A Data-Driven Approach for Automated Multi-Site Competitive Facility Location

First Author[a,*,1]**, Second Author**[b,1] **and Third Author**[b,c]

[a]Short Affiliation of First Author
[b]Short Affiliation of Second Author and Third Author
[c]Short Alternate Affiliation of Third Author
ORCID (First Author): https://orcid.org/....-....-....-...., ORCID (Second Author): https://orcid.org/....-....-....-....,
ORCID (Third Author): https://orcid.org/....-....-....-....

Abstract. This paper presents a data-driven approach for solving the Competition Facility Location (CFL) problem, in which a retailer wants to open new stores in a market where it has existing friendly and competitor stores. The goal is to select locations for new stores such that the overall potential consumers captured by the retailer is maximised. We apply Adaptive Large Neighbourhood Search (ALNS) with data enrichment techniques that includes community detection on road networks.Our contribution is the ability to handle the selection of a large number of candidate sites which is often limited by traditional facility location methods due to computational constraints and reliance on manual pre-selection. Conversely, our approach integrates the identification of urban population centres as candidate sites to efficiently process and consider an extensive number of candidate sites. We benchmarked against ArcGIS, a widely used commercial software for CFL problems. Our approach achieve an average of 2% improvement in captured consumer count when compared against ArcGIS, demonstrating ALNS effectiveness and potential in solving CFL problems.

1 Introduction

Facility location has a broad range of applications in the field of operations research problems. The more common location problem typically assumes only a single decision-maker without other competitors. Such a market is known as monopolistic where a single firm captures the entire market. This is applicable to public sector scenarios where the government opts to place public facilities such as hospitals or schools. In the private sector where location problems involve supermarkets or restaurants, market competition needs to be considered using Competitive Facility Location (CFL) models. Competitive facility location models address key concerns of decision-makers in a competitive market landscape. These models attempt to address questions of how businesses can discern optimal locations amidst intense competition and how the introduction of a new store might reshape the competitive dynamics. CFL can be addressed through exact methods such as mathematical programming formulations. Commonly used techniques include linear programming and mixed-integer programming. Metaheuristic algorithms such as Adaptive Large Neighbourhood Search (ALNS), Par-

ticle Swarm Optimisation (PSO) and Genetic Algorithms (GA) can be used to iteratively explore and improve solutions. In certain scenarios, Game Theory and Nash equilibrium are also used to achieve a steady state where no single participants can independently enhance their outcome.

Our data-driven approach offers a novel alternative to traditional methods that typically rely on manual preselection and mathematical optimization. Our contributions are as follows:

1. Elimination of manual human preselection and bias in the site identification process. Our method increases the efficiency of locating numerous facilities in multiple cities or expansive rural areas.
2. Scalability and generalizablity of approach for diverse use cases. By applying ALNS, our method allows for flexibility in defining the neighbour search space by using different destroy and repair strategies appropriate for different facilities.

We benchmark our approach against ArcGIS and obtain an average improvement of 2% in captured consumer count over ArcGIS. The effectiveness of the data-driven approach based on real-world observations of road network and mobility data suggests possible reapplication for different facility types, such as supermarkets, bike sharing stations.

2 Related Work

Static Foresight Dynamic -> changing demand but its about competitors moving away

Existing CPLs - what are the search algorithms - why are they not good?

2.1 Competitive Facility Location

In a survey paper by [**?**], CFL problems can be categorized based on type of competition behaviour, problem domain and search space.

Competitive facility location addresses the placement of facilities in a competitive environment that has numerous applications across different industries. These can include retail stores, warehouses, delivery hubs and even telecom towers. The primary objective is to optimise certain objective function that can incorporate elements such

---

as transportation costs, customer reach, demand and supply. Accurate demand forecasting enables a more realistic representation of the real world and is critical to optimising the facilities placement.

2.1.1 CFL Competition Behaviour

CFL problems can broadly be classified into 3 main approaches based on competitor behaviour - static, foresight and dynamic model.

- Static models ([?]) assume that competitors do not readily react to the entrance of new competition. Static models hold a short term view whereby the response time taken for competitors to react is long enough for the new facility to reap the benefits.
- Foresight models ([?]) look at locating facilities with the foresight of where a competitor will subsequently make an entry. This approach tries to ensure that the remaining market share is maximised even after the competitor reacts.
- Dynamic models allow the relocation of competitor facilities based on changing market demands. Such scenarios are commonly seen in disaster relief where the location and allocation ([?]) of search and rescue (SAR) assets are continually optimised to enhance the performance of SAR missions.

This paper focuses on static model, considering that the opening and closing of retail stores incur significant resources. Existing stores are unlikely to react to new competition in the short term.

2.1.2 CFL Problem Domain

Problem domains are spread across various industries, including telecommunications, supply chain management, electric vehicle charging infrastructure, and public facility provisioning. Some of the problem domains include the following:

1. Telecommunications: One study ([?]) proposes using Genetic Algorithm to find a suitable solution which optimises the placement of mobile antennas to maximise the network coverage while balancing the servicing costs.
2. Supply chain: Stackelberg game theory ([?]) is used to model the relationship between manufacturer and distribution center, distribution center and customer region, and customer region and retrieval center.
3. Charging Stations for electic vehicles: A multi-objective Particle Swarm Optimisation ([?] is used to consider population density, land cost and the site selection for recharging station and battery exchange station for e-scooter.
4. Public Facilities: Particle swarm optimisation (PSO) and artificial bee colony (ABC) to locate fire stations and resources ([?]) in different periods and emergency situations to account for dynamic needs.

2.1.3 Decision Space: Continuous vs Discrete Decision Space

The decision space in which facilities are located can be classified as either continuous location models, network location models and discrete location models. Discrete location models restrict facilities to specific points in the decision space and are particularly suited for scenarios where the initial possible locations are pre-selected. However, the pre-selection of these discrete points can be time consuming if the decision space is large.

In network location models, the facilities can be located on network nodes such as those of a traffic road network. One example ([?]) is the assignment problem for traffic police routine patrol vehicle on road network to optimise the response time to emergencies.

Continuous location models can be used in scenarios where facilities provide similar services, e.g., supermarkets and warehouses. Facility sites can potentially be sited anywhere within the decision space, without the constraints of predefined points. However, solving in continuous space is very computational intensive ([?]). Several heuristics have been developed to breakdown continuous space into discrete spaces ([?]). This includes the identification of hotspots and geographic tessellations.

In this paper, we aim to automate the site selection process by converting the continuous decision space into discrete decision space using community detection of road network ([?]). Each community detected then becomes a possible option for a retail store location.

2.2 Metaheuristics and Applications to CFL

Adaptive Large Neighbourhood Search (ALNS) was first proposed in 2006 ([?]). The ALNS framework is based on the extension of Large Neighbourhood Search ([?]) which deploys 2 kinds of operators - destroy and repair. An existing solution is first destroyed by destroy operators and subsequently repaired by repair operators. The performance of each operator is monitored and the selection of operators is based on their prior performance. Earlier works of ALNS focused on transportation optimisation problems such as pick-up and delivery vehicle routing problem. Later works of ALNS include domains such as health care ([?]) and job shop scheduling problem ([?]).

ALNS has been increasing in popularity as an optimisation framework as it provides the following advantages ([?]):

- ALNS can be applied in different domains.
- Ease of incorporating existing domain knowledge by experts into destroy and repair operators.
- Destroy and repair operators can be freely incorporated into the algorithm as ineffective operators will be less frequently executed.
- ALNS generates multiple neighbourhoods to increase the probability of finding the global optimal and avoiding being trapped in local optimal.

In a 2022 survey paper which reviewed ALNS applications ([?]) in 252 papers, 186 papers focused on routing, 141 papers focused on scheduling problems and 22 papers were related to location analysis. Of the 22 papers on location analysis, they focused on maximal coverage problems ([?]) and allocation of electronic vehicle charging stations ([?]). There are no papers which applies ALNS on CFL.

3 Problem Definition

How to find places where we maximise patronage with considerations of Competitor and Friendly Sites? How does "machine" able to consider more candidate site compared to human?

3.1 Huff Model

The Huff model ?] is a widely used model in retail location analysis since its introduction in 1963. It is a probabilistic model to estimate the consumer's patronage probability of a particular store, by comparing the attractiveness of the store to its nearby competitors. The

attractiveness of the store is usually defined as the floor size or actual sales.

The patronage probability, i.e., the probability that a customer located in community i will choose the store j can be expressed as:

$$P_{i,j} = \frac{A_j^\alpha D_{i,j}^{-\beta}}{\sum_{k=1}^{K} A_k^\alpha D_{i,k}^{-\beta}} \tag{1}$$

where:

- $P_{i,j}$ is the patronage probability that a customer located in community i will choose the store j.
- A_j is the measure of attractiveness of store j
- $D_{i,j}$ is the distance from community i to store j.
- α is the attractiveness parameter
- β is the distance decay parameter that reflects the rate at which the probability of choosing a store decreases as the distance between the customer and the store increases.
- K is the total number of stores, including store j

3.2 Formulation

In this paper, we consider a franchise with existing stores (i.e., friendly stores) which aims to maximize its profit by opening new stores in the presence of competitor stores. Both friendly and new stores are denoted by set J while competitor stores are denoted by set K. we define

$$v_i = \sum_{k=1}^{K} A_k / D_{ik}^{-\beta}, \quad \forall i \in I \tag{2}$$

As shown in Equation **??**, v_i is the utility provided by competitor stores at community i where D_{ik} is the distance between community i and competitor store k. The possible attractiveness levels, denoted by A, are presumed to be the same as all stores share similar floor plan and serve similar profile of consumers.

$$P_{i,j} = \frac{A_j^\alpha D_{i,j}^{-\beta}}{\sum_{j=1}^{J} (A_j^\alpha D_{i,j}^{-\beta}) + v_i} \tag{3}$$

In the presence of competitors, the patronage probability, P_{ij}, of customers at community i patronizing store j is then given by Equation **??**

The objective function **??** aims to maximize the total demand satisfied by the new friendly sites as well as existing friendly sites.

$$\max \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{A_j^\alpha D_{i,j}^{-\beta} \cdot x_j}{\sum_{j=1}^{J} (A_j^\alpha D_{i,j}^{-\beta}) + v_i} \cdot \text{Pop}_i \tag{4}$$

Where:

- I is the set of communities
- J is the set of friendly sites including new J_{new} and existing J_{existing} sites.
- K is the set of competitor sites
- $P_{i,j}$ is the probability that a customer from community i will visit the facility at friendly site j
- v_i is the utility provided by competitor stores at community i
- Pop_i is the population of community i
- x_j is the binary decision variable indicating whether a friendly facility is located at candidate site j (1 if located, 0 otherwise). For existing friendly sites, it will always be 1.

- D_{ij} is the distance between community i and friendly site j
- A_j is the attractiveness of friendly site j
- α and β are parameters in the Huff model

By multiplying the patronage probability with the population, Pop_i, in the community, we will be able to obtain the absolute amount demand, γ, in community i satisfied by store j.

$$\gamma_{i,j} = P_{i,j} \cdot Pop_i \tag{5}$$

To obtain the total demand satisfied by store j, we sum up all the demand, γ, across all communities assigned to store j.

$$\gamma_j = \sum_{i=1}^{I} P_{i,j} \cdot Pop_i \quad \forall j \in 1, 2, \ldots, J \tag{6}$$

To obtain the total demand satisfied by the franchise $1, ..., J$, we further sum up all the demand served across all stores in the franchise.

$$\gamma_{franchise} = \sum_{i=1}^{I} \sum_{j=1}^{J} P_{i,j} \cdot Pop_i \tag{7}$$

3.2.1 Constraint

The formulation for the constraints are as follows:

$$x_j \in 0, 1 \quad \forall j \tag{8}$$

This constraint ensures that the decision variable x_j can only take binary values of either 0 or 1 for all candidate sites j. It restricts the decision variables to be either 0 (no facility located at site j) or 1 (a facility is located at site j). Each community j can have at most 1 facility.

$$\min_{j \in J \cup K} D_{i,j} \geq D_{\min} \tag{9}$$

The distance between any community i and the nearest facility (either friendly or competitor) should not be below a minimum threshold $D_{\min}$. It is set as 100m to avoid instances of division by zero in the Huff model, which would result in infinite patronage probabilities.

$$\sum_{j \in J} x_j = J_{\text{new}} + J_{\text{existing}} \tag{10}$$

Sum of decision variables x_j across all potential store locations j from 1 to J must be equal to the number of stores to be opened plus existing stores

$$\sum_{j=1}^{J} P_{i,j} \cdot \text{Pop}_i = \text{Pop}_i \quad \forall i \in 1, \ldots, I \tag{11}$$

Total demand satisfied by all stores j must be equivalent to total population within each community i

4 Contribution

Not to rely on outdated city census data Domain expert view (pre-selection) not possible for large scale comparison and large n since it is continuous, it is not possible to find by manual methods (scalable) and need a smart way of approximation many applications

5 Case Study

The area of study shown in Figure **??** is the Metropolis Bangkok with 2 anonymised retailer chains A and B which have a significant presence in the city. Bangkok is chosen as it is a major Asian city which has enjoyed continual growth in recent years. The urban area experienced a recent notable population increase in recent years driving

Urbanisation in Thailand is concentrated within Bangkok which serves as the nation's economic, political and cultural hub. In the process, it has attracted much of the population and economic activity within the city, resulting in a core region with office skycrappers, shipping districts, luxury condominiums and cultural landmarks. Internal migration from rural areas into Bangkok continues today, latest estimate from United Nations shows that the city is growing between 1.5-2% in the recent 5 years. Thailand's National Statistical Office estimated a steady population increase in Bangkok, growing from 8 million in 2010 to 11 million in 2020.

Transportation infrastructure helps to connect the populace with different parts of the city but navigation within Bangkok's intricate urban network is challenging and complex. There has been efforts to alleviate traffic jams with innovative transportation solutions such as water taxis along the major river arteries coupled with mass rapid transit network such as the BTS Skytrain. But movement within the city continues to be time-consuming.

Land use allocation is essential to manage limited land resources to balance between competing land needs versus that of the burgeoning population growth. For cities with transportation difficulties like Bangkok, strategically locating amenities to increase accessibility is especially important. It reduces the need for extensive travelling and increase productivity at work and home. This contributes to a more efficient and sustainable urban environment in the future.

6 Proposed Approach

6.1 ALNS

Large scale CFL problems can be computationally expensive and time-consuming. Solution time grows exponentially as the problem size increases, e.g., addition of facilities, demand points etc. Hence exact methods may be impractical to solve real-world problems which require extensive data.

This paper uses ALNS to solve large scale CFL problems. ALNS iteratively destroys and repair parts of the solution to maximise the total demand served by the franchise, $\gamma_{franchise}$. Destroy operators will remove stores from the list of chosen candidate sites while the repair operators will open stores from the un-chosen list of candidate site. These operators compete with each other in each iteration and the operators with better performance are more likely to be used in future iterations.

6.2 Initialisation

Initialisation: The initial selection of new stores is based on the communities with the highest population.

6.3 Destroy Methods

Destroy operators remove parts of the current solution by identifying the stores to be eliminated based on the criteria stipulated within each operator. The destroy operation creates variation in in solution neighbourhoods and allows the subsequent repair operators to explore different parts of the search space. The destroy operators are as follows.

6.3.1 Destroy 1 - Random Destroy

Random destroy of N percentage of stores in a bid to minimise the risk of local optimal.

6.3.2 Destroy 2 - Remove Least Market Share

Remove store with the lowest market share. While computing the objective function for each iteration within ALNS, the market share of each store in the study will be computed, providing insights on which stores are likely to have low market share. Instead of taking a deterministic approach of removing the bottom n stores which may result in a cycle, we implement a probabilistic approach. Stores with lower market share will have higher probability of being removed. .

6.3.3 Destroy 3 - Remove Closest Store

The pair-wise distance between stores will be computed and each store-pair will be ranked based on the distance. Those with the shortest pair-wise distance will have a higher probability of being removed. One store from each of the selected pair will then be removed randomly.

6.4 Repair Methods

Repair operators restores the destroyed solution to create a new candidate solution. By reconstructing feasible solutions, the solution can be incrementally improved. The repair operators are as follows.

6.4.1 Repair 1 - Random Repair

Random Repair. Store locations will be selected randomly.

6.4.2 Repair 2 - Locate in Highly Populated Areas

Select areas which contains high population within an increasing probability weighted by population. Fig **??** shows communities which are darker with higher population.

6.4.3 Repair 3 - Locate in areas with high road network connectivity

Select areas with high network centrality. It is postulated that areas which are easily accessible will be attractive to customers. Closeness centrality of a road network provides insights into the accessibility from different areas. Nodes representing intersections with high closeness centrality are locations strategically positioned to efficiently reach other parts of the network, allowing for smoother traffic flow and easy navigation. When considering site selection, prioritizing those with high centrality will ensure better accessibility, creating a more connected and well-functioning urban environment. Fig **??** shows communities which are darker with highest closeness centrality. Central Bangkok has the highest closeness centrality.

7 Experimental Result

A dataset of 8 competitor stores and 8 friendly stores have been instantiated. 486 candidate sites are available for the selection of new stores. They are created using a community detection algorithm outlined by [? ?]. This algorithm utilizes the Louvain community detection algorithm on a road network to create communities of various sizes, mimicking the actual urban structure of the community.

The selection of new store locations is done in intervals of 10, ranging from 10 to 120. The same settings are applied to both ArcGIS and ALNS. These settings include a maximum consumer travel distance of 3km and a β value of 0.3. The store attractiveness values are set to 100, indicating equal attractiveness across all stores due to similar floor size and store layout. For ALNS, an iteration of 20 is performed, and the best state objective and store location are selected. ArcGIS has a maximum limit of 99 stores in a CFL analysis. In the case of ALNS, instead of randomly placing stores among the 486 candidate sites, the initial state is initialized by placing new stores in communities with the highest population count. ArcGIS produced 9 instances from 10 to 90 new stores, while ALNS produced 12 instances from 10 to 120 new stores.

To estimate the consumer count for each store, the Huff Model is utilized. This involves multiplying the patronage probability $\P_{ij}$ with the population count Pop_i of each community. The consumer count is then summed up for all new stores and existing friendly stores to obtain the total consumer count captured.

The results are depicted in Figure **??**, which showcases the comparison of consumer count between ALNS and ArcGIS. The figure demonstrates similar performance between the two methods, with ALNS continuing to recommend new store openings beyond the upper limit of 99 imposed by ArcGIS.

ifference in consumer count between ALNS and ArcGIS. On average, ALNS performs slightly better than ArcGIS with a 2% improvement. The instances generally show performance improvements within the range of 0-2

Figure **??** illustrates the evolution of the objective value in ALNS for a sample instance of 100 new store openings. The chart displays two lines: the Best Line and the Current Line. The Best Line represents the best objective value found thus far during the search process, and it shows the improvement in the objective value over time. The Current Line represents the objective value of the solution currently being evaluated or explored. By analyzing the gap between the best and current lines, we can assess the convergence and behavior of the algorithm.

Figure **??** provides insights into the performance of the repair and destroy operators in the instance of 100 new store openings. The most frequently chosen destroy operator is the least market share, which has a higher probability of removing stores with smaller consumer counts. On the other hand, the most frequently chosen repair operator is the top community population repair, which has a higher probability of opening stores in communities with higher populations.

These findings contribute to a better understanding of the performance, convergence, and operator behavior in the ALNS approach for new store location selection.

8 Conclusion

We have expanded the applications of ALNS on location analysis which are predominantly concentrated on maximal coverage. Our work pioneers the incorporation of competition in location allocation into the ALNS framework. ALNS has provided a framework which can easily incorporate additional layers of information that will contribute to high quality solutions for store allocations. For example, the availability of road network data allows for data enrichment for CFL.

The automation of site selection involves the integration of various factors, including the competition landscape, road network data and population data. This automation produces a more systematic and data-driven approach to selecting optimal store locations without any need for manual pre-selection of sites, reducing the potential for human bias and subjective judgment. We have also reduced the computational complexity through conversion of continuous to discrete decision spaces through community detection.

However, we acknowledge that the solution quality depends heavily on the quality of the data inputs. It is essential to continue refining the data collection processes to ensure that road network data and other relevant data sources are up-to-date and accurate. Other potential data sources can be changes land use data derived from satellite imagery which allows for near-real time analysis.