

EQUITABLE PUBLIC TRANSPORT NETWORK REDUCTION A REINFORCEMENT LEARNING APPROACH

SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE

RICCARDO FIORISTA
14012987

MASTER INFORMATION STUDIES
DATA SCIENCE
FACULTY OF SCIENCE
UNIVERSITY OF AMSTERDAM

2022-06-30

	Dimitris Michailidis	Fernando P. Santos
Title, Name	PhD cand.	Assistant Prof.
Email	d.michailidis@uva.nl	f.p.santos@uva.nl



Equitable Public Transport Network Reduction

A Reinforcement Learning Approach

Riccardo Fiorista

University of Amsterdam

Amsterdam, The Netherlands

riccardo.fiorista@student.uva.nl

ABSTRACT

Public Transport Networks (PTNs) are the backbone of urban socio-economic functioning and as urbanization increases globally, they represent a crucial driver of access to socio-economic opportunities. In this work, we formalize the Equitable PTN Reduction (E-PTNR) problem. Using this framework, we present an analysis of the inequality of access to educational opportunities between the population with western and non-western migration backgrounds in Amsterdam and show that the latter group is significantly disadvantaged in more than 91% of the times in terms of average travel time. We find that the number of PTN reductions and particularly the quality of such reductions may impact the equality distribution across the city additionally. Furthermore, we present a set of baselines on three synthetic datasets and a novel Deep MaxQ Network (DMaxQN)-learning approach to address the E-PTNR problem set. Our DMaxQN-learning achieves similar results as a Genetic Algorithm and the classical MaxQ-learning approach. With our contribution, real-world-sized E-PTNR problems can be addressed, resulting in PTN reductions which maximize socio-economic access equality.

KEYWORDS

Public Transport, Equality, Reinforcement Learning

1 INTRODUCTION

As a consequence of a globally growing urban population, demand for Public Transport (PT) grows [40], especially in light of policy goals to promote public mobility over car usage to reduce emissions [32]. Hence, PT is promoted as the backbone of socio-economic functioning [51]. However, PT providers face ever-stricter budgeting, governmental regulations, and diminished demand during exceptional situations such as the SARS-CoV-19 pandemic [10, 16]. Certain PT providers, such as the Gemeentevervoerbedrijf (GVB) in Amsterdam, now rely on external funds and see PT Network (PTN) reductions as the primary technique to adapt to the new financial limitations [23, 24].

PTN reductions, however, could increase socioeconomic inequality [26, 41] and widen a divide which is common in urban environments [27]. Such a divide could manifest, for example, in the accessibility of socio-economic opportunities for different population groups, such as individuals with western migration backgrounds (**w**) and those with non-western ones (**nw**). Hence, we consider migration background as our *protected attribute* in our work. Previous research has shown that especially lower-income neighborhoods, commonly located in cities' peripheral areas, are usually affected the most by PTN interventions [19, 28, 38]. Therefore, it is critical to plan PTN modifications with socio-economic

access equality in mind and quantify the impact of such changes on different population groups.

A problem related to the planning of PTN reductions is the Public Transport Network Design Problem (PTNDP) [14] which concerns the creation or extension of PTNs. Recent advances in this field have successfully employed Reinforcement Learning (RL) approaches, showing that policy-driven agents can provide strategies to find optimal solutions while simplifying the encoding of domain knowledge over previous methods such as linear programming and improving on Genetic Algorithms (GAs) [35, 50]. These novel approaches to solve the PTNDP have also considered *public equality of access*, i.e., the maximization of utility and equality in access provided by PT [53], allowing for an equality-driven design.

However, the PTNDP is more concerned with the design and extension of PTNs rather than their reductions. Research on such reductions mainly investigates their impact on graph properties such as connectivity, service robustness, and vulnerability. The PTN reductions considered in these studies, however, are caused by partial failures [7, 21] or random and directed attacks [5], hence not planned or voluntary.

Thus, to our knowledge, the current research landscape lacks a well-defined mathematical abstraction for systematic PTN reductions under the lens of public equality of socio-economic access.

This thesis addresses the identified research gap by proposing a novel framework and approach to PTN reduction, which we refer to as *Equitable Public Transport Network Reduction (E-PTNR)*. In our work, we first show the necessity of solving the E-PTNR problem by analyzing the city of Amsterdam's socio-economic access inequality and the impact of PTN reductions over the years 2019–2021. Subsequently, we show that the problem is combinatorial and contains a significantly sized problem space, especially in real-world settings. Due to its success in other domains such as the PTNDP and its well-defined properties for sequential decision making, we propose a maximum reward RL-driven approach to solve the E-PTNR problem.

The E-PTNR problem requires us to find a set of up to budget k interventions, i.e., edges to remove from the PTN, such that public access equality is *maximized*. To this end, we build on the works of Gottipati et al. [18] who introduce a novel MaxQ-learning approach for maximum reward maximization under the domain of maximum reward RL [34]. While we show that their approach is well suited for the E-PTNR problem, we identify its limited scalability due to its reliance on a look-up table for decision-making. Such a table could not be held in memory for real-world-sized E-PTNR problems. Hence, our work extends the MaxQ-learning approach with a Deep Q-Network (DQN), i.e., a Neural Network (NN) approximating the tabular policy, which we refer to as DMaxQN-learning.

1.1 Research Questions

Our main research question is:

How can reinforcement learning be used to identify public transport network reduction under a budget while maintaining or improving access equity throughout the serviced area and its population?

To address our main research question, we further subdivide our endeavor into the following sub-research questions:

- (1) How can the E-PTNR problem be mathematically formalized?
- (2) How can public access equality be numerically quantified and included in an objective function informing our E-PTNR approach?
- (3) Is there evidence that current PTNR strategies lead to inequality in accessibility for some population groups?
- (4) Is the application of standard RL applicable to the E-PTNR problem, and does it outperform other approximation algorithms, such as genetic algorithms?

Our first two questions are methodological, while our experiments will answer questions (3) and (4). Thus, our main contributions are:

- (1) A reusable mathematical framework for quantifying and evaluating E-PTNR problems.
- (2) An open-source Python3 framework for the computational evaluation of the above with access to our baselines and deep learning methods (`e-ptnr1`).
- (3) A novel deep-learning-based maximum reward RL, DMaxQN-learning, approach to address the E-PTNR problem.

2 RELATED WORK

PT is a cross-discipline topic that in the past has been the focus of major fields such as computer science, operational research, behavioral sciences, economics, and econometrics. On the one side, there is research originating from the mathematical and computational sciences [4]. Here, network theory, for example, introduced the notion of geographic transport *networks* and their representation as graphs [11]. This concept has then been central to the domain of operations research for the planning and scheduling of PT services [13]. On the other side, there is research concerning user behaviour, economics, and the socio-economic impact of PT on society [19, 28, 45, 47, 47]. However, recent research has skewed towards unifying these mathematical and societal domains, creating a cross-disciplinary interest in the computational design and optimization of PTNs under a socio-economic lens [29, 35, 50].

The SARS-CoV-19 pandemic gave rise to an accelerated, cross-disciplinary effort to contribute to the field of PT. Researchers and policymakers alike have called for the necessity of adaptability of PT [10, 16, 36]. This line of research sheds light on the lack of knowledge and methods for service adaptation and impact mitigation, among others, [13, 16, 46]. Our work addresses this knowledge gap and attempts to provide a mathematical formalization and a novel approach to efficiently find solutions to allow quick PTN size reductions while maximizing socio-economic access equality.

¹<https://github.com/EquiCity/thesis>

We first formally introduce the concept of equality and one possible formulation. Subsequently, we present the previous work on quantifying access to socio-economic opportunities across different demographic groups in the population. Finally, we consider existing research on RL in the PT domain and hence introduce the domain in which our work is rooted.

2.0.1 The Quantification of Equality. The best known and de-facto standard measure of (in)equality is the Gini index [8]. However, recent research has shown that this measure displays misrepresentations of the true inequality distribution [33]. Furthermore, it does not retain its properties with negative values or distributions centered at zero [20].

Hence, over the past decades, other (in)equality measures emerged from the field of information theory, specifically building on the concept of General Entropy (GE). These measures fulfill the basic requirements of inequality metrics, namely symmetry, scale independence, population independence, and the Pigou-Dalton principle [42]. However, they are also *additively decomposable* into within- and between-group inequality components. The most popular measures of this family are the Theil index, and the Atkinson index [3]. van Wee and Mouter [46] shows that these measures have also found their place in the PT literature as indices to evaluate transport equality despite their difficulty in communication to a broader audience and policy-makers.

2.1 Quantifying Public Accessibility and Equality

Geurs and van Wee identified four different approaches for the numerical quantification of public accessibility: infrastructure-based, location-based, person-based, and utility-based [15]. In this order, these measures vary from macroscopic to microscopic in that they are tailored to the needs of single users [4].

Infrastructure-based measures are generally most useful for holistic assessments of PTNs, such as overall service quality. **Location-based** measures, on the contrary, are vital to reason about the population's macroscopic experience of their socio-economic reach. These two categories include the following metrics, which we use throughout our work:

2.1.1 Location-based Accessibility Measures.

- **Average Travel Time:** the average time individuals of a certain group require to reach a certain type of socio-economic Points Of Interest (POIs) such as schools, hospitals, or workplaces, to name a few.
- **Average Journey Segment Count (Hops):** one *hop* here refers to one segment of a service, e.g., a bus, between two PT Stations (PTS). These journey segments directly affect accessibility as they indirectly indicate possible transport modality switches, crowdedness, and delay possibilities. Hence, the average number of hops is synonymous with the average number of segments a group has to traverse to reach a certain type of POI.
- **Cumulative Opportunity Measure (COM):** is the adequate number of POIs a population group can reach within a certain time threshold. Most recently, 15 minutes has been

coined as a critical threshold for future urban development [31].

In Section 3, we will elaborate further on the mathematical definitions of these measures and how we include them to inform our E-PTNR approaches.

2.2 Reinforcement Learning for Public Transport Optimization

Previous research on the PTND used techniques such as linear programming, integer linear optimization, Evolutionary Algorithms (EAs), and RL [10, 29, 50]. On the one side, a common pitfall of linear or dynamic programming approaches is the high amount of encoded expert knowledge required for the problem set up in the form of constraints, variable domains, and costs. On the other, EA-based algorithms, such as GAs, are generalized optimization techniques that can be applied to a wide set of problems without being specifically designed for them [52]. Thus, recent literature gravitates towards agent-based RL approaches, specifically designed for sequential decision making, which is central to the PTNDP [25, 50]. Here, like GAs, domain knowledge can be encoded in the reward formulation the algorithms attempt to maximize.

3 METHODOLOGY

In this section, we present the mathematical formalization of the E-PTNR problem. Furthermore, we define one possible access equality metric as an objective for the underlying optimization problem. Finally, we present the optimization and the datasets we will be applying our approach.

3.1 Definition of the Graph

We represent the PTN as a directed, cyclic, weighted multi-graph $\mathcal{G} = \langle V, E \rangle$ with vertex set V and edge set E . A directed multi-graph refers to a graph where the edge set is a multi-set of pairs of vertices, such that any two vertices could be connected through multiple edges [44].

Set V is composed of three types of vertices. First, the population-weighted Residential Centroids (RC) of neighbourhoods V^{rc} . These serve as a simplified representation of the population environment in the city. Second, the physical locations of the PTN's PTSs V^{pt} . Third, we summarize socio-economic POIs as the vertex set V^{poi} . Together they form the complete set $V = \{V^{rc}, V^{pt}, V^{poi}\}$.

All vertices in V^{rc} are the origin nodes from which citizens of different population groups commence their journey towards the POI nodes in V^{poi} either directly or through the use of the PTN, inherently visiting vertices in V^{pt} .

In Figure 1, we illustrate an example graph containing all of the vertices mentioned above. Here, all edges connecting the nodes in V^{rc} to those in V^{pt} (solid red) and V^{poi} (dotted violet), as well as the edges connecting the nodes in V^{pt} with those in V^{poi} (dashed green), are of type **WALK** and indicate a non-PT modality of transport. For the E-PTNR problem, we consider modifications on all other edges indicating a bus, tram, metro, or other modality connection [17] between two stations in V^{pt} . For simplicity, we refer to this set of edges as $E^{pt} \subset E$, which are extracted from a city's PTN specifications.

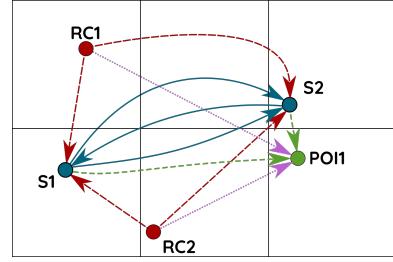


Figure 1: A sample graph, formalizing the key requirements of an E-PTNR setup.

Residential Centroid (RC), Public Transport Station (PTS) and Point of Interest (POI) vertices We illustrate the edges between the RCs and PTSs in long dashed red, the lines between PTSs in solid blue, the edges between RCs and POIs in dotted violet, and the edges between PTSs and POIs in short dashed green. The edges from RCs to PTSs, RCs to POIs, and PTSs to POIs represent the transport mode **walk** while the edges between the PTSs, i.e. the edges in the PT edge set E^{pt} , can be multi-modal (**bus, tram, metro, ...**)

3.1.1 Population Features as Node Features. Based on \mathcal{G} and the information about physical locations and travel times it encodes, we can compute the per-neighborhood travel statistics identified in Section 3.2.1 as illustrated in Figure 10. The computations of average travel time, average hops, and the 15-minute COM are conducted for each RC by averaging the shortest paths between each RC and all POIs. Once we obtain the travel metrics for each RC, we then use the population data for each RC to generate a city-wide sample of individuals and their respective travel metrics. A visual example of such a computation can be found in Appendix A, Figure 10.

3.2 Objective Formulation

Our objective is to identify the set K of edges $e_{i,j} \in E^{pt}$, where $k = |K|$ is the pre-defined budget of edges to be removed, such that the inequality of access between (and within) the different groups in a city is minimized. Hence, we here define the metrics, the measure of inequality we use to quantify this difference, and an appropriate reward for quantifying the quality of a solution.

3.2.1 Accessibility Metrics. For each of the following accessibility metrics, we consider the shortest paths between a residential centroid $v_i \in V^{rc}$ and all the POIs $v_j \in V^{poi}$ which we define as a function $h : (V^{rc} \times V^{poi}) \rightarrow P \subseteq E$. The resulting shortest path P is defined by a set of vertices $P = v_1^{rc}, v_2^{pt}, \dots, v_{n-1}^{pt}, v_n^{poi} \in V^{rc} \times V^{pt} \times \dots \times V^{pt} \times V^{poi}$ s.t. v_i is adjacent to v_{i+1} for $1 \leq i < n$. Given a function $f : E \rightarrow \mathbb{R}^+$ which returns the travel time of an edge $e_{i,j} \in E$, P is the path which minimizes the sum $\sum_{i=1}^{n-1} f(e_{i,i+1})$.

Following, we will mathematically quantify the accessibility metrics introduced in Section 2.1.1 based on the graph \mathcal{G} which we have defined above:

Group average travel time: This is the average time the residents of group $g \in G$ in neighbourhood $v_i^{rc} \in V^{rc}$ require to reach all POIs. We define it as the group-weighted average of the cumulative travel times over all shortest paths $f((v_i, v_j)), v_i \in V^{rc}, \forall v_j \in V^{poi}$:

$$m_{avg.tt}(v_i^{rc}) = \frac{1}{|V^{poi}|} \sum_{v_j^{poi} \in V^{poi}} \sum_{e \in h(v_i^{rc}, v_j^{poi})} f(e) \quad (1)$$

Group average segment count (hops): Every shortest path P between v_i^{rc} and a POI v_j consists of $|P| - 1$ edges or journey segments. We quantify the average segment count per neighborhood and group as:

$$m_{\text{avg.sc}}(v_i^{\text{rc}}; g) = \frac{1}{n_g} \sum_{v_j^{\text{poi}} \in V^{\text{poi}}} n_g^{v_i} |f((v_i^{\text{rc}}, v_j^{\text{poi}}))| \quad (2)$$

Group COM: The cumulative opportunity measure is one of the most used measures in the PT setting to quantify the accessibility to socio-economic opportunities (specified as POIs in our case). It is a threshold-based metric that considers the number of reachable POIs within a certain time t :

$$m_{\text{com}}(v_i^{\text{rc}}) = \sum_{v_j \in V^{\text{poi}}} [f((v_i^{\text{rc}}, v_j^{\text{poi}})) < t] \quad (3)$$

3.2.2 Computation of Inequality.

The Theil T index. is our inequality measure of choice due to its reduced limitations as well as its heavier weighting of long-tailed distributions (Appendix B). In Equation 4 we show the definition of the Theil T measure where n is the number of individuals in a population, \mathbf{y} the vector of a certain per-individual travel metric, and μ the mean of \mathbf{y} . As we are interested in the inequality between groups, we show in Equation 5 the definition of the one-level grouped Theil T index. Here, instead of a single vector \mathbf{y} , we pass a collection of vectors Y , each representing a group's distributional vector. For a more in-depth analysis we refer the reader to Appendix B.

$$T(\mathbf{y}; n) = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\mu} \ln \left(\frac{y_i}{\mu} \right) \quad (4)$$

$$T(Y; n) = T \left(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^G; n \right) \quad (5)$$

The egalitarian objective. is central to our E-PTNR under budget approach. The reward can be simply set as the overall reduction of the T across all metrics m . With this reward/objective, we quantify the overall reduction of inequality (entropy) in the PTN:

$$r_{\text{ega.}}(\mathcal{G}) = - \sum_m T \left(\hat{Y}_m \right) \quad (6)$$

Note that this objective is bound in $r_{\text{ega.}} \in [-|m| \ln N, 0]$ where $|m|$ is the number of accessibility metrics regarded. Total equality is reached with $r_{\text{ega.}} = 0$.

3.3 Optimization Formulation

Our objective is to find a set of edges $\mathcal{E} = \{e_1, e_2, \dots, e_i\} \subseteq E^{\text{pt}}, i \leq k$, up to budget k , such that the graph $\mathcal{G}' = \mathcal{G} - \mathcal{E}$ maximizes the objectives above. Hence, we can formalize the objective as:

$$J = \max_{s.t. |\mathcal{E}| \leq k} r(\mathcal{G}') \quad (7)$$

As our maximization formulation depends on \mathcal{E} , similarly to the formulation presented by Ramachandran et al. [35], it is non-differentiable and can therefore not be solved using gradient-based optimization methods. Furthermore, note that the order in which the edges are

removed does not influence the final reward value $r(\mathcal{G})$. Hence this is a combinatorial rather than a permutation-based problem.

3.4 Complexity Analysis

Following from the above, the size of graph \mathcal{G} is $|\mathcal{G}| = (|V|, |E|)$ where $|V| = |V^{\text{rc}}| + |V^{\text{pt}}| + |V^{\text{poi}}|$ and $|E| = |V^{\text{rc}}| \times |V^{\text{pt}}| + |V^{\text{rc}}| \times |V^{\text{poi}}| + |V^{\text{pt}}| \times |V^{\text{poi}}| + |E^{\text{pt}}|$. The last set of edges $|E^{\text{pt}}|$ is defined by the PTN and hence cannot be quantified a priori. The worst, yet unrealistic, case would be where $|E^{\text{pt}}| = |V_P| \times |V_P|$, i.e., the PTN is a fully connected network.

Due to the combinatorial nature of the E-PTNR problem, the possible solutions for $s = |E^{\text{pt}}|$ edges and budget k is $|\mathcal{S}| = \binom{s}{k}$. Thus, the space of candidate solutions grows polynomially in $|E^{\text{pt}}|$ and exponentially in budget k .

As a RL approach relies on taking sequential decisions, the state space consists of all the possible states up to budget k , i.e. $|\mathcal{S}| = \sum_{i=0}^k \binom{n}{i}$. Assume that we aim to remove $k = 4$ edges from E^{pt} , then a graph with $n = 11$ edges would result in $|\mathcal{S}| = 562$ states. However, in a realistic setup, we observe PTN edge sizes of $s = 1600$. Here, with the same budget, we would have a solution space of $|\mathcal{S}| \approx 2.7 \times 10^{11}$. With a budget $k = 10$, in turn, this would already be at $|\mathcal{S}| \approx 2.9 \times 10^{25}$ possible states.

3.5 Baselines

We present the baselines with which we compare and validate our RL approach. All baselines are run on the three synthetic E-PTNR problem-graphs presented in Figure 2 and the hyperparameters used are reported in Appendix I. Finally, we show the pseudo-code for each baseline algorithm in Appendix E.

3.5.1 Random Baseline. We devise this as our most naïve approach to verify the working of our setup and, two, to have a baseline to which to compare the informed baselines below. Here, k edges are chosen iteratively uniformly at random without repetition from the edge set E^{pt} . The chosen reward is evaluated at each removal. In contrast, the reward of the final solution \mathcal{E} where $|\mathcal{E}| = k$ is the indicative value for the performance on the E-PTNR problem.

3.5.2 Fully Informed. As our second baseline, we consider the entire solution space with the fully informed algorithm. Hence, we first generate the set of all possible solutions, i.e. the k -subsets of E^{pt} , $\mathcal{S} = \binom{E^{\text{pt}}}{k} \subseteq E^{\text{pt}}$. Subsequently, given reward r , all solutions in \mathcal{S} are evaluated, and the solution resulting in the maximum reward across the entire solution space is returned as solution \mathcal{E} .

3.5.3 Fully Informed, Greedy Baseline. Our first informed algorithm performs a breadth-first search by evaluating all possible edge removals of \mathcal{G} . Once the edge leading to the highest reward r after its deletion is identified, it is added to the solution set \mathcal{E} and a new graph $\mathcal{G}' = \mathcal{G} - \mathcal{E}$ is generated. The process is repeated until $|\mathcal{E}| = k$.

3.5.4 Genetic Algorithm. The method we use for this baseline uses the standard GA definition [30]. It builds on the notion of individuals, i.e., candidate solutions of length k (*genes*), which are created as part of a population at the beginning of each generation. Subsequently, their fitness, i.e., reward, is evaluated, and the solutions with the highest fitness are then, with a certain probability, crossed

over and their genes mutated. The resulting *children* are then replicated to fill the population at the beginning of the next generation. Finally, once the number of generations or a certain saturation, i.e., no variability in best fitness, is achieved, the algorithm terminates.

3.6 Reinforcement Learning Theory

To formulate the E-PTNR problem as a RL problem, we consider a classical Markov Decision Process (MDP) defined as a tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, T, d_0, r, \gamma)$ [2]. Here, \mathcal{S} is a discrete set of states $s \in \mathcal{S}$ which represents which edges of E^{pt} have been removed. \mathcal{A} is the set of discrete actions $a \in \mathcal{A}$ which are the edges that can be removed. In our case, we work with deterministic dynamics as there is no randomness in our system dynamics T . Hence, given a state s and an action a , the next state s_{t+1} is deterministically chosen using the policy π . However, given the epsilon-greedy search, which drives our policy, we encounter randomness during training in the form of a conditional probability distribution $\pi(a_t | s_t)$. The initial state is defined by the complete graph \mathcal{G} without any removed edges. Finally, the reward is defined as a function $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ and $\gamma \in (0, 1]$ is a scalar discount factor [2].

Eventually, the goal of RL is to learn a policy $\pi(a_t | s_t)$, i.e. a distribution over actions conditioned on states, that maximizes cumulative expected reward over time. A standard RL algorithm using this paradigm is Q-learning [49]. While we compare the results from the Q-learning algorithm to our novel DMaxQN-learning approach, we deem it standard literature and include an extended discussion in Appendix F.

3.6.1 Max Q-Learning. is a recent addition the maximum reward RL literature [34] introduced by Gottipati et al. [18]. Standard RL techniques such as Q-learning has a crucial drawback with regard to the E-PTNR problem, namely the goal of maximizing the expected *cumulative* reward. With any budget k , the RL agent will attempt to find a policy that maximizes the cumulative public access equality over the edge removals. This objective is contrary to our initial intention of finding the most equitable reduction of the PTN within a budget k . However, more importantly, it conveys no meaning as there is no interpretation of cumulative public access equalities over removals, i.e., once a reduced PTN is reached, the rewards of previous removals are irrelevant.

The reformulation of the Q-value update presented by Gottipati et al. fulfills all requirements of convergence imposed by the MDP while maximizing for maximum reward instead of the cumulative reward:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left(\max \left(r_t, \gamma \max_a Q(s_{t+1}, a) \right) - Q(s_t, a_t) \right) \quad (8)$$

Here, as in standard Q-learning, the Q-function $Q(s_t, a_t)$ is tabular and is learned by the agent during training. We have highlighted in red the newly introduced maximization between r_t , i.e., the return at time t , and the discounted action leading to the maximum return in the next state $\gamma \max_a Q(s_{t+1}, a)$. This setup optimizes the maximum return, and an ϵ -greedy approach guides the policy (Appendix F).

With this formulation, we can address the E-PTNR problem and find the reduced PTN configuration leading to the maximum

equality of access for all groups within a finite budget $0 < k < |E^{\text{pt}}|$. To verify this claim, we compare the results yielded by our MaxQ- and Q-learning approaches in Section 5.3.

3.6.2 Deep (Max)Q Network-Learning. The prohibitive size of the state and action space of real-world E-PTNR settings requires us to approximate the Q-function as, in its tabular form, it is hard or impossible to hold it in memory. Such approximations using NNs have been used previously in the Q-learning setting [1]. However, it has not yet been attempted with the MaxQ-learning approach.

For our approach, we use the same setup as presented by mni [1] for DQN-learning, namely *experience replay* as well as a *policy-target-network*. The former is a simple, limited, in-memory dataset that allows the policy network to train on a set of previously visited states and the obtained returns. The latter is a densely connected NN that allows a regularized approximation of the true Q-function. The input of the policy network is the current state s , here represented as a binary vector, while the output is a vector $\mathbb{R}^{|E^{\text{pt}}|}$ representing the Q-values for each action.

4 EXPERIMENTS

First, we assess whether, numerically, the egalitarian objective devised in Section 3.2.2 reflects the intended behavior. Subsequently, we compare the performance of the different baselines on our synthetic datasets to the performances of our RL approaches. Finally, we illustrate the difference between Q-learning and MaxQ-learning on Dataset 2.2. All hyperparameters are reported in Appendix I.

4.1 Datasets

This section presents the dataset we use to assess the distribution of access equality in Amsterdam over 2019-2021. We present three synthetic datasets we use to validate our reward formulation and our (Max)Q- and D(Max)QN-learning approaches.

4.1.1 Amsterdam. To generate this dataset, we collect the General Transit Feed Specifications (GTFS)² data from the open Transit-Feeds³ platform, containing stops, routes, and schedules of Amsterdam's PT, from 01.01.2019 to 31.12.2021. We then merge this data with information on the geographic location of neighborhoods⁴, POIs⁵, the street-network⁶, and the census data into an E-PTNR conformant representation of the city of Amsterdam. See Section C.1 for the complete data preparation pipeline.

The resulting Amsterdam E-PTNR graph, as illustrated in Appendix C, Figure 12, consists of 561 PTS with 1601 connections of type *bus*, *ferry*, *metro*, *rail*, and *tram* between them. These are the edges in E^{pt} , hence the edge-set which our RL approach will attempt to reduce. Furthermore, in Appendix C, we also include a map visualizing the 443 neighbourhoods with their respective population-weighted RCs (Figure 12a) as well as a map illustrating 1193 POIs which indicate the location of educational opportunities across the city (Figure 12c). In combination with the pedestrian network, we can then use these data points to compute the Theil T

²<https://gtfs.org/>

³<https://transitfeeds.com/p/ov/814>

⁴<https://maps.amsterdam.nl/gebiedsindeling/?LANG=en>

⁵<https://maps.amsterdam.nl/functiekaart/?LANG=en>

⁶<https://download.geofabrik.de/europe/netherlands/noord-holland.html>

inequality, as defined in Section 3.2.2, based on the per-group travel metrics as presented in Section 3.2.1.

4.1.2 Synthetic. Real-world datasets in the E-PTNR domain are complex and span across a significant amount of RCs and POIs while also containing a big solution-space due to the size of the underlying PTN (see previous Section 4.1.1). This complexity inherently obscures the ability to reason about an algorithm’s performance. Thus, to numerically verify our reward formulation, baselines, and novel DMaxQ-learning approach, we devise here three synthetic datasets.

All three datasets presented in Figure 2 follow the E-PTNR specification (according to Section 3.1) and contain geographic information on RCs, POIs, as well as the respective PTN. Furthermore, we also created synthetic census data for each dataset, describing the distribution of **red** and **purple** inhabitants. This census data for all three datasets are reported in Section D.1.

For Dataset 2, we present two reward schemes. Dataset 2.1 is based on the egalitarian reward as defined in Equation 6. For Dataset 2.2, displaying the same graph configuration, we hard-code the rewards over the possible reductions as reported in Section D.3.2, Figure 15.

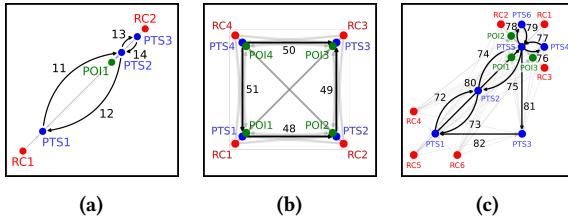


Figure 2: Three synthetic datasets.

Residential Centroids (RCs) in red, the Public Transport Stations (PTSs) in blue, and the Points of Interest (POIs) in green. Solid black lines indicate PTN connections by bus, tram, or train, while light gray lines indicate walking connections from RCs to POIs, RCs to PTSs, and PTSs to POIs. The synthetic census data can be found in Section D.1. a) Dataset 1: Simplest setup with advantaged red population (top right RC) and disadvantaged purple population (bottom left RC) b) Dataset 2: Equal population distribution across the RCs (500 purple / 500 red). Each PTS is close to its respective RC and POI. The shown topology is used for both Dataset 2.1 (Section D.3.1) and Dataset 2.2 (Section D.3.2). c) Dataset 3: Complex setup with one train and three bus lines. Disadvantaged purple RCs bottom left, advantaged red RCs top right Configurations and rewards: All possible reduced configurations and respective rewards see Appendix D.

4.2 Public Transport Equality in Amsterdam

In our related work section, we have justified the need for the formalization of the E-PTNR problem as well as the benefits of devising automated approaches to solving it (Section 2.2). To illustrate a real-world example of the impact of PTN reductions on the divide of access to socio-economic opportunities between groups, we analyze the inequality over three years (2019–2021) in Amsterdam. Specifically, we conduct two analyses on the inequality distribution between the population with *western* and *non-western* migration background.

First, we assess the week-by-week cumulative Theil T inequality, i.e., the summed inequalities across the entire city for each travel metric (Section 3.2), for the entire city of Amsterdam. We put this

inequality in relation to the reduction of the PTN size. This comparison allows us to reason about the impact of changes in PTN size on total inequality.

Second, we then reason about the accessibility distribution between the two groups, \mathbf{w} and \mathbf{nw} . We do so by ordering the per-neighborhood travel metrics by % \mathbf{w} population and fitting a linear model on the data points. The obtained correlation values (R^2) indicate the relation between the travel time and the percentage of \mathbf{w} population. As the percentages of each neighborhood’s \mathbf{w} and \mathbf{nw} populations add to 100%, 0% \mathbf{w} population indicates 100% \mathbf{nw} population and vice versa. We then report the week-by-week correlation values and put them again in relation to the PTN size.

4.3 Reward Validation Plan

Following the insight gathered on the necessity of the E-PTNR problem formulation in a real-world context, we validate the *egalitarian* reward defined in Section 3.2.2. We assess the resulting reward values for Dataset 1 and 2.2 as we progressively reduce the network.

For this and all following experiments we apply an exponential transformation to the egalitarian reward defined in Equation 6 (Section 3.2.2) to guarantee that $r_{\text{ega.}}(\mathcal{G}) \in [0, 100]$ and such that as $r_{\text{ega.}}(\mathcal{G}) \rightarrow 0$, the reward signal grows exponentially to give a strong weighting to PTN reductions which lead to higher equality:

$$r_{\text{ega.}}(\mathcal{G}) = 100 * e^{-5 * r_{\text{ega.}}(\mathcal{G})} \quad (9)$$

4.4 Baseline and (Deep) Reinforcement Learning Approaches

Once our egalitarian reward is validated, we apply each baseline and our (deep) Q- and MaxQ-learning algorithms to our three datasets, using the egalitarian reward as our optimization heuristic. We have chosen specific budgets to represent the most complex settings possible. For Dataset 1 that is $k_0 = 3$, for Dataset 2 it is $k_1 = 3$, and finally for Dataset 3 it is $k_3 = 9$. We show the optimal configurations of the respective PTN for these budgets in Appendix D.

As we present our E-PTNR approaches with problem datasets to which the solution is unknown during training, traditional evaluation metrics such as the Root Square Mean Error (RMSE) cannot be applied. Instead, we base our comparison on whether the algorithm in question finds the maximally equitable configuration under budget k and reward $r_{\text{ega.}}$ and whether its results are consistent over 15 different random seeds. While this approach is rather binary and limited, it allows for an initial comparison within our regulated environment which we then extend upon in our discussion.

We apply our (deep) Q- and MaxQ-learning approaches to all synthetic datasets. The hyperparameters for (Max)Q-learning are reported in Section I.3 and for D(Max)QN-learning in Section I.4, respectively. While we present the results on all datasets, we focus on Dataset 2.2 due to its hard-coded rewards and the resulting understandable Q-tables. Using this dataset, we show the differences between the policies learned by the Q- and MaxQ-learning agents and illustrate why the maximum reward RL approach is the most applicable to the E-PTNR problem set. Furthermore, we illustrate how the DQN- and DMaxQN-learning approaches reach the same optimal policies as their tabular-based counterparts. Finally, we

compare our DMaxQN-learning approach’s results to those provided by our baselines.

5 RESULTS

In this section, we first show that Amsterdam displays an inequality of socio-economic access between the two groups **w** and **nw**, disadvantaging the latter with nearly every PTN configuration between 2019 and 2021. Following, we show that our objective function correctly represents changes in public access equality depending on a PTN’s configuration and that it reflects the intended egalitarian behavior. Using the reward formulation, we then present our baseline, (Max)Q-learning, and D(Max)QN-learning results on our three synthetic datasets. Here we provide evidence that the standard RL approach is *not* applicable to the E-PTNR problem set, but rather the maximum reward RL approach offered by (D)MaxQ(N)-learning. Finally, we show how our DQN- and novel DMaxQN-learning approaches converge to the same true Q-tables as their tabular pendants.

5.1 Public Transport Equality in Amsterdam

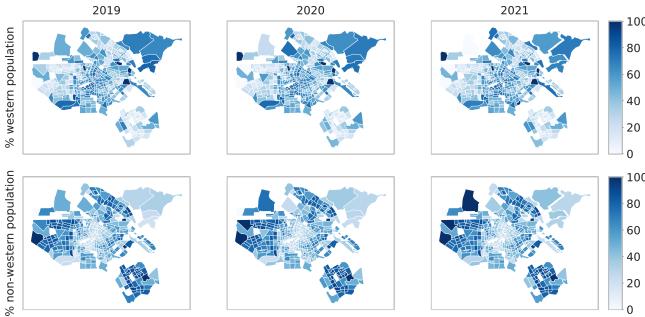


Figure 3: Per-neighborhood map of population with western (w) (top) and non-western (nw) (bottom) migration background for the years 2019, 2020, and 2021.

While the city center and the north-western neighborhoods of the city show a high density of **w** population, the peripheries of the city clearly show a higher **nw** population density.

In Figure 3 we show the spatial distribution of the **w** and **nw** groups from 2019 to 2021. We note that the spatial distribution of the two groups changes only marginally over time. Instead, the most apparent feature is the difference between the city’s center and outskirts, with a high % of **w** population in the former and a conversely high population % **nw** in the latter. This indicates that the two groups are spatially segregated and, implicitly, that the **w** population lives in an area with higher educational POIs (see Appendix C, Figure 12c).

In Figure 4, we then contrast the travel metrics (average travel-time, average hops, and 15-minute COM) to reach all educational POIs between the different neighborhoods for the year 2019 (for all years see Appendix G). Here we confirm our previous observation of the advantaged center in terms of lower average travel time, lower average hops, and a higher 15-minute COM. Implicitly, outskirts, i.e., areas with higher **nw** population density, are disadvantaged in terms of these travel metrics.

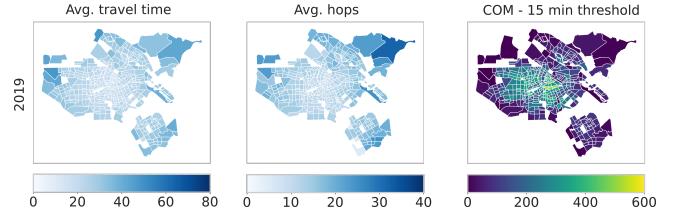


Figure 4: Per-neighborhood average travel time, avg. hops, and COM to all educational facilities in Amsterdam, displaying 2019.

For all distributions over the years 2019-2021 see Appendix G, Figure 5

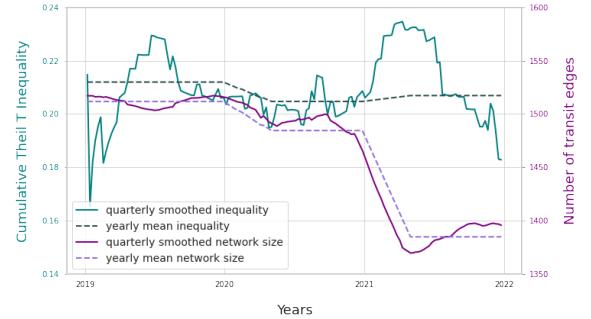


Figure 5: We show that the reduction of the Amsterdam Public Transport Network (PTN) relies more on the quality of removals than the quantity, as indicated by the cumulative Theil T inequality (where values closer to 0 indicate higher equality). We use a quarterly-smoothed Theil T inequality and PTN size (number of transit edges) sampled from 208 Mondays (07:00 am–09:00 am) in 2019–2021.

Left axis: Changes in access equality, evaluated using our definition of equality r , based on the Theil T and evaluated on travel time, hops, and the 15-minute Cumulative Opportunity Measure (COM). The considered journeys are from the Residential Centroids (RCs) to the Points of Interest (POIs) of type education. *Right axis:* The size, i.e., number of connections between Public Transport Stations (PTSs) of the PTN over the same period.

5.1.1 Public Transport Network-Induced Inequality. In Figure 5 we show how the quarterly smoothed cumulative Theil T inequality behaves over three years and in relation to fluctuations in the equally smoothed PTN’s size (E^{pt}). The data shows that while the size of a PTN reduction matters, as seen in mid-2019 and the first half of 2021, the quality, i.e., which transit edges are removed, can lead to much sharper and pronounced fluctuations in inequality. In this regard, we particularly note the beginning of 2019 and the second half of 2021, where changes in PTN size were marginal but fluctuations of equality significant. We also note that, although the reduction of the PTN in 2021 induced a rise in cumulative inequality, re-introducing a few edges led to a significant reduction.

In Figure 6 we then show how the two groups, **w** and **nw** relate to each other in terms of travel time over the same period of time. Here, we note that, on average, with a higher percentage of **w** population in a neighborhood, the average travel time (Figure 5) is low, implying that the average travel time is negatively correlated with the percent of **w** population.

This indicates that the **w** population experiences an, on average, *significantly* lower average travel time than the **nw** population.

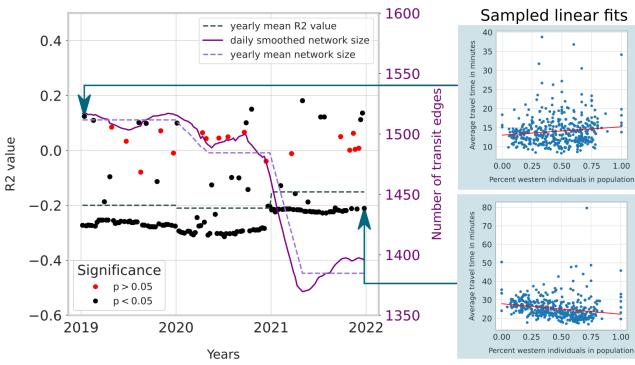


Figure 6:

Main plot (left): Left axis: R^2 values of linear fits on the distributions of travel time over the percentage of the population with western migration background over the years 2019-2021, sampled every Monday from the city of Amsterdam. The considered journeys are from the Residential Centroids (RCs) to Points Of Interest (POIs) of type education. Plots with the correlation of hops and COM over the same period are present in Appendix G. Right axis: Variation of Amsterdam's Public Transport Network (PTN) size over the years 2019-2021. Sub plots (right): Illustrates how linear fits are generated and how the R^2 values encode the correlation between the migration background and the travel time. Finally, the reduction of the PTN's average size in the year 2021 led to a notable increase of the R^2 value on average, indicating that while equality remained approximately equal (see Figure 5), the disadvantage experienced by the population group with non-western migration background reduced. 121 out of the 133 significant R^2 values ($\approx 91\%$) are below 0, i.e., indicating a higher average travel time for neighborhoods with higher percentages of the population with non-western migration background.

Furthermore, only 12 out of 25 values indicating positive correlation are statistically significant (i.e., the p-value of the fit is $p < 0.05$). Conversely, we find that in $\approx 91\%$ of the times, the nw population is disadvantaged. We report similarly significant observations on the other two travel metrics, avg. hops and 15-minute COM presented in Appendix G.

5.2 Reward Validation

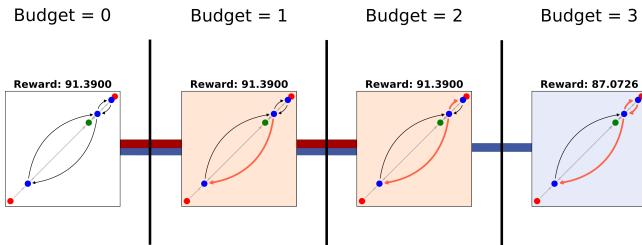


Figure 7: Dataset 1 egalitarian reward over removals up to budget $k = 3$.

Red dots indicate Residential Centroids (RCs), blue dots indicate Public Transport Stations (PTSs), and green dots indicate socio-economic Points Of Interest (POIs). Gray edges indicate modality WALK while black edges indicate modality BUS. Red edges are removed from the Public Transport Network (PTN). Configurations highlighted in orange represent maximum reward configurations, while blue indicates configurations leading to the highest cumulative reward. Edges connecting the configurations thus represent the paths taken by maximum cumulative reward and maximum reward algorithms, respectively. Higher rewards are better. For the visualization of all possible configurations see Section D.2

In Figure 7, we show how our devised reward function behaves with the reduction of the PTN in Dataset 1. As we regard only the equality of access to POIs, the PT edges leading from POIs back to RCs do not impact the equality measures, which is correctly reflected by the reward. Implicitly, with a budget $k = 1$ and $k = 2$, the edges from POIs to RCs are removed without penalty. With a budget of $k = 3$, the first two edge removals remain unchanged. However, the final removal is the edge from RC2 to the PTS close to POI1. As RC1 (bottom left) houses 1000 individuals of the **purple** inhabitants and RC2 (top right) 100 **red** inhabitants (see Section D.1), disadvantaging the smaller **red** population leads to a smaller increase of inequality than disadvantaging the **purple population** which is in-line with the egalitarian view of access equality we adopted.

While suggested by this example, the reduction inequality is not necessarily monotonously correlated with the reduction of the PTN, as shown by Dataset 2 (Section D.3.1). In the first instance, removing PT4 to PT1 leads to the highest reward (47.0568). However, with a budget of $k \geq 2$, a greedy approach would fail to reach the maximally equitable configuration, illustrating a non-monotonous behavior of the egalitarian reward.

5.3 Baselines & (Deep) Reinforcement Learning Results

In Table 1, we present the maximum rewards achieved by our baselines, and (deep) RL approaches on all synthetic datasets over 15 random seeds. We use the *optimal baseline*, i.e., the exhaustive search on all possible PTN reductions.

Starting from the random baseline, we note that the standard deviations for Dataset 1 and 2.2 are significantly higher than for 2.1 and 3. This is due to the chosen budget, $k = 3$ (1,2,1,2,2) and $k = 9$ (3) as well as the reward values (see Appendix D). While for Datasets 2.2 and 2.1, the mean (over runs) maximum rewards achieved by this algorithm are significantly lower than the optimum, for Dataset 1 and 3, the maximum rewards are within one standard deviation, hence frequently visited.

In contrast, our greedy approach shows no variation in its results as expected due to the static reward schemes. However, this approach fails when the reward is not monotonically decreasing, as in Dataset 2.1 and 2.2. This confirms our statement from Section 5.2.

The Q- and DQN-learning approaches show that, as expected, the maximization of the cumulative expected reward does not lead to the identification of the maximum reward reduction. However, as we show in Appendix D, the maximum reward reductions for Dataset 1 and 3 are part of the maximum cumulative reward solutions. Hence, the Q-learning algorithm identified the optimal reduction for Dataset 1 while DQN-learning identified both Dataset 1 and 3 in all runs.

The standard deviations over the runs are the most notable aspect of the Q- and DQN-learning results. DQN-learning converges in all runs and all datasets to the same results, although consistently missing the maximum cumulative reward in Dataset 2 (67.08 instead of 58.02). In comparison, Q-learning shows difficulties in reliably converging to the correct solutions with all datasets apart from Dataset 1.

Dataset	Random	Optimal	Greedy	GA	Q	MaxQ	DQN	DMaxQN
1	70.40 ± 43.99	91.39 ± 0.00						
2.1	59.46 ± 3.95	67.08 ± 0.00	58.02 ± 0.00	67.08 ± 0.00	65.87 ± 3.18	67.08 ± 0.00	58.02 ± 0.00	60.20 ± 4.30
2.2	53.33 ± 29.68	100.0 ± 0.00	50.0 ± 0.00	100.0 ± 0.00	50.0 ± 17.59	100.0 ± 0.00	50.00 ± 0.00	100.0 ± 0.00
3	71.94 ± 4.74	73.17 ± 0.00	73.17 ± 0.00	72.96 ± 0.78	70.72 ± 6.46	73.17 ± 0.00	73.17 ± 0.00	73.17 ± 0.00

Table 1: Performance of baseline and (deep) RL algorithms in terms of the maximum rewards achieved on Dataset 1, 2.1, 2.2, and 3 over 15 random seeds (Appendix J). The reward used for Dataset 1, 2.1, and 3 is the scaled egalitarian reward \hat{r}_{ega} , while the reward for Dataset 2.2 is hard-coded. Higher values are better and standard deviations smaller than 0.001 are indicated as 0.0. The hyperparameters for the respective algorithms are reported in Appendix I.

Finally, the GA baselines and the MaxQ- and DMaxQN-learning approaches behave nearly equally on all three datasets. MaxQ-learning outperforms the GA and the DMaxQN-learning as it converges to the maximum rewards on all datasets. However, this increased performance of the MaxQ-learning is only sustained by minimal deviations in the GA result on Dataset 3 and DMaxQN-learning result for Dataset 2.1.

5.4 D(Max)QN-Learning Convergence

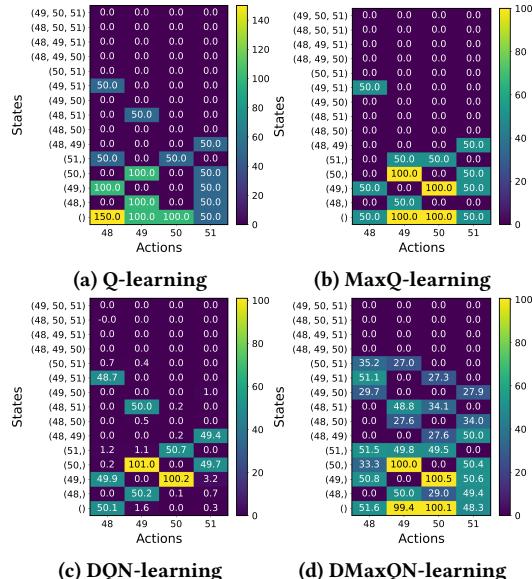


Figure 8: Q-tables learned by (a) Q-learning, (b) MaxQ-learning, (c) DQN-learning, and (d) DMaxQN-learning agents for Dataset 2.2 after 150 episodes.

Each numeric value is referred to as Q-value and describes the expected return for taking a certain action (x-axis) in a given state (y-axis). The states are represented as tuples of removed edges up to budget $k = 3$, while the actions are shown as the edge numbers to remove (see Figure 2 for an illustration of the edge-numbering). For (b) DQN-learning and (d) DMaxQN-learning, we constructed the table by extracting each Q-value tuple for a state from the policy NN while for (a) and (b), this table was created in-memory during training.

In Figure 8, we show the Q-tables learned by our different RL agents for Dataset 2.2 over 1,500 episodes to ensure value convergence. Starting from the Q-learning agent, we see that the Q-table correctly identifies the maximum *cumulative* return of 150. This

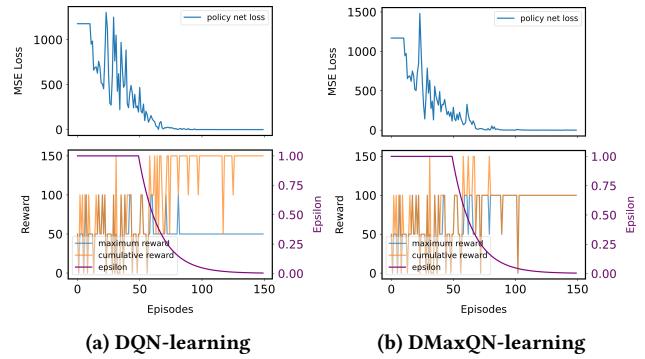


Figure 9: Policy network training losses (top) and agent cumulative and maximum reward as well as epsilon schedule (bottom) over 150 episodes on Dataset 2.2.

Left: The DQN-learning agent shows a clear convergence towards the maximum cumulative reward of 150 while the maximum reward remains constant at 50 as the Q-function converges (Dataset 2.2, removed edges: 48, 49, 51). *Right:* The DMaxQN-learning agent's cumulative and maximum rewards converge to the global maximum of 100 (Dataset 2.2, removed edges: 49, 50).

value is achieved by choosing edges 48, 49, 51 based on the maximum Q-values in each state, starting from the initial state. Note that the absolute values of the Q-tables are not required to match between our (Max)Q-learning and deep-learning approaches, but rather that the partial order of the values remains the same.

Contrasting the Q- and DQN-learning approaches are the MaxQ- and our novel DMaxQN-learning approaches. Despite considering the same dataset, we see that the Q-tables learned by the latter two reflect the path to the *maximum reward*, i.e., maximum egalitarian equality in our case, rather than maximizing the cumulative reward. For the MaxQ-learning agent, this results in the understanding that the optimal reduction for maximum return is to remove edges 49, 50 or 50, 49. The agent correctly identifies that the order in which edges are removed is irrelevant as the reward is specific to a *reduction of* the PTN (combinatorial) rather than the sequence of edges removed (permutational).

In Figure 9, we illustrate the convergence of the policy network loss and the cumulative and maximum rewards over 150 episodes on Dataset 2.2. Here, the difference between the DQN- and DMaxQN-learning approaches becomes apparent: the latter successfully identifies the most optimal, i.e., most equal reduction of the PTN within budget k (49, 50). Conversely, the DQN-learning approach correctly converges to the maximum cumulative reward of 150 (with the

removal of 48, 49, 51) while disregarding the configuration that maximizes the overall access equality.

6 DISCUSSION

In this section, we will elaborate on how our chosen methodology relates to existing works and how we interpret our results. First, we consider our novel formalization of the E-PTNR problem and the data used to quantify the access equality of PTNs.

To create an E-PTNR problem graph (Section 3.1), we rely on data describing the street and PT infrastructure as well as the available budget, similar to previous approaches [14]. Furthermore, as Ramachandran et al. [35], we additionally consider census data and socio-economic opportunities (POIs) and extend the author’s methods by including the population-weighted RCs as travel origins, similar to the works by Saxon and Snow [39] and [37] [37] as well as as the pedestrian connections for routing to and from POIs, RCs, and PTSs. Such a generalized formulation for equitable PTN interventions has not yet been presented in an extensive and replicable fashion.

Using this E-PTNR graph setup, we have then conducted our exploratory analysis of the inequality of access to educational opportunities in Amsterdam (Section 5.1). *We have shown that the size of PTN reductions and particularly the choice of removed segments can significantly impact access equality, resulting in significant variations of equality with minor changes of the underlying PTN size.* We highlight that, following a significant reduction in PTN size (Figure 5), the disadvantage of **new** individuals in terms of travel time reduced despite an increased overall inequality at the same time. We explain this observation by a significant increase in within-group inequalities in both groups and a reduction in between-group inequality, as shown by the increase of the R^2 values. These results align with previous research stating that certain population groups are significantly disadvantaged regarding access to socio-economic opportunities such as education and work [19, 28, 48].

Finally, we address our DMaxQN-learning results. While our novel approach did not outperform the GA nor its table-based pendant MaxQ-learning, it performed exceptionally well without dataset-specific hyperparameter tuning. This is not a trivial result as each dataset represents a completely different distribution of rewards and returns, which need to be approximated through gradient descent by the policy network. However, this finding is in line with previous findings by Wei et al. [50], namely that RL methods display cross-dataset applicability.

6.1 Limitations & Future Work

Following our discussion, we present our limitations and future work in a similar sequential layout. First, our mathematical formalization of the E-PTNR problem is limited to one type of socio-economic opportunity (POI). While we did so due to computational limitations, an E-PTNR representation accounting for multiple POIs could allow a more accurate representation of equality. Such a representation, combined with a reward formulation supporting group-necessity-driven weighting of these different POIs could result in an equity- rather than equality-driven optimization. This would lead to quantifying inequality closer to the population’s actual perception of it Carleton and Porter, Di Ciommo and Shiftan.

Second, in our work, we only consider an egalitarian reward regime, unlike other recent literature on equitable solutions to the PTNDP, which primarily focuses on utilitarian rewards [35, 50]. The exploration of E-PTNR settings considering utilitarian, elitarian, or welfare-based reward regimes could result in significantly diverse findings to ours without significant changes in the methodology.

Third, our analysis of the city of Amsterdam and our synthetic E-PTNR datasets only consider a two-group setup which is a significant limitation as it reduces the true diversity in urban spaces to a binary setting. Instead, a setup allowing for multiple groups based on features such as education, income, or age groups, as presented by Vinogradov in [48], could further add to a more representative approach.

Finally, a significant limitation of our work is that we were unable, due to time limitations, to apply our DMaxQN-learning approach to a real-world E-PTNR dataset such as the one generated for the city of Amsterdam. Together with external validation of the resulting solutions through tools such as Conveyal’s R5⁷, we hope to learn to which extent our presented methodology and novel DMaxQN-learning approach apply to real-world-sized datasets.

7 CONCLUSION

How can reinforcement learning be used to identify public transport network reduction under a budget while maintaining or improving access equality throughout the serviced area and its population?

In this thesis, we address the above research question in three parts. First, we formalize the Equitable Public Transport Network Reduction (E-PTNR) problem and present a comprehensive, reusable, graph-based representation of the required data. Furthermore, we define the quantification of access equality and formulate an egalitarian reward regime to guide optimization algorithms. Thus, we show how the E-PTNR problem can be formalized and how public access equality can be numerically quantified to inform optimization algorithms attempting to solve E-PTNR problems.

Second, we conduct an exploratory analysis on an E-PTNR representation of Amsterdam, showing that access to educational opportunities is unequally distributed across the two population groups: individuals with western and non-western migration backgrounds. More importantly, we show that even small reductions in the PTN can result in significant changes in access equality. Thus, we show evidence that current PTNR strategies lead to inequalities in accessibility and further justify our efforts in the formalization and algorithmic solution of the E-PTNR problem.

Third, we present a novel maximum Reinforcement Learning (RL) algorithm, Deep MaxQ Network (DMaxQN)-learning, which outperforms standard RL approaches and performs similarly well on our synthetic datasets as MaxQ-Learning and a Genetic Algorithm. This new formulation paves the way for applying maximum reward RL to bigger and more complex E-PTNR problems.

⁷<https://conveyal.com/>

ACKNOWLEDGMENTS

Thanks to Dimitris Michailidis for being an exceptional, always available, and constructively critical supervisor, and Fernando P. Santos for his exceptional teaching, incredible knowledge, and kindness. Thanks to Nicola Branchini for the uncountable hours of rubber-ducking, Nielja Knecht for being an incredibly supportive partner, my siblings for housing me when things got rough, and my parents for making the paths I have taken in the past possible in the first place.

Thank you - Riccardo

REFERENCES

- [1] 2013. Playing Atari with Deep Reinforcement Learning. <http://arxiv.org/abs/1312.5602> arXiv:1312.5602 [cs].
- [2] 2020. Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems. <http://arxiv.org/abs/2005.01643> arXiv:2005.01643 [cs, stat].
- [3] Anthony Barnes Atkinson and François Bourguignon. 2015. *Handbook of income distribution*. Handbooks in economics, Vol. 1. Amsterdam [u.a.].
- [4] Hamid Behbahani, Sobhan Nazari, Masood Jafari Kang, and Todd Litman. 2019. A conceptual framework to formulate transportation network design problem considering social equity criteria. *Transportation Research Part A: Policy and Practice* 125 (July 2019), 171–183. <https://doi.org/10.1016/j.tra.2018.04.005>
- [5] Bertrand Berche, Christian von Ferber, Taras Holovatch, and Yurij Holovatch. 2010. Public Transport Networks under Random Failure and Directed Attack. 1 (Feb. 2010), 1–10. <https://arxiv.org/abs/1002.2300v1> arXiv: 1002.2300.
- [6] Phillip R. Carleton and J. David Porter. 2018. A comparative analysis of the challenges in measuring transit equity: definitions, interpretations, and limitations. *Journal of Transport Geography* 72 (2018), 64–75. <https://doi.org/10.1016/j.jtrangeo.2018.08.012>
- [7] Oded Cats and Erik Jenelius. 2018. Beyond a complete failure: the impact of partial capacity gradation on public transport network vulnerability. *Transportmetrica B* 6, 2 (April 2018), 77–96. <https://doi.org/10.1080/21680566.2016.1267596> Publisher: Taylor & Francis.
- [8] Lidia Ceriani and Paolo Verme. 2012. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality* 10, 3 (Sept. 2012), 421–443. <https://doi.org/10.1007/s10888-011-9188-x>
- [9] Pedro NMII Conceicao and Pedro M. Ferreira. 2000. The Young Person's Guide to the Theil Index: Suggesting Intuitive Interpretations and Exploring Analytical Applications. *SSRN Electronic Journal* (2000). <https://doi.org/10.2139/ssrn.228703>
- [10] Yoran de Weert, Konstantinos Gkiotsalitis, Yoran de Weert, and Konstantinos Gkiotsalitis. 2021. A COVID-19 Public Transport Frequency Setting Model That Includes Short-Turning Options. *Future Transportation* 1, 1 (March 2021), 3–20. <https://doi.org/10.3390/futuretransp1010002> Publisher: Multidisciplinary Digital Publishing Institute.
- [11] Sybil Derrible and Christopher Kennedy. 2011. Applications of Graph Theory and Network Science to Transit Network Design. *Transport Reviews* 31, 4 (July 2011), 495–519. <https://doi.org/10.1080/01441647.2010.543709>
- [12] Floride Di Ciommo and Yoram Shiftan. 2017. Transport equity analysis. *Transport Reviews* 37, 2 (March 2017), 139–151. <https://doi.org/10.1080/01441647.2017.1278647> Publisher: Taylor & Francis.
- [13] Javier Durán-Micco, Pieter Vansteenwegen, Javier Durán-Micco, and Pieter Vansteenwegen. 2021. A survey on the transit network design and frequency setting problem. *Public Transport* 0123456789 (Oct. 2021). <https://doi.org/10.1007/s12469-021-00284-y>
- [14] Reza Zanjirani Farahani, Elnaz Miandoabchi, W.Y. Szeto, and Hannaneh Rashidi. 2013. A review of urban transportation network design problems. *European Journal of Operational Research* 229, 2 (Sept. 2013), 281–302. <https://doi.org/10.1016/j.ejor.2013.01.001>
- [15] Karst T. Geurs and Bert van Wee. 2004. Accessibility evaluation of land-use and transport strategies: Review and research directions. 12, 2 (June 2004), 127–140. <https://doi.org/10.1016/J.JTRANGEO.2003.10.005>
- [16] Konstantinos Gkiotsalitis and Oded Cats. 2020. Public transport planning adaptation under the COVID-19 pandemic crisis: literature review of research needs and directions. <https://doi.org/10.1080/01441647.2020.1857886> 41, 3 (2020), 374–392. <https://doi.org/10.1080/01441647.2020.1857886> Publisher: Routledge.
- [17] Transit API Google. 2020. Extended GTFS Route Types. <https://developers.google.com/transit/gtfs/reference/extended-route-types>
- [18] Sai Krishna Gottipati, Yashaswi Pathak, Rohan Nutall, Sahir, Raviteja Chunduru, Ahmed Touati, Sriram Ganapathi Subramanian, Matthew E. Taylor, and Sarath Chandar. 2020. Maximum Reward Formulation In Reinforcement Learning. (Oct. 2020). <http://arxiv.org/abs/2010.03744> Number: arXiv:2010.03744 arXiv:2010.03744 [cs, stat].
- [19] Diego Hernandez. 2018. Uneven mobilities, uneven opportunities: Social distribution of public transport accessibility to jobs and education in Montevideo. *Journal of Transport Geography* 67, September 2017 (2018), 119–125. <https://doi.org/10.1016/j.jtrangeo.2017.08.017> Publisher: Elsevier.
- [20] Sabiou Inoua. [n.d.]. Beware the Gini Index! A New Inequality Measure. ([n. d.], 26).
- [21] Erik Jenelius and Lars-Göran Mattsson. 2021. Resilience of Transport Systems. *International Encyclopedia of Transportation* (2021), 258–267. <https://doi.org/10.1016/b978-0-08-102671-7.10719-5>
- [22] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2015).
- [23] Marc Kruyswijk. 2021. Financiële problemen amsterdams ov: mogelijk minder bussen en trams 2021. *Parool.nl* (Feb. 2021). <https://www.parool.nl/amsterdam/financiele-problemen-amsterdams-ov-mogelijk-minder-bussen-en-trams-b11349pd/>
- [24] Marc Kruyswijk. 2022. Major cuts to Amsterdam's public transport under threat: fewer buses, streetcars and metros – Grote bezuiniging op Amsterdams ov dreigt: minder bussen, trams en metro's. *Het Parool* (June 2022). <https://www.parool.nl/amsterdam/grote-bezuiniging-op-amsterdams-ov-dreigt-minder-bussen-trams-en-metro-s-bed742c8/>
- [25] Gilbert Laporte and Marta M.B. Pascoal. 2015. Path based algorithms for metro network design. *Computers & Operations Research* 62 (Oct. 2015), 78–94. <https://doi.org/10.1016/j.cor.2015.04.007> Publisher: Pergamon.
- [26] Dinah Jane Lope and Anil Dolgun. 2020. Measuring the inequality of accessible trams in Melbourne. *Journal of Transport Geography* 83 (Feb. 2020), 102657. <https://doi.org/10.1016/j.jtrangeo.2020.102657>
- [27] Karel Martens. 2016. *Transport Justice*. Routledge, New York. <https://doi.org/10.4324/9781315746852>
- [28] Cristhian Figueiroa Martinez, Frances Hodgson, Caroline Mullen, and Paul Timms. 2018. Creating inequality in accessibility: The relationships between public transport and social housing policy in deprived areas of Santiago de Chile. *Journal of Transport Geography* 67, July 2016 (Feb. 2018), 102–109. <https://doi.org/10.1016/j.jtrangeo.2017.09.006> Publisher: Elsevier.
- [29] Dimitris Michailidis, Sennay Ghebreab, and Fernando P Santos. [n.d.]. Optimizing Fairness in Transport Network Design using Deep Reinforcement Learning. ([n. d.], 5).
- [30] Seyedali Mirjalili. 2019. Genetic Algorithm. In *Evolutionary Algorithms and Neural Networks: Theory and Applications*. Springer International Publishing, Cham, 43–55. https://doi.org/10.1007/978-3-319-93025-1_4
- [31] Carlos Moreno, Zaheer Allam, Didier Chabaud, Catherine Gall, and Florent Pratlong. 2021. Introducing the “15-Minute City”: Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities* 4, 1 (Jan. 2021), 93–111. <https://doi.org/10.3390/smartcities4010006>
- [32] Natalie Mueller, David Rojas-Rueda, Haneen Khreis, Marta Cirach, David Andrés, Joan Ballester, Xavier Bartoll, Carolyn Daher, Anna Deluca, Cynthia Echave, Carles Milà, Sandra Márquez, Joan Palou, Katherine Pérez, Cathryn Tonne, Mark Stevenson, Salvador Rueda, and Mark Nieuwenhuijsen. 2020. Changing the urban design of cities for health: The superblock model. *Environment International* 134 (2020), 105132. <https://doi.org/10.1016/j.envint.2019.105132>
- [33] Lars Osberg. 2017. On the Limitations of Some Current Usages of the Gini Index: On the Limitations of Some Current Usages of the Gini Index. *Review of Income and Wealth* 63, 3 (Sept. 2017), 574–584. <https://doi.org/10.1111/roiw.12256>
- [34] K.H. Quah and Chai Quek. 2006. Maximum reward reinforcement learning: A non-cumulative reward criterion. *Expert Systems with Applications* 31, 2 (Aug. 2006), 351–359. <https://doi.org/10.1016/j.eswa.2005.09.054>
- [35] Govardhana Sachiathanandam Ramachandran, Ivan Brugere, Lav R. Varshney, and Caiming Xiong. 2021. GAEA: Graph Augmentation for Equitable Access via Reinforcement Learning. In *AIES 2021 - Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, Inc, 884–894. <https://doi.org/10.1145/3461702.3462615> arXiv: 2012.03900.
- [36] Daniel Rhoads, Albert Solé-Ribalta, Marta C. González, and Javier Borge-Holthoefer. 2021. A sustainable strategy for Open Streets in (post)pandemic cities. *Communications Physics* 4, 1 (Dec. 2021). <https://doi.org/10.1038/s42005-021-00688-z> Publisher: Nature Research.
- [37] Tayebeh Saghpour, Sara Moridpour, and Russell G. Thompson. 2016. Public transport accessibility in metropolitan areas: A new approach incorporating population density. *Journal of Transport Geography* 54 (June 2016), 273–285. <https://doi.org/10.1016/j.jtrangeo.2016.06.019> Publisher: Elsevier Ltd.
- [38] Thomas W. Sanchez. 2002. The Impact of Public Transport on US Metropolitan Wage Inequality. *Urban Studies* 39, 3 (March 2002), 423–436. <https://doi.org/10.1080/00420980220112766>
- [39] James Saxon and Daniel Snow. 2020. A Rational Agent Model for the Spatial Accessibility of Primary Health Care. *Annals of the American Association of Geographers* 110, 1 (Jan. 2020), 205–222. <https://doi.org/10.1080/24694452.2019.1629870> SUPPL_FILE/RAAG_A_1629870_SM8256.DOCX Publisher: Taylor and Francis Ltd.
- [40] Andreas Schafer and David Victor. 1997. The Past and Future of Global Mobility. *Scientific American* 277, 4 (1997), 58–61. <http://www.jstor.org/stable/24995951>

- Publisher: Scientific American, a division of Nature America, Inc.
- [41] Jan Scheurer. 2013. Measuring Copenhagen's public transport accessibility and network performance in a European context. *Annual Transport Conference* (2013), 1–20.
 - [42] A. F. Shorrocks. 1980. The Class of Additively Decomposable Inequality Measures. *Econometrica* 48, 3 (April 1980), 613. <https://doi.org/10.2307/1913126>
 - [43] Richard S. Sutton and Andrew G. Barto. 2020. *Reinforcement learning: an introduction* (second edition ed.). The MIT Press, Cambridge, Massachusetts.
 - [44] Luca Trevisan. 2011. CS261 Lecture 3: TSP and Eulerian Cycles. <https://lucatrevisan.wordpress.com/2011/01/19/cs261-lecture-3-tsp-and-eulerian-cycles/>
 - [45] Job van Eldijk, Jorge Gil, Natalia Kuska, and Rashmita Sisinty Patro. 2020. Missing links – Quantifying barrier effects of transport infrastructure on local accessibility. *Transportation Research Part D: Transport and Environment* 85 (Aug. 2020). <https://doi.org/10.1016/j.trd.2020.102410> Publisher: Elsevier Ltd.
 - [46] Bert van Wee and Niek Mouter. 2021. Evaluating transport equity. In *Advances in Transport Policy and Planning*. Vol. 7. Elsevier, 103–126. <https://doi.org/10.1016/bs.atpp.2020.08.002>
 - [47] Giovanni Vecchio, Ignacio Tiznado-Aitken, and Ricardo Hurtubia. 2020. Transport and equity in Latin America: a critical review of socially oriented accessibility assessments*. *Transport Reviews* 40, 3 (May 2020), 354–381. <https://doi.org/10.1080/01441647.2020.1711828> Publisher: Taylor & Francis.
 - [48] R.A. Vinogradov. 2019. *Assessing the accessibility changes for public transport dependent user groups: An analysis of Amsterdam's newly adopted public transport system*. Ph.D. Dissertation. Wageningen University and Research Centre, Wageningen.
 - [49] C. J. C. H. Watkins. 1989. *Learning from Delayed Rewards*. PhD Thesis. King's College, Oxford.
 - [50] Yu Wei, Minjia Mao, Xi Zhao, Jianhua Zou, and Ping An. 2020. City Metro Network Expansion with Reinforcement Learning. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Aug. 2020), 2646–2656. <https://doi.org/10.1145/3394486.3403315> Publisher: Association for Computing Machinery ISBN: 9781450379984.
 - [51] Peter White. 2016. *Public transport: its planning, management and operation* (6 ed.). Routledge, London. <https://doi.org/10.4324/9781315675770>
 - [52] D.H. Wolpert and W.G. Macready. 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1, 1 (April 1997), 67–82. <https://doi.org/10.1109/4235.585893>
 - [53] Yunhan Zheng, Shenhao Wang, and Jinhua Zhao. 2021. Equality of opportunity in travel behavior prediction with deep neural networks and discrete choice models. *Transportation Research Part C: Emerging Technologies* 132 (Nov. 2021), 103410. <https://doi.org/10.1016/j.trc.2021.103410>

Appendix A TRAVEL METRIC DISTRIBUTIONS

In Figure 10 we show an example of how the distribution of a travel metric such as travel time is sampled across neighborhoods. We first use the RCs as departing locations to reach POI1 and compute the shortest path using Dijkstra's Algorithm. Based on this path's travel time (for example), we then sample and group the number of inhabitants of each group, here groups **purple** and **red**, in each RC, and assign them their neighborhood's travel metric value. This leads to a city-wide, grouped metric vector representing the distribution of a travel metric over the entire city.

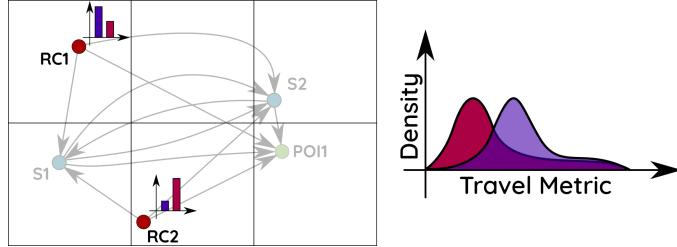


Figure 10: Visualization of travel time distribution per Residential Centroid (RC) and a group of inhabitants (here groups red and purple).

Left: The distribution of the two groups in terms of residents at the respective RC. Right: The city's cumulative distribution of the per-individual travel metric, such as average travel times, is split up into respective groups.

Appendix B THEIL T INDEX

Here we present an extensive definition of the Theil T inequality measure used in our work and additional mathematical background to the index.

Two Theil indices of inequality exist, namely Theil L or the Theil T index [42]. Both inequality indices are special cases of the General Entropy (GE) Equation 10. In its formulation, the parameter α weights the travel metric disparities across the travel metric distribution. Lower values lead to higher sensitivity in changes to the distribution's lower tail, whereas higher ones in changes to the upper tail [3, 9]. With $\alpha = 1$, we obtain Theil's T (Equation 11) which we use in our quantification of inequality due to its inherent sensitivity to worse travel metrics (higher travel time and hops).

$$GE(\alpha) = \frac{1}{\alpha(\alpha - 1)} \frac{1}{n} \sum_{i=1}^n \left(\left(\frac{y_i}{\mu} \right)^\alpha - 1 \right) \quad (10)$$

$$GE(1) = T_T(\mathbf{y}; n) = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\mu} \ln \left(\frac{y_i}{\mu} \right) \quad (11)$$

In Equation 11, Let $\mathbf{y} = (y_1, \dots, y_n)$ be the measure distribution vector for a population of n individuals and $\mu = \sum_{i=1}^n y_i / n$ the mean. As an example of the measure distribution vector, assume a city with $n = 10$ individuals living in pairs in different neighborhoods. We then, by the per-neighborhood travel metric computation, for example average travel time, obtain: $m_{avg.tt}(RC_1) = 1, m_{avg.tt}(RC_2) = 2, \dots, RC_5 = m_{avg.tt}(5)$. Now, we sample the respective travel metric for each individual in the population, resulting in a vector $\mathbf{y} = (1, 1, 2, 2, \dots, 5, 5)$ which is the input to the Theil T measure in combination with its size, i.e., n .

With a total population of size N , we have the boundaries $T \in [0, \ln N]$. Here, 0 represents perfect equality, i.e., every individual in the population has the same share of a certain metric (such as income), while $\ln N$ is the upper bound for perfect inequality.

Finally, the Theil T index can be decomposed into a within-group and between-group inequality. In Equation 16, \mathbf{y} is partitioned into G disjoint subgroups with $n_g \geq 1$, resulting in $\mathbf{y}^g = (y_1^g, \dots, y_{n_g}^g)$ and mean $\mu_g = \sum_{i=1}^{n_g} y_i / n_g$. We refer to the collection of the grouped measure distribution vectors as the matrix $\mathbf{Y} = [\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^G]$ (as presented in Section 3.2.2). By using this partitioning, Theil T can be written as in Equation 16 below.

We first compute each travel metric m for all RCs in V^{rc} to generate a quantification of the observed group-specific access metrics. Subsequently, using the census data, we assign each inhabitant of each RC the computed travel metric. By then, using the information on group membership, we first create the group-specific observed distribution vector:

$$\hat{\mathbf{y}}_m^g = \left[(m(v_1^{rc}))_{\times n_{v_1^{rc}}^g}, (m(v_2^{rc}))_{\times n_{v_2^{rc}}^g}, \dots, (m(v_s^{rc}))_{\times n_{v_s^{rc}}^g} \right] \quad (12)$$

Here, n_g is the total number of individuals in group $g \in G$, $n_{v_i^{rc}}^g$ is the number of individuals of group g at a specific RC v_i^{rc} , and $s = |V^{rc}|$.

B.1 Decomposition

As we were not able to find in any of the literature the formal (and complete) derivation of the decomposed Theil T index, we here show the steps omitted in [42]:

Proposition 1: The Theil T index applied to a partitioned distribution vector $\mathbf{y}^g = (y_1^g, \dots, y_{n_g}^g)$ can be decomposed into:

$$T(Y; n) = T(\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^G; n) \quad (13)$$

$$= \frac{1}{n} \sum_g \sum_{i=1}^{n_g} \frac{y_i^g}{\mu} \log \frac{y_i^g}{\mu} \quad (14)$$

$$= \sum_g \frac{n_g \mu_g}{n \mu} T(\mathbf{y}^g; n_g) + \sum_g \frac{n_g \mu_g}{n \mu} \log \frac{\mu_g}{\mu} \quad (15)$$

$$= \sum_g v_g T(\mathbf{y}^g; n_g) + \sum_g v_g \log \frac{\mu_g}{\mu} \quad (16)$$

$$= \textcolor{red}{T_{WG}} + \textcolor{blue}{T_{BG}} \quad (17)$$

$$\text{where } v_g = \frac{n_g \mu_g}{n \mu} \quad (18)$$

Here, we indicate the **within-group component** T_{WG} in red (first term), while the **between-group component** T_{BG} is in red (second term).

Proof:

$$T_T(\mathbf{y}; n) = \sum_g v_g T_T(\mathbf{y}^g; n_g) + \sum_g v_g \log \frac{\mu_g}{\mu} \quad (19)$$

$$= \sum_g \frac{n_g \mu_g}{n \mu} T_T(\mathbf{y}^g; n_g) + \sum_g \frac{n_g \mu_g}{n \mu} \log \frac{\mu_g}{\mu} \quad (20)$$

$$= \sum_g \frac{n_g \mu_g}{n \mu} \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \frac{y_i^g}{\mu_g} \ln \left(\frac{y_i^g}{\mu_g} \right) \right) + \frac{1}{n} \sum_g n_g \frac{\mu_g}{\mu} \ln \left(\frac{\mu_g}{\mu} \right) \quad (21)$$

$$= \sum_g \frac{1}{n \mu} \cancel{\frac{n_g \mu_g}{n_g \mu_g}} \left(\sum_{i=1}^{n_g} y_i^g \ln \left(\frac{y_i^g}{\frac{1}{n_g} \sum_{i=1}^{n_g} y_i^g} \right) \right) + \frac{1}{n} \sum_g \cancel{y_g} \frac{\cancel{\frac{1}{n_g} \sum_{i=1}^{n_g} y_i^g}}{\mu} \ln \left(\frac{\cancel{\frac{1}{n_g} \sum_{i=1}^{n_g} y_i^g}}{\mu} \right) \quad (22)$$

$$= \frac{1}{n \mu} \sum_g \sum_{i=1}^{n_g} y_i^g \ln \left(y_i^g \right) - y_i^g \ln \left(\frac{1}{n_g} \sum_{i=1}^{n_g} y_i^g \right) + \frac{1}{n \mu} \sum_g \sum_i y_i^g \ln \left(\frac{1}{n_g} \sum_{i=1}^{n_g} y_i^g \right) - y_i^g \ln (\mu) \quad (23)$$

$$= \frac{1}{n \mu} \sum_g \sum_{i=1}^{n_g} y_i^g \ln \left(y_i^g \right) - \cancel{y_i^g \ln \left(\frac{1}{n_g} \sum_{i=1}^{n_g} y_i^g \right)} + \cancel{y_i^g \ln \left(\frac{1}{n_g} \sum_{i=1}^{n_g} y_i^g \right)} - y_i^g \ln (\mu) \quad (24)$$

$$= \frac{1}{n \mu} \left[\sum_g \sum_i y_i^g \ln \left(y_i^g \right) - y_i^g \ln (\mu) \right] \quad (25)$$

$$= \frac{1}{n} \sum_g \sum_i \frac{y_i^g}{\mu} \ln \left(\frac{y_i^g}{\mu} \right) \blacksquare \quad (26)$$

Appendix C AMSTERDAM DATASET

C.1 Data Preparation Pipeline

In Figure 11, we illustrate the pipeline used to extract the necessary components of the Amsterdam dataset. We have used four different data sources, which our pipeline, part of our e-ptnr Python3 package, wrangles into one E-PTNIR graph representation. In Figure 12, we illustrate the dataset resulting from the devised pipeline, split into its components, namely RCs, POIs, and the PTN of Amsterdam in the year 2022.

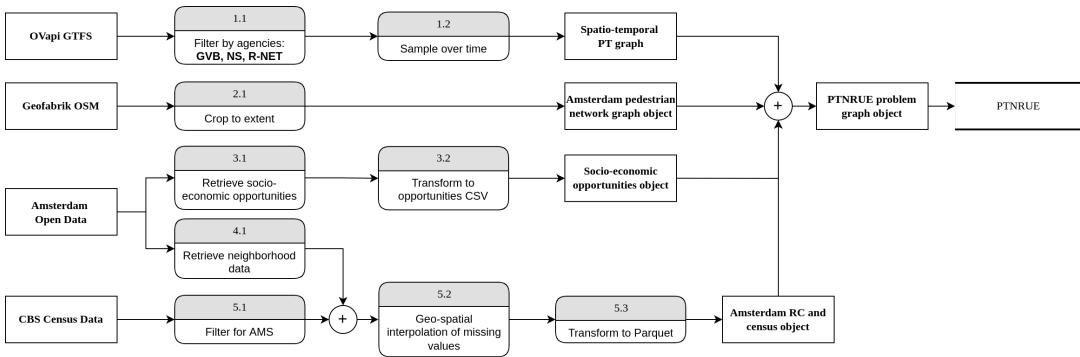


Figure 11: Pipeline for the extraction and preparation of the E-PTNR problem graph based on real-world data for the city of Amsterdam

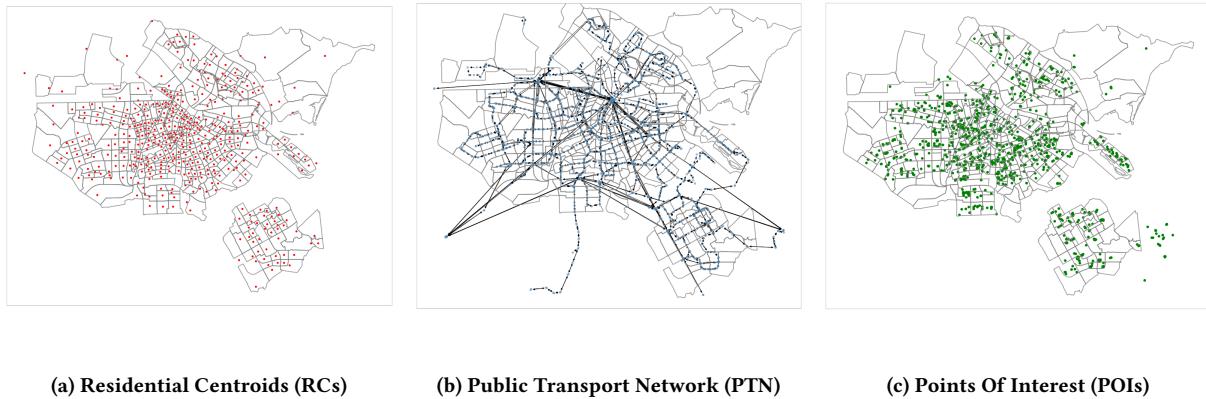


Figure 12: Map of Amsterdam neighborhoods and educational opportunities with the local Public Transport Network (PTN) inscribed, illustrative (state of 2022).

a) 443 red points represent the population-weighted Residential Centroids (RCs) of each neighborhood. These are used as departing locations for the shortest path generation. b): The blue points represent 561 Public Transport Stations (PTSs) while the 1193 lines connecting them are multi-modal connections by *bus, ferry, metro, rail, and tram.* c) 1193 green points indicating the locations of educational POIs such as university and school buildings.

Appendix D SYNTHETIC DATASETS

D.1 Population Specifications

Dataset	RC	n_inh total	n_inh purple	n_inh red
1	1	1000	1000	0
	2	100	0	100
2.1 & 2.2	1	1000	500	500
	2	1000	500	500
	3	1000	500	500
	4	1000	500	500
3	1	1000	100	900
	2	1000	200	800
	3	1000	200	800
	4	1200	1000	200
	5	800	600	200
	6	1000	1000	0

Table 2: Synthetic census data for the three synthetic datasets defined in Section 4.1.2

D.2 Dataset 1

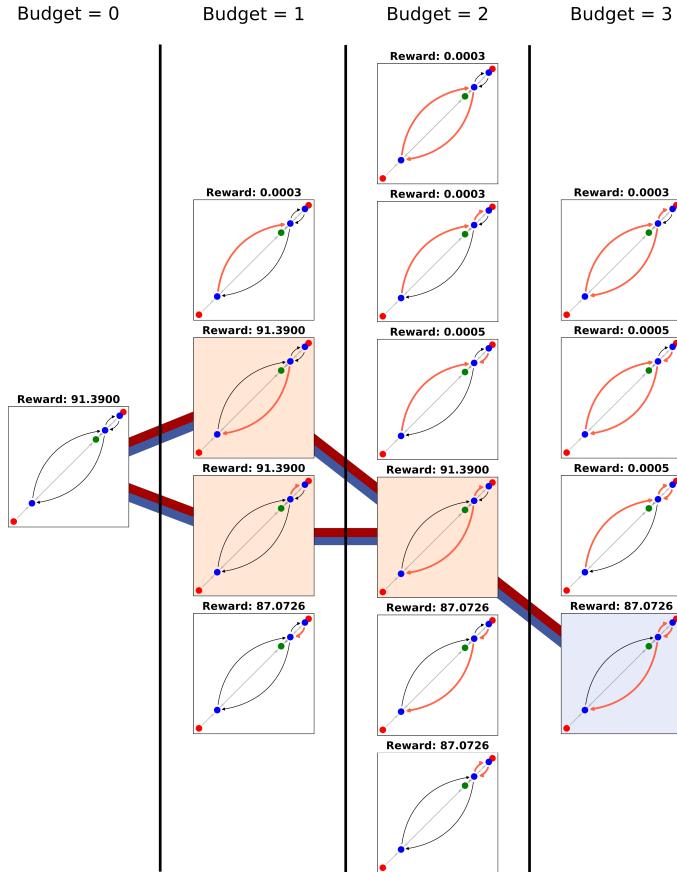


Figure 13: Caption

D.3 Dataset 2

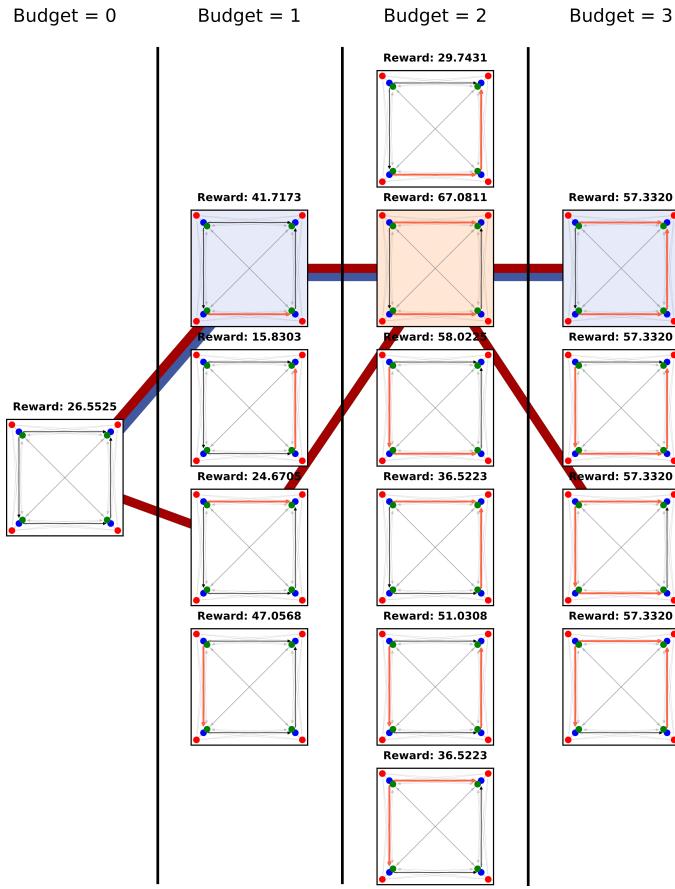


Figure 14: Caption

D.3.1 Dataset 2.1.

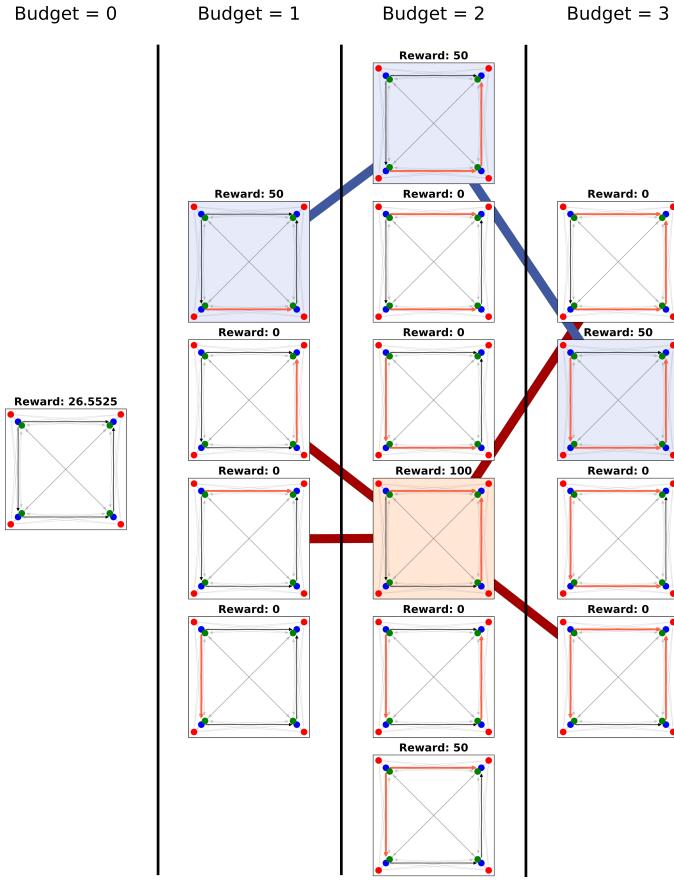


Figure 15: Caption

D.3.2 *Dataset 2.2.*

D.4 Dataset 3

With $|E^{pt}| = 11$, this synthetic dataset is too big to illustrate all possible combinations of removed edges. However, in Figure 16, we show the edges which can be removed, in any order, without a penalty in reward. Here, the egalitarian reward is, for all 8 of the 11 edges removed, 73.1734.

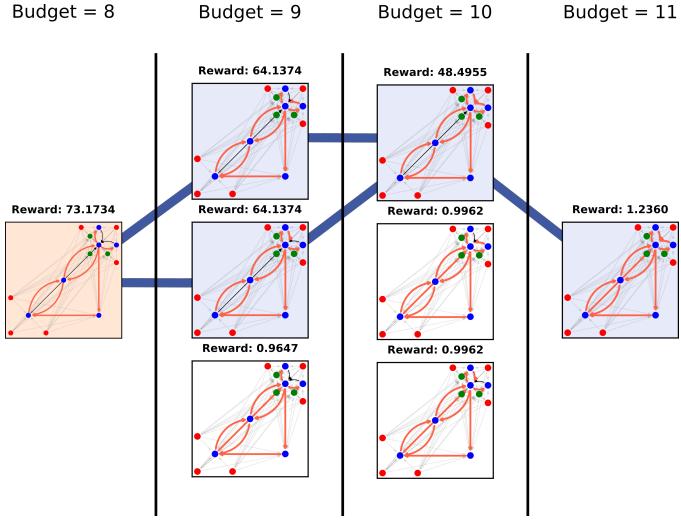


Figure 16: Caption

Appendix E BASELINE ALGORITHMS

Algorithm 1: Random Baseline

```

Input: Graph  $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ , budget  $k \geq 1$ 
Output: A set of edges to remove  $\mathcal{E}$ 
 $\mathcal{E} := \emptyset$  for  $i \leftarrow 0$  to  $k$  do
     $e \leftarrow \text{sample\_wo\_replacement}(E_m, m \in M \setminus \{\text{WALK}\})$ 
     $\mathcal{E} \leftarrow \text{append}(\mathcal{E}, e)$ 
end

```

Algorithm 2: Fully Informed, Greedy Baseline

```

Input: Graph  $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ , budget  $1 \leq k < |E_{M \setminus \{\text{WALK}\}}|$ , reward function  $r$ 
Output: A set of edges to remove  $\mathcal{E}$ 
 $\mathcal{E} := \emptyset$ 
for  $i \leftarrow 0$  to  $k$  do
    /* Evaluate reward of all graphs with one PTN edge removed */
     $e \leftarrow \underset{e_m \in E_{M \setminus \{\text{WALK}\}}}{\text{argmax}} (r(\mathcal{G} - e_m))$ 
     $\mathcal{E} \leftarrow \text{append}(\mathcal{E}, e)$ 
end

```

Algorithm 3: Genetic Algorithm

Input: Graph $\mathcal{G} = \langle V, E \rangle$, budget $1 \leq k < |E_{M \setminus \{\text{WALK}\}}|$, reward function r
Output: A set of edges to remove \mathcal{E}
 $(\mathcal{R} \leftarrow \emptyset, \mathcal{X} \leftarrow \emptyset)$
 $L \leftarrow \phi_{ME}$ **for** $iter = 0$ **to** I **do**
 if $iter < L$ **then**
 | $\mathbf{x}' \leftarrow \text{random_solution}()$
 else
 | $\mathbf{x} \leftarrow \text{random_selection}(\mathcal{X})$
 | $\mathbf{x}' \leftarrow \text{random_variation}(\mathbf{x})$
 end
 | $\mathbf{b}' \leftarrow \text{feature_descriptor}(\mathbf{x}')$
 | $r' \leftarrow r(\mathbf{x}')$
 if $\mathcal{R}(\mathbf{b}') = \emptyset$ **or** $\mathcal{R}(\mathbf{b}') < p'$ **then**
 | $\mathcal{R}(\mathbf{b}') \leftarrow p'$ $\mathcal{X}(\mathbf{b}') \leftarrow \mathbf{x}'$
 end
end

Appendix F Q-LEARNING

This section represents a continuation of the introduction of standard RL in Section 3.6. As alluded to in the main text, Q-learning was first introduced by Watkins [49]. It guarantees that the resulting policy is optimal with infinite episodes, i.e., attempts at finding a path from a start state to a final form using a finite set of discrete actions. To conduct this optimization, the agent explores the environment and, while doing so, updates the Q-values, indicating the expected return for a specific state-action pair. This leads to a tabular representation of the optimal policy. At test time, the agent follows the Q-table (Q-function) deterministically, i.e., in each state, the next action with the highest expected return is chosen. Finally, this leads to the maximum expected cumulative reward.

The exploration and exploitation of the agent during training are guided through a regime on the randomness of action choice at each step. In our case, we rely on an ϵ -greedy approach, which is the most used in literature [43]. Our chosen schedule for the regime consists of a static $\epsilon_0 > 0$ up until a specified episode, followed by an exponential decrease to $0 < \epsilon_1 < \epsilon_0$ which remains constant until the final episode. At test time, we set $\epsilon = 0$ to deterministically obtain the solution of edge removals with the highest cumulative reward.

Appendix G AMSTERDAM EQUALITY ANALYSIS

In Figure 17 (average hops) and Figure 18 (15 minute COM), we present the results on the **w** and **nw** travel metric correlation analysis presented in Section 5.1. We see a very similar behavior as with average travel time in the hops, namely that the percentage of **w** population in a neighborhood is negatively correlated with the travel metric. In the 15-minute COM, however, we see that there is a positive correlation with the % **w** population which can be interpreted as this group having access to more opportunities within 15 minutes than the **nw** group.

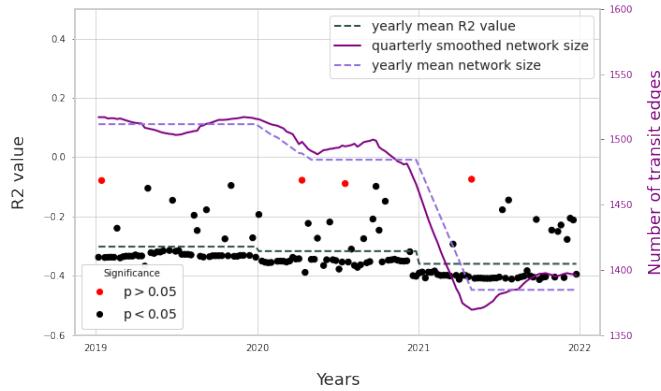


Figure 17: R^2 values of the linear fits on the distributions of hops over the percentage of the population with western migration background. The considered journeys are from the RCs to the POIs of type education. Similar to the travel time-based version of this plot in Figure 3, we see a clear disadvantage imposed on the non-western population. In this case, the population with western migration background required, on average, fewer hops than the one with a non-western migration background. This is the case in 100% of the examined period (146/146 statistically significant R^2 values).

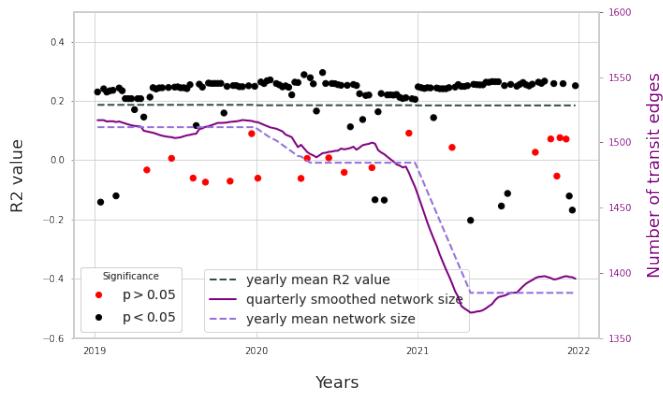


Figure 18: R^2 values of the linear fits on the distributions of 15-minute COM over the percentage of the population with western migration background. The considered journeys are from the RCs to the POIs of type education. Similar to Figure 3 and Figure 17, we here too see evidence for the same disadvantage. In this case, however, the correlation with a higher percentage of the population with a western migration background is positive. This indicates that neighborhoods with a higher percentage of individuals with western migration background tend to have access to more socio-economic opportunities, in this case, education, within 15 minutes of travel. This is the case in $\approx 93\%$ of the examined period (122/131 statistically significant R^2 values).

Appendix H OBJECTIVE VALIDATION

H.1 Dataset 1

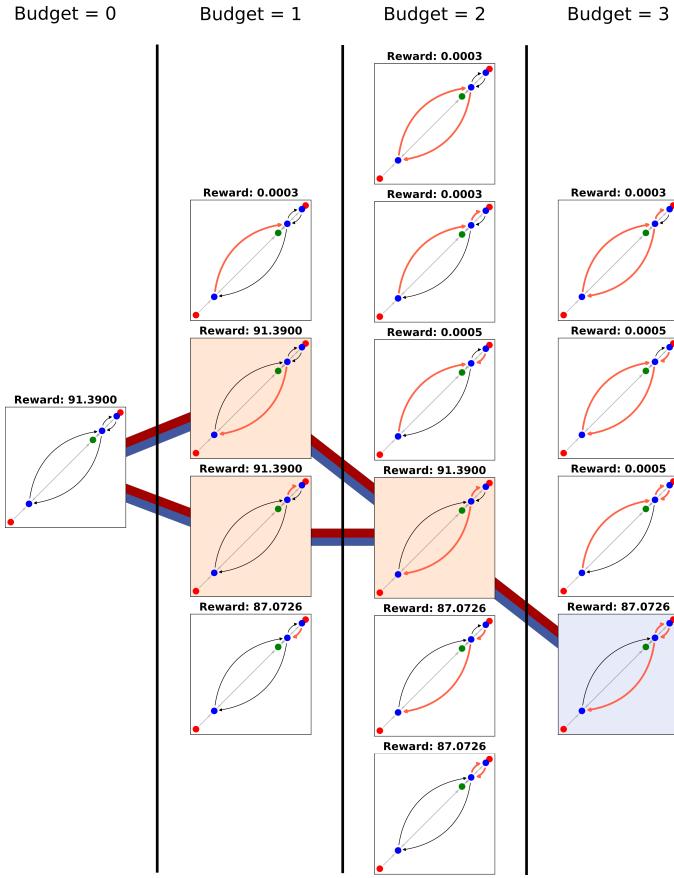


Figure 19: Caption

H.2 Dataset 2.2

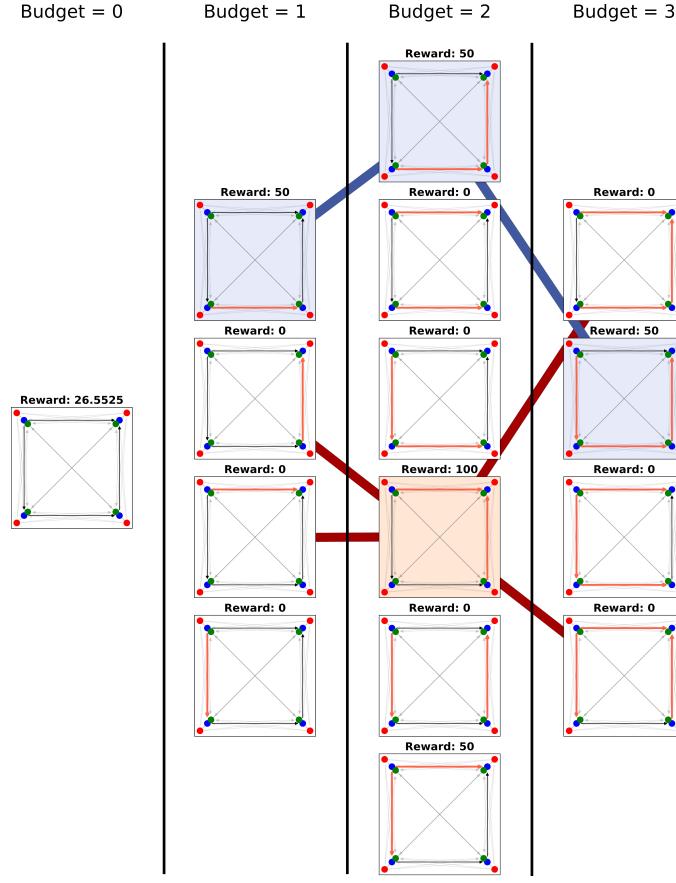


Figure 20: Caption

Appendix I HYPERPARAMETERS

For the sake of reproducibility, we here report the hyperparameter settings of all baselines, as well as RL, approaches applied in this work. All algorithms were applied with exactly these hyperparameters on all synthetic datasets.

I.1 Random Baseline

Here the choice of the edge to remove is uniformly at random. As the Python3 internal random number generator is pseudo-random, the sequence depends on the used random seed (Appendix J).

I.2 GA Baseline

Genetic Algorithm (GA) Hyperparameters	
Hyperparameter	Value
Number of generations	150
Number of mating parents	10
Solutions per population	30
Crossover probability	0.5
Mutation probability	0.5
Iterations until saturation	20

I.3 (Max)Q-Learning

Q-learning Hyperparameters	
Hyperparameter	Value
Number of episodes	150
Step-size (α)	1.0
ϵ -start	1.0
ϵ -end	0.01
ϵ -decay	500
ϵ -static	100

I.4 D(Max)QN-Learning

D(Max)QN-learning Hyperparameters	
Hyperparameter	Value
Number of episodes	150
Step-size (α)	1.0
Policy and target network structure	IL: $\# E^{pt} $ HL1: $\# E^{pt} + 10$ OL: $\# E^{pt} $
Optimizer	ADAM [22]
Loss	Mean Squared Error
Learning rate	0.1
ϵ -start	1.0
ϵ -end	0.01
ϵ -decay	500
ϵ -static	100
Batch size	32
Replay memory size	512
Target network update step-size	50

Table 3: IL: Input Layer, HL1: Hidden Layer 1, OL: Output Layer

Appendix J RANDOM SEEDS

We used the following random seeds throughout our work for the packages `random`, `numpy`, and `torch`. Note that all single experiments were run with random seed 2048 while we used the remaining random seeds for comparative runs:

Random Seed
2048
4096
6144
8192
10240
12288
14336
16384
18432
20480
22528
24576
26624
28672
30720