Walkability Optimization: Formulations, Algorithms, and a Case Study of Toronto

Weimin Huang, Elias B. Khalil

Department of Mechanical & Industrial Engineering, University of Toronto
Scale AI Research Chair in Data-Driven Algorithms for Modern Supply Chains
cheryl.huang@mail.utoronto.ca, khalil@mie.utoronto.ca

Abstract

The concept of walkable urban development has gained increased attention due to its public health, economic, and environmental sustainability benefits. Unfortunately, land zoning and historic under-investment have resulted in spatial inequality in walkability and social inequality among residents. We tackle the problem of Walkability Optimization through the lens of combinatorial optimization. The task is to select locations in which additional amenities (e.g., grocery stores, schools, restaurants) can be allocated to improve resident access via walking while taking into account existing amenities and providing multiple options (e.g., for restaurants). To this end, we derive Mixed-Integer Linear Programming (MILP) and Constraint Programming (CP) models. Moreover, we show that the problem's objective function is submodular in special cases, which motivates an efficient greedy heuristic. We conduct a case study on 31 underserved neighborhoods in the City of Toronto, Canada. MILP finds the best solutions in most scenarios but does not scale well with network size. The greedy algorithm scales well and finds high-quality solutions. Our empirical evaluation shows that neighbourhoods with low walkability have a great potential for transformation into pedestrian-friendly neighbourhoods by strategically placing new amenities. Allocating 3 additional grocery stores, schools, and restaurants can improve the "WalkScore" by more than 50 points (on a scale of 100) for 4 neighbourhoods and reduce the walking distances to amenities for 75% of all residential locations to 10 minutes for all amenity types. Our code and paper appendix are available at https://github.com/khalil-research/walkability.

1 Introduction

The concept of walkability in urban planning has gained increased attention as research has shown that good walkability contributes to physical health, economic development, and environmental sustainability (Zapata-Diomedi et al. 2019). Highly walkable neighbourhoods allow for residents to easily access amenities such as retail and food in the vicinity, giving rise to the concept of the "15-minute city" (Whittle 2020). However, the zoning regulations introduced in the early 20th century that separate industrial, commercial, and residential areas have prevented walkable development and contributed to automobile-reliant communities (Levine 2005; Fischel 2003) – residents have to travel outside their local communities in order to meet their daily needs. Moreover, historic disinvestment in segregated neighbourhoods including low-income groups and racial minorities has led to spatial inequality in walkability. Evidence shows that disadvantaged groups live in neighbourhoods with less accessible physical infrastructure and services (Massey 1990). With the COVID-19 pandemic reshaping the relationship between cities and the quality of life, addressing inequities and improving access to services and amenities such as healthcare and green spaces for vulnerable groups has also accelerated the need for walkable development (Mouratidis 2021).

Improving walkability is also a powerful way to reduce greenhouse gas emissions (GHG) and tackle climate change. Dense and walkable neighbourhoods encourage active transport (walking, cycling), thus reducing automobile dependence (McIntosh et al. 2014; Brand et al. 2021). Research shows that technological measures (e.g., increasing the use of electric vehicles) alone will not be sufficient in reducing GHG, whereas a shift to a more sustainable mode of transportation can result in a quicker and more significant reduction of emissions from vehicles, particularly in urban areas (Creutzig et al. 2018; Neves and Brand 2019). Shifting from motorized transport to active transport is considered one of the most promising ways to reduce GHG. A study shows that active walking as a lifestyle change of residents can significantly reduce emissions related to private vehicles, even in European cities that are already highly walkable (Brand et al. 2021). Building inclusive, safe, resilient, and sustainable cities has been highlighted in the Sustainable Development Goal 11 of the United Nations.

Towards this goal, researchers have been interested in improving walkability in cities by converting certain underused spaces into easily accessible amenities. For instance, some urban planning research relies on simulation and inspection, such as converting high-density regions into amenities and straightening busy routes (Yang, Samaranayake, and Dogan 2020). On the other hand, one common optimization-based approach is to use a Genetic Algorithm (GA) that encodes the potential allocation locations (e.g., empty lots where a grocery store can be built) as fixed-sized vectors and iteratively generate child solutions from a population of candidate solutions (Cichocka 2015; Rakha and Reinhart 2012; Sonta and Jain 2019; Nagy, Villaggi, and Benjamin 2018;

Indraprastha 2019). However, genetic algorithms lack convergence guarantees and tend to lead to suboptimal solutions (Kim, Kim, and Oh 1997). Moreover, GAs do not handle constraints directly (Yang 2020), making them less applicable in the context of modern city planning, given the existing city layout and property ownership. Despite this interest in walkability optimization, there have not been robust and efficient methods for this problem.

Contributions: To our knowledge, the problem of *Walkability Optimization* has not yet been examined from an algorithmic perspective, particularly one that considers the problem's realistic aspects. In this paper, we contribute to this question along multiple axes:

1. We formulate *Walkability Optimization* as the combinatorial optimization problem of selecting the locations of new amenities to maximally improve residents' access to basic necessities (Section 2). Our formulation extends the standard facility location problem and models residents' behaviour realistically – we consider multiple facility types, multiple potential choices for the same type, and an objective function with respect to the travel distances that represent the proximity to residents. Also, compared to existing work on Walkability Optimization, our formulation takes into account existing amenities rather than designing a layout from scratch and is flexible in capturing additional constraints on allocation.
2. We analyze the complexity of the problem, showing that (a) its decision version is NP-Complete in general and that (b) the objective function is submodular in special cases (Section 3). The latter property motivates an efficient greedy algorithm presented in Section 4.
3. We derive Mixed-Integer Linear Programming (MILP) and Constraint Programming (CP) models (Section 4).
4. We perform a case study of 31 underserved neighbourhoods identified by the City of Toronto, Canada (Section 5). In most neighborhoods, significant reductions in walking distances can be obtained by optimizing the placement of several new amenities. MILP outperforms CP and the simple greedy algorithm achieves a good tradeoff between running time and solution quality.

2 Problem Formulation

At a high level, the Walkability Optimization problem (WALKOPT) is defined as follows. Given a set N of residential locations, a set M of candidate allocation locations, and a set L of existing amenities, we seek a set of locations where new amenities of different types A are allocated so that the average "Walkability Score" f_i for residential locations $i \in N$ is maximized. f_i is a function of walking distances and can be interpreted as the proximity of A to residents. The maximum number of instances to be allocated for each amenity type $a \in A$ is denoted as k_a. The set of locations M, N, and L correspond to nodes on a network where edges represent the walkable paths in the neighbourhoods. The walking distances between these locations are the shortest-path distances, which we denote as d_{ij} for $i \in N, j \in M \cup L$. Note that our formulation improves

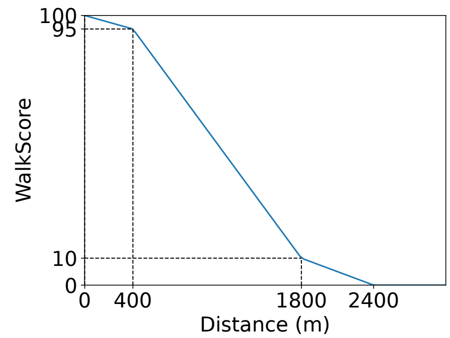


Figure 1: Distance (meters) v.s. WalkScore.

walkability by introducing amenities instead of other potential approaches such as improving network connectivity by adding edges. The latter is a challenging intervention since road networks are resistant to change compared to the speed of city expansion (Rakha and Reinhart 2012).

Building Blocks

Walkability Score To quantify walkability, we adopt the WalkScore methodology (Walk Score 2011), used in prior quantitative analyses (Duncan et al. 2011; Verbas et al. 2015), that assigns a score f_i for each residential location i based on the weighted walking distances, denoted as l_i, from i to different amenity types. The WalkScore assigns a near full score (100) for distances within 400m with scores decreasing with respect to l_i after 400m. Distances above 2400m (about a 30-minute walk) are not rewarded any points. The WalkScore f_i is originally a smooth function that models this nonlinear decay with respect to l_i; see Fig. 2 (left) in (Reyer et al. 2014). For computational purposes, nonlinear functions can be approximated with Piecewise Linear functions (PWLs) (Ngueveu 2019): f_i is represented as a PWL that is parameterized by breakpoints $\bar{t}$ (Fig. 1). The parameters $\bar{t}$ are shown in Appendix E.

Weighted Walking Distances The weighted walking distance l_i is a linear combination of distances to multiple amenity types $a \in A$. Each type is given a weight w_a based on its level of necessity, e.g., grocery stores are given the highest weight as they are the most frequent walking destination (Lee and Moudon 2006). For most amenity types, we assume that residents at location i will walk for $D_{i,a}$ meters to the nearest instance of the amenity a; we denote amenities of this type with A^{plain}. For amenity types for which variety and options are important (e.g., restaurants, coffee shops), we account for the resident's "depth of choice" by considering a weighted combination of the distances to the top-r nearest instances (e.g., $r = 10$ for restaurants): the p^{th} nearest instance (e.g., of a restaurant) has distance $D^p_{i,a}$ and weight w^p_a. We denote such amenities with A^{depth}. We have $A^{plain} \cup A^{depth} = A$. Finally, l_i is expressed as

$$l_i = \sum_{a \in A^{plain}} w_a D_{i,a} + \sum_{a \in A^{depth}} \sum_{p \in P_a} w^p_a D^p_{i,a}, \forall i \in N, \quad (1)$$

where P_a is the set of top-r nearest instances of amenity type $a \in A^{depth}$. Note that not all r options in P_a may be available in the network, as the sum of the number of existing and allocated instances may be smaller than r. For example, there may be 2 existing restaurants and a budget of 3 additional restaurants to be allocated, but the number of choices being considered is $10 > 3 + 2$. The set of *available* choices of type a is denoted as P_a^Y, where $|P_a^Y| = min(k_a + |L_a|, |P_a|)$. The set of choices that we consider but are not available is denoted as $P_a^N = P_a \setminus P_a^Y$. Choices $p \in P_a^N$ have distances $D_{i,a}^p = D^\infty$, where $D^\infty = 2400$m (i.e., WalkScore is zero), so that a neighbourhood with no amenity has $f_i = 0$.

Objective Function We maximize the average WalkScore across all residential locations

$$F = \frac{1}{|N|} \sum_{i \in N} f_i. \tag{2}$$

f_i is a PWL function of l_i parameterized by breakpoints $\bar{t}$:

$$f_i = \text{PiecewiseLinear}(l_i, \bar{t}), \forall i \in N. \tag{3}$$

Existing and Candidate Amenity Locations We introduce new amenities to built neighbourhoods with consideration of existing amenities, instead of allocating from scratch. The set of locations with existing instances of amenity type a is denoted as L_a. For candidate allocation locations M, we set a capacity on the number of amenity instances that each candidate location j can accommodate (denoted as c_j, $j \in M$) to represent potential physical constraints. Given the context of built cities, M can be underused spaces such as parking lots, as freeing up parking spaces and converting them into amenities has been recognized as an urban renewal model that reduces air pollution and improves the quality of life (Natalie Marchant 2020; Eric Reguly 2020). However, our framework can accommodate various types of locations.

3 Theoretical Properties

Computational Complexity

We prove that WALKOPT is computationally hard even with a single amenity type and without depth of choice. This is achieved through a reduction from the widely studied and NP-Complete k-median problem to the decision version of WALKOPT. The proof is deferred to Appendix A.

Submodularity

Seeking a polynomial-time approximation algorithm for WALKOPT, we analyzed its objective function for submodular structure. Formally, a set function $F: 2^V \to \mathbb{R}$ is submodular if it satisfies the *diminishing returns* property: for every $S \subseteq T \subseteq V$ and $e \in V \setminus T$ it holds that $\Delta_F(e|S) \geq \Delta_F(e|T)$, where $\Delta_F(e|S) := F(S \cup \{e\}) - F(S)$ is the *discrete derivative* of F at S w.r.t. e.

We show that the objective (2) is indeed submodular when depth of choice is not considered. This motivates the use of a greedy algorithm as a $(1 - \frac{1}{e})$-approximation (Nemhauser, Wolsey, and Fisher 1978) when there is a single amenity type and no depth of choice, and as a heuristic (with no guarantees) otherwise. We also show that submodularity does not hold when considering depth of choice.

Submodularity in the SingleChoice Case

Theorem 1. *Objective* (2) *is submodular when* $A^{depth} = \emptyset$.

Proof. First, we represent a solution as a set of actions S where each element $e = (a, m) \in S$ consists in allocating an instance of amenity type a to a candidate location m. Note that it is feasible to have more than one instance of type a allocated to the same location, which can introduce identical (a, m) pairs. Since sets cannot contain duplicates, we construct an equivalent set $\bar{M}$ by duplicating each node $j \in M$ for c_j times. Then, the ground set is $V = \{(a, m) : a \in A^{plain}, m \in \bar{M}\}$ and objective (2) is a set function, $F: 2^V \to \mathbb{R}$. Let S and T be solution sets such that $S \subseteq T \subseteq V$, and let $e = (a', m') \in V \setminus T$. We show that $\Delta_F(e|S) \geq \Delta_F(e|T)$.

We denote the weighted walking distances at i under solution set S as l_i^S and express $\Delta_F(e|S)$ in terms of l_i^S:

$$\Delta_F(e|S) = \frac{1}{|N|} \sum_{i \in N} (f(l_i^{S \cup \{e\}}) - f(l_i^S)). \tag{4}$$

In this SingleChoice case, we have $A = A^{plain}$ and we consider the nearest choice for each a. Then, the distances are:

$$l_i^{S \cup \{e\}} = w_{a'} \min_{j \in J_{a'}^S \cup L_{a'} \cup \{m'\}} d_{ij} + u, \tag{5}$$

$$l_i^S = w_{a'} \min_{j \in J_{a'}^S \cup L_{a'}} d_{ij} + u. \tag{6}$$

u is the weighted distance to types $a \in A \setminus \{a'\}$, which is not affected by $e = (a', m')$:

$$u = \sum_{a \in A \setminus \{a'\}} w_a \min_{j \in J_a^S \cup L_a} d_{ij}.$$

$J_{a'}^S$ is the set of locations allocated for type a' under solution set S. $L_{a'}$ is the set of existing instances for a'. For simplicity, we denote the minimum distance to a' under S as $D_{ia'}^S$:

$$D_{ia'}^S = \min_{j \in J_{a'}^S \cup L_{a'}} d_{ij}.$$

Then, from Eqn. (6) we have $l_i^S = w_{a'} D_{ia'}^S + u$. From Eqn. (5), when the new location m' does not reduce the distance to a' under S (i.e., $d_{im'} \geq D_{ia'}^S$), we have $l_i^{S \cup \{e\}} = l_i^S$ and thus $f(l_i^{S \cup \{e\}}) = f(l_i^S)$. When m' results in a new minimum distance (i.e., $d_{im'} < D_{ia'}^S$), we have $l_i^{S \cup \{e\}} = w_{a'} d_{im'} + u$. Therefore, $\Delta_F(e|S)$ in Eqn. (4) is:

$$\Delta_F(e|S) = \frac{1}{|N|} \sum_{d_{im'} \geq D_{ia'}^S} (f(l_i^{S \cup \{e\}}) - f(l_i^S)) + \frac{1}{|N|} \sum_{d_{im'} < D_{ia'}^S} (f(l_i^{S \cup \{e\}}) - f(l_i^S))$$

which simplifies to

$$\Delta_F(e|S) = \frac{1}{|N|} \sum_{d_{im'} < D_{ia'}^S} (f(w_{a'} d_{im'} + u) - f(w_{a'} D_{ia'}^S + u)).$$

Similarly, for solution set T, we have

$$\Delta_F(e|T) = \frac{1}{|N|} \sum_{d_{im'} < D^T_{ia'}} (f(w_{a'} d_{im'} + u) - f(w_{a'} D^T_{ia'} + u)),$$

where $J^T_{a'}$ is the set of locations allocated for type a' under T. Since $S \subseteq T$, we have $J^S_{a'} \subseteq J^T_{a'}$, which leads to $D^T_{ia'} \leq D^S_{ia'}$. $\Delta_F(e|S)$ can then be further grouped into two cases:

$$\Delta_F(e|S) = \frac{1}{|N|} \left[\sum_{d_{im'} < D^T_{ia'}} (f(w_{a'} d_{im'} + u) - f(w_{a'} D^S_{ia'} + u)) + \sum_{D^T_{ia'} \leq d_{im'} < D^S_{ia'}} (f(w_{a'} d_{im'} + u) - f(w_{a'} D^S_{ia'} + u)) \right].$$

We compute and re-arrange $\Delta = \Delta_F(e|S) - \Delta_F(e|T)$:

$$\Delta = \frac{1}{|N|} \left[\sum_{d_{im'} < D^T_{ia'}} (f(w_{a'} D^T_{ia'} + u)) - f(w_{a'} D^S_{ia'} + u)) + \sum_{D^T_{ia'} \leq d_{im'} < D^S_{ia'}} (f(w_{a'} d_{im'} + u) - f(w_{a'} D^S_{ia'} + u)) \right].$$

We know that $(w_{a'} D^T_{ia'} + u) \leq (w_{a'} D^S_{ia'} + u)$ from $S \subseteq T$ and that $(w_{a'} d_{im'} + u) \leq (w_{a'} D^S_{ia'} + u)$ by definition of the second summation. Since WalkScore $f()$ is monotonically decreasing, we have $\Delta \geq 0$. We've thus proved that $\Delta_F(e|S) \geq \Delta_F(e|T)$, as desired. □

No Submodularity with Depth of Choice We show that the objective function is not submodular by providing a counter-example in Appendix B.

4 Models and Algorithms

Mixed-Integer Linear Programming (MILP)

Variables Our MILP model has four sets of variables. First, for allocation, integer variable y_{ja} indicates the number of amenities of type a allocated to location j. Second, binary variables are used to indicate the assignment of amenities to residents. For amenity types where only the nearest instance is considered, $x_{ija} = 1$ indicates that residents at location i visit location j for type a. For types where depth of choice is considered, a fourth index is used and $x^p_{ija} = 1$ indicates that residents at location i visit location j for the p^{th} nearest instance of type a. There are also two sets of continuous variables: l_i and f_i represent the weighted distance and WalkScore for i, respectively. The number of discrete decision variables in the model is $O(|M||N|(|A^{plain}| + |A^{depth}|h))$, where $h = \max_{a \in A^{depth}} |P^Y_a|$.

Constraints First, we enforce the requirements on the maximum number of amenities to be allocated and the capacity of each candidate allocation location:

$$\sum_{j \in M} y_{ja} \leq k_a, \forall a \in A,$$

$$\sum_{a \in A} y_{ja} \leq c_j, \forall j \in M.$$

Second, we describe the assignment of amenities to residents. We ensure that each resident is assigned to one instance for types $a \in A^{plain}$ and to one instance for each available choice for types $a \in A^{depth}$:

$$\sum_{j \in M \cup L_a} x_{ija} = 1, \forall i \in N, a \in A^{plain},$$

$$\sum_{j \in M \cup L_a} x^p_{ija} = 1, \forall p \in P^Y_a, \forall i \in N, \forall a \in A^{depth}.$$

Note that for $a \in A^{depth}$, each choice $p \in P^Y_a$ should be a different instance of a. When the choice is assigned to an existing instance of an amenity type ($j \in L_a$), we ensure that the instance appears only once among all choices for each resident. When the choice corresponds to candidate locations ($j \in M$), we ensure that the number of choices provided at j does not exceed the number of instances allocated to j:

$$\sum_{p \in P^Y_a} x^p_{ija} \leq 1, \forall j \in L_a, \forall i \in N, \forall a \in A^{depth},$$

$$\sum_{p \in P^Y_a} x^p_{ija} \leq y_{ja}, \forall j \in M, \forall i \in N, \forall a \in A^{depth}.$$

Additionally, we ensure that all the amenities are allocated before they are assigned:

$$x_{ija} \leq y_{ja}, \forall i \in N, \forall j \in M \cup L_a, \forall a \in A^{plain},$$

$$x^p_{ija} \leq y_{ja}, \forall i \in N, \forall j \in M \cup L_a, \forall p \in P^Y_a, \forall a \in A^{depth}.$$

Finally, we describe the weighted walking distances based on Eqn. (1) and the PWL WalkScore:

$$l_i = \sum_{a \in A^{depth}} \left(\sum_{p \in P^Y_a} w^p_a \sum_{j \in M \cup L_a} x^p_{ija} d_{ij} + \sum_{p \in P^N_a} w^p_a D^\infty \right) + \sum_{a \in A^{plain}} w_a \sum_{j \in M \cup L_a} x_{ija} d_{ij}, \quad \forall i \in N.$$

For the PWL in Eqn. (3), commercial MILP solvers provide the functionality for linearizing PWL functions (Gurobi Optimization 2022).

Objective The objective is to maximize F in Eqn. (2).

Constraint Programming Model (CP)

We also provide a CP model for WALKOPT. CP allows for an index-based formulation in which decision variables indicate the index of the location of each instance of $a \in A$. This significantly reduces the number of discrete variables compared to the binary formulation in MILP, particularly in the case without depth of choice. Specifically, the number of discrete decision variables is $O(k|A| + h|N||A^{depth}|)$ where $k = \max_{a \in A} k_a$ and $h = \max_{a \in A^{depth}} |P^Y_a|$. The model is deferred to Appendix C.

Algorithm 1: Greedy Algorithm

$S \leftarrow \emptyset$.
$n_a \leftarrow 0$ for all $a \in A$. $\{n_a$: Number of allocated instances for type $a\}$
$c[j] \leftarrow c_j$ for all $j \in M$.
while $\exists n_a < k_a$ **and** $\max_{j \in M}(c[j]) > 0$ **do**
  $(a, j) \leftarrow \operatorname{argmax}_{\substack{a \in A \\ n_a < k_a, j \in M \\ c[j] > 0}} \sum_{i \in N} f(l_i^{S \cup \{(a,j)\}})$
  $S \leftarrow S \cup (a, j)$
  $n_a \leftarrow n_a + 1$
  $c[j] \leftarrow c[j] - 1$
end while
return S

Greedy Algorithm (Greedy)

Motivated by the submodularity in the SingleChoice case, we use a greedy algorithm that iteratively selects the (amenity type, location) pair that maximizes the increase in the objective. We also use Greedy as a heuristic when considering depth of choice and multiple amenity types. Greedy runs in time $O((\sum_{a \in A} k_a)|M||N||A|h)$ and is shown in Algorithm 1. The solution set S contains (a, j) pairs that represent the action of allocating an instance of type a to location j. We denote the weighted walking distance of i under solution set S as l_i^S. The WalkScore function is denoted as $f()$.

5 Case Study

We perform a case study for 31 underserved neighbourhoods in Toronto, Canada. The City of Toronto has identified 31 of its 140 neighbourhoods as *Neighbourhood Improvement Areas* (NIAs) that are facing the most inequitable outcomes under the Toronto Strong Neighbourhood Strategy (TSNS) 2020 (City of Toronto 2022). TSNS aims to provide equitable social, economic, and cultural opportunities for all residents by partnering with agencies to invest in services and facilities in neighbourhoods that face historic underinvestment, and the NIAs capture areas of the city with a significant concentration of disadvantaged and equity-seeking groups, particularly visible minorities. Low walkability and limited access to amenities in the physical surroundings are important criteria in the selection of the NIAs (Social Policy Analysis and Research, City of Toronto 2014).

Data

Neighbourhood Improvement Areas (NIAs) We create instances of WALKOPT from each NIA. The geographical boundary of the NIAs is publicly available from The City of Toronto's Open Data Portal (Social Development, Finance and Administration, City of Toronto 2019).

Pedestrian Network To obtain the network of walkable paths, we use a publicly available *Pedestrian Network* (PedNet) of Toronto which includes various pedestrian assets such as sidewalks, crosswalks, and pedestrian-controlled crossings that are topologically consistent (City of Toronto 2019a). We precompute the shortest-path pairs based on PedNet for each NIA. An alternative to PedNet is OpenStreetMap which provides walking routes data worldwide, which may enable case studies in other cities in the future. We use PedNet in our case study since it undertook quality assurance and has been used in a walking time assessment report by the Transportation Services and Information and Technology Division (City of Toronto 2019b).

Residence and Candidate Locations The locations of residential areas and potential allocation candidates are obtained from OpenStreetMap (OpenStreetMap contributors 2017) and mapped to the nearest nodes in the PedNet. Residential nodes N are the set of nodes that map to at least one residential address. As mentioned in Section 2, our candidate allocation nodes M are parking lots. For this case study, the capacity of each candidate node c_j is the number of parking lots mapped to the node j.

Amenity Weights In this case study, we consider 3 types of amenities: grocery stores, restaurants, and schools, for which the locations of existing instances are also obtained from OpenStreetMap. These 3 types are chosen because they are the major categories that the WalkScore methodology considers (Walk Score 2011) and the data provided by OpenStreetMap is relatively rich for these 3 types based on our visual inspection. The weights w_a (for different types) and w_a^p (for different choices) are obtained from the WalkScore methodology documentation (Walk Score 2011), and the values are listed in Appendix D.

Computational Setup

We perform experiments under two scenarios. In the first (MultiChoice), the distance to the nearest instance is considered for groceries and schools, while the distances to the top 10 nearest choices are considered for restaurants. In the second (SingleChoice), the distance to the nearest instance is considered for all 3 amenity types. MultiChoice is consistent with the original WalkScore methodology (Walk Score 2011), while the SingleChoice scenario helps assess the effect of depth of choice on walkability and solving difficulty.

Instances Throughout the case study, we use the same upper bound for the 3 amenity types considered: $k_a = k, \forall a \in A$. We create 9 instances for each NIA with $k \in \{1, 2, ..., 9\}$. The 31 NIAs are split into 4 groups according to the size ($|M| + |N|$). The number of NIAs (# NIA) and the number of instances (# Inst.) for each NIA group are shown in Table 1.

Setup All methods are implemented in Python. MILP and CP models are solved with Gurobi and CP Optimizer, respectively, and 8 threads. Experiments were run on Intel E5-2683 v4 Broadwell at 2.1GHz CPUs and a memory limit of 32GB. Each solving run is limited to 5 hours. We use a relatively large time limit since a few hours of computation is tiny compared to the service time (in years) of an amenity.

6 Results

Comparison of Solution Methods

Group	Network Size $\lvert M\rvert + \lvert N\rvert$	# NIA	# Inst.	Groceries	Restaurants	Schools
1	[0,200)	11	99	1.45	5.73	3.18
2	[200,400)	13	117	2.00	11.85	5.46
3	[400,600)	4	36	4.50	14.50	9.25
4	[600,1200)	3	27	6.67	19.00	14.33

Table 1: Instance statistics. The last 3 columns are averages.

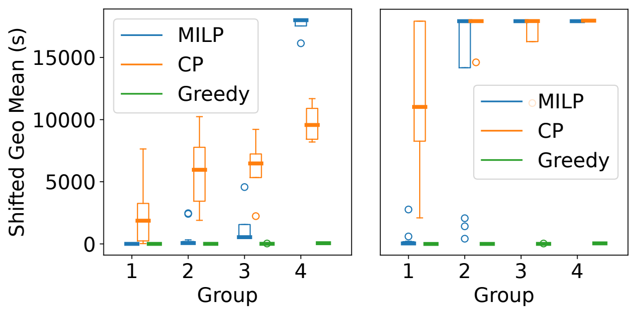


Figure 2: Shifted Geometric Mean in seconds across $k \in \{1, 2, ..., 9\}$ in SingleChoice (left) & MultiChoice (right).

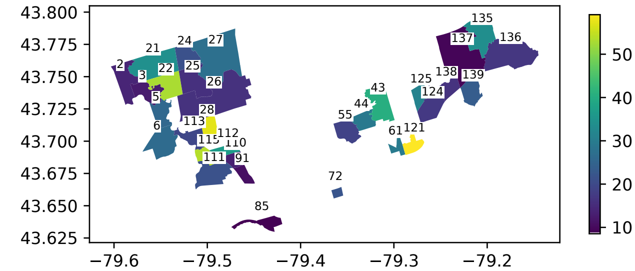


Figure 3: Change in average WalkScore with MultiChoice WALKOPT ($k = 3$). NIAs are labeled by their IDs (neighborhood names in Appendix F).

Solving time We first study how the computational running times scale with increasing network size ($|M| + |N|$). For each NIA, we measure the Shifted Geometric Mean of the solving time across $k \in \{1, 2, ..., 9\}$ (see (ACHTERBERG 2007), section A.3, for a definition of this widely used summary statistic). Fig. 2 shows the results for each NIA group. Greedy is orders of magnitude faster than MILP and CP and scales well in both scenarios. The solving times for CP and MILP in the two scenarios show that the difficulty of the problem significantly increases when considering depth of choice. In SingleChoice, the solving time for MILP is shorter than CP in medium and small instances but increases rapidly for large instances, performing worse than CP in Group 4. A possible explanation is that the number of discrete variables of the CP formulation does not depend on $|M|$ or $|N|$ in this scenario, as discussed in Section 4. In MultiChoice, MILP scales better than CP overall.

Solution Quality Next, we compare the methods in terms of three metrics: the Mean Relative Error (MRE), the number of instances for which the method found a feasible solution, and the number of instances for which the method proved optimality (Table 2). MRE measures the gap between the best solution found by a method and the best solution found across all methods, normalized by the latter. MILP has the lowest MRE for all groups in both scenarios. However, MILP struggles to find feasible solutions for large instances in MultiChoice (Group 4), while CP and Greedy find feasible solutions to all instances. The MRE of Greedy is lower than 0.7% for all groups in both scenarios and is significantly lower than CP for medium and large instances in MultiChoice. This shows that Greedy produces high-quality solutions as a heuristic even when submodularity does not hold.

Empirical Evaluation

What is the effect of solving WALKOPT on the WalkScore and travel distances from residents to amenities, in terms of different neighborhoods and residential locations? To answer this question, we use solutions of MILP for instances where it is feasible since MILP has the lowest MRE on average. For instances where MILP did not find a feasible solution, Greedy's solutions are used.

Impact on WalkScore Fig. 3 shows the change in the average WalkScore for each NIA when additional amenities are introduced by optimization. Most NIAs lie in the mostly-industrial northwest and the suburban/rural northeast of Toronto, where infrastructure is limited. In contrast, few NIAs lie in the urban core in the south where amenities are dense, which is consistent with the reported walkability of Toronto (Toronto Public Health 2012; City of Toronto 2019b). The change in WalkScore varies across NIAs. Adding 3 amenities of each type improves the WalkScore by more than 50 for 4 NIAs. We observe that NIAs with low current WalkScore show greater improvement after allocation; current WalkScore for each NIA in Appendix G.

Additionally, we show how the average WalkScore across all NIAs changes w.r.t. the value of k (Fig. 4). The objective exhibits diminishing returns as k increases in the SingleChoice scenario but not in MultiChoice; this agrees with the submodularity analyses of Section 3. Looking at the travel distances, we see that the framework effectively reduces walking distances for all types/choices considered.

		SingleChoice			MultiChoice		
Group	Method	MRE (%)	Feas	Opt	MRE(%)	Feas	Opt
	MILP	**0.00**	99	**99**	**0.00**	99	**97**
1	CP	**0.00**	99	59	0.43	99	14
(99)	Greedy	0.29	99	N/A	0.66	99	N/A
	MILP	**0.00**	117	**116**	**0.00**	117	**29**
2	CP	0.01	117	31	3.34	117	1
(117)	Greedy	0.34	117	N/A	0.51	117	N/A
	MILP	**0.00**	36	**33**	**0.00**	36	0
3	CP	0.04	36	9	12.41	36	**1**
(36)	Greedy	0.38	36	N/A	0.57	36	N/A
	MILP	**0.00**	27	1	**0.05**	21	0
4	CP	0.91	27	**3**	26.13	**27**	0
(27)	Greedy	0.51	27	N/A	0.20	**27**	N/A

Table 2: Mean relative error (MRE), number of instances for which the method found a feasible solution (Feas) and proved optimality (Opt) for each NIA group. The total number of instances in each group is shown in brackets.

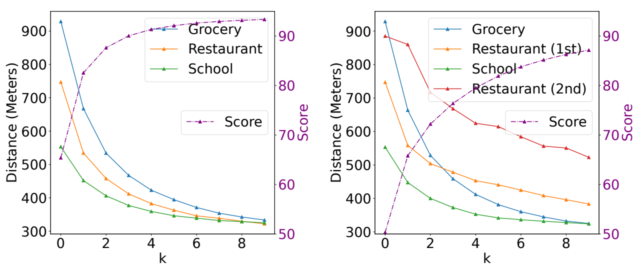


Figure 4: Average WalkScore and travel distances to different amenity types w.r.t. k in SingleChoice (left) & MultiChoice (right). For MultiChoice, distances to the 1st and 2nd restaurants are shown.

Individual Residential Locations We analyze the impact of WALKOPT on individual residential locations using a histogram of walking times to the 3 amenity types for all residential nodes in all 31 NIAs (Fig. 5). Distance-to-time conversion is done using a walking speed of $1.2m/s$ (Traffic Management, City of Toronto 2018). In the MultiChoice case, an allocation with $k = 3$ reduces the walking distances of the 75th percentile of all individual residential locations to 10 minutes for all amenity types. According to Toronto Public Health (2012), a residential preference survey reveals that a 10-minute walking distance to stores and services characterizes a walkable neighbourhood. For grocery stores, this reduction in distance, of up to half relative to the current state, is significant. For schools, we do not observe a large improvement in the mean or the 75th percentile; most residents can walk to a school within 10 minutes currently. As schools are non-commercial, their locations may have been well-optimized historically. However, we do observe a large reduction in the *maximum* walking times to schools.

Visualization of allocated amenities Fig. 6 illustrates a WALKOPT solution for NIA Victoria Village along with existing residential locations, candidate allocation locations, and existing amenities. Allocated amenities seem to fall at the heart of residential clusters. Perhaps more interestingly, some newly allocated amenities are very close to the locations of other types of amenities (existing or allocated) and seem to form an urban center with a mix of different types.

7 Related Work

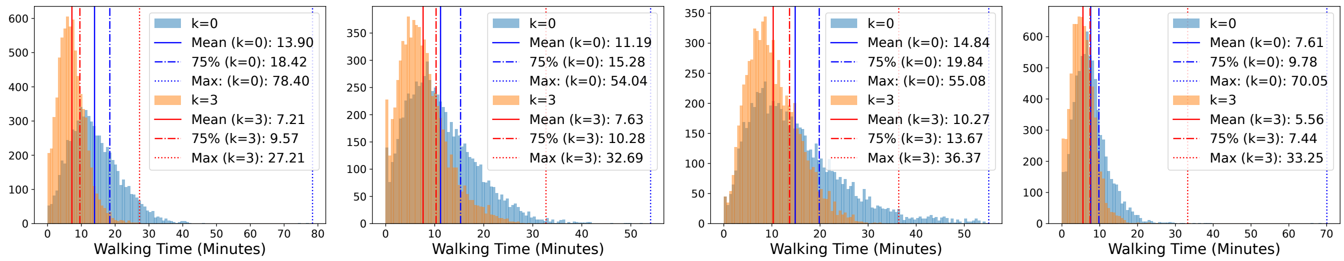
Walkability Optimization

The Introduction already discusses some of the most relevant work that uses genetic algorithms. To further elaborate, Cichocka (2015) optimizes for each different amenity type independently, which may hinder optimality. Moreover, existing works suffer from unrealistic assumptions that limit the applicability of the framework and the quality of empirical evaluation such as designing street grid patterns from scratch (Lima, Brown, and Duarte 2022) and allocating amenities to empty street layout (Rakha and Reinhart 2012). Sonta and Jain (2019) perform a case study in an existing city but only use randomly sampled residential units and ignore current amenity locations, generating solutions that override existing infrastructure.

k-Median and Facility Location Problem (FLP)

Compared to the FLP, WALKOPT considers multiple facility types and depth of choice. Without these, WALKOPT is equivalent to the Submodular FLP defined in (Lindgren, Wu, and Dimakis 2015), and previous work has shown that objective (2) is submodular in this case (Frieze 1974).

Moreover, the objective function WalkScore does not satisfy the properties of a metric space, in contrast to the closely related k-median problem. Algorithms for k-median that provide better approximation ratios than standard greedy include reverse greedy (Chrobak, Mathieu, and Young 2005), local search (Arya et al. 2001), LP relaxation (Charikar et al. 1999; Charikar and Li 2012), and Lagrangian relaxation



(a) Grocery (b) 1st Nearest Restaurant (c) 2nd Nearest Restaurant (d) School

Figure 5: Histogram of walking times to different amenity types across all residential nodes in all neighbourhoods. Adding $k = 3$ amenities shifts the histogram of walking times to the left (in orange), which also translates into smaller mean/maximum/75th percentile walking times relative to not adding any amenities (in blue).

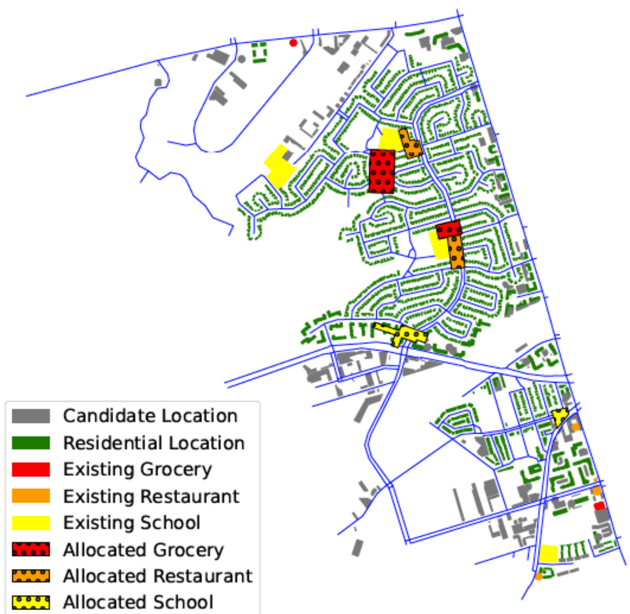


Figure 6: Allocated amenities at Victoria Village: 2 newly introduced groceries, restaurants, and schools.

(Jain and Vazirani 2001; Jain et al. 2003). However, these algorithms assume that the objective function is defined on a metric space and are not applicable to our problem.

8 Conclusion and Discussion

Automobile-reliant communities with limited access to amenities in their vicinity have a great potential for transformation into more walkable and sustainable neigbhbourhoods. We formulate the problem of Walkability Optimization where amenities are introduced at strategic locations to improve the proximity to residents. Our WALKOPT formulation realistically models residents' behaviour by integrating multiple amenity types, depth of choice, and an objective function representing the proximity to amenities. We also take into account existing amenities in the context of built cities. We provide MILP and CP formulations and an efficient greedy algorithm motivated by the submodular structure of the WALKOPT objective (without depth of choice). An experimental evaluation on high-quality data from Toronto shows that MILP and Greedy are effective at producing high-quality solutions, with a scalability advantage for the latter. Our framework produces solutions that significantly improve the walkability in underserved neighbourhoods on average and reduce the walking distances for individual residential locations.

While we have prioritized incorporating realistic facets of walkability optimization into our formulation, more can potentially be done by: considering the population at each residential location, the construction cost at each candidate allocation location, the area/size of candidate locations, and the service capacity of amenities. If the data is available, these can be easily integrated into the objective and constraints. In addition, WALKOPT quantifies walkability in terms of travel distances without considering other factors that may affect accessibility such as the safety/quality of walking paths, which can potentially be incorporated into the formulation by applying penalties appropriately. Lastly, our experiments were based on neighbourhood-scale instances; testing our methods at full city-scale might be of future interest.

References

ACHTERBERG, T. 2007. Constraint Integer Programming. *Ph. D. Thesis, Technische Universitat Berlin*.

Arya, V.; Garg, N.; Khandekar, R.; Meyerson, A.; Munagala, K.; and Pandit, V. 2001. Local search heuristic for k-median and facility location problems. In *STOC '01*.

Brand, C.; Götschi, T.; Dons, E.; Gerike, R.; Anaya-Boig, E.; Avila-Palencia, I.; de Nazelle, A.; Gascon, M.; Gaupp-Berghausen, M.; Iacorossi, F.; Kahlmeier, S.; Int Panis, L.; Racioppi, F.; Rojas-Rueda, D.; Standaert, A.; Stigell, E.; Sulikova, S.; Wegener, S.; and Nieuwenhuijsen, M. J. 2021. The climate change mitigation impacts of active travel: Evidence from a longitudinal panel study in seven European cities. *Global Environmental Change*, 67: 102224.

Charikar, M.; Guha, S.; Tardos, E.; and Shmoys, D. 1999. A constant-factor approximation algorithm for the k-median problem. *Proceedings of 31st Annual ACM Symposium on Theory of Computing*, 1–10.

Charikar, M.; and Li, S. 2012. A Dependent LP-Rounding Approach for the k-Median Problem. In *ICALP*.

Chrobak, M.; Mathieu, C.; and Young, N. 2005. The reverse greedy algorithm for the metric k-median problem. *Inf. Process. Lett.*, 97: 68–72.

Cichocka, J. M. 2015. Generative design optimization in urban planning: walkability-optimized city concept. *Architectus*, 119–128.

City of Toronto. 2019a. Toronto Walkability Project. https://github.com/gcc-dav-official-github/dav_cot_walkability. Accessed: 2022-07-17.

City of Toronto. 2019b. Walkability Report. https://ckan0.cf.opendata.inter.prod-toronto.ca/dataset/4b5c7a84-dea1-4137-875d-71d7f662c83f/resource/d546f5d1-ecc2-4237-bcc9-84000715b60a/download/Toronto%20Walkability%20Analysis.pdf. Accessed: 2022-08-01.

City of Toronto. 2022. Neighbourhood Improvement Area Profiles. https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/nia-profiles/. Accessed: 2022-07-17.

Creutzig, F.; Roy, J.; Lamb, W. F.; Azevedo, I. M. L.; de Bruin, W. B.; Dalkmann, H.; Edelenbosch, O. Y.; Geels, F. W.; Grubler, A.; Hepburn, C.; Hertwich, E. G.; Khosla, R.; Mattauch, L.; Minx, J. C.; Ramakrishnan, A.; Rao, N. D.; Steinberger, J. K.; Tavoni, M.; ürge Vorsatz, D.; and Weber, E. U. 2018. Towards demand-side solutions for mitigating climate change. *Nature Climate Change*, 8: 260–263.

Duncan, D. T.; Aldstadt, J.; Whalen, J.; Melly, S.; and Gortmaker, S. L. 2011. Validation of Walk Score® for Estimating Neighborhood Walkability: An Analysis of Four US Metropolitan Areas. *International Journal of Environmental Research and Public Health*, 8: 4160 – 4179.

Eric Reguly. 2020. Bikes, pedestrians and the 15-minute city: How the pandemic is propelling urban revolutions. https://www.theglobeandmail.com/canada/article-bikes-pedestrians-and-the-15-minute-city-how-the-pandemic-is/?_sp=64b0e81e-9061-4f86-94b9-c74d400683e3.1619052392166. Accessed: 2022-07-17.

Fischel, W. A. 2003. An Economic History of Zoning and a Cure for its Exclusionary Effects. *Urban Studies*, 41: 317 – 340.

Frieze, A. M. 1974. A cost function property for plant location problems. *Mathematical Programming*, 7: 245–248.

Gurobi Optimization. 2022. Model.addGenConstrPWL(). https://www.gurobi.com/documentation/9.1/refman/py_model_agc_pwl.html. Accessed: 2022-08-14.

Indraprastha, A. 2019. Informed Walkable City Model - Developing A Multi-Objective Optimization Model for Evaluating Walkability Concept. *Proceedings of the 24th Conference on Computer Aided Architectural Design Research in Asia (CAADRIA) [Volume 2]*.

Jain, K.; and Vazirani, V. V. 2001. Approximation Algorithms for Metric Facility Location and K-Median Problems Using the Primal-Dual Schema and Lagrangian Relaxation. *J. ACM*, 48(2): 274–296.

Jain, K. K.; Mahdian, M.; Markakis, E. K.; Saberi, A.; and Vazirani, V. V. 2003. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *ArXiv*, cs.DS/0207028.

Kim, B. M.; Kim, Y. B.; and Oh, C. H. 1997. A study on the convergence of genetic algorithms. *Computers & Industrial Engineering*, 33: 581–588.

Lee, C.; and Moudon, A. V. 2006. Correlates of walking for transportation or recreation purposes. *Journal of Physical Activity and health*, 3(s1): S77–S98.

Levine, J. 2005. *Zoned Out: Regulation, Markets, and Choices in Transportation and Metropolitan Land Use*. Hoboken: Taylor and Francis.

Lima, F. T.; Brown, N. C.; and Duarte, J. P. 2022. A grammar-based optimization approach for walkable urban fabrics considering pedestrian accessibility and infrastructure cost. *Environment and Planning B: Urban Analytics and City Science*, 49: 1489 – 1506.

Lindgren, E. M.; Wu, S.; and Dimakis, A. G. 2015. Sparse and greedy: Sparsifying submodular facility location problems. In *NIPS Workshop on Optimization for Machine Learning*.

Massey, D. S. 1990. American Apartheid: Segregation and the Making of the Underclass. *American Journal of Sociology*, 96: 329 – 357.

McIntosh, J.; Trubka, R.; Kenworthy, J.; and Newman, P. 2014. The role of urban form and transit in city car dependence: Analysis of 26 global cities from 1960 to 2000. *Transportation Research Part D: Transport and Environment*, 33: 95–110.

Mouratidis, K. 2021. How COVID-19 reshaped quality of life in cities: A synthesis and implications for urban planning. *Land Use Policy*, 111: 105772.

Nagy, D.; Villaggi, L.; and Benjamin, D. 2018. Generative urban design: integrating financial and energy goals for automated neighborhood layout. In *Proceedings of the Symposium for Architecture and Urban Design Design, Delft, the Netherlands*, 265–274.

Natalie Marchant. 2020. Paris halves street parking and asks residents what they want to do with the space. https://www.weforum.org/agenda/2020/12/paris-parking-spaces-greenery-cities/. Accessed: 2022-07-17.

Nemhauser, G. L.; Wolsey, L. A.; and Fisher, M. L. 1978. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14: 265–294.

Neves, A.; and Brand, C. 2019. Assessing the potential for carbon emissions savings from replacing short car trips with walking and cycling using a mixed GPS-travel diary approach. *Transportation Research Part A: Policy and Practice*.

Ngueveu, S. U. 2019. Piecewise linear bounding of univariate nonlinear functions and resulting mixed integer linear programming-based solution methods. *Eur. J. Oper. Res.*, 275: 1058–1071.

OpenStreetMap contributors. 2017. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org. Accessed: 2022-07-17.

Rakha, T.; and Reinhart, C. 2012. Generative urban modeling: a design work flow for walkability-optimized cities. *Proceedings of SimBuild*, 5(1): 255–262.

Reyer, M.; Fina, S.; Siedentop, S.; and Schlicht, W. 2014. Walkability is Only Part of the Story: Walking for Transportation in Stuttgart, Germany. *International Journal of Environmental Research and Public Health*, 11: 5849 – 5865.

Social Development, Finance and Administration, City of Toronto. 2019. About Neighbourhood Improvement Areas. https://open.toronto.ca/dataset/neighbourhood-improvement-areas/. Accessed: 2022-07-17.

Social Policy Analysis and Research, City of Toronto. 2014. TSNS 2020 NEIGHBOURHOOD EQUITY INDEX Methodological Documentation. https://www.toronto.ca/legdocs/mmis/2014/cd/bgrd/backgroundfile-67350.pdf. Accessed: 2022-07-17.

Sonta, A. J.; and Jain, R. K. 2019. Optimizing Neighborhood-Scale Walkability. *Computing in Civil Engineering 2019*.

Toronto Public Health. 2012. The Walkable City: Neighbourhood Design and Preferences, Travel Choices and Health. https://www.toronto.ca/wp-content/uploads/2017/10/9617-TPH-walkable-city-report.pdf. Accessed: 2022-08-01.

Traffic Management, City of Toronto. 2018. Pedestrian Timing at Signalised Intersections. https://www.toronto.ca/wp-content/uploads/2019/01/96a8-TS_Traffic-Management_Pedestrian-Timing-Guidelines.pdf. Accessed: 2022-07-17.

Verbas, I. Ö.; Frei, C.; Mahmassani, H. S.; and Chan, R. 2015. Stretching resources: sensitivity of optimal bus frequency allocation to stop-level demand elasticities. *Public Transport*, 7: 1–20.

Walk Score. 2011. Walk Score Methodology. http://pubs.cedeus.cl/omeka/files/original/b6fa690993d59007784a7a26804d42be.pdf. Accessed: 2022-07-17.

Whittle, N. 2020. Welcome to the 15-minute city. *Financial Times*, 17.

Yang, X.-S. 2020. *Nature-inspired optimization algorithms*. Academic Press.

Yang, Y.; Samaranayake, S.; and Dogan, T. 2020. An adaptive workflow to generate street network and amenity allocation for walkable neighborhood design. Proceedings of the Symposium on Simulation for Architecture and Urban Design.

Zapata-Diomedi, B.; Boulangé, C. L.; Giles-Corti, B.; Phelan, K.; Washington, S.; Veerman, J. L.; and Gunn, L. D. 2019. Physical activity-related health and economic benefits of building walkable neighbourhoods: a modelled comparison between brownfield and greenfield developments. *The International Journal of Behavioral Nutrition and Physical Activity*, 16.