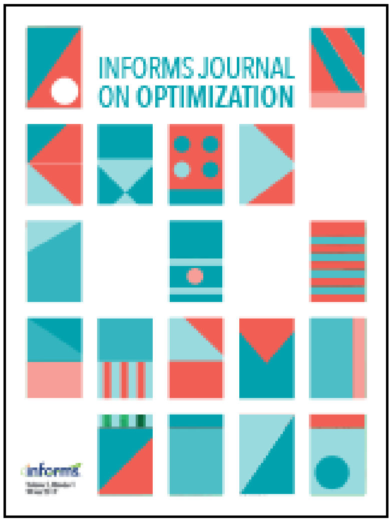


INFORMS Journal on Optimization

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Spatial Separability in Hub Location Problems with an Application to Brain Connectivity Networks

Taghi Khaniyev, Samir Elhedhli, Fatih Safa Erenay

informs
http://pubsonline.informs.org/journal/ijoo

INFORMS JOURNAL ON OPTIMIZATION
Vol. 2, No. 4, Fall 2020, pp. 320–346
ISSN 2575-1484 (print), ISSN 2575-1492 (online)

Spatial Separability in Hub Location Problems with an Application to Brain Connectivity Networks

Taghi Khaniyev,[a] Samir Elhedhli,[b] Fatih Safa Erenay[b]

[a] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [b] Department of Management Sciences, University of Waterloo, Waterloo N2L 3G1, Ontario, Canada

Contact: khaniyev@mit.edu, https://orcid.org/0000-0002-2160-0553 (TK); elhedhli@uwaterloo.ca, https://orcid.org/0000-0003-1922-8286 (SE); ferenay@uwaterloo.ca, https://orcid.org/0000-0002-3408-0366 (FSE)

Received: October 15, 2018
Revised: December 5, 2019
Accepted: February 9, 2020
Published Online in Articles in Advance: July 31, 2020

https://doi.org/10.1287/ijoo.2019.0031

Abstract. Motivated by the need to solve large hub location problems efficiently and accurately, we discover an important characteristic of optimal solutions to p-hub median problems that we call spatial separability. It refers to the partitioning of the network into allocation clusters with nonoverlapping convex hulls. We illustrate numerically that the property persists over a wide range of randomly generated instances and propose a data-driven approach based on an insight from the property to tackle very large problem sizes. Computational experiments corroborate the effectiveness of the proposed approach in generating high-quality solutions within reasonable computational times. We then explore a new application area of hub location problems in brain connectivity networks and introduce the largest and the first set of three-dimensional instances in the literature. Computational results demonstrate the capability of hub location models in successfully depicting the hub organization of the human brain, as validated by the medical literature, thus revealing that hub location models can play an important role in investigating the intricate connectivity of the human brain.

Keywords: hub location • brain connectivity • spatial separability

1. Introduction

The recent interest of the medical and neuroscience community in building a "network map," also known as "connectome," of the human brain lead to the rapid development of medical imaging technologies that can probe the neural circuitry of human brains with high spatiotemporal resolution. One of these technologies is diffusion tensor/spectrum imaging (DTI/DSI), which is used to map putative white matter pathways between different regions of the brain based on the diffusion of water molecules in the brain (Oldham and Fornito 2018). Despite recent advances in the resolution of such imaging technologies, it remains impractical to achieve a spatial resolution on the single-neuron level given the billions of neurons in an adult human brain. Hence, concepts like *region*, *parcel*, *cluster*, and *community* are widely used in the neuroscience literature to lump together nodes (brain regions) that are anatomically and functionally similar and are spatially contiguous. Moreover, using DTI/DSI data, researchers have shown that some regions in the brain cortex play a *hub* role and act as an intermediary to transmit signals between different regions (van den Heuvel and Sporns 2013). This is believed to provide an evolutionary advantage in reducing the size of the brain (Bullmore and Sporns 2012). As damage to these regions causes serious problems (van den Heuvel et al. 2010, Zalesky et al. 2011), it is important to identify these hub regions as precisely as possible.

Researchers have hitherto used descriptive graph-theoretical methods (Crossley et al. 2013, Bassett and Sporns 2017) to study the connectivity networks of the brain and identify hub regions. Although it is generally understood that identifying clusters is directly related to identifying hubs in a given network, the current literature often handles the two problems separately. Specifically, to determine the clusters, methods such as community detection, which try to maximize intracluster connection density, are used (Newman 2006). On the other hand, to identify hub regions, node attributes, such as degree, strength, coreness, centrality, and efficiency, are used without explicitly accounting the clusters identified based on the former analysis. This disconnect is a major handicap to fully understand the very intricate nature of brain connectivity networks (BCN), which encompasses traits of both high local modularity and interregional integration at the same time (Bullmore and Sporns 2012). Therefore, a good approach to identifying hub regions in BCN should be able to account for the interdependence of cluster and hub choices, which could be tackled using mathematical optimization. In particular, hub location problems (HLPs) are a family of mathematical models that are designed to account for such interrelated choices. To the best of our knowledge, identifying hub-and-cluster topology in BCN has never been studied under an optimization lens, which is what we study in this paper.

Hub location is a well-established area of operations research with over 30 years of contributions. Following the seminal work by O'Kelly (1986), the field enjoyed a proliferation of new models, solution methodologies, and applications. Current models are capable of incorporating practical features, such as capacity constraints (Campbell 1994, Ebery et al. 2000, Boland et al. 2004, Labbé et al. 2005, Correia et al. 2010, Contreras et al. 2011c); congestion (Elhedhli and Hu 2005, de Camargo et al. 2009b, Elhedhli and Wu 2010); incomplete hub network topologies (Campbell et al. 2005a, b; Alumur et al. 2009; Calık et al. 2009; Contreras et al. 2017); and uncertainty (Marianov and Serra 2003, Contreras et al. 2011b, Alumur et al. 2012).

Proposed solution methodologies range from decomposition-based methods, such as Lagrangian relaxation (Aykin 1994, Pirkul and Schilling 1998, Elhedhli and Hu 2005), branch-and-price (Contreras et al. 2011c), and Benders decomposition (de Camargo et al. 2009a, Contreras et al. 2011a, Ghaffarinasab and Kara 2019) to heuristics, such as genetic algorithms (Topcuoglu et al. 2005), tabu search, greedy randomized adapted search procedure (GRASP) (Klincewicz 1992, Mayer and Wagner 2002), Variable Neighborhood Search (VNS) (Ilić et al. 2010), and Euclidean projection (Meier and Clausen 2017). On the application side, HLPs have been introduced in the context of airline transportation (O'Kelly 1987, Martin and Román 2004), postal delivery (Ernst and Krishnamoorthy 1996), and telecommunication (Klincewicz 1998). The reviews by Alumur and Kara (2008), Farahani et al. (2013), and Contreras (2015) provide a thorough survey of the state of the art in the field, whereas Campbell and O'Kelly (2012) discuss the past and the future of hub location research and possible fertile research avenues.

One of the main challenges of using HLP in certain promising domains, such as BCN, is the inability of current solution approaches to handle large instances (e.g., networks with thousands of nodes). We believe that HLP can be a useful tool in identifying the precise locations of hub regions in BCN provided that the resulting large instances can be solved efficiently. Most state-of-the-art methods, with the exception of Ilić et al. (2010) and Guan et al. (2018)), are capable of solving instances with only a few hundred nodes within a reasonable amount of time. Considering the number of neurons in an adult human brain, one can argue that being able to optimally solve a problem with a few hundred nodes is not appealing enough by itself from the practical point of view as it would yield very low precision. To make the mathematical HLP models more relevant for practical purposes, general-purpose methods that can efficiently tackle very large problem instances should be available to end users. We propose one such method in Section 3.

Another potential issue with the application of HLPs to BCN is a structural pattern consistently observed in the neuroscience literature regarding the hub-and-cluster topology: Clusters, irrespective of their sizes, always appear to be spatially contiguous (Arslan et al. 2018). An alternative way of expressing this is that the volumes enclosed by the sets of nodes in different clusters do not intersect with each other. Hence, it is important that any proposed methodology for identifying the hub-and-cluster topologies in BCN incorporates this structural property in some way. Although spatial contiguity has been studied in the operations research literature in application areas such as forest planning (Carvajal et al. 2013) and political districting (Shirabe 2009, King et al. 2012), to the best of our knowledge, it has not been explicitly studied in the context of hub location. In Section 3, we analyze the spatial patterns in the known HLP instances, and discover that the optimal hub-and-cluster topology is typically composed of disjoint clusters with nonoverlapping convex hull which is a special case of spatial contiguity we introduce as "spatial separability." Although it is not a general property, extensive testing on randomly generated instances reveals that spatial separability holds in 99.5% of the time. For the remaining 0.5% of the cases, we propose a set of constraints that can be used to mathematically impose this observed form of spatial contiguity in the HLP models. In Section 4, we exploit an insight derived from this structural property to devise a new data-driven approach to solve large HLPs.

It is not uncommon that methodologies that work well on moderate-size instances become intractable as the problem size increases. Hence, it is essential to test the performance of proposed methodologies not only on small-to-moderate-size problems but also on very large problem instances. The largest frequently cited benchmark instance has only 200 nodes (Ernst and Krishnamoorthy 1996), and the largest ever proposed has 1,000 nodes (Ilić et al. 2010, Guan et al. 2018). As there is always an interest in new large data sets originating from emerging practical applications, we introduce the BCN data set as one of the largest (with $n = 998$ nodes) and the first three-dimensional (3D) data set in the literature so far. The BCN data set, originally generated by Hagmann et al. (2008), is reintroduced in Section 2 in a way amenable to existing hub location models. In Section 5, we generate benchmark instances using the BCN data set to evaluate the performance of the proposed approach.

1.1. Problem Description

In this study, we focus on a specific subclass of HLPs and its application to BCN: uncapacitated single-allocation p-hub median problem (USApHMP) in which all the hubs are assumed to be fully interconnected; each node in the network is allocated to a single hub, the hubs are uncapacitated, and the total number of hubs, p is specified a priori. Motivated by the BCN application, we would like to incorporate the aforementioned spatial separability property to USApHMP, a fundamental subclasses of HLP that only requires application-specific data on the coordinates of the nodes, and the volumes of pairwise flows between the nodes. In Section 2, we describe the data available for the BCN application, which are just sufficient to build a USApHMP model. Note that other HLP variants may need additional data, for example, capacitated HLP models would require data on the capacities of each node/arc.

One of the first linear mixed integer programming formulations for USApHMP was introduced by Skorin-Kapov et al. (1996), which had $O(n^4)$ variables and $O(n^3)$ constraints. Although the formulation has been shown to provide very tight linear programming (LP) bounds, it does not scale well with the number of nodes. Even for the problem instances of moderate size (between 50 and 100 nodes), commercial solvers reach the memory limit. The formulation proposed by Ernst and Krishnamoorthy (1996) requires a smaller number of variables and constraints ($O(n^3)$ and $O(n^2)$, respectively). Although it has a looser LP bound compared with the formulation of Skorin-Kapov et al. (1996), commercial solvers can solve problem instances up to 100 nodes using this formulation. Ebery (2001) presented another formulation that requires $O(n^2)$ variables and $O(n^2)$ constraints. This formulation uses fewer variables than all of the other models previously presented in the literature. However, in practice, the computational time required to solve this new formulation was shown to be greater than that required to solve the formulation by Ernst and Krishnamoorthy (1996). Therefore, in this paper, we use the formulation by Ernst and Krishnamoorthy (1996) as the mathematical model for USApHMP. Table 1 provides the list of important notations used throughout the paper, and Definition 1 formally states the USApHMP and provides the mixed integer programming (MIP) formulation proposed by Ernst and Krishnamoorthy (1996).

Definition 1. Let $\mathcal{N} = \{1, 2, \ldots, n\}$ be a set of nodes in a given coordinate system. For each node pair (i, j), let w_{ij} represent the volume of flow and d_{ij} represent the Euclidean distance between the two nodes adhering to triangular inequality. The traffic (flow) between nodes is to be routed via a set of nodes designated as hubs. All hubs are assumed to be fully interconnected, and each node must be allocated to exactly one hub. A collection (distribution) cost is incurred to route traffic from (to) an origin (destination) node to (from) a hub. Moreover, a discounted transfer cost is incurred to send flow between hubs, where the discount factor α is constant. Given p, the number of hubs to locate on the network, the USApHMP is the problem of finding the optimal hub locations as well as the allocation of the remaining nodes to hubs such that the overall cost of satisfying the flow demand is minimized and can be formulated mathematically as follows:

$$
\begin{aligned}
[\text{USA}p\text{HMP}] \quad \min \ & \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{N}}\sum_{m\in\mathcal{N}} \alpha d_{km} x_{ikm} + \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{N}} (O_i + D_i) d_{ik} y_{ik}, && && \text{(1a)}\\
\text{s.t.} \ & \sum_{k\in\mathcal{N}} y_{kk} = p, && && \text{(1b)}\\
& \sum_{k\in\mathcal{N}} y_{ik} = 1 && \forall i \in \mathcal{N}, && \text{(1c)}\\
& y_{ik} \le y_{kk} && \forall i, k \in \mathcal{N}, && \text{(1d)}\\
& \sum_{m\in\mathcal{N}} x_{ikm} - \sum_{m\in\mathcal{N}} x_{imk} = O_i y_{ik} - \sum_{j\in\mathcal{N}} w_{ij} y_{jk} && \forall i, k \in \mathcal{N}, && \text{(1e)}\\
& x_{ikm} \ge 0 && \forall i, k, m \in \mathcal{N}, && \text{(1f)}\\
& y_{ik} \in \{0, 1\} && \forall i, k \in \mathcal{N}. && \text{(1g)}
\end{aligned}
$$

Table 1. Descriptions of Parameters and Decision Variables

Notation	Description
α	Interhub discount factor
w_{ij}	The amount of flow from node i to node j
d_{ij}	The distance from node i to node j
$O_i = \sum_j w_{ij}$	Total amount of outflow from node i
$D_i = \sum_j w_{ji}$	Total amount of inflow to node i
x_{ikm}	Total amount of flow emanating from node i that is routed between hubs k and m
y_{ik}	Binary variable indicating whether node i is allocated to hub k

It is worth mentioning that despite the limitations mentioned above on using these formulations with general-purpose solvers for large problem instances, it is well known that decomposition algorithms usually work well to handle such formulations. In fact, there has been recent progress in the development of algorithms capable of optimally solving large-size instances for the considered problem (Meier and Clausen 2017, Ghaffarinasab and Kara 2019) and heuristically solving even larger ones (Ilić et al. 2010, Guan et al. 2018).

1.2. Contributions

Motivated by the need to solve large practical problem instances, our goal in this study is to investigate the spatial patterns observed in the optimal solutions of USApHMP instances, elucidate the connection between USApHMP and several other problems, and propose a new approach exploiting these insights to tackle the largest problem instance in the literature, efficiently. Specifically, our contributions can be listed as follows:

- Identify and mathematically express a new spatial pattern in which nodes are allocated to hubs in optimal solutions of USApHMP and illustrate its consistent appearance over a wide range of random problem instances.
- Introduce brain connectivity networks as a new application area for HLPs together with the data set for the largest (998 nodes) and first three-dimensional problem instance in the literature.
- Propose a data-driven approach that uses the insights generated from the solution of a smaller problem (low-resolution representation) to find high-quality solutions for the original problem.
- Conduct comprehensive computational experiments to provide evidence for the effectiveness of the proposed approach and discuss the implications of the observed pattern and potential extensions to the proposed approach from modeling and BCN application perspectives.

The rest of the paper is organized as follows: In Section 2, we introduce brain connectivity networks as a new application for the HLPs. In Section 3, we formally define the spatial separability concept. In Section 4, we propose a data-driven approach to tackle large problem instances. In Sections 5, we report results from computational experiments. Brief discussions on future research avenues and conclusion are provided in Sections 6 and 7, respectively.

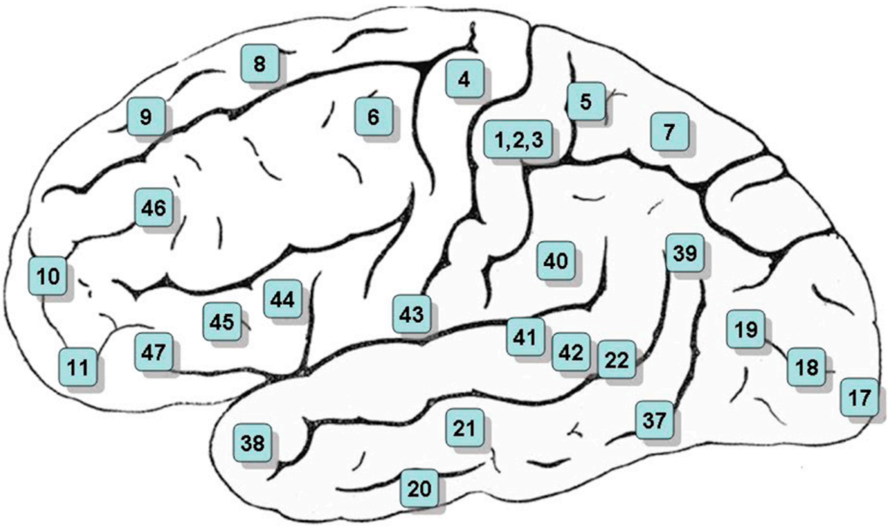
2. Brain Connectivity Networks

The human brain is one of the most complex systems known. It has approximately 86 billion neurons in a human brain, and each neuron, in turn, is connected to a number of other neurons. The cell bodies of neurons are located at the outer shell (cortex, or gray matter) of the brain, whereas the axons (the fibers that connect cell bodies) are located inside the cortex (white matter). It is now understood that every single activity we do, every thought that comes to our minds, every emotion we experience, and every memory we have is coded in our brains. Evidence suggests that cognitive processes ranging from visual recognition (Behrmann and Plaut 2013) and language (Friederici and Gierhan 2013) to emotion (Pessoa 2012) are results of dynamic interactions, via instantaneous electrical impulses known as action potential, between different regions in our brains. Therefore, a complete understanding of how our brains work will only be possible once we are able to map the entire brain network.

The Human Connectome Project (HCP), initiated in 2013, aiming to create a complete map of the anatomic and functional connections in the human brain is probably one of the biggest collective projects involving scientists from various disciplines. The two biggest challenges of this project are (1) to probe the resolution levels that allow distinguishing every single neuron and (2) to store and process the data emanating from such high-resolution imaging. One of the methods that proved very useful in abating the enormity of the task is to cluster neurons based on their locations and the observed functions. This allows us to work on the resolution level of brain areas instead of single neurons. Figure 1 shows an example of how the brain is divided into different areas commonly referred to as Brodmann areas. This parcellation of the brain cortex into 52 areas based on histological properties of neurons was proposed by Brodmann (1909) and has since been widely accepted by neuroscientists. The recent advances in imaging technologies allowed scientists to improve the parcellation proposed by Brodmann (1909). Glasser et al. (2016) proposed a new parcellation with 180 areas per hemisphere bounded by sharp changes in cortical architecture, function, connectivity, and/or topography using multimodel magnetic resonance images from the HCP database. Based on the current trends, it is safe to say that as our understanding of the brain improves, our need for higher-resolution data acquisition and processing will increase proportionally too. Hence, tools that allow processing high-resolution data will be in high demand.

One of the revelations of the recent advances in brain studies is the realization that the functional specialization of brain regions alone cannot fully account for most aspects of brain function, and those dynamic

Figure 1. (Color online) Brodmann Areas in the Human Brain (Commons 2008)



interactions across multiple regions are required to accomplish cognitive processes (van den Heuvel and Sporns 2013). One aspect of brain organization, particularly significant in enabling integration of distributed neural information is that important integrative functions are performed by a specific set of brain regions and their anatomical connections. In recent years, brain's anatomical and functional organization has been approached from the perspective of complex networks, and numerous studies described neural systems in terms of graphs or networks comprising nodes (neurons and/or brain regions) and edges (synaptic connections, interregional pathways) (Fox et al. 2006, Honey et al. 2007, Vincent et al. 2007). The comprehensive network map of the nervous system of a given organism is called its *connectome* and represents a structural basis for dynamic interactions to emerge between its neural elements (Le Bihan et al. 2001, Hagmann et al. 2003). Several studies have independently concluded that human connectome combines attributes pertaining to specialization with attributes that ensure efficient communication (integration). At the center of the latter lies network elements that are often referred to as hub nodes/regions, which are generally characterized by their dense connectivity to other regions and their central placement in the network.

Identifying the locations of hub regions as precisely as possible is very important from the clinical perspective because any damage to these regions is likely to cause more serious problems, especially in cognitive functions requiring high integrative processes. This knowledge could have immediate implications for neurosurgery when deciding the location of surgical opening, identifying the trade-off between removing tumour regions and keeping hub regions, and applying/avoiding radiation to hub regions. Current literature on identifying the hub regions of brain relies highly on graph theoretical proxy measures (such as strength and betweenness) of connectivity and centrality. The advantage of using these proxy measures is the computational efficiency in that it takes only seconds to provide a "hubness" rank for all the network nodes. However, it is very common that a measure pertaining to connectivity provides completely different hubness ranks than those obtained by a measure of centrality. Researchers, then, typically average the ranks obtained by various measures to calculate an overall hubness rank.

Reconciling the discrepancy between connectivity-based and centrality-based ranks by a simple average carries the risk of being an oversimplification of the underlying relationship. In fact, it is known in the hub location literature that a high discrepancy between centrality and connectivity in a network is typically an indicator of an underlying multicenter hub topology, that is, not a single contiguous hub region but multiple hub regions across the network and a linear combination of the two measures often performs very poorly in identifying the optimal hub structure Klincewicz (1992). Therefore, one may argue that the mathematical models for the HLPs could potentially be a more proper way to account for the aforementioned discrepancy. HLP models provide not only a list of hub locations but also a complete list of flow pathways between pairs of nodes that elucidates the physical reason why a certain node is more likely to be a hub.

In Section 5.3, we provide comparative computational experiments to explore the potential of HLP models as an alternative tool in explaining the underlying hub organization of the human brain. We attempt to do that by using an existing data set from the literature by Hagmann et al. (2008). The data set is generated by dividing the brain cortex into 998 regions of interest (ROIs) with an average size of $1.5cm^3$ and measuring the density of white matter between ROIs using DSI technology. It includes the following information:

- *FiberDensity*: A 998×998 connection matrix of 998 ROIs measured as the thickness/density of white matter between every pair of ROIs.
- *EdgeLength*: A 998×998 distance matrix of 998 ROIs measured as the length of fiber between every pair of ROIs with nonzero fiber density. To account for the distances between ROI pairs with no fiber connection, another matrix *ShortestLength* is generated by calculating the shortest distances between every pair.
- *ROICoordinates*: A 3×998 matrix that consists of the three-dimensional coordinates of ROIs.
- *ROILabels*: A vector of size 998 that shows the Brodmann area each ROI belongs to.

Note that, henceforth, we will be using *FiberDensity* as a proxy for the flow matrix (**W**) used in hub location literature. Ideally, the flow matrix should represent the frequency of action potential between each pair of ROIs. This would be a good measure of the information flow between ROIs. However, such a measurement is not possible with the current technology. Hence, to construct connectivity networks, researchers typically resort to proxies like *FiberDensity* to represent the connection strength (edge weight) between the node pairs (Hagmann et al. 2008). This is often justified from an evolutionary point of view by asserting that if there was not a high frequency of information flow between two regions, natural selection would not have favored a high density of fiber tracts between these regions. In addition, a recent study by Kwon et al. (2019) presents a methodological pipeline for identifying influential nodes and edges in human brain networks based on the self-regulating biological concept adopted from the Physarum model. The basic concept underlying the Physarum model is that short and dense fiber tracts are the most effective for fluid transmission of information. This provides another justification for the use of *FiberDensity* as a proxy for the flow matrix.

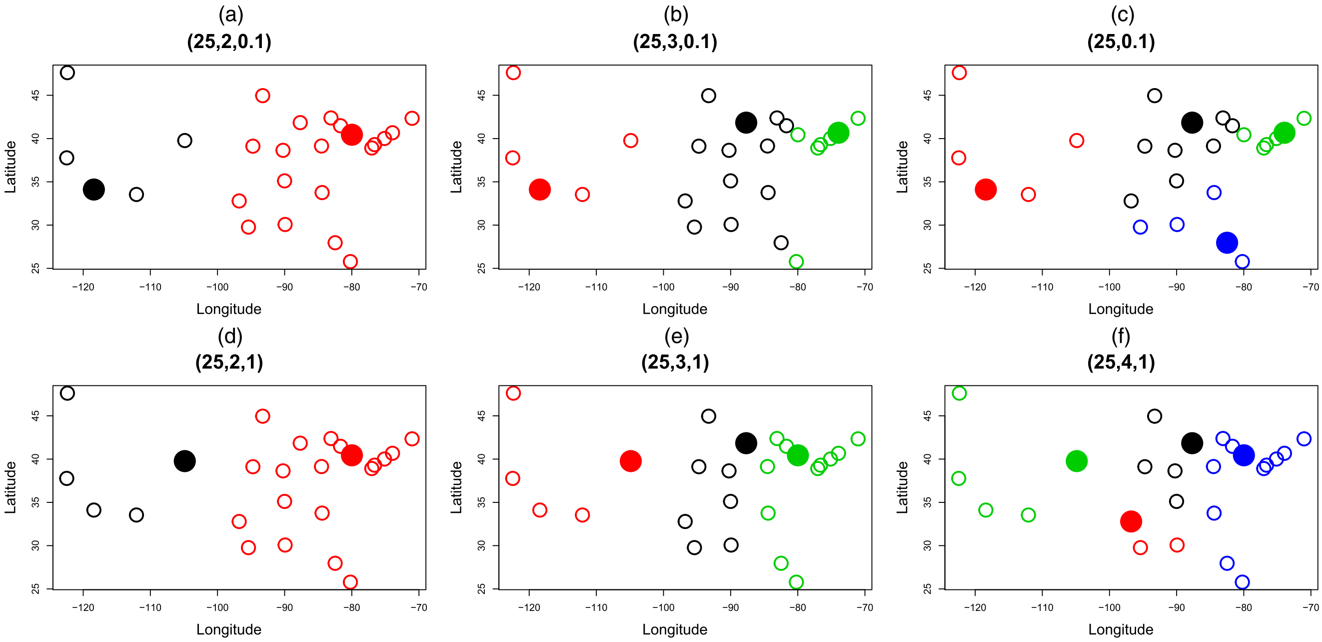
3. Spatial Contiguity and Separability

As mentioned in Section 1, the spatial contiguity of node clusters is an implicit requirement in the context of the BCN application. Although there exist studies in the literature which incorporate spatial contiguity into mathematical models, this is typically done for planar graphs/maps for which neighbors of each node/region are readily available. On both Civil Aviation Board (CAB) and Australia Post (AP) data sets, however, each node is represented by a point in two-dimensional (2D) space from which it is not straightforward to deduce the neighborhood information. Similarly, on the newly introduced BCN data set, each node is represented by a point in 3D space. Hence, in this section, we attempt to derive a concept similar to spatial contiguity when neighborhood information of the nodes is not explicitly available, and the nodes are represented as points on Cartesian coordinates, that is, data on the areas/volumes they occupy is not available.

Campbell and O'Kelly (2012, p. 166) argue that "the purpose of hub location research is insight, not numbers, and better insight is likely to result from the combination of improved understanding of the properties of hub location problems, better algorithms, new solutions, and a clearer understanding of the practical implications of the solutions." Several studies in the literature attempt to identify structures and patterns that would help mitigate the inherent difficulty of HLPs. O'Kelly (1986), for example, observes that in most cases nonhub nodes are allocated to the closest hub. However, a later paper by O'Kelly (1992), which exploits the parallelism between clustering problems and planar hub location problems, illustrates that the closest-hub allocation assumption is too strong of an assumption that could potentially lead to high optimality gaps. Campbell (1990) studies the spatial conditions under which single allocation (to the nearest hub) or multiple allocations (to the hub minimizing the distance between origin and destination) are optimal for a nonhub node. They find analytical expressions for the spatial regions around hub nodes where the nearest hub allocation is optimal. Peker et al. (2016) propose several spatial measures to evaluate node importance as a proxy for the likelihood of being a hub in the optimal solution and develop a heuristic that is based on restricting the solution space to sets most likely to contain hubs. One of the main contributions of this study is to reveal another spatial pattern of optimal USApHMP solutions, which has not been studied in the literature. The observed pattern motivates the solution methodology that is developed with the goal of tackling the newly introduced BCN instances with $n = 998$ nodes.

We start our analyses with the re-examination of a statement from O'Kelly (1992, p. 342): "The property that all observations which are closer to centroid A than to centroid B are assigned to the same group does not hold for HLPs. Therefore, we cannot use the strategy of examining only the nonoverlapping partitions." Specifically, we are interested in redefining the nonoverlapping partitions concept. To this end, we begin with a visual inspection of the optimal solutions of known problem instances and look for spatial patterns in these solutions. Figure 2 depicts the optimal solutions of six USApHMP instances with ($n = 25$) numbers of nodes from the CAB data set provided by O'Kelly (1986). Figure 1, (a)–(c) show the solutions for $p \in \{2, 3, 4\}$ hubs

Figure 2. (Color online) Optimal Solutions of Various CAB Instances



with interhub discount factor $\alpha = 0.1$, whereas Figure 1, (d)–(f), shows the solutions for $p \in \{2, 3, 4\}$ hubs with interhub discount factor $\alpha = 1$. In each map, hub nodes are marked with filled circle points, whereas nonhub nodes are marked with empty circle points. Moreover, nodes that are allocated to the same hub are shown with the same shade of gray. We observe that the nodes allocated to the same hub are geographically (spatially) clustered. To formalize this observation, we proceed with the following definition:

Definition 2. The set of nodes that are allocated to the same hub in a USA*p*HMP solution is called an allocation cluster.

Now the observation from Figure 2 can be rephrased as follows: Allocation clusters are nonoverlapping, that is, they are separated from each other with clear boundaries.

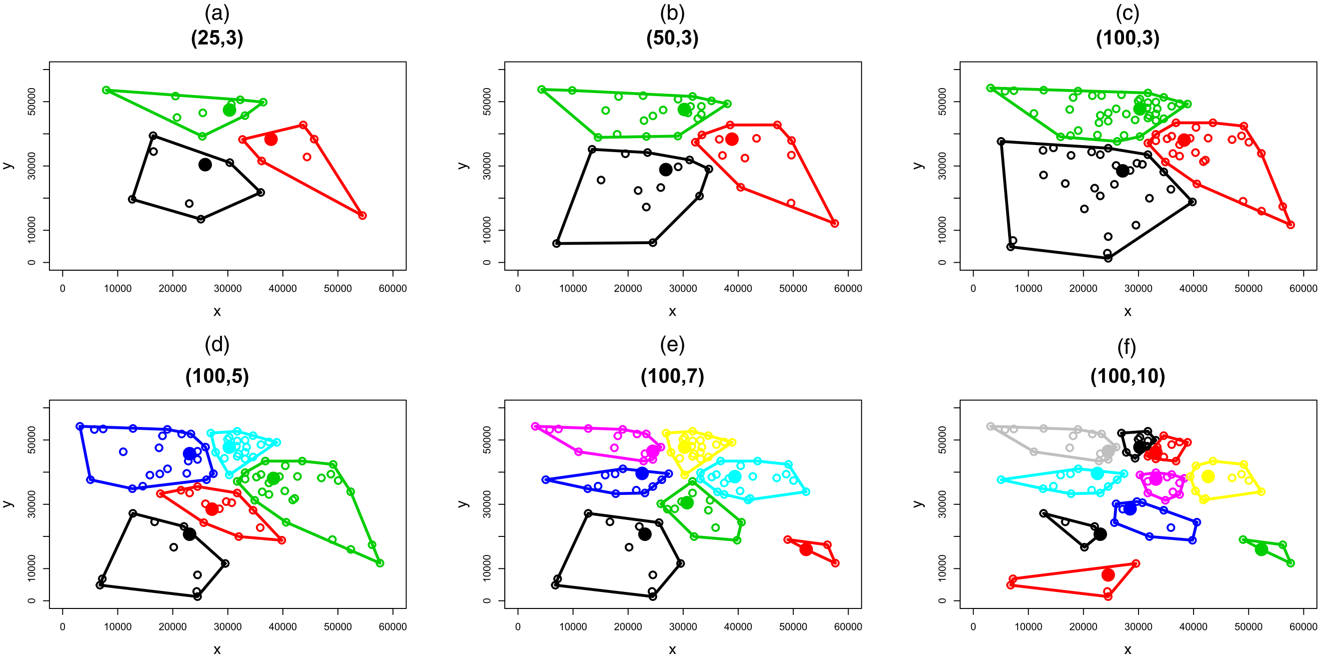
The aforementioned statement in O'Kelly (1992) appears to be in contradiction with the observation above. However, the subtle difference lies in how the clusters are defined. O'Kelly's (1992) definition of a partition involves clusters with a centroid and the subset of nodes for which that centroid is the closest among all centroids in the given partition. They experimentally show that such partitions are often nonoverlapping. By contrast, partitioning into allocation clusters does not specify a centroid for clusters. A consistent pattern whereby the regions encapsulating the nodes in each allocation cluster appear to be nonoverlapping has already been illustrated in Figure 2 for CAB instances. In an attempt to extend its validity for problem instances other than the CAB data set, and to provide a concise mathematical expression for the aforementioned pattern, the optimal solutions of USA*p*HMP instances from the AP data set with different numbers of nodes (n) and hubs (p) are obtained and depicted in Figure 3 (with $\alpha = 0.75$, $\chi = 3$, and $\delta = 2$ as given by Ernst and Krishnamoorthy (1996)). Figure 3, (a)–(c) show the solutions for three-hub problem with $n \in \{25, 50, 100\}$ nodes, whereas Figure 3, (d)–(f) show the solutions for $p \in \{5, 7, 10\}$ hubs with $n = 100$ nodes.

Each polygon in Figure 3 represents an allocation cluster. Observe that irrespective of the n and p values, the allocation clusters are always nonoverlapping. More specifically, the convex hulls of allocation clusters always appear to be disjoint. It is worth mentioning that we tested the validity of this observation with different (n, p, α) parameters in both CAB and AP data sets and observed no exception. We call this property *spatial separability* and formally define it as follows.

Definition 3. Let $A^i = (a_{i1}, a_{i2}, \ldots, a_{ik}), \forall i \in \mathcal{N} \equiv \{1, 2, \ldots, n\}$ be the coordinate vector of node i in k-dimensional space. A partition of nodes $\mathcal{N}$ into p (allocation) clusters $\mathcal{P} = \{P_1, P_2, \ldots, P_p | \bigcup_{i=1}^{p} P_i \equiv \mathcal{N}$ and $P_i \cap P_j \equiv \varnothing, \forall i, j \neq i\}$ is said to be spatially separable if the convex hulls of the clusters are (pairwise) disjoint, that is, $\text{Conv}(P_i) \cap \text{Conv}(P_j) \equiv \varnothing, \forall i, j \neq i$.

Definition 3 helps us remove the ambiguity in describing the observed spatial pattern by providing a mathematical expression for the spatial separability property. So far, we have consistently observed spatial

Figure 3. (Color online) Optimal Solutions of Various AP Instances



separability in the optimal solutions of the known USA*p*HMP instances. Figure 4, on the other hand, shows that there exist some extreme cases under which spatial separability is violated in the optimal solutions. Figure 4(a) is the simplest counter example in 2D space for which the optimal solution can be made to violate spatial separability by selecting appropriate values for the flows, for example, $w_{14} = 100, w_{13} = 1, w_{24} = 1$, and $w_{ij} = 0$ for all other pairs of nodes (where nodes are numbered from left to right). An interesting research question would be how likely these violating cases are to occur in real-life settings. To this end, in Section 5.1, we conduct an extensive computational analysis on randomly generated USA*p*HMP instances and show that these extreme cases occur so infrequently that overlooking them would not be completely unreasonable as long as this provides a computational advantage. Figure 4(b) is a typical example of nonseparable instances where spatial separability is violated only by a handful of nodes and relocation of those nodes so that spatial separability is restored can be done with a negligible increase in the objective value.

Partitioning into clusters with disjoint convex hulls has been studied in the literature. Chakravarty et al. (1985) study a related problem: Partition a given set of indexed items $(1, 2, \ldots, n)$ into p subsets $(P_1, P_2, \ldots, P_p)$ to minimize an objective function $g(P_1, P_2, \ldots, P_p)$. They provide several sufficient conditions on $g(.)$ that would guarantee that there is an optimum partition in which each subset consists of consecutively indexed items and show that this partitioning problem can be solved in polynomial time. Barnes et al. (1992)

Figure 4. (Color online) Counter Examples Violating Spatial Separability

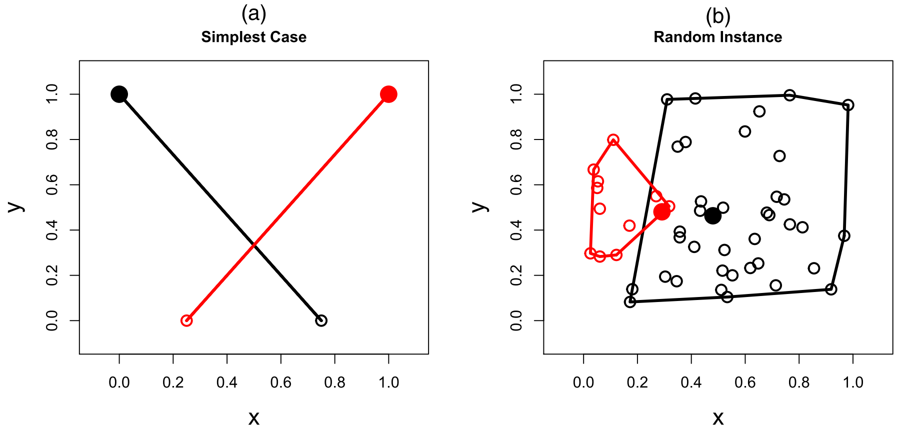
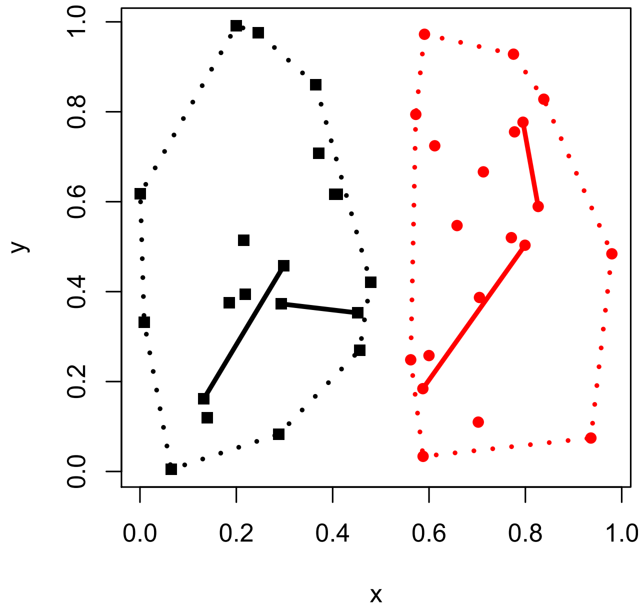


Figure 5. (Color online) A Spatially Separable Partition

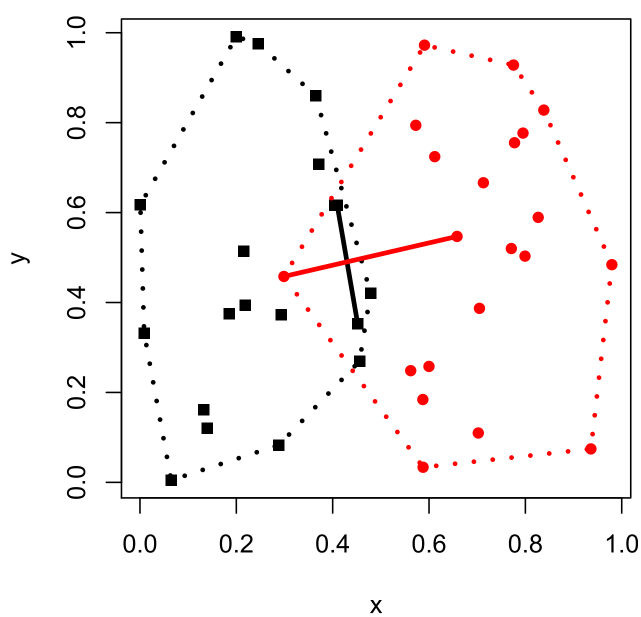


use results from Chakravarty et al. (1985) on consecutive optimizers and derive necessary and sufficient conditions on $g(.)$ such that there exists an optimum partition, which is spatially separable.

It is worthwhile to comment on an important implication of the aforementioned results for our study. The results of Chakravarty et al. (1985) can be applied to a restricted version of USApHMP in which an additional set of constraints is introduced to impose spatial separability. This would imply that the size of the feasible region (i.e., the search space) becomes polynomial in n. In other words, the computational effort to completely enumerate all feasible partitions would be reduced from exponential complexity $O(p^n)$ to polynomial complexity $O(n^p)$. Considering that many efficient heuristics for HLPs, such as tabu search and simulated annealing, depend on some kind of clever enumeration of feasible solutions, such a reduction in search space would have a substantial impact on computational performance of these heuristics.

We are, however, mainly interested in making practical use of spatial separability property. To this end, next we introduce the set of constraints that would impose spatial separability to HLPs. Figures 5 and 6 show a spatially separable and nonseparable partition, respectively, of the same set of nodes. An observation

Figure 6. A Nonseparable Partition



from Figures 5 and 6 allows us to impose spatial separability as a set of linear constraints to the MIP formulation of USApHMP: In a spatially separable partition, no line segment contained in one cluster intersects with a line segment contained in another cluster. Lemma 1 (given without proof) can be used to check whether two given line segments intersect.

Lemma 1. *Let (x_1, y_1) and (x_2, y_2) be the two endpoints of the line segment LS_1 and (x_3, y_3) and (x_4, y_4) be the endpoints of the line segment LS_2. Line segments LS_1 and LS_2 are said to intersect if the following conditions hold:*

$$\sqrt{(x-x_1)^2+(y-y_1)^2}+\sqrt{(x-x_2)^2+(y-y_2)^2}=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}, \tag{2a}$$

$$\sqrt{(x-x_3)^2+(y-y_3)^2}+\sqrt{(x-x_4)^2+(y-y_4)^2}=\sqrt{(x_3-x_4)^2+(y_3-y_4)^2}, \tag{2b}$$

where (x, y) is the intersection point of the *lines* defined by the endpoints of LS_1 and LS_2 and can be written as follows:

$$(x,y)=\left(\frac{(x_1y_2-y_1x_2)(x_3-x_4)-(x_1-x_2)(x_3y_4-y_3x_4)}{(x_1-x_2)(y_3-y_4)-(y_1-y_2)(x_3-x_4)}, \frac{(x_1y_2-y_1x_2)(y_3-y_4)-(y_1-y_2)(x_3y_4-y_3x_4)}{(x_1-x_2)(y_3-y_4)-(y_1-y_2)(x_3-x_4)}\right).$$

Note that Lemma 1 excludes the degenerate cases where the line segments are parallel to each other. Figure 7 illustrates different possible ways the lines defined by the endpoints of LS_1 and LS_2 can intersect, only one of which implies the intersection of LS_1 and LS_2.

To impose spatial separability, one needs to identify all pairs of intersecting line segments (i.e., quadruplets of nodes) and add constraints that prevent these line segments from being in different clusters. This would require $O(n^6)$ additional constraints: $O(n^4)$ for all possible intersecting line segments and $O(n^2)$ for all possible cluster pairs. This is a formidable number of additional constraints that would unnecessarily complicate the MIP formulation. An alternative approach would be similar to adding subtour elimination constraints for the Travelling Salesman Problem (Applegate 2006). Following Applegate's (2006) approach, one may opt to add only the cuts necessary to prevent the violating intersections in a given solution instead of adding all constraints pertaining to spatial separability at once. For example, assume that the two line segments, LS_1 and LS_2, are intersecting. Let nodes i, j and s, t be the endpoints of the LS_1 and LS_2 line segments, respectively. Assume also that in the given solution, the nodes i, j are included in cluster k and nodes s, t are included in a different cluster m. This is a case where spatial separability is violated, and we need to remove this solution from the feasible set by ensuring at least one of the two cases listed below is satisfied:

- Nodes i and j are not both in cluster k.
- Nodes s and t are not both in cluster m.

Introducing the following set of constraints —hereinafter called *Intersection Elimination Constraints (IEC)*— would eliminate this violation:

$$y_{ik}+y_{jk}-Mu\le 1 \tag{3a}$$

$$y_{tm}+y_{sm}-M(1-u)\le 1 \tag{3b}$$

$$u\in\{0,1\} \tag{3c}$$

In constraints (3a)–(3c), M is a very large number and u is an auxiliary decision variable to ensure that at least one of the two cases is satisfied. As one may notice, to eliminate each nonseparable solution, we need to add two constraints and one binary variable. Hence, ideally, we would like to use as few IECs as possible. In Section 5.1, we use IECs to impose spatial separability on the randomly generated nonseparable USApHMP instances to be able to calculate optimality gaps. A major drawback of using IECs is that it requires the ability

Figure 7. Different Possibilities of Intersections

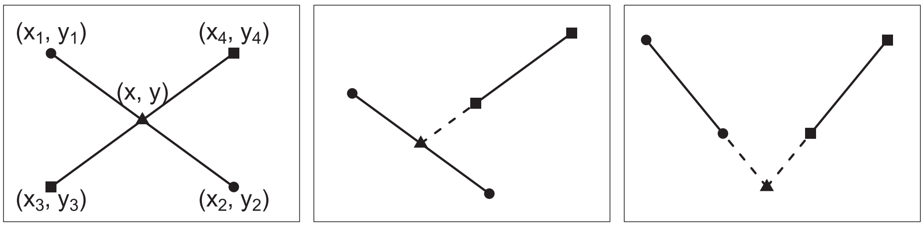
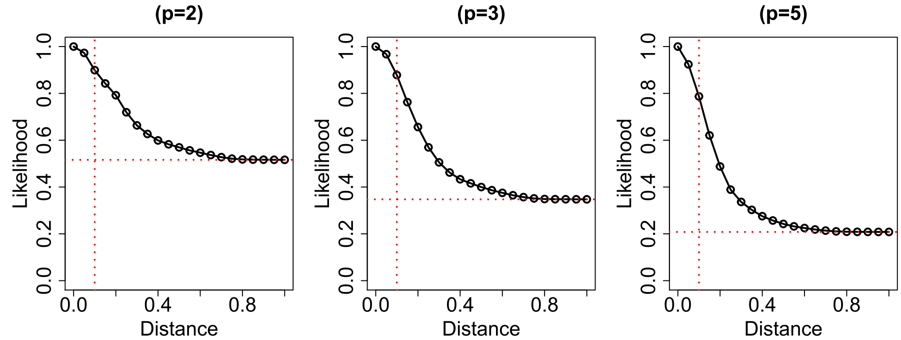


Figure 8. (Color online) Likelihood of Two Nodes Being in the Same Cluster as a Function of Normalized Distance for the Optimal Solutions of Various AP Instances



to input and process MIP formulation of an instance by a solver. This hinders the use of IECs for very large problem instances (e.g., BCN instances) as the number of variables and constraints becomes too large to process. Therefore, we focused on using an insight from this property (the nodes that are spatially closer are more likely to be in the same allocation cluster than those farther) to develop a heuristic. Figure 8 illustrates the likelihood of two nodes' being in the same allocation cluster with respect to the distance between them (normalized with the maximum pairwise distance) at the optimal solutions of several AP instances ($p = \{2, 3, 5\}$). It is evident from the figure that the likelihood of being in the same cluster is highly correlated with the distance between two nodes, especially if the distance between two nodes is very small (i.e., ~ 10% of the maximum distance). Beyond a certain distance, the correlation fades out and the likelihood converges to the base likelihood. In the next section, we develop a new approach that exploits this insight.

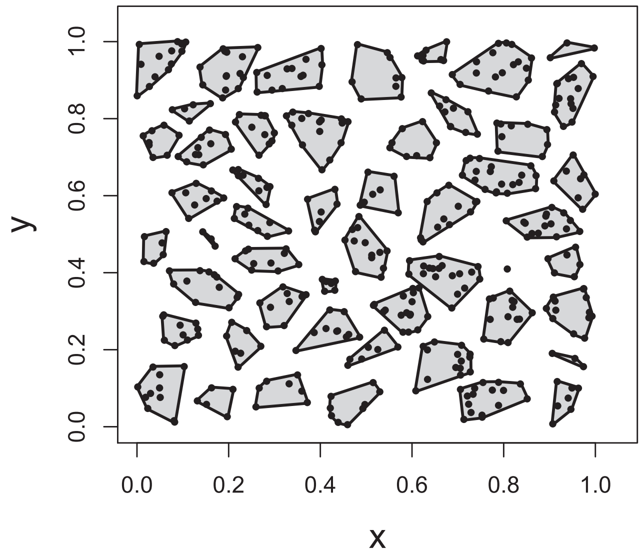
4. Approach

Despite the abundance of efficient algorithms to find feasible solutions and bounds for small-to-moderate size USApHMP instances, one important component appears to be missing in the literature: drawing insights on the optimal solution of a large instance from an (approximate) smaller representation of the same instance. Ernst and Krishnamoorthy (1996) briefly touch this issue while describing how the smaller instances are generated from the original 200-node AP data set. They generate smaller instances by dividing the coordinate system into n disjoint regions with an equal number of nodes ($200/n$ nodes each) and merge the nodes within the same region into a single centroid node. This procedure is often referred to as parcellation or segmentation in different fields. The resulting maps are indeed good approximations of the original map, and the optimal solutions of these smaller instances can be used as approximation to the original problem. However, these smaller instances fail to provide either an upper or a lower bound (LB) on the original problem. They do not provide an upper bound because their solution is not feasible to the original problem. On the other hand, they do not provide a lower bound because representing a set of nodes with a single centroid node results in overestimation of at least one inter-region transportation cost. The lack of interest in the literature to draw insight from smaller instances may be attributed to their failure to provide a lower/upper bound. To bridge this gap, we propose a new approach that efficiently makes use of the solutions of smaller instances to find high-quality feasible solutions, that is, upper bounds for the original USApHMP problem.

The main implication of the previous section can be summarized as follows: The nodes that are in spatial proximity to each other tend to behave similarly in their "choice" of allocation cluster, and this tendency becomes more pronounced as the distance between the two nodes decreases. In other words, on the microlevel, the effect of distance (coordinates) become the dominant factor in determining the coallocation decision of two nodes. One way to interpret and make use of this insight is to lump together several spatially neighboring nodes and assume they will be allocated to the same hub. Doing this for all the nodes in a given map will create spatially disjoint parcels containing lumped nodes for which the hub location and allocation decisions are yet to be made (see, for example, Figure 9). Notice that the intraparcel distances are typically smaller than 10% of the maximum distance which, based on Figure 8, makes it very likely that the nodes lumped into the same parcel will be allocated to the same cluster in the optimal solution.

This procedure is, in fact, very common in various fields and can be phrased as reducing the problem resolution. Solving this *low-resolution* problem (LRP) may not give us the optimal solution to the original problem, but by construction, it is expected to get us very close to it. One may argue that for very large

Figure 9. Dividing the AP200 Instance into 50 Parcels Minimizing the Intraparcel Distances



problem instances where MIP formulations become astronomically large to process, an approximate solution is the best one can hope for. The aforementioned insight derived from spatial separability combined with the property of USApHMP that all flow must be routed through hubs makes a properly designed parcellation approach less prone to approximation error. In other words, the nonexistence of direct parcel-to-parcel connections enables a significant reduction of the dimension of the problem by simply aggregating the parcels into a single node, without deviating much from the optimal solution of the original problem. The challenge here is to identify which nodes to aggregate, as this obviously depends on the optimal solution, which is not known beforehand. The proposed approach starts with handling this challenge at the *parcellation* step. Our ultimate goal with the proposed approach to be able to tackle 1,000-node USApHMP instances and find near-optimal solutions with negligible optimality gaps in a reasonable amount of time. We now proceed with describing the building blocks (modules) of the approach.

4.1. Parcellation

The first module, which is called *parcellation*, is used to identify an efficient partition of the nodes into a predetermined number of parcels. A parcellation is considered efficient if all parcels consist only of nodes that are in the same allocation cluster at the optimal solution. As evident from the description, measuring the exact efficiency of a parcellation requires the knowledge of the optimal solution to the original problem. Hence, such parcellations will have to be obtained using heuristics. As argued earlier, distance becomes the dominant force in determining the coallocation of two nodes on the microlevel. Therefore, if we require a large enough number of parcels (e.g., $r \geq 50$), the original map will be divided into small enough parcels that will justify making node allocations to parcels solely based on distance. This, in turn, relates the parcellation problem to the well-known p-median facility location problem.

Next, we provide the formal definition and the mathematical formulation for a variant of p-median problem we call r-Parcellation Problem (rPP). Given r, the number of parcels to divide a spatial network of nodes into, the rPP is the problem of finding the optimal parcel medoids as well as the allocation of the remaining nodes to parcels such that each node is allocated to exactly one parcel and the sum of node-to-medoid distances is minimized:

$$[r\text{PP}] \qquad \min \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}} d_{ij}y_{ij}, \tag{4a}$$

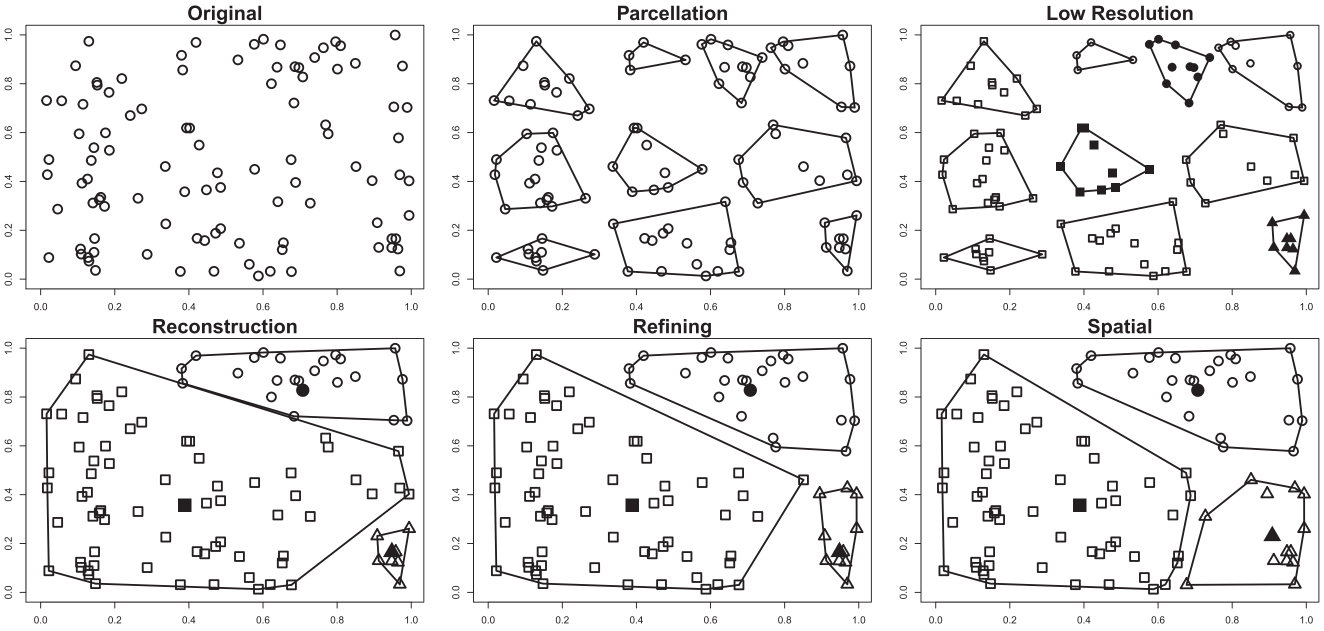
$$\text{s.t.} \sum_{j\in\mathcal{N}} x_j = r, \tag{4b}$$

$$\sum_{j\in\mathcal{N}} y_{ij} = 1 \qquad \forall i \in \mathcal{N}, \tag{4c}$$

$$y_{ij} \leq x_j \qquad \forall i,j \in \mathcal{N}, \tag{4d}$$

$$x_j, y_{ij} \in \{0,1\} \qquad \forall i,j \in \mathcal{N}, \tag{4e}$$

Figure 10. Step-by-Step Flow of SPATIAL Algorithm for a Random Instance with $n = 100$ Nodes and $p = 3$ Hubs



where the decision variables are defined as follows:

$$x_j = \begin{cases} 1 & \text{if a medoid is located at node } j \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij} = \begin{cases} 1 & \text{if node } i \text{ is allocated to parcel } j \\ 0 & \text{otherwise.} \end{cases}$$

Figure 10(b) shows an example of parcellation. The efficient parcellation of a map is not unique, that is, there are multiple different ways an optimum allocation cluster can be divided into parcels. Therefore, *r*PP is not the only method yielding parcels that will lead to a quality solution. One may use a different method to find an efficient parcellation. In fact, to speed up this step, we use a heuristic by Kaufman and Rousseeuw (2009) to solve the *r*PP rather than solving it optimally. Once an efficient parcellation is found, the next step is to find the hub-parcel locations and parcel-to-hub allocations, which is handled by the *low-resolution* module.

4.2. Low Resolution

Given an efficient parcellation $\mathcal{P} = \{P_1, P_2, \ldots, P_r\}$ of nodes N such that

$$P_i \cap P_j = \varnothing, \ \forall P_i \neq P_j \in \mathcal{P}$$
$$\bigcup_{i=1}^{r} P_i = \mathcal{N},$$

let $P(i)$ represent the parcel node i belongs to in $\mathcal{P}$. Assuming that the distances between two nodes in the same parcel are negligible, that is, $d_{uv} \leq \epsilon, \forall u, v \in \mathcal{N} : P(u) = P(v)$, we can approximate the original problem by treating each parcel as a single supernode and duplicating the decisions made for a parcel to all the nodes contained in it.

An approximation of the original USA*p*HMP can be done in several different ways depending on how the resulting output will be used. For example, by assuming the distances between nodes within the same parcel is equal to zero, one may construct an approximate problem to obtain a lower bound to the original problem. Alternatively, inter- and intraparcel distances may be set in a way to approximate the objective value of the original USA*p*HMP as accurately as possible. Because the main goal of this study is to produce high-quality feasible solutions to the original problem, we opt to use the latter approximate problem for which the distances and flows between the parcels are determined as follows:

$$\widetilde{d}_{ij} = \frac{\sum_{u\in P_i}\sum_{v\in P_j} w_{uv}d_{uv}}{\sum_{i\in P_i}\sum_{j\in P_j} w_{uv}} \qquad \forall i,j \in \{1,2,..,r\}, \tag{5a}$$

$$\widetilde{w}_{ij} = \sum_{i\in P_i}\sum_{j\in P_j} w_{uv} \qquad \forall i,j \in \{1,2,..,r\}. \tag{5b}$$

Note that the distance between parcels i and j is calculated as the weighted average of the distances between every node pair $\{(u,v) : u \in P_i, v \in P_j\}$, and the flow between the parcels is simply the sum of flows between all node pairs. Consequently, the unit cost of sending flow within a parcel is not zero unlike in the original USApHMP. We call this approximate problem the LRP and formally define it as follows:

$$[\text{LRP}] \qquad \min \sum_{i=1}^{r}\sum_{k=1}^{r}\sum_{m=1}^{r} \alpha\widetilde{d}_{km}\widetilde{x}_{ikm} + \sum_{i=1}^{r}\sum_{k=1}^{r}\left(\widetilde{O}_i + \widetilde{D}_i\right)\widetilde{d}_{ik}\widetilde{y}_{ik}, \tag{6a}$$

$$\text{s.t.} \sum_{k=1}^{r} \widetilde{y}_{kk} = p, \tag{6b}$$

$$\sum_{k=1}^{r} \widetilde{y}_{ik} = 1 \qquad \forall i \in \{1,2,\dots,r\}, \tag{6c}$$

$$\widetilde{y}_{ik} \le \widetilde{y}_{kk} \qquad \forall i,k \in \{1,2,\dots,r\}, \tag{6d}$$

$$\sum_{m=1}^{r} \widetilde{x}_{ikm} - \sum_{m=1}^{r} \widetilde{x}_{imk} = \widetilde{O}_i\widetilde{y}_{ik} - \sum_{j=1}^{r} \widetilde{w}_{ij}\widetilde{y}_{jk} \qquad \forall i,k \in \{1,2,\dots,r\}, \tag{6e}$$

$$\widetilde{x}_{ikm} \ge 0 \qquad \forall i,k,m \in \{1,2,\dots,r\}, \tag{6f}$$

$$\widetilde{y}_{ik} \in \{0,1\} \qquad \forall i,k \in \{1,2,\dots,r\}, \tag{6g}$$

where $\widetilde{O}_i$ ($\widetilde{D}_i$) is the total amount of outflow (inflow) from (to) parcel P_i, $\widetilde{w}_{ij}$ is the total amount of flow from parcel P_i to parcel P_j, and the decision variables are defined as follows:

$\widetilde{x}_{ikm}$ = Total flow emanating from parcel P_i that is routed between hub parcels P_k and P_m

$$\widetilde{y}_{ik} = \begin{cases} 1 & \text{if parcel } P_i \text{ is allocated to hub parcel } P_k \\ 0 & \text{otherwise.} \end{cases}$$

The term low-resolution is adopted from the image-processing field where a similar procedure, namely, assigning the same color to predetermined sets of neighboring pixels, is a very common practice to achieve computational efficiency at the expense of image quality. Figure 10(c) illustrates the optimal solution to LRP for a given parcellation. Hub parcels are marked with filled symbols and nonhub parcels are marked with hallow symbols of the hubs they are allocated to.

LRP is an essential module in the proposed approach as it approximates the original problem with fewer supernodes and thus allows us to make use of our ability to solve smaller USApHMP instances in solving larger problem instances. Once LRP is solved for a given parcellation, what remains is to obtain a feasible solution to the original problem. Next, we describe the third module, which constitutes the link between the low-resolution problem and the original problem.

4.3. Reconstruction

Notice that the solution for the low-resolution problem would give us a partitioning of the parcels (and, consequently, of the nodes) into p clusters. However, instead of p hub nodes, we have p hub parcels, which typically add up to more than p nodes. In order to reconstruct a feasible solution to the original problem, one needs to identify the exact locations of p hub nodes. Recall that finding the locations of the hub nodes for a given p-partition of nodes $\mathcal{N}$ is also an existing problem in the literature introduced by Sung and Jin (2001). Wagner (2007) calls the problem Cluster Hub Location Problem (CHLP) and provides the most efficient formulation for the problem. We next provide the formulation provided by Wagner (2007), which constitutes our third module *reconstruction*.

Let $\mathscr{C} = \{C_1, C_2, \dots C_p\}$ be a partition of $\mathscr{N}$ into allocation clusters such that

$$C_i \cap C_j = \emptyset, \ \forall C_i \neq C_j \in \mathscr{C}$$
$$\bigcup_{i=1}^{p} C_i = \mathscr{N}, \tag{7}$$

$$\text{[CHLP]} \quad \min \sum_{c=1}^{p} \sum_{k \in C_c} \sum_{m \in \mathscr{N} \setminus C_c} \alpha d_{km} F_{km} x_{km} + \sum_{c=1}^{p} \sum_{k \in C_c} \sum_{i \in C_c} (O_i + D_i) d_{ik} y_k,$$

$$\text{s.t.} \sum_{k \in C_c} y_k = 1 \qquad \forall c = \{1, \dots, p\}, \tag{8a}$$

$$\sum_{m \in C_c} x_{km} = y_k \qquad \forall k \in \mathscr{N} \text{ and } \forall c = \{1, \dots, p\} \text{ with } k \notin C_c, \tag{8b}$$

$$\sum_{k \in C_c} x_{km} = y_m \qquad \forall m \in \mathscr{N} \text{ and } \forall c = \{1, \dots, p\} \text{ with } m \notin C_c, \tag{8c}$$

$$0 \le x_{km} \le 1 \qquad \forall k, m \in \mathscr{N}, \tag{8d}$$

$$y_k \in \{0, 1\} \qquad \forall k \in \mathscr{N}, \tag{8e}$$

where F_{km} is the total amount of flow from the cluster of node k to the cluster of node m and the decision variables are defined as follows:

$$x_{km} = \begin{cases} 1 & \text{if both nodes } k \text{ and } m \text{ are hubs} \\ 0 & \text{otherwise} \end{cases}$$
$$y_k = \begin{cases} 1 & \text{if node } k \text{ is a hub} \\ 0 & \text{otherwise.} \end{cases}$$

The solution of CHLP gives us the optimal hub locations for the given partitions. However, considering the possible information loss due to shrinking the original map into a low-resolution one, the input partition may not be the optimal one. To avoid suboptimal solutions resulting from resolution errors, we devise a final module to recursively refine the feasible solutions.

4.4. Refining

One way to improve a given feasible solution is to enumerate all its neighborhood solutions, which is what heuristics such as tabu search do. Another approach would be to make use of a property of optimum solutions to USApHMP: The solution of the CHLP with the optimum allocation clusters will give the optimum hubs and, similarly, the solution of another related problem known as the p-Hub Allocation Problem (pHAP) with the optimum hubs will give the optimum allocation clusters. Hence, by recursively solving CHLP and pHAP, one may reach a local optimum. Considering the fact that the feasible solution generated by the first three modules is not any random solution, but it is found by taking the problem data into account, our hope is that the local optima hit at the end of *refining* will be very close to the global optimum. Next, we provide the mathematical formulation of pHAP for a given subset $\mathscr{H} \subseteq \mathscr{N}$ of p hubs:

$$[p\text{HAP}] \quad \min \sum_{i \in \mathscr{N}} \sum_{k \in \mathscr{H}} \sum_{m \in \mathscr{H}} \alpha d_{km} x_{ikm} + \sum_{i \in \mathscr{N}} \sum_{k \in \mathscr{H}} d_{ik} (O_i + D_i) y_{ik}, \tag{9a}$$

$$\text{s.t.} \ y_{kk} = 1 \qquad \forall k \in \mathscr{H}, \tag{9b}$$

$$\sum_{k \in \mathscr{H}} y_{ik} = 1 \qquad \forall i \in \mathscr{N}, \tag{9c}$$

$$y_{ik} = 0 \qquad \forall i \in \mathscr{H} \text{ and } k \in \mathscr{H} \setminus i, \tag{9d}$$

$$\sum_{m \in \mathscr{H}} x_{ikm} - \sum_{m \in \mathscr{H}} x_{imk} = O_i y_{ik} - \sum_{j \in \mathscr{N}} W_{ij} y_{jk} \qquad \forall i \in \mathscr{N} \text{ and } \forall k \in \mathscr{H}, \tag{9e}$$

$$0 \le x_{ikm} \le 1 \qquad \forall i \in \mathscr{N} \text{ and } \forall k, m \in \mathscr{H}, \tag{9f}$$

$$y_{ik} \in \{0, 1\} \qquad \forall i \in \mathscr{N} \text{ and } \forall k \in \mathscr{H}. \tag{9g}$$

Notice that pHAP is the same formulation as USApHMP, only reduced in size. The constraint sets (1b) and (1d) become redundant in pHAP. For the sake of consistency, (9b) and (9d) are added to the formulation of

Table 2. Descriptions of Functions

Function	Output	Description
GET.FEASIBLE$(\tilde{y})$	y	Generates a feasible solution from the optimal solution of USApHRP
GET.CLUSTERS(y)	$\mathscr{C}$	Extracts the allocation clusters from a given feasible solution
GET.HUBS (y)	$\mathscr{H}$	Extracts the hub nodes from a given feasible solution
SOLVE.RPP $(\mathscr{I}, r)$	$\mathscr{P}$	Divides the map into r spatial parcels and allocates each node into one parcel
SOLVE.LRP$(\mathscr{I}, \mathscr{P})$	$\tilde{y}$	Finds the optimal solution to low-resolution problem for a given parcellation
SOLVE.CHLP$(\mathscr{I}, \mathscr{C})$	y, z	Finds the optimal solution to cluster hub problem for given clusters
SOLVE.PHAP$(\mathscr{I}, \mathscr{H})$	y, z	Finds the optimal solution to node allocation problem for given hubs

USApHAP; however, the model can be further reduced by dropping the variables whose values are fixed by constraints (9b) and (9d).

This concludes the description of modules in the proposed approach. Hereinafter, we refer to this approach as SPATIAL and outline its flow in Algorithm 1. Table 2 provides brief descriptions of the functions used in Algorithm 1 and Figure 10 illustrates a sample flow of the approach at each module. In Sections 5.2 and 5.3, we first illustrate the efficiency of the approach on known USApHMP instances from the AP data set and then use it to find high-quality solutions for the very large BCN instances.

Before moving to the computational experiments, we clarify the reason spatial separability imposition does not explicitly appear in the Algorithm 1. In virtually all real benchmark instances, the resulting solution from the SPATIAL algorithm did end up satisfying spatial separability. Thus, we did not find it useful to add a separate step for this. This is because spatial separability also—almost always—holds for the optimal solutions of *pHAP* which constitutes the *refining* step of the SPATIAL algorithm. A nice illustration can be seen in Figure 10. The intermediate solution found after the *reconstruction* step violates spatial separability. However, this is immediately corrected in the *refining* step, and the final solution ends up satisfying the spatial separability. In the unlikely case of the final solution not satisfying the spatial separability, one needs to go back to the *parcellation* step to make sure that the nodes taking part in the violating intersection are properly parceled to avoid this violation. Because this had never happened in our computational experiments with larger instances, we did not devise a rigorous parcellation update procedure.

Algorithm 1 (SPATIAL Algorithm)

function SPATIAL$(\mathscr{I}, r = 50)$
  $\mathscr{P} \leftarrow$ SOLVE.RPP$(\mathscr{I}, r)$
  $\tilde{y} \leftarrow$ SOLVE.LRP$(\mathscr{I}, \mathscr{P})$
  $y \leftarrow$ GET.FEASIBLE$(\tilde{y})$
  $\mathscr{C} \leftarrow$ GET.CLUSTERS(y)
  $z_C \leftarrow \infty$
  $z_A \leftarrow 0$
  while $z_C \neq z_A$ **do**
    $[y, z_C] \leftarrow$ SOLVE.CHLP$(\mathscr{I}, \mathscr{C})$
    $\mathscr{H} \leftarrow$ GET.HUBS(y)
    $[y, z_A] \leftarrow$ SOLVE.PHAP$(\mathscr{I}, \mathscr{H})$
    $\mathscr{C} \leftarrow$ GET.CLUSTERS(y)
  end while
  return y, z_A
end function

5. Computational Experiments

The main purpose of the computational experiments is to illustrate the noncoincidental existence of the *spatial separability* and the remarkable computational gain resulting from its incorporation in a novel approach to solve USApHMP instances. Following the brief description of the experimental framework, we first show the validity of the spatial separability property in two-dimensional space. We then test the performance of the proposed approach on the problem instances from well-known data sets. Finally, we conduct computational experiments to illustrate the effectiveness of the hub location problems in modeling the brain hub organisation, and to provide further evidence on the scalability of the proposed approach.

The testing is done on benchmark instances (AP data set) provided in Ernst and Krishnamoorthy (1996), BCN instances introduced in Section 4, and randomly generated instances. For a given set of nodes $\mathcal{N}$, a *map* is defined by the coordinates of the nodes in two or three-dimensional space and the flows W_{ij} between each pair of nodes. Random maps are generated by randomly assigning a value to each coordinate $\forall i \in \mathcal{N}$ and w_{ij} $\forall i, j \in \mathcal{N}$ from a uniform distribution between zero and one. For a given map, a USApHMP *instance* is defined by setting the values for two parameters, namely, p (the number of hubs to open) and α (interhub discount factor).

We used the R programming environment for the implementation of the proposed algorithm and Gurobi 6.0.2, with default parameters and a single thread, to solve the MILP instances. All experiments were carried out on an Intel(R) Core(TM) i7-4510U PC (2.60GHz, 8GB memory) running Windows 10. We limited the processing time of all experiments to 3,600 seconds.

5.1. Spatial Separability

In Figure 2, we showed that the optimal solutions of a few instances from the AP data set possess what we call spatial separability property. In this section, we try to generalize this observation. To this end, we first checked 100 instances generated using the AP data set (with different parameters) and validated that the said property holds for all of them. Next, we wanted to validate that this property is not specific to AP instances. To do so, we resorted to random maps and instances. The generation process of random maps was described at the beginning of this section. We generated 100 random maps for each $n \in \{25, 50\}$. Then, for each map, we generated 12 instances with different values of $p \in \{2, 3, 4, 5\}$ and $\alpha \in \{0.1, 0.5, 0.9\}$. Figure 11 shows an example of the spatial separability of 12 instances generated from a random map with different parameters. As clearly seen from the figure, the convex hulls of the allocation clusters in optimal solution are disjoint in all 12 instances. Similar plots are generated for all random maps. The observations mostly complied with Figure 11.

Figure 11. (Color online) An Example of Spatial Separability of Optimal Solutions on a Random Map

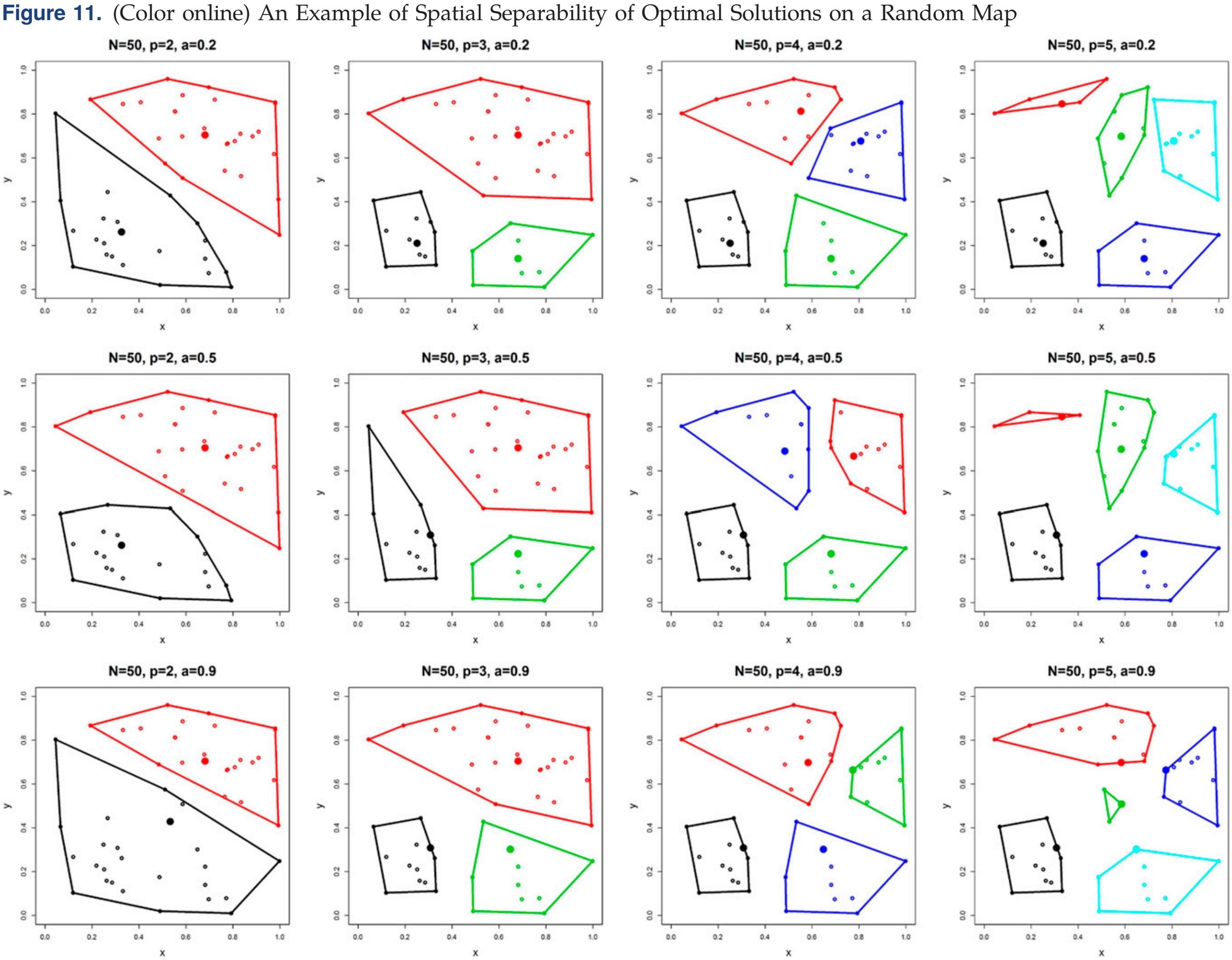


Table 3. Fraction of Separable Instances for Different Parameter Values

	$n = 25$			$n = 50$		
p	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
2	100/100	100/100	99/100	100/100	98/100	90/100
3	100/100	100/100	100/100	100/100	100/100	100/100
4	100/100	100/100	100/100	100/100	100/100	100/100
5	100/100	100/100	100/100	100/100	100/100	100/100

However, we also observed several instances for which the spatial separability was violated. Table 3 shows the fraction of instances for which the spatial separability holds.

Several insights are derived from Table 3. First, spatial separability holds for 99.5% (2,387/2,400) of the randomly generated instances. Second, $p = 2$ is the only p-value for which spatial separability is violated. Moreover, although stronger evidence is needed for such generalizations, we observe that the number of nonseparable instances increases with n and α. Unfortunately, such mass computational experiments are not possible for larger n values because of exponentially increasing processing time to solve USApHMP instances optimally. Nevertheless, the existence of violating instances suffices to refute the claim that spatial separability in 2D space is a property of USApHMP.

Although there are few instances for which spatial separability is violated, the overwhelming majority of the instances are spatially separable. This brings forward a question: How much would we lose from optimality if we looked only for the spatially separable solutions? More specifically, is there an upper bound on the optimality gap of feasible solutions restricted to hold spatial separability property? Because, for the majority of the violating instances in our experiments, we could establish spatial separability by adding only a few intersection elimination constraints (3a)–(3f), we were able to check their optimality gaps manually. Not surprisingly, the optimality gaps were very small, ranging between 0.1% and 0.5%, which indicates that even if we imposed spatial separability a priori, we would not lose much from optimality. Whether such small optimality gaps are consistent throughout different parameter sets and maps is subject to a more detailed computational experimentation, which we refrain from in this study.

5.2. Performance on AP Data Set

We implemented the proposed approach on instances selected from the AP data set with 100 nodes and $p \in \{2, 3, \ldots, 10\}$ hubs. Table 4 shows the optimal solutions (OPT), the LP relaxation (LP) using the MILP formulation given in Ernst and Krishnamoorthy (1996), and the results from the proposed approach with a different number of parcels ($r \in \{25, 50\}$). Because the optimal solutions to the instances listed above could be found by Gurobi, both LP gap (lower bound) and SPATIAL gap (upper bound) are calculated with respect to the optimal solution.

It is evident that the proposed approach generates high-quality solutions in a reasonable amount of time. In terms of CPU time, Spatial takes significantly less time compared with both OPT and LP, which hints about the scalability of the proposed approach. As expected, Spatial($r = 25$) requires much less time than Spatial($r = 50$). It is, however, interesting to observe that the gaps for Spatial($r = 25$) are very similar to those for Spatial($r = 50$), which implies that the resolution ratio of 1:4 is almost as efficient as the ratio of 1:2. Finding

Table 4. Computational Results on Selected Problem Instances from AP Data Set with 100 Nodes

	OPT		LP		SPATIAL ($r = 25$)		SPATIAL ($r = 50$)	
p	CPU (sec)	z	CPU (sec)	Gap (%)	CPU (sec)	Gap (%)	CPU (sec)	Gap (%)
2	158.26	180,223.80	40.54	0.46	0.70	0.00	3.12	0.00
3	406.41	160,847.00	93.30	0.76	0.78	0.00	3.34	0.00
4	647.69	145,896.58	77.00	0.96	0.75	0.00	6.67	0.00
5	2,636.83	136,929.44	132.97	1.77	1.72	0.00	19.38	0.58
6	7,250.79	129,214.65	126.18	1.88	1.79	0.63	24.01	0.63
7	3,210.05	122,577.13	130.08	2.16	1.42	1.04	20.79	1.04
8	2,260.60	116,410.90	116.00	1.76	1.67	1.33	19.05	2.01
9	2,335.53	111,108.79	102.20	1.37	1.17	2.24	29.65	2.37
10	1,620.84	106,469.57	85.84	1.32	1.71	2.83	25.46	1.38

the resolution ratio that provides the optimum trade-off between the computation time and the solution quality would be an interesting research question, though it is out of the scope of this study.

Next, we conduct experiments to compare the performance of Spatial ($r = 50$) with Gurobi in terms of optimality gap and CPU time. Table 5 shows the results from Gurobi and Spatial ($r = 50$) for different parameter values. For each (n, p, α) parameter set considered, the following experiment is conducted: First, Spatial ($r = 50$) is run and the CPU time and objective value z_S are recorded. Then, Gurobi is run for the same amount of time, and the LB and the upper bound z_G (i.e., the best feasible solution found within the same amount of time) are recorded. Because the optimal solutions are not known for these instances, the gaps are calculated as $\delta_S = 100\ (z_S - LB)/z_S$ and $\delta_G = 100\ (z_G - LB)/z_G$, respectively, and the best of the two gaps is written in bold. Typically for instances with $n = 200$, Gurobi was not able to find a feasible solution within the time required for running Spatial. For those cases, we let Gurobi run until it found a second feasible solution. This was done to avoid any unfairness in comparison, because it is typical for commercial solvers like Gurobi to find a drastically improved second feasible solution within a few seconds after the first one. Nevertheless, it becomes clear in Table 5 that Spatial significantly outperforms Gurobi across all parameter sets, which points to the global efficiency of the proposed approach. A final remark on the results from Table 5 is that the performance of Spatial deteriorates with increasing α both in terms of CPU time and in terms of optimality gap. This observation concurs with the results presented in Table 3. Namely, because the spatial separability is more likely to be violated for larger values of α, it is expected that the performance of the proposed approach, which is based on spatial separability, deteriorates with increasing α.

One might argue that a comparison with a general-purpose solver is not sufficient, especially considering the recent developments in the literature to develop efficient heuristics to tackle large problem instances. Therefore, we compare the performance of Spatial algorithm ($r = 25$) with state-of-the-art heuristics. We consider three heuristics: (1) the VNS of Ilić et al. (2010), (2) Euclidean Projection Method of Meier and Clausen (2017), and (3) Bender decomposition of Ghaffarinasab and Kara (2019). Table 6 provides a comparison in terms of the optimality gap and computational time. It is evident that the best heuristic among the four is VNS. However, the results of Spatial, especially on instances with smaller p, are quite comparable with those of VNS and significantly better than the other methods. To be fair, the other two methods are geared toward proving optimality, and it is well-known that proving optimality typically requires substantially more time than finding a good solution. Similarly, the Spatial algorithm incorporates optimization in three of its steps, which slows it down.

When comparing algorithms, it is important to identify the potential benefits of adopting the strong elements of one algorithm in another. We believe the highly modular structure of the Spatial algorithm is one of its strengths. When we checked the computational bottlenecks of the algorithm, we found out that optimally solving the LRP appeared to be the main bottleneck for the above AP instances. One way to alleviate this

Table 5. Performance Comparison of Spatial and Gurobi on AP Instances

		$\alpha = 0.1$				$\alpha = 0.5$				$\alpha = 0.9$			
n	p	CPU	LB	δ_S (%)	δ_G (%)	CPU	LB	δ_S (%)	δ_G (%)	CPU	LB	δ_S (%)	δ_G (%)
100	2	3.4	68,471.5	**0.06**	0.16	7.3	77,766.5	**1.16**	13.45	20.4	85,382.7	**1.98**	3.17
	3	3.5	59,000.6	0.35	**0.32**	16.9	70,239.7	**1.91**	**1.91**	58.8	79,090.9	**4.22**	7.54
	4	4.4	51,956.3	**0.42**	1.28	17.9	65,045.8	**1.73**	2.79	96.5	75,478	**3.95**	15.48
	5	9.4	47,373.3	1.27	**0.32**	45.9	61,476.5	**3.26**	11.96	182.6	73,054.7	**4.98**	8
	6	7.6	43,543.4	0.78	**0.42**	44.3	58,765.7	**3.71**	14.78	752.3	71,334.1	**5.38**	22.81
	7	11.8	40,640.8	**1.75**	4.05	36.2	56,448.4	**3.85**	14.7	260.0	69,969.4	**5.11**	24.82
	8	8.0	38,041	**0.88**	1.01	50.0	54,511.1	**3.42**	17.14	584.0	68,792.7	**5.55**	23.25
	9	11.5	35,616.3	**1.4**	2.74	59.5	52,907.6	**3.34**	17.4	1,312.4	67,836.9	**4.42**	18.59
	10	11.4	33,484.4	**1.34**	3.96	90.4	51,480	**4.06**	12.54	1,618.7	67,017.4	**4.33**	24.31
200	2	4.3	69,188.4	**0.05**	**0.05**	9.9	78,526.6	**1.31**	12.36	24.7	85,741.3	**1.85**	12.05
	3	7.4	59,618.1	0.25	**0.2**	30.6	70,918.7	**2.16**	2.38	67.4	79,584.9	**4.76**	9.16
	4	4.6	52,605.4	0.49	**0.11**	26.3	65,594.9	**2.03**	3.42	116.6	75,867.1	**5.06**	17.65
	5	15.1	48,546.3	**0.67**	1.47	42.0	62,208.8	**3.64**	16.12	380.9	73,476.3	**5.97**	21.86
	6	4.9	44,880.8	**0.85**	5.32	44.8	59,584.9	**4.46**	31.16	1,076.2	71,760.6	**5.94**	12.87
	7	16.3	41,793.2	**0.82**	7.86	63.7	57,301.3	**4.46**	20.09	1,096.9	70,418.0	**6.57**	21.34
	8	22.0	39,103.9	**1.32**	4.47	113.6	55,359.9	**3.92**	31.37	1,004.4	69,312.2	**5.94**	24.03
	9	17.6	36,800.9	**1.62**	18	141.2	53,750.5	**5.06**	19.62	1,359.2	68,391.2	**6.21**	20.91
	10	15.4	34,679.0	**2.06**	24.83	154.5	52,401.5	**5.18**	23.7	1,460.4	71,141.8	**4.99**	22.06

Table 6. Performance Comparison of SPATIAL and State-of-the-Art Methods on AP Instances

		Gap (%)				CPU (sec)			
n	p	SPATIAL	VNS	Euclidean	Benders	SPATIAL	VNS	Euclidean	Benders
100	2	0.00%	0.00%	< 1%	*NA*	< 1	< 1	6.50	*NA*
	3	0.00%	0.00%	< 1%	*NA*	< 1	< 1	8.57	*NA*
	4	0.00%	0.00%	< 1%	*NA*	1.5	< 1	10.34	*NA*
	5	0.58%	0.00%	< 1%	0.00%	2.5	< 1	12.63	313.80
	10	1.69%	0.00%	*NA*	0.00%	2.8	< 1	*NA*	109.18
200	2	0.00%	0.00%	< 1%	*NA*	6.90	< 1	94.49	*NA*
	3	0.08%	0.00%	< 1%	*NA*	3.55	< 1	141.51	*NA*
	4	0.03%	0.00%	< 1%	*NA*	8.66	< 1	131.86	*NA*
	5	0.27%	0.00%	< 1%	2.00%	6.18	5.16	629.78	~20,000
	10	1.92%	0.00%	*NA*	1.50%	9.95	5.60	*NA*	~7,500

Note. NA, not available.

bottleneck is to use VNS in the *low-resolution* step, which succeeds in finding near-optimal solutions (for moderate-size instances) very quickly. Furthermore, resolving this computational bottleneck would allow us to increase the resolution of the LRP, which, in turn, is expected to improve the accuracy of the solution to the original problem. On the other hand, as mentioned before, the spatial separability property could be used to reduce the number of neighborhood solutions explored at each step of VNS. These are all interesting potential research directions; however, our main focus on this study is not computational efficiency, but the analysis of the newly introduced BCN application, which we tackle in the next subsection.

5.3. Analysis of Brain Hub Organisation

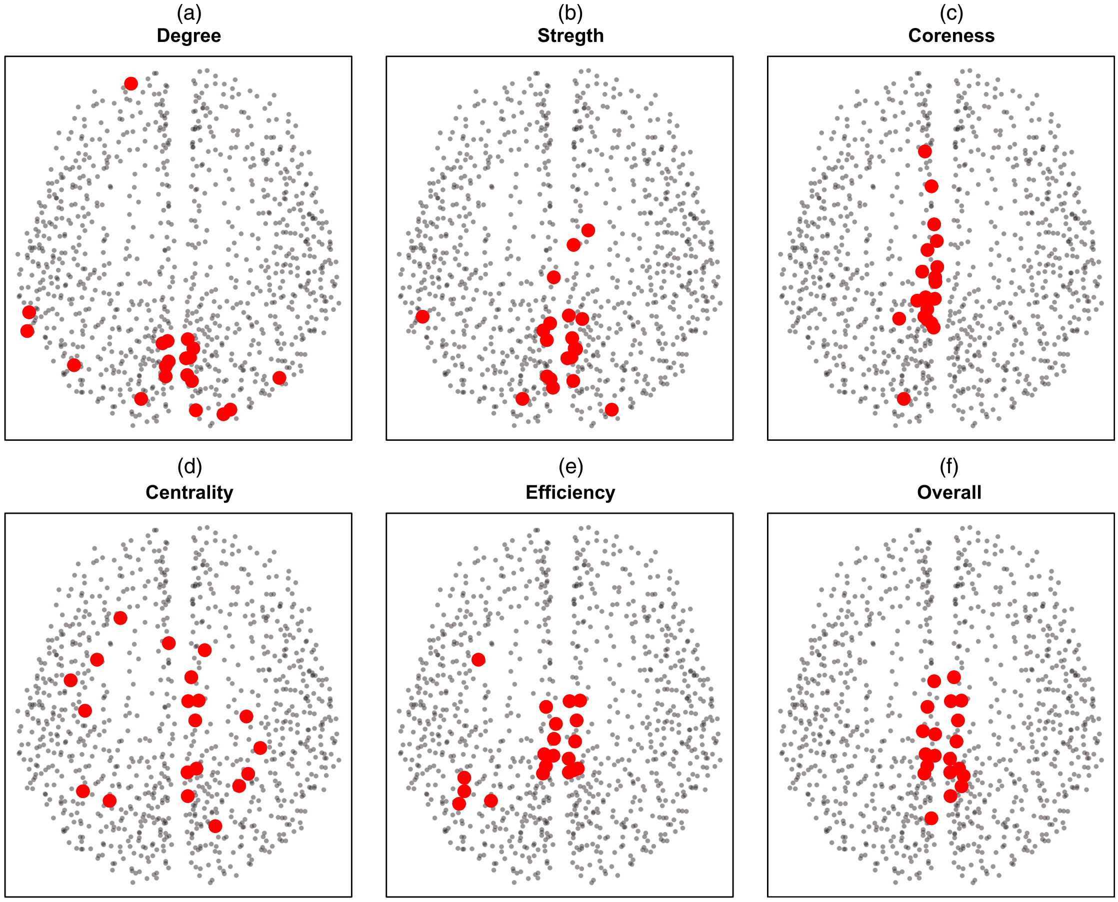
Devising a mathematical model that captures the underlying mechanisms of hubs is an essential step toward understanding the true nature of the brain hub organisation. This is what we attempt to accomplish in this section by testing the validity of USApHMP in modeling the brain hub organisation and the effectiveness of Spatial algorithm in providing near-optimal solutions to large USApHMP instances. We conduct experiments mainly to compare the results obtained by the state-of-the-art techniques in the medical literature (Hagmann et al. 2008) and those obtained by the Spatial algorithm. Throughout this section, we use the real BCN data provided in Hagmann et al. (2008), which constitute the largest hub location problem instances (with $n = 998$ nodes) in the literature to the best of our knowledge.

There are several issues to address before presenting the experiment results. Firstly, the spatial separability property, which the Spatial approach is based upon, assumes Euclidean distances, whereas the distances between two brain ROIs are recorded as fiber length, which does not necessarily satisfy triangular inequality. Moreover, when we compared the recorded fiber lengths between two ROIs against the Euclidean distances calculated based on the ROI coordinates, we noticed for some ROI pairs, the fiber length was smaller than the Euclidean distance, which indicates that there are measurement errors in the recorded data as Euclidean distance is the smallest possible distance between the two points in space. By excluding these pairs, we calculated the correlation coefficient between the Euclidean distances and the fiber lengths as 0.904. This high correlation implies that the fiber connections between two ROIs roughly follow a straight line, which justifies using Euclidean distances as a proxy for the fiber lengths.

We begin by briefly introducing the methodology adopted by Hagmann et al. (2008) to determine hub ROIs in the brain network. They calculate the network measures, such as degree, strength, coreness, efficiency and centrality, for each node and then rank the nodes based on their overall score. They suggest that this overall ranking could be interpreted as the hubness of a node. In the paper, they report results on the low-resolution brain map with 66 Brodmann areas. Following the same the steps, we first regenerated similar results for the high-resolution brain map with 998 ROIs. Figure 12 shows the 20 highest-ranking ROIs with respect to each network measure mentioned above and the overall hubness rank.

Several observations can be made from Figure 12. It is argued by Hagmann et al. (2008) that a hub ROI is expected to rank among the highest for all the attributes represented by the aforementioned proxy network measures. It becomes evident from Figure 12, however, that the rankings with respect to these network measures do not necessarily concur. Although there seems to be some overlap (along the area between two hemispheres), discrepancies outweigh the similarities. For example, the *centrality* measure seems to suggest a more distributed hub organisation across the brain map, whereas *coreness* suggests a very concentrated hub region along the area between two hemispheres. On the other hand, *degree* and *strength* provide very similar

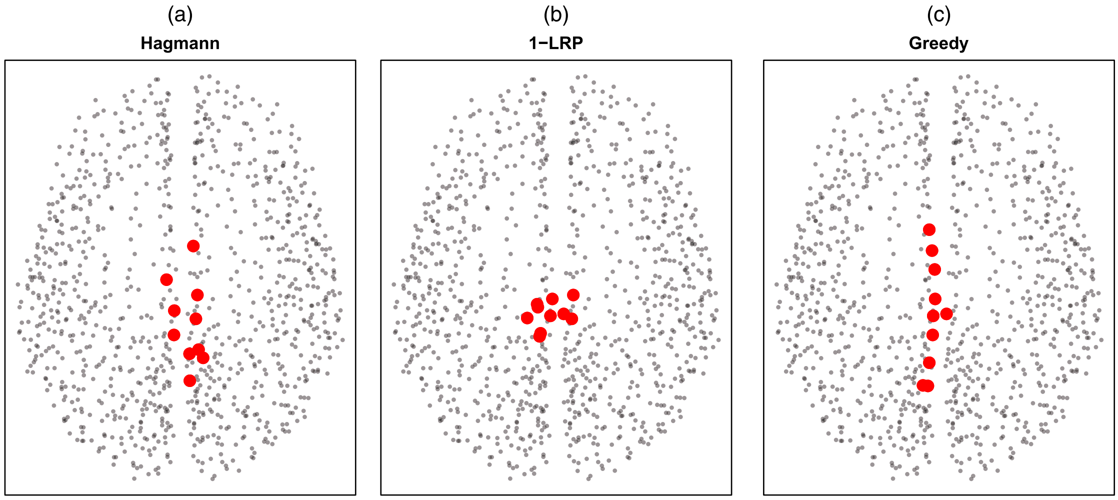
Figure 12. (Color online) Hub ROIs Found by Each Network Measure



results in terms of the distribution of hub ROIs across the brain map. This brings up the question: Aren't we biasing the *overall* hubness rank toward the distribution provided by *degree* and *strength* as we essentially account for the "same" distribution twice? When combining multiple conflicting measures into a single overall measure, taking a simple average is usually considered a naive approach because it ignores possibly different weight factors associated with each measure. Although using different weights for each proxy measure may alleviate this issue to some extent, there are many other issues with using proxy measures that are hard to address. Are these five proxy measures all that one would expect a hub ROI to rank good at? An important measure that seems to be missing, for instance, is the measure of interaction between hubs. Because hubs, by definition, are interacting entities, one would expect the hubness of a node in a given network to depend on the set of other hub nodes. This, however, is not a trivial measure to account for by a simple proxy measure because of its inherent nonlinearity. USA*p*HMP model is known to account for this nonlinearity which is, in large part, the reason why it is very challenging to solve very big problem instances.

Next, we provide a comparison between the main results of Hagmann et al. (2008) and the solutions found by the USA*p*HMP model. It is worth mentioning that unless otherwise stated, throughout the paper the results pertaining to the USA*p*HMP models are obtained by using $\alpha = 0.5$ because they appear to be in more agreement with the results of Hagmann et al. (2008). In the original paper, they identify 16 Brodmann areas (*posterior cingulate cortex, the precuneus, the cuneus, the paracentral lobule, the isthmus of the cingulate, the banks of the superior temporal sulcus, and the inferior and superior parietal cortex*, in both hemispheres) as hub regions, which corresponds to 236 ROIs in total. This implies that almost one-quarter of the whole brain cortex assumes a hub role. Moreover, this also assumes that the ROIs within the same Brodmann area will all either be hub or nonhub, which does not help much in isolating the specific hub regions. Finally, using USA*p*HMP to find $p = 236$ optimal hub ROIs in a network of $n = 998$ nodes would be computationally infeasible. Hence, we resort to some simplifications. First, we select only the top 10 hub ROIs identified by Hagmann et al. (2008) (shown in Figure 13(a)) because with $p = 10$ the problem becomes feasible to work on. Notice that hub ROIs are spatially adjacent to each other. This implies that there is a single contiguous hub region which is made up

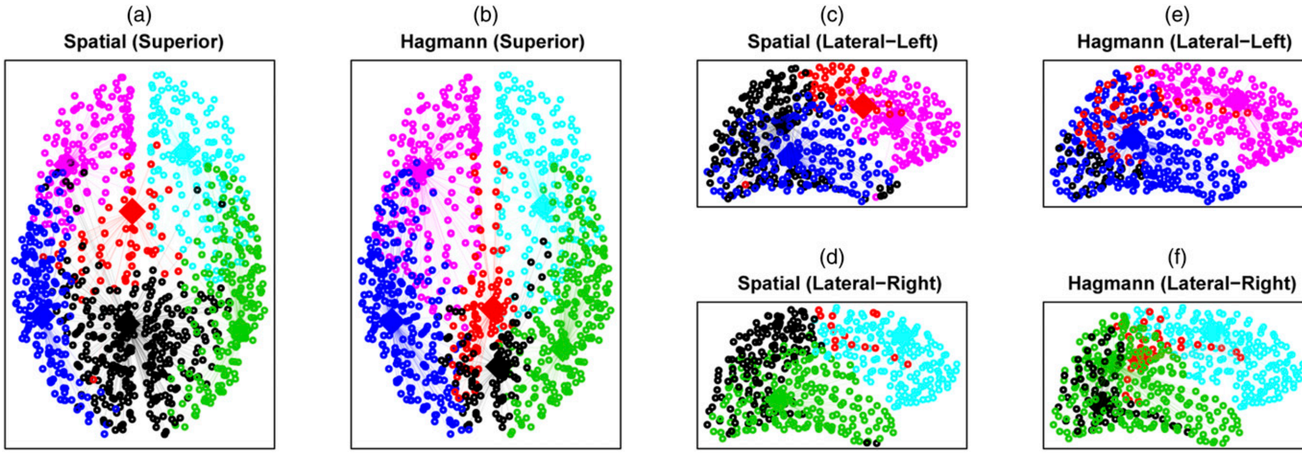
Figure 13. (Color online) Comparison of Hub ROIs Found by Hagmann et al. (2008) and USApHMP



of 10 ROIs. However, it is known that USApHMP model typically distributes the hubs all over the given map. Hence, it is very unlikely that USApHMP with $p = 10$ hubs will reproduce similar results in terms of hub locations. On the other hand, the spatial contiguity of hub region signals a connection with one of the problems introduced in Section 3: the LRP. Specifically, if the original brain map with 998 ROIs is reduced to a lower-resolution map with approximately 100 contiguous regions, then the 1-hub LRP is expected to produce similar results to the one reported in Hagmann et al. (2008). Figure 13(b) shows the optimal solution to 1-hub LRP. The 10 ROIs identified as hubs by LRP belong to the following brain areas: *paracentral lobule, posterior cingulate cortex, and precuneus* in both the right and left hemispheres. As expected, the locations of the hub ROIs are roughly in agreement with those found by Hagmann et al. (2008). Slight differences may be attributed to the approximate nature of both approaches; in principle, they can be calibrated in a way that produces more similar results. For example, a different parcellation technique for the LRP could be used so that one of the resulting parcels contains exactly the same ROIs as the hub ROIs of Hagmann et al. (2008). In fact, just to confirm, we manually placed Hagmann's hub ROIs in a single parcel (P_1) and divided the rest of the ROIs into 99 parcels ($P_2, P_3, \ldots P_{100}$). As argued, the resulting LRP finds P_1 as the hub region. When we calculated the objective value of the solution with hubs identified by Hagmann et al. (2008) ($z_a = 29883.75$ for Figure 13(a)), it turned out to be better than that of the original 1-LRP ($z_b = 30906.03$ for Figure 13(b)), which indicates that the objective function of USApHMP is suitable to assess the *quality* of a given subset of hubs and also implies that there may be an even better subset with a better objective value. To this end, we implemented another variant of USApHMP as follows: First, we find a single hub ROI that is selected as the optimal hub for USA1HMP (i.e., single hub problem). Then, we iteratively add new ROIs to the subset of hubs until we have $p = 10$ hub ROIs. In doing so, we restrict the candidate ROI set only to those that are spatially contiguous to the current hubs and greedily select the candidate addition of which to the current hub set improves the objective function value the most. Figure 13(c) shows this solution that resembles the Hagmann's more both in terms of the locations of hub ROIs (i.e., stretched along the vertical axis) and in terms of the objective function value ($z_c = 28481.40$).

These preliminary results are important as they imply that a variant (LRP or Greedy-contiguous) of USApHMP model can be used to reproduce the results similar to those reported in the medical literature, which are solely based on network topological measures. The significance of this implication lies in what more the USApHMP model can provide. A mathematical model can tell us not only which ROIs are likely to behave as hubs but also what makes them more favorable candidates for hubness compared with the others. To elaborate, there are two main factors that determine which nodes are selected to be hubs in a given network: spatial centrality and flow-based centrality. The former simply indicates how centrally a node is located in the network, whereas the latter pertains to the amount of flow that goes through the close proximity of a node. The two are not always in perfect agreement and often need to be reconciled. In fact, of the six network measures used in Hagmann et al. (2008), some such as *efficiency* pertain to spatial/topological centrality and some such as *strength* to flow-based centrality. We observed that there is a significant discrepancy between the node ranks calculated according to different network measures. This makes the assumption that *average rank*

Figure 14. Comparison of the Six Brain Modules Identified by SPATIAL and Hagmann et al. (2008)



of a node is a good proxy for its hubness an oversimplification. On the other hand, USA*p*HMP provides an alternative framework under which the two factors can be reconciled in a way to minimize the "transportation" cost of the overall system.

Moreover, hub ROIs identified by network measures will likely always correspond to a monocentric hub structure because typically two neighboring nodes have similar scores on these network measures. However, the monocentricity of brain hub structures has not yet been established. In fact, in the same study, Hagmann et al. (2008) consider another model where ROIs are clustered into different modules and hub ROIs are identified for each module. This model is more similar to USA*p*HMP in the general sense, yet it, too, is just a descriptive model. Namely, it does not provide information on what is the advantage of having a certain set of hub ROIs as opposed to a different set. Conversely, by providing the pathways all pairwise flows are routed through for a given set of hubs, USA*p*HMP can provide deeper insight on the physical/economical advantage of one set over another. Hence, it is worthwhile to illustrate the parallelism between the result of the module-based approach of Hagmann et al. (2008) and the results of the Spatial($p = 6, \alpha = 0.5, r = 66$) algorithm. Figure 14, (a) and (b) show the six allocation clusters identified by the Spatial algorithm and by Hagmann et al. (2008), respectively, on the superior view of the brain. To have a better understanding of how the clusters look in 3D, we also included the lateral views of left and right hemispheres for both Spatial and Hagmann in Figure 14, (c)–(f). ROIs in each cluster (module) are marked with the same color in both figures. Moreover, a straight line is drawn between each ROI and the hub it is allocated to illustrate the pathways.

We observe a clear resemblance in the distribution of ROIs to clusters and modules. First of all, both figures reflect the highly symmetrical nature of the brain. Specifically, both figures show that there are two clusters (marked with black and red colors) that span across the left and right hemispheres, whereas the remaining four clusters consist of ROIs only from a single hemisphere (blue and pink from left hemisphere, green and turquoise from right hemisphere, respectively). A symmetry is also apparent among these four clusters; specifically, blue-green and pink-turquoise clusters are approximately pairwise symmetrical. The only visible difference between Hagmann's modules and Spatial's allocation clusters is the span of interhemispheric clusters (black and red). In particular, Spatial's interhemispheric clusters appear to be more widely spread than Hagmann's, which is mostly constricted along the axis between the two hemispheres. To quantify the difference between the two clusterings, we solved a CHLP with Hagmann's modules and compared its objective value ($z_H = 29936.71$) against the objective value from the Spatial ($z_S = 29248.25$). A 2.35% relative gap between the two objective values can be interpreted as a quantitative indicator of the similarity between the two solutions.

Whether the ideal segmentation of the human brain indeed consists of six modules as suggested by Hagmann et al. (2008) is an interesting research question. Unfortunately, it requires a more elaborate computational analysis ideally with a bigger data set collected from multiple individuals to answer this question. Based on the results presented so far, however, it seems like the USA*p*HMP model in general and the solution algorithm Spatial in particular could serve as a useful tool in such analyses. To this end, we finally provide the objective values and the CPU times of Spatial ($r = 66$) for several BCN instances with different p and α values in Table 7.

Table 7. Objective Values and CPU Times of SPATIAL Algorithm for Different BCN Instances

	$\alpha = 0.1$		$\alpha = 0.5$		$\alpha = 0.9$	
p	z_S	CPU (sec)	z_S	CPU (sec)	z_S	CPU (sec)
2	38,220.1	186.3	38,658.2	486.2	39,303.8	384.3
3	33,002.6	261.0	33,933.6	183.4	34,767.9	196.0
4	28,554.5	325.2	29,556.1	345.7	30,460.6	275.6
5	25,822.7	236.8	27,143.7	177.9	28,061.6	278.2
6	24,426.2	232.7	25,724.9	267.8	26,804.2	209.3
7	23,326.3	389.5	24,568.1	294.0	25,645.0	433.0
8	22,561.1	413.0	23,760.8	658.4	24,664.8	1,769.9
9	21,443.7	497.4	22,728.3	1,271.9	24,154.4	4,198.9
10	20,386.4	519.9	22,769.8	2,608.7	24,088.3	3,933.4

These results may not provide any important clinical insight, nevertheless, they serve as a benchmark for future studies that aim to tackle very large instances such as BCN. Similar to the observation from Table 5, the performance of the SPATIAL algorithm deteriorates with increasing α in terms of CPU time. Another observation is that increasing the number of hubs from p to $p+1$ improves the objective value significantly for the small values of p; however, past a certain threshold, which seems to be around $p_t = 8$, the improvement is quite small. Potentially, this threshold point may be used to figure out the optimum number of modules the human brain should be segmented into. The justification for such an approach would be that there is no "economic" advantage of further segmentation beyond p_t where the objective value seems to converge to a certain value. Obviously, to make such an assertion, one needs to first establish the validity of the HLP model to effectively represent the underlying economic model of the human brain. The validity of the USA*p*HMP model to represent the underlying economic nature of brain hub organisation needs to be tested with larger data sets and compared against the models already proposed in the literature (Bullmore and Sporns 2012, Avena-Koenigsberger et al. 2018), which is beyond the scope of this study.

6. Discussion

It is important to emphasize that we do not assert that the USA*p*HMP or its variants explain the underlying dynamics of how information flows between different regions in the human brain. Our goal is to introduce optimization models as a potential tool to help investigate the complex hub organization of the human brain. The fact that a single, integrated tool is able to reproduce the insights/results derived by manually combining the often conflicting outcomes of multiple descriptive techniques may, in itself, be appealing to the neuroscience community. We believe the vast and constantly growing body of research in hub location could further contribute to the field of neuroscience.

There is a plethora of future research directions in this line of research. For example, spatial separability seems to prevail in the vast majority of the hub-and-spoke networks. This raises the question of whether the few cases that violate this feature can be singled out by some mild assumptions so that the objective function of the USA*p*HMP satisfies the necessary conditions of Barnes et al. (1992). The two-dimensional space in which we have sought spatial separability consists only of the information on the location of the nodes and completely overlooks the flow information. Is it possible to introduce some extra dimensions that encapsulate the flow of information such that incorporation of these dimensions eliminates the few nonseparable cases and restores the spatial separability (in higher dimensions) as a property of the optimal solutions of USA*p*HMP? If so, what would be the physical meaning of these dimensions and what is the best way to express these extra dimensions mathematically? The simplest idea that comes to mind is to calculate for each node i the fraction of total flow that is between the nodes falling to the left (bottom) of node i and the nodes falling to the right (top) of node i. When we rechecked the spatial separability with these additional dimensions, all the nonseparable instances we had among our random instances hold the separability property. This indicates that there could be a way to impose spatial separability as a rule (rather than just a consistent pattern) by introducing proper flow-based dimensions.

The similarity of the results obtained by 1-hub LRP and those of Hagmann et al. (2008) signals that there may be a need for a mathematical model of a related problem. The problem could be phrased as follows: Given p the number of distinct hub regions and a limit r on the total number of ROIs which can be hubs, what is the best way to distribute r ROIs among p spatially contiguous regions. Such a model would combine the strengths of both approaches, that is, the tendency of USA*p*HMP to distribute hubs over the whole map and the ability of network measures to result in similar hubness for spatially adjacent nodes. Similar studies where imposing connectivity constraints on existing models can be found in the literature (King et al. 2012, Carvajal et al. 2013).

Finally, although some of the results obtained in this paper coincided with the findings in the literature, it also became evident that a more specialized model of the hub location problems may be required to best represent the dynamics of brain networks. For example, the sparsity in the actual fiber connections should be taken into account. Moreover, the fiber densities could be considered as capacities of the arcs between ROIs, and the actual flows could be represented by the functional connectivity data sets, which were not included in this study. Relating the architecture of functional and structural connectivity networks is one of the most challenging topics within the human connectome project for which such a specialized model could provide a fresh perspective.

7. Conclusions

In this study, we introduced brain connectivity networks as a new application of hub location problems. Motivated by an additional requirement—spatial contiguity—of this new application area, we investigated the optimal hub-and-cluster networks and discovered an important property that holds in almost all instances: a special case of spatial contiguity, which we coined as *spatial separability*. We then used an insight derived from this property to devise a solution methodology to solve large USApHMP instances. The effectiveness of the proposed methodology was tested on existing benchmark instances as well as a new set of instances originating from brain connectivity networks. Computational experiments showed that the results established in the medical literature can be reproduced to a certain degree using the proposed approach.

Acknowledgments

The authors thank Olaf Sporns, professor of psychological and brain sciences at Indiana University, Bloomington, for sharing the brain connectivity network data set used in their original study (Hagmann et al. 2008), which significantly enriched the authors' computational experiments. The authors also express their gratitude to Sahin Hanalioglu, MD-PhD of the faculty of medicine at Hacettepe University; Professor James F. Campbell at the University of Missouri, St. Louis; and the anonymous reviewers for their valuable feedback and comments that greatly improved the original manuscript.

References

Alumur S, Kara BY (2008) Network hub location problems: The state of the art. *Eur. J. Oper. Res.* 190(1):1–21.

Alumur S, Kara BY, Karasan OE (2009) The design of incomplete single allocation hub networks. *Transportation Res B: Methodological* 43(10):936–951.

Alumur SA, Nickel S, Saldanha-da Gama F (2012) Hub location under uncertainty. *Transportation Res. Part B: Methodological* 46(4):529–543.

Applegate DL (2006) *The Traveling Salesman Problem: A Computational Study* (Princeton University Press, Princeton, NJ).

Arslan S, Ktena SI, Makropoulos A, Robinson EC, Rueckert D, Parisot S (2018) Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage* 170:5–30.

Avena-Koenigsberger A, Misic B, Sporns O (2018) Communication dynamics in complex brain networks. *Nature Rev. Neuroscience* 19(1):17–33.

Aykin T (1994) Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *Eur. J. Oper. Res.* 79(3):501–523.

Barnes ER, Hoffman AJ, Rothblum UG (1992) Optimal partitions having disjoint convex and conic hulls. *Math. Programming* 54:69–86.

Bassett DS, Sporns O (2017) Network neuroscience. *Nature Neuroscience* 20(3):353–364.

Behrmann M, Plaut DC (2013) Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends Cognitive Sci.* 17(5):210–219.

Boland N, Krishnamoorthy M, Ernst AT, Ebery J (2004) Preprocessing and cutting for multiple allocation hub location problems. *Eur. J. Oper. Res.* 155(3):638–653.

Brodmann K (1909) *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* (Barth, Leipzig, Germany).

Bullmore E, Sporns O (2012) The economy of brain network organization. *Nature Rev. Neuroscience* 13(5):336–349.

Calık H, Alumur SA, Kara BY, Karasan OE (2009) A tabu-search based heuristic for the hub covering problem over incomplete hub networks. *Comput. Oper. Res.* 36(12):3088–3096.

Campbell JF (1990) Freight consolidation and routing with transportation economies of scale. *Transportation Res. Part B: Methodological* 24(5):345–361.

Campbell JF (1994) Integer programming formulations of discrete hub location problems. *Eur. J. Oper. Res.* 72:387–405.

Campbell JF, O'Kelly ME (2012) Twenty-five years of hub location research. *Transportation Sci.* 46:153–169.

Campbell JF, Ernst AT, Krishnamoorthy M (2005a) Hub arc location problems: Part II—formulations and optimal algorithms. *Management Sci.* 51(10):1556–1571.

Campbell JF, Ernst AT, Krishnamoorthy M (2005b) Hub arc location problems: Part I—introduction and results. *Management Sci.* 51(10):1540–1555.

Carvajal R, Constantino M, Goycoolea M, Vielma JP, Weintraub A (2013) Imposing connectivity constraints in forest planning models. *Oper. Res.* 61(4):824–836.

Chakravarty AK, Orlin JB, Rothblum UG (1985) Consecutive optimizers for a partitioning problem with applications to optimal inventory groupings for joint replenishment. *Oper. Res.* 33(4):820–834.

Commons W (2008) Lateral surface of left cerebral hemisphere, viewed from the side. Accessed February 28, 2020, https://commons.wikimedia.org/wiki/File:Gray726-Brodman.svg.

Contreras I (2015) Hub location problems. Laporte G, Nickel S, Saldanha da Gama F, eds. *Location Science* (Springer, Cham), 311–344.

Contreras I, Cordeau J-F, Laporte G (2011a) Benders decomposition for large-scale uncapacitated hub location. *Oper. Res.* 59(6):1477–1490.

Contreras I, Cordeau J-F, Laporte G (2011b) Stochastic uncapacitated hub location. *Eur. J. Oper. Res.* 212(3):518–528.

Contreras I, Díaz JA, Fernandez E (2011c) Branch and price for large-scale capacitated hub location problems with single assignment. *INFORMS J. Comput.* 23(1):41–55.

Contreras I, Tanash M, Vidyarthi N (2017) Exact and heuristic approaches for the cycle hub location problem. *Ann. Oper. Res.* 258:655–677.

Correia I, Nickel S, Saldanha-da Gama F (2010) The capacitated single-allocation hub location problem revisited: A note on a classical formulation. *Eur. J. Oper. Res.* 207(1):92–96.

Crossley NA, Mechelli A, Vértes PE, Winton-Brown TT, Patel AX, Ginestet CE, McGuire P, Bullmore ET (2013) Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci. USA* 110(28):11583–11588.

de Camargo RS, de Miranda G Jr, Luna HP (2009a) Benders decomposition for hub location problems with economies of scale. *Transportation Sci.* 43(1):86–97.

de Camargo RS, de Miranda G Jr, Ferreira RPM, Luna HP (2009b) Multiple allocation hub-and-spoke network design under hub congestion. *Comput. Oper. Res.* 36(12):3097–3106.

Ebery J (2001) Solving large single allocation *p*-hub problems with two or three hubs. *Eur. J. Oper. Res.* 128(2):447–458.

Ebery J, Krishnamoorthy M, Ernst A, Boland N (2000) The capacitated multiple allocation hub location problem: Formulations and algorithms. *Eur. J. Oper. Res.* 120(3):614–631.

Elhedhli S, Hu FX (2005) Hub-and-spoke network design with congestion. *Comput. Oper. Res.* 32(6):1615–1632.

Elhedhli S, Wu H (2010) A Lagrangian heuristic for hub-and-spoke system design with capacity selection and congestion. *INFORMS J. Comput.* 22(2):282–296.

Ernst AT, Krishnamoorthy M (1996) Efficient algorithms for the uncapacitated single allocation *p*-hub median problem. *Location Sci.* 4(3):139–154.

Farahani RZ, Hekmatfar M, Arabani AB, Nikbakhsh E (2013) Hub location problems: A review of models, classification, solution techniques, and applications. *Comput. Indust. Engrg.* 64(4):1096–1109.

Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME (2006) Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl. Acad. Sci. USA* 103(26):10046–10051.

Friederici AD, Gierhan SM (2013) The language network. *Curr. Opin. Neurobiol.* 23(2):250–254.

Ghaffarinasab N, Kara BY (2019) Benders decomposition algorithms for two variants of the single allocation hub location problem. *Networks Spatial Econom.* 19(1):83–108.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, et al. (2016) A multi-modal parcellation of human cerebral cortex. *Nature* 536(7615):171–178.

Guan J, Lin G, Feng H-B (2018) A learning-based probabilistic tabu search for the uncapacitated single allocation hub location problem. *Comput. Oper. Res.* 98:1–12.

Hagmann P, Cammoun L, Gigandet X, Meuli R, Christopher J Honey, Van J Wedeen, Sporns O (2008) Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6(7):e159.

Hagmann P, Thiran J-P, Jonasson L, Vandergheynst P, Clarke S, Maeder P, Meuli R (2003) DTI mapping of human brain connectivity: Statistical fibre tracking and virtual dissection. *Neuroimage* 19(3):545–554.

Honey CJ, Kötter R, Breakspear M, Sporns O (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. USA* 104(24):10240–10245.

Ilić A, Urošević D, Brimberg J, Mladenović N (2010) A general variable neighborhood search for solving the uncapacitated single allocation *p*-hub median problem. *Eur. J. Oper. Res.* 206(2):289–300.

Kaufman L, Rousseeuw PJ (2009) *Finding Groups in Data: An Introduction to Cluster Analysis*, vol. 344 (John Wiley & Sons, Hoboken, NJ).

King DM, Jacobson SH, Sewell EC, Tam Cho WK (2012) Geo-graphs: An efficient model for enforcing contiguity and hole constraints in planar graph partitioning. *Oper. Res.* 60(5):1213–1228.

Klincewicz JG (1992) Avoiding local optima in the *p*-hub location problem using tabu search and grasp. *Ann. Oper. Res.* 40:283–302.

Klincewicz JG (1998) Hub location in backbone/tributary network design: A review. *Location Sci.* 6(1–4):307–335.

Kwon H, Choi Y-H, Lee J-M (2019) A Physarum centrality measure of the human brain network. *Sci. Rep.* 9:5907.

Labbé M, Yaman H, Gourdin E (2005) A branch and cut algorithm for hub location problems with single assignment. *Math. Programming* 102:371–405.

Le Bihan D, Mangin J-F, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H (2001) Diffusion tensor imaging: Concepts and applications. *J. Magnetic Resonance Imaging* 13(4):534–546.

Marianov V, Serra D (2003) Location models for airline hubs behaving as M/D/*c* queues. *Comput. Oper. Res.* 30(7):983–1003.

Martin JC, Román C (2004) Analyzing competition for hub location in intercontinental aviation markets. *Transportation Res. Part E: Logist. Transportation Rev.* 40(2):135–150.

Mayer G, Wagner B (2002) Hublocator: An exact solution method for the multiple allocation hub location problem. *Comput. Oper. Res.* 29(6):715–739.

Meier JF, Clausen U (2017) Solving single allocation hub location problems on Euclidean data. *Transportation Sci.* 52(5):1141–1155.

Newman MEJ (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* 103(23):8577–8582.

O'Kelly ME (1986) The location of interacting hub facilities. *Transportation Sci.* 20(2):92–106.

O'Kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. *Eur. J. Oper. Res.* 32(3):393–404.

O'Kelly ME (1992) A clustering approach to the planar hub location problem. *Ann. Oper. Res.* 40:339–353.

Oldham S, Fornito A (2018) The development of brain network hubs. *Developmental Cognitive Neuroscience* 36:100607.

Peker M, Kara BY, Campbell JF, Alumur SA (2016) Spatial analysis of single allocation hub location problems. *Networks Spatial Econom.* 16(4):1075–1101.

Pessoa L (2012) Beyond brain regions: Network perspective of cognition–emotion interactions. *Behavior Brain Sci.* 35(3):158–159.

Pirkul H, Schilling DA (1998) An efficient procedure for designing single allocation hub and spoke systems. *Management Sci.* 44(12-part-2): S235–S242.

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

Skorin-Kapov D, Skorin-Kapov J, O'Kelly M (1996) Tight linear programming relaxations of uncapacitated p-hub median problems. *Eur. J. Oper. Res.* 94(3):582–593.

Sung CS, Jin HW (2001) Dual-based approach for a hub network design problem under non-restrictive policy. *Eur. J. Oper. Res.* 132(1):88–105.

Topcuoglu H, Corut F, Ermis M, Yilmaz G (2005) Solving the uncapacitated hub location problem using genetic algorithms. *Comput. Oper. Res.* 32(4):967–984.

van den Heuvel MP, Mandl RCW, Stam CJ, Kahn RS, Hulshoff Pol HE (2010) Aberrant frontal and temporal complex network structure in schizophrenia: A graph theoretical analysis. *J. Neuroscience* 30(47):15915–15926.

van den Heuvel MP, Sporns O (2013) Network hubs in the human brain. *Trends Cognitive Sci.* 17(12):683–696.

Vincent JL, Patel GH, Fox MD, Snyder AZ, Baker JT, Van Essen DC, Zempel JM, Snyder LH, Corbetta M, Raichle ME (2007) Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* 447(7140):83–86.

Wagner B (2007) An exact solution procedure for a cluster hub location problem. *Eur. J. Oper. Res.* 178(2):391–401.

Zalesky A, Fornito A, Seal ML, Cocchi L, Westin C-F, Bullmore ET, Egan GF, Pantelis C (2011) Disrupted axonal fiber connectivity in schizophrenia. *Biol. Psychiatry* 69(1):80–89.