Scenario Analysis with Facility Location Optimization

Shen-Tat Goh, Shudong Liu, Tracy Yong, Eric Foo
*Institute for Infocomm Research, A*STAR*
1 Fusionopolis Way, Singapore 138632
(stgoh, liush, tracy-yong, foosk)@i2r.a-star.edu.sg

***Abstract*— In a densely populated city, effective land uses and redevelopment is critical to enable growth, development and renewal in all aspects of city planning. For instance, sufficient housing provision, adequate supporting infrastructure and accessible amenities are required to support population growth. Urban planners face the issue of where to locate facilities (e.g. clinics, elder/childcare centres and schools) such that they can best serve their intended target population. This presents the problem of arriving at a distribution of facilities that maximally optimizes coverage to target populations while minimizing the number of facilities to build. In this paper, we proposed a planning decision support platform for city planning with new flexible mixed integer programming models at the core and an interactive interface at the front-end that is integrated with a Geo-Information System. The system can generate scenarios, calculate travel time/distance based on road network, call and solve optimization models, and do scenario comparison and sensitivity analysis. Using existing and projected data for population growth and schools as an example, we examined the effectiveness and performance of our decision system and optimization approach for city planning. The results show that our optimization model and approach can solve real problems within a reasonable timeframe and can provide valuable insights for planners to inform land use planning decisions.**

Keywords—Site Search, City Planning, Location Optimization, Decision Support Systems

I. Introduction

Urban City Planning is a continuous process to enhance liveability and sustainability in any densely populated city. Rapid population growth drives innovation in creating efficient resource renewal and management within the city. For instance, land is a valuable resource required for the development of critical infrastructure, key amenities and housings. In the city, there may exists many competing resource issues like road congestions; under supply of public services; inaccessibility or inefficient connectivity to key amenities; insufficient green coverage or social interaction space etc. Proper planning to determine the supply and distribution for key land uses is essential. This entails a provision and coverage analysis of existing land uses, as well as plans for future capacities and placements of infrastructure such as roads, transport hubs and utilities; amenities such as healthcare centers, schools, childcare and greeneries; and other land uses such as housing, commerce and industries.

In this paper, we will examine the facility placement problem making use of a diverse set of city planning data. This paper is structured as follows. In Section II, the problem formulation will be introduced together with the related works. In Section III, a new flexible optimization method will be proposed and discussed. This is followed by the experiments setup and results in Section IV and the conclusion in Section V.

II. Problem Formulation and Related Works

A. Data Richness and Analysis

The collection of large amount of data by organizations for data analysis has been a runaway trend in this century. With the proliferation of data processing technology adoption to extract insights from these data, the organizations have gathered much anticipated gains.

Data is further categorized into many diverse forms: structured or unstructured; simple or complex; transactional or analytical; text or media; attributed or free form; locational or location-free etc. Meta-data, annotations and other relationship information further enrich the pool of available data one can collect over a long period of time, and the sampling rate at which the data are collected gives the granularity or resolution for the data set. The analysis of this large and diversified set of data can provide insights showing both positive and negative outcomes, which are equally valuable if the relevance and confidence levels are sufficiently high to confirm or influence a decision.

In our analysis, we consider data received from multiple planning sources/agencies and the data comes in a wide range of formats as describe above, varying from geo-spatial spread and resolution, to temporal snapshots and future projections.

B. Related Works

Early works on location analysis like the Weber problem and its variants [1, 8] is based on measuring the interrelated factors and agglomeration (transportation, labor, suppliers etc.) to determine the extent of locating an industry to maximize the economic value of the production line. Likewise in urban planning, a facility like the industry looks for interrelated factors around it to provide services for a cluster of residents or demand points. Each demand point is assigned to a single facility, and the problem is to locate a number of facilities to serve these demand points. An example to reduce the carbon footprint in city planning has been the use of shared vehicles [13] and rented bicycles [12]. In [12], various parameters, such as the human mobility, point of interest and public transportation accessibility, are considered in order to balance the pick-ups and the drop-off at the bicycle rental stations.

Discrete Facility location problem (FLP) on general road network is an NP-hard problem, and one solution is to

This research is supported by the Singapore Ministry of National Development and the National Research Foundation, Prime Minister's Office under the Land and Liveability National Innovation Challenge (L2 NIC) Research Programme (L2 NIC Award No: L2NICTDF1-2016-2).

approximate it as a probabilistic clustering problem [5, 6]. The solutions for such problems have been defined by models and there is no single general location model that can be used for all applications as indicated in [2]. There are two categories on single demand point to facility mapping: one based on maximum or cover distance [4] and the other on the total or average distance, for instance the use of k-median algorithm as presented in [7, 9]. In [4], Current et al describe a set of maximal covering location models to the problem of locating emergency warning sirens. There is another category, known as location routing, which optimizes multiple demands along a serving route [3, 11]. This is usually applied in a logistics problem where one facility can service the demands from multiple clients along an optimized route. In [10], a simulation model based on Monte Carlo and swarm optimization is presented to study the coverage location problem in health services.

In this paper, we consider city planning problems for social and community facilities. We present a platform to assist city planning in which a new flexible Mixed Integer Programming model is developed considering actual planning scenarios. The main characteristics of our approach is as follows: a) in the literature most researchers assume all facilities are new, while we consider the facility planning with existing facilities, which can be phased out or remain to provide service. b) We consider the land price in the model, as land is a very important factor in city planning. c) As a social and community facility, the facility cost may fall on the government, while the residence's transportation cost and other indirect cost are on residence. This is unlike other facility problems where the fixed cost and transportation cost are both on the company. So, our problem is in fact a multi-criteria optimization problem. We use a weight to combine multiple objectives and then change the weight to obtain the Pareto frontiers. d) Uncertainty in demand or other parameters exists in most of problems. The dynamic change, due to the demand shift from the multitude of city development and other factors, poses a challenge for planners to plan effectively, especially in identifying changes in demands for land uses. We address this uncertainty issue in the way of generating multiple scenarios and doing scenario analysis. This is easier to understand and apply in practice.

Our model will run against a diverse set of existing and long-term planning data to investigate the effectiveness of our approach. In the next section, we will present the optimization model and solution approach.

III. Facility Optimization Methods

There are numerous facilities that are of interest to city planners. For instance, the placements of schools to cater for enough enrolments, childcare or eldercare centers for the caring of children and elderly, and other public servicing amenities like clinics, supermarkets, banks, public transport hubs, parks and communal areas. In the following, we will present the necessary assumptions, notations and how we can address the facility placement challenges.

A. *Optimzation Methods*

The proposed methods include considerations for the following scenario composition variations:

- Cost of travelling from the demand points to the facilities based on: (a) Direct distance, (b) road network and (c) travelling time model matrix
- Use of capacitated and Un-capacitated facilities
- Presence of phase-in(add) and phase-out(remove) of a number of facilities over a period of time
- Recommendation of the minimum number of facilities with the associated capacities to optimize cost

We first define an example to the facility location problem. Given that there are n potential sites for building schools (existing and new) and m demand points. For each new site j, there is a yearly activation cost f_j, i.e., an annual leasing expense that is incurred for using it, independent of the volume it services. Each facility has a capacity for the amount that may be handled yearly, M_j. Additionally, there is a transportation cost c_{ij} per unit serviced from facility j to the demand point i.

The demand of a demand point can be satisfied by more than one facility. The change in demand quantity for each demand point or the number of demand points drives new allocations with new school additions and removals. If we remove an existing facility j, then there is a one-time cost $\underline{G_j}$. Let T be the planning horizon (years). We can allocate the dismantle cost G_j over the whole horizon. We want to find out how many new facilities to build and their locations, and which existing facilities to remove by minimizing total cost which includes the facility costs and transportation costs. Table 1 shows the notations used by the proposed models.

Parameters and Constraints	
m	number of demand points
I	$\{1, \ldots, m\}$, set of demand points
n	number of potential points for facility
J	$\{1, \ldots, n\}$, set of potential points for facilities including existing and new ones
c	$\{c_{ij} \mid i \in I, j \in J\}$, the matrix of travelling distance/time/cost from demand point i to facility location j, or transportation cost per unit serviced from facility j to the demand point i
d	$\{d_i \mid i \in I\}$, set of weights (e.g. demands, population) at demand points
J_e	$J_e \in J$, set of locations where an existing facility is located, assuming there is at most one facility located at one location point. If there is more than one facility at one location point, we can regard them as one big facility whose capacity is the sum of those capacities
J_n	$J \backslash J_e$, set of potential locations for new facilities. J_n has no overlap with J_e, based on the assumption here is at most one facility located at one location point
g	$\{g_j \mid j \in J_e\}$, average facility cost per year for dismantling an existing facility at location j, by allocating the one-time dismantling cost G_j over planning horizon T. So, $\underline{g_j = G_j/T}$
f	$\{f_j \mid j \in J_n\}$, fixed cost of opening a new facility at location j, an annual leasing expense that is

	incurred for using it, independently of the volume it services
M	$\{M_j \mid j \in J\}$, capacity of facilities at locations
SL	Service level
Wt	Weight for travel cost in objective
Decision Variables	
x_{ij}	$x_{ij} \geq 0$, $i \in I$, $j \in J$, demand amount at demand point i serviced by facility j
y_j	$y_j \in \{0, 1\}$, $j \in J_n$, 1 if a new facility is established at location j, 0 otherwise
z_j	$z_j \in \{0, 1\}$, $j \in J_e$, 1 if an existing facility at location j is dismantled, 0 otherwise

Table 1: Model Notations

The model considers the costs for opening new facilities, costs for removing existing facilities and travelling cost. The formulation is as follows:

$$\min \quad \sum_{j\in J_n} f_j y_j + \sum_{j\in J_e} g_j z_j + wt \sum_{i\in I}\sum_{j\in J} c_{ij} x_{ij} \tag{1}$$

$$\text{s.t.} \quad \sum_{j\in J} x_{ij} = d_i, \forall i \in I \tag{2}$$

$$\sum_{i\in I} x_{ij} \leq M_j y_j, \forall j \in J_n \tag{3}$$

$$\sum_{i\in I} x_{ij} \leq M_j(1 - z_j), \forall j \in J_e \tag{4}$$

$$x_{ij} \leq d_i y_j, \forall i \in I, j \in J_n \tag{5}$$

$$x_{ij} \leq d_i(1 - z_j), \forall i \in I, j \in J_e \tag{6}$$

$$\sum_{i\in I}\sum_{j\in J} c_{ij} x_{ij} \Big/ \sum_{i\in I} d_i \leq SL \tag{7}$$

$$x_{ij} \geq 0, \forall i \in I, j \in J \tag{8}$$

$$y_j \in \{0,1\}, \forall j \in J_n \tag{9}$$

$$z_j \in \{0,1\}, \forall j \in J_e \tag{10}$$

Equations: Model Formulation

The objective (1) of the problem is to minimize the weighted sum of facility costs (opening new facilities and removing old ones) and transportation costs. Note that the cost of facility is on government and the transportation cost is on residence/students. If we minimize only facility cost, the transportation cost of residence will increase, and vice versa. In fact, this is a multi-criteria optimization problem. We combine these two criteria using a weight (*Wt*) to allow adjustable emphasis on each criterion during different scenarios to address the multi-criteria optimization problems. We will change the weight to obtain the Pareto frontier.

The first constraint (2) requires that each customer's demand must be satisfied. The capacity of each facility j is limited by the constraints (3-4). The constraints (5-6) provide the variable upper bounds which are used to reduce computational time. Constraint (7) makes sure the average travel cost/time is less than a certain value which is determined according to current practice. In other words, in the future, the service level will be higher than current level. Lastly, constraints (8-10) are for variable definition.

The algorithm is shown below which includes data input and preparation, rules and model setup, and the return values for the planning interface.

Algorithm:

1. Initialize input data:
   a. Setup demand points parameters: m, d
   b. Setup facility points parameters: n, M
   c. Setup cost parameters: f, g, c
   d. Setup locations of demand points I, facilities J: $\{J_e, J_n\}$

2. Form up optimization model:
   a. Setup constraints related to (2-7)
   b. Setup optimization model based on (1-10)

3. Run model:
4. Return results:
   a. x_{ij} ($\forall x_{ij} \geq 0$), $\{y_i, z_i\}$ ($\forall\, y_j, z_j = 1$)
   b. location mappings of demand point i to facility location j

In addition to the above constraints set for the model optimization, there are other decision parameters to be considered:

- Preferences in co-location with other facilities
- Capacity size variation in comparison with nearby facilities
- Minimum utilization required to justify for operating the facilities
- Maximum overlapping coverage to minimize over servicing
- Behavioral changes in demand points like travelling mode, working locations, varying service time requirement

All these are part of the planning parameters in consideration when strategizing the placement of the facilities to be included in the plans. These considerations/factors can also be encoded into the model as constraints. One issue is the computation time when the model becomes more and more complex and has more constraints and decision variables.

In the following section, we will discuss the experiments that we have conducted to understand the needs of adding or removing facilities from the set of existing establishment.

IV. EXPERIMENTS AND RESULTS

The experiment is setup with a set of land parcel boundaries with each parcel quantified into a set of demographics properties. As a case study for this paper, the student population in each land parcel acts as the demand point, and the schools as the facilities. The demand points are representative centroids within each parcel with an associated housing and population density. Similarly, the facilities are representative centroids within each available parcel with an associated capacity and opened to service without boundary to the demand points.

The list of datasets and record tuples used to implement the FLP optimization models are as follows:

- Zone/boundary demography with distributed housing and population data
  tuple{zone_id, sub_zone_id, gis_polygon, size_area, centroid(x, y), housing_type, total_housing_units}
- Public transport connectivity and reach by time matrix
  tuple{zone_id_origin, zone_is_destination, transport_model:: time_travel}
- Road network connectivity and distance matrix
  tuple{zone_id_origin, zone_id_destination, road_network:: shortest_distance}
- Existing and projected facility (hospital, elder/child_care, schools) capacities
  tuple{zone_id, sub_zone_id, gis_polygon, size_area, centroid(x, y), facility_type, capacity, dismantle_cost, build_cost}
- Vacant land parcels for new facilities placement
  tuple{zone_id, sub_zone_id, gis_polygon, size_area, centroid(x, y), facility_type, capacity, fixed_cost}

The zones are bounded, for instance, by waterbodies, major roads and jurisdiction boundaries. There are approximately 1400 zones/boundaries, 200 existing schools, 3500 potential sites (each site with capacity between 1200-2000 students enrolment). We aggregated 3500 potential sites into about 900 sites as some sites are very close to each other. In this way, we reduce the problem size. Even after that, there are still more than 700,000 variables (most are continuous variables).

The proposed model has been implemented using Python and solved by the state-of-art open source solver SCIP [14]. The experiments are run on a notebook with Intel Core i7 CPU @2.60GHz.

We need to obtain a lot of parameters to run the optimization model. One important parameter is the travel cost/distance matrix c between demand points and facilities. We can estimate it in different ways: by real road network or by direct distance. We use the method of real road network. This method gives us a more accurate estimate in terms of true reachability by eliminating ground obstacles, such as waterbodies, walls/fences and highways. This elimination is not considered when using the direct distance.

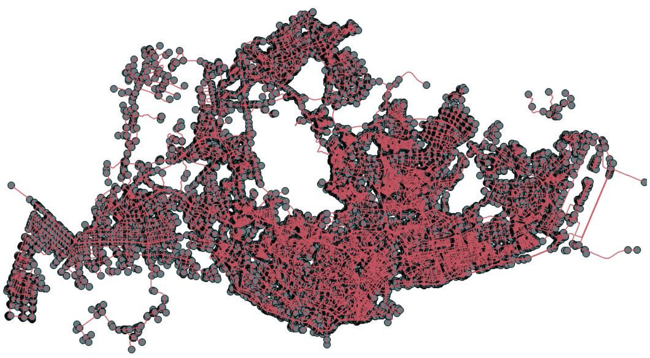


Fig. 1: Road Network Connectivity

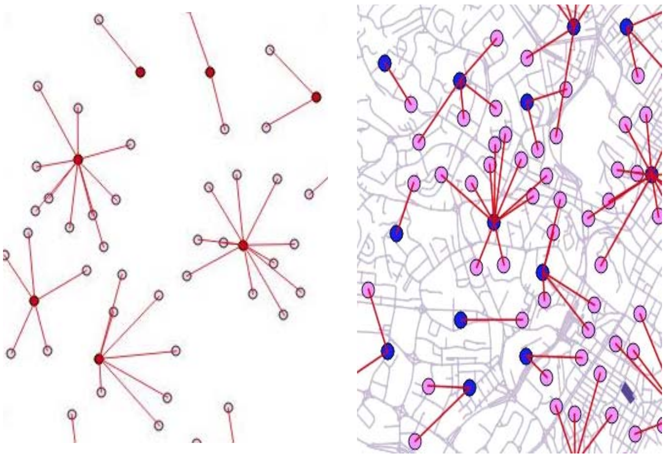
The actual road network in Singapore is shown in Fig. 1. The facilities and the demand points are mapped to the closest nodes in the network for the routing algorithm to compute the connectivity distances. From this figure, we can see that in some areas, the direct distance is not a good estimation for travel time (e.g. points across waterbodies depicted by white spaces in Fig. 1).

Fig 2: Central Area Mapping and Influence on Land Price

Another important parameter is the cost f of building a new facility. A very large part of this cost is the cost of land. In the city, the land costs at different areas are significantly different. The distribution of the influence from the central district area is shown in Fig 2. The fixed cost f from the annual land leasing is estimated taking reference from this influence. However, with better accessibility and the push for a distributed land use instead of a centralized area, the land lease cost at the peripherals has become more competitive in recent years. The cost of removing an existing facility is also affected by the land cost, as removing a facility means recalling land for other usage. So the cost of removing an existing facility can be a negative value, after considering other costs for dismantling buildings.

We need to obtain the service level (*SL*) of current practice as a benchmark for future facility planning. We use the optimization model above to obtain it, by setting the parameters with (i) no new facilities, (ii) no removal of existing facilities and (iii) the facility/school capacity as its current student enrolment. The optimization model will assign students to their nearest school and each school has the same number of students as its current practice. In this way, we can get the average travel time/distance/cost for each student. Then we use it as a standard service level for future facility planning.

The main purpose of this study is for planning facilities considering future population growth. We consider three scenarios: population increases by 10%, 20%, and 30% in the following years, assuming the whole population distribution pattern is almost stable. Of course, the optimization model can also address other scenarios. For each scenario, we vary the weight (*Wt*) for travel cost in the objective in the following levels: 0.5, 1, 2, 4, 6, 8 and 10. Then we run the optimization model for each case to obtain the optimal number of facilities (new and the remaining existing facilities) and their locations, travel costs, facility costs and service levels. Most cases required less than one hour to get the optimal solutions. The results are shown in Figures 3-6.



(a) Euclidean distance (b) Road network

Fig. 3: Assignment of Demand Points to Facilities

Fig. 3 shows part of results of assigning demand points to facilities based on different distance/time matrix *c*: (a) based on direct Euclidean distance where red dots are facilities; (b) based on the shortest route distance on the road network where blue dots are facilities. From the figure we can see that each demand point is assigned to its nearest facility. In the experiments that follow, we use the travel matrix based on the road network.

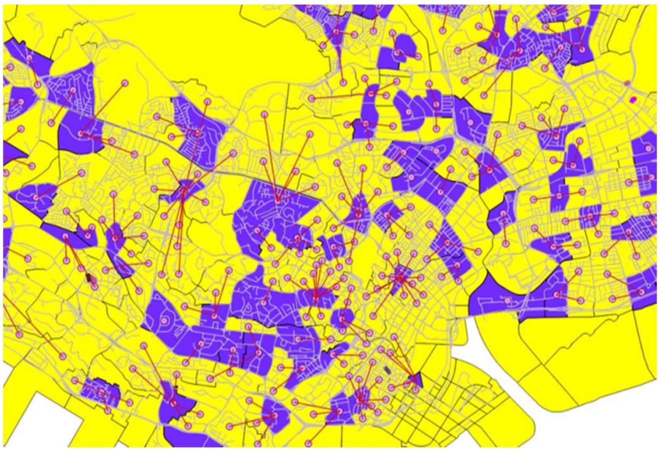


Fig. 4: Mapping to Potential Zones for Facilities

Fig. 4 shows the assignment of demand points to facilities on the background of road network and available land parcels (in blue) for new and existing facilities. Note that not all available land parcels will be used to build new facilities if the existing and planned new facilities can satisfy the demand.

Fig. 5 shows the number of new facilities for the three scenarios: population increasing by 10%, 20% and 30%. The results show that all existing facilities will not be removed in these cases. This is intuitive because when the population increases in an area, existing facilities should not be removed unless it is removed from an area with extremely high land price. From the figure, we can see that when we put more weight on the travel cost, the number of new facilities in general will increase so that the average travel cost will be reduced.

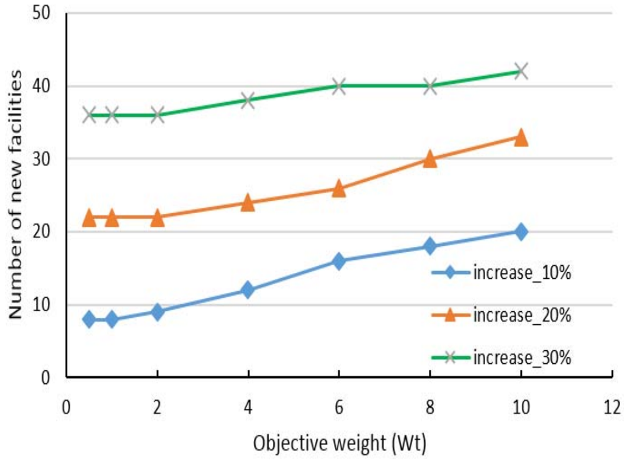


Fig 5. Number of New Facilities

Fig. 6 shows the *Pareto frontier* for the three scenarios. In the figure, the facility and travel costs are normalized. The travel costs (normalized) are the real travel costs divided by the travel cost of the benchmark (current practice). The facility costs (normalized) are the real facility costs divided by the facility cost of respective scenario when Wt=0.5. Note that for different scenarios, the facility costs "100%" have different real values.

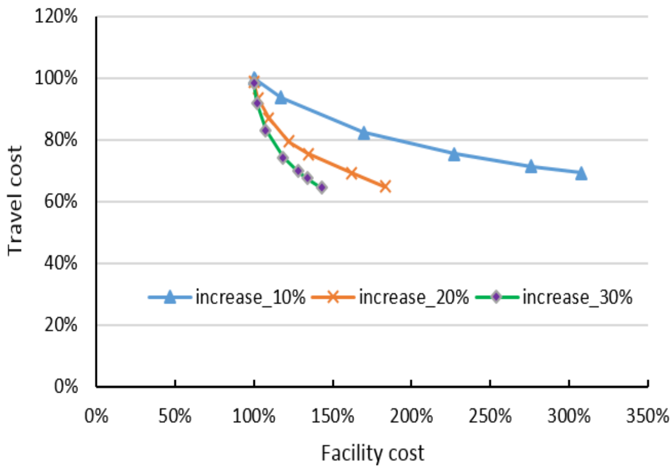
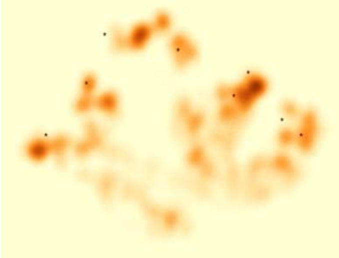


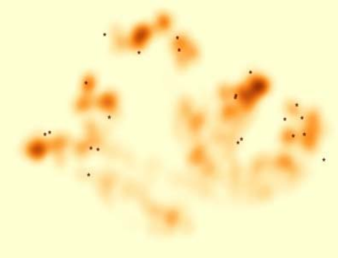
Fig 6. Pareto Frontier

From the figure, when Wt=0.5, facility costs of three scenarios (increasing population by 10%, 20% and 30%) are 100%, the travel cost are: 100%, 98.75%, 98.27%, respectively. So, the average travel distances are very close to the required service level (current practice).

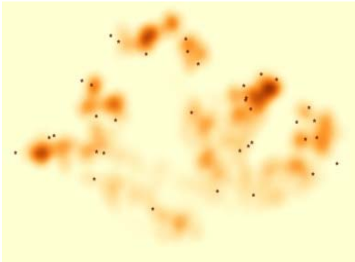
Also from the figure, when we increase the facility cost, the travel cost will decrease. This is intuitive as more and more facilities are built, the average travel distance will decrease. The curves of travel cost vs facility cost are convex. From the figure, we can also see the sensitivity of facility cost when we change the service level. The sensitivity is different for different scenarios, and this is useful for the city planner to understand the interlinking relationship in parameters variation.



(a) 10% population increase



(b) 20% population increase



(c) 30% population increase

Fig. 7: Selected Facility Locations vs %Population Increase

In Fig. 7, we mapped the selected locations for the new facilities (as red dots), generated from the model against the percentage of population increase. The facilities are clearly congregated around population with higher densities (population heat-map), taking into consideration their accessibilities via road network and available land parcels. An observation on the locations distribution shows that there are a lesser number of new provisions in denser area as compared to those in less dense area. This is largely due to the greater number of existing facilities in the denser area that can accommodate the initial population increase and that the denser area will have lesser available land parcels for placing these new facilities.

We can generate many other scenarios, such as changing population distribution pattern due to new town development. In a typical planning exercise, numerous scenarios are generated with all the feasible parameters, taking into considerations tangible and intangible benefits, and across each planning period for both short and long term future development. The planners may also consider other factors such as co-locating multiple types of facilities. In these cases, we may add more constraints for these practical requirements into the above core model.

One issue on developing a feasible operational system is to have a reasonable computation time for solving the optimization models. Considering the problems with a large and diversified number of constraints, in the future, we will use the commercial solvers such as CPLEX [15] to improve the speed of our model.

V. Conclusion

In this paper, we presented a platform for assisting city planning with new flexible optimization models which can address the addition and removal of existing facilities, multi-criteria decision, with and without capacity limits, and various travelling cost from (a) direct distance; (b) road network routing distance and (c) transport model travel time. The optimization model was ran against actual planning data including the residential population data, land use planning boundaries, existing facility locations, large number of potential facility sites, other planning projection data and approximated cost. The scenario settings are limited to parameters as close to the planning norms and as feasible as possible for actual ground development. Urban city planners can leverage on geo-analyzing the large amount of diverse and complex planning data to understand the dynamic growth and movement within the city. They can do so by setting different weighted planning parameters to extract scenario insights for understanding the interlinking relationships between these parameters. This enables them to map out the future land, infrastructure and amenities needs for different growth scenarios, and to draw up strategies and facility placement plans to guide future developments to maximize tangible and intangible land use benefits.

Further, there is some future work in the planning. We will develop other flexible models to address other realistic and practical requirements in city planning, such as co-locating different types of facilities with the possibility of phasing out existing facilities, and considering the impact from the demand uncertainty arising from new modes of transportation. Secondly, we will try to reduce the computation time by developing advanced algorithms such as customized ones for specific mixed integer programming models.

[1] Weber A. (edited by C. J. Friedrich), "Theory of Location of Industries", The University of Chicago Press, July 1929.

[2] Current J., Daskin M. and D. Schilling, "Discrete Network Location Models", chapter 3 in Facility Location Theory: Applications and Methods, Z. Drezner and H. Hamacher eds., Springer-Verlag, Berlin, pp. 81-118. 2002.

[3] Daskin M., Snyder L. and Berger R., "Facility Location in Supply Chain Design",Northwestern University, December 2003.

[4] Current J. and O'Kelly M., "Locating Emergency Warning Sirens", Decision Sciences, June 2007.

[5] Shmoys D., Tardos E. and Aardal K., "Approximation Algorithms for Facility Location Problems", In the Proceedings of the 29th Annual ACM Symposium on Theory of Computing, May 1997.

[6] Lin J. and Vitter J., "e-approximations with minimum packing constraints violation", In the Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 1992.

[7] Dohan D., Karp S. and Matejek B, "K-median Algorithms: Theory in Practise", Princeton University, Jan 2015.

[8] Hansen P., Miladenovic N. and Taillard E., "Heuristic Solution of the Multisource Weber Problem as a p-median problem", Operations Research Letters, March 1998.

[9] Arya V., Meyerson A., Munagala K, and Pandit V., "Local Search Heuristics for k-median and Facility Location Problems", In the Proceedings of the 33rd Annual ACM Symposium on Theory of Computing July 2001.

[10] Elkady S. and Abdelsalam H., "A Simulation-based Optimization Approach for Healthcare Facility Location Allocation Decision", Science and Information Conference 2015.

[11] Papadimitriou D., Colle D. and Demeester P., "Mixed-Integer Optimization for the Combined Capacitated Facility Location-Routing Problem", The 12th International Conference on the Design of Reliable Communication Networks 2016.

[12] Liu J., Li Q., Qu M., Chen W., Yang J., Xiong H., Zhong H. and Fu Y., "Station Site Optimization in Bike Sharing Systems", International Conference on Data Mining 2015.

[13] Quilliot A. and Sarbinowski A., "Facility Location Models for Vehicle Sharing Systems", Proceedingsof the Federated Conference on Computer Science and Information Systems 2016.

[14] Gleixner A., Eifler L., Gally T., Gamrath G., Gemander P., Gottwald R. L., et al. The SCIP Optimization Suite 5.0, Tech Report, *ZIB,* 2017.

[15] CPLEX 12.6 Manuals, IBM Inc., 2015.