p} \sum_{c \in \mathcal{C}_p} \frac{d_{u,p}}{s_c} \cdot Y_{c,u} \right) \tag{3}$$
$$+ a_2 \left(\sum_{u \in \tilde{\mathcal{U}}} \sum_{(v,w) \in \Omega_u} \beta_{u,v,w} \cdot Z_{u,v,w} \right)$$

subject to

$$\sum_{u \in \tilde{\mathcal{U}}_c} Y_{c,u} = 1 \qquad \forall c \in \mathcal{C}, \tag{4}$$
$$\sum_{c \in \mathcal{C}_u} Y_{c,u} \leq 1 \qquad \forall u \in \tilde{\mathcal{U}}, \tag{5}$$
$$\sum_{c \in \mathcal{C}_u} Y_{c,u} \geq X_{u,v} + \tilde{X}_{u,v} \qquad \forall u \in \tilde{\mathcal{U}}, v \in \mathcal{K}_u, \tag{6}$$
$$\sum_{c \in \mathcal{C}_u} q_c \cdot Y_{c,u} \leq \sum_{v \in \mathcal{K}_u} r_v (X_{u,v} + \tilde{X}_{u,v}) \quad \forall u \in \tilde{\mathcal{U}}, \tag{7}$$
$$Y_{c,u} \leq \sum_{v \in \mathcal{K}_u \cap \mathcal{U}^{\text{street}}} X_{u,v} \qquad \forall u \in \tilde{\mathcal{U}}, c \in \mathcal{C}_u, \tag{8}$$
$$\sum_{v \in \mathcal{U}_u^{\text{fix}}} X_{u,v} = \left| \mathcal{U}_u^{\text{fix}} \right| \cdot \sum_{c \in \mathcal{C}_u} Y_{c,u} \qquad \forall u \in \tilde{\mathcal{U}}, \tag{9}$$
$$X_{u,v} + \tilde{X}_{u,v} \leq \sum_{w \in \tilde{\mathcal{N}}_{u,v}} X_{u,w} \qquad \forall u \in \tilde{\mathcal{U}}, v \in \mathcal{K}_u | v \neq u, v \notin \mathcal{N}_u, \tag{10}$$
$$\sum_{u \in \tilde{\mathcal{U}} | v \in \mathcal{K}_u} (X_{u,v} + \tilde{X}_{u,v}) \leq 1 \qquad \forall v \in \mathcal{U}, \tag{11}$$
$$X_{u,v} + X_{u,w} \geq 2 \cdot Z_{u,v,w} \qquad \forall u \in \tilde{\mathcal{U}}, v \in \mathcal{K}_u, w \in \mathcal{K}_u \cap \mathcal{N}_v | v < w, \tag{12}$$
$$X_{u,v} \in \{0,1\},\ 0 \leq \tilde{X}_{u,v} \leq 1 \qquad \forall u \in \tilde{\mathcal{U}}, v \in \mathcal{K}_u, \tag{13}$$
$$Y_{c,u} \in \{0,1\} \qquad \forall c \in \mathcal{C}, u \in \tilde{\mathcal{U}}_c \tag{14}$$

Objective function (3) combines the objectives L_0 and L_1 introduced in Section 3.1. Coefficients a_1, a_2 are weighting parameters.

Constraints (4)-(5) ensure that we assign each camp exclusively to one start cell. Layout variables X are coupled with the camp decisions Y via the constraints (6)-(9). By (6), source cells without an associated camp cannot construct a layout around them. If there is a camp associated with a start cell, (7) ensures that the layout provides sufficient space for the camp. Constraints (8) require each camp to include at least one cell with direct access to the pedestrian walkways (camp entrance). According to constraints (9), the solution must include the cell assignment predefined in $\mathcal{U}_u^{\text{fix}}$ if cell u serves as a camp center. These subsets include at least their source cell u itself. Planners may also predefine layout structures through these constraints. Contiguity constraints (10) prevent any assignment of a cell to a source cell if no neighbor cell closer to the source cell is assigned as well. This block of constraints effectively enforces the construction of a path between the source and each of its assigned satellite cells within the layout network. Note that the constraints establish a path through binary variables only. Cells, assigned by the relaxed variables $\tilde{X}_{u,v}$, cannot extend a path. Hence, a camp may partially occupy a cell along its borders only. Such cells may also be split between two different camp zones. The exclusive assignment of cells, a typical feature in districting problems (Kalcsics, 2015), is limited to the core area of the camps. To prevent over-subscription, constraints (11) limit the total relative assignment of a cell v to its potential sources to 100%. Finally, constraints (12) set up the counting variable $Z_{u,v,w}$. For a given source cell u and a potential connection (v, w), constraints (12) set $Z_{u,v,w}$ to 1 if u captures both cells v and w exclusively for the camp. The variable domains are defined in (13) and (14).

4. Solution Approach and Application

Districting problems are known to be difficult to solve by exact methods (Kong et al., 2019). Cordone and Maffioli (2004) demonstrated that its generalized variant, the constrained graph partition problem, is NP-hard. Computational experimentation on problem **P1** with real Hajj data of 2017 (Table 1) confirmed that computation times are prohibitive when using a standard solver.

Table 1: Problem size

			Cells				Variables		
Entity	Blocks	$\lvert P \rvert$	$\lvert \mathcal{C} \rvert$	$\lvert \mathcal{U} \rvert$	$\lvert \tilde{\mathcal{U}} \rvert$	$\lvert \mathcal{U}^{\text{street}} \rvert$	$X, \tilde{X}$	Y	F, H
Size	62	6	282	3,058	983 (32%)	1,378 (45%)	53,852	56,463	12,146

To tackle the computational difficulties, we employ a heuristic two-stage solution approach. In the first stage, we seek to identify a feasible starting solution to **P1**. We generate an assignment of camps to street blocks that is feasible with respect to space demand and camp contiguity. In the second stage, we fix the assignment of camps to blocks and improve the solution by solving **P1** for each block separately.

We use a network flow model to construct a feasible solution. In the model, flow represents the space requirement of a camp. Each camp initiates a flow in the network by assigning its full space requirement to one start cell. We use flow balancing constraints to establish each cell's space capacity limit. Additional constraints are introduced to ensure unique solutions.

Model **P2** incorporates non-negative flow variables $F_{u,v}$ and the binary indicator variables $H_{u,v}$:

Minimize

$$L_3 = \sum_{p \in \mathcal{P}} \sum_{u \in \tilde{\mathcal{U}}_p} \sum_{c \in \mathcal{C}_p} d_{u,p} \cdot q_c \cdot Y_{c,u} + \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{N}_u} e_{u,v} \cdot F_{u,v} \tag{15}$$

subject to (4),(5),(14),

$$\sum_{c \in \mathcal{C}_u} (q_c - r_u) \cdot Y_{c,u} \leq \sum_{v \in \mathcal{N}_u} F_{u,v} \qquad \forall u \in \tilde{\mathcal{U}}, \tag{16}$$

$$\sum_{v \in \mathcal{N}_u} F_{v,u} - \sum_{v \in \mathcal{N}_u} F_{u,v} \leq r_u \qquad \forall u \in \mathcal{U}, \tag{17}$$

$$\sum_{v \in \mathcal{N}_u} F_{v,u} \leq 2 \cdot q^{\max} \cdot (1 - \sum_{c \in \mathcal{C}_u} Y_{c,u}) \qquad \forall u \in \tilde{\mathcal{U}} \tag{18}$$

$$F_{u,v} \leq q^{\max} \cdot H_{u,v} \qquad \forall u \in \mathcal{U}, v \in \mathcal{N}_u \tag{19}$$

$$\sum_{v \in \mathcal{N}_u} \frac{H_{u,v}}{|\mathcal{N}_u|} + \sum_{v \in \mathcal{N}_u} H_{v,u} \leq 2 \qquad \forall u \in \mathcal{U} \tag{20}$$

$$H_{u,v} \in \{0, 1\} \qquad \forall u \in \mathcal{U}, v \in \mathcal{N}_u \tag{21}$$

$$F_{u,v} \geq 0 \qquad \forall u \in \mathcal{U}, v \in \mathcal{N}_u \tag{22}$$

The objective function (15) minimizes the total distance-weighted flows. Constraints (16) initialize the flow for each camp at the source cells according to the source-camp assignment $Y_{c,u}$. Flow balance constraints in (17) ensure that the flow residing in a cell $u \in \mathcal{U}$ is bound by its space capacity. To ensure that camp centers are assigned exclusively, we prevent any inflow into selected start cells through (18). Constraints (20) in combination with coupling constraints (19) limit the number of predecessor cells that provide inflow into a cell to two. Additionally, outflow from a cell would be forced to zero, if it received flows from two different neighbor cells. If, on the other hand, only one or zero neighbors provided inflow into the cell, it would be able to transmit flow to all connected neighbors. As a cell can receive flow from two different adjacent cells, its maximum acceptable inflow is $2q^{\max}$ where $q^{\max} = \max_{c \in \mathcal{C}} \{q_c\}$. Due to (20) cells cannot bundle and transmit flows from different start cells. Hence, each individual flow cannot be greater than $q^{\max}$, which we use as *big-M* value in the coupling constraints (19).

We solve **P2** by an iterative fix-and-optimize heuristic. We sort the camps by the value of their sequence parameter s_c in ascending order. In each iteration, the algorithm moves camps from the top of the sorted list to the problem instance of **P2**. After solving, we fix the start position decisions $Y_{c,u}$ and move to the next iteration. The procedure stops after solving the partially fixed **P2** instance with all camps. Constraints (18)-(20) are imposed to make the flows uniquely identifiable. Hence, we can convert a solution of **P2** to a solution of **P1**.

Using the block assignment of **P2**, we solve **P1** for each block using a lexicographic approach. First, we optimize for the layout objective stated in (2) by setting $a_1 = 0$ and $a_2 = 1$. We then fix the value for L_2, set $a_1 = -1$, $a_2 = 0$ and solve for the sequence objective stated in (1).

We have implemented models **P1**, **P2**, and the solution algorithm in the algebraic modeling system GAMS 25.1. The real data set (Table 1) was processed with the CPLEX solver on a workstation with 256GB of memory and two Intel Xeon E5-2667v3 3.2-GHz processing units.

By applying the described approach all 62 subproblems were solved to optimality in the first run (layout objective L_2) with an average computation time of 6.2 cpu seconds and a coefficient of variation of 2.67cpu seconds. After fixing the value of L_2 in the first stage, 2 of the 62 subproblems were not solved to optimality after a maximum computation time of 300s. The maximum optimality gap was 7.8%.

Details on computation times for **P1** are shown in Table 2. Column "Req/Cap" displays the space requested by the assigned camps divided by the space capacity for every block. This ratio is close or equal to 1 for the instances with the highest computation times. It seems intuitive that solutions to these instances also peak in the number of split cells (i.e., not exclusively assigned cells). Allowing the partial assignment of cells to a camp effectively prevents the generation of infeasible block instances which would otherwise occur during the two-stage solution approach. Table 2 also shows that instance size, expressed in the number of variables (#Var) and equations (#Equ), was not the dominant factor for computation times in our data set. Some of the largest instances, for example, blocks 80, 43, 64, did not exhibit particularly high computation times.

The fix and optimize heuristic for **P2** finished after 38 iterations and 701.38 cpu seconds. In each of the iterations between five and ten camps were added to the problem. We have employed P2 for the construction phase, because it contains binary and nonnegative flow variables for adjacent cells only. Compared to **P1**, which contains variables for each feasible pair of source cell and consumable cell, the number of binary variables in P2 reduces by more than 77%. In contrast to, for example, a more simple camp-to-block assignment model with aggregated capacity consideration, contiguity and flow constraints of **P2** always ensure that its solutions are convertible to feasible starting points for **P1**.

Table 2: Characteristics and computation time of **P1** instances

Block	Camps	Units	Req/Cap	Splits	#Equ	#Var	Cpu sec [L2]	Cpu sec [L1]
BL93	8	112	0.92	2	7892	7961	102.9	225.3
BL67	12	81	0.96	5	8361	8643	83.1	305.1
BL57	7	52	1.00	10	4204	4325	29.6	300.7
BL83	7	101	0.97	2	11756	11924	15.8	81.8
BL70	9	35	1.00	10	1320	1380	13.3	72.4
BL91	5	34	1.00	8	1536	1579	13.2	32.5
BL55	6	41	0.97	2	1705	1738	8.8	5.9
BL77	7	31	0.92	6	1262	1315	7.7	15.3
BL56	4	40	1.00	5	3457	3549	7.1	7.5
BL96	4	65	0.90	0	4016	4040	6.9	45.5
BL37	5	49	1.00	7	2962	3014	6.8	31.3
BL47	4	50	0.94	0	6251	6388	5.6	6.3
BL64	4	81	0.89	0	8646	8725	4.8	5.1
BL69	7	35	1.00	9	1994	2081	4.7	7.0
BL79	8	57	0.83	0	4026	4137	4.6	62.2
BL78	8	55	0.88	2	5190	5367	4.5	7.5
BL54	4	40	1.00	5	2160	2204	4.4	4.0
BL68	8	58	0.93	0	5094	5255	4.3	6.1
BL36	3	60	0.94	2	4559	4600	4.0	5.0
BL38	7	68	0.93	2	3839	3896	4.0	116.6
BL75	7	44	0.89	0	1873	1923	3.1	3.8
BL65	8	50	0.88	0	4105	4259	3.0	176.0
BL84	4	64	0.94	0	3707	3728	2.8	3.7
BL76	10	49	0.92	0	3181	3325	2.7	4.4
BL53	6	40	0.98	4	2693	2792	2.6	1.1
BL81	4	72	0.61	0	5569	5608	2.2	1.7
BL97	6	99	0.90	3	4563	4653	2.2	1.6
BL43	4	90	0.32	0	9208	9268	2.1	3.0
BL40	5	40	1.00	6	1206	1218	1.9	1.7
BL63	6	86	0.69	0	7342	7426	1.8	2.6
BL62	3	66	0.79	0	5157	5190	1.6	1.6
BL50	5	23	0.95	4	947	991	1.4	1.2
BL89	3	54	0.85	0	4509	4562	1.4	1.5
BL42	4	55	0.36	0	4181	4235	1.4	1.8
BL41	3	68	0.41	0	5134	5154	1.3	1.9
BL58	5	24	1.00	6	705	729	1.2	1.3
BL46	4	31	0.96	2	1076	1093	1.1	1.1
BL95	3	51	0.60	0	5700	5788	1.0	1.9
BL60	4	27	0.99	5	984	1010	1.0	1.4
BL74	5	35	0.96	4	1064	1081	0.9	1.3
BL85	3	35	0.52	0	1280	1289	0.9	0.5
BL48	2	37	0.90	0	2977	3017	0.8	0.5
BL86	3	48	0.70	0	2626	2642	0.8	0.8
BL80	1	104	0.09	0	10504	10474	0.7	0.4
BL61	4	20	0.88	0	662	691	0.7	0.8
BL49	4	46	0.86	0	3811	3887	0.7	1.2
BL66	5	36	0.63	0	1721	1763	0.6	0.9
BL71	5	39	0.76	0	884	878	0.6	0.2
BL45	2	49	0.97	0	2124	2119	0.6	0.7
BL72	5	31	0.88	0	1019	1043	0.6	0.4
BL90	2	36	0.88	0	1606	1616	0.6	0.3
BL52	2	37	1.00	2	2100	2116	0.5	0.4
BL88	2	40	0.85	0	917	898	0.4	0.3
BL73	3	21	0.77	0	631	651	0.4	0.3
BL51	4	19	0.90	0	484	499	0.4	0.4
BL92	2	38	0.91	0	1450	1447	0.3	0.3
BL94	2	26	0.62	0	1319	1343	0.3	0.3
BL59	2	29	0.67	0	437	422	0.3	0.3
BL39	1	25	0.91	0	659	652	0.2	0.0
BL82	1	31	0.76	0	1286	1284	0.2	0.1
BL98	1	43	0.45	0	963	935	0.1	0.1

(a) Flow assignment based on P2 (b) Improved layout solution based on P1

Figure 6: Construction and improvement

Figure 6 highlights the difference between the network flow solution and the final solution of **P1** for two blocks. In Figure 6a, camps consume minimum space but might form irregular shapes like the green camps no. 222 and 192. The second solution in Figure 6b creates a more rectangular and regular camp pattern. Moreover, all camps in 6b have access to a street or road entry point.

A complete solution is presented in Figures 9 and 10. Generated layout plans can serve as a basis to engage in constructive discussions with the stakeholders. Before finalizing the layout, planners might adjust the calculated solution for several reasons. First, a small fraction of the camps may still not feature the preferred shape. The planner can rearrange most of the camps as there is excess space in many blocks. Second, cells may be rendered unusable on short notice. In the past, this happened due to construction and cable works for example. Additionally, non-pilgrim entities might require space close to the metro on short notice as well. Examples for such entities are logistics operations for security, first aid staff, and pilgrim organizational staff.

Figure 10 also shows the camp alignment according to the scheduling sequence. In the illustration, dark zone colors correspond to high values of sequencing parameter s_c. According to objective L_1, we locate light-colored camps close to the stations. Note that we consider the shortest path distances in the network for L_1, which may contain detours, stairs, and ramps. Moreover, the solution space is constrained by minimum space requirements, the network structure and the consideration of a second objective. As a result, most but not all camps are aligned correctly with the scheduling sequence in Figure 10. As the number of misaligned camps on each road segment is limited, no immediate danger for crowd safety arises. During the Hajj operation, crowd management staff is deployed to critical points in the network to observe and react to the crowd flows.

5. Final Remarks

In this work, we formulate an unrooted regionalization problem with requirements on the spatial distribution, minimum area, and shape of the constructed regions. This regionalization problem arises in the context of pilgrim camp layout planning during the annual Hajj proceedings. We introduce a mathematical model and a two-stage solution approach to solve the problem for practical data sets. In the model, contiguity of each camp zone is established by path construction constraints, which we enhance to allow partial assignments of cells. As partially assigned cells cannot be used to extend a path, each camp zone constructs a core of fully assigned cells. The partial allocation of cells only occurs at the border between two camps. To the best of our knowledge, this approach has not been proposed in the literature yet. We have demonstrated that our solution algorithm can solve large, practical data instances within reasonable computation time. Moreover, we have discussed potential problems of an auto-generated layout solution.

In past Hajj seasons, a team of engineers constructed camp layouts manually in constant interaction with the pilgrim authorities. They usually planned under immense time pressure and began the process only days before the Hajj proceedings. Based on their layout solution, pilgrim dispatching schedules were then generated to prepare the departure operation from the Arafat stations for up to 500,000 pilgrims. As such the layout planning step is a central, time-critical aspect in the crowd management system of Hajj (Haase et al., 2016). The approach described in this paper is well-suited to be integrated into the pilgrim dispatching planning to streamline the whole process of Hajj pilgrim flow planning. Auto-generated plans may be the basis for a faster generation of final layout plans in future operations.

Further development of the approach may include a mechanism to control the spatial grouping of camps based on criteria such as nationality or common service providers. From a computational perspective, we encourage further research on the efficiency of alternative construction and improvement heuristics.

References

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

Cordone, R. and F. Maffioli (2004). On the complexity of graph tree partition problems. *Discrete Applied Mathematics 134*(1), 51 – 65.

Cova, T. J. and R. L. Church (2000, October). Contiguity constraints for single-region site search problems. *Geographical Analysis 32*(4), 306–329.

Drexl, A. and K. Haase (1999). Fast approximation methods for sales force deployment. *Management Science 45*(10), 1307–1323.

Fischer, D. T. and R. L. Church (2003). Clustering and compactness in reserve site selection: An extension of the biodiversity management area selection model. *Forest Science 49*(4), 555–565.

Fleischmann, B. and J. N. Paraschis (1988). Solving a large scale districting problem: a case report. *Computers & Operations Research 15*(6), 521 – 533.

Haase, K., H. Z. Al Abideen, S. Al-Bosta, M. Kasper, M. Koch, S. Müller, and D. Helbing (2016, February). Improving pilgrim safety during the hajj: An analytical and operational research approach. *Interfaces 46*(1), 74–90.

Hess, S. W., J. B. Weaver, H. J. Siegfeldt, J. N. Whelan, and P. A. Zitlau (1965). Nonpartisan political redistricting by computer. *Operations Research 13*(6), 998–1006.

Jafari, N. and J. Hearne (2013). A new method to solve the fully connected reserve network design problem. *European Journal of Operational Research 231*(1), 202–209.

Kalcsics, J. (2015). *Location Science*, Chapter Districting Problems, pp. 595 – 622. Springer, Cham.

Kalcsics, J., S. Nickel, and M. Schröder (2005, Jun). Towards a unified territorial design approach — applications, algorithms and gis integration. *Top 13*(1), 1–56.

Kong, Y., Y. Zhu, and Y. Wang (2019). A center-based modeling approach to solve the districting problem. *International Journal of Geographical Information Science 33*(2), 368 – 384.

Li, W., M. F. Goodchild, and R. Church (2013). An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *International Journal of Geographical Information Science 27*(6), 1227–1250.

Müller, S. (2015). *Spaces of Rites and Locations of Risk: The Great Pilgrimage to Mecca*, pp. 841–853. Dordrecht: Springer Netherlands.

Onal, H., Y. Wang, S. Dissanayake, and J. D. Westervelt (2015, June). Optimal design of compact and functionally contiguous conservation management areas. *European Journal of Operational Research 251*(3), 957–968.

Ricca, F. and B. Simeone (2008). Local search algorithms for political districting. *European Journal of Operational Research 189*(3), 1409 – 1426.

Shirabe, T. (2005). A model of contiguity for spatial unit allocation. *Geographical Analysis 37*(1), 2–16.

Shirabe, T. (2009). Districting modeling with exact contiguity constraints. *Environment and Planning B: Planning and Design 36*(6), 1053–1066.

Wang, Y., H. Önal, and Q. Fang (2018). How large spatially-explicit optimal reserve design models can we solve now? an exploration of current models' computational efficiency. *Nature Conservation 27*, 17–34.

Webster, G. R. (2013). Reflections on current criteria to evaluate redistricting plans. *Political Geography 32*, 3 – 14.

Williams, J. C. (2002, October). A zero-one programming model for contiguous land acquisition. *Geographical Analysis 34*(4), 330–349.

Williams, J. C., C. S. ReVelle, and S. A. Levin (2005, Sep). Spatial attributes and reserve design models: A review. *Environmental Modeling & Assessment 10*(3), 163–181.

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

Zoltners, A. A. and P. Sinha (2005). Sales territory design: Thirty years of modeling and implementation. *Marketing Science 24*(3), 313–331.

Appendix

A. Model Symbol Enumeration

(a) sets

$\mathcal{P}$ partition (combination of station and platform), indexed p,

$\mathcal{U}$ layout cells,indexed u, v, w,

$\tilde{\mathcal{U}}$ origin cells for layout construction, $\tilde{U} \subset U$

$\tilde{\mathcal{U}}_c$ feasible start cells for camp c (partition),

$\mathcal{C}_u$ feasible camps for layout cell u

$\mathcal{K}_u$ layout cells in range for layout construction from origin cell u

$\mathcal{N}_u$ adjacent cells of cell u

$\tilde{\mathcal{N}}_{u,v}$ adjacent cells of cell v closer to potential source cell u than cell v itself

$\mathcal{U}_u^{\textsf{fix}}$ set of cells fixed to start cell u

$\mathcal{U}^{\textsf{street}}$ set of cells with direct access to a street

(b) parameters

q_c space demanded by camp c

s_c requested departure sequence of camp c

a_1, a_2 objective function weight coefficients

$\beta_{u,v,w}$ weight of pair (v, w) for layout from origin u

(c) variables

$Y_{c,u}$ 1 if camp c is assigned to start cell u

$X_{u,v}$ 1 if cell v is fully occupied by camp assigned to start cell u

$\tilde{X}_{u,v}$ share of space from cell v allocated to camp assigned to start cell u

$Z_{u,v,w}$ 1 if the pair (v, w) is part of camp assigned to start cell u

$F_{u,v}$ flow (of space-requirement) between cell u and v ($v \in \mathcal{N}_u$)

$H_{u,v}$ 1 if solution includes flow between cell u and v

B. Additional Figures