ABSTRACT

Title of thesis:	Phased replacement of bus lines by rail in a transit network with passenger flow modeling
	Arturo Ramirez, Master of Science, 2024
Thesis directed by:	Professor Paul M. Schonfeld, Department of Civil and Environmental Engineering

New urban rail lines are often built gradually, aiming to leverage the revenue from operating on built infrastructure to increase the available budget for further extensions. Each candidate inter-station link competes with the others in funding allocation and offers different benefits in travel time reduction and increased fare revenue. Previous studies have relied on travel paths that are fixed over time for the users of each origin-destination pair, thus failing to capture their actual dependence on network development in redundant networks.

A bi-level optimization method is proposed here to solve the Rail Link Building Sequencing Problem with replacement of bus lines. The first level uses a genetic algorithm to optimize the sequence in which new links from multiple rail lines of a network will be built, replacing preexisting bus lines, with the objective of maximizing the present value of the total societal net benefit (Y) over a time horizon. As the evolution of the network caused by a building sequence is modeled, the second level of optimization maximizes each year's undiscounted Y by means of an iterative process

to algebraically reoptimize the headways of individual lines for that year, while simultaneously recomputing user travel paths, travel times and flows in response to those headways. Travel time computations are performed on a graph with several edges per link to ease the correct inclusion of waiting times at transfer stations.

The algorithm was used to optimize the rail link-building sequence of a proposed urban transit network in San Jose, Costa Rica, under two different user path modeling strategies: 1) dynamic re-computation each year, and 2) adherence of the users of each O-D pair throughout the entire timeframe to the path that will be the fastest upon rail network completion. In both cases, the objective function was found to be maximized by replacing each bus line by rail in a mostly uninterrupted way. As the main finding, the results show how dynamic path modeling yields sequences 3% better in terms of Y and more strongly favors completing radial lines before circular lines.

PHASED REPLACEMENT OF BUS LINES BY RAIL IN A TRANSIT NETWORK WITH PASSENGER FLOW MODELING

by

Arturo Ramirez

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2024

Advisory Committee:
Professor Paul M. Schonfeld, Chair
Professor Cinzia Cirillo
Professor Xianfeng Yang

Dedication

To God, who has been my foundational stone.

To my wife and son, who have been my warm light and reason to strive.

To my family, who has been my joyous port of departure and source of fond memories.

Acknowledgements

I wish to express my gratitude to my advisor, Dr. Paul Schonfeld, whose knowledgeable advice that goes beyond the technical matters places him as an exceptional mentor.

To committee members Dr. Cinzia Cirillo and Dr. Xianfeng Yang, for being inspiring and reachable tutors both in the classroom and outside.

To the University of Maryland and the A. James & Alice B. Clark Foundation, for believing in my project and pushing me to break my boundaries.

Table of contents

List of tables

List of figures

Chapter 1: Introduction

1.1. <u>Background</u>

Efficient urban transit often relies on trunk or high-capacity bus lines to provide relatively fast access between sectors of a city. As demand grows over time, rail service may become a justified replacement for those lines since it provides higher travel speeds and typically lower headways. The societal benefit from faster travel progressively outweighs the naturally higher investment, operation, and maintenance costs of rail service. Therefore, the construction of urban rail lines often builds upon preexisting main bus lines, replacing their operating path as bus stations are upgraded to rail and links (tracks) are built to connect them.

The timing for the execution of projects (the construction of a link) is a complex problem because each of them 1) competes with the others in terms of funding, 2) determines the topological feasibility of other projects due to network and service connectivity requirements, and 3) often has a characteristic impact on future common budget availability through the fare revenue from the ridership it enables. These three latter effects change throughout the development of the network in response to the completion of other projects.

The project-timing problem is reduced to finding an optimal link-building sequence under the assumption that yearly budget is the binding constraint for building rail links at a given time. Even with that simplification there is no unique way to judge between sequence alternatives, and sometimes the prioritization is guided by expert opinions or adaptation of development patterns from rail systems in other cities (Li, Liang, Han, Wang, & Xu, 2022). A good modeling-based viewpoint would typically include not only the costs for the provider of the service, but also an assessment of the value of the time spent by the users in the system and the benefit they perceive from traveling. Defining a link-building sequence that maximizes an objective function that includes these two components poses some challenges during optimization if the demand is modeled as responsive to travel time, which includes service headway whose optimization in turn depends on passenger flows. Additionally, a proper analysis should focus on the cumulative net present value (NPV) of the economic measure utilized during the entire studied timeframe, such as the discounted total societal net benefit.

1.2. Literature review

Phase prioritization has been studied as a central task for the incremental construction or enhancement of components of transportation networks, when the state of those components mutually affects the utilization of others. Recent examples include Montesinos-Valera, Aragonés-Beltrán and Pastor-Ferrando (2017), who propose a multi-criteria decision analysis to prioritize different line-wise improvements for the railway network of Valencia, Spain. Expert judgment from the dispatcher is used to quantify the interrelations between criteria, one of which is user travel time. Rui (2017) develops a decision support system for the phasing of rail maintenance tasks in two Danish interregional corridors, considering the ridership- and revenue-decreasing effects of track closures ("possessions"). Finally, even if devoted to assessing past and planned development instead of optimizing it, it is worth mentioning studies such as that of Xu et al (2018), who compare the benefits of high-speed rail expansion between cities by means of city-level measures of regional connectivity, which are based on reachability under a travel distance or time threshold.

In the specific context of phased construction of urban transit rail lines, Song and Yang (2008) rank the construction of stages of new urban rail lines in Wuhan, China (each line is to be built in one or two predefined stages) according to their benefit-cost ratio, with benefit being linear with passengers served. Qualitative assertions on city connectivity drive the suggested building sequence. Luo, Guo and Wu (2011), and Li et al (2022) optimize the construction sequence of complete urban rail lines in the Chinese cities of Xi'an and Chengdu, respectively. The former study uses weighted gray correlation, while the latter employs an improved version of TOPSIS, the *Technique for Order Preference by Similarity to an Ideal Solution* (Zavadskas, Mardani, Turskis, Jusoh, & Nor, 2016). Both methodologies account for static line-wise indicators like potential passenger flow, urban location or impact on land value. The latter study further determines construction timing then by aiming to follow a "reasonable" growth curve (total kms vs. year), whose parameters are calculated from the growth curves of existing urban rail networks in select comparable cities. Notably, neither of these three works optimize the link construction sequence of each line, and in all of them travel behavior is irresponsive to construction and operation decisions beyond the mere connection of stations.

Traffic assignment or flow simulation is a commonly implemented lower-level optimization in algorithms whose first-level decisions relate to the construction sequence or timing of transit network components. Examples are numerous in fields like highways or waterways. However, to the best of the author's knowledge, recent research on phased development of rail lines has typically omitted dynamic user pathing in favor of pursuing conclusions of more general applicability about link prioritization, which are easier to derive from simpler settings like single lines or networks with predetermined paths for the O-D pairs. The rest of this subsection describes the five studies that more closely encapsulate the modeling interests that guide this thesis, with the differences and similarities remarked in Table 1 near the end of the subsection.

Kuby, Xu and Xie (2001) use mixed integral programming to optimize the selection of construction or upgrade projects to be executed in each 5-year planning period of a short timeframe of three periods only. Those decisions are greedily optimized to minimize the total cost of each period, with the resulting network and budget constraining the decisions for the previous period, because periods are processed backwardly in time. Their framework is more comprehensive in terms of construction project types than this thesis because it differentiates between constructing single-track links, upgrading them to double-track, and upgrading them to electrified. The modeling tradeoff is omitting demand elasticity and headway optimization, and simplifying the involvement of demand satisfaction in the objective by only accounting for the cost of trips lost out of travel impossibility. Unlike other models, it seemingly does consider more than one possible path for each O-D pair, although they have to be input and it is not clear how the path selection (or flow distribution) occurs.

Cheng and Schonfeld (2015) optimize the scheduling of the construction of additional stations to extend a single rail transit line in one direction, over a year-discretized timeframe. The objective is to maximize the present worth of net societal benefit, which equals the users' gross benefit minus travel time, operation costs and maintenance costs. The solution method is simulated annealing, where the decision variables are binaries specifying completion state of each link at each year. Stations are pre-located, and economies of building multiple stations in a single period are accounted for. Flow between any two stations grows exponentially over time at a constant rate that increases after the two stations are rail-connected. However, flows (and therefore gross used benefit) are assumed to be inelastic with respect to travel time and cost, and consequently, the headway of each year can simultaneously be optimized to maximize the objective function, and capped to satisfy the year's static demand. The case study includes a budget-and-revenue-constrained example, as well as an unconstrained one which isolates the question of when the demand between two nodes is large enough for the user benefit to offset the additional costs of connecting them.

In a similar setting, Wu and Schonfeld (2022) use a genetic algorithm with the same objective function, but the existing rail line can be extended to both sides and demand is elastic with respect to fare and travel time. Chromosomes encode both the link-building sequence and their grouping into phases as decision variables. Since time is continuously modeled, phase duration is determined by the time required for the available budget (which considers external funding and internal revenue) to reach the amount needed to build the links scheduled for that phase, which occurs at the end of the phase. Travel in alternative modes is allowed, and its in-vehicle time accounted for, if necessary to partially cover the distance between two nodes as long as one of them is served by rail. The headway for each year was not optimized to maximize the objective function, but it was used to enforce vehicle compliance by solving together the ridership equation (which depends on the headway through wait time) with the equation for the maximum allowed headway based on vehicle capacity (which depends on ridership). A restriction was imposed to only allow extension to a single side of the track in each phase, which created the need for rules for crossover, mutation and generation of random chromosomes to avoid unacceptable solutions.

Peng, Li and Schonfeld (2019) extend the analysis to a multi-line rail network, but their objective function includes discounted supplier and user costs only, excluding user benefit. Demand between nodes in the network is inelastic to fare and travel time but increases with time and rail connectivity status. Travel by other modes is considered for nodes that are not connected end-to-end by rail. Rail travel times are immutable for each O-D pair. The links are sequentially built in continuous time as soon as the available budget allows, and in the order encoded in the chromosomes of the genetic algorithm used. Headway is capped by vehicle capacity, and reoptimized each year to minimize the objective function.

Finally, Sun et al. (2017) specifically optimizes the evolution of a bus into rail service over time, starting from the CBD outwards. Their single-line study considers the CBD as the only demand attractor, while demand origination is assumed continuous along the transit line according to a linear demand density function. Discounted cost over the time horizon is minimized by optimizing the decision variables which pertain to extension length at a given time.

The following table summarizes the modeling and optimization differences between the last five mentioned studies and the current thesis (only the first author is shown in the header):

Table 1. Differences between previous models used in the literature.

Criterion	Kuby (2001)	Cheng (2015)	Sun (2017)	Peng (2019)	Wu (2022)	This work
Rail replaces bus lines			✓			✓
Multiple rail lines	✓			✓		✓
Travel in alternative modes			✓	✓	✓	✓
Built stations increase max. demand				✓	✓	
Travel paths are dynamically calculated	✓					✓
Demand elasticity					✓	✓
Headways optimized		✓	✓	✓		✓
Headways capped by train capacity		✓	✓	✓	✓	
Obj. function includes user benefit		✓			✓	✓
Link construction may be delayed	✓	✓	✓			
Time is modeled as continuous			✓	✓	✓	
Bunched construction is supported	✓	✓	✓		✓	✓
Economies of bunched construction	✓	✓	✓			

To reduce the number of iteration layers, this thesis omits vehicle capacity constraints on headways, because as hinted in the table, it is computationally complex to optimize headways while simultaneously 1) enforcing that constraint, 2) using cost-responsive demand, and 3) using a demand-sensitive objective function in a multi-line case such as the one presented here. Similarly, allowing the deferral of construction until it is economically justified (as opposed to building as soon as funds allow) would require further components in the chromosomes, expanding the search space and reducing algorithm efficiency. Therefore, this work leaves out some criteria from those two groups in order to not overshadow the analysis of the component that more solidly sets apart this work from previous implementations on urban network evolution, which is the dynamic re-computation of user travel paths in response to the rail network development.

Additionally, the proposed model considers the following aspects which are used in all or most of the last five cited sources:

1. Physical continuity for the development of the rail network
2. Demand increasing over time
3. Budget constraints
4. Internal revenue

1.3. Scope of study

This thesis presents a framework for optimizing the construction sequence of the links and stations that compose a rail transit network involving multiple lines, which gradually replace preexisting equivalent bus lines. The replacement of a bus line can start at any link of its path, which splits it into two lines if the first link replaced is not an extreme one. The objective function is the present value of the net societal benefit over the analyzed timeframe, which is the user surplus minus the supplier costs reduced by the revenue, or conversely, the user gross benefit minus the sum of supplier costs and travel time costs. Headways of all operating bus and rail lines are jointly optimized each year to maximize that year's net societal benefit.

Time is discretized in periods of equal length, years in this case, where the economic and operational variables of each year are computed deterministically as a result of building the links and stations as soon as the available *balance* (unspent budget) allows it. Construction, operation and maintenance costs are considered on the supplier's side (the developer of the network), while the user surplus is affected by the cost of travel time and fares paid. The developer has its budget constrained by yearly internal revenue from fares and external funding. Hourly passenger flow for each O-D pair is determined by a demand function that is linear with respect to travel time and fare, and its parameters increase over time.

The solution method is a genetic algorithm whose chromosomes are the building priorities of the links, which are also the decision variables during optimization, although the conceptual model uses binary "built at year" variables as independent variables. Station construction is directly determined by link construction, and infrastructure construction continuity is enforced by the decoding algorithm. The start of operations of lines on links is largely dependent on link construction, but constrained to prevent line discontinuity that could arise even with contiguous link building, as illustrated later in Figure 4.

The main contributions of the model presented here relate to network modeling:

1. **Support for line redundancy**: The users' travel path for each O-D pair is dynamically reoptimized at each stage of the network development.
2. **Realistic modeling of transfer times:** Transfers between lines and modes are unlimited and assume uncoordinated dispatching. Accounting for them required eight edges per transit link in the network graph used for travel time computations. This, along with the previous contribution, enables more

realistic modeling of travel times, which dominates the cost per passenger-km in the included case study.

Some omitted aspects are:

1. Economies of completing several stations in a single year
2. Vehicle capacity considerations
3. Continuous time formulation
4. Influence of rail connectivity status of a node over its demand parameters, particularly the maximum theoretical flow.
5. Optimization of the feeder bus system or other components of the broad transit system managed by the same authority.

Chapter 2 of this thesis explains all the components of the multimodal transit model proposed and the relationship between variables. The algorithm designed to optimize the system is described in Chapter 3. Chapter 4 showcases an application of the solution algorithm for a real-life scenario, and Chapter 5 closes with some conclusions and proposals for future work.

Chapter 2: Problem formulation

A transit network is considered with the following components:

1. n_N nodes $i\epsilon\{1, \dots, n_N\}$ which are predefined station locations
2. n_L links $l\epsilon\{1, \dots, n_L\}$ between the stations, to be used by rail and bus lines
3. n_{RR} rail lines $r\epsilon\{1, \dots, n_{RR}\}$ to be progressively developed to serve the stations

4. n_{RR} initial bus lines $r \epsilon \{n_{RR} + 1, \dots, 2n_{RR}\}$ using the existing roads to serve the same sequence of stations as the planned rail lines that will gradually replace them, and
5. At most n_{RR} additional bus lines $r \epsilon \{2n_{RR} + 1, \dots, 3n_{RR}\}$ that might appear when the start of operations of a rail line splits its corresponding bus line into two

A link is understood here as the immaterial, bidirectional connection between two stations visited consecutively by any line. Terminology-wise, more than one planned rail line may *run* in the same link, but that does not represent two separate links: just the inclusion of a same link in the operating path of two lines. The gradual extension of rail lines and the replacement of bus by rail service occurs link-wise, as the initially inexistent rail stations are built in the nodes and rail tracks are built in the links.

The network development resulting from decisions of when to build rail infrastructure is analyzed over a timeframe discretized into n_T complete periods $t \epsilon \{0,1, \dots, n_T\}$, which in this study have a duration of one year. The objective is to maximize the performance of the network in terms of the discounted total societal net benefit Y it provides over the timeframe.

2.1.<u>Simplifying assumptions</u>

1. No rail link or station is built at the beginning of the analyzed timeframe.
2. Rail links are built in a contiguous (connected) way.

3. Rail links are built as soon as there is balance available at the beginning of a year, without subtracting from the funds required to operate and maintain the existing network during that entire year.
4. In any year, an unlimited number of rail links and stations is allowed to be built, only constrained by fund availability.
5. All rail link and station constructions scheduled for a year begin and end during that year, regardless of their number.
6. A rail station on a node can also be a bus station, where intermodal transfers are possible, if bus lines still use it.
7. For dispatching purposes at the beginning of each shift, trains can switch between non-colinear tracks in (and only in) any transfer station, which implies the existence of the necessary connecting tracks at those stations.

2.1.1. Transit service generalities

8. All transit lines are rail or bus services operated here by a single authority that in the long run does not earn any profit.
9. All the nodes are served by transit in every year, be it bus and/or rail, and thus every node is always reachable from each other node.
10. Lines are bidirectional, following the same link sequence in reverse order for the inbound and outbound legs of each round trip, in both cases stopping at the same stations and using the same track alignment (for rail) or road routes (for buses).

11. At the beginning of the timeframe, there is one bus line operating for every planned rail line, replicating its line path. It is gradually replaced by the rail line as links are built.
12. Rail lines start to operate on links of their maximum operating path the year immediately after each of them is built, except if line the connectivity limitation explained in section 2.6 delays the start of operations of the link
13. All service days of the year have the same number of service hours.

2.1.2. Transit operation specifics

14. Trains are dispatched from one or more depots from which the entire built network is accessible.
15. Bus and rail headways remain steady during all operating hours of all operating days of the year, but they can differ between them and vary from a year to the next.
16. Bus and rail headways are jointly optimized every year to maximize that year's societal benefit, but are capped by a minimum and maximum bound.
17. Headways and dispatching are not synchronized among lines.
18. Each year, every transit line has its own exclusive fleet of vehicles.
19. Fleets are computed as the minimum integer number that satisfies the computed headways. Also, no spare rolling stock is accounted for.
20. Bus and rail travel speeds represent an average applicable to links of any length and they account for local speed changes.
21. There is no additional time required for turning around a bus or train at a terminal station.

22. Dwell times are invariable across stations regardless of passenger flows in them.

2.1.3. Supplier costs and balances

23. Unit construction, operation and maintenance costs do not vary over time.
24. The unit construction costs of rail link and stations represent an average considering the construction of necessary viaducts, tunnels, rail tracks, platforms and electrical and mechanical systems, as well as land acquisition and externalities such as road closures.
25. The cost of building, operating and maintaining depots, additional facilities for terminal stations, or other structural components, is disregarded.
26. The cost of building a rail station in a node is invariable across nodes, and it is independent from the number of links adjacent to the node that are being upgraded to rail links. As an extreme implication of this, all platforms in a planned rail transfer station are completely built even from the moment it becomes the terminal station of one single rail line.
27. Operation costs factor in labor, energy consumption, and vehicle acquisition, maintenance and replacement, as well as any other cost that is positively correlated with vehicle-hours of operation.
28. Hourly operating costs per vehicle are assumed to relate to buses or trains large enough to meet the hourly demand on every link of the operating path of their respective line, for the assigned service frequencies.
29. Maintenance costs for rail links and stations begin the year immediately after they are built.

30. Bus stations maintenance costs are not considered.

31. External funding increases exponentially over time.

32. A flat fare f_t is charged per trip at year t, regardless of origin and destination, and allows unlimited transfers even between nodes.

33. Fare increases linearly year by year at the same rate as the maximum willingness to pay (explained later).

34. All revenue collected is destined to offset the construction, maintenance and operation costs. Costs and revenue of the feeder bus network or other components of the urban transit system at large are independent.

35. The arrival of a year's external funding and the setting aside of the year's construction funds are both assumed to occur at the beginning of the year.

36. Operation costs, maintenance costs and revenue are incurred or received by the operator in a prorated way throughout the year. Since balance must not be negative at any point of any year, this assumption increases the available funds for construction at the beginning of the year.

2.1.4. User behavior

37. A linear, hourly demand function is assumed for every O-D pair in each year t, with two O-D specific parameters: maximum hourly demand and maximum willingness to pay.

38. The maximum willingness to pay of every O-D pair is assumed to increase linearly every year from its respective initial value, and maximum hourly demands are assumed to increase appropriately so that the slope of every demand function does not change.

39. The hourly demand of an O-D pair is uniform during all operating hours of the day and throughout all operating days of the year.
40. Arrivals of users to origin stations are uniformly distributed in time during operation hours.
41. Access costs or time to or from the transit stations are disregarded.
42. All users know how to individually optimize their travel time, which causes all users of an O-D pair to follow the same “travel path” in a given year.

Simplifying assumptions for the generation of demand parameters are not listed here, since the object of this thesis is to develop and solve the optimization model.

2.2.Notation

The following table lists all the variables used by the model or the equations in this document, as well as the units used by each of them, and the parameter values used for the case study presented in Chapter 4: Numerical results. For parameters specific to links or origin-destination pairs, their average is provided instead.

Table 2. Variables used.

Symbol	Description	Units	Value
a_B	Additional acceleration or deceleration delay for bus vehicles	hours	$20\ s$
a_R	Additional acceleration or deceleration delay for rail vehicles	hours	$20\ s$
b_B	Alighting time for bus vehicles	hours	$20\ s$
b_R	Alighting time for rail vehicles	hours	$20\ s$
B_{St}	Supplier net benefit for year t	$	
B_{Ut}	User surplus for year t	$	
c_L	Construction cost per km of rail connections	$/km	25,000,000

Symbol	Description	Units	Value
c_N	Construction cost of a rail station, upgrading it from a bus station	$	$25{,}000{,}000$
C_t	Total annual construction cost for year t	$	
d_{Bl}	Length of link l for bus vehicles	km	$3.23\ (avg)$
d_{Rl}	Length of link l for rail vehicles	km	$2.15\ (avg)$
D_{Et}	Balance at the end of year t	$	
D_{St}	Balance at the beginning of year t	$	
$e_{ii'0}$	Maximum acceptable impedance from node i to node i' at the beginning of the analyzed timeframe	$/trip	$28.89\ (avg)$
$e_{ii't}$	Maximum acceptable impedance from node i to node i' in year t	$/trip	
f_0	Flat transit fare at the beginning of the analyzed timeframe	$/trip	2
f_t	Flat transit fare for year t	$/trip	
F_1	Annual external funding for the first year of the analyzed timeframe	$	$50{,}000{,}000$
F_t	Annual external funding for year t	$	
g_e	Linear annual growth rate for the economy (fares and maximum acceptable impedances)	$/year	0.05
g_f	Exponential annual growth rate for the external funding	%	1
h_{min}	Minimum allowed headway	hours	$3\ min$
h_{max}	Maximum allowed headway	hours	$15\ min$
h_{rt}	Decision variable: headway for transit line r in year t	hours	
$h_{rt}{}^*$	Optimized headway for transit line r in year t	hours	
$\hat{h}$	Set of headways for every year and every transit line	hours	
$\hat{h}^*$	Set of optimized headways for every transit line and year	hours	
i	Index used for the station nodes	-	

Symbol	Description	Units	Value
l	Index used for the transit links	-	
m_L	Annual maintenance cost per km of rail tracks	\$/year	600,000
m_R	Annual maintenance cost for a rail station	\$/year	600,000
M_t	Total annual maintenance cost for year t	\$	
n_L	Number of transit links in the network	-	57
n_N	Number of nodes in the network	-	51
n_R	Number of all possible transit lines	-	12
n_{RR}	Number of planned rail lines	-	4
n_T	Number of complete years in the analyzed timeframe	-	50
o_B	Hourly operation cost for one bus	\$/hour	100
o_R	Hourly operation cost for one train	\$/hour	4000
O_t	Total annual operating cost for year t	\$	
p_{lr}	Travel and dwell time for link l for transit line r		
$q_{ii't}$	Hourly passenger flow from node i to node i' in year t	trips/hour	
$\mathbb{Q}$	Total hourly round trips of the total catchment region (target value for generating first year demand)	trips/hour	30000
$Q_{ii'0}$	Maximum theoretical hourly passenger flow from node i to node i' at the beginning of the analyzed timeframe	trips/hour	11.75 (avg)
$Q_{ii't}$	Maximum theoretical hourly passenger flow from node i to node i' in year t	trips/hour	
r	Index used for the transit lines	-	
s_B	Dwell time at stations for buses	hours	20 s
s_R	Dwell time at stations for rail vehicles	hours	20 s
t	Index used for the years	years	
u	Value of time for passengers	\$/hour	15
v_B	Average cruise speed for buses	km/h	65
v_R	Average cruise speed for trains	km/h	25
V_t	Total annual revenue for year t	\$	

Symbol	Description	Units	Value
x_{lt}	"Link built binary" telling if link l in a built state in year t	-	
$\hat{x}$	Set of "built binaries" for every link and year	-	
$\hat{x}^*$	Set of optimized "built binaries" for every link and year	-	
Y	Net present value of total societal net benefit	\$	
Y_t	Undiscounted societal net benefit for year t	\$	
y_{lrt}	"Operational binary" telling if link l is part of the operating path of transit line r in year t	-	
$\hat{y}$	Set of "operational binaries" for every link, transit line and year	-	
$\hat{y}^*$	Set of optimized "operational binaries" for every link, transit line and year	-	
$z_{lrii't}$	User-optimized variable: "Travel path binary" telling if link l of transit line r is part of the optimal path for traveling from node i to node i' in year t	-	
$\hat{z}$	Set of "path binaries" for every link, transit line origin, destination and year	-	
$\hat{z}^*$	Set of optimized "path binaries" for every link, transit line origin, destination and year	-	
α_{il}	"Link connectivity binary" telling if link l connects to node i	-	$(varies)$
β_{lr}	"Line path binary" telling if link l belongs to the intended line path of transit line r	-	$(varies)$
γ	Annual discount rate	%	4
ε_B	Baseline impedance tolerance	\$/trip	20
ε_M	Additional impedance tolerance per km of Euclidean distance	\$/km/trip	1
$\delta_{ii'}$	Euclidean distance between nodes i and i'	km	5.68 (avg)
η	Number of operating hours per year	h/year	6205
θ_r	"Transit type binary" telling if line r is a rail line (otherwise it is a bus line)	-	$(varies)$
κ_i	Relative attractive potential of the catchment zone of node i	-	0.02

Symbol	Description	Units	Value
ρ_{rt}	Round trip time for transit line r in year t	hours	
$\tau_{ii't}$	Travel time from node i to node i' in year t	hours	
χ_{it}	"Station built binary" telling if the station in node i is in a built state in year t	-	
ω_i	Relative generative potential of the catchment zone of node i	-	0.02

2.3.<u>Network and line topology inputs</u>

The paths of the planned rail lines are externally defined using a set of link connectivity binaries α_{il} and a set of line path binaries β_{lr}. The α_{il} binaries specify whether link l connects to node i. They must comply with the following three constraints:

Internodal connections, equation (1): every link l connects exactly two nodes.

$$\sum_{1\le i\le n_N} \alpha_{il} = 2 \quad \forall l \in \{1,\dots,n_L\} \tag{1}$$

Exclusive connections, equation (2): no two different links l and l' connect the same two nodes.

$$\sum_{1\le i\le n_N} \alpha_{il}\alpha_{il'} \le 1 \quad \forall l,l' \in \{1,\dots,n_L\}, l \ne l' \tag{2}$$

O-D connectivity, equation (3): every node must be reachable from every other node. In other words, for every pair of nodes i and i' there must exist a sequence (ordered subset) of $n_{L_{ii'}}$ different links $\left\{l_{ii'1},\dots,l_{ii'n_{L_{ii'}}}\right\} \subset \{1,\dots,n_L\}$ that connects them. For that to happen the following three conditions must hold collectively true, which happens if their respective multiplied parentheses in equation (3) all equal 1.

1. Exactly one link in the sequence connects to node i
2. Exactly one link in the sequence connects to node i' (these two first conditions guarantee that i and i' are ends in the sequence), and
3. Each link shares exactly one node with the previous link in the sequence, which guarantees that the sequence is connected if it contains more than one link (although it can be branched).

$$\forall i, i' \in \{1, \dots, n_N\} \; \exists \left\{l_{ii'1}, \dots, l_{ii'n_{L_{ii'}}}\right\} \subset \{1, \dots, n_L\} \mid$$
$$\left(\sum_{1 \le \lambda \le n_{L_{ij}}} \alpha_{i l_{ii'\lambda}}\right)\left(\sum_{1 \le \lambda \le n_{L_{ij}}} \alpha_{i' l_{ii'\lambda}}\right)\left(\prod_{1 < \lambda \le n_{L_{ii'}}} \sum_{1 \le i'' \le n_N} \alpha_{i'' l_{ii'\lambda}} \alpha_{i'' l_{ii'(\lambda-1)}}\right) = 1 \tag{3}$$

Each planned rail line $r \in \{1, \dots, n_{RR}\}$ belongs to a "line triad" composed by itself, a *right-end bus line* $r + n_{RR}$ and a *left-end bus line* $r + 2n_{RR}$. All three lines of every triad are modeled as existing in every year, though not necessarily operating. Consequently, the total number of transit lines n_R in any year is always three times the number of planned rail lines n_{RR}. A set of transit type binaries θ_r specify whether line r is a rail line, otherwise it is a bus line. Rail lines are indexed first and bus lines afterward.

$$n_R = 3n_{RR} \tag{4}$$

$$\theta_r = \begin{cases} 1 \text{ if } r \le n_{RR} \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

The distinction between the right-end and the left-end bus line of a triad originates from each rail line having an arbitrarily defined outbound direction, with the "rightmost link" being the last one traversed by the line in outbound trips when the line is fully built and operational. The left-end bus line is the one that starts operating at the beginning of the timeframe, while the right-end one starts with an empty operating path (i.e., not operating). If the rail line starts operating on any other link that is not its rightmost one, then the right-end bus line begins to operate on the links to the right of the rail line's current operating path. Figure 1 exemplifies this by showing the gradual start of operations of a rail line r (thick solid line) through years $t_1 < t_2 < t_3 < t_4 < t_5 < t_6$, to replace its preexisting left-end bus line $r + n_{RR}$, (thin solid line), and the appearance of a right-end bus line $r + 2n_{RR}$ (thin dashed line) because the replacement starts at an intermediate link. The maximum operating path of the triad in this example consists of the generic link sequence l_1, l_2, l_3, l_4, l_5.

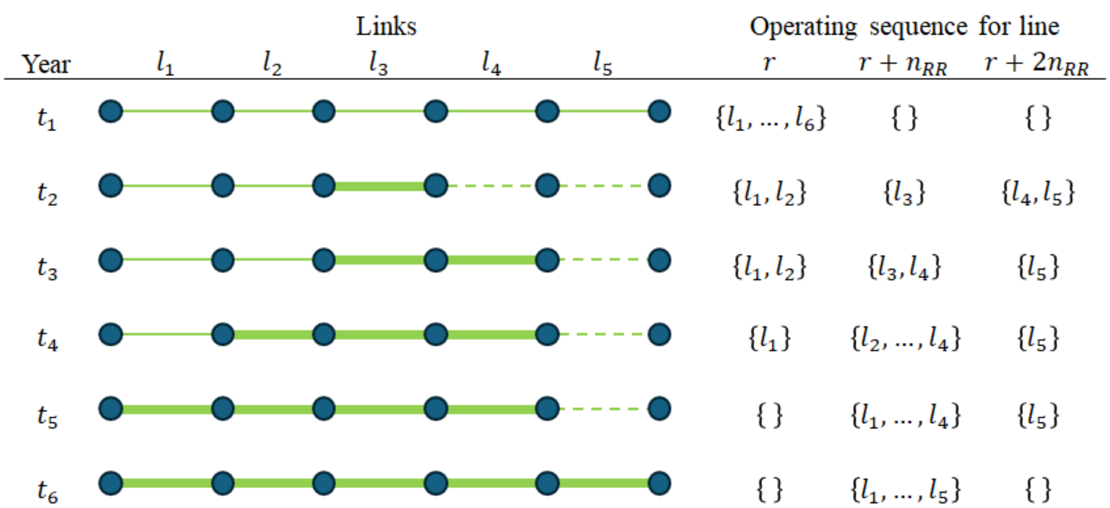

Figure 1. Bus line splitting during the gradual replacement of a bus line by a rail line.

The set of line path binaries β_{lr} specifies whether link l belongs to the "maximum path" of transit line r. The maximum paths are externally given for the n_{RR} rail lines, signifying the intended path of the line when the network's development is completed, while for the bus lines they replicate their corresponding rail line's. Association between bus and rail lines is ingrained in the definition of those path binaries β_{lr}, with each bus line having its "maximum path" tied to its corresponding rail line's, as shown in equation (6).

$$\beta_{lr} = \begin{cases} \text{externally given if } r \leq n_{RR} \\ \beta_{l(r-n_{RR})} \text{ otherwise} \end{cases} \tag{6}$$

There are restrictions on the externally provided β_{lr} binaries.

Total rail coverage, equation (7): each link l must be part ($\beta_{lr} = 1$) of the maximum path of at least one rail line r.

$$\sum_{1 \leq r \leq n_{RR}} \beta_{lr} > 1 \quad \forall l \in \{1, \dots, n_L\} \tag{7}$$

No line branching, equation (8): no node i can be connected ($\alpha_{il} = 1$) to more than two links l that are part ($\beta_{lr} = 1$) of the maximum path of a rail line r.

$$\sum_{1 \leq l \leq n_L} \beta_{lr} \alpha_{il} \leq 2 \quad \forall i \in \{1, \dots, n_N\}, r \in \{1, \dots, n_{RR}\} \tag{8}$$

Line connectivity, equation (9): the maximum line path of any rail line r must form a connected sequence with exactly two extreme nodes.

$$n_N - \sum_{1 \leq i \leq n_N} \left(-1 + \sum_{1 \leq l \leq n_L} \beta_{lr} \alpha_{il} \right)^2 = 2 \tag{9}$$

$$\forall\, r \in \{1, \dots, n_{RR}\}$$

The logic operations of the above expression unfold as follows:

1. The innermost summation evaluates to 0, 1 or 2, depending on the number of links l that are part ($\beta_{lr} = 1$) of the maximum line path of rail line r, that are

connected ($\alpha_{il} = 1$) to a node i. It cannot reach 3 or more because maximum line path branching is forbidden by equation (8).

2. Every of those ternary results is subtracted by one and squared, which converts it into a binary with a value of zero if the node i is an extreme node of the maximum line path of rail line r, and a value of one otherwise (internal or non-visited nodes).
3. The outermost summation yields the number of nodes i in the network that are not extremes of the maximum line path of rail line r.
4. Subtracting the total number of nodes of the network minus the above result yields the total number of extreme nodes of the maximum line path of rail line r. If equal to two, the condition is met.

Equation (9) still fails to prevent the maximum line path of rail line from having disconnected, circular segments (sequences of nodes), additional to the main one that connects the origin to the destination. Explicitly formulating such constraint would result in very cumbersome equations.

Finally, it is worth noting that despite this last constraint, circular rail lines can still be modeled by setting both ends in a single node. However, current algorithm limitations will fail to add *through* edges in the *SAN* (*Service and Actions Network*, explained in section 3.2) in that node, thus incorrectly imposing unnecessary alighting and waiting time for users who just want to traverse it.

2.4.Origin-destination travel times

A set of binary variables $z_{lrii't}$, called "path binaries", is employed to specify whether link l of transit line r is part of the travel path followed by users of O-D pair ii' in year t. Conceptually, those path binaries are not pure decision variables because they do not embody independent alternatives for the decision-making authority to choose among and needing to be optimized. Instead, they refer to an expected, self-optimized user behavior that occurs as a response of the transit authority's decisions in an assumed, procedurally predictable manner, which is mimicked by a deterministic quickest path algorithm (Dijkstra's, referenced in section 3.2) that guarantees optimality. However, because of the complexity of properly expressing optimal travel times in terms of any network-state variables in networks with redundant paths for O-D pairs, the path binaries $z_{lrii't}$ will be registered as additional decision variables in the model formulation and termed as "used-optimized variables".

$$p_{lr} = \frac{d_{Rl}\theta_r + d_{Bl}(1-\theta_r)}{v_R\theta_r + v_B(1-\theta_r)} + 2\big(a_R\theta_r + a_B(1-\theta_r)\big) + \big(s_R\theta_r + s_B(1-\theta_r)\big) \tag{10}$$

$$\tau_{ii't} = \sum_{1\le r\le n_R}\left(\left(\sum_{1\le l\le n_L} z_{lrii't}\,p_{lr}\right) + \mathbb{1}_{rii't}\frac{h_{rt}}{2} + \mathbb{1}_{rii't}\big(b_R\theta_r + b_B(1-\theta_r)\big)\right) \tag{11}$$

$$\mathbb{1}_{rii't} = 1 - \prod_{1\le l\le n_L}(1 - z_{lrii't}) \tag{12}$$

Equation (11) shows that the travel time $\tau_{ii't}$ from a node i to a node i' in a given year t is composed of one "line travel time" component for each transit line r that is part of the travel path, as represented by the outer summation of the equation. Each of those components is in turn the sum of a link-wise part and a line-wise part, as shown in the equation:

1. Link-wise part: it is the summation of one sub-component p_{lr} for each link l of the transit line's path that is part of the travel path for the O-D pair in year t, with that condition being verified by means of the $z_{lrii't}$ binary of the respective link. Each of those sub-components p_{lr} reflects the moving and dwell time of the link for that transit line, which includes the following three terms as shown in equation (10):
    a. Traversing time: time that would be required for a vehicle of the transit line to traverse the link without stopping at any of its two nodes, i.e., the length of the link (d_{Rl} or d_{Bl}, depending on whether it is a rail or a bus line) divided by some externally given average speed v_R or v_B depending on whether it is a rail or a bus line. That last condition is accounted for with a "transit type" binary variable θ_r that has a value of 1 if for rail and 0 for bus.
    b. Acceleration and deceleration delays: two times an externally given additional acceleration and deceleration delay a_R or a_B depending on whether it is a rail or a bus line. This input must represent only the additional time required for the vehicles to go from rest to the cruise

speed or vice versa, compared to the time required to travel the same acceleration/deceleration distance at cruise speed.

c. Station dwell time: an externally given station dwell time s_R or s_B depending on whether it is a rail or a bus line. It is differentiated from the acceleration and deceleration delays for modeling reasons: the SAN network graph used to optimize the user's travel paths as will be explained in section 3.2 distinguishes between alighting and dwell edges, but the users that use any of them still must incur the line's acceleration and deceleration delay.

2. Line-wise part: this part consists in the sum of two terms that are added only once, if line r is part of the travel path for O-D pair ii' in year t. That condition is verified by the indicator $\mathbb{1}_{rii't}$, which as shown in equation (12) is just a logical "OR" that yields one if $z_{lrii't} = 1$ for any l, for that line r, nodes i and i', and year t, and 0 otherwise. The two added terms if that condition is true are:

   a. Waiting time: it represents the time spent by the users waiting in a station for the bus or train of line r to arrive, regardless of whether it is the wait to take the first bus or train of the travel path between the nodes or a subsequent one (i.e., a wait in a transfer node). The term consists of half the headway h_{rt} of that transit line, to express an average waiting time.

   b. Alighting time: an externally given average alighting time b_R or b_B depending on whether it is a rail or a bus line. It should be noted that boarding time does not need to be considered because the passengers must still wait the entire duration of the stop (s_r) in the station where they board the vehicle after it arrives at the station, regardless of how fast they board.

2.5. User costs, demand and benefits

Actual hourly passenger flow $q_{ii't}$, meaning the number of trips initiated during a period of one hour from origin node i to destination node i' during year t, is assumed to follow the underlying demand function shown in equation (13). This function is linear with respect to the trip impedance, also called generalized trip cost, for that O-D pair in that year. Trip impedance includes: 1) the trip fare f_t, which is uniform for all O-D pairs but grows linearly by g_Q each year as expressed in equation (22), and 2) the monetary equivalent of the time spent in the trip, calculated as the trip's travel time $\tau_{ii't}$ multiplied by a fixed value of time u that does not differentiate between in-vehicle and waiting times. Importantly, the demand function is actually defined by parts, evaluating to zero if $f_t + u\tau_{ii't}$ is outside of the $[0, e_{ii't}]$ range.

$$q_{ii't} = \begin{cases} Q_{ii't}\dfrac{e_{ii't} - f_t - u\tau_{ii't}}{e_{ii't}} & if\ 0 \leq f_t + u\tau_{ii't} \leq e_{ii't} \\ 0\ otherwise \end{cases} \tag{13}$$

The linearity constants for equation (13) are specific for the O-D pair in each year t:

1. $e_{ii't}$ represents the maximum acceptable impedance that causes any flow in year t. In simple words, it is the maximum gross benefit any user of the network assigns to the trip between nodes i and i' and is therefore willing to pay in the form of trip impedance.
2. $Q_{ii't}$ represents the maximum theoretical hourly flow at zero generalized cost.

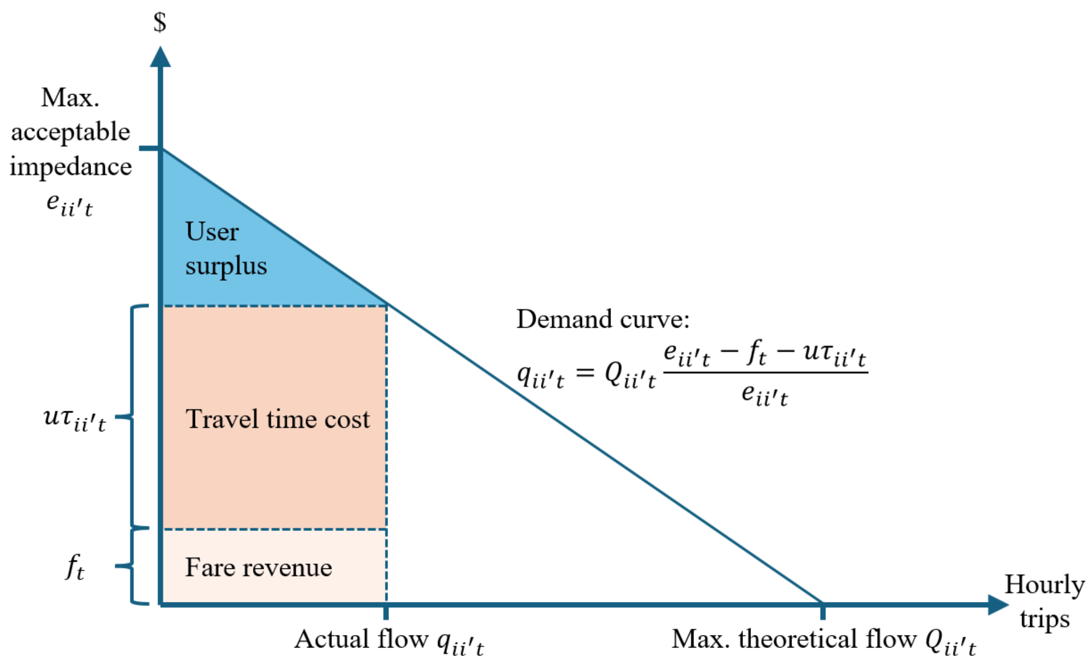

Figure 2. Hourly demand function for an O-D pair in a specific year.

Gross user benefit for the $q_{ii't}$ passengers of an O-D pair ii' who choose to travel during a period of one hour in some year t as per the passenger flow equation (13) is defined as the area under that function in Figure 2 from zero to $q_{ii't}$. The total, yearly gross benefit therefore aggregates this result for all O-D pairs in all the operating hours of the year. However, neither equations for the hourly or yearly gross user benefit nor those for the user cost (impedance, mentioned inside equation (13)) are separately formulated here, because it is their difference that is important for this study. This difference of gross user benefit minus the impedance is known as the user surplus (or user net benefit) and is illustrated as a triangular area in Figure 2, for the case of a single O-D pair. The area of that triangle is easily determined with the operations inside the summations in equation (14). In full, this equation expresses the annual user surplus

B_{Ut} for a year t, which is the sum of the user surplus for all O-D pairs for all the η operating hours of the year.

$$B_{Ut} = \eta \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} q_{ii't} \frac{(e_{ii't} - f_t) - u\tau_{ii't}}{2} \tag{14}$$

The O-D-specific demand constants $e_{ii't}$ and $Q_{ii't}$ increase over time to reflect a yearly "vertical rise" of every demand curve by a fixed amount of g_e monetary units. This demand growth model was chosen for ease of computation, and implies a constant annual growth of g_e monetary units for $e_{ii't}$, and a constant annual growth of $g_e \frac{Q_{ii'0}}{e_{ii'0}}$ users for $Q_{ii't}$. This causes the demand function of an O-D pair ii' to maintain a fixed slope $\frac{Q_{ii'0}}{e_{ii'0}}$ throughout the timeframe. In other words, every year the willingness to pay of every user of the network (i.e., his or her potential individual gross benefit from making the trip) rises by the fixed amount g_e, and also $g_e \frac{Q_{ii'0}}{e_{ii'0}}$ new potential users appear each with an individual willingness to pay ranging between zero and g_e. The yearly introduction of those new riders reflects the combined effect of population growth and increase in the willingness of existing population to travel at all, induced by economic expansion. Figure 3 illustrates this vertical rise of a demand function from year t to year $t + 1$, where the user benefit (colored triangular areas, explained later) is unchanged if travel time does not change ($\tau_{ii't} = \tau_{ii'(t+1)}$).

$$e_{ii't} = e_{ii'0} + g_e t \tag{15}$$

$$Q_{ii't} = Q_{ii'0} + \left(g_e \frac{Q_{ii'0}}{e_{ii'0}}\right) t \tag{16}$$

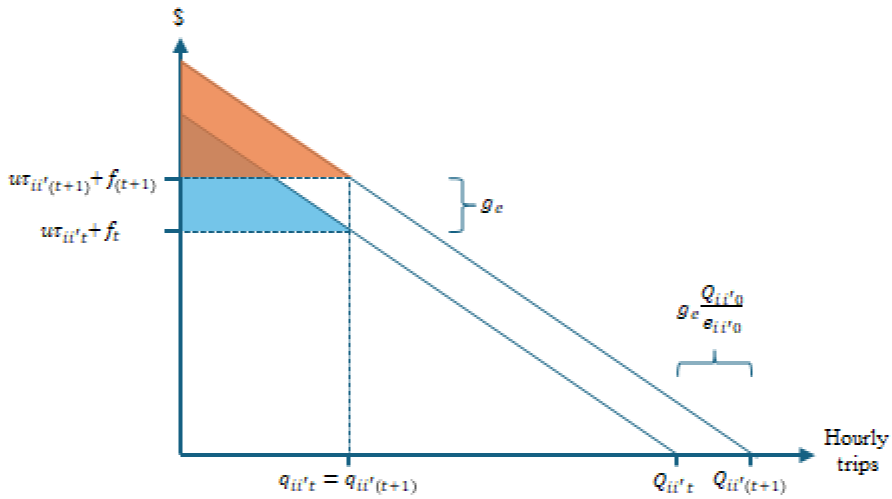

Figure 3. Assumed yearly economic growth which determines demand and fares.

Note that g_e is also the fixed annual growth of the flat transit fares as formulated in equation (22). This double usage is deliberate, so that unless there are travel time alterations induced by infrastructural changes, fares "catch up" with yearly demand growth so the equilibrium passenger flow and total net user benefit (explained later) remain unchanged for each O-D pair. This makes it unnecessary to reoptimize headways in years that do not see any new operational links compared to the previous year, saving valuable computation time.

2.6.<u>Supplier costs, benefits and balances</u>

Costs for the supplier of the transit service depend on building decisions for the infrastructure and operating decisions for the transit lines.

- Infrastructure building: Two sets of binary variables x_{lt} and χ_{it} specify, respectively, whether a rail connection for link l or a rail station in node i is in a "built state" in year t, which means that its construction or upgrading occurred in year t or earlier. Consequently, the subtractions $x_{lt} - x_{l(t-1)}$ and $\chi_{it} - \chi_{i(t-1)}$ yield 1 if the construction or upgrading (respectively) occurred exactly in year t, and 0 otherwise.

- Transit operations are determined by a set of binary variables y_{lrt}, which specify whether link l is part of the "operating outbound path" of transit line r in year t, i.e., the sequence of links traversed during an outbound trip of the line (the opposite sequence is followed during inbound trips).

On the infrastructure side, "link built" binaries x_{lt} are the ones considered as independent variables, while "station built" binaries χ_{it} are dependent on them. χ_{it} is equal to 1 if the x_{lt} of any link l that connects to node i is equal to 1 for that year t, i.e., if any link that reaches the station is by that year, as per equation (17).

$$\chi_{it} = 1 - \prod_{1 \le l \le n_L} (1 - \alpha_{il} x_{lt}) \tag{17}$$

On the line operations side, the binaries y_{lrt} that refer to rail operations could in theory be seen as dependent on the link building binaries x_{lt}. The latter would then be optimized under the assumption that at any given moment, all rail lines operate in all the rail-enabled links of their intended path, because it is wasteful to build rail tracks in a link and not immediately starting to operate on it with all rail lines that use it. However, there is a limitation explained later which prevents the binaries y_{lrt} of rail lines from being easily computed as 1 if the link l is part of the operating path of line r and it is in a built state in year t, and 0 otherwise. Additionally, the binaries y_{lrt} that refer to bus operations are completely dependent on the ones that refer to rail operations, because bus lines must service the links not covered by their respective rail lines at a given time. Because of the complexity of translating those limitations and rules for rail and bus line operations into closed-form equalities, those operation binaries will be registered instead as additional decision variables in the model formulation, and the rules and limitations are included as restrictions on those variables.

The aforementioned limitation on the immediate start of operations on built links relates to line connectivity. In some cases, even a contiguous development of the rail network can leave multiple isolated rail-enabled segments in a rail line's intended path. This is illustrated in Figure 4, where the red rail line could operate in the links of either of the two segments of its intended path that are rail-enabled (a and b) in that year, but not in both because they do not form a connected sequence. Since rail lines are not allowed here to stop operating on links on which they started operating, the segment that was already (at least partially) rail-enabled in the previous year is the only one that can be operated and expanded until there is a continuous rail connection between segments a and b. If none of the rail-enabled segments was so in the previous year, the decision could be solved by arbitrarily choosing to operate in one of the segments.

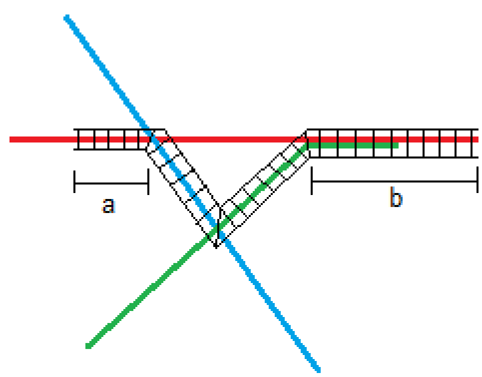

Figure 4. Line continuity constraint on rail operation

For any given year t, the total annual construction cost C_t is the sum of the costs incurred by the transit agency for building new infrastructure in that year. As shown in equation (18), a parameter c_L is used to encompass an average cost per kilometer for building rail connections. Another parameter c_N represents the average cost of building one rail station, this is, upgrading it from being a bus station.

$$C_t = \sum_{1 \le l \le n_L} \left(x_{lt} - x_{l(t-1)}\right) c_L d_l + \sum_{1 \le i \le n_N} \left(\chi_{it} - \chi_{i(t-1)}\right) c_N \tag{18}$$

The total annual maintenance cost M_t is the sum of the costs incurred for maintaining existing infrastructure in that year. As shown in equation (19), two parameters m_R and m_L are used to represent the average annual maintenance cost per built rail station and per built kilometer of rail connection, respectively. Bus operation does not generate any maintenance costs.

$$M_t = \sum_{1 \le l \le n_L} x_{l(t-1)} m_L d_l + \sum_{1 \le i \le n_N} \chi_{i(t-1)} m_R \tag{19}$$

For any given year t, the total annual operation cost O_t is the sum of the hourly costs incurred for operating all the existing transit lines in that year, multiplied by the η operating hours of the year. The hourly cost for a transit line r in a given year t is defined as the hourly operation cost for one vehicle of that line (o_R or o_B respectively for rail and bus lines) multiplied by the fleet required that year, which corresponds to the round trip time (ρ_{rt}) divided by the headway (h_{rt}) of that line in that year, rounded up to the nearest integer as shown in equation (20). The round trip time ρ_{rt} is calculated as the sum of the travel and dwell time p_{lr} (as per equation (10)) of all links l that are part of the operating outbound path of that transit line r in that year t (this condition is verified with the y_{lrt} binary), multiplied by two to include both the inbound and outbound trips.

$$O_t = \eta \sum_{1 \le r \le n_R} \left\lceil \frac{\rho_{rt}}{h_{rt}} \right\rceil \left(o_R \theta_r + o_B (1 - \theta_r) \right) \tag{20}$$

$$\rho_{rt} = 2 \sum_{1 \le l \le n_L} y_{lrt} p_{lr} \tag{21}$$

The flat fare charged to users for making any trip increases linearly by g_e each year (whose magnitude is related to the growth of the demand curve as was outlined in section 2.5), with a theoretical fare of f_0 at the beginning of the analyzed timeframe ($t = 0$).

$$f_t = f_0 + g_e t \tag{22}$$

The benefit that the supplier receives to offset those costs is the revenue from providing the service. For any year t, the total revenue aggregates the fares collected for the trips of all O-D pairs during the η operating hours of the year.

$$V_t = \eta f_t \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} q_{ii't} \tag{23}$$

The balance D_{Et} at the end of year t considers all income (revenue and external funding) and expenditures (construction, maintenance and operation) incurred during the year and all years prior. The balance D_{St} at the beginning of year t is not equal to $D_{E(t-1)}$, but instead it refers to the funds available right after the arrival of the year's external funding and the setting aside of the funds needed for the year's construction, both of which are assumed to occur at the very beginning of the year. In other words,

D_{St} equals D_{Et} minus the year's "net continued cost": revenue minus maintenance and operation costs.

$$D_{Et} = \sum_{1 \le t' \le t} (F_{t'} + V_{t'} - C_{t'} - M_{t'} - O_{t'}) \tag{24}$$

$$D_{St} = D_{Et} - (V_{t'} - M_{t'} - O_{t'}) \tag{25}$$

Similarly to the analysis for users, the supplier's net benefit consists of its total revenue minus the costs it incurs in to provide the service. Market-driven suppliers typically aim to operate with the former outweighing the latter, resulting in a positive net benefit that is called profit. However, in this case negative profits are permitted since the supplier is a subsidized agency tasked to maximize social benefits even at internal monetary losses, which is almost universally the case for mass transit systems. For any given year t, the annual supplier net benefit B_{St} for that year is then the total annual revenue (V_t) minus the sum of the total annual construction (C_t), operating (O_t) and maintenance (M_t) costs, whose result can be positive or negative.

$$B_{St} = V_t - C_t - O_t - M_t \tag{26}$$

If negative, this difference between revenue and supplier cost is covered (even can be exceeded in some years) with public funding F_t external to the operations apparatus, which is made available yearly at the very beginning of each year t, and whose value increases exponentially at yearly rate g_f from some value F_1 for the first year.

$$F_t = F_1(1 + \gamma)^{g_f} \tag{27}$$

Although ultimately also representing an economic burden for society in the form of taxation, this external funding is not counted in any benefit assessment. This is because of the assumption that in practice it will be readjusted to match the supplier's yearly deficit once all links are built, if that happens before the analyzed timeframe ends. Therefore, all public spending in the long run is already considered by the yearly negative benefit of the supplier of the service.

2.7.<u>*Objective function and constraints*</u>

The undiscounted societal net benefit J_t provided by the transit system during year t is the sum of the user's net benefit B_{Ut} and the supplier's net benefit B_{St} for that year. The latter is included because in the long term the transit agency is not supposed to cash in any profit, and its yearly operational deficits are ultimately carried over to society through the public-sourced external funding.

$$\mathrm{J}_t = B_{Ut} + B_{St} \tag{28}$$

The optimization problem consists in finding the optimal sets of decision variables ($\hat{x}$, $\hat{y}$, $\hat{h}$ and $\hat{z}$) that maximize objective function (29), which is the total *net present value* (J, or *NPV*) of the societal benefit over the analyzed time horizon, brought to present value with an annual discount rate γ.

$$\mathrm{J} = \sum_{1 \le t \le n_T} \frac{1}{(1+\gamma)^t} \mathrm{J}_t \tag{29}$$

$$\hat{x}^*, \hat{y}^*, \hat{h}^*, \hat{z}^* = \underset{\hat{x},\hat{y},\hat{h},\hat{z}}{\operatorname{argmax}} \mathrm{J} \tag{30}$$

It is subject to the following series of constraints related to budget, construction, operations, and travel path. Although expressed with sometimes intricate equalities and inequalities, they represent intuitive limitations enforced by the optimization algorithm, whose explanation accompanies the equations.

2.7.1. Budget and construction constraints

Nonnegative balances, equations (31) and (32): the balance for the transit agency must be nonnegative at every moment of every year t. This is achieved by verifying it is nonnegative at the start (D_{St}) and at the end (D_{Et}) of the year.

$$D_{St} \geq 0 \quad \forall\, t \in \{1, \dots, n_T\} \tag{31}$$

$$D_{Et} \geq 0 \quad \forall\, t \in \{1, \dots, n_T\} \tag{32}$$

The first of those two equations is not redundant with the second, because as mentioned, D_{St} is evaluated right after the arrival of the year's external funding and the setting aside of the funds needed for the year's construction, both of which are assumed to occur at the very beginning of the year. If not appropriately restricted, high construction costs could drive D_{St} into negative values, which could still be compensated with a positive enough net continued cost (revenue minus maintenance and operating cost) that year, so that the final balance constraint is not violated. Therefore, this additional balance verification for D_{St} is performed.

Rail infrastructure persistence, equation (33): for every year t, rail tracks remain built on links l that had rail tracks in the previous year. This also ensures that the same happens with rail stations.

$$x_{l(t-1)}(1 - x_{lt}) = 0 \quad \forall\, t \in \{1, \dots, n_T\}, l \in \{1, \dots, n_L\} \tag{33}$$

Rail development continuity: for every year t, the rail-enabled links must form a connected network. An equation to express this simple requirement in terms of state variables would be unnecessarily complicated and is therefore omitted, since the algorithm procedurally enforces this restriction.

Initial rail development, equation (34): no links l are rail-enabled ($x_{l1} = 1$) in the first year of the timeframe.

$$x_{l1} = 0 \quad \forall\, l \in \{1, \dots, n_L\} \tag{34}$$

2.7.2. Operation constraints

Intended operating path, equation (35): for every year t, transit lines r can only operate ($y_{lrt} = 1$) on links l that are part of their intended line path defined by the β_{lr} input binaries.

$$y_{lrt}(1 - \beta_{lr}) = 0 \quad \forall\, t \in \{1, \dots, n_T\}, r \in \{1, \dots, n_R\}, l \in \{1, \dots, n_L\} \tag{35}$$

Infrastructure requirement, equation (36): for every year t, rail lines r can only operate ($y_{lrt} = 1$) on links l whose construction has already occurred by that year ($x_{lt} = 1$).

$$y_{lrt}(1 - x_{lt}) = 0 \quad \forall\, t \in \{1, \dots, n_T\}, r \in \{1, \dots, n_{RR}\}, l \in \{1, \dots, n_L\} \tag{36}$$

Rail service persistence, equation (37): for every year t, rail lines r do not cease to operate ($y_{lrt} = 1$) on links l on previously operating links.

$$y_{lr(t-1)}(1 - y_{lrt}) = 0 \quad \forall\, t \in \{1, \dots, n_T\}, r \in \{1, \dots, n_{RR}\}, l \in \{1, \dots, n_L\} \tag{37}$$

Complementary coverage, equation (38): for every year t, every link l of the intended path of every transit line r must be served ($y_{lrt} = 1$) by one and only one transit line of the triad: either the rail line, or its left-end bus line, or its "right-end bus line" ($\lambda = 0, 1, 2$, respectively). The latter is achieved if either the inner summation or the $1 - \beta_{lr}$ term yields one, with the latter preventing links that do not belong to the intended path from intervening in the verification.

$$(1-\beta_{lr})+\sum_{0\le\lambda\le 2} y_{l(r+\lambda n_{RR})t}=1 \qquad (38)$$

$$\forall\, t\in\{1,\dots,n_T\}, r\in\{1,\dots,n_R\}, l\in\{1,\dots,n_L\}$$

Initial bus operation, equation (39): for every rail line r, neither itself nor its corresponding right-end bus line $r+2n_{RR}$ are operating ($y_{lrt}=0$) at the beginning of the timeframe ($t=1$). The coverage restriction stated later ensures that the left-end bus line $r+1n_{RR}$ is operating and services the entire operating path during the first year.

$$y_{lr1}=0\quad \forall\, r\in\{1,\dots,n_{RR}\}\cup\{2n_{RR}+1,\dots,n_R\}, l\in\{1,\dots,n_L\} \qquad (39)$$

Operational connectivity, equation (40): for every year t, the operating path (given by the y_{lrt} binaries) of every transit line r must form a connected sequence if it includes any links. This is enforced by verifying that the number of extreme nodes of the line is either zero (the line is not operating or is circular) or two, which rules out disjointed operating segments.

$$\left(-1+n_N-\sum_{1\le i\le n_N}\left(-1+\sum_{1\le l\le n_L} y_{lrt}\alpha_{il}\right)^2\right)^2=1 \qquad (40)$$

$$\forall\, t\in\{1,\dots,n_T\}, r\in\{1,\dots,n_R\}$$

The logic operations for every year and transit line unfold as follows:

1. The innermost summation evaluates to 0, 1 or 2, depending on the number of links l that are part ($y_{lrt}=1$) of the operating path of line r, that are connected ($\alpha_{il}=1$) to a node i during year t. It cannot reach 3 or more because line branching was forbidden by equation (8) in the restrictions on the β_{lr} input variables (section 0), which determine the intended operating paths.

2. Each of those ternary results is reduced by one and squared, which converts it into a binary with a value of zero if the node i is an extreme node of the operating path of line r, and a value of one otherwise (internal or non-visited nodes).
3. The outermost summation yields the number of nodes i in the network that are not extremes of line r in year t.
4. Subtracting the total number of nodes of the network minus the above result yields the total number of extreme nodes of line r in year t.
5. The result is reduced by one and squared, which yields one if and only if line r has either zero or two extreme nodes in year t. If so, the condition is met.

Prohibited bus-to-bus transfers, equation (41): for every year t, a node i cannot be served by two bus lines of the same triad, i.e., corresponding to the same rail line r. In other words, service in a non-rail-enabled subset of the links of a rail line's path cannot be split between the two corresponding bus lines. In the equation, $\lambda \in \{1,2\}$ serves as an auxiliary index to iterate over the two bus lines of the triad. Therefore, if the result of that innermost summation is nonzero for both bus lines at that node, then their product will be so as well. The verification is carried out for every year t, rail line triad r, and node i.

$$\prod_{1\le\lambda\le2}\sum_{1\le l\le n_L} y_{l(r+\lambda n_{RR})t}\alpha_{il} = 0 \qquad (41)$$
$$\forall\ t \in \{1,\dots,n_T\}, r \in \{1,\dots,n_{RR}\}, i \in \{1,\dots,n_N\}$$

The logic operations for every year, transit line and node unfold as follows:

1. The innermost summation evaluates to either 0, 1 or 2 depending on the number of links l that are part ($y_{lrt} = 1$) of the operating path of bus line $r + \lambda n_{RR}$, that are connected ($\alpha_{il} = 1$) to node i during year t.
2. The constraint is violated if the result of that innermost summation is nonzero for both bus lines in that node, which will make their product be nonzero as well.

Figure 5 exemplifies three violations to this constraint for a triad composed of rail line r (thick solid line), left-end bus line $r + n_{RR}$, (thin solid line), and right-end bus line $r + 2n_{RR}$ (thin dashed line). The red node in each case is the where the prohibited bus-to-bus transfer occurs.

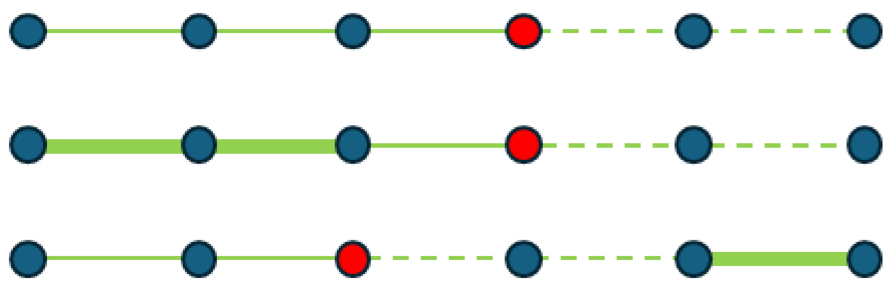

Figure 5. Violations of the bus-to-bus transfer prohibition in a triad.

Headway bounds, equations (42) and (43): during every year t, the headway h_{rt} of every transit line r must comply with globally set minimum h_{min} and maximum h_{max}.

$$h_{rt} \geq h_{min} \quad \forall\, t \in \{1, \ldots, n_T\}, r \in \{1, \ldots, n_R\} \tag{42}$$

$$h_{rt} \leq h_{max} \quad \forall\, t \in \{1, \ldots, n_T\}, r \in \{1, \ldots, n_R\} \tag{43}$$

It is worth noting that none of the operation constraints implies that a rail line *must* operate in all the rail-enabled links of its path. However, the chromosome decoding algorithm does ensure that rail lines always make use of the available built network, only limited by the continuity and no-subsets rules.

2.7.3. User travel constraints

Service requirement, equation (44): for every year t, users of every O-D pair ii' can only use ($z_{lrii't} = 1$) a transit line r in a link l as part of their travel path, if that link is part of that line's operating path in that year ($y_{lrt} = 1$).

$$z_{lrii't}(1 - y_{lrt}) = 0$$
$$\forall\, t \in \{1, \dots, n_T\}, r \in \{1, \dots, n_{RR}\}, l \in \{1, \dots, n_L\}, \quad (44)$$
$$i \in \{1, \dots, n_N\}, i' \in \{1, \dots, n_N\}$$

Link-wise line exclusivity, equation (45): for every year t, users of every O-D pair ii' can only use ($z_{lrii't} = 1$) at most one transit line r to traverse a link l.

$$\sum_{1 \le r \le n_R} z_{lrii't} \le 1 \quad (45)$$
$$\forall\, t \in \{1, \dots, n_T\}, l \in \{1, \dots, n_L\}, i \in \{1, \dots, n_N\}, i' \in \{1, \dots, n_N\}$$

No travel path branching, equation (46): for every year t, no node i'' along the travel path of any O-D pair ii' can be connected ($\alpha_{i''l} = 1$) to more than two links l that are part ($z_{lrii't} = 1$) of that travel path.

$$\sum_{1 \le l \le n_L} \sum_{1 \le r \le n_R} z_{lrii't}\alpha_{i''l} \le 2 \quad (46)$$
$$\forall\, t \in \{1, \dots, n_T\}, i \in \{1, \dots, n_N\}, i' \in \{1, \dots, n_N\}, i'' \in \{1, \dots, n_N\}$$

Origin to destination connection, equations (47) and (48): for every year t, exactly one link l that is part ($z_{lrii't} = 1$) of the travel path of any O-D pair ii' connects to origin node i, and exactly one connects to destination node i'.

$$\sum_{1 \le l \le n_L} \sum_{1 \le r \le n_R} z_{lrii't} \alpha_{il} = 1 \quad \forall\, t \in \{1, \dots, n_T\}, i \in \{1, \dots, n_N\}, i' \in \{1, \dots, n_N\} \tag{47}$$

$$\sum_{1 \le l \le n_L} \sum_{1 \le r \le n_R} z_{lrii't} \alpha_{i'l} = 1 \quad \forall\, t \in \{1, \dots, n_T\}, i \in \{1, \dots, n_N\}, i' \in \{1, \dots, n_N\} \tag{48}$$

Travel path connectivity, equation (49): for every year t, the travel path (given by the $z_{lrii't}$ binaries) of any O-D pair ii' must form a single connected sequence with exactly two extreme nodes.

$$n_N - \sum_{1 \le i'' \le n_N} \left(-1 + \sum_{1 \le l \le n_L} \sum_{1 \le r \le n_R} z_{lrii't} \alpha_{i''l} \right)^2 = 2 \tag{49}$$

$$\forall\, t \in \{1, \dots, n_T\}, i \in \{1, \dots, n_N\}, i' \in \{1, \dots, n_N\}$$

The logic operations for every year, origin and destination unfold as follows:

5. The two innermost summations evaluate to 0, 1 or 2, depending on the number of links l that are part ($z_{lrii't} = 1$) of the travel path of O-D pair ii', that are connected ($\alpha_{i''l} = 1$) to a node i'' during year t. It cannot reach 3 or more because travel path branching is forbidden by equation (46).
6. Each of those ternary results is reduced by one and squared, which converts it into a binary with a value of zero if the node i'' is an extreme node of the operating path of line r, and a value of one otherwise (internal or non-visited nodes).
7. The outermost summation yields the number of nodes i'' in the network that are not extremes of the travel path of O-D pair ii'.

8. Subtracting the total number of nodes of the network minus the above result yields the total number of extreme nodes of the travel path of O-D pair ii' in year t. If equal to two, the condition is met.

Equation (49) still fails to prevent a travel path from having disconnected circuit segments, additional to the linear one that connects the origin to the destination. Explicitly formulating that constraint would result in very cumbersome equations, so it is assumed that any algorithm chosen for optimizing the travel paths will by itself rule out satellite circuits for being useless travel time increasers.

Chapter 3: Solution method

The sets of construction ($\hat{x}$), operation ($\hat{y}$), headway ($\hat{h}$) and user travel path ($\hat{z}$) decision variables are optimized primarily by a genetic algorithm (GA). What it optimizes is the link-building sequence, which translates into the $\hat{x}$ and $\hat{y}$ variables by means of rules described later to comply with the constraints mentioned in section 2.7. When decoding each chromosome, the GA implements an iterative greedy heuristic involving a quickest path algorithm to optimize its $\hat{z}$ and $\hat{h}$ for each year, as will be explained. The main characteristics and protocols of the GA used are the following:

Encoding: each chromosome contains n_L positions, one for each link, specifying its building priority as a number between 0 and 1: the greater the priority, the sooner the link will be built, subject to network connectivity Mutations described later may push positional values of the chromosomes out of the $[0,1]$ range, but this is not considered inconvenient because they are just compared in terms of their signed magnitude.

Initial generation: all the positions of each chromosome in the first generation is randomly drawn from an $U(0,1)$ distribution.

Fitness values: each chromosome's fitness is the discounted societal net benefit Y that its building priority values generate. The decoding process is described in section 3.1.

Parent selection: the two best chromosomes of each generation are kept for the next one, and the rest are created by drawing from the previous population pairs of "parent" chromosomes and producing from each pair two child chromosomes using the crossover protocol, until reaching the population size again. Each parent is individually drawn from the population by randomly sampling, with equal probability, three chromosomes from the population, and selecting as a parent the one with the highest fitness.

Crossover: two child chromosomes are created by taking the sequences of values of their parent chromosomes and performing a two-point crossover by swapping a random slice of them. Crossover occurs with some probability, otherwise each of two child chromosomes inherit one of their parent's untouched genomes.

Mutation: with some probability, each individual child chromosome can be mutated after being generated, by having some of its values independently altered (each with a 5% probability) by being added a random *gaussian noise* value sampled from $N(0,0.2)$.

3.1.Genetic algorithm chromosome decoding

The decoding of a chromosome starts with three preparatory steps:

1. Translating the priorities into a feasible *construction sequence* $\mathbb{S}$ that is contiguous (connected) in all years. For this, the algorithm iteratively adds to $\mathbb{S}$ the link with the highest priority that has not been added yet and is either 1) the first link added, or 2) adjacent to any link that has already been added to the sequence. This action is repeated until all links have been added.
2. Initialize all link and node construction binaries as zero (unbuilt).
3. Set all the left-end bus routes as operating in all links of their respective rail line's maximum operating path, and all other lines as not operating, in year 1. This involves setting the operational binaries y_{lr1} and initializing the SAN graph (which will be explained in section 3.2) to contain only nodes and edges

related to the left-end bus routes in all the links they will operate in the first year.

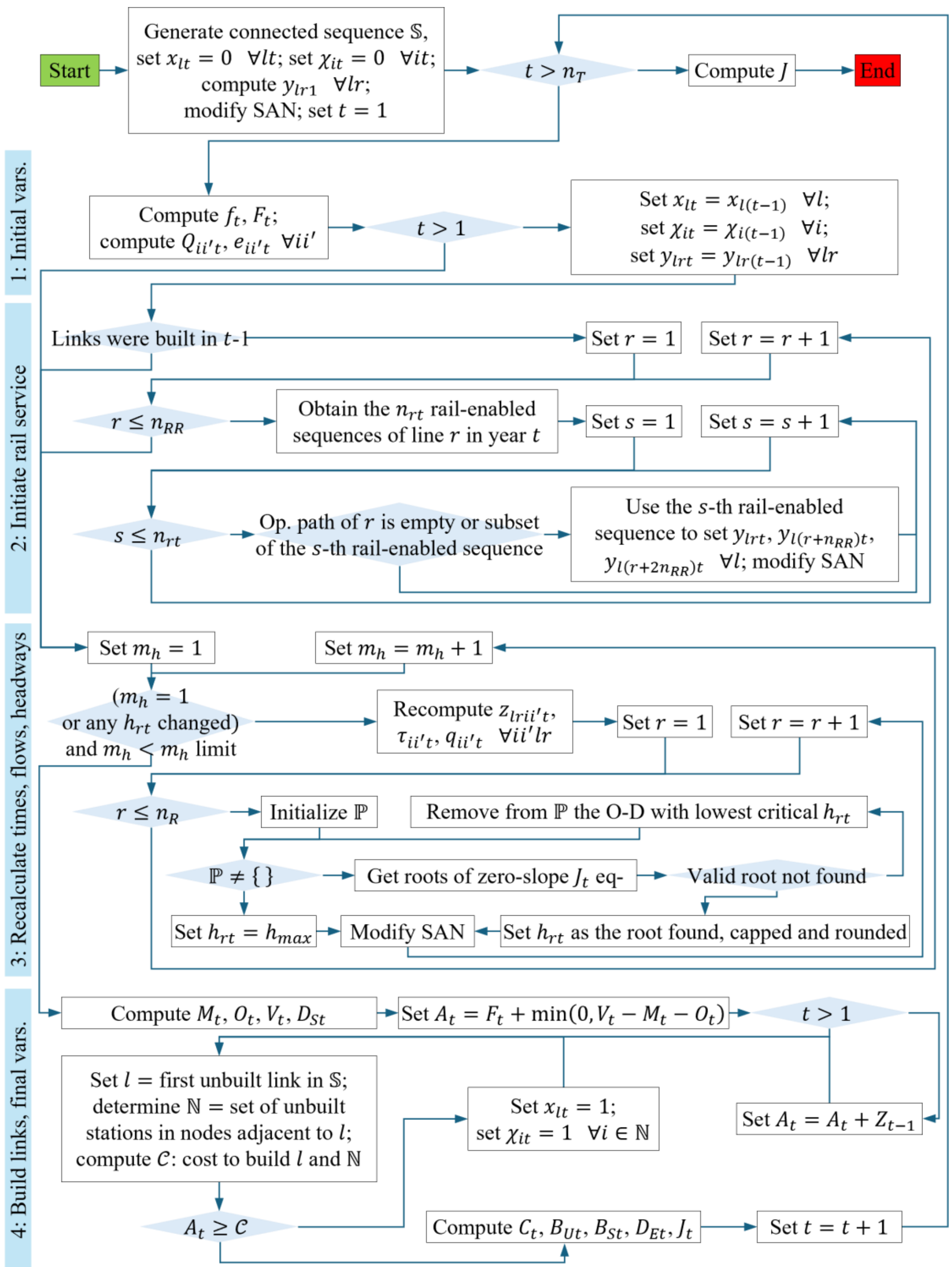

Figure 6. Flow chart of the chromosome-decoding algorithm.

Then, the building sequence is further decoded to generate its associated $\hat{x}$, $\hat{y}$, $\hat{h}$, $\hat{z}$ and all intermediate variables by iterating years 1 through n_T of the timeframe and obtain all calculated variables for each year. The process for each year is subdivided into four steps as shown in the flowchart in Figure 6, in which the right arrows from decisions boxes represent “True” and the down arrows are “False”.

3.1.1. Compute variables that only depend on t

Fare f_t, funding F_t, maximum theoretical demand $Q_{ii't}$ for every O-D pair, and maximum acceptable impedance $e_{ii't}$ for every O-D pair are calculated for current year t. Link construction (x_{lt}), node construction (χ_{it}) and line-link operation (y_{lrt}) binaries for the current year are initialized as copies of their respective versions of the previous year, if t is not the first year.

3.1.2. Initiate rail service on links built the previous year

If any link was built in year $t-1$, which means t is not the first year, then the algorithm iterates over all rail lines r, and for each of them iterates over all its isolated rail-enabled segments: sequences of adjacent built links that are disconnected from other sequences.

Three cases are possible when determining the new operating path of a transit line r:

1. The line was already operating, in which case one and only one of the isolated segments contains (expands) its current operating path (links l for which $y_{lr(t-1)} = 1$). That segment will be the line’s new operating path for year t.
2. The line was not operating and there is only one isolated segment. That segment will be the line’s new operating path for year t.

3. The line was not operating and there is more than one isolated segment. In this is a rare event, the first of the segments (the ordering proceeds from the indexing of the links) is arbitrarily chosen as the line's new operating path for year t.

The new operating path is set for rail line r, and the operating paths of its corresponding bus lines $r + n_{RR}$ and $r + 2n_{RR}$ are also updated (typically shortened) to comply with the line coverage restrictions mentioned in section 2.7. This involves two actions: modifying the respective operational binaries y_{lrt} and modifying the SAN, which is a graph representation of the network that is used for user travel time computations and will be explained in section 3.2. New edges are created in the SAN for rail line r in the links where it will start operating, and the edges of the corresponding bus lines in those edges are removed. If the rail line was not operating previously and its new operation splits its existing bus line in two, then the operating path of the existing bus line (the "left-end" bus line of the triad, i.e. line $r + n_{RR}$) is modified, both in terms of its y_{lrt} binaries and its edges in the SAN, to operate only in the links to the left of the newly initiated rail line, while the "right-end" bus line ($r + 2n_{RR}$) will operate in the links to the right of the rail line.

3.1.3. Recompute travel times, flows and headways

First-level iterations are performed, in each of which the headways of all transit lines are sequentially optimized. The reason to perform this more than once is because modifying the headway of a line might turn suboptimal a previously determined one. Iterations stop when preset iterations limit is reached or until none of the headways changed with respect to the previous iteration. On each iteration, the following steps are followed:

1. The travel paths $z_{lrii't}$ are optimized for all O-D pairs to minimize travel times $\tau_{ii't}$ for the current operating state of the transit lines. This is done by implementing Dijkstra's algorithm on the SAN graph, as will be referenced in section 3.2.
2. Hourly flows $q_{ii't}$ are computed for every O-D pair with equation the demand elasticity equation, (13).
3. Second-level iterations are performed on the transit lines. In each of them, the headway of transit line r is individually optimized by treating the headways of the other lines as constants and using a cubic equation solving algorithm (explanation omitted) to find a *valid* (real and positive) root for equation (73), which will be derived in section 3.3. To find the valid root for a transit line r, third-level iterations are performed, following these rules:
    a. The first third-level iteration attempts to find the valid root using the set ($\mathbb{P}$) of all O-D pairs that use r in that year, as per the travel paths computation of the current first-level iteration. If two real, different, positive roots are found, the smallest corresponds to the headway that maximizes the net societal benefit Y for that year (the other is guaranteed to be a local minimum). If a valid root is not found the algorithm discards from $\mathbb{P}$ the O-D pair with the lowest critical headway for line r, i.e., the one that causes zero hourly flow for that pair. Discarding a pair effectively means that the optimal headway that will be found causes zero flow in it.

b. The next and subsequent third-level iterations reattempt to find a valid root with the updated set $\mathbb{P}$ of O-D pairs that use line r.

c. When a valid root is first found, it is capped with the preset bounds h_{min} and h_{max}, rounded to the nearest half minute, and set as the headway of line r for year t. Third-level iterations end.

d. If at any moment $\mathbb{P}$ is an empty set, which means that no O-D pair uses line r as part of its user travel path for any headway value, then its headway is arbitrarily set as h_{max} and third-level iterations.

e. After completing the third-level iterations, the weights of boarding edges of line r in the SAN are updated as half the new headway found.

3.1.4. Build new links and calculate the remaining variables

Four main operations are performed in order:

1. Maintenance costs (M_t), operating costs (O_t) and as fare revenue (V_t) are calculated for the current year using equations (19), (20) and (23) respectively. Initial balance D_{St} is computed as the year's external funding plus previous year's ending balance, if $t > 1$.
2. The construction budget A_t is computed, i.e. the funds that are available for construction without compromising the ones required to cover the year's net expenses. It is calculated as the year's initial balance D_{St} plus the current year's "net continued cost" (revenue minus maintenance and operation costs) if negative. The year's net continued cost cannot be considered for construction budgeting if it is positive because construction costs are incurred at the

beginning of the year, before money flows for the year's operation, maintenance and revenue take place.

3. As many links and stations as allowed by A_t are built, in the order specified by the link construction sequence $\mathbb{S}$. This is determined iteratively by considering each time the first unbuilt link l in $\mathbb{S}$ and its unbuilt adjacent stations i (which might be zero, one or two stations). If A_t equals or exceeds their combined building costs, the construction binaries of the link x_{lt} and the stations χ_{it} are set as 1.
4. The year's construction costs (C_t), user surplus (B_{Ut}), supplier net benefit (B_{St}), final balance (D_{Et}) and societal net benefit (Y_t) are computed with equations (18), (14), (26), (24) and (28) respectively.

After computing all years, each with the four steps described in the sections above, the net present value (Y) of the total societal benefit is determined with equation (29). This is the fitness value of the chromosome.

3.2.User travel path optimization: Service and Actions Network (SAN)

To ease the computation of user paths and travel times, a directed graph called *Service and Actions Network* (SAN) is created and updated as the network is developed, to accurately represent the actions done by the users to move in the network using the service provided by the transit lines at any particular stage of the network's development.

There are two types of nodes in the SAN: "station nodes" which are all the n_N nodes from all previous mentions, and "dummy nodes" which help separate the edges that refer to different movements by the users: waiting, alighting, dwelling at stations, and moving from one station to another in a transit vehicle. At every year, each transit line has four dummy nodes for each link in its current operating path: an "outbound departure" and an "inbound arrival" node connected to the upstream station node of the link (considering its arbitrarily defined outbound direction), and an "outbound arrival" and an "inbound departure" node connected to the downstream station node of the link.

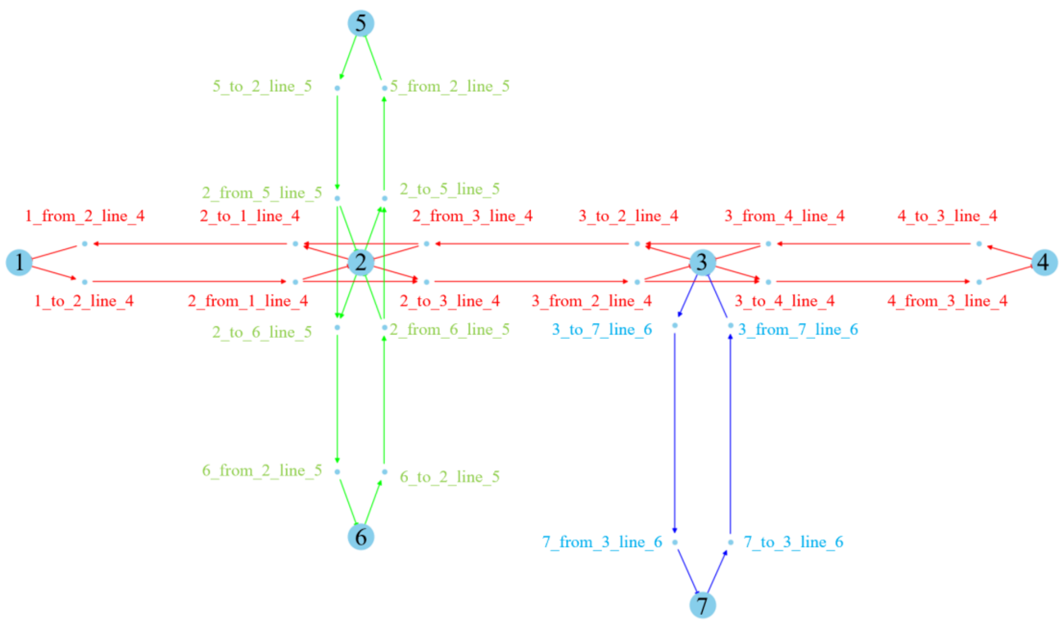

Figure 7. Example Service and Actions Network (SAN).

The small circles in Figure 7 represent the dummy nodes of the initial state a network composed of three (extremely short) bus lines $4, 5, 6$ (red, green and blue), whose indices starts at 4 according to indexing established for line triads in equation (5) (their respective rail lines would be $1, 2, 3$). Their names, which are colored according to their line r, follow the node naming convention used in the solution algorithm programmed, for a link connecting upstream station node n_1 with downstream station node n_2:

1. Outbound departure dummy node: $n_1\_to\_n_2\_line\_r$
2. Outbound arrival dummy node: $n_2\_from\_n_1\_line\_r$
3. Inbound departure dummy node: $n_2\_to\_n_1\_line\_r$
4. Inbound arrival dummy node: $n_1\_from\_n_2\_line\_r$

The following six edges are defined for each link of the operating path:

1. Two waiting/boarding edges, one from the upstream station node to the outbound departure node, and another from the downstream station node to the inbound departure node. Their weights are set as half the line's headway.
2. Two moving edges, one from the outbound departure node to the outbound arrival node, and one from the inbound departure node to the inbound arrival node. Their weights are set as the sum of the first two terms of the expression for p_{lr} in equation (10), which refer to the traversing time of the link and the acceleration and deceleration delay of the transit line.
3. Two deboarding edges, one from the outbound arrival node to the upstream station node, and another from the inbound arrival node to the downstream station node. Their weights are set as equal to the alighting time, but even if it was zero the existence of these links is necessary to prevent alightings to incorrectly incur dwell time.

Additionally, between any two adjacent links of the line's operating path there are two dwell edges that connect the outbound arrival node of the downstream link with the outbound departure node of the upstream link, and the inbound arrival node of the downstream link with the inbound departure node of the upstream link. Their weights are set as the third term of the expression for p_{lr} in equation (10), which refers to the dwell time of the transit line depending on whether it is a bus of a rail line.

The selection of the quickest travel path along the network for users from every node i to every node i' in year t (i.e., the $z_{lrii't}$ binaries) and the associated travel times $\tau_{ii't}$ are determined by executing the directed-network version of Dijkstra's algorithm in the SAN at its developmental state by that year, once for every origin node i'. Dijkstra's is a broadly used deterministic shortest path algorithm that yields the optimized paths and travel times (or cumulative weight, in general) from a source node to every other reachable node of the network. Its explanation is omitted here for conciseness, since it was implemented using the function `single_source_dijkstra()` from the Python library `networkx`.

Note that the existence of dedicated waiting and dwell edges is what allows Dijkstra's algorithm to determine quickest paths while accurately assessing travel times in a network that allows transfers. Alternatively, the quickest path algorithm could have instead been performed on a network with just the station nodes and single bidirectional edges between them, with the moving times as weights, and the dwell and waiting times could have been accounted for after the travel path optimization. However, the resulting optimized paths would have been different and potentially suboptimal, which is why this more computationally demanding network modeling approach was favored since execution time is typically not a constraint at transit planning level.

As mentioned before, the SAN is initialized with edges representing the operation of the left-end bus lines in year 1, and they are progressively replaced link-wise with edges pertaining to their respective rail lines (and right-end bus lines, if necessary).

3.3.Headway optimization

We are interested in formulating how the societal net benefit in a year t responds to the headways of the lines that operate on that year, with the intention of jointly optimizing these headways. Since line ridership depends on the headways of various lines, the algebraic manipulations are more intricate than more conventional headway-optimizing approaches.

3.3.1. Sensitivity of User Surplus to headways

Let's recall equations (13) and (14) for the hourly passenger flows $q_{ii't}$ and the annual user surplus B_{Ut} of a year t, respectively. When inserting (13) in (14), the latter turns into (50).

$$q_{ii't} = Q_{ii't} \frac{e_{ii't} - f_t - u\tau_{ii't}}{e_{ii't}} \tag{13}$$

$$B_{Ut} = \eta \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} q_{ii't} \frac{e_{ii't} - f_t - u\tau_{ii't}}{2} \tag{14}$$

$$B_{Ut} = \eta \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \left(\frac{Q_{ii't}}{2e_{ii't}} (e_{ii't} - f_t - u\tau_{ii't})^2 \right) \tag{50}$$

Meanwhile, the travel time $\tau_{ii't}$ for O-D pair ii' in a given year t as per equation (11) can be rearranged into the convenient form displayed in equation (51), which isolates the part of the travel time that does not depends on transit lines headways from the part that does. The former is denoted as $\tilde{\tau}_{ii't}$, and as shown in (52) it comprises the in-vehicle travel time and the alighting time for all transit lines employed by users going from node i to node i' in year t. As a reminder, the indicator $\mathbb{1}_{rii't}$ equals one if line r is part of the travel path for O-D pair ii' in year t, as was defined in equation (12).

$$\tau_{ii't} = \tilde{\tau}_{ii't} + \sum_{1 \leq r \leq n_R} \frac{\mathbb{1}_{rii't} h_{rt}}{2} \tag{51}$$

$$\tilde{\tau}_{ii't} = \sum_{1 \leq r \leq n_R} \left(\left(\sum_{1 \leq l \leq n_L} z_{lrii't}\, p_{lr} \right) + \mathbb{1}_{rii't} \big(b_R \theta_r + b_B (1 - \theta_r) \big) \right) \tag{52}$$

Inserting (51) in (50) produces the following:

$$B_{Ut} = \eta \sum_{1 \leq i \leq n_N} \sum_{1 \leq i' \leq n_N} \left(\frac{Q_{ii't}}{2e_{ii't}} \left(e_{ii't} - f_t - u\tilde{\tau}_{ii't} - u \sum_{1 \leq r \leq n_R} \frac{\mathbb{1}_{rii't} h_{rt}}{2} \right)^2 \right) \tag{53}$$

To expand the quadratic expression above and isolate terms based on their powers of headways, the identity presented in (54) is used to express the square of a summation of terms that include headways from different lines.

$$\left(\sum_{1 \leq r \leq n_R} \frac{\mathbb{1}_{rii't} h_{rt}}{2} \right)^2 = \sum_{1 \leq r \leq n_R} \frac{\mathbb{1}_{rii't} {h_{rt}}^2}{4} + \sum_{1 \leq r \leq n_R} \sum_{\substack{1 \leq r' \leq n_R \\ r' \neq r}} \frac{\mathbb{1}_{rii't} \mathbb{1}_{r'ii't} h_{rt} h_{r't}}{2} \tag{54}$$

Note that this identity introduces not only squared headways, but also crossed terms where the headways of two different lines are multiplied. The latter responds to the fact that the marginal effect of increasing an individual line's headway by some amount is dependent on the ridership of the line, which is a function of not only the headway of that line but of other lines' as well.

After expanding the squared parenthesis in (53), employing (54), and inserting into the resulting inner terms the two summations that refer to origin and destination nodes, expression (55) is reached.

$$B_{Ut} = \eta \left(A_t - \sum_{1 \le r \le n_R} B_{rt} h_{rt} + \sum_{1 \le r \le n_R} \sum_{\substack{1 \le r' \le n_R \\ r' \neq r}} C_{rr't} h_{rt} h_{r't} + \sum_{1 \le r \le n_R} D_{rt} {h_{rt}}^2 \right) \tag{55}$$

$$A_t = \frac{1}{2} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \left(\frac{Q_{ii't}}{e_{ii't}} \right) (e_{ii't} - f_t - u \tilde{\tau}_{ii't})^2 \tag{56}$$

$$B_{rt} = \frac{u}{2} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \mathbb{1}_{rii't} \left(\frac{Q_{ii't}}{e_{ii't}} \right) (e_{ii't} - f_t - u \tilde{\tau}_{ii't}) \tag{57}$$

$$C_{rr't} = \frac{u^2}{4} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \mathbb{1}_{rii't} \mathbb{1}_{r'ii't} \left(\frac{Q_{ii't}}{e_{ii't}} \right) \tag{58}$$

$$D_{rt} = \frac{u^2}{8} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \mathbb{1}_{rii't} \left(\frac{Q_{ii't}}{e_{ii't}} \right) \tag{59}$$

Its four line-year-specific coefficients A_t, B_{rt}, $C_{rr't}$, D_{rt}, respectively multiply the expression terms according to the four types in which they can be categorized according to their dependance on headways:

1. One constant term: independent from any headway.
2. At most n_R first-order terms: each of them is proportional to the headway of one line.

3. At most $n_R(n_R - 1)$ second-order "crossed" terms: each of them is proportional to the product of the headways of two (nonequal) lines.
4. At most n_R second-order "squared" terms: each of them is proportional to the square of the headway of one line.

The actual number of terms of each type is determined by the line usage indicators $\mathbb{1}_{rii't}$, which as explained after equation (12) are binaries that evaluate to 1 if O-D pair ii' uses line r as part of its travel path in year t.

Equation (55) is a hyper-paraboloid $\mathbb{R}^{n_R} \to \mathbb{R}$ function where n_R is the number of all possible transit lines, including rail lines and the two corresponding bus lines of each of them. This hyper-paraboloid is convex (proof is omitted) and therefore guaranteed to have only one local and global minimum at the optimal set of values of the n_R decision variables (the lines' headways h_{rt}). As mentioned in section 3.1.3, solution logic chosen to jointly optimize the headways is iterative, optimizing the headway of one line at a time. Therefore, equation (55) needs to be expressed in terms of one single generic headway h_{rt}, treating the headways of the other lines as constant, as shown in equation (60). Then, its partial derivative with respect to a particular h_{rt} is determined as in (61).

$$B_{Ut} = \eta \left(A_t - B_{rt}h_{rt} - \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} B_{r't}h_{r't} + h_{rt} \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} C_{rr't}h_{r't} \right.$$
$$+ \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} \sum_{\substack{1 \le r'' \le n_R \\ r'' \ne r'}} C_{r'r''t}h_{r't}h_{r''t} + D_{rt}{h_{rt}}^2 \tag{60}$$
$$\left. + \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} D_{r't}{h_{r't}}^2 \right)$$

$$\frac{\partial B_{Ut}}{\partial h_{rt}} = \eta \left(-B_{rt} + \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} C_{rr't}h_{r't} + 2D_{rt}h_{rt} \right) \tag{61}$$

3.3.2. Sensitivity of Supplier Net Benefit to headways

The only two components of the supplier's net benefit B_{St} formulated in equation (26) that depend on headways are the operating cost as defined in equation (20) and revenue as defined in equation (23). Inserting those two equations in (26), and removing the fleet size rounding up for simplicity, gives:

$$B_{St} = \eta \left(-C_t - M_t - \eta \sum_{1 \le r \le n_R} \frac{\rho_{rt}\big(o_R\theta_r + o_B(1-\theta_r)\big)}{h_{rt}} \right.$$
$$\left. + \eta f_t \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} q_{ii't} \right) \tag{62}$$

In the above equation, each O-D flow $q_{ii't}$ can be expressed in terms of the travel times using (13), and the use of $\tilde{\tau}_{ii't}$ as defined in (52) allows to separate each of those travel times into a component that depends on headways and another than do not.

$$q_{ii't} = \frac{Q_{ii't}}{e_{ii't}}(e_{ii't} - f_t - u\tilde{\tau}_{ii't}) - \frac{Q_{ii't}}{e_{ii't}}\left(u \sum_{1 \le r \le n_R} \frac{\mathbb{1}_{rii't} h_{rt}}{2}\right) \tag{63}$$

Thus, the partial derivative of (62) with respect to the headway h_{rt} of one transit line is:

$$\frac{\partial B_{St}}{\partial h_{rt}} = \eta\left(\frac{\rho_{rt}\big(o_R\theta_r + o_B(1-\theta_r)\big)}{{h_{rt}}^2} - \frac{uf_t}{2}\sum_{1 \le i \le n_N}\sum_{1 \le i' \le n_N} \mathbb{1}_{rii't}\frac{Q_{ii't}}{e_{ii't}}\right) \tag{64}$$

3.3.3. Sensitivity of a year's societal net benefit to headways

Finally, the partial derivative of the societal net benefit Y_t of year t with respect to the headway h_{rt} of one particular transit line r during that year is found by adding (61) and (64):

$$\begin{aligned}\frac{\partial(B_{Ut} + B_{St})}{\partial h_{rt}} = \eta\Bigg(&-B_{rt} + \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} C_{rr't} h_{r't} + 2D_{rt}h_{rt} \\ &+ \frac{\rho_{rt}\big(o_R\theta_r + o_B(1-\theta_r)\big)}{{h_{rt}}^2} - \frac{uf_t}{2}\sum_{1 \le i \le n_N}\sum_{1 \le i' \le n_N} \mathbb{1}_{rii't}\frac{Q_{ii't}}{e_{ii't}}\Bigg)\end{aligned} \tag{65}$$

Equating the above expression to zero produces equation (66).

$$E_{rt} + F_{rt}h_{rt} + \frac{G_{rt}}{{h_{rt}}^2} = 0 \tag{66}$$

Two of the three coefficients of the above equation can be straightforwardly expressed in terms of the input or calculated variables:

$$F_{rt} = 2D_{rt} = \frac{u^2}{4} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \mathbb{1}_{rii't} \left(\frac{Q_{ii't}}{e_{ii't}} \right) \tag{67}$$

$$G_{rt} = \rho_{rt} \big(o_R \theta_r + o_B (1 - \theta_r) \big) \tag{68}$$

The third coefficient E_{rt} deserves additional attention. It is initially expressed as in (69) below, which is then expanded to (70). By introducing $\tilde{\tau}_{rii't}$ as the travel time for O-D pair ii' in year t including all waiting times involved *except* that of transit line r, as defined in (71), the expression for E_{rt} can be simplified as (72).

$$E_{rt} = -B_{rt} + \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} C_{rr't} h_{r't} - \frac{u f_t}{2} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \mathbb{1}_{rii't} \frac{Q_{ii't}}{e_{ii't}} \tag{69}$$

$$\Rightarrow E_{rt} = -\frac{u}{2} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \mathbb{1}_{rii't} \left(\frac{Q_{ii't}}{e_{ii't}} \right) \left(e_{ii't} - f_t - u\tilde{\tau}_{ii't} \right.$$
$$\left. - u \left(\sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} \mathbb{1}_{r'ii't} \frac{h_{r't}}{2} \right) + f_t \right) \tag{70}$$

$$\tilde{\tau}_{rii't} = \tilde{\tau}_{ii't} + \sum_{\substack{1 \le r' \le n_R \\ r' \ne r}} \mathbb{1}_{r'ii't} \frac{h_{r't}}{2} \tag{71}$$

$$\Rightarrow E_{rt} = -\frac{u}{2} \sum_{1 \le i \le n_N} \sum_{1 \le i' \le n_N} \mathbb{1}_{rii't} \left(\frac{Q_{ii't}}{e_{ii't}} \right) (e_{ii't} - u\tilde{\tau}_{rii't}) \tag{72}$$

To find the roots of equation (66), the following cubic equation is considered, which shares the same roots.

$$F_{rt} {h_{rt}}^3 + E_{rt} {h_{rt}}^2 + G_{rt} = 0 \tag{73}$$

The sought minimum occurs at the smallest of the two real, positive roots of the equation. If no such two roots exist, it is due to the hourly flow of at least one O-D pair becoming negative, which of course makes no sense in real life. If that happens, it means that the optimal headway of line r in year t occurs at a value that causes zero flow for at least one O-D pair (proof of biimplication between both situations is omitted in this document). The iterative logic described in section 3.1.3 sequentially discards those O-D pairs from the calculation until an optimal headway is found.

Chapter 4: Numerical results

4.1.<u>Context of the case study</u>

A real-life scenario is analyzed to showcase the model's performance in optimizing the building sequence for the rail links of a multi-line network. The studied case is located in the most densely populated part of the western Greater Metropolitan Area in Costa Rica, comprising the cities and surroundings of San Jose, Heredia, and Alajuela.

Historically, satellite towns in this area have grown organically to the point of merging together with little to no region-wise urban planning. Thus, the roads that connect them are sinuous, narrow, highly interrupted by intersections, and often congested by local traffic because they run through town centers. Moreover, scarce budget and long-term planning to build bridges over the many rivers running westwards has pushed the bus system to favor radial, branched connections towards the center of San Jose, the largest and most populous city in the area. These two factors often hinder achieving adequate bus speeds and make trips between some suburbs (even adjacent ones) highly circuitous. National and local transportation authorities are working together to sectorize bus lines, which in the near future will enable a hub-and-spoke service with fast and frequent service in trunk lines, and ease of transfers between them and with shorter collector lines thanks to integrated electronic payment.

On the other hand, the urban railroad currently runs over three repurposed freight rail branches that meet in San Jose, built more than a century ago. Travel speeds are low due to the track condition and geometry, as well as the many grade crossings, and dwell times are high because some of the rolling stock includes refurbished commuter rail cars with smaller door area than metro cars. The national railways authority has made efforts to roll out a project to enhance service with better stations and cars, fewer grade crossings and slightly improved geometry along the existing right of way. However, critics have pointed out how this current layout is dissociated from major urban axes, and thus sticking to it will still constrain optimal station placement.

4.2.Inputs of the case study

4.2.1. Network geometry and topology parameters

The proposed metro-type urban rail network consists of 51 stations placed near major population, industry and commerce centers of the studied region, while also aiming to reduce average access distance by collector buses (not calculated). The total number of inter-station transit links is 57. To reduce construction costs, above-grade construction was assumed for most stations and links, with some underground segments mostly through important and dense town centers. Therefore, the tentative rail alignment proposed intuitively intends to minimize expropriation costs and benefit cut-and-cover construction along some few straight roads. The length of the bus connection between two stations using the existing roads was estimated to be 1.5 times its respective rail distance.

The geolocations of all nodes, and the rail and bus length for the links are shown in Table 9 and Table 10 in the appendices, respectively. For descriptive purposes, Table 3 summarizes the distribution of the values used for the rail (d_{Rl}) and bus (d_{Bl},) distance of the set of links used.

Table 3. Statistical descriptors of the inter-station links used as input for the model.

Measure	d_{Rl} (km)	d_{Bl} (km)
Minimum	0.71	1.06
Maximum	5.59	8.39
Average	2.15	3.23
Standard deviation	0.98	1.47

4.2.2. Transit service parameters

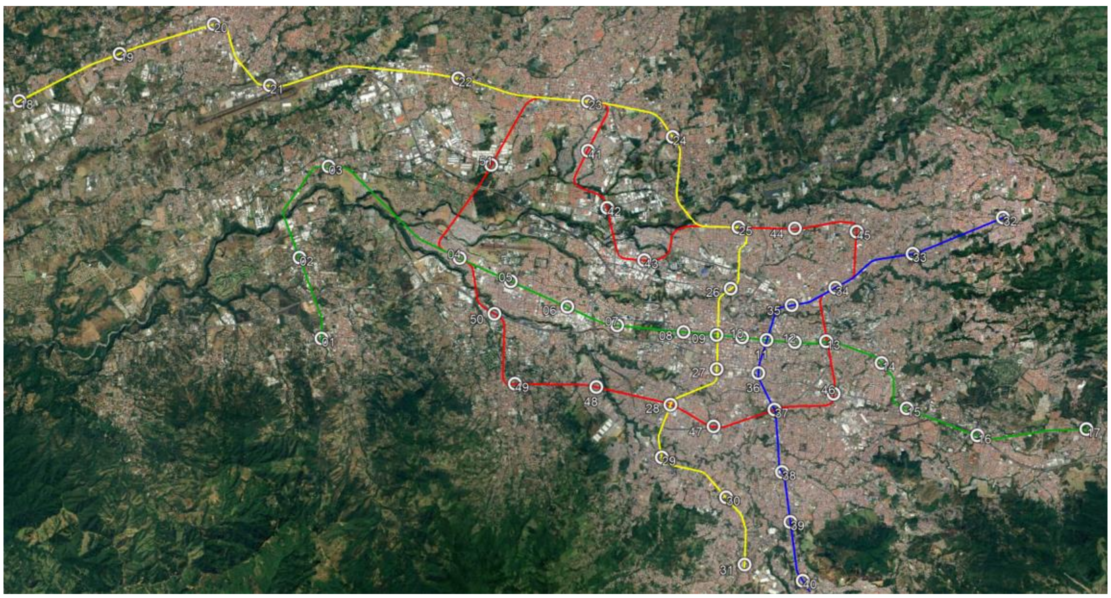

Figure 8. Rail lines proposed for the case study.
Background obtained from Google Earth (2024)

Four lines are proposed: three *through* lines that meet in San Jose, plus a fourth, circular line that visits additional outskirts. This layout creates nine transfer stations, each servicing two different lines. The lines would start operating as bus lines, each servicing the future stations of its respective rail line in the same order, and will be progressively replaced as the links and stations are built. Riding in the network would cost a flat fare per one-way trip, regardless of origin and destination, with unlimited transfers even between modes. The lines were designed so that no line visits two adjacent stations that are also visited adjacently by another line, because the model cannot yet acknowledge the reduced waiting times that this situation induces for some O-D pairs with the introduction of interchangeable alternative lines.

Table 4. Characteristics of the transit lines used as input for the model.

Line	Color	Number of stations	Total rail distance (km)	Total bus distance (km)
1	Green	16	30.25	45.38
2	Yellow	14	37.32	55.99
3	Blue	9	15.07	22.61
4	Red	18	39.92	59.88
Total		51 (unique)	122.57	183.85

4.2.3. Demand parameters

To generate credible $Q_{ii'0}$ (initial passenger flow) and $e_{ii'0}$ (initial maximum acceptable impedance) demand parameters to each of the 2,550 O-D pairs, a circular catchment zone with a radius of 2.5 km was created around each station, with Voronoi tessellation employed for their intersections. The sum of all catchment zones is termed the "total catchment region". Then, data from the latest (2011) nation-wide census were used to estimate the relative potential of each catchment zone to be the origin (<u>generative potential</u>) and the destination (<u>attractive potential</u>) of round trips, e.g., where people reside and where people travel to, respectively. This approach makes the simplistic assumption that all daily commuting is strictly two-legged, meaning that travelers depart from their residence to some destination zone in the city and then return home from that same zone, thus using transit only two separate times every day without intermediate destinations.

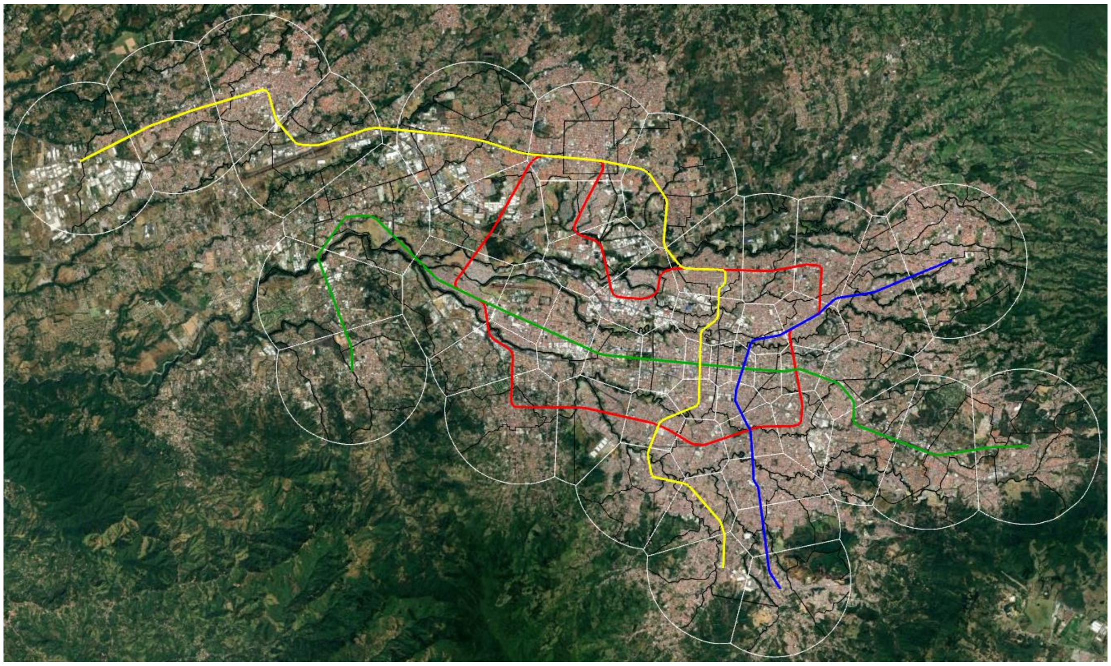

Figure 9. Geographic districts (black) and station catchment zones (white).

Background obtained from Google Earth (2024)

Generative potential: for obtaining the generative potential $\omega_i \in [0,1[$ of each node i relative to the other nodes', the number of employed persons in the 109 geographic districts totally or partially contained in the total catchment region was used, according to information from the 2011 census (Costa Rican Institute of Statistics and Census, 2016). This metric excludes non-work-related trips, but it was deemed useful to compare overall trip generation across zones in relative terms. Then, two assumptions were made:

1. 100% of the employed persons of each district reside homogeneously in the area of that district that is covered by the catchment zones.

2. The number of daily round trips per person to destinations that are in the total catchment region is homogeneous across persons from different districts and within the same district, say one round trip per day.

The generative potential ω_i of each node was then calculated as the number of daily round trips departing from its catchment zone, divided by the sum of the daily round trips departing from the total catchment region. Note that $\sum_{1\leq i\leq n_N} \omega_i = 1$.

Attractive potential: an analogous reasoning was followed for the relative attractive potential $\kappa_i \in [0,1[$ of each node i, but using the number of daily job-related round trips destined at each of the 22 cantons (territorial subdivisions larger than districts) totally or partially contained in the total catchment region, gathered from the same census (Pujol, Perez, & Aguero, 2016), because district-level attraction data was not available. Similarly, $\sum_{1\leq i\leq n_N} \kappa_i = 1$

Initial passenger flows: finally, the initial (at "year zero") maximum hourly passenger flow $Q_{ii'0}$ from node i to node i' was estimated by taking an input parameter $\mathbb{Q}$ that is meant to represent the number of total daily round trips generated in the total catchment region divided by the number of daily operating hours, and multiplying that parameter by the sum of the "joint trip potential" $\omega_i\kappa_{i'}$ of outbound trips from i to i' and the joint trip potential $\omega_{i'}\kappa_i$ of returning trips from i' to i. The opposite hourly flow $Q_{i'i0}$ is identical, making the hourly flow matrix symmetrical. Flows in its diagonal are forced to zero since trips within catchment zones do not use the transit lines evaluated, and therefore the sum of all flows does not add up to $\mathbb{Q}$. All hourly flows were rounded to the nearest integer.

$$Q_{ii'0} = Q_{i'i0} = \begin{cases} \mathbb{Q}(\omega_i \kappa_{i'} + \omega_{i'} \kappa_i) & if\ i \neq i' \\ 0 & if\ i = i' \end{cases} \tag{13}$$

Initial maximum acceptable impedance: the initial (at "year zero") maximum acceptable impedance $e_{ii'0}$ from node i to node i' was modeled as a baseline impedance tolerance ε_B plus an additional impedance tolerance ε_M per each kilometer of Euclidean distance $\delta_{ii'}$ between the stations.

$$e_{ii'0} = e_{i'i0} = \begin{cases} \varepsilon_B + \varepsilon_M \delta_{ii'} & if\ i \neq i' \\ 0 & if\ i = i' \end{cases} \tag{74}$$

Summary of values used for demand parameters: an exhaustive enumeration of the values of $Q_{ii'0}$ and $e_{ii'0}$ for all 2550 O-D pairs is omitted for conciseness. For reference, however, the ω_i and κ_i values of each of the 51 nodes, from which $Q_{ii'0}$ stem, are listed on Table 9 in the appendices. The same table also lists the coordinates of each node in the CRTM05 coordinate system, which is cartesian and in meters, allowing to directly derive all internode Euclidean distances $\delta_{ii'}$ from which $e_{ii'0}$ were calculated. For descriptive purposes about the inputs used, Table 5 and Figure 10 summarize the distribution of the values used for demand parameters $Q_{ii'0}$ and $e_{ii'0}$.

Table 5. Statistical descriptors of the demand parameters of "year zero".

Measure	$\delta_{ii'}$ (km)	$Q_{ii'0}$ (trips/hr)	$e_{ii'0}$ (\$/trip)
Minimum	0.71	0	$20.71
Maximum	32.20	76	$52.20
Average	8.89	11.75	$28.89
Standard deviation	5.68	9.02	$5.68

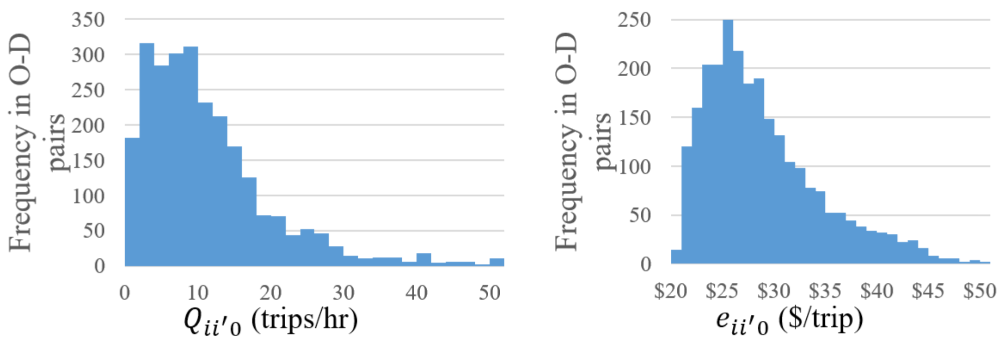

Figure 10. Histograms of the demand parameters used for “year zero”.

Figure 11 shows that while 80% of the travel demand is concentrated in O-D pairs closer than 15 km, the demand of individual O-Ds tends to increase with distance. This is consistent with documented travel patterns of this metropolitan area, with many users commuting between peripheral areas through the center of San Jose.

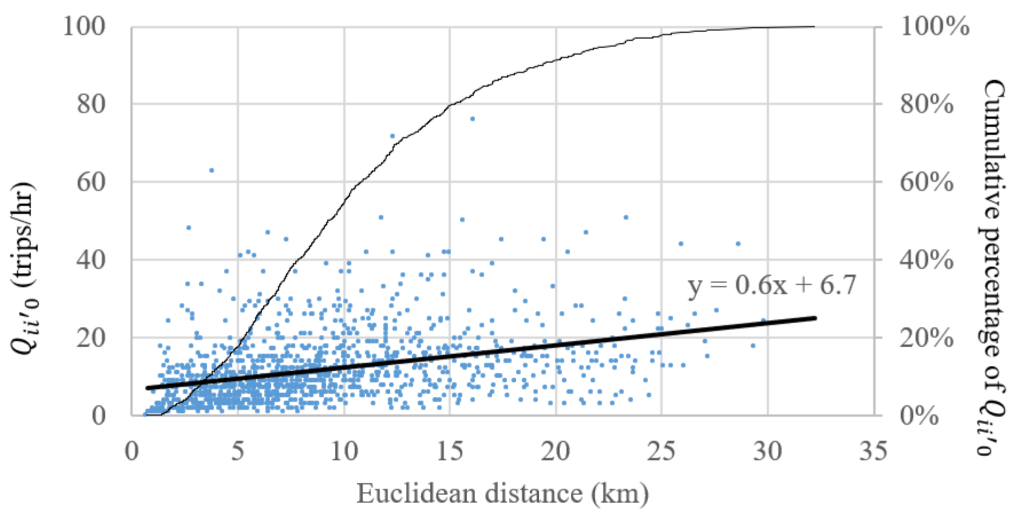

Figure 11. Relation between Euclidean distance and estimated maximum demand.

4.2.4. Other input parameters

The values used for all other input parameters are the ones that were provided in Table 2 in section 2.2. The unit construction, maintenance and operating costs employed are optimistically lower than values found in the literature, as a way to offset the diminished revenue from assuming comparatively lower fares ($2 per one-way trip, considering the zone's economic development). The 6,205 operating hours per year come from assuming 365 operating days per year and 17 operating hours per day.

4.3.Results of the evolutionary optimization

The optimization algorithm was coded in Python 3.12.7 and executed in the Zaratan High Performance Computer Cluster of the University of Maryland in 1000 CPU cores across 8 different nodes, each core with 2 GB of RAM. A population of 5000 chromosomes was evolved across 100 generations, with all the CPUs decoding the chromosomes of each generation in parallel before proceeding to the next generation. Therefore, during each generation each core sequentially decoded on average five chromosomes. Crossover and mutation probabilities were set as 0.9 and 0.1 respectively, and the maximum number of optimizations of all headways per simulated year was set as 10.

The algorithm ran in 57 minutes and 9 seconds. Figure 12 shows the fitness percentiles in ranges of ten, for the chromosomes of all generations. The fitness of the best solution of the first generation was improved by only 5.5% by the end of the algorithm, which hints that the population size was adequately large to allow for a lot of variation in the first generation. The chromosomes became increasingly similar in fitness to each other, to the point of near homogeneity (with few low-fitness outliers) after the first third of the generations.

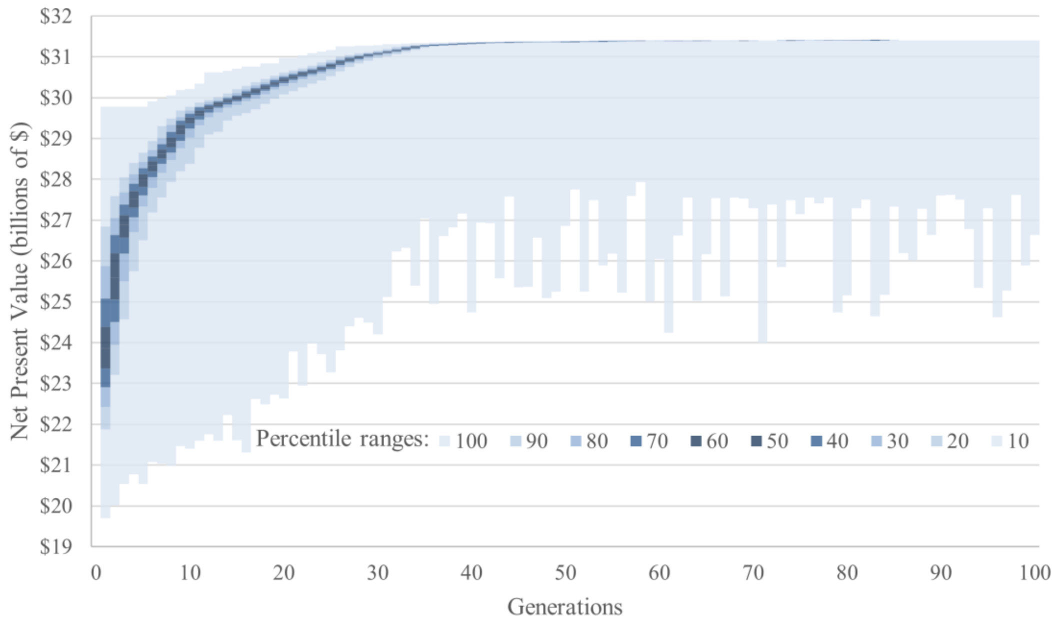

Figure 12. Percentiles of objective function of the chromosomes of each generation.

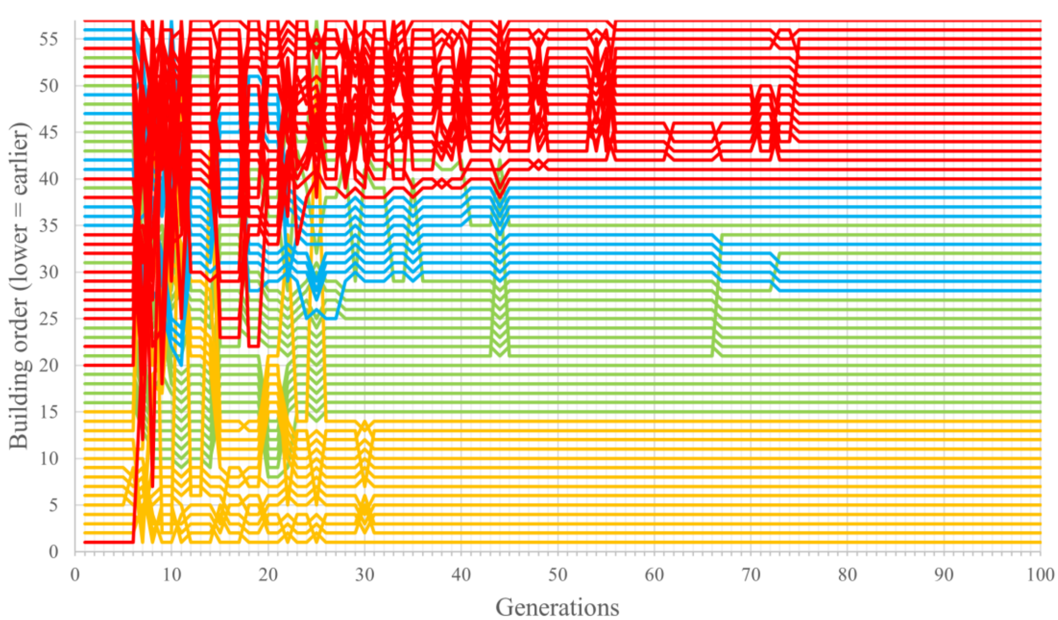

Figure 13. Link building priorities for the best solution of each generation.

Figure 13 shows a line for every link of the network colored according to the transit line it belongs to, depicting the evolution of its building priority across generations, with lower meaning built earlier. Remarkably, in the first generations the construction of the red line is intermingled among the other two lines, but eventually settles at the very end of the building sequence in later generations. This is attributed to this line potentially having the lowest demand between its own O-D pairs, or the other three lines providing already fast enough connections between peripheral towns.

4.4.Network development for the optimized solution

This subsection discusses the chronological development of the network across the timeframe and its state in the final year, associated with the best chromosome of the last generation, i.e., the optimized solution. The optimal link-building sequence is:

$$21 - 20 - 19 - 22 - 23 - 24 - 25 - 26 - 27 - 28 - 29 - 30 - 18 - 17 - 8 - 7$$
$$-6 - 5 - 4 - 3 - 9 - 10 - 11 - 12 - 13 - 14 - 15 - 34 - 33 - 32 - 31 - 16$$
$$-35 - 2 - 1 - 36 - 37 - 38 - 39 - 57 - 56 - 46 - 45 - 48 - 49 - 43 - 42$$
$$-41 - 40 - 52 - 53 - 54 - 44 - 55 - 47 - 50 - 51$$

Figure 14 shows a schematic of the network rotated 90° clockwise for larger visualization. Each link is shown with a number representing its building order, and in a color ranging from red (built first) to blue (built last). The demand of each node for the last year (year 50) is shown as circles whose area represents the sum of the hourly inbound and outbound trips, both potential (light gray) and satisfied (dark grey).

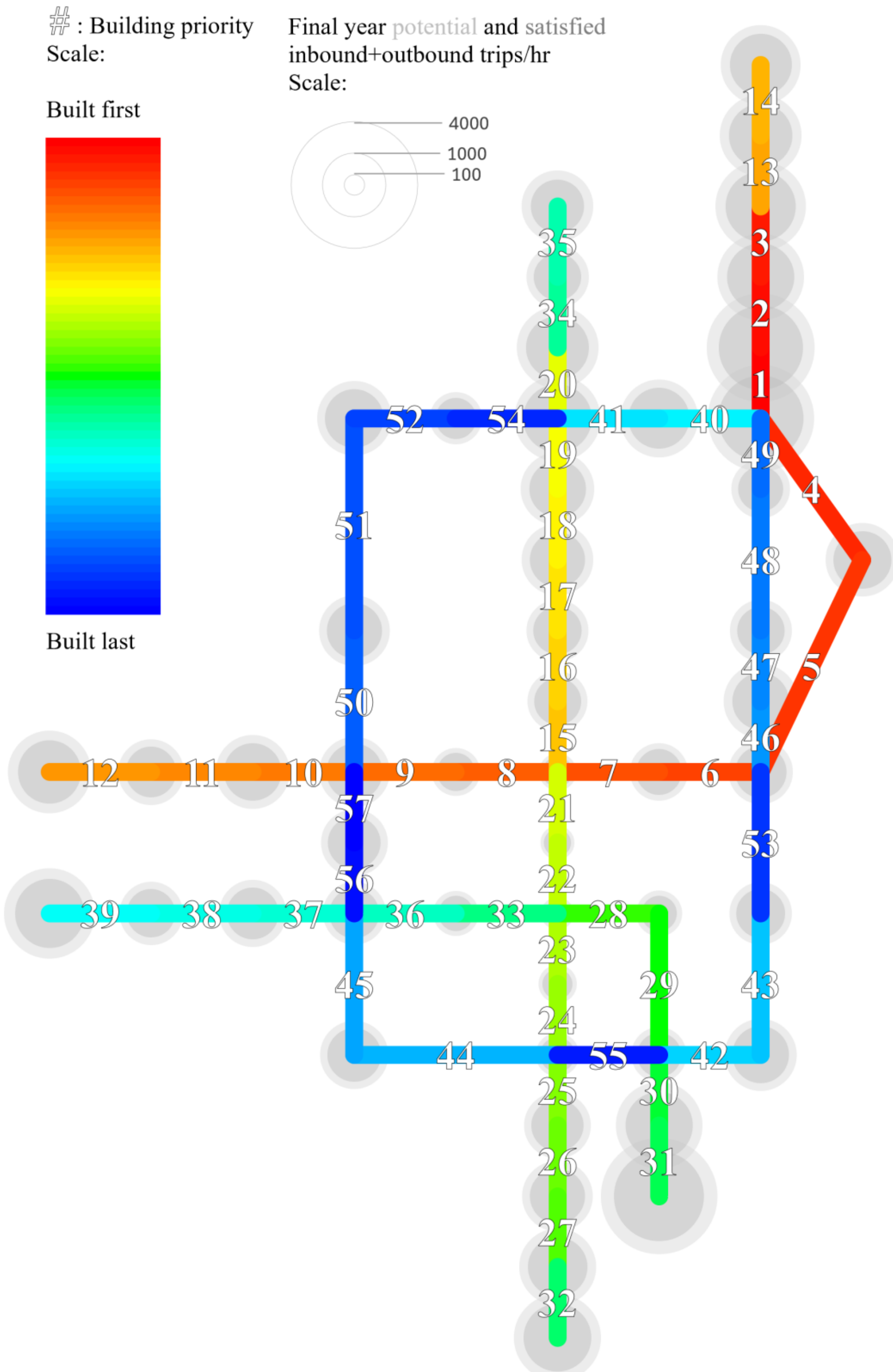

Figure 14. Link building order and satisfied node demand.

As expected, the optimization algorithm prioritized providing rail connectivity for nodes with higher overall demand importance, since the inputs provided established flows as strictly dependent on that factor alone. In other words, the method used for generating demand parameters ruled out atypically large potential demands between nodes whose overall demand importance is otherwise not that high. However, that situation could happen if the model is fed with surveyed O-D-specific demand parameters, causing the optimization to behave differently.

Table 6 shows the evolution of the economic variables caused by building the links in the optimized order. The links built each year are in a font color according to their rail line. All monetary amounts are in (rounded) millions of dollars, except the fare which is in dollars. As a reminder, B_{Ut} and B_{St} are not discounted to present value but J_t is, and the meanings of the column headers are the following.

t: year	l: links built	f_t: fare	B_{Ut}: user surplus
V_t: revenue	C_t: construction cost	O_t: operating cost	M_t: mainten. cost
B_{St}: supplier profit	F_t: external funding	D_{Et}: final balance	J_t: NPV

Table 6. Economic variables for each year for the optimized solution.

t	l	f_t	B_{Ut}	V_t	C_t	O_t	M_t	B_{St}	F_t	D_{Et}	J_t
1		2.05	592	172	-	30	-	142	50	192	720
2	21	2.10	592	176	146	30	-	1	51	243	570
3	20,19	2.15	623	186	257	51	3	-126	51	168	469
4	22,23	2.20	691	201	219	102	10	-130	52	90	519
5	24,25	2.25	795	219	141	113	15	-50	52	92	675
6	26,27	2.30	855	231	119	126	18	-32	53	112	732
7	28,29	2.35	880	240	147	139	21	-68	53	98	708
8	30	2.40	899	248	77	152	25	-6	54	146	764
9	18	2.45	904	254	97	153	27	-23	54	177	739
10	17,8,7	2.50	922	261	228	157	29	-153	55	78	631
11	6,5	2.55	956	271	133	183	34	-80	55	54	706

t	l	f_t	B_{Ut}	V_t	C_t	O_t	M_t	B_{St}	F_t	D_{Et}	J_t
12	4	2.60	984	281	66	215	38	-37	56	72	748
13		2.65	1000	289	-	212	39	38	56	167	804
14	3,9	2.70	1000	294	188	212	39	-144	57	80	650
15	10,11,12	2.75	1023	304	135	248	44	-122	57	15	671
16	13	2.80	1062	316	69	260	47	-59	58	14	733
17	14	2.85	1080	324	66	266	48	-56	59	16	733
18		2.90	1093	332	-	274	50	8	59	83	773
19	15,34	2.95	1093	338	135	274	50	-121	60	22	670
20	33	3.00	1102	345	59	287	53	-55	60	28	708
21	32	3.05	1122	354	87	293	55	-81	61	8	690
22		3.10	1150	364	-	317	57	-9	62	60	741
23	31	3.15	1150	370	94	317	57	-97	62	25	671
24		3.20	1174	380	-	322	59	-2	63	87	732
25	16	3.25	1174	386	103	322	59	-98	63	52	659
26	35	3.30	1186	394	53	333	62	-54	64	62	680
27	2	3.35	1182	399	110	340	63	-114	65	13	629
28		3.40	1194	407	-	339	65	3	65	81	691
29	1,36	3.45	1194	413	140	339	65	-131	66	16	602
30	37	3.50	1209	422	70	345	69	-62	67	21	637
31	38	3.55	1222	431	61	352	70	-53	67	35	637
32	39	3.60	1230	438	68	356	72	-57	68	46	626
33		3.65	1242	447	-	359	74	15	69	129	658
34	57	3.70	1242	453	121	359	74	-100	69	98	586
35	56,46	3.75	1246	460	159	387	76	-163	70	6	545
36		3.80	1266	471	-	400	80	-9	71	68	621
37	45	3.85	1271	478	75	400	80	-77	72	63	578
38	48,49	3.90	1271	484	114	400	82	-111	72	24	551
39		3.95	1271	490	-	400	85	6	73	103	594
40	43	4.00	1271	497	107	400	85	-94	74	82	537
41	42	4.05	1271	503	79	400	87	-63	74	93	541
42	41,40	4.10	1271	509	131	400	89	-111	75	58	509
43	52	4.15	1271	515	80	400	92	-57	76	77	523
44	53	4.20	1271	522	84	400	94	-56	77	98	513
45	54,44	4.25	1271	528	125	400	96	-94	77	81	487
46	55,47	4.30	1271	534	93	400	99	-58	78	101	492
47	50,51	4.35	1288	544	106	436	102	-99	79	81	473
48		4.40	1327	560	-	497	104	-41	80	120	502
49		4.45	1327	566	-	497	104	-35	81	166	494
50		4.50	1327	573	-	497	104	-28	81	219	487

4.4.1. Societal benefits and costs

The algorithm is expected to generally prioritize building rail connections in links that provide a higher absolute increase of the undiscounted net societal benefit Y (user net benefit minus supplier net costs). However, the network topology imposes connectivity constraints on the independent selection of links to abide by this criterion. Apart from that, the observed tendency of the optimization algorithm to cluster the construction of links into complete lines also responds to the fact that in general, starting to operate a segment of a completely new line involves a steeper increase in operation costs than extending an existing one over a segment of comparable length and demand. This is because the former option involves introducing a new fleet for the new line, which will typically require more vehicles than the additional ones required for extending a line. This effect is accentuated the longer the existing line is.

Nonetheless, if we visualize four sequential phases in the development of the rail network for the optimized solution, the trend to favor higher absolute increases of Y does manifest in that they occur in ascending order of Y-increase. The phases are:

1. Years 1 to 10: line 2 (yellow) is the developed to near completion.
2. Years 11 to 18: line 1 (green) is developed to near completion, and line 2 gets a final extension of one link.
3. Years 19 to 33: line 3 (blue) is developed to completion, and line 1 gets a final extension of four links.
4. Years 34 to 50: line 4 (red) is developed to completion.

Figure 15 shows the four phases in a graph that plots the undiscounted Y_t for every year, where each connection between years t and $t + 1$ is colored according to the rail link or links that start operating in year $t + 1$. Semi-transparency in the colors helps visualize when two or more links start operating in the same year.

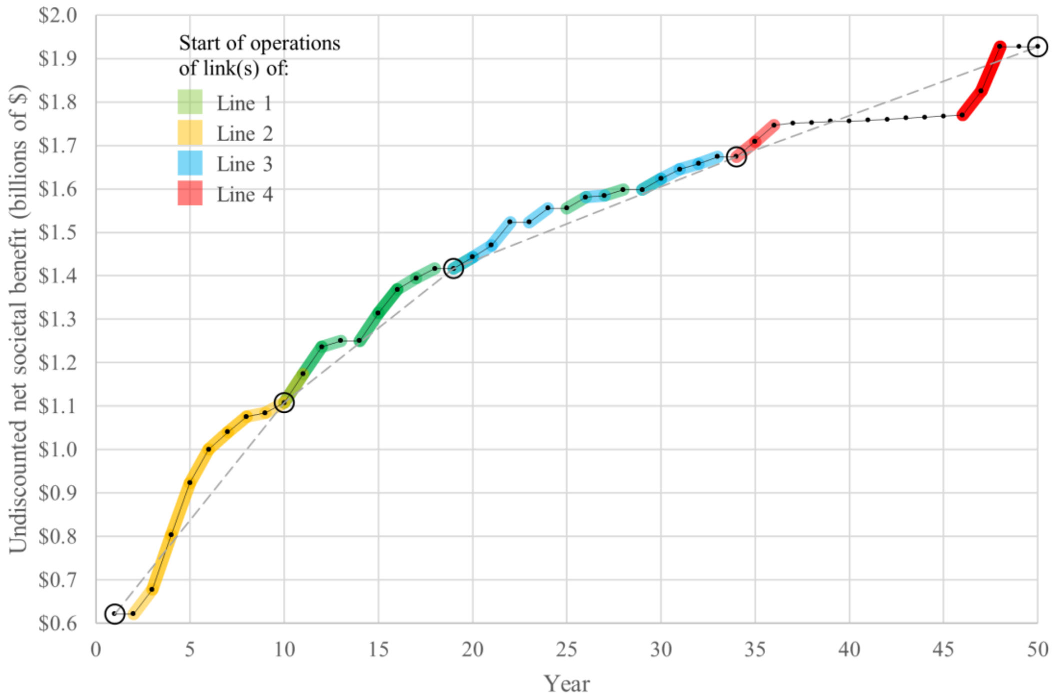

Figure 15. Net societal benefit for every year.

Remarkably, 12 links from the red line built between years 35 and 45 did not start operating until year 45, after being finally connected to the line's two working links in year 44. Figure 4 in section 2.6 illustrates the reason why even with contiguous infrastructure development, the model allowed this 12-link rail segment to be built isolated from its line's operating path. The preference of the algorithm to do that instead of contiguously extending the existing red line might signal that demand was not sufficiently high in year 35 to justify operating costs. Alternatively, it could be due to most of the red line not being cost-efficient unless its full circular operating path is completely developed. Finally, travel time overestimations can't be ruled out as a possible cause, because of the limitation in the chromosome decoding algorithm mentioned in section 0 that imposes unnecessary alighting and waiting time in the node where a circular line starts and ends, for users who just want to traverse it.

An alternative outlook for the marginal cost-effectiveness of the start of operations of links is obtained if instead of subtracting gross user benefit minus total societal cost (supplier cost plus travel time cost) for each year, the ratio of the two quantities is examined. In Figure 16, each point represents the gross user benefit and the societal “continued” cost (maintenance + operation + cost of time) of a year. The greater the slope of the line that connects two years, the more marginally cost-efficient are the starts of operations that occurred in the second year. Diagonal lines have a $1:$1 slope.

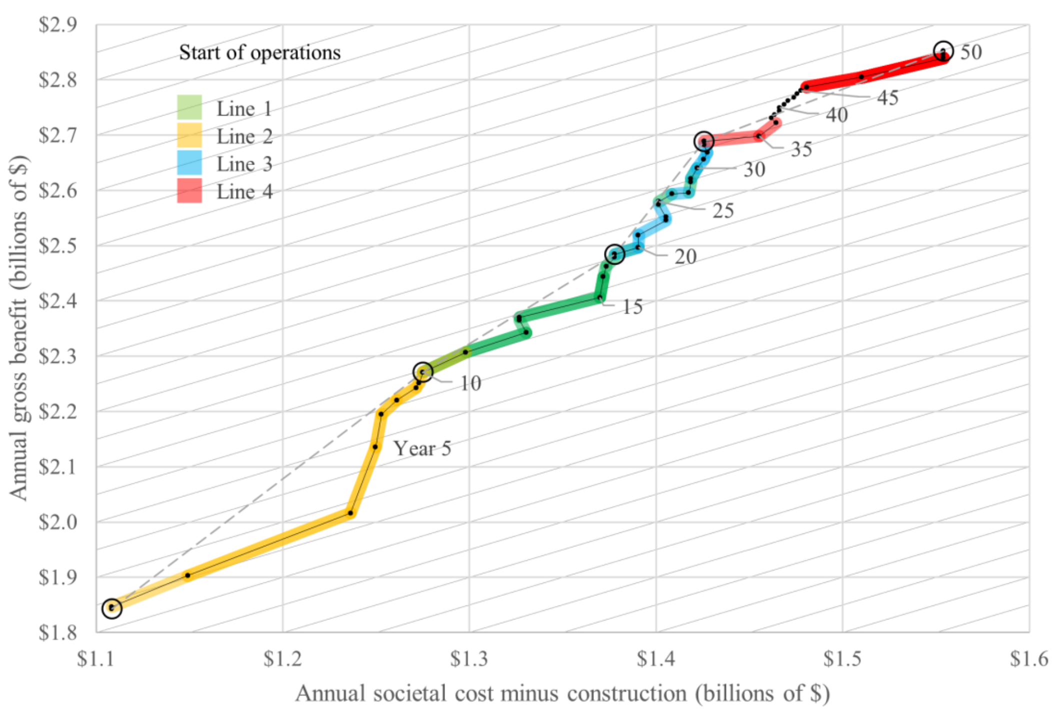

Figure 16. Annual gross benefit vs. societal “continued” cost for every year.

When employing this benefit-cost ratio, the four development phases of the network mentioned earlier occur in decreasing order of cost-effectiveness, except for the third phase which primarily builds the blue rail line. Despite this line providing the best benefit-cost ratio, it was not built first because the construction of the yellow and green lines grants a higher net societal benefit in absolute terms.

4.4.2. Cost structure

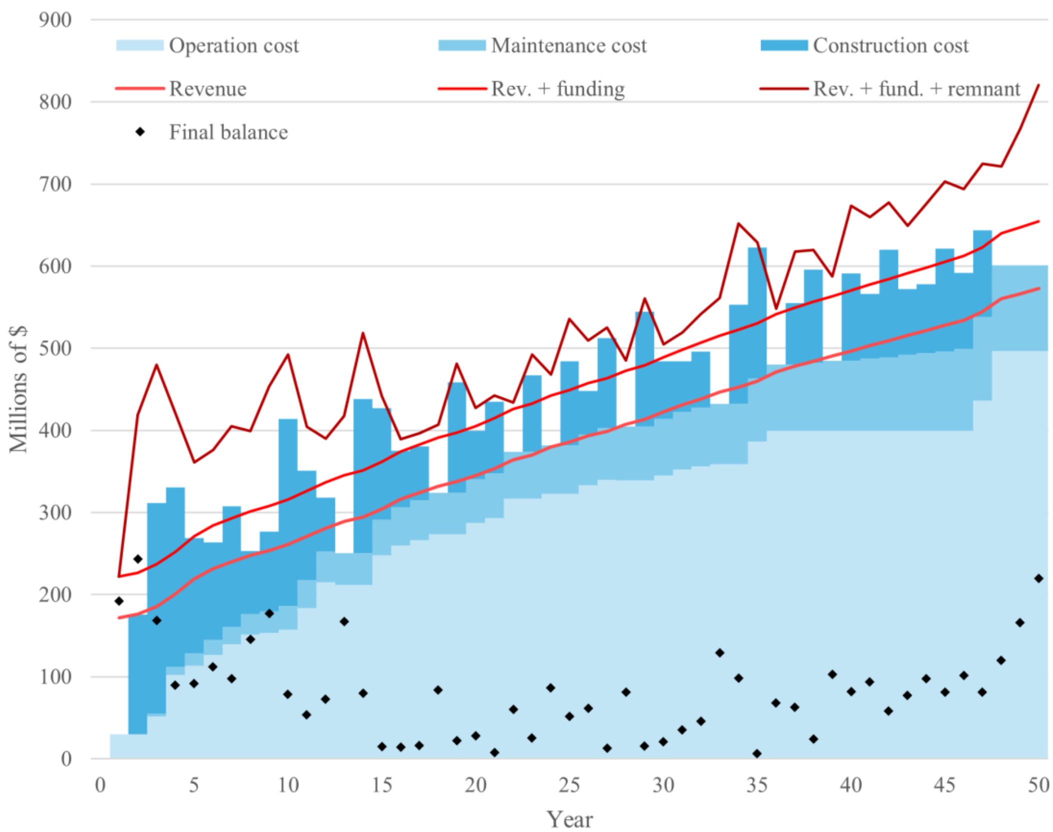

Figure 17. Monetary inflows and outflows for every year.

Figure 17 compares the monetary inflows and outflows for each year. For cash flow computations, the total inflow of a year includes not only the revenue collected and the external funding received that year, but also the remaining balance at the end of the previous year. The total outflow includes the cost of operation, maintenance and construction for that year. The difference between the two is the year's final balance. Congruent with the unit costs and the size of the network, operation costs quickly start dominating over construction and maintenance costs after the fifth year.

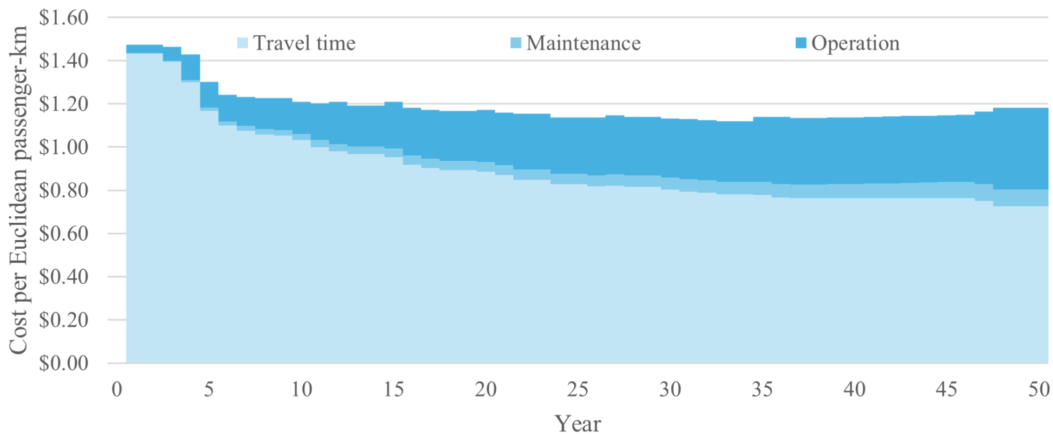

Figure 18. Cost decomposition per Euclidian passenger-km for every year.

Figure 18 presents the evolution across the years of the societal cost composition of every kilometer traveled by the users of the system. To provide a better measure of overall urban mobility costs, the Euclidean distance between the origin and destination of each trip was used instead of its in-network distance, a decision that factors in the goodness of the alignment proposed. As expected, average distance-adjusted travel times consistently decline as the rail network is built, while maintenance and operation costs increase (the latter increases marginally for a few isolated years), and all these three costs remain steady after the rail network is finalized. However, the aggregated cost starts to increase after year 33 when line 4 (red) starts to be built, which might be an argument against the economic pertinence of rail-enabling that line at all.

4.4.3. Ridership and travel times

Figure 19 displays the hourly trips that use each of the four rail lines and their corresponding bus lines. Bus ridership values are combined whenever the two bus lines of the same triad are simultaneously operating. As expected, the ridership of a line often fluctuates as a response to the extensions of the operation of other lines, marked by colored dots. This proves the capabilities of the model to dynamically reroute passengers among their different travel path alternatives, seeking the shortest one with Dijkstra's algorithm.

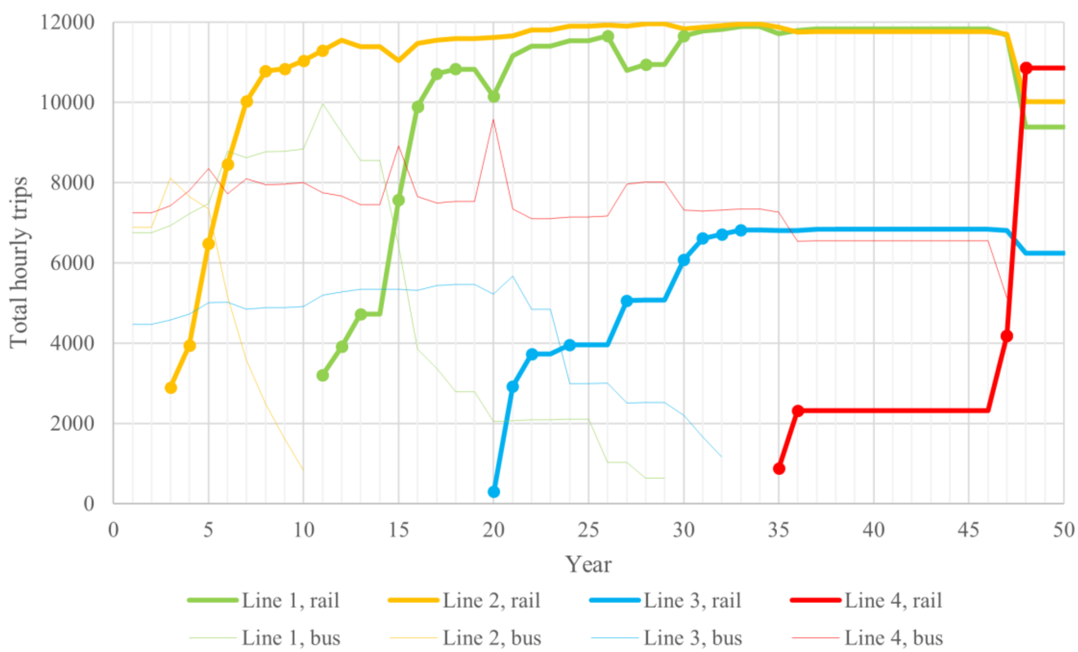

Figure 19. Total hourly line ridership for every year.

Figure 20 confirms the expected downward evolution (49% reduction) of the average travel time per Euclidean km of travel in the network as it is built over the years. It also evidences the almost strictly increasing behavior of demand satisfaction, as the percentage of total realized trips with respect to total potential trips of all O-D pairs for each year. Demand satisfaction is illustrated both in unadjusted terms as well as adjusted for O-D Euclidean distance (realized passenger-km divided by potential passenger-km).

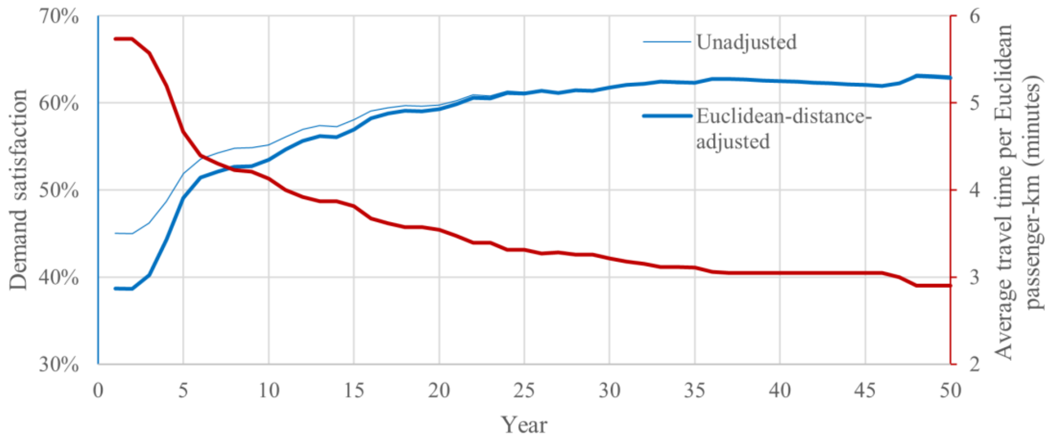

Figure 20. Network-wide demand satisfaction and average travel time for every year.

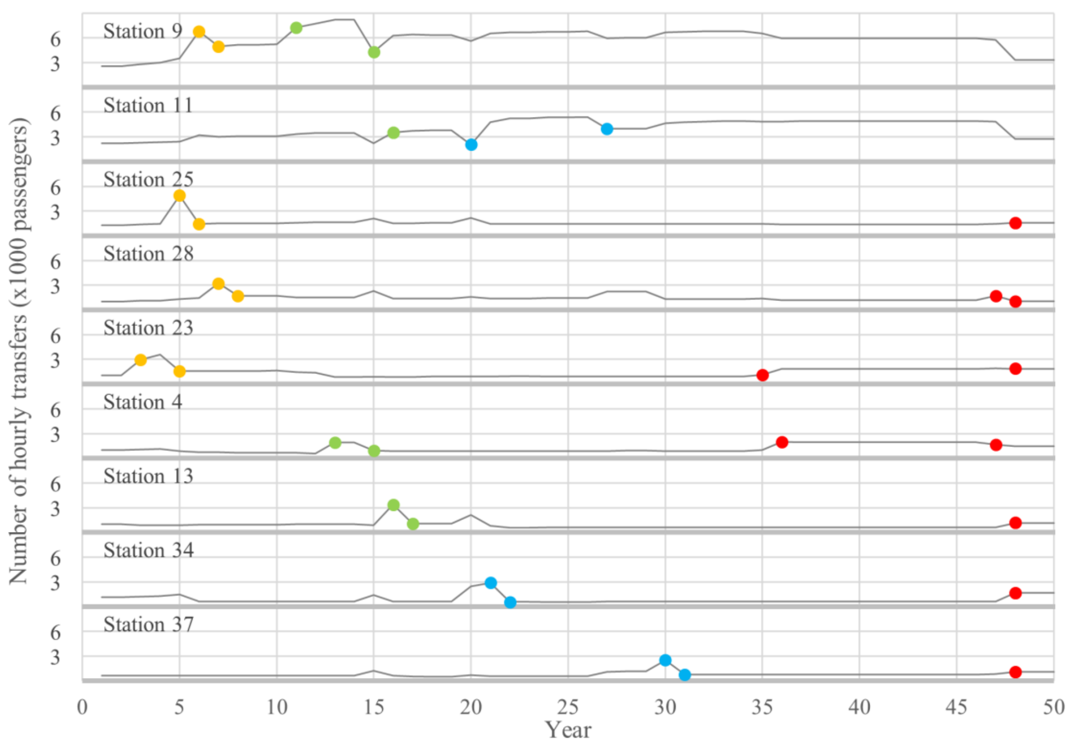

Figure 21. Hourly transfers at transfer stations for every year.

Figure 21 shows how the number of hourly transfers at each of the nine transfer nodes remains fairly steady throughout the years, with notable perturbances whenever the construction of any of the four rail links that connect each node occurs (marked with colored dots in the chart). In particular, the construction of the first of the two links of a line always increases transfers at the node because new bus-rail transfers are created.

Finally, Figure 22 shows the hourly link-wise ridership of each rail line throughout the timeframe. Neither of the lines exhibit a clear bell-shaped pattern, but instead the demand is distributed very unequally across the length of the lines.

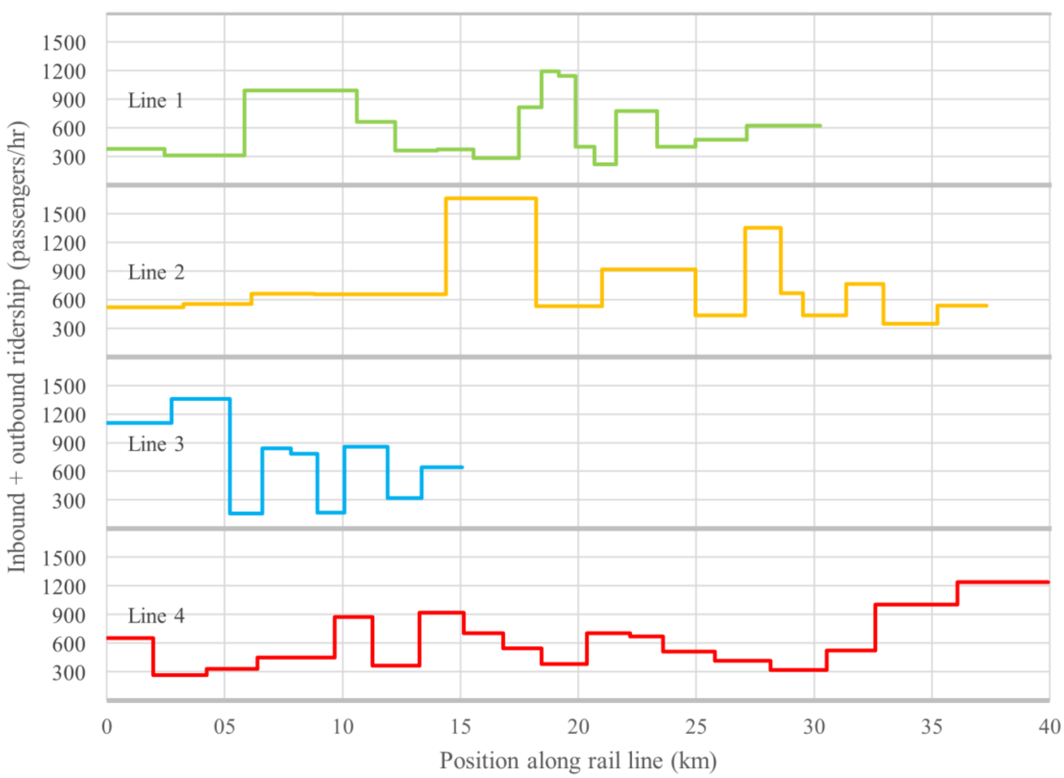

Figure 22. Line hourly load profiles for the last year of the timeframe.

4.5.<u>*Comparison between predefined and dynamic user travel paths*</u>

A major improvement in the chromosome decoding algorithm employed in this work compared to previous models is the addition of the possibility for users to reoptimize their travel paths every year in response to changes in the services available. The alternative is strict adherence to link-wise travel paths for each origin-destination pair throughout the entire timeframe, which could greatly simplify the computation of travel times by simply summing the individual travel and dwell times of all links of the preset path, each according to its built state, and then adding the necessary transfer times. The latter approach is commonplace in literature when the link- (or station-) building sequence optimization problem is solved for an individual transit line or a network without redundancy. This subsection provides evidence of the impact on the quality of the link-building sequences produced when the optimization is performed allowing users to reoptimize their paths every year *("dynamic pathing")* as opposed to not allowing them to *("predefined pathing")*, in a network with redundancy like the one studied.

To examine this potential effect, 10 separate genetic optimization processes were run with predefined pathing (control group) and 10 were run with dynamic pathing. Each optimization process used the same network inputs from section 4.2 and the same genetic algorithm described in section 4.5 with the same parameters (including using 100 generations), except that a population of size 500 instead of 5000 was used.

For predefined pathing, the decoding algorithm does not use Dijkstra's shortest path optimization in step 1 of the computation of travel times every year (section 3.1.3). Instead, a "completed state network" is temporarily created at the beginning of each decoding process, with all rail lines operating to their full extent and no bus lines operating. Dijkstra's algorithm is performed only once in this completed network to obtain a set of predefined travel path binaries $z_{lrii'C}$, with the C subscript standing for "Completed" and replacing the t subscript. Then, when modeling the evolution of the network in each year t, the actual path binaries $z_{lrii't}$ are obtained with this simple logic for each O-D pair ii' and link l:

- If $z_{lrii'C} = 0\ \forall r$ it means that the predefined path for ii' does not use link l, so $z_{lrii't}$ is set as zero for all transit lines r.
- Otherwise, the predefined path uses one (and only one) rail line r' on link l. Therefore, $z_{lr'ii't}$ is set as one for the line of the triad $(r', r' + n_{RR}, r' + 2n_{RR})$ that operates on that link, and zero for all other transit lines.

The first comparison between predefined and dynamic pathing is in terms of the "Optimized NPV", the present value of the total societal net benefit associated with the best link-building sequence of the last generation. Data series 1 in Table 7 shows this number for each of the 10 separate optimization processes that were conducted with predefined pathing during optimization. Specifically, these values in Data series 1 are the result of evaluating those 10 independently optimized sequences by employing *predefined pathing during the evaluation as well*. On the other hand, data series 2 shows the result of evaluating those same 10 sequences that were obtained with predefined pathing during optimization, *but using dynamic pathing during this post-optimization evaluation*. Since for now we are evaluating the same exact set of ten link-building sequences under two different analysis types, the difference in the average optimized NPV between the two modes indicates that, in general, analyzing a generic link-building sequence with dynamic pathing yields a higher (by 5.6% in the studied network) NPV compared with evaluating the same sequence with predefined pathing. The gap between the points in the two data series (Figure 23) also solidifies this assertion. This is a fairly obvious conclusion because travel times in a given year cannot improve (at best will be equal, but will likely worsen) if users are prohibited from choosing a better path than the one that was predetermined for their O-D pair from the start, compared to allowing for that possibility. The increased generalized cost and its associated decreased ridership cause a reduction in the year's user net benefit which immediately worsens the NPV, but the decreased revenue also delays the construction of new links and further lowers the NPV.

Table 7. NPV of independent optimization executions conducted with dynamic and predefined pathing.

Data series		1	2	3
Symbol in chart		●	▲	▲
Pathing during optimization		Predefined	Predefined	Dynamic
Pathing during evaluation		Predefined	Dynamic	Dynamic
Optimized NPV (billions of $) of optimization process #	1	28.70	30.58	31.36
	2	28.70	30.13	31.39
	3	28.81	30.20	31.21
	4	28.92	30.73	31.34
	5	28.72	30.53	31.27
	6	28.75	30.31	31.05
	7	28.83	30.31	31.27
	8	28.93	30.71	31.32
	9	28.78	30.25	31.32
	10	28.52	29.89	31.19
Average NPV		**28.77**	**30.36**	**31.27**
Percent difference		-	5.6%	3.0%

Although users might not always know how to minimize their travel time, or might involve other factors in their path decisions, it can be argued that evaluating a link-building sequence with dynamic pathing achieves a more realistic modelling of riderships, independently from how a sequence was obtained (through an optimization process, or even arbitrarily defined). Therefore, from this point on dynamic pathing will be used for evaluating all optimized sequences, even the ones obtained with predefined pathing (i.e., series 2), which also guarantees a fair comparison.

Data series 3 in the table suggests that even if using the same pathing mode when evaluating post-optimization, sequences optimized with dynamic pathing are intrinsically better than those optimized with predefined pathing (data series 2). This apparent trend is highlighted by the fact that all 10 points from data series 3 are clustered at higher NPV values than those of data series 2 for the experiment performed, as shown in Figure 23. Despite the improvement being only 3.0% for the studied network, it is notable that a difference of $907 million in net present value over a timeframe of 50 years is not negligible.

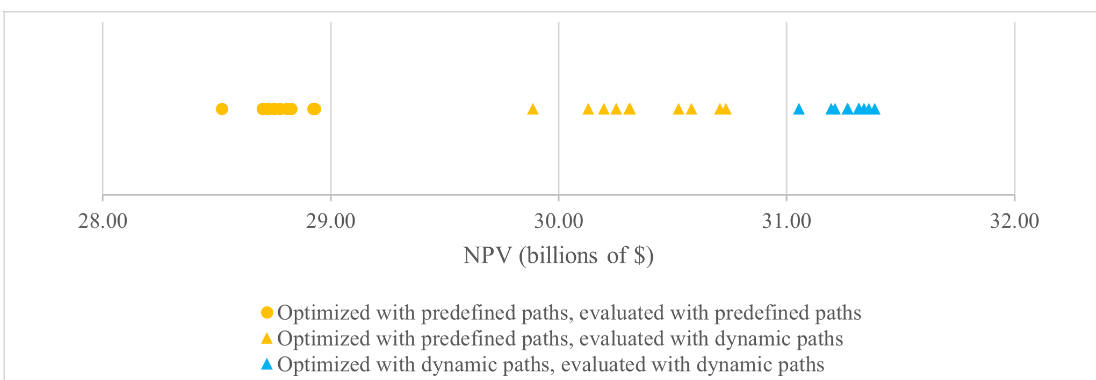

Figure 23. NPV of independent optimization executions conducted with dynamic and predefined pathing.

For the next comparison, each link was assigned a priority number for each of the 20 optimization processes, according to the order in which links are built in the optimized sequence of the optimization process, with one signifying the highest priority. Then, an *average priority* was calculated for each link for the predefined and the dynamic pathing case, by averaging the 10 respective priority numbers for each case. Finally, the average priority of each link for the dynamic case was compared with that of the predefined case of that same link. In Figure 24, positive numbers and red colors mean that the priority of a link increased when dynamic pathing was used compared to predefined, and negative numbers and blue colors mean its priority decreased. It should be emphasized that the numbers do not represent average absolute priorities, but only the magnitude of the priority difference between the two cases for a link.

Table 8. NPV of independent optimization executions conducted with dynamic and predefined pathing.

Line	Link-wise average for the average priority, for optimization with		Change
	Predefined pathing	Dynamic pathing	
1	29	21	+8
2	10	9	+1
3	48	35	+13
4	34	48	-14

Table 8 uses the same sign convention, with positive meaning higher priority in dynamic pathing. It shows how, in general, rail line 2 is built first in both pathing types, while line 1 is built before line 3 and both of them after line 2. The difference, however, is that with predefined pathing line 4 is generally prioritized before line 3.

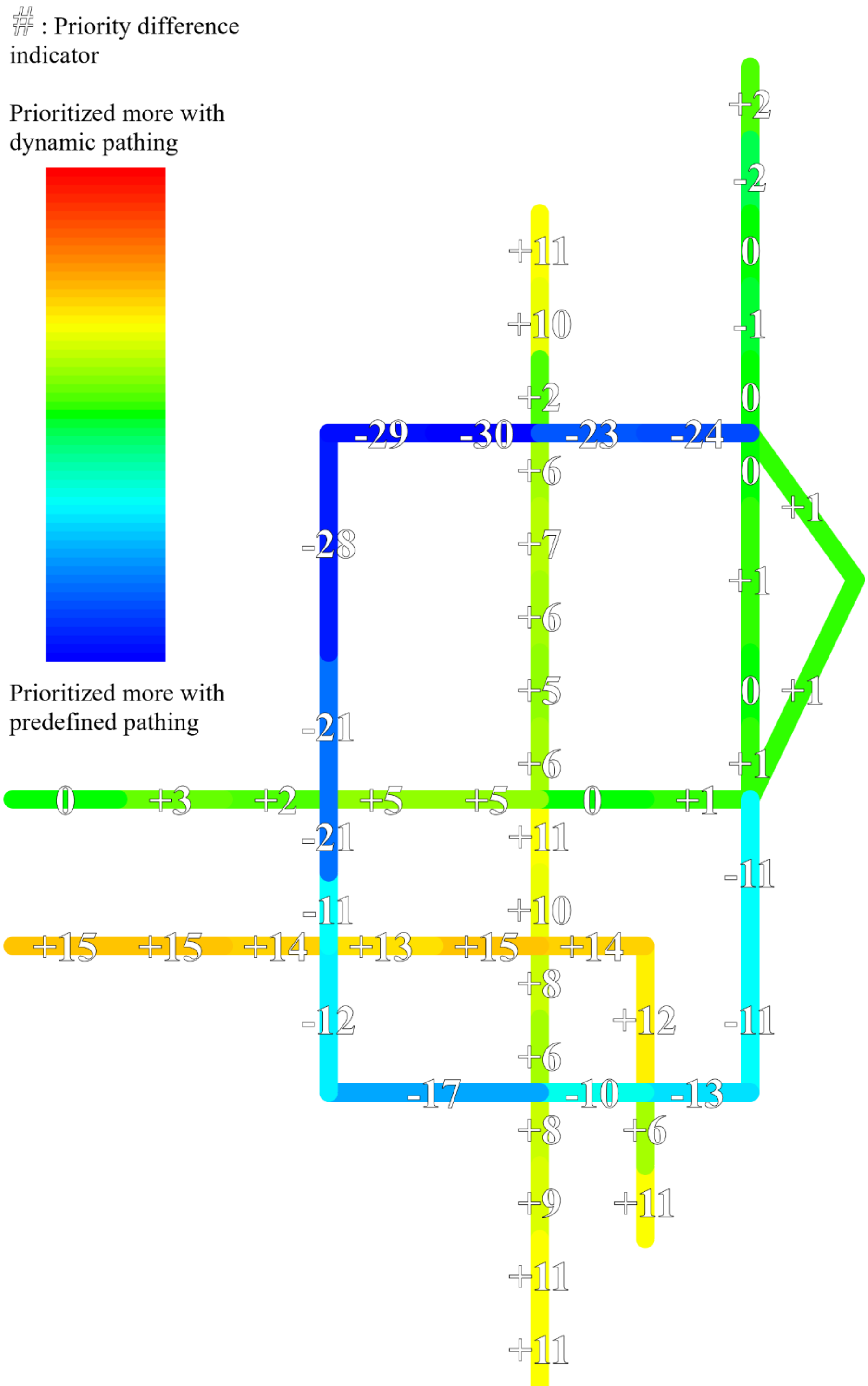

Figure 24. Differences in link priorities between dynamic and predefined pathing.

The discussion of Figure 15, Figure 16 and Figure 17 in section 4.4 documented the overall less cost-effectiveness of building line 4, which is a circular line, or at least completing it before the three radial lines. The question arises then of why predefined pathing during optimization fails to deprioritize the circular line. Drawing potential reasons by delving deeper into relations between the inputs (particularly demand concentration, distances involved, transfer times, or the mere topology of the network) might be the object of a future study. However, the scope of this thesis is limited to quantify, for the studied network, the enhanced optimization quality observed when using dynamic pathing.

Chapter 5: Conclusions

An algorithm has been developed for optimizing the phased replacement of bus by rail service in a multi-line network in order to maximize the present value of societal net benefit over a time horizon, considering demand elasticity to fares and travel times, and accurately modeling (shortest) travel times of the users. Continuity of infrastructure construction and of operating paths are enforced, and a yearly supplier budget affected by internal revenue, external funding, and operation and maintenance expenses constrain the network expansion.

Distinctive elements included in the model are:

1. A heuristic to guarantee rail line continuity, which may otherwise not be guaranteed even if rail infrastructure is continuous.
2. An original subroutine to guarantee the finding of an optimal headway, by iteratively removing from the objective function the O-D pairs for which the optimized headway causes zero flow, since they would otherwise preclude finding real solutions if not discarded.
3. Dynamic pathing which enables users to use the shortest path at every developmental state of the network.

Optimal link-building sequence and year-by-year headways were obtained for a proposed rail network in San Jose, Costa Rica by optimizing link building priorities with a Genetic Algorithm coded with Python 3.12.7, with a run time of 57 minutes in 1000 parallel CPUs of a computer cluster. The optimized solution after 100 generations of 5000 chromosomes indicate that is generally beneficial to sequentially build each planned rail lines in a mostly uninterrupted way. However, the presence of line interweaving in the link-building sequence is evidence in favor of link-wise over line-wise sequencing, which was the approach used in some cited previous work.

Another finding was that when evaluated post-optimization in the same way (namely with dynamic pathing), the sequences produced by 10 independent optimization executions with dynamic user pathing were found to yield societal net benefits 3.0% higher than those yielded by sequences produced with predefined user pathing. From an academic standpoint for future research, this result facilitates evidence-backed weighing about the shortcomings of excluding this component, against the potential reduction in computation time. The deprioritization of the circular line against the radial ones in the studied case when evaluating with dynamic pathing (which is the accurate way to evaluate link-building sequences) is consistent with notions of shortest path in a region with relatively homogeneous demand distribution. However, further analysis might solidify conjectures about which network or demand inputs caused the circular line to not be completely deprioritized when evaluating with predefined pathing.

Other suggestions for future work include:

1. Modify the routines that create and modify the SAN, to add the *through* edges at the start and end node of circuit lines.
2. Model the reduced wait time for trips featuring legs for which the choice among two alternative parallel lines is indifferent.
3. Constrain headway optimization by vehicle capacity, including a peak hour factor.

4. Model the timeframe in continuous time and allow for different construction durations for the links. This would require predicting the time when the available balance will reach the amount needed for building the next link or group of links.
5. Encode construction-bunching for links into the chromosomes, and model economies of scale when building several links simultaneously.
6. Model an increase in the maximum theoretical demand of nodes as a response to rail station construction, to reflect a potential urban densification near the station after it is built.
7. Factor in externalities as additional revenue, like increased tax revenue from land appreciation (might require a probabilistic approach if for budgeting).
8. Test the model with directly surveyed O-D maximum demand parameters.
9. Model train turnaround time.
10. Model dwell time as demand dependent.
11. Model and maintenance costs as ridership dependent.
12. Separate rolling stock acquisition costs from operation costs.
13. Establish the location of the train depots, which would constrain the initial expansion points of the rail network.
14. Model and optimize the number, alignment and headways of the feeder bus routes, and merge their cost structures with the costs of the "trunk" lines (the ones addressed in the current model), to conform an integrated urban transit system.
15. Add additional layers of optimization for fares and external funding.

Appendices

Table 9. Geographic and demand characteristics of the nodes of the modelled network.

i	Station name	GPS coords. Lon. (°)	GPS coords. Lat. (°)	CRTM05 coords. Lon. (m)	CRTM05 coords. Lat. (m)	ω_i	κ_i
1	Santa Ana	-84.1887	9.9336	479307	1098402	0.019	0.020
2	Lindora	-84.1947	9.9546	478655	1100730	0.013	0.017
3	Belén	-84.1868	9.9783	479517	1103349	0.018	0.034
4	Metrópolis	-84.1521	9.9544	483329	1100709	0.019	0.038
5	Pavas	-84.1385	9.9484	484810	1100042	0.015	0.027
6	Loreto	-84.1236	9.9418	486444	1099311	0.014	0.026
7	Estadio	-84.1105	9.9370	487880	1098780	0.006	0.023
8	Paseo Colón	-84.0931	9.9353	489789	1098590	0.008	0.020
9	Hospital	-84.0845	9.9345	490736	1098504	0.003	0.007
10	Avenida Central	-84.0778	9.9339	491470	1098435	0.002	0.006
11	Asamblea	-84.0714	9.9333	492173	1098364	0.002	0.006
12	Escalante	-84.0640	9.9327	492978	1098303	0.003	0.008
13	San Pedro	-84.0558	9.9329	493879	1098326	0.006	0.008
14	Freses	-84.0414	9.9273	495465	1097708	0.017	0.017
15	Curridabat	-84.0348	9.9157	496184	1096416	0.022	0.016
16	Pinares	-84.0164	9.9087	498207	1095648	0.025	0.020
17	Tres Ríos	-83.9883	9.9104	501284	1095839	0.040	0.020
18	Garita	-84.2691	9.9955	470495	1105256	0.017	0.033
19	Coyol	-84.2423	10.0077	473440	1106601	0.025	0.027
20	Alajuela	-84.2171	10.0152	476199	1107430	0.040	0.024
21	Aeropuerto	-84.2023	9.9993	477820	1105676	0.034	0.025
22	Flores	-84.1523	10.0009	483300	1105849	0.047	0.045
23	Heredia	-84.1182	9.9948	487047	1105170	0.051	0.038
24	San Pablo	-84.0959	9.9856	489481	1104153	0.027	0.014
25	Tibás	-84.0787	9.9623	491367	1101575	0.016	0.016
26	Cinco Esquinas	-84.0808	9.9465	491144	1099826	0.013	0.013
27	Pacífico	-84.0845	9.9258	490740	1097534	0.006	0.012
28	Hatillo	-84.0966	9.9165	489412	1096514	0.018	0.021
29	Alajuelita	-84.0988	9.9029	489162	1095012	0.029	0.014
30	San Rafael Abajo	-84.0820	9.8927	491010	1093875	0.024	0.009
31	San Rafael Arriba	-84.0772	9.8754	491533	1091969	0.033	0.020
32	Ipís	-84.0101	9.9646	498895	1101832	0.064	0.044
33	Purral	-84.0333	9.9554	496354	1100815	0.032	0.026
34	Guadalupe	-84.0534	9.9467	494143	1099852	0.013	0.010
35	Calle Blancos	-84.0648	9.9422	492893	1099354	0.009	0.007

i	Station name	GPS coords.		CRTM05 coords.		ω_i	κ_i
		Lon. (°)	Lat. (°)	Lon. (m)	Lat. (m)		
36	Plaza Víquez	-84.0735	9.9247	491938	1097416	0.005	0.010
37	La Paz	-84.0693	9.9153	492405	1096376	0.013	0.018
38	Desamparados	-84.0673	9.8992	492615	1094596	0.021	0.014
39	Calle Fallas	-84.0652	9.8863	492854	1093173	0.019	0.011
40	San Miguel	-84.0620	9.8713	493200	1091512	0.040	0.025
41	Guararí	-84.1182	9.9821	487043	1103772	0.012	0.014
42	Valencia	-84.1132	9.9674	487593	1102141	0.009	0.020
43	Uruca	-84.1034	9.9539	488660	1100643	0.016	0.027
44	Llorente	-84.0640	9.9620	492985	1101543	0.017	0.013
45	Moravia	-84.0479	9.9611	494748	1101445	0.023	0.016
46	Zapote	-84.0538	9.9194	494098	1096830	0.015	0.021
47	San Sebastián	-84.0852	9.9109	490652	1095896	0.016	0.018
48	Tiribí	-84.1160	9.9211	487277	1097020	0.021	0.017
49	Escazú	-84.1374	9.9219	484933	1097115	0.020	0.022
50	Guachipelín	-84.1429	9.9400	484335	1099109	0.008	0.013
51	Ulloa	-84.1437	9.9785	484246	1103375	0.015	0.032

Table 10. Topological and geometric characteristics of the links of the modeled network.

r	j	Node 1	Node 2	d_{Rl} (m)	d_{Bl} (m)
1	1	1	2	2450	3675
1	2	2	3	3391	5087
1	3	3	4	4771	7157
1	4	4	5	1621	2432
1	5	5	6	1789	2684
1	6	6	7	1539	2309
1	7	7	8	1921	2882
1	8	8	9	951	1427
1	9	9	10	740	1110
1	10	10	11	705	1058
1	11	11	12	808	1212
1	12	12	13	906	1359
1	13	13	14	1745	2618
1	14	14	15	1638	2457
1	15	15	16	2162	3243
1	16	16	17	3113	4670
2	17	18	19	3241	4862
2	18	19	20	2882	4323

r	j	Node 1	Node 2	d_{Rl} (m)	d_{Bl} (m)
2	19	20	21	2678	4017
2	20	21	22	5590	8385
2	21	22	23	3828	5742
2	22	23	24	2789	4184
2	23	24	25	3952	5928
2	24	25	26	2108	3162
2	25	26	9	1527	2291
2	26	9	27	925	1388
2	27	27	28	1837	2756
2	28	28	29	1577	2366
2	29	29	30	2305	3458
2	30	30	31	2085	3128
3	31	32	33	2740	4110
3	32	33	34	2485	3728
3	33	34	35	1360	2040
3	34	35	11	1227	1841
3	35	11	36	1121	1682
3	36	36	37	1153	1730
3	37	37	38	1813	2720
3	38	38	39	1454	2181
3	39	39	40	1718	2577
4	40	23	41	1971	2957
4	41	41	42	2259	3389
4	42	42	43	2160	3240
4	43	43	25	3266	4899
4	44	25	44	1619	2429
4	45	44	45	1986	2979
4	46	45	34	1885	2828
4	47	34	13	1673	2510
4	48	13	46	1616	2424
4	49	46	37	1930	2895
4	50	37	47	1835	2753
4	51	47	28	1394	2091
4	52	28	48	2196	3294
4	53	48	49	2352	3528
4	54	49	50	2399	3599
4	55	50	4	2044	3066
4	56	4	51	3492	5238
4	57	51	23	3844	5766

References

Cheng, W.-C., & Schonfeld, P. (2015). A Method for Optimizing the Phased Development of Rail Transit Lines. *Urban Rail Transit*, 227-237.

Costa Rican Institute of Statistics and Census. (2016, March 29). *Costa Rica: Indicadores económicos según cantón y distrito.* Retrieved November 4, 2024, from INEC, Censo 2011: https://admin.inec.cr/sites/default/files/media/repoblaccenso2011-06.xls_3.xlsx

Google. (2024). Google Earth aerial imagery from the San Jose Metropolitan Area. Retrieved November 2024

Kuby, M., Xu, Z., & Xie, X. (2001). Railway network design with multiple project stages and time sequencing. *Journal of Geographical Systems, 3,* 25-47. doi:10.1007/PL00011465

Li, S., Liang, Q., Han, K., Wang, H., & Xu, J. (2022). A Double-Level Calculation Model for the Construction Schedule Planning of Urban Rail Transit Network. *Applied Sciences, 12*(10), 5268. doi:10.3390/app12105268

Luo, X., Guo, Y., & Wu, Y. (2011). The Timing of Urban Rail Transit Construction based on the weighted gray correlation. *Applied Mechanics and Materials, 97-98,* 1189-1194.

Montesinos-Valera, J., Aragonés-Beltrán, P., & Pastor-Ferrando, J.-P. (2017). Selection of maintenance, renewal and improvement projects in rail lines using the analytic network process. *Structure and Infrastructure Engineering, 13*(11), 1476-1496. doi:10.1080/15732479.2017.1294189

Peng, Y.-T., Li, Z.-C., & Schonfeld, P. (2019). Development of rail transit network over multiple time periods. *Transportation Research Part A*, 235-250.

Pujol, R., Perez, E., & Aguero, J. (2016, March 29). *Actividad Económica, características sociales y demanda de transporte.* Retrieved November 11, 2024, from

INEC, Censo 2011: https://admin.inec.cr/sites/default/files/media/anempleocenso2011-04.pdf_2.pdf

Rui, L. (2017). *Phase-based Planning for Railway Infrastructure Projects.* [Doctoral dissertation, Technical University of Denmark]. DTU Research Database. Retrieved from https://orbit.dtu.dk/en/publications/phase-based-planning-for-railway-infrastructure-projects

Song, P., & Yang, Q.-F. (2008). Empirical Research on Construction Scheduling for Urban Rail Transit Network. *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing*, (pp. 1-5). Dalian, China. doi:10.1109/WiCom.2008.2092

Sun, Y., Guo, Q., Schonfeld, P., & Li, Z. (2017). Evolution of public transit modes in a commuter corridor. *Transportation Research Part C, 75*, 84-102.

Wu, F., & Schonfeld, P. (2022). Optimized two-directional phased development of a rail transit line. *Transportation Research Part B*, 424–447.

Xu, W., Zhou, J., Yang, L., & Li, L. (2018). The implications of high-speed rail for Chinese cities: Connectivity and accessibility. *Transportation Research Part A, 116*, 308-326. doi:10.1016/j.tra.2018.06.023

Zavadskas, E., Mardani, A., Turskis, Z., Jusoh, A., & Nor, K. (2016). Development of TOPSIS Method to Solve Complicated Decision-Making Problems: An Overview on Developments from 2000 to 2015. *International Journal of Information Technology & Decision Making, 15*(3), 645-682. doi:10.1142/S0219622016300019