University of Hertfordshire UH 25 Research Archive

Citation for published version:
Soheil Davari, Kemal Kilic, and Siamak Naderi, 'A heuristic approach to solve the preventive health care problem with budget and congestion constraints', *Applied Mathematics and Computation*, Vol. 276, pp. 442-453, March 2016.

DOI:
https://doi.org/10.1016/j.amc.2015.11.073

Document Version:
This is the Accepted Manuscript version.
The version in the University of Hertfordshire Research Archive may differ from the final published version.

Copyright and Reuse:
© 2015 Elsevier Inc.
This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/.

Enquiries
If you believe this document infringes copyright, please contact the Research & Scholarly Communications Team at rsc@herts.ac.uk

A heuristic approach to solve the preventive health care problem with budget and congestion constraints

Soheil Davari *[1], Kemal Kilic[2], and Siamak Naderi[3]

[1]School of Mathematics, Cardiff University, Cardiff, United Kingdom
[2,3]Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey

Abstract

Preventive health care is of utmost importance to governments since they can make massive savings on health care expenditure and promote the well-being of the society. Preventive care includes many services such as cancer screenings, vaccinations, hepatitis screenings, and smoking cessation programs. Despite the benefits of these services, their uptake is not satisfactory in many countries in the world. This can be attributed to financial barriers, social issues., and other factors. One of the most important barriers for preventive care is accessibility to proper services, which is a function of various qualitative and quantitative factors such as the distance to travel, waiting time, vicinity of facilities to other attractive facilities (such as shopping malls), and even the cleanliness of the facilities. Statistics show that even a small improvement in people's participation can save massive amounts of money for any government and improve the well-being of the people in a society.

This paper addresses the problem of designing a preventive health care network considering impatient clients, and budget constraints. The objective is to maximize the accessibility of services to people. We model the problem as a mixed-integer programming problem with budget constraints, and congestion considerations. An efficient variable neighborhood search procedure is proposed and computational experiments are performed on a large set of instances.

Keywords. Preventive health care, Facility location, Equity, Variable neighborhood search, Network design.

1 Introduction

Efficient management of the health care supply chain is becoming a more active research field every passing day. Scarcity of resources, increasing customer expectations in terms of quantity (due to the aging population) and quality, increasing costs (both due to the increase in demand and investment costs of new health-related technologies and drug discovery) are among the factors which make this topic a significant one for health care service providers, insurance companies, as well as governments who are constantly striving to become as efficient as possible in delivery of their services. Various methods from mathematics and operations research fields have been used to facilitate management of health care systems, such as Markov chains, mathematical programming, and simulation.

*School of Mathematics, Cardiff University
Senghennydd Road, Cardiff, Wales, UK, CF24 4AG
Email: davaris@cardiff.ac.uk
Tel: +44 (0) 2920 870 850

Many costly, and disabling problems such as cancer, cardiovascular diseases, diabetes, and chronic respiratory problems are linked by some risk factors which are preventable to a rather high extent and/or better managed by regular checkups. A preventive health care program is the set of actions taken to avoid or to delay the onset of diseases. The old idiom "an ounce of prevention is worth a pound of cure" summarizes the benefits of preventive health care in a perfect way. Generally speaking, there are three categories in preventive health care. Primary interventions are those that reduce the risk of disease for healthy individuals (e.g. immunization programs, diet schemes, autism screenings for children, etc.). Secondary interventions are those that are designed for early detection of diseases for the individuals who are in the risk groups (e.g., screening for breast cancer for females over a certain age, cholesterol control, screenings for osteoporosis, colonoscopies, consultancy services provided for pregnant women, audiometric tests). Neither the individuals who receive primary prevention services nor those who receive secondary services have any obvious sign of the disease. On the other hand, tertiary interventions are designed for individuals who have been already diagnosed clinically for a disease and the goal of the preventive health care service is reducing the complications that might be caused by it (e.g., for individuals that have diabetes, regular retinal checks are performed). Preventive health care programs can bring about substantial reduction in the overall health care spending of the society (Maciosek [1] reports savings of $3.7 billion for USA in 2006). Efficient management of preventive health care programs and increasing the public participation would definitely benefit the society and patients simultaneously. For instance, studies in USA show that for every HIV infection prevented, an estimated $355,000 is saved in the cost of providing lifetime HIV treatment [2]. Moreover, a 5 percent reduction in the prevalence of hypertension would save $25 billion in 5 years [3].

Preventive health care is inherently different from programs for acute ailments. In contrast to sick people who need urgent medical attention, people who seek preventive services have more flexibility as to when and where to receive preventive health care services. Even though the benefits of preventive services are clear both in terms of cost and health, most of the people are reluctant about their own health status and often prefer not to participate. Therefore, the achievement of the desired participation level continues to be a challenge to many preventive health care programs. The maximal participation levels lead to economies of scale in the operational costs of preventive health care facilities. An increase in participation levels in such preventive programs reduces the overall burden of health for the society and increases the expected benefits from the health care delivery.

Preventive health care Facility Network Design Problem (PHFNDP), briefly speaking, which deals with where to locate the facilities and determination of their capacities, is among the most significant strategic level decisions in any preventive health care program. The goal is to establish a set of facilities among a set of potential locations, so that the participation level is maximized. Empirical research in health economics literature deals with the concept of attractiveness of health care facilities and suggests that there are various factors that influence attractiveness. For example, Bjorn and Godager [4] as well as Gravelle et al. [5] demonstrate that the attractiveness of health care facilities is not only influenced by the proximity but also by other qualitative factors such as quality, availability of other facilities in the neighborhood (e.g. shopping malls, restaurants, etc.), amenities near the facility. etc. On the other hand, Muller et al. [6] determines that in urban areas, distance influences the decision on which kind of medical services (e.g. a medical doctor or a hospital) the patients use, whereas in rural areas of developing countries, distance is the decisive factor whether or not to use medical services at all. Other research such as Varkevisser et al. [7] and Haynes et al. [8] also reveal evidence that distance plays a pivotal role in the attractiveness of facilities. Note that, these empirical evidences are not from studies that focus particularly for preventive health care services but for general health care services. Normally, in the literature of the PHNFND problem, distance is used as the only factor affecting the attractiveness of a facility (e.g., Zhang et al. [9]; Zhang et al. [10]; Zhang et al. [11]; Gu et al. [12]). The relation between distance and attractiveness can be modeled as linear or non-linear based on some demographic issues, traffic conditions, etc. Figure 1 depicts a network with four facilities established on a plane and an exponential attractiveness function in which darker areas are areas with good accessibility to facilities and the accessibility wears out as clients move away from the facilities. As it is clear from this figure, the attractiveness of facilities can decrease dramatically by an increase of the distance (Distance can be replaced by travel time which is a more realistic measure for our problem like many other health care related problem, both in rural areas owing to possible difficulties of access and in urban areas because of the traffic conditions. For more information, interested readers can refer to various sources such as Phibs and Luft [13], Schuurmann et al. [14]).

Insert Figure 1 around here

Chronic diseases such as cancer, diabetes, and heart diseases in US account for 75% of the country's health expenditure as reported by the Centers for Disease Control and Prevention while they are largely preventable. An optimal provision of preventive health care services can save money up to $590,000/QALY for governments as the health care providers [15]. As discussed in Daskin [16], the implications of poor location of health care facilities can be well beyond cost and customer service considerations, increasing the prevalence of diseases, and mortalities. That said, health care providers can benefit massively from a higher participation rate of individuals in addition to citizens. The main goal of the problem in this paper is to maximize the participation level of people as a measure of the distance between facilities and population centers in order to efficiently manage the health care spending, and promote the health status of a society. We contribute to the literature by proposing an efficient heuristic procedure to solve the PHNDP with budget and capacity constraints.

The outline of this paper is as follows: It proceeds with a literature review of relevant publications in section 2. The mathematical model of the paper is presented in section 3. In section 4, our proposed solution procedure is elaborated. Numerical experiments and some analysis appear in section 5, and finally, conclusions and some future research avenues are provided in section 6.

2 Background and Literature Review

The publication of first papers regarding to the design of health care networks date back to early 60s when [17] presented the problem of location on a graph with applications in telecommunication, police stations, and hospitals. Later, many papers have been published in the context of health care network design such as locating organ transplant centers (Belien et al. [18]), locating health care facilities assuming moving populations (Ndiaye et al. [19]), locating ambulances (Shariat et al. [20]), and real-world case studies in Brazil (Galvao et al. [21]), Burkina Faso (Cocking et al. [22]), and Malaysia Shariff et al. ([23]).

The problem of designing preventive health care network with congestion considerations is not new to the literature. The first publication in this area was Verter et al. [24] where the problem of locating preventive facilities was presented and case studies in Georgia, USA and Montreal, Canada were reported. Recently, [9] presented the problem of preventive health care network design on a graph with optimal choice allocation and an objective of maximising participation level. They presented four different heuristic methods for their problem. Later, Zhang et al. [10], addressed a similar problem and modeled it as a bi-level nonlinear optimization model. In order to solve their problem, they developed a lower level problem and an upper level problem and proposed gradient projection method and an efficient tabu search procedure. Gu et al. [12], developed an accessibility measure for PHFLP and presented an efficient interchange algorithm to solve it. The impact of client choice behavior on the network and the participation level of people was considered in Zhang et al. [11]. Their decision variables are the location of facilities and also the number of facilities in each location. They presented a genetic algorithm to solve the problem. In another recent paper, Gunes et al. [25], the physician allocation to health care centers is modeled and an illustrative case study in Turkey has been presented.

Unlike sick people who usually take an appointment (unless it is a medical emergency situation) from a health care service provider before a visit to the facility, usually individuals participate in the prevention health care programs whenever they have time and they are in the vicinity of the facility. That is to say, the prevention programs execute in a non-appointment setting and the participants "walk-in" a center. Therefore, in line with some of the existing literature (Gunes et al. [26]; Zhang et al. [9]; Zhang et al. [10]), in this paper the preventive service provider facility is modeled as a queuing system due to the stochastic demand and limited capacity.

There is little discussion regarding the health benefits of prevention. However, there is an ongoing debate both in the scientific literature and public regard in terms of the cost-effectiveness of the prevention programs. Substantial evidence exits that it is more beneficial to prevent rather than cure (Holland et al. [27]; Woolf et al. [28]; Maciosek et al. [1]). However, some researchers argue that it might be more beneficial to treat a few sick patients rather than trying to prevent the whole population. Cohen et al. [29] conducted a systematic review of 599 articles from cost-effectiveness literature and concluded that opportunities for efficient investment for prevention and treatment are roughly equal. Therefore, the policy makers should be more selective while allocating limited resources to prevention programs and only initiate those that would be more cost-efficient. That is to say, as opposed to the existing literature that focus on preventive health care management, budget should be considered as a concern in the model. Hence, we have included budget in our model. Note that, as it is the case in most of the public service investment decision making problems (e.g., Aktas et al. [30]; Angulo et al. [31]; Keranshahi et al. [32]), we also incorporated budget as a constraint to the model rather than an objective function and conducted sensitivity analysis regarding to different budgetary levels.

From above and to the best of our knowledge, the case of preventive health care network design with equity considerations and budget constraints has not been addressed in the literature. Hence, this paper presents a model for such a problem and proposes an efficient Skewed Variable Neighborhood Search algorithm which is able to reach solutions with errors not worse than 1.78%.

3 Mathematical model

Consider a region (say a city) with demands for preventive health care spread over the network nodes and transportation links between nodes. The decision maker is interested in maximizing the participation level of people throughout the region by establishing facilities in the network. There are candidate locations to establish facilities and the proximity of clients to facilities is the key factor in making facilities more attractive. Besides, there are budget constraints and also congestion considerations in the problem. The sets, parameters, and decision variables of the problem are as follows (please note that for the sake of integrity, we use index i for the demand nodes and index j for the potential nodes to establish facilities.)

Sets

N Set of nodes (population centers, index i is used for this set)

$F \subseteq N$ Set of potential nodes to establish facilities (index j is used for this set)

$H = \{1, 2, ..., H_{max}\}$ Set of available number of servers

Parameters

m Number of population centers (cardinality of the set N)

n Number of potential nodes to establish facilities (cardinality of the set F)

λ The number of clients requiring service per unit of time (following a Poisson distribution)

p_i Population living at node $i \in N$

t_{ij} The shortest path from node $i \in N$ to node $j \in F$

$\overline{\lambda_k}$ Maximum participation rate such that the system does not explode with k servers

η The attractiveness coefficient

a_{ij} The attractiveness of facility $j \in F$ to the client at node $i \in N$ ($a_{ij} = e^{-\eta t_{ij}}$)

B The budget available to establish facilities

c^v The unit cost of adding a server to a facility

c_j^f The fixed establishment cost for a facility $j \in F$

Decision Variables

x_{ij} A binary variable taking a value 1 if node $i \in N$ gets served by facility $j \in F$ and 0 otherwise

h_{jk} Binary variable taking a value of 1 if node $j \in F$ has $k \in H$ or more servers and 0 otherwise

We assume that each facility $j \in F$ has k servers, each providing an exponentially distributed service at a rate of μ service per unit of time. Moreover, $\overline{\nabla\lambda_k} = \overline{\lambda_k} - \overline{\lambda_{k-1}}(k \in H)$ and $\overline{\lambda_0}$=0. Now, the mathematical model of the problem as an integer programming model is as follows:

$$\max \lambda \sum_{i \in N} p_i \sum_{j \in F} a_{ij} x_{ij} \tag{1}$$

$$\sum_{j \in F} x_{ij} = 1 \qquad \forall i \in N \tag{2}$$

$$x_{ij} \leq h_{j1} \qquad \forall i \in N,\ j \in F \tag{3}$$

$$h_{j,k+1} \leq h_{jk} \qquad j \in F, k \in H \backslash \{H_{max}\} \tag{4}$$

$$t_{ij} x_{ij} \leq t_{ip} + M(1 - h_{p1}) \qquad \forall i \in N, j \in F, p \in F \tag{5}$$

$$\sum_{j \in F} c_j^f h_{j1} + c^v \sum_{j \in F} \sum_{k \in H} h_{jk} \leq B \tag{6}$$

$$\lambda \sum_{i \in N} p_i a_{ij} x_{ij} \leq \sum_{k \in H} \overline{\nabla\lambda_k} h_{jk} \qquad \forall j \in F \tag{7}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i \in N, j \in F \tag{8}$$

$$h_{jk} \in \{0, 1\} \qquad \forall j \in F, k \in H \tag{9}$$

Objective function (1) maximizes the total participation level of clients. Constraint (2) stipulates that each region is served by one and only one facility (optimal choice). Constraint (3) guarantees that an allocation is only possible to open facilities. Moreover, constraint (4) ensures that h_{jk} is a non-increasing function of k. The allocation of clients to the nearest open facility is guaranteed using equation (5). In this inequality constraint, M is a sufficiently large number. Constraint (6) is the budget constraint in which the total investment is shown as the sum of fixed establishment costs and the variable cost to add servers. Constraint (7) ensures that the number of clients referring to an open facility with k servers should be below the capacity of the facility to satisfy the maximum waiting time constraint. Finally, the set of constraints (8-9) are integrality constraints on x and h variables.

The proposed mathematical model is a mixed-integer programming model for which finding optimal solutions needs extended times if the problem size exceeds a certain threshold. That is why we propose an efficient heuristic to find near-optimal solutions quickly.

4 Solution procedure

Numerical experiments clearly show that exact algorithms are handicapped to solve instances with a rather large number of potential nodes in reasonable times. Hence, we propose an efficient variable neighborhood search in this paper. One may argue that since location problem is a strategic problem, there is not enough justification for using a non-exact method. However, the time to solve the problem to optimality grows exponentially which makes solving real-world problems almost impossible even in extended periods of time. Moreover, the problem discussed in this paper is applicable to cases such as mobile clinics where facilities are dynamic during the planning period (e.g. a day) which means that solutions should be found quickly. The benefit of using a heuristic is finding a near-optimal solution within a reasonable time. Although there are a myriad of procedures available to solve this problem, we proposed a VNS algorithm which is fast and has a high potential to avoid getting stuck in local optima. In this section, we will present an introduction to the classical VNS algorithm and discuss our developed version of that which has been tailored for PHDNP.

4.1 Variable Neighborhood Search

Variable Neighborhood Search (VNS) has been a popular solution approach since its introduction by Mladenovic et al. [33]. A clear indication of its popularity is the large number of its published applications such as those in scheduling (Karimi et al. [34]), vehicle routing (Belhaiza et al. [35]), and facility location (Davari et al. [36]). A VNS performs based on three assumptions (please note that the notations in this section are independent from those in section 3):

- A local optimum found by neighborhood structure N_i is not necessarily optimum for another N_j
- In many cases, the local optima of several neighborhood structures are close to each other
- When a local optimum is optimum with respect to all neighborhood structures, it is called a global optimum

The skewed VNS is a variant of VNS which can outperform the traditional one in many instances. In a SVNS, a similar approach to *simulated annealing* is pursued by accepting worse solutions with a certain probability.

In order to equip the classical VNS with the ability of exploring farther areas of solution space, solutions with worse fitness values which are good enough are accepted, provided that they are far enough from the current solution. In other words, in SVNS, the fitness of a new solution s'' is found as $f(s'') - \alpha\rho(s, s'')$ where ρ is the distance function which can be defined as the hamming or euclidean distance between solutions or any other rational distance function by the user. The pseudo-code of SVNS is given below.

Initialization. Define a set of neighborhood structures $N_k(k = 1, 2, ..., k_{\max})$, initialize a solution s, set $s^* \leftarrow s$ and $f^* \leftarrow f(s)$.

Repeat the following steps until the stopping criteria are met:

1. Set $k = 1$
2. ***Repeat*** the following steps until $k = k_{\max}$:
   (a) ***Shaking*** Generate a solution s' from the k^{th} neighborhood structure of s.
   (b) ***Local Search*** Using the local search procedure, try to improve s' and denote the improved solution as s''
   (c) ***Improvement Checking*** If $f(s'') < f^*$,set $f^* \leftarrow f(s'')$
   (d) ***Move or not*** If $f(s'' - \alpha\rho(s', s'')) < f(s)$, set $s \leftarrow s''$; otherwise set $k \leftarrow k + 1$

In the proposed SVNS, we have used two ρ functions, namely the hamming distance and the modified euclidean distance. The hamming distance d_{Ham} of two solutions is simply the number of different bits between two solutions. Moreover, we propose a modified euclidean distance which is defined as follows. Assuming S_1 as the set of opened facilities in the first solution and S_2 as the set of opened facilities in the second, the modified euclidean distance is defined as follows:

$$d_{MEuc}(S_1, S_2) = \kappa \frac{\sum\limits_{X_i \in S_1 \backslash S_2} \sum\limits_{X_j \in S_2 \backslash S_1} d(X_i, X_j)}{|S_1 \backslash S_2| \quad |S_2 \backslash S_1|} \tag{10}$$

in which X_i is the set of facilities opened in the ith solution, d is the euclidean distance measure between two locations on the plane, κ is the scaling factor, and $|.|$ is the cardinality of a set. In other words, our proposed measure finds the average distance between the uncommon located facilities of the two solutions.

4.2 Solution Representation

Solution representation plays a pivotal role in success of any heuristic method including VNS. The structure we have used in this paper represents a solution as a string of numbers. Assuming m population zones and n potential facilities, the representation consists of $l \leq n$ elements each containing an integer number between 1 and n and showing the index of the facility located. Needless to say, the values in a solution should be non-duplicate. For instance, a feasible solution is [1, 4, 5] which represents a solution in which three facilities are located at nodes 1, 4, and 5. As the problem is solved for the optimal-choice where each zone is allocated to its nearest facility, knowing the facilities located, the allocation scheme and the number of servers in each facility are easily determined.

4.3 Initialization

Needless to say, the initialization procedure is another vital element of any heuristic algorithm. In our proposed SVNS, we employed two distinct procedures to initialize the solutions. While in the first one, locations are generated randomly, the second one employs a roulette-wheel selection module to generate initial locations. To this end, assuming m population zones, both x and y planes are discretised into a number of identically-sized areas u. Then, each of the u^2 zones is given a value which equals the sum of the demands in that specific area. Then, using a roulette-wheel selection procedure, facilities are located in these zones. Figure 2 depicts how this procedure works for one of our problems with a uniform distribution of nodes, m=1000 where crosses show the location of facilities. In the following sections of the paper, the two initialization algorithms will be mentioned as RND for the random initialization and RWS for the roulette-wheel selection respectively.

Insert Figure 2 around here

4.4 Neighborhood search

The set of neighborhoods used for shaking is at the heart of the VNS. Each neighborhood should strike a proper balance between perturbing the incumbent solution and retaining the good parts of the incumbent solution [37]. Therefore, search operators should be selected meticulously to equip the algorithm with the ability of searching the space efficiently. The neighborhood search structures employed in this paper are explained in the following sections.

4.4.1 addFacility operator

This operator adds a random number of new facilities to the list of opened facilities and updates the assignment vector based on the new set of facilities. Assuming $l < n$ facilities open, the number of facilities to be added should be in the range $[1, n-l]$ to keep the solution feasible. However, if n facilities are located, this move is excluded from the list of moves in that iteration.

4.4.2 removeFacility operator

This operator deals with removing a random number of facilities from the list of open facilities and updating the assignment vector based on the new set of facilities. Assuming l facilities open, the number of facilities to remove should be in the range $[1, l-1]$ to keep the solution feasible.

4.4.3 swapFacility operator

This operator is to contribute to the algorithm by removing one of the facilities from the list of facilities located and selecting one of the closed ones to become open.

Figure 3 schematically represents the performance of these three operators on a single solution where modified bits in a solution are highlighted.

Insert Figure 3 around here

In any well-designed heuristic, there should be a balance between two strategies, namely diversification and intensification. In simple terms, diversification deals with exploration of the search space and intensification refers to exploitation of the current solution. In our procedure, the addFacility and removeFacility operators could be regarded as moves which diversify the solutions and the swapFacility is the operator used for intensification. Hence, addFacility, and removeFacility are used as shaking mechanisms respectively and swapFacility is the local search mechanism. The initialization procedure is also run if the procedure fail to improve the solution in a certain number of iterations.

4.5 Stopping Criteria

Our experiments showed that running the procedure for more than 60 seconds rarely results in an improvement in the solution quality. Hence, the proposed procedure stops after running for 60 seconds regardless of the problem size and parameters.

5 Numerical experiments

In order to assess the performance of the proposed algorithm, a set of 54 hypothetical test problems with different characteristics was developed. For each one of the 54 hypothetical test problems, ten random replications were generated. Hence in total 540 experiments were conducted. These 54 test problems were generated for $m = 100, 150, 250$; $n = 25, 50, 75$; and $B = \lfloor \frac{n}{5} \rfloor * \delta$ for $\delta = 3000, 4000$ and 5000. For each of these 27 possible combinations, node coordinates were generated following a uniform distribution and also a normal distribution. In all of the instances, the x-coordinate and the y-coordinate values were generated in a way that nodes get a value in the range $[0, 30]$. The travel times between nodes are assumed to be equal to the distance (i.e, the unit cost of traveling equals to 1) and found using the euclidean measure. Fixed costs to establish facilities were generated using a uniform distribution between 1000 and 5000.

5.1 Parameter tuning

In order to fine-tune the proposed SVNS, we have done statistical analysis (Analysis of Variance, i.e., ANOVA) on the effects of three possibly effective parameters, namely the value of α ($\alpha_1 = 0.01$, $\alpha_2 = 0.03$, $\alpha_3 = 0.05$), the initialization procedure ω (ω_1: RND and ω_2: RWS), and also the ρ function (ρ_1: hamming and ρ_2: euclidean). Table 1 demonstrates the results of the ANOVA test and also the significance of all the three parameters in the performance of the algorithm where $\times$ shows the interaction between parameters. In this table, the rows containing a significant parameter are highlighted in gray. Results show that all the three parameters tested have considerable effects on the algorithm performance. Besides, the interaction of the initialization method and the value of α is significant at the 5% level. In order to find the optimal values of these parameters, ten runs of each of the twelve possible settings for the problem with normal distribution of nodes and $(m, n, Q_{\max}) = (250, 75, 20)$ were conducted. The Box-Whisker diagram of the results of various combinations is presented in Figure 4 where settings are shown on the horizontal axis as (α, ω, ρ). The results suggests that $\alpha = 0.01$, initializing solutions using the RWS methodology and using the eclidean distance as the distance measure of two solutions as the best parameter setting.

Insert Figure 4 around here

Table 1: ANOVA results

Source	**Sum Sq.**	**d.f.**	**Mean Sq.**	F	$Prob \geq F$
ω	660.08	1	660.08	5.11	0.0296
ρ	2523	1	2523	19.54	0.0001
α	7626.12	2	3813.06	29.54	0
$\omega \times \rho$	0.75	1	0.75	0.01	0.9396
$\omega \times \alpha$	1424.29	2	712.15	5.52	0.0079
$\rho \times \alpha$	465.12	2	232.56	1.8	0.1789
Error	4905.63	38	129.1		

6 Results and discussion

In this section, we present the results of the computational experiments for a set of test problems. The proposed SVNS was coded in C++ language, compiled using Visual Studio 2012 and implemented on a laptop with Intel Core 2 Duo 2.33 GHz processor with 4 GB RAM. Moreover, the exact results were obtained using IBM ILOG CPLEX Studio 12.5.

Table 2 and Table 3 report the performance of our proposed SVNS for normal and uniform data sets respectively. In these tables, column 5 shows the value of δ which represents the budget parameter. Moreover, the sixth and seventh columns show the optimal value obtained using CPLEX and the required time. The columns 8-10 show the worst, the average and the best fitnesses found in ten runs using SVNS respectively. Besides, column 11 is the run time of the proposed SVNS which is 60 seconds for all instances regardless of the problem characteristics. Moreover, the last three columns show the Relative Percentage Deviation (RPD) of the worst, the average and the best performance of the algorithm. Please note that RPD is found using the below formulation:

$$RPD = \frac{Z_{CPLEX} - Z_{BEST}}{Z_{CPLEX}} \tag{11}$$

where Z_{BEST} and Z_{CPLEX} are the best value found using the proposed VNS and the optimal values found using CPLEX respectively. Please note that Z_{CPLEX} can be the optimal value of the problem or an upper bound for those problems for which CPLEX was unable to reach optimality in 6000 seconds.

For those instances where the proposed SVNS was able to reach the optimal fitness in all the ten runs, the whole row is highlighted as gray. Results show that such rows are more common for those instances with a normal distribution of nodes. Such an observation makes sense as in the Normal data set, demand nodes tend to be located in the center, so solutions are most robust to change. In addition, in some instances, CPLEX was unable to reach the optimal solution in 6000 seconds and we have reported the gap between the output of the proposed SVNS and that upper bound. For those rows for which CPLEX was unable to reach optimal solution in more than 6000 seconds, an asterisk has been added as a superscript to the index of the problem in the respective row. Another interesting finding is the fact that the value of n affects the runtime much more than the value of m. Knowing that, we divide the problems in three groups as small scale ($n = 25$), medium scale ($n = 50$) and large scale ($n = 75$). Figure 5 depicts the performance of the algorithm on these three subcategories of data. Results show that on average, the error of the proposed procedure is 0.36% which is a clear indication of its good performance.

Another analysis is about the role of the budget parameter δ on the solution and how it affects the quality of solutions found by the proposed heuristic. Results show that although there is not a strong correlation between the budget available and the quality of solutions found by VNS, in more than 70% of cases (13 out of 18), the quality of the best solution found is worse for those cases with a higher budget (a higher value for the parameter δ). On the other hand, the time to solve test problems to optimality is normally higher or those cases with a higher value for the budget parameter. This can be attributed to the fact that an increased budget extends the feasible region which means a larger region to explore for both the algorithms.

Insert Figure 5 around here

Results clearly show that the proposed SVNS achieves results with an error up to 1.80%. Furthermore, in 50% of test instances, SVNS could reach the optimal solution in at least one of the runs and in almost 20% of the cases, it gets the optimal results in all the ten runs. Therefore, the proposed algorithm is capable in reaching high quality solutions in considerably less time than CPLEX.

7 Conclusion and future research

In this paper, we have addressed the problem of preventive health care network design. We developed an efficient SVNS to solve this problem and examined its performance on some randomly generated data. The approach proposed in this paper can be used in designing a preventive health care network from scratch or improving the status-quo for an already established network. However, this calls for having a data warehouse comprising the population of each residential area, travel times, and other factors as well. The set of candidate locations can be defined using various techniques prior to being used in the proposed method as a function of various social, political, and economical issues, using qualitative/quantitative methods. Case studies similar to ours were carried out in Canada (such as Gu et al. [12]) and other countries (interested readers are referred to these reports for further information).

We believe that future research stems from considering probabilistic choice environment, assuming uncertain travel times or developing other heuristics. Another appealing future research is to assume a case in which preventive facilities are dynamic, like immunization programs. Then, the problem is to locate facilities dynamically and to decide on the time the locations should be changed. As another possible extension to the PHNFDP problem, other qualitative factors such as quality of the health care facility, availability of amenities near the facility etc. which also influence the attractiveness of the health care facility besides the proximity, can also be incorporated while modeling participation to preventive programs. A possible way of such an extension is modeling participation by means of fuzzy numbers rather than crisp numbers and utilize fuzzy mathematical programming approaches instead of traditional mathematical programming. As another extension, one can conduct empirical analysis into using other solution methods such as different heuristics such as evolutionary algorithms or simulation-based optimization procedures in case of uncertainty in the problem parameters. Last but not least is the possibility of integrating preventive, screening and treatment decisions. In most of the

Table 2: Results of CPLEX vs. SVNS for the Normal data

Instance Specifications					CPLEX		SVNS				Gaps		
Number	**Distribution**	m	n	δ	Z_{CPLEX}	**Time**	**Worst**	**Average**	**Best**	**Time**	**Worst**	**Average**	**Best**
1	Normal	100	25	3000	45.863	10	45.863	45.863	45.863	60	0%	0%	0%
2	Normal	100	25	4000	66.223	45	66.223	66.223	66.223	60	0%	0%	0%
3	Normal	100	25	5000	86.281	80	86.281	86.281	86.281	60	0%	0%	0%
4	Normal	100	50	3000	45.157	86	44.917	45.123	45.157	60	0.53%	0.08%	0%
5	Normal	100	50	4000	70.952	30	70.834	70.858	70.952	60	0.17%	0.13%	0%
6	Normal	100	50	5000	91.421	46	91.256	91.331	91.421	60	0.18%	0.1%	0%
7	Normal	100	75	3000	45.742	322	45.441	45.647	45.742	60	0.66%	0.21%	0%
8	Normal	100	75	4000	67.4	1222	66.949	67.057	67.129	60	0.67%	0.51%	0.4%
9*	Normal	100	75	5000	89.366	6000	88.565	88.706	88.774	60	0.9%	0.74%	0.66%
10	Normal	150	25	3000	44.543	34	44.519	44.532	44.543	60	0.05%	0.02%	0%
11	Normal	150	25	4000	68.732	21	68.732	68.732	68.732	60	0%	0%	0%
12	Normal	150	25	5000	89.38	30	89.38	89.38	89.38	60	0%	0%	0%
13	Normal	150	50	3000	45.547	228	45.339	45.486	45.547	60	0.46%	0.13%	0%
14	Normal	150	50	4000	67.768	274	67.357	67.448	67.698	60	0.61%	0.47%	0.1%
15	Normal	150	50	5000	87.727	3011	86.945	87.409	87.661	60	0.89%	0.36%	0.08%
16	Normal	150	75	3000	45.04	1345	44.237	44.812	44.871	60	1.78%	0.51%	0.38%
17	Normal	150	75	4000	70.32	187	69.598	69.992	70.212	60	1.03%	0.47%	0.15%
18*	Normal	150	75	5000	90.476	6000	89.646	90.021	90.084	60	0.92%	0.5%	0.43%
19	Normal	250	25	3000	45.854	59	45.854	45.854	45.854	60	0%	0%	0%
20	Normal	250	25	4000	66.741	193	66.741	66.741	66.741	60	0%	0%	0%
21	Normal	250	25	5000	87.197	539	86.923	87.133	87.197	60	0.31%	0.07%	0%
22	Normal	250	50	3000	45.519	766	45.31	45.4	45.519	60	0.46%	0.26%	0%
23	Normal	250	50	4000	70.458	186	70.152	70.272	70.458	60	0.43%	0.26%	0%
24*	Normal	250	50	5000	89.4	6000	88.875	89.11	89.313	60	0.59%	0.32%	0.1%
25*	Normal	250	75	3000	46.907	6000	46.217	46.275	46.321	60	1.47%	1.35%	1.25%
26*	Normal	250	75	4000	45.077	4309	44.432	44.918	44.973	60	1.43%	0.35%	0.23%
27*	Normal	250	75	5000	86.782	6000	85.522	86.107	86.446	60	1.45%	0.78%	0.39%

Table 3: Results of CPLEX vs. SVNS for the Uniform data

Instance Specifications					CPLEX		SVNS				Gaps		
Number	**Distribution**	m	n	δ	Z_{CPLEX}	**Time**	**Worst**	**Average**	**Best**	**Time**	**Worst**	**Average**	**Best**
28	Uniform	100	25	3000	10.257	4	10.257	10.257	10.257	60	0%	0%	0%
29	Uniform	100	25	4000	10.94	4	10.94	10.94	10.94	60	0%	0%	0%
30	Uniform	100	25	5000	11.206	3	11.199	11.201	11.206	60	0.06%	0.04%	0%
31	Uniform	100	50	3000	41.153	244	41.046	41.061	41.153	60	0.26%	0.22%	0%
32	Uniform	100	50	4000	61.553	527	61.126	61.295	61.451	60	0.69%	0.42%	0.17%
33	Uniform	100	50	5000	77.375	2412	76.688	76.891	77.204	60	0.89%	0.63%	0.22%
34	Uniform	100	75	3000	41.781	1177	41.098	41.332	41.665	60	1.63%	1.07%	0.28%
35	Uniform	100	75	4000	62.213	2439	61.426	61.68	62.152	60	1.27%	0.86%	0.1%
36*	Uniform	100	75	5000	81.27	6000	80.377	80.544	80.749	60	1.1%	0.89%	0.64%
37	Uniform	150	25	3000	41.57	49	41.406	41.499	41.57	60	0.39%	0.17%	0%
38	Uniform	150	25	4000	60.077	84	59.873	59.994	60.077	60	0.34%	0.14%	0%
39	Uniform	150	25	5000	75.393	335	74.996	75.249	75.393	60	0.53%	0.19%	0%
40	Uniform	150	50	3000	41.495	559	41.317	41.384	41.495	60	0.43%	0.27%	0%
41	Uniform	150	50	4000	61.371	1443	61.145	61.254	61.361	60	0.37%	0.19%	0.02%
42	Uniform	150	50	5000	76.886	9789	76.302	76.487	76.685	60	0.76%	0.52%	0.26%
43	Uniform	150	75	3000	41.674	3321	40.925	41.327	41.427	60	1.8%	0.83%	0.59%
44*	Uniform	150	75	4000	63.207	6000	62.56	62.646	63.095	60	1.02%	0.89%	0.18%
45*	Uniform	150	75	5000	80.906	6000	79.709	80.169	80.741	60	1.48%	0.91%	0.2%
46	Uniform	250	25	3000	41.627	179	41.511	41.579	41.627	60	0.28%	0.12%	0%
47	Uniform	250	25	4000	59.677	971	59.491	59.521	59.644	60	0.31%	0.26%	0.06%
48	Uniform	250	25	5000	75.528	1187	75.326	75.441	75.528	60	0.27%	0.12%	0%
49	Uniform	250	50	3000	41.532	1078	41.231	41.465	41.532	60	0.72%	0.16%	0%
50*	Uniform	250	50	4000	62.007	6000	61.475	61.644	61.894	60	0.86%	0.59%	0.18%
51*	Uniform	250	50	5000	79.747	6000	78.997	79.365	79.48	60	0.94%	0.48%	0.33%
52*	Uniform	250	75	3000	41.677	6000	41.205	41.255	41.329	60	1.13%	1.01%	0.83%
53*	Uniform	250	75	4000	62.689	6000	61.737	62.261	62.509	60	1.52%	0.68%	0.29%
54*	Uniform	250	75	5000	81.518	6000	80.936	81.001	81.172	60	0.71%	0.63%	0.42%

studies to date, these three are assumed to be independent. However, in reality, these three have an impact on a patient's outcomes such as quality adjusted life span and total cost of health care.

Acknowledgement

This work was supported by The Scientific and Technological Research Council of Turkey under grant TUBITAK-2216.

References

[1] Michael V. Maciosek, Ashley B. Coffield, Thomas J. Flottemesch, Nichol M. Edwards, and Leif I. Solberg. Greater use of preventive services in us health care could save lives at little or no cost. *Health Affairs*, 29(9):1656–1660, 2010.

[2] Centers for Disease Control, Prevention (CDC), et al. Hiv prevention in the united states: At a critical crossroads. *Retrieved February*, 22:2012, 2009.

[3] Barbara A Ormond, Brenda C Spillman, Timothy A Waidmann, Kyle J Caswell, and Bogdan Tereshchenko. Potential national and state medical care savings from primary disease prevention. *American Journal of Public Health*, 101(1):157–164, 2011.

[4] Erik Biørn and Geir Godager. Does quality influence choice of general practitioner? an analysis of matched doctor–patient panel data. *Economic Modelling*, 27(4):842–853, 2010.

[5] Hugh S Gravelle, Carol Propper, and Rita Santos. Does quality affect patients? choice of doctor? evidence from the uk. Technical report, CEPR Discussion Papers, 2013.

[6] Ivo Müller, Tom Smith, Steve Mellor, Lawrence Rare, and Blaise Genton. The effect of distance from home on attendance at a small rural health centre in papua new guinea. *International Journal of Epidemiology*, 27(5):878–884, 1998.

[7] Marco Varkevisser, Stéphanie A. van der Geest, and Frederik T. Schut. Do patients choose hospitals with high quality ratings? empirical evidence from the market for angioplasty in the netherlands. *Journal of Health Economics*, 31(2):371 – 378, 2012.

[8] Robin Haynes, Andrew Lovett, and Gisela Sunnenberg. Potential accessibility, travel time, and consumer choice: geographical variations in general medical practice registrations in eastern england. *Environment and Planning A*, 35(10):1733–1750, 2003.

[9] Yue Zhang, Oded Berman, and Vedat Verter. Incorporating congestion in preventive healthcare facility network design. *European Journal of Operational Research*, 198(3):922–935, 2009.

[10] Yue Zhang, Oded Berman, Patrice Marcotte, and Vedat Verter. A bilevel model for preventive healthcare facility network design with congestion. *IIE Transactions*, 42(12):865–880, 2010.

[11] Yue Zhang, Oded Berman, and Vedat Verter. The impact of client choice on preventive healthcare facility network design. *OR spectrum*, 34(2):349–370, 2012.

[12] Wei Gu, Xin Wang, and S. Elizabeth McGregor. Optimization of preventive health care facility locations. *International journal of health geographics*, 9(1):17, 2010.

[13] Ciaran S Phibbs and Harold S Luft. Correlation of travel time on roads versus straight line distance. *Medical Care Research and Review*, 52(4):532–542, 1995.

[14] Nadine Schuurman, Robert S Fiedler, Stefan CW Grzybowski, and Darrin Grund. Defining rational hospital catchments for non-urban areas based on travel-time. *International Journal of Health Geographics*, 5(1):43, 2006.

[15] Joshua T Cohen, Peter J Neumann, and Milton C Weinstein. Does preventive care save money? health economics and the presidential candidates. *New England Journal of Medicine*, 358(7):661–663, 2008.

[16] Mark S Daskin and Latoya K Dean. Location of health care facilities. In *Operations research and health care*, pages 43–76. Springer, 2004.

[17] S. Louis Hakimi. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3):450–459, 1964.

[18] Jeroen Beliën, Liesje De Boeck, Jan Colpaert, Stijn Devesse, and Filip Van den Bossche. Optimizing the facility location design of organ transplant centers. *Decision Support Systems*, 54(4):1568–1579, 2013.

[19] Malick Ndiaye and Hesham Alfares. Modeling health care facility location for moving population groups. *Computers & Operations Research*, 35(7):2154–2161, 2008.

[20] Afshin Shariat-Mo.man, Mohsen Babaei, Saeed Moadi, and Sayyed Mahdi Amiripour. Linear upper-bound unavailability set covering models for locating ambulances: Application to tehran rural roads. *European Journal of Operational Research*, 221(1):263–272, 2012.

[21] Roberto D. Galvao, Luis Gonzalo Acosta Espejo, and Brian Boffey. A hierarchical model for the location of perinatal facilities in the municipality of rio de janeiro. *European Journal of Operational Research*, 138(3):495–517, 2002.

[22] Cara Cocking, Steffen Flessa, and Gerhard Reinelt. Improving access to health facilities in nouna district, burkina faso. *Socio-Economic Planning Sciences*, 46(2):164–172, 2012.

[23] S. S. Shariff, Noor Hasnah Moin, and Mohd Omar. Location allocation modeling for healthcare facility planning in malaysia. *Computers & Industrial Engineering*, 62(4):1000–1010, 2012.

[24] Vedat Verter and Sophie D. Lapierre. Location of preventive health care facilities. *Annals of Operations Research*, 110(1-4):123–132, 2002.

[25] Evrim D. Gunes, Hande Yaman, B. Cekyay, and Vedat Verter. Matching patient and physician preferences in designing a primary care facility network. *Journal of the Operational Research Society*, 2012.

[26] Evrim D Güneş, Stephen E Chick, and O Zeynep Akşin. Breast cancer screening services: trade-offs in quality, capacity, outreach, and centralization. *Health care management science*, 7(4):291–303, 2004.

[27] Walter Werner Holland and Susie Stewart. *Screening in disease prevention: what works?* Radcliffe Publishing, 2005.

[28] Steven H Woolf, Corinne G Husten, Lawrence S Lewin, James S Marks, JE Fielding, and EJ Sanchez. *The economic argument for disease prevention: distinguishing between value and savings*. Partnership for Prevention, 2009.

[29] Joshua T Cohen, Peter J Neumann, and Milton C Weinstein. Does preventive care save money? health economics and the presidential candidates. *New England Journal of Medicine*, 358(7):661–663, 2008.

[30] Emel Aktaş, Özay Özaydın, Burçin Bozkaya, Füsun Ülengin, and Şule Önsel. Optimizing fire station locations for the istanbul metropolitan municipality. *Interfaces*, 43(3):240–255, 2013.

[31] Eusebio Angulo, Enrique Castillo, Ricardo García-Ródenas, and Jesús Sánchez-Vizcaíno. A continuous bi-level model for the expansion of highway networks. *Computers & Operations Research*, 41:262–276, 2014.

[32] Shahab Kermanshahi, Yousef Shafahi, and Mehdi Bagherian. Application of a new rapid transit network design model to bus rapid transit network design: case study isfahan metropolitan area. *Transport*, (ahead-of-print):1–10, 2013.

[33] Nenad Mladenović and Pierre Hansen. Variable neighborhood search. *Computers & Operations Research*, 24(11):1097–1100, 1997.

[34] Hamid Karimi, Seyed Habib A. Rahmati, and M. Zandieh. An efficient knowledge-based algorithm for the flexible job shop scheduling problem. *Knowledge-Based Systems*, 2012.

[35] Slim Belhaiza, Pierre Hansen, and Gilbert Laporte. A hybrid variable neighborhood tabu search heuristic for the vehicle routing problem with multiple time windows. *Computers & Operations Research*, 2013.

[36] Soheil Davari, Mohammad Hossein Fazel Zarandi, and Ismail Burhan Turksen. A greedy variable neighborhood search heuristic for the maximal covering location problem with fuzzy coverage radii. *Knowledge-Based Systems*, 2013.

[37] Vera Hemmelmayr, Karl F. Doerner, Richard F. Hartl, and Martin W. P. Savelsbergh. Delivery strategies for blood products supplies. *OR spectrum*, 31(4):707–725, 2009.

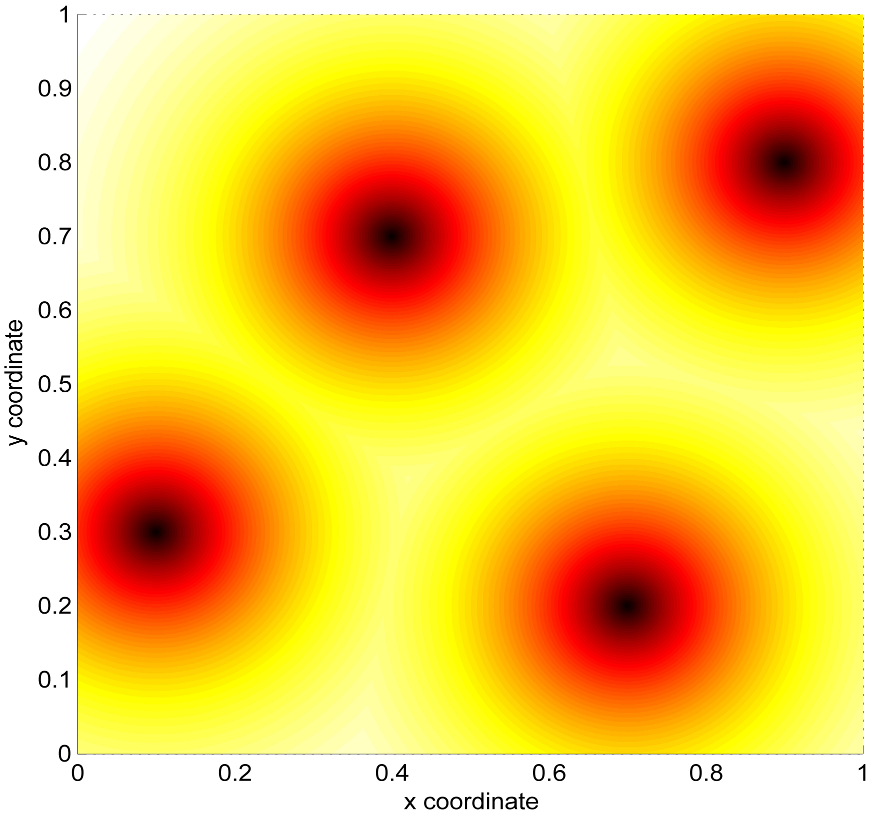

Figure 1: The attractiveness on the plane with an exponential attractiveness function

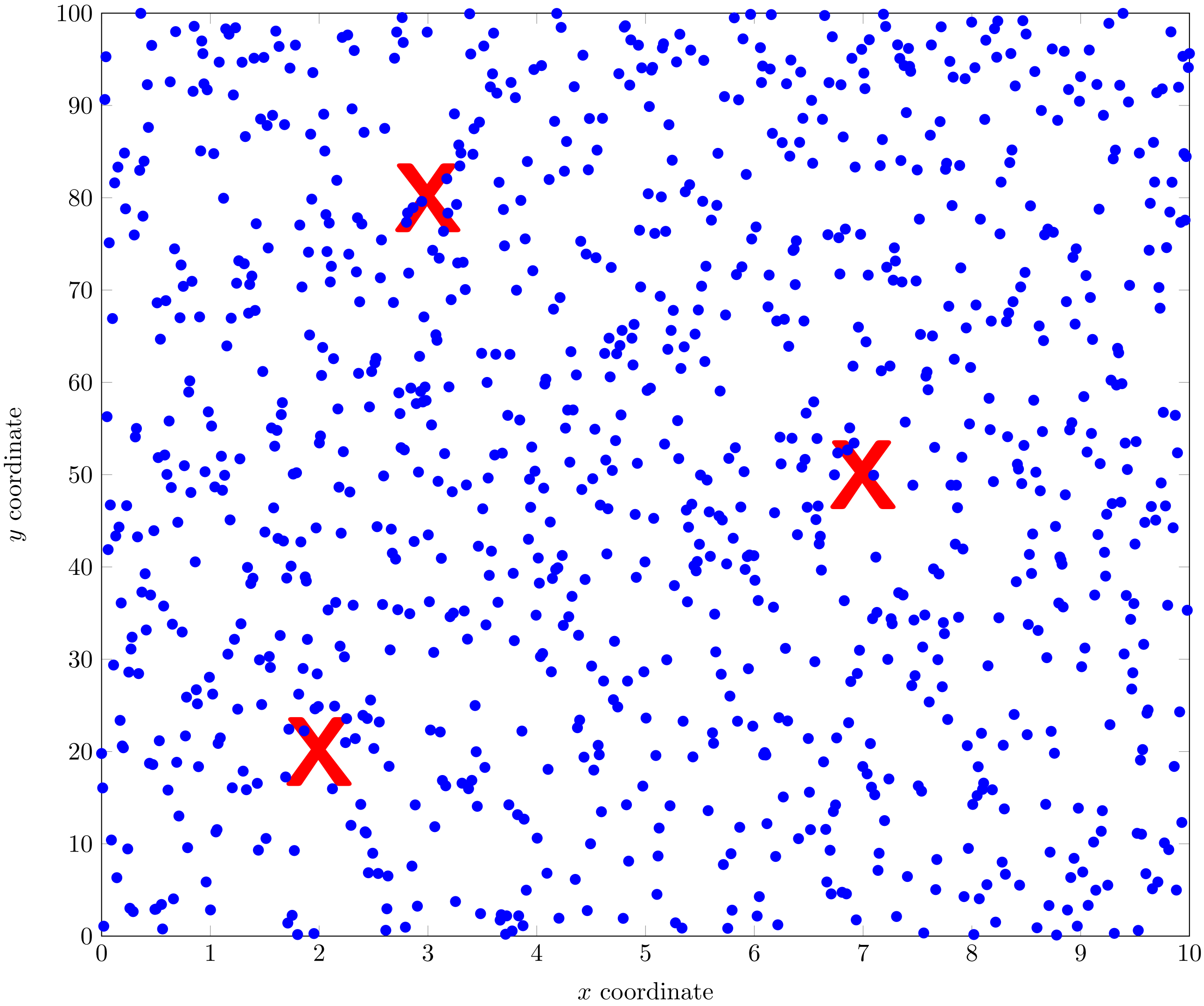

Figure 2: A sample of the initialization procedure

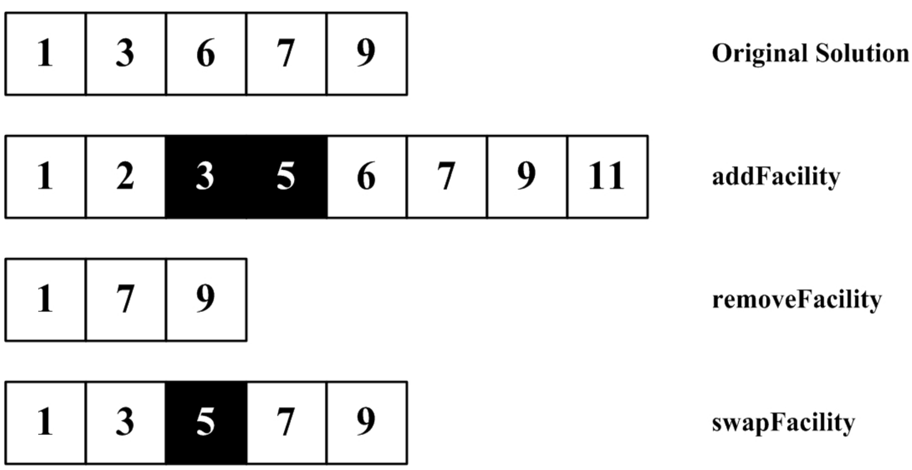

Figure 3: The schematic view of the neighborhood search structures

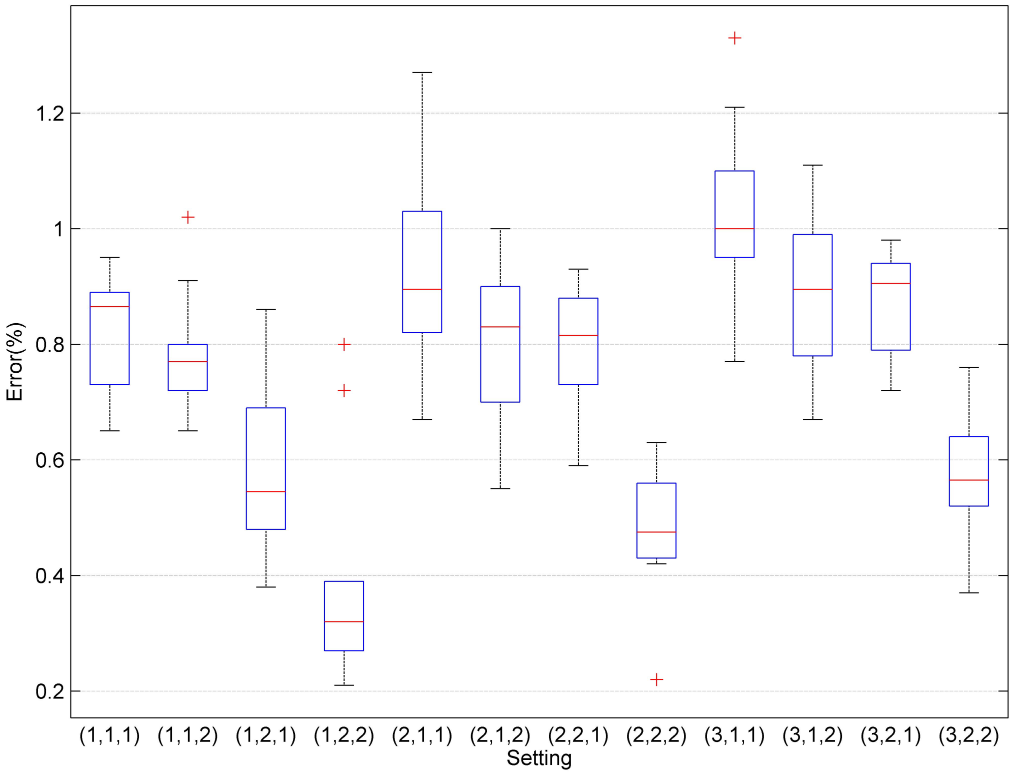

Figure 4: Parameter Tuning

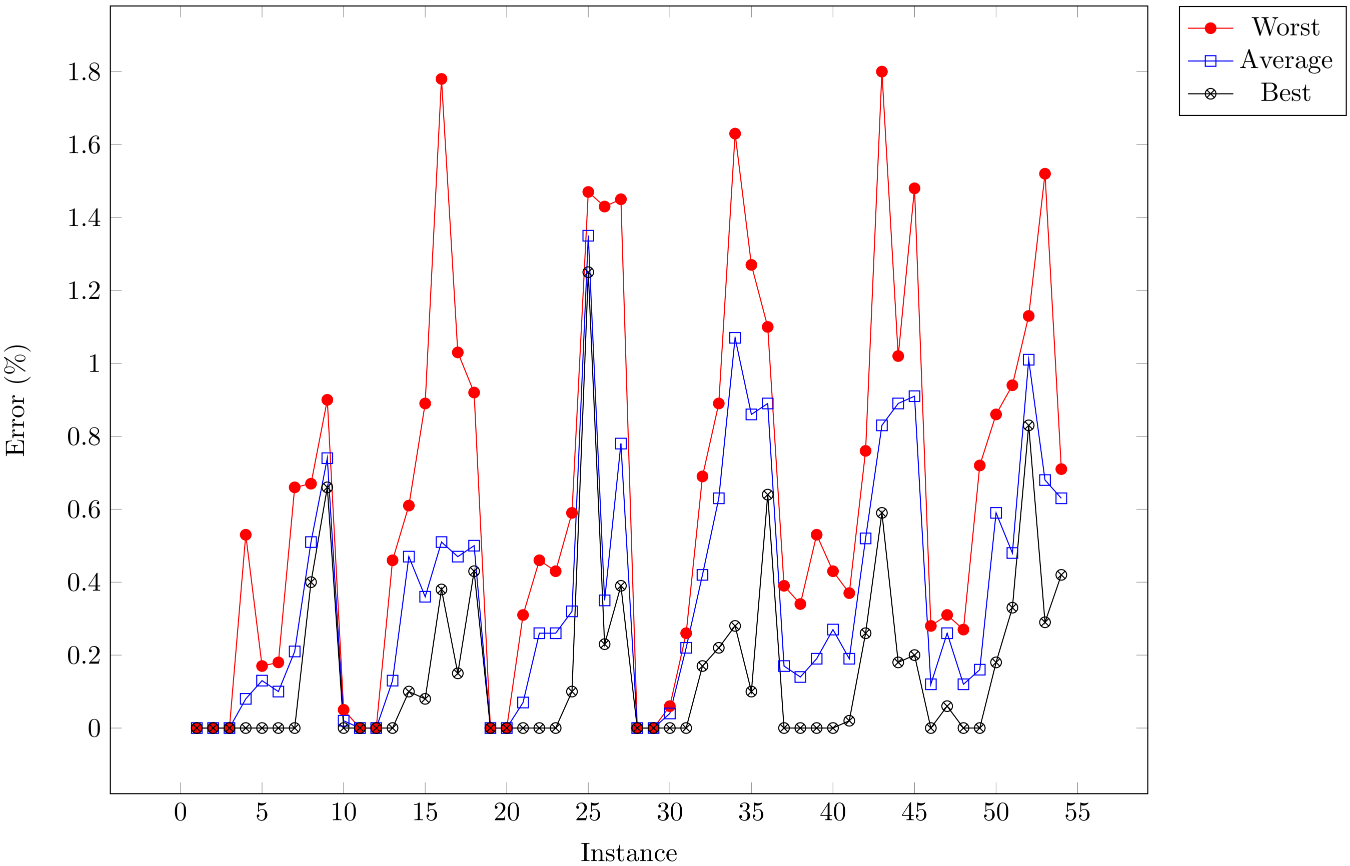

Figure 5: Comparing the performance of the algorithms for different instances based on the problem size