TECHNISCHE UNIVERSITÄT MÜNCHEN

Ingenieurfakultät Bau Geo Umwelt
Lehrstuhl für Verkehrstechnik

A Methodology to Design Multimodal Public Transit Networks: Procedures and Applications

Zain Ul Abedin

Vollständiger Abdruck der von der Ingenieurfakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Gebhard Wulfhorst

Prüfer der Dissertation:

1. Prof. Dr.-Ing. Fritz Busch
2. Associate Prof. David Wang Zhiwei, Ph.D.

Die Dissertation wurde am 17.04.2019 bei der Technischen Universität München eingereicht und durch die Ingenieurfakultät Bau Geo Umwelt am 09.07.2019 angenommen.

Acknowledgements

Throughout my stay at TUMCREATE, I came across, some of the most intelligent, kind-hearted and encouraging people. They have touched my life in so many ways, that I would not have been able to do this without their support.

Firstly, I would like to express my gratitude to Dr. Andreas Rau who believed in me and supported me for so many years through my stay at TUMCREATE. Secondly, I would like to thank Prof. Fritz Busch, whose continuous support and pragmatic view about research helped me to broaden my vision. It was him who told me, “*follow your heart and the mind will follow, believe in yourself and your ideas*”. This gave me immense confidence that I needed, and I will always be thankful to him. Thirdly, I would like to thank Prof. David Z. W. Wang for his continuous support for my thesis, and I appreciate the freedom he gave me to pursue my research ideas. Fourthly, Prof. Avishai Ceder, with whom I have had some of the most productive and delightful discussions about public transit, its challenges and prospects. Not only that, I have also enjoyed his company enriched with anecdotes, riddles and jokes, which always have an interesting transport angle to them. I am grateful to him for his support and help during my PhD studies.

To my colleagues, I would like to thank TUMCREATE in general and especially RP-8b, RP-10 and RRT teammates. The journey of my PhD would not have been possible without you guys. You were always there for me, whenever I needed you. I learnt a great deal, not just the scientific knowledge but also about the culture, customs, beliefs and good food.

A special thanks to PTV AG for their continued support and for providing me a research license of VISUM software for my research!

To my parents, I thank my father Iftikhar and my mother Rizwana, who always prayed and still pray for my success with their every living breath. They always guided me to make the right decision, and not once, I regretted any of my decisions.

To my friends, who listened my sermons about transport, future, history and politics so patiently for all these years. They are and will be a great backing for years to come.

Finally, to my wife Sara, who has stuck with me through thick and thin. I am thankful to you for cheering me up all these times, and pushing me to finish the thesis, whenever, I procrastinated.

Abstract

This thesis deals with public transit network design problem, one of the core issues facing the existing planning and operation of public transit systems. More specifically, this problem involves designing an efficient/optimal set of routes and their associated frequencies for public transit networks, given certain constraints. A novel route overlap-based approach is developed for this design undertake considering the perspectives of passengers and agencies. The concept of route overlap entails the ratio of common segments/arcs that each route shares with the remaining set of the routes of the transit network. Varying this ratio results with a variety of transit-network configurations such as a sparse network with lower ratios and dense network with higher ratios.

The main part of the developed methodology consists of four major components. It starts with transit route creation and network construction where, at first, feasible routes are created by using the traditional k-shortest path algorithm, and then route overlap concept is used to design a public transit network. The second component is the analysis of assigning transit demand including headway derivation; this is used to assign demand onto the transit network and to generate travel times and suitable transit modes. The third component is the holistic evaluation of the transit network; this is done by using performance metrics to reflect the viewpoints of passengers, agencies and authorities. The last component is a metaheuristic search engine employed to explore feasible search space for attaining improved transit network configurations from both passengers and agencies perspectives.

The new methodology undergone testing of different networks including benchmark and real-life networks. The outcomes of the experiments are used for the evaluation and validation of the proposed methodology including a comparison with other research studies. The results show that instead of using many overlapping redundant routes, it is more efficient and productive to have lesser number of routes with a reduced number of overlapping route segments; this will be more suitable and efficient for both passengers and agencies. Finally, the viability of the developed methodology is examined for the Singapore’s public transit network; the results suggest that this network can be optimised further with even lesser number of routes.

Kurzfassung

Die vorliegende Dissertation befasst sich mit dem Problem des Entwurfs öffentlicher Verkehrsnetze - einem der Kernprobleme in Bezug auf die bestehende Planung sowie den Betrieb von öffentlichen Verkehrssystemen. Im Detail beinhaltet dieses Problem den Entwurf einer optimalen Auswahl an Fahrtrouten für öffentliche Verkehrsnetze sowie deren zugehörigen Taktfrequenz unter bestimmten Randbedingungen. Für diesen Entwurf wird ein neuartiger, auf die Variation von Routenüberlagerungen basierender Ansatz entwickelt, welcher die verschiedenen Perspektiven von Fahrgästen und Betreibergesellschaften berücksichtigt. Das Konzept der Routenüberlagerungen beinhaltet den Anteil der gemeinsamen Streckensegmente, die jede Route mit den übrigen Routen des öffentlichen Verkehrsnetzes teilt. Die Variation dieses Anteils führt zu einer Vielzahl an Konfigurationen des öffentlichen Verkehrsnetzes wie beispielsweise einem ausgedünnten Netzwerk bei niedrigem Anteil oder einem dichten Netzwerk bei hohem Anteil an Überlagerungen.

Der Hauptteil dieser entwickelten Methodik besteht aus vier wesentlichen Komponenten. Es beginnt mit der Erstellung von Routen und dem Aufbau des Netzes. Dabei werden zunächst mögliche Routen mit Hilfe des traditionellen k-shortest path Algorithmus erstellt. Anschließend wird das Konzept der Routenüberlagerungen zum Entwurf des öffentlichen Verkehrsnetzes angewendet. Die zweite Komponente ist die Analyse der Umlegung der Verkehrsnachfrage einschließlich der Ableitung der Taktzeiten. Dies dient dazu, für die erstellten Netze Umlegungen durchzuführen sowie die Fahrzeiten und die geeigneten Formen des öffentlichen Verkehrs abzuleiten. Die dritte Komponente ist die holistische Bewertung des öffentlichen Verkehrsnetzes. Dies geschieht durch die Verwendung von Leistungsmetriken, welche die Standpunkte von Fahrgästen, Betreibern und Behörden widerspiegeln. Die letzte Komponente ist eine metaheuristische Suchmaschine, die verwendet wird, um einen realisierbaren Suchraum zu erschließen, um dadurch verbesserte Konfigurationen des öffentlichen Verkehrsnetzes zu erreichen, und zwar sowohl aus Sicht der Fahrgäste als auch aus Sicht der Betreiber.

Diese neue Methodik wurde mit unterschiedlichen Netzen getestet - gängigen Benchmark-Netzen wie auch realen Netzen. Die Ergebnisse der Experimente werden für die Bewertung und Validierung der vorgeschlagenen Methodik verwendet, einschließlich eines Vergleichs mit anderen Forschungsstudien. Die Resultate zeigen, dass eine geringere Anzahl an Routen mit einer geringeren Anzahl an sich überlagernden Streckensegmenten effektiver und produktiver ist als eine Vielzahl sich überschneidender und redundanter Routen und zwar sowohl für die Fahrgäste als auch für die Betreiber. Zum Ende wird geprüft, ob die entwickelte Methodik für das öffentliche Nahverkehrsnetz von Singapur geeignet ist. Die Ergebnisse legen nahe, dass dieses Netz durch die Reduzierung der Routen weiter optimiert werden kann.

Table of Contents

1 Introduction

Majority of the public transit (PT) systems in major mega cities around the world are facing an image problem. There are many reasons for that, however, the two most fundamental problems are: 1) the PT systems do not serve all the regions adequately within the service area; 2) the PT systems have severe reliability issues. As a potential passenger, one might ask, why would I use the PT if it doesn’t serve the region well where I want to go. Even if, I manage to use the PT service, it turns out to be extremely unreliable i.e., cannot adhere to its published schedule and prone to delays. The first problem stems from bad PT network design, although the second problem is related to operational aspects of the PT service. But a bad network design also contributes to PT service unreliability. Often, a PT network is considered good if it consists a lot of high frequency overlapping routes. However, this is not true in most cases, maintaining schedule and reliability for such a PT network is difficult. Moreover, it leads to bad schedule adherence, uneven loads, bus bunching and reduced stop service levels.

Unfortunately, apart from few exceptions, the PT systems never get attention in the past, since private cars were the ultimate choice for personal mobility, attributed to higher speeds, door to door access, free of spatial and temporal limitations. But in the last few decades, mass migration from rural areas to cities and preference of private cars over PT services for daily commute is on the rise. This urban sprawling and the preference of private cars leads to more traffic on the roads and cause congestions and delays. It gets much worse, when the expansion of road infrastructure can’t keep up with growing traffic demand. Therefore, the problem of traffic congestion persists. This issue is partly result of poor PT service design, a significant portion of the residents might choose PT services provided they offer a high spatio-temporal connectivity.

More recently, climate change, pollution, local air quality, smog, road accidents, depleting fossil fuels, and lack of funds to maintain and built new road infrastructure forced the city municipalities and local transport authorities to rethink, how the city’s transport should function. Although, the technologies in cars such as autonomous driving, battery powered vehicles and low emission engines are improving day by day. These technologies will surely help in curtailing a lot of these issues. Still, there is plenty of time before achieving high penetration of such vehicles and technologies. No matter what the future may hold, the PT services will stay relevant. Instead, they will take the advantages of these new technologies, and they will be able to offer even more flexible, improved, reliable and seamless services to masses. For the time being, well planned and functioned PT systems can still help in curtailing transport issues, faced by the cities.

The two main modes of urban transport (private cars and PT services) have different dynamics and level of service. For instance, given the road infrastructure supply is always extended the traffic demand increases. This supply-demand gap is reduced to minimum as the traffic increase on the roads especially in peak hours. With increased traffic, the level of service drops, reduction in travelling speeds and increase in the density of the traffic is witness. Therefore, increase in the supply doesn’t mean permanent service improvement. However, in case of PT system, as the demand increases, the level of service increases as well. This is due to use of high capacity PT services, operating at lower headways with improved speeds, faster fare collection and quicker boarding/alighting at stops. These dynamics of private cars and PT system indicate that, PT services should be given equal importance as they can aide in reducing traffic congestion and offer an attractive alternative to commuters.

The PT system is an extremely important component of urban transport system. It helps in serving the diverse needs of the passengers by offering them reliable service round the clock. Moreover, it also facilitates social and economic activities in a city. Therefore, the PT system for a given city must be designed carefully according to the mobility needs and the aspiration of the city residents. Ceder & Wilson [1986] proposed five sequential steps for PT system planning. These include: 1) network design; 2) frequency setting; 3) timetable development; 4) vehicle scheduling; 5) crew scheduling. Therefore, covering the strategic, tactical and operational aspects of PT system planning. These sequential steps have strong influence on each other. For instance, the number of routes and their topology will dictate how the frequencies of the routes will be determined. These frequencies and route structures are then used to create timetables. In similar fashion, each preceding step will influence the subsequent step. Therefore, it is extremely important that each step is performed carefully. Likewise, the importance of step is directly related to the level of planning it belongs to. For instance, the network design and frequency setting of the routes are the most important strategic steps in PT system planning. If these two steps are not performed properly, the quality of the end results will be subpar.

At present, many transit agencies and authorities are developing and reconfiguring the PT networks around the world. They use procedures, which are authority specific, developed inhouse, and are based on the local conditions and regulations. Not to mention, a lot of such planning is based on expert opinions and past experiences. These procedures and expert opinions are basic heuristics and judgements, often used to create a new route or extend a current route to cater the needs of a newly developed residential or commercial area. Such procedures and expert opinions might bring local improvements, but this is not true at system level, it might lead to system level degradation instead of improvement. Additionally, the human expert opinion and basic heuristics can’t see the picture for a moderately average network of 250 odd stops, as the number of combinations for different network configurations are endless. For large networks with up to 5000 stops, it is far too complicated, to design/redesign a network with just basic heuristics and opinions.

For large PT networks, an algorithmic approach is required, to help and aide transport planning agencies/authorities to create/reconfigure better PT networks. Such an approach considers the whole network at once, and it can generate a number of different network configurations serving different objectives. Later, the planners can select any variation of these network configurations. However, using such algorithmic approach does not mean that the planning knowledge or expert opinion is not important. Such opinions and knowledge are naturally the next steps after the algorithmic approach-based PT network is created. Since, there are a lot of assumptions and simplifications involved in the algorithmic approach, the expert opinions and planning knowledge can be used to improve the created network, reduce the number of anomalies and tailored to the local needs.

The PT network design involves creating a set of routes and their frequencies by incorporating certain objectives such as minimisation of total travel time for the passengers or minimisation of the number of transfers. These objectives are subjected to a number of different constraints such as fleet size limit, operational cost limit, percentage of unsatisfied demand. More formally, it is called transit network design problem (TNDP) if only the routes and terminals are considered. In case, route frequencies are also involved then, it is called transit network design and frequency setting problem (TNDFSP). A lot of research has been conducted in the last few decades in solving TNDP and TNDFSP, with a wide variation of approaches and methodologies.

The approaches and methodologies adopted to solve TNDP and TNDFSP involve a lot of simplifications, such as simplified objective functions and constraints. These simplifications help in reducing the complexity of the problem, however, the end solutions (PT networks) often tend to be non-realistic with little practical viability. The transit agencies and authorities cannot rely on such solutions, instead a more realistic approach is required, which uses realistic objectives, and which can easily be adopted in creation, enhancement and reconfiguration of the transit networks.

This thesis presents a concept of route overlap ratio, which can be used to create a variety of different PT network configurations, such as a sparse network for high demand corridors and a dense network for low demand routes. Moreover, a hybrid of the two networks is also possible with such concept. Through this concept, different configurations of PT network, serving different objectives can be created. This concept is further enhanced by introducing demand-based multiclass modes. These modes are assigned to the routes based on the demand requirements within the PT network. In order to capture more realistic behaviour of the passengers' route choice, instead of simple all or nothing (AON) assignment, headway-based capacity constrained transit demand assignment is used. Besides that, a holistic evaluation of the PT network, considering perspectives of passengers, agencies and authorities is used for the assessment of the PT networks.

The central idea is to: 1) first create feasible routes, among these routes, a subset is selected to form the PT network; 2) this subset is created by employing a route overlap ratio concept, where a route can only be selected if it doesn't share more than a predefined route course with any other route. Through such concept, PT network is constructed, route by route. The process continues until the PT network feasibility criteria are met. These criteria include node duplication, route duplication, node coverage and network connectivity; 3) Once the feasible PT network is created, transit demand is assigned to the PT network. This provides both supply and demand side information such as travel time components, headways and mode types for each route in the PT network; 4) with this information, the evaluation of the created PT is performed on a set of different performance measures. The quality of the PT network is judged on the selected performance measures; 5) The process of improving the quality of the PT network is performed by employing a multi-objective evolutionary algorithm (MOEA) called nondominated sorting genetic algorithm-II (NSGA-II) to obtain a set of pareto optimal PT networks.

1.1 Motivation

PT network design is predominantly designed by transit planning agencies and authorities based on the analysis of current PT system, adopted master plans and adherence to local service level guidelines. Additionally, expert opinions are also given importance when designing, restructuring or extending the PT network. Such planning mechanism is extremely important and should be given its due importance. However, human eyes, basic planning guidelines and service standards are not capable to see the big picture, and to encapsulate system level PT network planning.

The reason why, such PT network planning is still conducted in this way is due to a number of factors: 1) lack of data availability, which can be used for PT network planning; 2) too much effort required to capture data via surveys and data gathering; 3) lack of quality methods and algorithms that can be used to design PT network; 4) lack of commercial software packages' availability, which can create PT networks given certain constraints and requirements. From an academic point of view, the topic of PT network design problem received a lot of attention in the past three decades. There exist plenty of various approaches and methodologies to solve this problem. Some of these approaches, require simplification of this problem to an extent that it is no more viable for real world implementation. Other type of approaches can solve the problem within acceptable time, but the quality of the end result is not up to the required standard.

The reason why such approaches and methodologies developed by academia are not used by the PT agencies and authorities are because of: 1) too much simplification of the problem; 2) inability to handle actual size PT networks; 3) unrealistic PT network designs. The issues and problems faced by the PT agencies and academia are getting less day by day. Thanks to the advent of new technologies such as low-cost high-performance computing (HPC), cloud computing (CC), internet of things, smart phones, smart card, information communication technologies, automatic passenger counting and automatic vehicle location. Through such technologies, data capturing is possible. Such data includes demand data about passengers as well as the supply data for PT services. This data from different sources can be utilised for modelling and simulation purposes. On top of that, with HPC and CC, data computation cost involved in such modelling and simulation can be reduced significantly.

Such data and information availability provide an interesting opportunity to fill the gap between the academia and the PT authorities/agencies to work together to solve the problems faced by PT systems. A methodology or a set of procedures which can use this data to derive a PT network, which is realistic, and whose quality can be measured and compared with the existing PT system, is worth investigating. Moreover, such methodology should not compete with the existing procedures and PT network planning techniques, used by many PT planning agencies. Instead, it should aide the PT planning agencies in improving the quality of PT networks.

1.2 Goals and contributions

The aim of this thesis is to investigate the TNDFSP and to come up with a PT network design methodology, which is based on a simple idea yet highly capable i.e., understandable and acceptable to PT planning agencies. Furthermore, such methodology should easily be adopted by these agencies in their procedures and processes. Apart from that, all the components of the proposed methodology should be able to overcome the limitations of current PT network design methodologies. These includes: 1) lack of realistic demand assignment; 2) lack of demand-based multiclass modes for different routes; 3) lack of holistic evaluation representing the perspectives of all stakeholders; 4) inability to tackle large urban PT networks.

The thesis contributions include: 1) a novel route overlap ratio-based PT network creation concept; 2) realistic multi-objective approach considering passenger and PT agency cost; 3) demand-based multiclass mode selection; 4) capacity constrained headway-based assignment; 5) holistic PT network evaluation; 5) real world case study for a large urban network.

1.3 Organisation

The main thesis is presented in seven chapters. In chapter 2, literature review is presented with detailed listing of different approaches, methodologies and their combinations used to solve TNDFSP. Moreover, different demand assignments and PT network evaluations are also discussed. Chapter 3 introduces considered objective functions used to represent perspectives of two different stakeholders of the PT system, constraints and assumptions used in the proposed approach to solve TNDFSP. In chapter 4, details about the proposed NSGA-II based approach using route overlap ratio concept is explained. This includes route creation procedures, transit network creation procedures, transit demand assignment, transit network evaluation and NSGA-II for optimisation. The implementation of the proposed methodology on a set of benchmark networks along with and its sensitivity analysis is performed in chapter 5. A case study of a real PT network of Singapore is conducted in chapter 6. This includes data preparation and implementation of the proposed methodology. In chapter 7, conclusions are drawn from the thesis along with outlook.

1.4 Summary

This chapter has discussed the importance of PT systems and why they should be given more attention. Apart from that, the challenges which are being faced by PT agencies and authorities were also briefly discussed. Moreover, the mismatch between the academia and the PT planning agencies was also highlighted. Besides that, the opportunity to utilise the available data to help the PT agencies and authorities for improving their existing PT networks was mentioned. Finally, the goals and the contribution of this thesis was mentioned.

The next chapter will look into the literature related to the TNDFSP, where the literature is divided into different approaches, methodologies and their combinations. Other important components such as demand assignment, PT network evaluation and limitation of the current literature will also be discussed.

2 Literature Review

2.1 Transit network design and frequency setting problem (TNDFSP)

The TNDFSP mainly consist of two phases: 1) initially, the design of PT routes with terminal identification and stop sequences is acted on; 2) finally, the frequencies of the created routes are calculated. Formally, it can be stated as the determination of PT routes with associated frequencies, subjected to constraints in search of selected objectives.

2.1.1 Objectives

The resultant PT network is highly dependent on the selected objective functions, decision variables, and constraints. The objectives usually represent the interests of passenger, agency and authority. Here the term ‘agency’ refers to the operator, while the ‘authority’ refers to the governmental agencies. The passenger wants a fast, direct, and reliable service with minimum waiting time. Whereas, the agency wants to minimise the required fleet size, PT network, PT vehicle kilometres and operating costs. Authority prefers high PT utilisation, better area coverage, and decent policy headway maintenance. In reality, multiple objectives are desired, these objectives are conflicting in nature. Minimisation of passenger costs lead to an increase in agency costs and vice versa.

For instance, passenger desires a PT service which offers direct connection with minimum transfers and waiting time. This will yield the desired PT network but at the cost of many lengthy routes, requiring large fleet to provide the required service. An increase in the number of routes, their lengths, and short headways tend to be operationally expensive for the agency. In similar fashion, lesser number of routes with long headways tend to be operationally viable and profitable for the agency. However, such PT service is not attractive for the passengers as it involves long waiting times and higher number of transfers. However, the authorities prefer the middle ground where the PT network is attractive for the passengers at the same time, it is operationally viable and requires minimum subsidy.

The most commonly used objectives in TNDFSP related literature can be categorised into three categories: 1) passenger cost minimisation and 2) agency cost minimisation; 3) passenger and agency cost minimisation. Typical objectives for passengers and agencies include:

$$\min O_1 = OWT + InVT + TWT + tp.NoT \tag{2.1}$$

$$\min O_2 = TNL \tag{2.2}$$

Where

OWT	Origin waiting time
$InVT$	In-vehicle time
TWT	Transfer waiting time
NoT	Number of transfers
tp	Transfer penalty
TNL	Total network length

2.1.2 Decision variables

Although there are a number of decision variables but the key decision variables in the TNDFSP are the routes and their associated frequencies.

2.1.3 Constraints

The PT network design for a given objective is heavily dependent on the constrains and parameters. These constraints help in attaining a desired feasible PT network. For the TNDFSP, following constraints can be considered:

2.1.3.1 *Number of routes*

The number of bus routes are decided in advance by the planning agencies. This decision depends upon the context, available fleet size and PT design guidelines. In most of the studies, the number of routes is predefined. However, studies which deal with real networks, define a range for minimum and maximum allowed number of routes.

2.1.3.2 *Public transit network connectivity*

To assure that all the stops in the study area are served, the designed PT network must be able to offer the following two properties: 1) every stop in the PT network must be served by at least one route; 2) every route must share at least one stop from other route. The former property assures that all the stops are served by the PT network whereas, the latter promises that for any given OD pair, a trip can be made from origin stop to destination stop.

2.1.3.3 *Route length and stops*

The route length range is predefined and heavily dependent on the local environment. Lengthy routes are avoided due to reliability and scheduling issues. Similarly, short routes are also avoided for the sake of efficiency. Besides the route length, the number of stops in a route also plays a decisive role. A route with too many stops is prone to schedule delays.

2.1.3.4 *Route Detour*

To improve the PT network image, the routes should serve the major demand centres as direct as possible, with little deviation from the shortest path. The detour is a ratio of the route path length to shortest path length. The typical values lie between 1.2-1.5 times the shortest path (Ceder [2016]).

2.1.3.5 *Route stop duplicity*

The route duplicity means, the occurrence of same stop twice in a route. This is generally true for circular routes. However, majority of the studies prefer no duplicate stops in the routes, simply because of the limitations associated with the path finding algorithms.

2.1.3.6 *Route overlap*

A route overlap makes sure that no two routes share more than a predefined route course with each other. Too much route overlapping makes the transit network confusing for the passengers and it is considered as network redundancy.

2.1.3.7 *Route frequency*

The route frequencies are determined by many different aspects. The policy frequency comes from the transit authority, which provides the lower and upper bound for the route frequency. The demand load on the route defines the actual route frequency for a given transit vehicle capacity.

2.1.3.8 *Route load factor*

The route load factor shows the allowed load on a route. This represents different comfort levels for the passengers on the route. Higher route load factors mean much crowded routes, whereas lower values show adequate loads on the route.

2.1.3.9 *Fleet size*

This represents the available vehicle fleet that can be used for determining the frequencies of the routes. The fleet size has significant impact on the frequencies, mostly the resultant route frequency is a compromise between available vehicles, the policy headway and the route load profile.

2.2 Approaches to solve TNDFSP

Several approaches have been used to tackle TNDFSP, however, these approaches can be divided into to three main categories: 1) heuristic; 2) metaheuristic; 3) exact methods. The heuristics are the simplest procedures used in a systematic way to solve the TNDFSP. In metaheuristics, TNDFSP can easily be represented by existing metaheuristic search models. These models are composed of heuristic procedures and stochastic optimisation algorithms. These are not problem specific and are extremely flexible. Thus, any objective, and constraint can easily be adopted. They can solve large networks in acceptable time with satisfactory solution quality. However, there is no guarantee of achieving an optimal solution. In exact methods, the TNDFSP needs to be formulated in to known mathematical models. These can be solved either by gradient based approaches or by integer optimisation. Such models can find the optimal solution, but they are less flexible and can only be implemented on generalised small networks.

2.2.1 Heuristic approaches

The use of heuristic techniques are the most common types of approaches used to solve the TNDFSP. These are widespread and have been used in the PT design for quite some time. These are mostly domain specific and depends upon the knowledge and experience of PT planners. The results achieved from these procedures usually aligned well with the established guidelines and quality indicators.

Lampkin & Saalmans [1967] used a sequential approach to tackle TNDFSP, first by designing the PT routes and later setting the frequencies. First a skeleton transit network is created by using heuristic procedures with the objective of maximising the direct trips. Once the routes are created and given the available fleet size, the frequencies are derived such that the total travel time is minimised.

Rea [1971] proposed a service specific model, where no specific objective function was listed instead the focus was on satisfactory solution. A solution which meets the performance levels, set by agencies. The selection of links to construct routes was based on a service specific model. Moreover, single path assignment was used, with the assumption that all passengers will travel on the least weighted cost path under fixed demand.

Silman, Barzily, & Passy [1974] presented a similar two-staged methodology with the objective of minimising the sum of journey times. However, their model also considered transfer time between vehicles and discomfort penalties for those passengers who couldn’t find seats. In first phase, the candidate route set was created through route addition and deletion processes. In second phase, the frequencies were decided for the created route set by complying with the available fleet size.

Mandl [1980], presented a heuristic algorithm to derive the optimal routes with the assumption of fixed frequency for all the routes. First, a feasible set of routes is generated, later rule-based heuristics are applied to modify the route structures, and to improve the average transportation costs. The process continues until no further significant improvements can be achieved. Mandl was one of the first author to propose a benchmark network for solution methodology evaluation.

Ceder & Wilson [1986] proposed a two-level methodological approach for the design of the bus route network and their frequencies. Two levels were considered, with the first one focusing on passenger and the second one focusing on passenger and agency. For passenger, the focus was on minimising the total travel cost while for passenger and agency, a balanced travel time and waiting time together with required fleet size. In addition, the route frequencies and timetables were also set in second level. Israeli & Ceder [1989] introduced an enhanced approach to tackle TNDP with seven sequential steps. It starts by creating all the feasible routes serving direct as well as transfer-based trips. Later demand assignment is carried to determine the passenger waiting time, in-vehicle time and empty seat hours. Furthermore, the total required fleet size is also calculated. In similar fashion, multiple solutions were created, to be presented to decision maker, to choose the most suitable solution.

Nes, Hamerslag, & Immers [1988] solved the TNDFSP by simultaneous selection of routes, their frequencies and the number of passengers being served directly, for a given fleet size. They proposed an analytical model, where mathematical methods were applied. They used route marginal efficiency concept in an incremental way to improve the quality of the routes.

More recently Baaj & Mahmassani [1991] and Baaj & Mahmassani [1995] proposed an artificial intelligence-based approach to solve the TNDFSP problem. At first, an initial set of routes were created by heuristic route generation algorithm. Later, frequencies were assigned to the created routes. The route analyst procedure was employed to evaluate the created routes and their associated frequencies. Finally, route improvement procedure was used to improve the created routes by altering the routes structures. Shih & Mahmassani [1994]; Shih & Mahmassani [1997] proposed an enhanced approach to tackle TNDFSP. The difference between them and the earlier research work was primarily the schedules and variable vehicle fleet size. They proposed solutions for different types of schedules mainly, uncoordinated and coordinated time transfer systems. They implemented the proposed solution methodology on the transit network of Austin, Texas.

Carrese & Gori [2002] proposed a heuristic hierarchal urban transit-network design methodology, composed of express and feeder routes. A heuristic based procedure was employed to create different layers of public transit routes. First, the express routes were created by following an objective function of minimising passenger and agency costs. Later these were fixed as the main routes. In the final step, feeder lines were created, and improvement procedures were used to improve the transit-network performance such as reduction in transfers and operating costs. The authors tested their methodology on the transit network of Rome, Italy.

Lee & Vuchic [2005] used an iterative approach to solve the TNDFSP with varied demand. First, an initial set of feasible routes was generated. Second, demand was assigned to the routes with route modification and elimination procedures. The process terminated when there was no notable change in the network-performance attribute values.

Mauttone & Urquhart [2009] extended the work of Baaj & Mahmassani [1991] and introduced a pair insertion algorithm to further improve the set of routes. The proposed algorithm can provide better results with respect to the viewpoints of both passenger and agency. Moreover, the proposed algorithm showed improved results for agency perspectives, when compared to previous approach. The algorithm was tested on the real city of Rivera, Uruguay.

Kim, Kim, Kho, & Lee [2016] developed a bi-level model to solve the TNDP, which simultaneously decides the mode type, route configuration and its frequency. They showed that the proposed integrated decision model can provide better solution compared to sequential decision methods. They also tested their methodology on a simple benchmark network.

2.2.2 Metaheuristic approaches

Metaheuristic approaches are advanced version of heuristics. Unlike the traditional heuristics, where domain knowledge is extremely important, they do not require such knowledge. Additionally, they are extremely flexible and can easily be adopted to any given problem. These are ideal techniques when the problem search space is too large and efficient searching of the whole space is not possible. Although, it can find feasible solutions with increased search efficiency but with one major limitation, that it does not guarantee optimality of the solution.

In the past two decades, many researchers have developed and used metaheuristic techniques to solve large instances of this problem, with acceptable solution qualities and within acceptable time. According to Talbi [2009], typical metaheuristic techniques include genetic algorithms (GA), tabu search (TS), simulated annealing (SA), bee colony optimisation (BCO), ant-colony optimisation (ACO), and particle swarm optimisation (PSO).

Pattnaik, Mohan, & Tom [1998] and Tom & Mohan [2003], used two level approach where a heuristic based candidate route set generation algorithm was used to create candidate routes and GA was employed for objective specific route subset selection as the final solution. Chakroborty & Wivedi [2002] and Ngamchai & Lovell [2003], also used two-tier strategy for network design. An initial-route-set generation algorithm was used to create feasible route set. Subsets of these routes were then fed to the route-evaluation procedure. After this procedure, a GA-based route modification and improvement procedures were employed to derive an optimised route set.

Petrelli [2004] proposed a solution methodology to solve TNDFSP in three phases by using heuristic and metaheuristic procedures. In phase one, a heuristic algorithm was used to generate a set of feasible routes. In phase two, GA was employed to find the optimal sub-set of routes. Finally, improvement of the network configuration was performed to improve the quality of the transit network. Zhao & Zeng [2006] proposed a methodology, where a stochastic search scheme, based on an integrated SA and GA was proposed. The methodology was implemented on a large-scale real network.

Fan & Machemehl [2006] proposed a framework with three main components to solve the TNDFSP. The first component was a heuristic-based initial candidate route set creation procedure. The second was a network-analysis procedure, which assigned the demand to the selected routes, and later derive route frequencies and their performance. The last component was a GA-based selection procedure, used for the selection of the optimum set of routes.

Fan & Mumford [2008] proposed a framework for solving the TNDP. First, the initialisation procedure was used to create feasible routes. Later, the modification procedure was used to make small adjustments to the created routes. A feasibility procedure was utilised to check whether the selected routes were constraints compliant or not. Successful candidates were then evaluated based on a set of performance indicators. The selection procedure, which was based on HC and SA, was used to select the best set of routes. Fan, Mumford, & Evans [2009] solved the TNDP by using evolutionary multi-objective optimisation technique. They used similar approach to Fan & Mumford [2008], however, they considered multiple objectives. Zhang, Lu, & Fan [2010] And Mumford [2013], used an improved version of the methodology employed by the previous two studies. Moreover, the latter used larger benchmark networks to test their algorithms.

A hierarchal network-design methodology for actual-size networks was proposed by Bagloee & Ceder [2011]. First the categorisation of the potential stops was performed by using a clustering concept. Once stops were identified, Newton’s theory of gravity, together with a manipulated shortest-path algorithm, was used to create hierarchal (mass, feeder, and local) routes. In the last stage, a metaheuristic search engine was launched to search for the best solution among the candidate routes.

Szeto & Wu [2011], proposed a methodology for solving the TNDFSP. They used GA to solve the network-design part of the problem and neighbourhood-search heuristics to optimise the frequency setting of the problem. In the beginning, the initial solutions of route structures were randomly generated. Later, these solutions were evaluated by the heuristics-based frequency-setting procedure. Some custom GA operators were also used to improve the robustness of the solution quality.

Cipriani, Gori, & Petrelli [2012] proposed a two-step methodology to solve TNDFSP. At first, heuristics were employed for route generation, which were based on flow concentration process. Later, a parallel GA was used to find the suboptimal route set with their respective frequencies. The authors tested the developed methodology on the real network of Rome, Italy. Hosapujari & Verma [2013] proposed a hub and spoke model for solving TNDFSP. First the potentials hubs were identified, later nodes were assigned to these hubs at optimal locations. Inter hub and feeder routes were designed, and GA algorithm was employed to find the objective specific optimal route set.

Nikolić & Teodorović [2013] developed a BCO-based methodology for TNDFSP. In the initial stage, heuristics were used to create feasible routes and later BCO was used to investigate the solution search space, to find the best set of routes. Chew, Lee, & Seow [2013] proposed a two-level methodology for solving TNDFSP problem. An initial set of feasible routes was generated after going through feasibility checks via shortest path algorithm. Later GA was applied over the feasible candidate route set to find the optimal subset of routes. Kuo [2013] also proposed a two-level network design methodology. In the initial stage, a set of routes was generated and in the final stage, SA based optimisation was used to attain an optimised route set.

Kechagiopoulos & Beligiannis [2014], developed a methodology where the initial set of feasible routes were generated by heuristics while a PSO-based algorithm was applied to select the best possible subset of routes from the feasible route set. Nayeem, Rahman, & Rahman [2014] used a greedy algorithm to generate initial set of routes. Later, they used two versions of GA with elitism to solve TNDP; 1) with fixed size and population; 2) with increasing population size. Kiliç & Gök [2014] solved the TNDFSP by using a two steps methodology. In the first step, heuristics were used to create the initial set of routes and in the second step, local search algorithms (hill climbing and TS) were used to search for the best solution in the solution search space.

Majima, Takadama, Watanabe, & Katuhara [2014] and Majima, Takadama, Watanabe, & Katuhara [2015] tackled the TNDFSP with a multi agent system (MAS). First, initial set of routes were generated through a community detection algorithm. Later, the evolution of MAS started for line agents to achieve the objective specific set of routes. Zhao, Xu, & Jiang [2015] proposed a memetic algorithm-based optimisation for solving TNDFSP. They used custom operators for improving the quality of the chromosome and used try-and-error procedure to improve the search efficiency.

Arbex & Cunha [2015] employed two level strategy for solving TNDFSP. In the first step, heuristics were used to create feasible routes database. In the second step, this database was then used by alternating objective GA (AOGA) to search the best solution. The AOGA cyclically altered the objectives to improve the search and to tackle the multi objective nature of the TNDFSP. Buba & Lee [2016] proposed a differential evolution algorithm for solving the TNDFSP with the objective of minimising the average travel time of all passengers. The authors introduced a new repair procedure called sub-route reversal mechanism, which was used to reduce the infeasibility among candidate route sets. The authors tested the efficiency of their proposed repair procedure by comparing it with other solutions for a selected benchmark network.

Huang, Liu, Fu, & Blythe [2018] introduced a three-stage method to solve TNFSP. In the first stage, a novel clustering algorithm was designed to identify the hubs in the considered network. In the second stage, routes were designed by considering the viewpoints of passengers and operators. In the last stage, bi-level programming problem was adopted to describe the concept of optimal lines and their frequencies. The upper level objective function was to minimise the sum of passenger and agency cost, the lower level dealt with passenger route choice behaviour. Due to non-deterministic polynomial-time (NP)-hard (see section 2.2.4) nature of the problem, the authors used BCO to solve the problem. Moreover, they also tested their methodology on both benchmark and real network instances.

2.2.3 Exact method approaches

The mathematical approaches for the TNDFSP (TND + TNFSP) tend to focus on specific aspects of the problem, mainly due to its NP-hard nature and their tendency to get more inefficient as the problem instance grows.

Schéele [1980] approach the problem of travel pattern determination and required route frequencies in a given PT network by formulating it as a non-linear programming problem in the form of a compound minimisation problem. The model was implemented on the bus network of Linköping, Sweden with 80,000 inhabitants.

Furth & Wilson [1981] proposed a model to solve the problem of bus allocation to the PT routes to maximise net social benefits. They treat problem as a constrained resource-allocation problem. The algorithm was developed to solve the resulting mathematical program, and a case study of Arborway Garage of MBTA was chosen to illustrate the capabilities of the model.

Constantin & Florian [1995] formulated a bi-level Min-Min problem, to optimise the route frequencies with the main objective of minimising the total travel time. Bussieck [1998] formulated the TNDP as a non-linear integer program and solve it via relaxation and branch and bound method. The author proposed cost optimal planning to determine routes with maximum direct trips.

Wan & Lo [2003] solved the TNDFSP by modifying the existing transit system, they presented a mixed integer formulation for this problem with the objective of minimising the sum of the operating costs. The adopted approach was only suitable for small networks and the authors believed that heuristics algorithms are crucial for solving large networks.

Schöbel & Scholl [2005] approached the TNDFSP problem by formulating it as a mixed integer linear programming (MILP) model with the main objective of minimising the total travel cost for the passengers. Since the TNDFSP is a NP-hard problem, simple solution to linear programming (LP) relaxation was not possible, therefore, the authors used Dantzig-Wolfe decomposition for solving the LP-relaxation.

Guan, Yang, & Wirasinghe [2006] tried to model the PT routes and passenger transfers by a linear binary integer program, which was solved by standard branch and bound method. The objective was to minimise the total PT network length, total number of PT routes and total length travelled by passengers. The authors implemented their model on simplified Hong Kong mass transit railway network. The authors acknowledged that, dealing with large size instance can be cumbersome therefore, metaheuristic approaches should also be explored.

Cancela, Mauttone, & Urquhart [2015] proposed a MILP formulation to solve TNDFSP with the objective of minimising passenger cost while adhering to limited fleet size constraint. They used two graphs (infrastructure and trajectory) in one model, to capture the aspects of planners and the passengers. They also tested their model on a number of networks including benchmark as well as real networks. The authors agreed that heuristics are by far the best approaches when it comes to solve a realistic PT network instance.

Chu [2018] proposed an innovative mixed-integer programming (MIP) model to solve TNDFSP for minimising the weighted sum of agency operating cost, passenger cost and unsatisfied demand. They solved the problem by developing parallel branch-price-and-cut algorithm. The authors also conducted a computational study to evaluate, and to understand the performance of the proposed methodology along with the effects of the proposed model's parameters on the results. Moreover, they also tested their model on a small benchmark network.

The ability of exact methods to find the optimal solution makes it obvious choice in many domains. However, in TNDP and TNDFSP, these methods do not scale well to actual sized transit networks. Most of the mathematical models were tested either on small test networks or simplified real transit networks.

2.2.4 Complexity of TNDFSP

All the above-mentioned approaches to TNDP and TNDFSP showed that solving these problems to get optimal solution for actual size transit network is extremely difficult, complex and time consuming. According to Baaj & Mahmassani [1991], five sources of complexity prevents seeking an optimum solution for TNDFSP.

- Difficulty of defining the decision variables related to TNDFSP.
- The non-linearities and non-convexities exhibit by TNDFSP formulation.
- The combinatorial explosion related to discrete nature of TNDFSP.
- The multi-objective nature of the TNDFSP.
- Formal characterisation and incorporation of routes' spatial layout.

The sources mentioned by Baaj & Mahmassani [1991], prevent an exact method optimisation solution for TNDP. Therefore, a hybrid of heuristics and metaheuristics approaches seem suitable for solving actual sized transit network instances. This is evident from the literature as well, because many of such hybrid techniques are tested on real networks. However, the optimality of the results is still an issue, but the generated solutions are a compromise between the computation cost and desired solution quality.

Schöbel [2012] reviewed literature related to TNDFSP. The author discussed the complexity of the TNDFSP in terms of the objective functions and basic constraints. According to the author, even the very basic version of TNDFSP, where a subset of feasible routes is searched by complying to minimum and maximum allowed route frequencies, is NP-hard. Even if one finds special cases which can be solved easily, the problem becomes NP-hard, the moment the cost of each route is involved.

2.2.5 Comparison of heuristics, metaheuristics and exact methods

Each approach has its own advantages and limitations, for instance, heuristics are perhaps the easiest and widely adopted approaches among all others. But they are scenario specific and cannot be generalised. While, metaheuristics are upper level heuristics, they try to overcome the limitations attributed to basic heuristics by providing the generality that heuristics lack. They guide heuristics to search for a better solution to an optimisation problem. These techniques can find good quality solutions with less computation effort when compared to exact methods. But then again, they too have one major drawback, they do not guarantee optimality of the created solution. While, the exact methods can solve the problem with guaranteed global optimum solution. But, the size of the problem instance is an issue. The TNDFSP belongs to NP-hard category, and there exist no known methods which can solve the problems belong to this category efficiently. A brief comparison of these three approaches is given in the Tab 2.1 below.

Approach	Advantages	Disadvantages
Heuristics	Simple Computationally efficient Widespread utilisation	No guarantee of optimality Cannot be generalised Requires domain knowledge
Metaheuristics	High flexibility Easy implementation Guaranteed feasible solution	No guarantee of solution optimality Requires fine tuning of parameters
Exact methods	Guaranteed solution optimality Availability of commercial solvers	Computationally inefficient Lack flexibility

Tab 2.1 Comparison of heuristics, metaheuristics and exact methods

2.2.6 Summary

Each approach has its advantages and limitations. However, in the past two decades, the focus is shifting more towards metaheuristics. There are far too many studies on TNDFSP, which used metaheuristics as compared to exact methods. But in real world, most of the transit agencies and authorities are still using heuristics to: 1) identify PT network problems; 2) design or restructure PT network; 3) improve existing PT network. There are many commercial software packages available to the agencies for PT system management. However, up till now, there isn’t a single commercial software package, which can deal with network design of PT system. This means either the developed methodologies are too simple, i.e., lacks the details that a typical transit agency requires, or they can’t handle actual size PT network.

2.3 Methodological procedures and combinations

In literature, a wide variety of methodological approaches have been used to solve the TNDFSP. Most of these methodologies are composed of a set of heuristic and metaheuristics procedures. These procedures have some unique features, but the general ideas are rather similar. Therefore, majority of these procedures can be grouped into four general categories: 1) route creation; 2) route selection; 3) route configuration; 4) network improvement. Moreover, in terms of their combinations they can be combined in three different ways: 1) candidate route creation and selection; 2) candidate route creation and configuration; 3) network creation and improvement.

2.3.1 Route creation

The route-creation procedure is generally heuristic in nature and it is used to create a set of feasible routes for the selected stop pairs.

2.3.1.1 *Shortest path algorithms*

In most cases, a route between two stops points is created based on the shortest path between them. However, the shortest path-based route might not be an optimal route. Therefore, a set of k-shortest paths for each stop pair is created to cover a broader range of alternate feasible routes. Apart from that, manipulated shortest paths-based routes are also created between the given stop pair to service more nodes in-between. The most widely used algorithms to create routes, based on shortest paths include Dijkstra, Floyd and Yen's algorithms Dijkstra [1959]; Floyd [1962]; Yen [1971].

2.3.1.2 *Neighbourhood stop search*

Another approach to create the feasible routes is via neighbourhood stop search procedure. At first, a stop is selected, and all its neighbours are evaluated. A suitable neighbour, based on defined criteria will be added into the route. Then the newly added stop's neighbours are evaluated, and in this way new stops are added into the route. The process terminates once the considered constraints such as maximum route length and/or number of stops per route are violated.

2.3.2 Route selection

In route selection procedure, the selection is made using an iterative process or via metaheuristic-based search algorithms. The objective is to find the best possible subset of routes from the feasible route set.

2.3.2.1 *Iterative process*

The iterative process involves sequential derivation of routes and their associated frequencies.

2.3.2.2 *Metaheuristic-based search algorithms*

The metaheuristic-based search algorithms are advanced form of iterative-based approaches, where they guide the selection process coupled with demand assignment (see section 2.4) and network evaluation (see section 2.5). There are many such algorithms which can be used to search for the best possible subset of routes (see section 2.2.2).

2.3.3 Route configuration

The route-configuration procedure involves simultaneous route selection and route modification. There are several types of route configuration procedures. At a single-route level, where individual route structures are modified to enhance their performance. At network level, where two routes are selected, and their stops are exchanged to create two new routes, or they are merged to create a single new route.

2.3.3.1 *Single route level*

At a single route level, there are many heuristics procedures that are used to enhance the route performance. The main types are as follows:

A *route expansion* takes an establish route and tries to expand it by incorporating new stop at the end or at the beginning of the route. If inclusion of new stop enhances the route then the stop is added into the route or else, it is discarded.

A *route contraction* is the opposite of the route expansion. It tries to enhance the route performance by removing a stop from the route.

A *stop insertion* inserts a stop between two consecutive stops only if there exist an arc from and to a new stop form existing stop pairs. Furthermore, the added stop should not induce detour more than the predefined detour limit.

2.3.3.2 *Network level*

At network level, heuristics are used to change the route structure to enhance the performance of multiple routes. Some of the common heuristics are as follows:

A *route swap* takes place between two routes, where stops are exchanged to see if the routes' performance can be enhanced. Two routes are selected at random and if there exist a common stop between them, then the route stop sequences are swapped, from and to common stop. The swap is carried in such way that the stops of one route will be attached to other route and vice versa. This procedure is used to bring the diversity in the route course and to improve the route performance.

A *route merger* takes place between two routes, where two routes are merged to form a new route. Two routes are selected at random, their stop sequences are analysed, and if they share significant path with each other, routes are merged to form a single route.

2.3.4 Network improvement

The network improvement procedure starts by employing the route-creation procedure to create the most promising routes first. These routes are then used to create a skeleton PT network. Once the skeleton is created, the network is enhanced by adding more routes and at the same time modifying the existing ones.

The four basic procedures discussed above can be used in a variety of combinations for solving the TNDFSP problem. In literature, three types of combinations are mostly used, these are as follows:

2.3.5 Candidate route creation and selection

In the candidate route creation and selection, a set of feasible candidate routes is generated, and a selection is made to obtain the most suitable objective-specific subset of the routes. This is the most commonly used combination for solving TNDFSP.

Ceder & Wilson [1986] used two-level methodology, where the demand-driven shortest paths were used to create feasible routes in the first step. In the second step, a set of routes was selected, and other planning attributes were calculated. This process continued until the predefined criteria were met and an evaluation of the end solution was performed. Pattnaik, Mohan & Tom [1998] and Tom & Mohan [2003] also used two level approach, where a heuristic-based candidate route set generation algorithm was used to create candidate routes. Later GA was used for objective specific route subset selection as final solution. Fan & Machemehl [2006] proposed a framework, which has three main components, for solving the TNDFSP. The first component was a heuristic-based, initial-candidate-route-set creation procedure. The second was a network-analysis procedure for demand assignment. The last component was a GA-based selection procedure for optimum route set selection.

A hierarchal network-design methodology for actual-size networks was proposed by Bagloee & Ceder [2011]. First the categorisation of the potential stops was performed by using a clustering concept. Once stops were identified, Newton's theory of gravity, along with a manipulated shortest-path algorithm, was used to create hierarchal (mass, feeder, and local) routes. In the last stage, a metaheuristic search engine was launched to search for the best solution among the candidate routes. Kechagiopoulos & Beligiannis [2014], developed a methodology where the initial set of feasible routes were generated by heuristics while a PSO-based algorithm was applied to select the best possible subset of routes from the feasible route set. Arbex & Cunha [2015] employed a two-level strategy for solving the TNDFSP. In the first step, heuristics were used to create a feasible routes database. In the second step, this database was used by alternating objective-GA (AOGA) to search for the best solution. The AOGA cyclically altered the objective to improve the search and tackle the multi-objective nature of the TNDFSP.

2.3.6 Candidate route creation and configuration

In the candidate route creation and configuration, firstly, a set of feasible candidate routes are generated. Secondly, a metaheuristic-based selection is made, and thirdly, a heuristic-based route configuration (altering route structures) is applied. In the last step, the selected routes are evaluated against the desired objective values. The process of selection and configuration continues until the desired solution is obtained or the predefined number of iterations are reached.

Mandl [1980] was one of the first authors to use a benchmark network to evaluate the network-design methodology. First, a feasible set of routes was generated, later rule-based heuristics were applied to modify the route structures to improve the average transportation costs. The process continued until no further significant improvement can be achieved. Baaj & Mahmassani [1991]; Baaj & Mahmassani [1995] used heuristics and artificial intelligence to solve the TNDFSP. First, feasible routes were generated by using a heuristic-based, route-creation algorithm. Second, route analysis procedure was used to calculate the performance of selected routes. Third, the route-improvement algorithm was employed to improve the selected routes by changing the route structure, such as route splitting and route joining. Chakroborty & Wivedi [2002] and Ngamchai & Lovell [2003] also used the two-tier strategy for network design. An initial-route-set creation algorithm was used to generate a feasible route set. Subsets of these routes were then fed to the route-evaluation procedure. After this procedure, a GA-based route modification and improvement procedure was employed to generate an optimised subset of routes.

Nikolić & Teodorović [2013] developed a bee-colony optimisation (BCO) methodology for the TNDFSP. In the initial stage, heuristics were used to create feasible routes. Later solution modification procedures and BCO were employed to investigate the solution search space to find the best set of routes. Nayeem, Rahman & Rahman [2014] used a greedy algorithm to generate an initial set of routes. They used two versions of the GA with elitism: 1) with a fixed population size; 2) with an increasing population size, to solve the TNDFSP. Moreover, they used a special mutation operator based on small and big modification procedure. Kiliç & Gök [2014], solved the TNDFSP, using a two-step methodology. In the first step, heuristics were used to create the initial set of routes, and in the second step, local search algorithms (hill climbing and TS) with modify solution procedure were used to search for the best solution.

2.3.7 Network creation and improvement

In the network creation and improvement, three steps are involved: 1) initially, a skeleton network is created by selecting limited routes; 2) the skeleton network is improved by incorporating new routes as well as changing the existing route structures; 3) after each step of improvement, the network is evaluated against a set of indicators and the selected objective. The process of improvement and evaluation continues until the desired network is achieved.

Carrese & Gori [2002] proposed a heuristic hierarchal urban transit-network design methodology. Initially a heuristics-based procedure was employed to generate a skeleton transit network. First, the express lines were identified and created by following an objective function of minimising passenger and agency costs. Later, feeder lines were created and improvement procedures were used for transit network performance enhancement (e.g. a reduction in transfers and operating costs). Lee & Vuchic [2005] used an iterative approach to solve the TNDFSP with varied demand. First, an initial set of feasible routes was generated, later demand was assigned to the routes with route modification and elimination procedures for network improvement. The process terminates when there is no notable change in the network performance.

Fan & Mumford [2008] and Fan, Mumford, & Evans [2009] proposed a framework for solving the TNDFSP. First, the initialisation procedure was used to create feasible routes. Later, the modification procedure was used to make small adjustments to the created routes. A feasibility procedure checked whether the selected routes were constraints compliant or not. Successful candidates were then evaluated based on a set of performance indicators. The selection procedure, which was based on HC and SA, was used to select the best set of routes. Zhang, Lu,

& Fan [2010] and Mumford [2013] used an improved version of the methodology employed by the previous two studies. Furthermore, the latter used new and larger benchmark networks to test their algorithms. Szeto & Wu [2011] proposed a methodology for solving the TNDFSP. They used GA to solve the network-design part of the problem and neighbourhood-search heuristics to optimise the frequency setting of the routes. In the beginning, the initial solutions of the route structures were randomly generated. Later, these solutions are evaluated by the heuristic-based frequency-setting procedure. Custom GA operators were also used to improve the robustness of the solution quality.

2.3.8 Summary

Different variations and combinations of the methodological procedures were used to solve TNDFSP by researchers. The results of these studies include:1) route set; 2) route frequencies; 3) performance indicators. The criteria used to evaluate the adopted procedures were the comparative results. These studies didn’t discuss the impact of the solution methodologies on the end solutions such as routes, their properties, functions and headways. It is worth investigating, as this might provide interesting insights into these procedures and end solutions.

2.4 Demand assignment

Irrespective of which TNDFSP methodological approach is used, the demand assignment is an integral part of the entire process. It is the assignment procedure which provides the basic travel-time components, which are later used in assessing the quality of the end solution. Generally, there are two types of assignment models (1) headway-based and (2) schedule-based transit assignment (Ortuzar & Willumsen [1994]). The suitability of an assignment model depends upon the objective, available information and resources. A detailed comparison and suitability of different assignment models is discussed in Lam & Bell [2002]. According to them, the transit assignment models allow the transit authorities to obtain route loads and other PT service attributes. However, it all depends upon the selected transit demand assignment model. For instance, if the schedule-based assignment is used. All the individual service runs can be considered, this provides detailed results such as individual PT vehicle arrival/departure time, its loads and detailed level of service attributes. In contrast, for more aggregated results, headway-based assignment is used. This type of assignment considers PT routes and not individual service runs per route. With such assignment, average values of route loading and route load profiles can be calculated.

In the literature, only some of the studies used headway-based assignment. The majority of the studies used simplified demand assignment (arc times and transfer penalties) procedures with the assumption of unlimited PT supply capacity. The main reason for using such simple assignment procedure is for simplification and reduction in computation cost. Such simplified assignment procedures are acceptable for small networks however, these models are not ideal for real networks. Because, they miss one of the core cost components of passenger travel time cost (waiting time).

2.4.1 Comparison between headway-based and schedule-based assignment

Selection of the transit demand assignment depends upon the objective, data availability and computational time. Both assignment procedures have their own advantages and disadvantages, some of them are listed in Tab 2.2.

Assignment type	Advantages	Disadvantages
Headway-based	Ideal for strategic planning	Unable to identify peak load
	Minimum data requirement	Not suitable for low frequency PT network
	Computationally efficient	Lacks detailed attributes of PT service
Schedule-based	Ideal for detailed spatio-temporal planning	Requires extensive information
	Suitable for low frequency PT network	Computationally expensive
	Coordination of timetable is possible	

Tab 2.2 Headway and schedule-based assignment comparison

2.4.2 Summary

In the reviewed literature, most of the studies used simplistic assignment procedures, which is oversimplification of the problem. Therefore, not considering the impact of waiting time, limited capacity of PT supply and interdependence of route performance to its frequency. The demand assignment procedure is perhaps one of the most important components of any methodology involved in solving the TNDFSP. This is due to: 1) the evaluation of the demand on PT network depends upon demand assignment results; 2) the identification and evaluation of the PT supply also depend upon the selected demand assignment model; 3) the demand assignment also incorporates the route choice behaviour of the passengers. Therefore, the most suitable assignment procedure based on the available data must be selected for TNDFSP. In order to capture the realism in the results, one must consider both cases: 1) unlimited PT supply capacity; 2) limited PT supply capacity. Often, at peak hours, the loads on the PT vehicles are near to their capacity limit and not all passengers are able to board the first arriving vehicle, instead they must wait for the next vehicle.

2.5 Public transit network evaluation

To judge the quality of a PT network, a holistic evaluation representing viewpoints of the related stakeholders is required. The evaluation is composed of performance indicators, where some represent passenger viewpoints such as average travel time and number of trips with 0, 1, 2 transfers. While others represent agency viewpoints such as total PT network length and fleet size. Indicators that represents authority viewpoints includes unsatisfied demand and unutilised offered capacity. The most common performance indicators used in the literature to evaluate transit networks are listed below:

2.5.1 Average travel time

Average travel time refers to the average time a passenger spends in the PT system for commuting. It is calculated by dividing the total travel time of all OD pairs by the total demand. It includes waiting time, in-vehicle time, and transfer penalties. However, in some cases the waiting time is not considered in the travel time calculation.

2.5.2 Trips with transfer number

Trips with transfers is defined as the number of transfers required to complete the trip from the origin stop to the destination stop for all OD pairs. This includes trips that requires zero, one and two transfers. In most of the studies, trips which require more than two transfers are considered as unsatisfied demand. This indicator strictly represents the viewpoint of the passengers and the authorities.

2.5.3 Total network length

It refers to the combined length of all the routes in a network. This indicator is mainly considered when the TNDFSP is solved for minimising agency costs. Moreover, it strictly represents the viewpoint of the agencies.

2.5.4 Total fleet size

It refers to the number of vehicles required to provide the designed service for the given routes and associated frequencies. Almost all studies used same type of vehicles and capacities to calculate the fleet size. This also represents the viewpoint of the agencies.

2.5.5 Summary

In the reviewed literature, studies used different indicators for the evaluation of the created solutions. However, these indicators are limited and doesn’t truly represent the perspectives of different stakeholders, such as the agency, authority and passenger. For instance, in the literature, the focus was mostly on how much percentage of the trips are served directly, this can be one of the indicators to judge the quality of a created PT. However, using just a single indicator to judge the overall quality of PT network is not correct. New and more realistic indicators which encompass perspectives of other involved stakeholders should also be considered for PT network evaluation.

2.6 Overview of the approaches to TNDFSP

In the previous sections, different approaches, methodologies and procedures, used to solve TNDFSP were discussed. To provide a basic overview of all the all these, and to summarise the key features, all the selected studies are classified by eight features, and are listed in Tab 2.3. These features include: 1) from the problem perspective (objective type, decision variables, considered constraints); 2) from the approach and methodological perspective (approaches and procedures); 3) assignment and network evaluation; 4) experiments type (network instances). For the sake of simplicity and arrangement of the information, the following terms are used in Tab 2.3:

AB	Available budget
AON	All-or-nothing
BM	Benchmark
CAM	Custom assignment model
CCHWAM	Capacity constrained headway-based assignment model
CDAM	Combined distribution assignment model
DS	Demand satisfaction
DS 0/1/2	Demand satisfaction through 0, 1, 2 transfers
DT	Direct trips
FiC	Fictious
FS	Fleet size
H	Heuristics
HWAM	Headway-based assignment model
LF	Load factor
M	Mathematical
MH	Metaheuristics
ND	Network directness
NI	Network improvement
NoR	Number of routes
NoS	Number of stops
NoT	Number of transfers
OC	Operator cost
P&OC	Passenger and operator cost
PAM	Probabilistic assignment model
PC	Passenger cost
PD	Path deviation
RBT	Route backtracking
RC	Route creation
RCO	Route configuration
RE	Route efficiency
RH	Route headway
RL	Route length
RS	Route set
RSL	Route selection
RsC	Route set connectivity
SPAM	Shortest path-based assignment model
TNL	Total network length
TTT	Total travel time
USD	Unsatisfied demand
VC	Vehicle capacity
WT	Waiting time

Study	Objective “To minimise”	Decision variables	Constraints	Approach	Procedures				Demand assignment	Evaluation indicators	Network
					RC	RSL	RCO	NI			
Lampkin & Saalmans [1967]	PC	RS, RH	FS	H			●	●	-	Operating deficit	Real
Rea [1971]	-	RS, RH	-	H				●	AON	-	-
Silman, Barzily, & Passy [1974]	PC	RS, RH	FS	H	●			●	-	-	Real
Mandl [1980]	P&OC	RS, RH	FS	H				●	AON	DS 0/1/2, TTT	BM
Schéele [1980]	PC	RH	FS	M	-	-	-	-	CDAM	TTT, DS 0/1/2, RH	Real
Furth & Wilson [1981]	PC	RH	FS, RH, LF	M	-	-	-	-	HWAM	WT	Real
Ceder & Wilson [1986]	P&OC	RS, RH	RH, NoR, PD, FS	H	●	●			-	-	FiC
Nes, Hamerslag & Immers [1988]	PC	SR, RH	AB, FS	H	●	●			-	RS, FS, DT	Real
Israeli & Ceder [1989]	P&OC	RS, RH	RH, FS	H	●	●			PAM	-	-
Baaj & Mahmassani [1991]	P&OC	RS, RH	RH, FS, ND	H	●	●		●	HWAM	DS 0/1/2, TTT, FS	Real
Shih and Mahmassani [1994]	P&OC	RS, RH	RH, FS, NoS	H	●	●		●	HWAM, CAM	DS 0/1/2, TTT, FS	Real
Constantin & Florian [1995]	PC	RH	FS	M	-	-	-	-	HWAM	TTT	Real
Baaj & Mahmassani [1995]	P&OC	RS, RH	RH, FS, ND	H	●	●		●	HWAM	DS 0/1/2, TTT, FS	Real
Pattnaik, Mohan, & Tom [1998]	P&OC	RS, RH	LF, RH	MH	●	●			HWAM	DS 0/1/2, TTT, FS, USD	Real
Carrese & Gori [2002]	P&OC	RS, RH	DS, NoS, FS	H	●		●	●	HWAM	RS, TTT, FS, TNL	Real
Chakroborty & Wivedi [2002]	PC	RS	-	MH	●		●		AON	DS 0/1/2, TTT	FiC
Tom & Mohan [2003]	P&OC	RS, RH	LF, RH	MH	●	●			HWAM	DS 0/1/2, TTT, FS, USD	Real
Ngamchai & Lovell [2003]	P&OC	RS, RH	RH, LF	MH	●		●		HWAM	-	FiC

Study	Objective “To minimise”	Decision variables	Constraints	Approach	Procedures				Demand assignment	Evaluation indicators	Network
					RC	RSL	RCO	NI			
Petrelli [2004]	P&OC	RS, RH	-	MH	●		●		HWAM	-	-
Schöbel & Scholl [2005]	PC	RS, RH	VC	M	-	-	-	-	CCHWAM	TTT, NoT	FiC
Lee & Vuchic [2005]	P&OC	RS, RH	-	H	●		●	●	HWAM	RS, TTT, TNL	FiC
Zhao & Zeng [2006]	PC	RS	-	MH				●	-	DS 0/1/2, RS	Real
Guan,Yang, & Wirasinghe [2006]	P&OC	RS	RL, NoT	M	●	●			CAM	TNL, NoT, TTT	Real
Fan & Machemehl [2006]	P&OC	RS, RH	RH, LF, FS, RL	MH	●	●			HWAM	-	FiC
Fan & Mumford [2008]	PC	RS	RBT, NoS, RsC	MH	●			●	SPAM	TTT, USD 0/1/2	-
Antonio & Urquhart [2009]	P&OC	RS, RH	-	H	●	●		●	HWAM	DS 0/1/2, TTT, FS	Real
Fan, Mumford, & Evans [2009]	P&OC	RS, RSL	RsC, DS, RL	MH	-	-	-	-	SPAM	TTT, TNL, DS 0/1/2	BM
Zhang, Lu & Fan [2010]	P&OC	RS, RSL	RBT, RsC, NoS	MH	-	-	-	-	-	TTT, TNL, DS 0/1/2	BM
Bagloee & Ceder [2011]	PC	RS, RH	PD,		●	●			CCHWAM		Real
Szeto & Wu [2011]	PC	RS, RH	RBT, FS, RH	MH			●	●	HWAM	TTT, NoT	Real
Cipriani, Gori, & Petrelli [2012]	P&OC	RS, RH	RL, RH	MH	●	●			HWAM	TTT, NoT, USD	Real
Mumford [2013]	P&OC	RS, RSL	RsC, DS, RL, RBT	MH				●	SPAM	TTT, TNL, DS 0/1/2, USD	BM
Hosapujari & Verma [2013]	P&OC	RS, RH	FS, LF	MH	●	●		●	-	TTT, DS 0/1/2, FS	BM
Nikolic & Teodorovic [2013]	PC	RS, RH	-	MH	●	●			SPAM	TTT, DS 0/1/2, UDS	BM
Chew, Lee, & Seow [2013]	P&OC	RS	NoS, RBT, RsC	MH	●			●	SPAM	TTT, DS 0/1/2, UDS	BM

Study	Objective "To minimise"	Decision variables	Constraints	Approach	Procedures				Demand assignment	Evaluation indicators	Network
					RC	RSL	RCO	NI			
Kuo [2013]	PC	RS	RL, NoR	MH	●			●	-	TTT, DS 0/1/2, UDS	BM
Kechagiopoulos & Beligiannis [2014]	PC	RS	RL, RBT, NoR	MH	●		●		-	TTT, DS 0/1/2, UDS	BM
Nayeem, Rahman, & Rahman [2014]	PC	RS	-	MH	●		●		SPAM	TTT, DS 0/1/2, UDS	BM
Kilic & Gok [2014]	P&OC	RS, RH	RBT, NoS	MH	●			●	SPAM	TTT, DS 0/1/2, RSL	BM
Majima, Takadama, Watanabe, & Katuhara [2014]	P&OC	RS, RH	-	MH	●			●	SPAM	TTT, DS 0/1/2, FS	BM
Majima, Takadama, Watanabe, & Katuhara [2015]	P&OC	RS, RH	-	MH	●			●	SPAM	TTT, DS 0/1/2, FS	BM
Zhao, Xu, & Jiang [2015]	PC	RS, RH	FS, RH, RS	MH	●			●	SPAM	TTT, DS 0/1/2, FS	BM
Cancela [2015]	P&OC	RS, RH	FS, VC	M	●	●			CCHWAM	TTT, DS 0/1, RH	Real
Arbex & Cunha [2015]	P&OC	RS, RH	RsC, RBT, NoR, RH, RL, FS	MH	●	●			HWAM	TTT, DS 0/1/2, FS	BM
Kim, Kim, Kho, & Lee [2016]	P&OC	RS, RH	RH	H	●		●	●	-	-	BM
Buba & Lee [2016]	PC	RS	RL, NoR, RsC,	MH	●			●	-	TTT, DS 0/1/2, USD	BM
Huang, Liu, Fu, & Blythe [2018]	P&OC	RS, RH	LF, RL, FS, RH	MH	●			●	HWAM	TTT, DS 0/1/2, RE	Real
Chu [2018]	P&OC	RS, RH	RH, FS	M	-	-	-	-	CCHWAM	TTT, DS 0/1, UDS, RH, NoR	BM

Tab 2.3 Classification of studies related to TNDFSP

2.7 Limitation of current approaches

The TNDFSP has received a lot of attention from researchers in the past three decades, due to its key role in mitigating a wide variety of transport related issues. However, the research mostly focused on solving simplified networks with limited performance indicators for the evaluation purpose. Most of the authors believe that a good PT network is the one which offers direct service to the passengers. Ideally this is true, but in real world there exist many PT layers including bus routes, bus rapid transit (BRT), light rail transit (LRT) and mass rapid transit (MRT). These layers are linked with the help of seamless transfers. Additionally, the unlimited PT supply capacity assumption helps in simplification of the problem, and reduction in computation cost. However, imagine a PT network which doesn't explains which mode will be most suitable for a given route in the network and how much waiting time on average does a passenger spends in the PT network. Furthermore, simplified assignment procedures are used to derive values for performance indicators. First, these assignment procedures don't provide meaning full results and second, the values derived from selected performance indicators (mostly represent passengers' viewpoint) from such results do not show a holistic picture.

2.7.1 Lack of solution quality evaluation

An extensive amount of literature is available about the TNDFSP, including a number of recent reviews by Farahani, Miandoabchi, Szeto, & Rashidi [2013]; Guihaire & Hao [2008]; Ibarra-Rojas, Delgado, Giesen, & Muñoz [2015]; Kepaptsoglou & Karlaftis [2009]; Schöbel [2012]. These reviews discussed the methodology type, inputs, outputs, constraints, and general aspects of the TNDFSP. Although these reviews provided sufficient information on several TNDFSP studies, they lacked quality evaluation of the solutions (PT routes) produced by TNDFSP methodologies. The lack of quality evaluation is due to two reasons: 1) different networks were used for the experiments, in the related TNDFSP studies; 2) the environment settings that were used, such as different objectives, network constraints, and demand matrices, were variable. Fortunately, one benchmark network Mandl [1980] with same demand matrix, network configuration, and route set size was used in 27 TNDFSP studies. Some of these studies also used a set of large benchmark networks Mumford [2013] as well. These benchmark networks provide an excellent platform to conduct a fair evaluation and comparison of the solution qualities, produced by different solution methodologies. Additionally, by analysing their solution qualities, limitations of current solutions methodologies can be identified.

2.7.2 Solution quality evaluation of selected studies

The solution quality evaluation of all these 27 studies is performed by creating a quality-evaluation platform that enables a comparison of their solution qualities. The developed evaluation framework consists of an integration component between the criteria and objectives of the passengers, the agencies, and the authorities. More details about the evaluations can be found in Ul Abedin, Busch, Wang, Rau, & Du [2018]. The evaluation framework revealed the following:

- The proposed solutions could perform according to their specific objective functions for smaller networks but failed to show same quality for larger networks.

- The solutions' quality was not consistent with the variation in the network size and other inputs.
- Realistic demand assignment procedures should be used instead of simple procedures.
- Mode type should be considered an integral part of the TNDFSP, especially for large networks.
- For smaller networks, there were many comprehensive (performs good on all related indicators) solutions, but for larger networks, only a single solution showed partial comprehensiveness.

2.8 Summary

In this chapter, a brief overview about the TNDFSP along with the most commonly used objectives, decision variables and constraints was provided. Different approaches adopted by authors to deal with this problem were highlighted. A number of methodological procedures and their combinations were also explained. Demand assignment, which is one of the core parts of TNDFSP, was also briefly discussed. Apart from that, commonly used performance indicators used for PT network evaluation were also mentioned.

The solution quality evaluation was conducted on a number of selected studies in a systematic way to get some insights of the TNDFSP methodologies. The results revealed that the current methodologies cannot cope up with the changes in the network instances, and the quality of the generated solutions is rather inconsistent. Furthermore, lack of multiple modes consideration and use of basic assignment procedure is oversimplification of the problem. Additionally, the objective functions used for the agency and passenger cost are not realistic.

In this thesis, the proposed methodology will consider the above-mentioned limitations. The focus will be on using realistic objective functions, multimodality in PT network design and capacity constrained demand assignment. Moreover, a holistic evaluation compromised of the different viewpoints will be used for PT network evaluation. Apart from that, the consistency of the PT networks for different network variations will be monitored to guarantee consistent quality PT networks.

The next chapter will focus on problem definition and model formulation. First, the graph definitions and concepts are introduced. Second, the notations used in this thesis are listed and third, the proposed objective functions, constraints and assumptions are explained.

3 Problem Definition and Model Formulation

3.1 Graph and networks

A graph $G=\{N,A\}$ consists of a finite set points called nodes N and a set of unordered pair of points taken from N called arcs A (see Fig 3.1). Each arch is a line joining a pair of nodes.

$$G=(N,A) \tag{3.1}$$

$$N=\{n_0,n_1,n_2,n_3,n_4,n_5,n_6,n_7\} \tag{3.2}$$

$$A=\{(n_0,n_2),(n_1,n_3),(n_2,n_3),(n_2,n_4),(n_3,n_5),(n_4,n_5),(n_4,n_6),(n_4,n_7)\} \tag{3.3}$$

A network is a graph with additional information such as numbers assigned to nodes represent stops, junctions and traffic source. The numbers assigned to arcs represents costs, distances, time and capacity. In the TNDFSP context, nodes represent stops and arcs represent sequence stop pairs, usually called links.

A route r is a progressive path starts from a terminal and terminated at a certain node while traversing given arcs in sequence.

$$\begin{aligned} r_1 &= \{(n_0,n_2),(n_2,n_4),(n_4,n_7)\} \\ r_2 &= \{(n_1,n_3),(n_3,n_5),(n_5,n_4),(n_4,n_6)\} \end{aligned} \tag{3.4}$$

$$R=(r_i,r_{i+1},r_{n-1},r_n) \tag{3.5}$$

$$\begin{aligned} N_{r_1} &= (n_0,n_2,n_4,n_7) \\ N_{r_2} &= (n_1,n_3,n_5,n_4,n_6) \end{aligned} \tag{3.6}$$

A path p is an arc progression form in which all arcs are different.

$$p_{n_1 n_4}=\{(n_1,n_3),(n_3,n_5),(n_5,n_4)\} \tag{3.7}$$

$$P=(p_i,p_{i+1},p_{n-1},p_n) \tag{3.8}$$

$$N_{p1,4}=(n_1,n_3,n_5,n_4) \tag{3.9}$$

A transfer path tp is a progressive path which uses at least two routes.

$$tp_{n_0 n_6}=\left(\{(n_0,n_2),(n_2,n_4)\},\{(n_4,n_6)\}\right) \tag{3.10}$$

$$TP=(tp_i,tp_{i+1},tp_{n-1},tp_n) \tag{3.11}$$

$$N_{tp0,6} = (n_0, n_2, n_4, n_6) \tag{3.12}$$

A shortest path sp is a progressive path between two nodes which uses least number of arcs

$$sp_{n_1 n_6} = \{(n_1, n_3), (n_2, n_4), (n_4, n_6)\} \tag{3.13}$$

$$SP = (sp_i, sp_{i+1}, sp_{n-1}, sp_n) \tag{3.14}$$

$$N_{sp1,6} = (n_1, n_3, n_2, n_4, n_6) \tag{3.15}$$

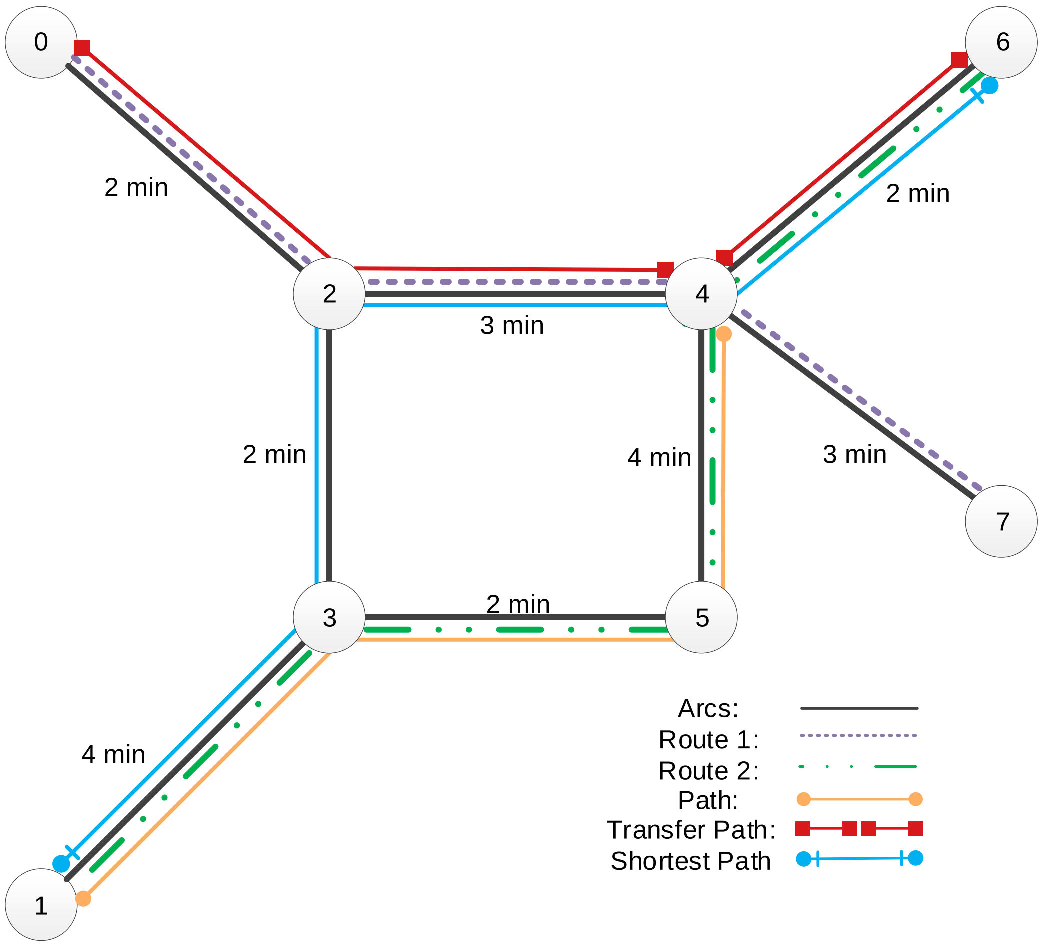


Fig 3.1 Sample transit network

3.2 TNDFSP

The TNDFSP involves determining a PT network configuration with a set of routes and their associated frequencies, to achieve certain objective/s, by adhering to given constraints.

3.3 Formulation

Consider a connected network composed of an undirected graph with a finite number of nodes, connected by arcs.

3.3.1 Notations

G	A network with N nodes and A arcs
N	Set of nodes in the network.
A	Set of arcs in the network.
r	A progressive path starts from a terminal and terminated at a certain node while traversing given arcs in sequence is called route.
p	A progressive path starts from the origin node and terminated at destination node, on same route while traversing given arcs in sequence is called path
tp	A progressive path which uses at least two routes is called transfer path.
sp	A progressive path between two nodes which uses least number of arcs.
R	Set of transit routes; $R=(r)$
PR	Set of potential transit routes; $PR=(pr)$
P	Set of paths; $P=(p)$
TP	Set of transfer paths; $TP=(tp)$
SP	Set of shortest paths; $SP=(sp)$
N_r	Set of nodes located on route r
N_{pr}	Set of nodes located on potential route pr
$N_{\max}$	Maximum number of nodes allowed on route r
$N_{\min}$	Minimum number of nodes allowed on route r
N_p	Set of nodes located on path p
N_{tp}	Set of nodes located on transfer path tp
N_{sp}	Set of nodes located on shortest path sp
A_r	Set of arcs located on route r
A_p	Set of arcs located on path p
A_{tp}	Set of arcs located on transfer path tp
A_{sp}	Set of arcs located on shortest path sp
TNA	Set of arcs in a transit network
d_{ij}	Passenger demand between node i and node j
d_{ij}^{r}	Passenger demand between node i and node j on route r ; $i, j \in N$
d_{ij}^{p}	Passenger demand between node i and node j on path p; $i, j \in N$
d_{ij}^{tp}	Passenger demand between node i and node j along the transfer path tp
d_{ij}^{sp}	Passenger demand between node i and node j along the shortest path sp
t_{ij}^{r}	Travel time from node i to node j on route r
t_{ij}^{p}	Travel time between node i and node j on path p
t_{ij}^{tp}	Travel time between node i and node j on transfer path tp

t_{ij}^{sp}	Travel time between node i and node j on shortest path sp
T_r	Travel time on route r from start to end.
L_r	Length of route r from start to end
TL	Total network length; $TL = \sum_{r\in R} L_r - \sum_{a\in TNA} (k-1)\cdot L_a^k$
L_a^k	Length of the arc within the transit network which occurs k times
a_{tr}^{r}	$\begin{cases} 1, & \text{transfer tr moves through route } r \\ 0, & \text{otherwise} \end{cases}$
F_r	Vehicle frequency associated with route r ; $F_r = \frac{MLS_r}{d_0\cdot C_v}$
MLS_r	Maximum passenger load on a route r
d_0	Desired occupancy on each vehicle; $d_0 = C_v.\delta$
δ	Load factor; $0 \le \delta \le 1$
C_v	Capacity of the vehicle
$F_{\min}$	Minimum frequency required for routes
$F_{\max}$	Maximum frequency required for routes
W_r	Passenger waiting time on route r ; $W_r = \frac{1}{2\cdot F_r}$
T_v	Type of the vehicle
S_v	Travel speed of the vehicle
CT_r	Cycle time of route r ; $CT_r = RT_r(1+\gamma)$
RT_r	Round trip time on route r ; $RT_r = 2T_r$
RL_r	Round length of route r; $RL_r = 2L_r$
γ	Terminal time coefficient
N_r^v	Number of vehicles required for given route r ; $N_r^v = \left[\frac{F_r\cdot CT_r}{60}\right]^+$
tr_{ij}^{tp}	Number of transfers involved in a transfer path from node i to node j
$tr_{\max}$	Maximum number of transfers allowed for transfer path
$A_{pr,r}^{c}$	Common arcs of a potential route pr that it shared with the transit route r ; $A_{pr,r}^{c} = A_{pr} \cap A_r$;
$ROR_{A_{pr}}^{A_{pr,r}^{c}}$	Route overlap ratio is the ratio of the common arcs of $A_{pr,r}^{c}$ to the total arcs of the potential route pr
β	Maximum allowed overlap percentage between pr and r
C_r^{vhr}	Vehicle operating cost per hour
C_r^{vkm}	Vehicle cost per kilometre
C_r^{mkm}	Route maintenance cost per kilometre
C_r^{ocp}	Overhead operating cost as percentage mark-up on all other costs
φ	Maximum allowed deviation from shortest path
μ	Penalty for transfer in minutes
ε_r^{ij}	Disutility equivalent to the in-vehicle time

Vol_r^{ij}	Passenger volume on the stop section ij of the route r
Cap_r^{ij}	Offered supply capacity on the stop section ij of the route r

3.3.2 Objective functions

The objectives functions that most of the studies considered in solving TNDFSP are rather simple and do not show true viewpoints of the different stakeholders. Therefore, the selected objective functions must be realistic and must represent the viewpoints of the involved stakeholders. Thus, two set of objectives functions are used in this thesis: 1) the first one (Z_1) represents the passenger costs such as waiting time, in-vehicle time, transfer waiting time and transfer penalty (as a proxy for the inconvenience caused during transfers); 2) the second (Z_2) represents the operating costs that an agency bears to offer a required service. This includes crew cost, fuel cost, infrastructure maintenance cost and other overhead costs.

The objective is to minimise the passenger cost and agency cost. The objective functions are as follows:

$$\min Z_1 = \frac{1}{\sum\limits_{i,j\in N} d_{ij}} \left(\sum_{r\in R} \sum_{i,j\in N_r} d_{ij}^r t_{ij}^r + \sum_{tp\in TP} \sum_{i,j\in N_{tp}} d_{ij}^{tp} t_{ij}^{tp} + \sum_{r\in R} \frac{1}{2F_r} \left(\sum_{i,j\in N_r} d_{ij}^r + \sum_{i,j\in N_{tp}} d_{ij}^{tp} a_{tr}^r \right) + \sum_{tp\in TP} \sum_{i,j\in N_{tp}} d_{ij}^{tp} tr_{ij}^{tp} \mu \right) \tag{3.16}$$

$$\min \mathrm{Z}_2 = \sum_{r\in R} \left\{ \left[(N_r^v \cdot C_r^{vhr}) + (2F_r . L_r . C_r^{vkm}) + (RL_r \cdot C_r^{mkm}) \right] . (1 + C_r^{ocp}) \right\} \tag{3.17}$$

Subject to

$$F_{\min} \le F_r \le F_{\max} \quad \forall r \in R \tag{3.18}$$

$$N_{\min} \le N_r \le N_{\max} \qquad \forall r \in R \tag{3.19}$$

$$\bigcup_{r\in R} N_r = N \tag{3.20}$$

$$r_i \cup r_j \ne \varnothing \qquad \forall r_i \in R,\ \exists r_j \in R \tag{3.21}$$

$$r_i \ne r_j \quad r_i, r_j \in R \tag{3.22}$$

$$t_{ij}^r \le (1+\varphi) t_{ij}^{sp} \quad \forall r \in R, \forall ij \in N_r \tag{3.23}$$

The first term of the first object is to minimise the travelling cost for the passenger, it includes the travel cost on direct routes as well as on transfer paths. The second term's objective is to minimise the passenger waiting time. This includes origin waiting time and transfer waiting time. The third term is the transfer penalty for transfer paths. All these terms strictly represent the viewpoint of the passenger. The second objective is to minimise the total operating cost for the agency for given set of routes. The first term is the direct vehicle operating cost per hour for the given route. This cost includes salaries of the drivers and onboard vehicle crew. The second term is the cost per vehicle kilometre, this includes fuel and maintenance cost. The third term includes the infrastructure maintenance cost such as track maintenance, right-of-way (ROW) and signalling cost. The last term includes the overhead operating cost, such as scheduling, duty rostering, supervision and depot related costs.

3.3.3 Constraints

There are six constraints that were considered for TNDFSP. Constraint (3.18) ensures the frequencies for the routes stay within the defined range. Constraint (3.19) defines the range of the number minimum and maximum stops allowed in a route. Constraint (3.20) guarantees all stops are served by at least one route and none of the stops is left vacant. Constraint (3.21) assures that the transit network is connected and there is always a path for any given OD pair. Constraint (3.22) assures that there are no duplicate routes in the transit network. Constraint (3.23) limits the path deviation of any given route to its shortest path.

3.3.4 Assumptions

For the sake of simplicity and brevity of the study, there are number of assumptions considered in this study. These are as follows:

- The transit demand is considered as stops level demand.
- The transit demand matrix is fixed and symmetric.
- The in-vehicle time is constant and traffic conditions are not considered.
- The transfer walk time from/to stop is not considered.

3.4 Summary

This chapter focused on the problem definition and model formulation for TNDFSP. The selected objectives were formulated as multi-objective optimisation problem with different cost components for passengers and agencies. Moreover, considered constraints and assumptions used in the optimisation problem were also listed in this chapter.

The next chapter will introduce the proposed approach to solve the TNDFSP. The approach is based on a novel concept of route-overlap ratio, for PT network creation. Moreover, NSGA-II is used as the main search engine for the exploration of feasible search space. The methodology is composed of four components starting from: 1) transit route and network creation; 2) transit demand assignment; 4) transit network evaluation; 5) metaheuristic search engine.

4 An NSGA-II Based Approach to TNDFSP Using Route Overlap Ratio Concept

This chapter focuses on route overlap ratio (ROR) based network design methodology, which aims at solving large urban transit network instances. The proposed methodology is composed of heuristic and metaheuristic techniques. It consists of four main components: 1) transit route and network construction (TRNC) which at first, generates all the feasible routes. Later, a subset of these routes is used to create a transit network; 2) transit demand assignment and headway derivation (TDAHD) is used to assign the demand and to generate travel time components and headways for the routes; 3) transit network evaluation (TNE) is used to holistically evaluate the created transit network, considering the viewpoints of passengers, agencies and authorities; 4) NSGA-II based metaheuristic search engine (NSGA-II MSE) is used to explore the search space in order to find a set of pareto optimal solutions.

The preference of metaheuristic-based approach over exact methods to solve TNDFSP is based on: 1) metaheuristic techniques' high flexibility and applicability to several combinatorial optimisation problems; 2) its capability of tackling large network instances with satisfactory solution quality within acceptable time. In the proposed methodology, the central idea is to generate a set of feasible transit routes by incorporating shortest path algorithms. Once all the feasible routes are generated, ROR-based approach is used to create feasible PT networks. These networks are then subjected to demand assignment for calculating travel time components, route headways and mode types. Later, these PT networks are evaluated based on different criteria representing the viewpoints of all the stakeholders (passengers, agency and authority). Finally, NSGA-II is employed to create and search for the optimal set of PT networks for selected objective functions. A schematic overview of the proposed methodology is depicted in Fig 4.1 where, blue colour modules represent contributions from other authors/studies.

The key features of the proposed methodology are as follows:

- The solution approach considers a ROR-based TRNC. This concept can provide a whole spectrum of different PT network configurations.
- Realistic objectives for passenger and agency costs are considered.
- NSGA-II is used to search pareto optimal solutions for the selected objectives.
- The solution approach utilises all the feasible modes and their associated properties to provide a heterogenous PT network.
- Headway-based capacity constrained demand assignment is used to derive more realistic travel time components and route loads.
- Holistic evaluation of the transit network considering all three stakeholders' viewpoints is performed.

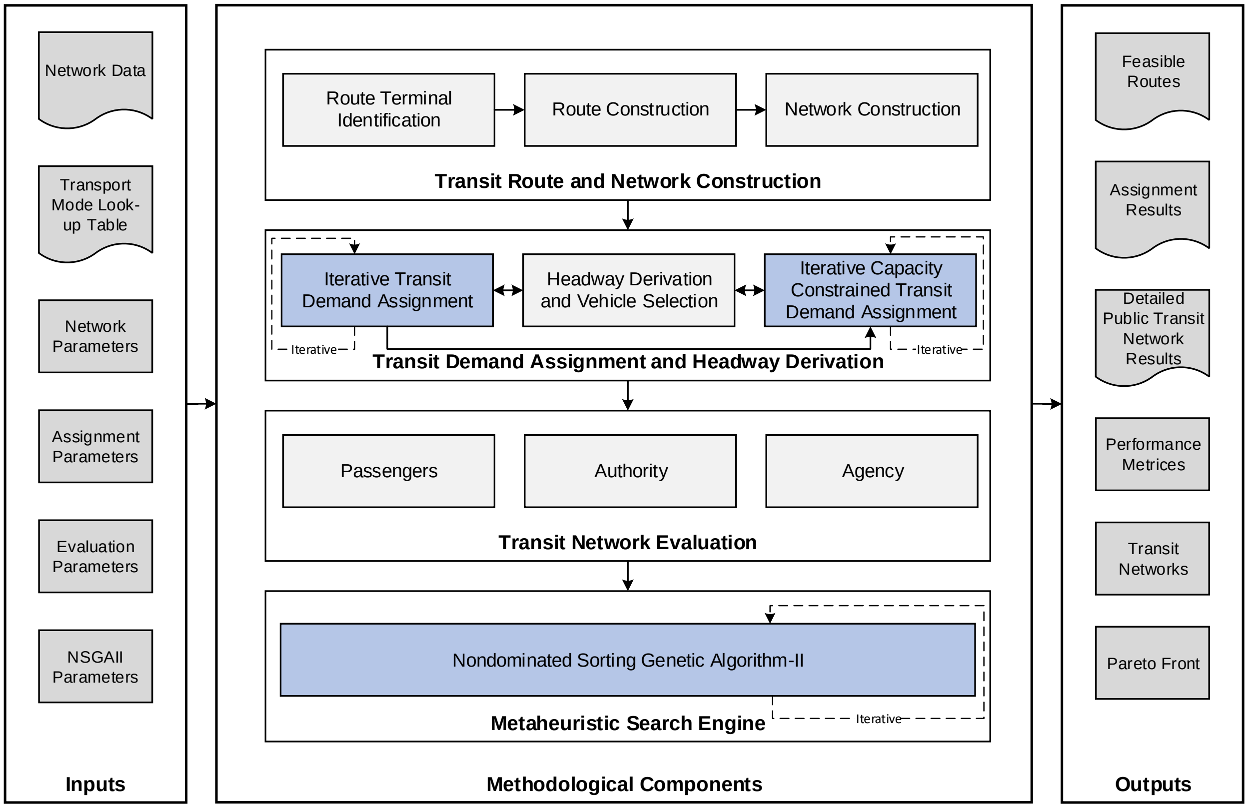


Fig 4.1 Schematic overview of the proposed transit network design concept

4.1 Transit route and network construction

In TRNC component first, the route terminals are identified. Second, once the terminals are identified, route construction (RC) module is used to create the feasible transit route set (FTRS), these routes are created by employing shortest paths algorithm starting and/or ending at the identified terminals. Third, the network construction (NC) module is used to construct the transit network by selecting and inserting routes from the created FTRS. Fourth, transit network improvement (TNI) and transit network feasibility (TNF) procedures are employed for transit network improvement and feasibility check respectively. A schematic overview of TRNC component is illustrated in Fig 4.2.

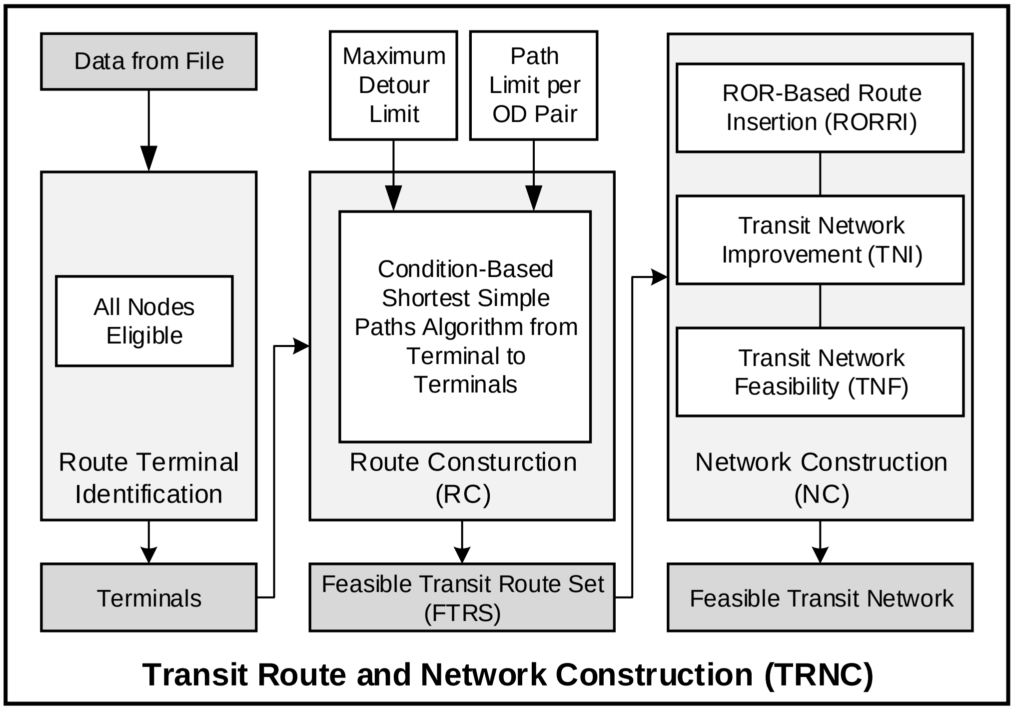


Fig 4.2 Schematic overview of transit route and network construction component

4.1.1 Route terminal identification

The terminal act as the starting point of a route and in many cases, it also acts as an ending point of the route. The terminals can either be predefined or based on the high demand generation and culmination node pairs. Identification of terminals is extremely complicated exercise and several factors need to be taken into consideration. In real world, such factors include the surrounding land use, availability of the land, connectivity with the existing street network and its density, existing transit routes and demand flow patterns in the surrounding area, other terminals in the surrounding area, future demand and land use projections and most importantly the cost of building a terminal. Mostly, the decision of identification, selection and construction is taken by the PT authorities, backed by the governments.

In case of artificial benchmark networks, which are often derived from the real networks. It is one of the inputs that are already given. However, in case such information is not given, it is not considered as part of network design. If the terminals are not given, in such cases two approaches are followed: 1) every node in the network can act as a terminal; 2) The highest demand node pairs are selected to act as terminals. In this study, if the terminal data is available, the given terminals will be used. In cases, where no such data is available, the first approach is followed.

4.1.2 Route construction

The routes are created by employing RC module. There are many shortest path creation algorithms available, such as Dijkstra algorithm (Dijkstra [1959]), Floyd-Warshall algorithm (Floyd [1962]), A* search algorithm (Hart, Nilsson, & Raphael [1968]) and Yen k-shortest path algorithm

(Yen [1971]). In the proposed RC module, condition-based shortest simple paths algorithm adopted from Yen [1971] is employed to create feasible routes.

4.1.2.1 *Condition-based shortest simple paths algorithm from terminal to terminals*

A condition-based shortest simple paths algorithm is employed to generate feasible routes. The algorithm is fast on small sparse networks, but it quickly leads combinatorial explosion for large networks. Therefore, to reduce the computation cost, the procedure sets two conditions on the path finding algorithm: 1) a path between a node pair is only accepted if its cost is less than or equal to the maximum detour limit; 2) the maximum number of paths per node pair is set to predefined values e.g., 2000 paths.

The procedure used to create routes between terminal pairs is as follows:

1. The procedure starts by creating simple paths between a pair of terminals.
2. For each route insertion into FTRS, two conditions are checked:
    - The maximum detour path limit check, if it is within the limit, route is saved into the FTRS, or else it is discarded.
    - The maximum number of routes for a given pair of terminals must be less than predefined value.

After employing this procedure, all the feasible routes are created, and inserted into FTRS. The main procedure used to create feasible routes is given in Alg 4.1.

```
Condition-based shortest simple paths algorithm
Input G (Network graph), KPaths (maximum number of routes between two nodes), Detour
(maximum detour limit from shortest path), Source (starting node), Sink (destination node),
Infinity (very large number)
FeasiblePaths = GenerateArray()
K = GenerateArray(Infinity)
Paths = GenerateArray()
PP = GenerateArray()
SP = CalculateShortestPathDijkstra(G, Source, Sink)
Paths[0] = AssignPath(SP)
FeasiblePaths = AssignPath(SP)
FeasiblepathsRequired = True
For each k in K[1: ]:
    For each index in GetLength(Paths[k-1]):
        RemovedEdges = GenerateArray()
        RemovedPathNodes = GnerateArray()
        SpurNode = Paths[k-1][index]
        RootPath = Paths[k-1][Source: index])
        For each path in Paths:
            IF RootPath==path[Source, index]:
                RemovedEdges = RemoveEdgeFromGraph(G, index, index+1)
            End IF
        For each rootpathnode in RootPath[:-1]:
            RemovedPathNodes = RemoveRootPathNodeFromGraph(G, rootpathnode)
        SpurPath = CalculateShortestPathDijkstra(G, SpurNode, Sink)
        TotalPath = RootPath + SpurPath
        PP = AddPath(TotalPath)
        AddEdgesToGraph(G, RemovedEdges)
        AddRootPathNodesToGraph(G, RemovedPathNodes)
    IF Not PP:
        Break
    Sort(PP)
    Paths[k] = PP[0]
    For potentialpath in PP:
        IF potentialpath <= SP * Detour:
            FeasiblePaths = FeasiblePaths(potentialpath)
        End IF
        IF len(FeasiblePaths)==Kpaths:
            FeasiblepathsRequired = False
            Break
        End IF
    IF Not FeasiblepathsRequired:
        Break
    End IF
Output FeasiblePaths (A set of feasible paths for given node pair)
```

Alg 4.1 Condition-based shortest simple paths procedure

4.1.3 Network construction

In Network Construction (NC) module, three procedures are used, the first two are used to create and improve the transit network, whereas the third one is used to check the feasibility of the transit network. These are: 1) ROR-based route insertion (RORRI); 2) TNI; 3) TNF. At first, a single route is inserted into the transit network. Later, only ROR complaint routes are inserted into transit network. The created transit network is then subjected to improvement procedures. Finally, the feasibility of the transit network is checked.

4.1.3.1 *Route overlap ratio-based route insertion*

The RORRI procedure is used to insert the routes into the transit network. In the beginning, a route at random is drawn from the FTRS, and inserted into transit network, which is empty at this point. After the successful insertion of the first route, the FTRS is sorted according to the route length. Now the second route is drawn from the sorted FTRS, and it is compared with first route to check whether the second route shares significant path with the first route or not. This is done by checking ROR between the new route and the routes present in the network. The insertion of new routes continues as long as the drawn routes' ROR values are within the predetermined range.

A ROR is simply the common route arcs that a potential route shares with the rest of the routes inside the transit network. The term $A^c_{pr,r}$ represent the common arcs of the potential route pr that it shares with the transit route r. The term $ROR^{A^c_{pr,r}}_{A_{pr}}$, is the ratio of the number of the common arcs to the total number of arcs in a potential transit route pr. The pr is checked against all the transit routes R for ROR. The pr will only be accepted if its value is less than the β for all transit routes.

$$A^c_{pr,r} = A_{pr} \bigcap A_r;\ \forall r \in R, \forall pr \in PR \tag{4.1}$$

$$ROR^{A^c_{pr,r}}_{A_{pr}} = \frac{A^c_{pr,r}}{A_{pr}}; \qquad ROR \le \beta, \forall r \in R, \forall pr \in PR \tag{4.2}$$

To understand the concept, let's take a simple road network example depicted in Figure 1. Now assume that there are four feasible routes r_1, r_2, r_3 and r_4 with route arcs A_{r1}, A_{r2}, A_{r3} and A_{r4} and the β value of 0.3. Moreover, it is assumed that no routes are present in the network. Apart from that, for this example the routes will not be sorted according to length.

$$A_{r1} = \{(n_0,n_2),(n_2,n_4),(n_4,n_7)\} \tag{4.3}$$

$$A_{r2} = \{(n_1,n_3),(n_3,n_2),(n_2,n_4),(n_4,n_6)\} \tag{4.4}$$

$$A_{r3} = \{(n_0,n_2),(n_2,n_4),(n_4,n_6)\} \tag{4.5}$$

$$A_{r4} = \{(n_1,n_3),(n_3,n_5),(n_5,n_4),(n_4,n_7)\} \tag{4.6}$$

Let's say, a random route, r_1 with arcs A_{r1} is selected and inserted in the network. Since, it is the first route to be inserted therefore, no comparison is required. For the second route r_2 with arcs A_{r2}, it will become a potential route pr and will be checked against transit routes R, which in this case is only r_1. There is only one common arc between r_1 and r_2 with the ROR value of 0.25. This is acceptable for inserting into the network.

$$A^c_{pr,r1} = \left((n_2, n_4)\right) \tag{4.7}$$

$$ROR^{A^c_{pr,r1}}_{A_{pr}} = 0.25 \tag{4.8}$$

Now there are two routes in the transit network and the third route r_3 becomes the potential route pr and the same procedure will be used. However, the potential route is checked against two routes $r_1 \text{ and } r_2$. In both cases, the ROR is 0.66 and 0.66 respectively, which is above the threshold value of β. Therefore, r_3 will be discarded and it will not be inserted into the transit network.

$$A^c_{pr,r_1} = \left((n_0, n_2), (n_2, n_4)\right), A^c_{pr,r_2} = \left((n_2, n_4), (n_4, n_6)\right) \tag{4.9}$$

$$ROR^{A^c_{pr,r_1}}_{A_{pr}} = 0.66, \qquad ROR^{A^c_{pr,r2}}_{A_{pr}} = 0.66 \tag{4.10}$$

At the end, the last route r_4 becomes the potential route and same procedure is repeated. The route r_4 is checked against $r_1\, and\, r_2$ and the value of ROR is 0.25 and 0.25 respectively. Therefore, r_4 will be inserted in the transit network. Moreover, with the insertion of this route, all of the nodes in the network are covered by the transit routes. It has now become feasible (see section 4.1.3.3), with no more insertion required.

$$A^c_{pr,r_1} = \left((n_4, n_7)\right), \qquad A^c_{pr,r_2} = \left((n_1, n_3)\right) \tag{4.11}$$

$$ROR^{A^c_{pr,r_1}}_{A_{pr}} = 0.25, \qquad ROR^{A^c_{pr,r2}}_{A_{pr}} = 0.25 \tag{4.12}$$

The flowchart (see Fig 4.3) of ROR-based route insertion is listed below:

1. At first, the transit network route container (TNRC) is checked. If it is empty, a random route from FTRS is drawn and placed (InsertRoute) directly into the transit route container. Once the first route is removed from FTRS, the rest of the routes are sorted (Sort) according to their lengths in descending order.
2. The second route is selected from FTRS, and it is compared (RORInsertion) with the rest of the routes in the TNRC to check, whether the new route exceeds route overlap ratio (β) or not. If it does, it is discarded; else it is inserted into TNRC.

3. In such manner, the routes from the FTRS are evaluated. This process stops if one of the two conditions are satisfied:
    a. The TNF (CheckRouteSetFeasibility) is satisfied
    b. All the routes in FTRS are evaluated

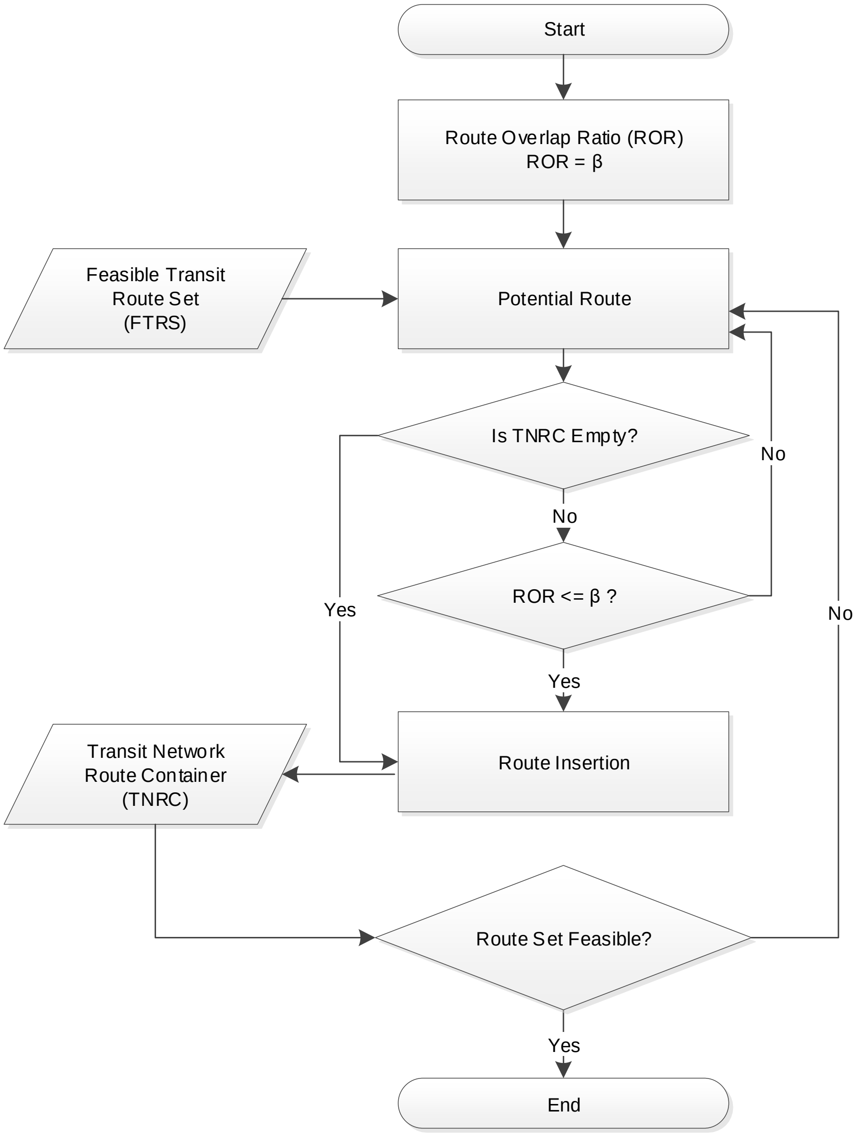


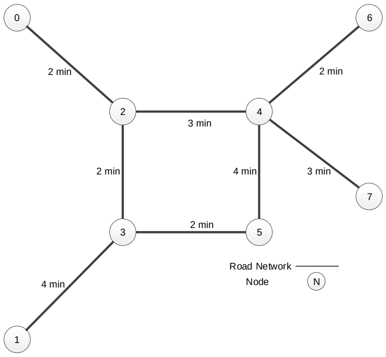
Fig 4.3 Flow chart of route insertion based on route overlap ratio concept

The basic procedure of ROR-based route insertion is listed in Alg 4.2.

```
Route overlap ratio-based route insertion algorithm
Input FTRS (feasible transit route set), ROR (route overlap ratio), Incr (ROR increment),
TNRC (transit network route container)
Size=calculateSize(FTRS)
IF Not TNRC:
    RN = GenerateRandomNumber(Size);
    TNRC = InsertRoute(FTRS[RN])
    Sort(FTRS)
End IF
TNRCComplete = False
While Not TNRCComplete:
    For each r in FTRS:
        RORInsertion(r, RouteSet, ROR)
        IF CheckRouteSetFeasibility(TNRC):
            TNRCComplete = True
            Break
        End IF
    IF Not TNRCComplete:
        TNI(TNRC)
        IF CheckRouteSetFeasibility(TNRC):
            TNRCComplete = True
        ELSE:
            ROR+=Incr
        End IF
    End IF
End While
Output TNRC (A ROR complaint feasible route set)
```

Alg 4.2 Route overlap ratio-based route insertion algorithm

Let’s assume a small road network, depicted in Fig 4.4a and the demand matrix, presented in Fig 4.4b. Now let’s consider a small set of examples (see Fig 4.5) for different types of PT networks by using route overlap concept. The considered assumptions in the example are: 1) fleet size is set to 10 buses with a capacity of 80 person/bus; 2) the speed levels are set to 15km/h for normal bus service and 22km/h for trunk bus service.



(a) Base road network with travel times

	0	1	2	3	4	5	6	7
0	0	100	50	75	75	25	25	15
1	100	0	25	100	50	50	75	75
2	50	25	0	75	25	100	75	15
3	75	100	75	0	75	75	50	60
4	75	50	25	75	0	50	100	25
5	25	50	100	75	50	0	50	25
6	25	75	75	50	100	50	0	75
7	15	75	15	60	25	25	75	0

(b) Demand matrix

Fig 4.4 Base road network and demand matrix

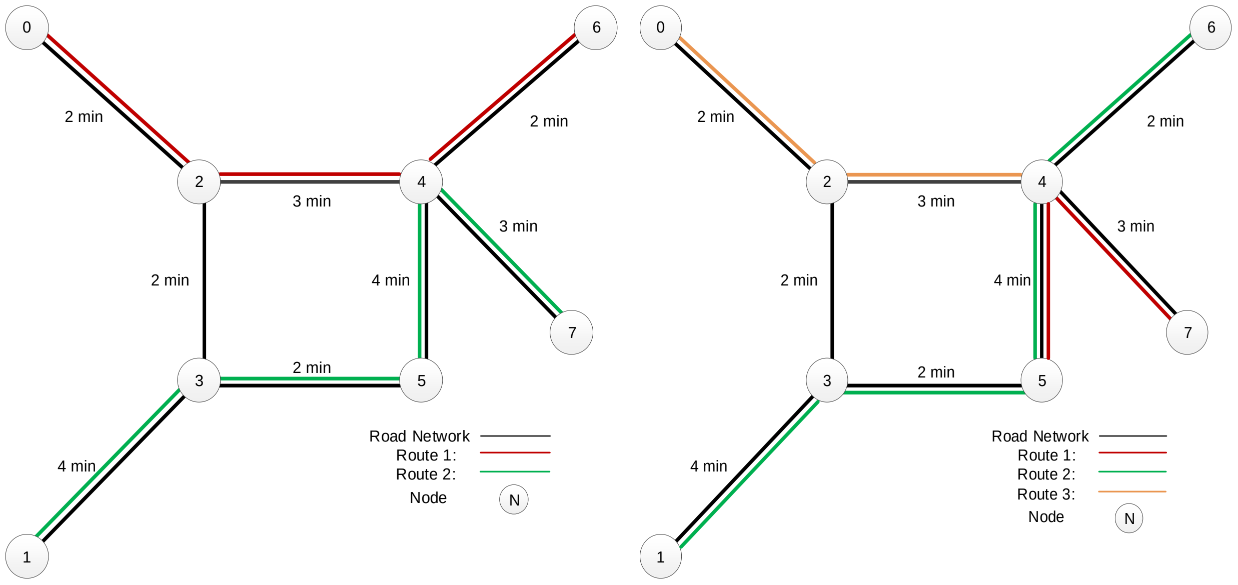
The main idea for these illustrations is to show: 1) the capability of the ROR-based route insertion procedure in creating different PT networks configurations; 2) the properties of different PT networks; 3) the performance evaluation of different PT networks (see section 4.3 for PT evaluation measures).

In terms of ROR-based PT network creation. Three networks are created with ROR-based route insertion with different values of route overlap ratios. The first PT network is created with ROR value of 0%, called the minimum route overlap network (see Fig 4.5a). The second PT network is created with a maximum ROR value of 25%, called the partial route overlap network (see Fig 4.5b). The third PT network is created with ROR value of 100%, called the maximum route overlap network (see Fig 4.5c). These three types cover a wide variety of PT networks configurations.

The minimum route overlap resembles most of the MRT networks, where the objective is to provide fast, uninterrupted and smooth travel experience along with minimum route overlap. The objective here is to provide connections among the routes and not to run parallel competing services. Such networks are considered as the backbone of PT networks, where the bulk of demand moves through, especially during peak demand periods.

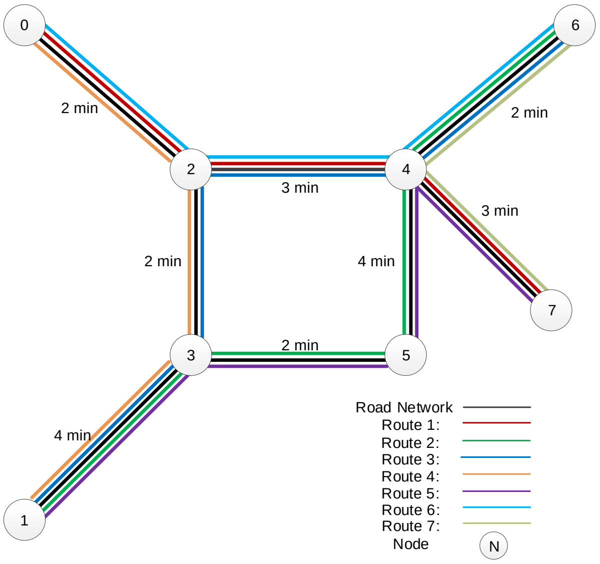
In the partial route overlap concept, this is a typical trunk-feeder service. A trunk line offers fast and high frequency service with right of way category A/B. Whereas, feeders feed the demand to the trunk route to maximise its utilisation, and to achieve a network effect. In such a concept, seamless transfers play an important role in enhancing operational efficiency and viability of the transit system.

The last one is the maximum route overlap, it offers a direct service among different nodes with many redundant routes running in parallel. Most bus-based PT networks in the cities are similar to such network. The reason lies in the historical evolution of the PT network. In most cases, the initial PT network is less redundant with fewer routes. However, with the passage of time and due to new developments and land use changes, more routes are introduced in the PT network.



(a) Minimum route overlap PT network

(b) Partial route overlap PT network



(c) Maximum route overlap PT network

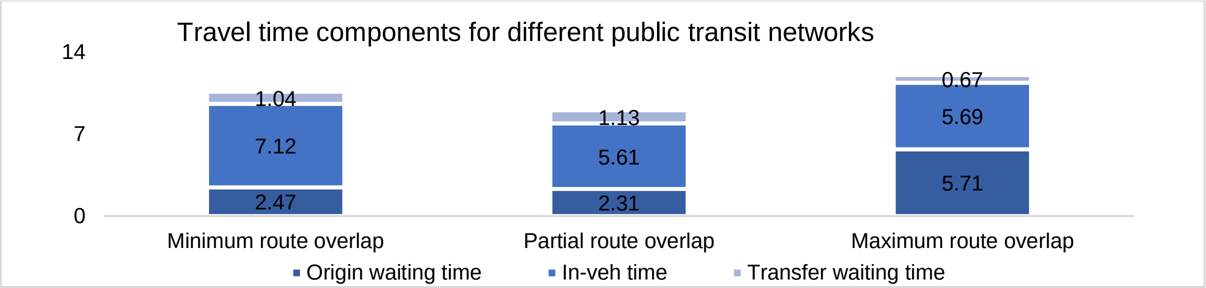
Fig 4.5 Different public transit networks

In terms of attributes, these three PT networks have certain characteristics. For instance, the minimum route overlap network is perhaps the simplest PT network, it is easier to manage, tend to offer high frequencies, and it’s easier for passengers to follow. The only, downside of such network is the transfers involved in the journeys. The partial route overlap network offers similar service to minimum route overlap one. However, with much lower total travel times and waiting times. The utilisation of such PT network can be maximised, given efficient transfer stations, single fare collection system and coordinated headways. The maximum route overlap network offers least amount of transfers to complete the trip. However, offering such service leads to redundant routes with low frequencies.

In terms of performance evaluation (see Fig 4.6), all these three PT networks are evaluated on set of indicators listed in the proposed TNE component (see section 4.3) except unsatisfied demand percentage. As, 100% demand was satisfied by all three PT networks. The travel time components are compared in Fig 4.6a, it shows that minimum and partial route overlap offers minimum waiting time, whereas maximum route overlap offers maximum waiting time. In terms of in-vehicle time, maximum route overlap and partial route overlap offers similar values. However, the reason for similar in-vehicle time is different for both, as for partial route overlap, it is achieved by faster in-vehicle time due to trunk route operation. But, for maximum route overlap, its simply because of direct routes for most of OD pairs. For minimum route overlap, the increased in-vehicle time is due to detours involved for some of the trips. The transfer waiting time is minimum for maximum route overlap network, as majority of the trips are served directly. For minimum and partial route overlap networks, the transfer waiting times are similar. The increase in transfer waiting time is simply because of transfers, required to complete the trips.

The listed performance indicators in Fig 4.6 suggests the following:

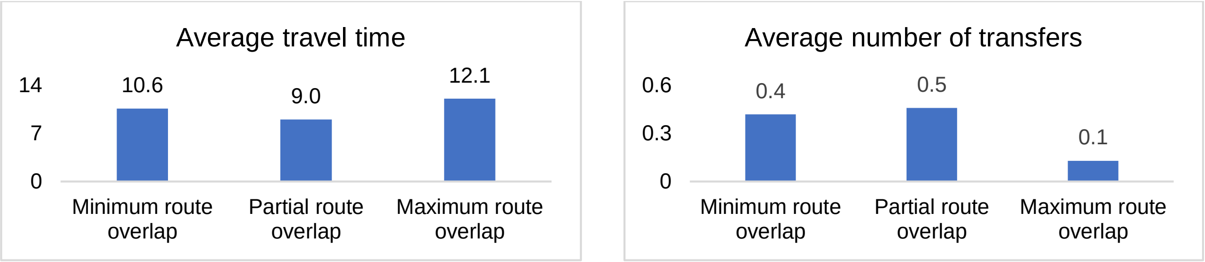
- The route overlap index (see Fig 4.6b) shows that the minimum route overlap has the minimum value whereas the maximum route overlap network has three times the minimum value.
- In terms of the directness of the network, all three networks have similar directness values (see Fig 4.6c).
- The average travel time is least for partial route overlap network and then minimum route overlap, and finally the maximum route overlap network (see Fig 4.6d). This shows that, transfer based network offers improved travel times compared to a redundant PT network with many overlapping routes.
- The maximum route overlap network offers least number of transfers compared to other two networks (see Fig 4.6e). However, the results from other indicators already showed that lesser number of transfers don’t translate into reduced travel times.
- In terms of operating cost, minimum and maximum route overlap, cost the same however, partial route overlap cost a fraction higher than the two (see Fig 4.6f). This is due to high operational costs associated with trunk route.
- The empty seat hours show mixed results, as minimum route overlap offers highest capacity utilisations, whereas partial and maximum route overlap shows lower capacity utilisation (see Fig 4.6g).
- In terms of required seating capacity, it is fixed and same for the all three networks (see Fig 4.6h).
- The minimum and partial route overlap networks require least transit network length whereas, maximum route overlap requires three times the length of other two networks (see Fig 4.6i).



(a) Travel time components different public transit networks in minutes

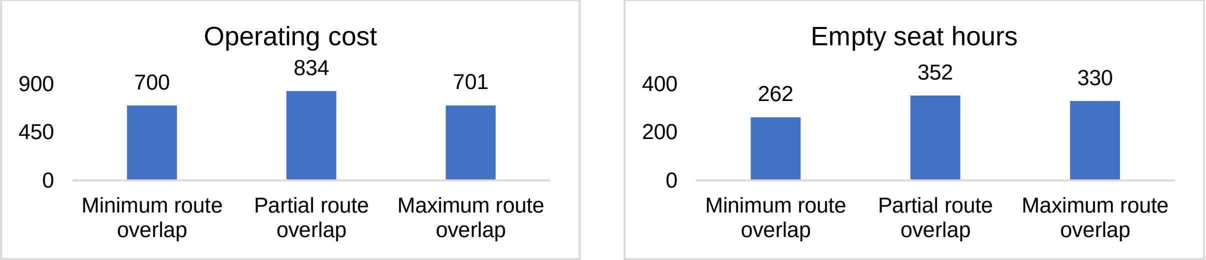
(b) Route overlap index

(c) Transit network directness



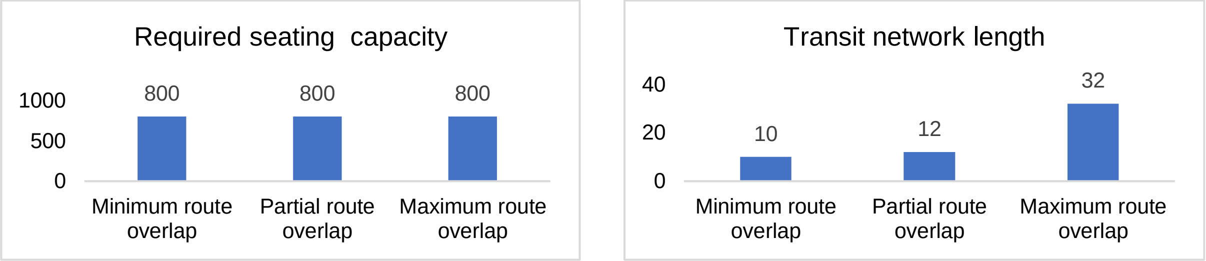
(d) Average travel time in minutes

(e) Average number of transfers



(f) Operating cost in $/hour

(g) Empty seat hours



(h) Required seating capacity

(i) Transit network length in km

Fig 4.6 Comparison of different PT network types based on selected performance measures

4.1.3.2 *Transit network improvement*

After the RORRI procedure, TNI procedure is used for PT network improvement. The TNI is an important procedure, its main job is to improve created transit network and to incorporate missing nodes into the transit network. It consists of two sub procedures: 1) transit route extension (TRE); 2) transit route node insertion (TRNI).

In the *transit route extension (TRE)* sub procedure, it is checked whether a route can be extended or not. This depends upon the presence of unserved stops in the neighbourhood of the last/first node of the route. The potential node is only inserted if the arc cost is less than or equal to the maximum deviation from the shortest path between the given node and the potential node. The TRE is applied on all the routes in the transit network.

1. The TRE sub procedure (see Alg 4.3) starts by checking the missing nodes (CheckMissingNodes) in the PT network (TNRC).
2. Later, for each missing node, the first and last node of every route in the TNRC is checked. Whether the neighbour nodes (CheckNeighbours) of the first node of the selected route includes the missing node (*PotentialNode*) or not.
3. In case, the node is present, shortest path (CalculateShortestPathDijkstra) between the first node and the missing node is calculated.
4. If the arc cost (*ArcCost*) between the first node and the missing node is within the allowed deviation from the shortest path. It is appended (InsertNode) at the start of the route and the node is removed from the missing nodes (RemoveMissingNode).
5. The same process is followed for the last node of the route as well.
6. The process of route extension stops when all the missing nodes are checked against all the routes in TNRC.

```
Transit route extension algorithm
Input TNRC (transit network route container), G (network graph), Detour (maximum detour
limit from shortest path)
MissingNodes = CheckMissingNodes(TNRC)
IF MissingNodes:
    For missingnode in MissingNodes:
        For each route in TNRC:
            Start = route[-1]
            End = route[1]
            Neighbours = CheckNeighbours(Start)
            PotentialNode = {missingnode U Neighbours}
            IF PotentialNode:
                SP = CalculateShortestPathDijkstra(Start, PotentialNode)
                ArcCost = G[Start, PotentialNode]
                IF ArcCost <= Detour * SP:
                    InsertNode(PotentialNode,route)
                    RemoveMissingNode(potentialnode, MissingNodes)
                    Break
                End IF
            End IF
            Neighbours = CheckNeighbours(End)
            PotentialNode = {missingnode U Neighbours}
            IF PotentialNode:
                SP = CalculateShortestPathDijkstra(End, PotentialNode)
                ArcCost = G[End, PotentialNode]
                IF ArcCost <= Detour * SP:
                    InsertNode(PotentialNode,route)
                    RemoveMissingNode(PotentialNode, MissingNodes)
                    Break
                End IF
            End IF
End IF
Output TNRC (TNRC after the TRE sub procedure)
```

Alg 4.3 Transit route extension procedure

A small example is depicted in the Fig 4.7, the potential nodes (dark colour) present in the vicinity of the last node of a route (Fig 4.7a). Based on the considered node, which is highlighted in light grey colour, the shortest path from last node to the considered node is calculated, which goes through node x with total cost as 4 min. The arc cost between last node and considered node is calculated, which is 5 min. It is within the allowed deviation (1.5 times shortest path), the node is selected, and the route structure is changed (Fig 4.7b).

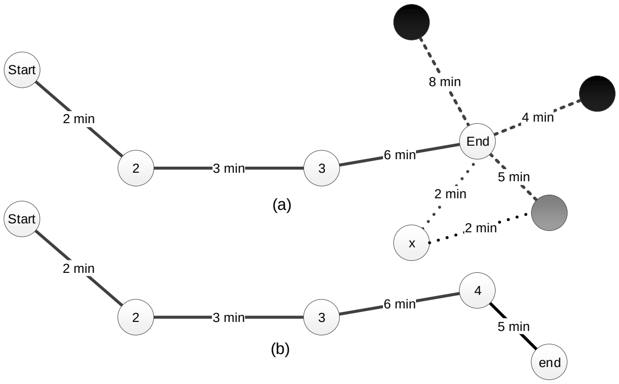


Fig 4.7 Transit route extension

In the ***transit route node insertion (TRNI)*** sub procedure, it is checked whether there exists an unserved node in the neighbourhood of the route node pairs (arcs) or not. If there is one, then the detour cost is check against the original cost. If the detour is within the detour limit, the node is inserted into the route. All routes in TNRC are subjected to TRNI procedure.

1. The TRNI sub procedure (see Alg 4.4) starts by checking the missing nodes (CheckMissingNodes) in the PT network.
2. If there are missing nodes present in the PT network, then for each missing node, each route and its node pairs are check, for the neighbours (CheckNeighbours).
3. If the missing node has a route node pair in its neighbourhood (*PairNodeExist*). In that case, the arc cost of the current node pair (*ArcCost*) is compared with the combined arc cost of both from node pair's first node to missing node and from missing node to node pair's second node.
4. If the value of combined arc cost (CalculateSum) is less than or equal to the maximum detour limit. The missing node is inserted (InsertNode) between the consecutive node pair, and it is removed (RemoveMissingNode) from missing nodes.
5. The process moves to the next missing node and continues until all missing nodes are evaluated.

```
Transit route node insertion algorithm
Input TNRC (transit network route container), G (network graph), Detour (maximum detour
limit from shortest path)
MissingNodes = CheckMissingNodes(TNRC)
IF MissingNodes:
    For missingnode in MissingNodes:
        MissingNodeServed = False
        For each route in TNRC:
            For each pair in route:
                Neighbours = CheckNeighbours(missigNode)
                PairNodeExist = {pair U Neighbours}
                IF PairNodeExist:
                    ArcCost = G[pair [1], pair [2]]
                    PotentailArcCost1 = G[pair[1], missingnode]
                    PotentailArcCost2 = G[missingnode, pair[2]]
                    TotalCost = CalculateSum(PotentialArcCost1, PotentialArcCost2)
                    IF TotalCost <= ArcCost * Detour:
                        InsertNode(missingnode,route)
                        RemoveMissingNode(missingnode, MissingNodes)
                        MissingNodeServed = True
                        Break
                    End IF
                End IF
            IF MissingNodeServed:
                Break
            End IF
End IF
Output TNRC (TNRC after the TRNI sub procedure)
```

Alg 4.4 Transit route node insertion procedure

A small example about TRNI is depicted in Fig 4.8. A route with two potential nodes near the stop pair (2,3) are shown in Fig 4.8a. Based on the calculation, the node which offers least detour, and which lies within the limits of maximum detour, will be selected for insertion. In this example, the bottom node is selected as the detour is within the defined range (1.5 times shortest path), and the route structure is changed after new node insertion (see Fig 4.8b).

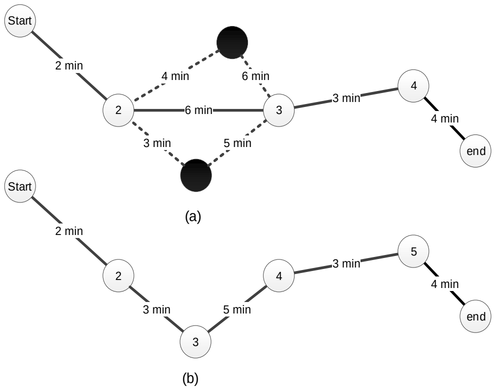


Fig 4.8 Transit route node insertion

4.1.3.3 *Transit network feasibility*

The TNF procedure checks whether a PT network (TNRC) is feasible or not. Four conditions need to be satisfied to consider a transit network as feasible: 1) There is no duplication of node in a route; 2) There are no duplicate routes in the transit network; 3) all the nodes should be covered in the transit network; 4) all the routes are connected in such a way that there is always a transit path between a given OD pair.

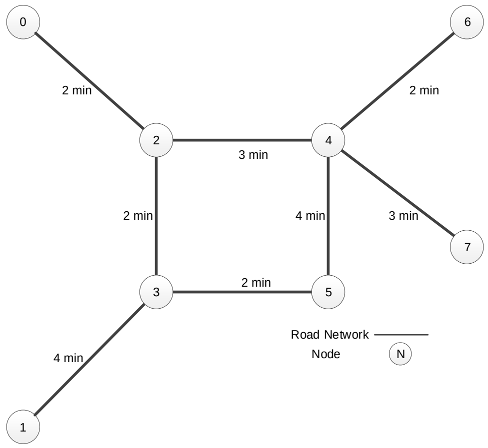
A *node duplication* means that a route sequence cannot repeat a node twice. Although, in real world there exist many loop routes. However, due to algorithmic limitation, loop routes or node duplication is not permitted.

A *route duplication* means each route in the transit network is unique. In the transit network, no two routes can have the same node sequence. For any given two routes r_i and r_j, they must hold $r_i \neq r_j$.

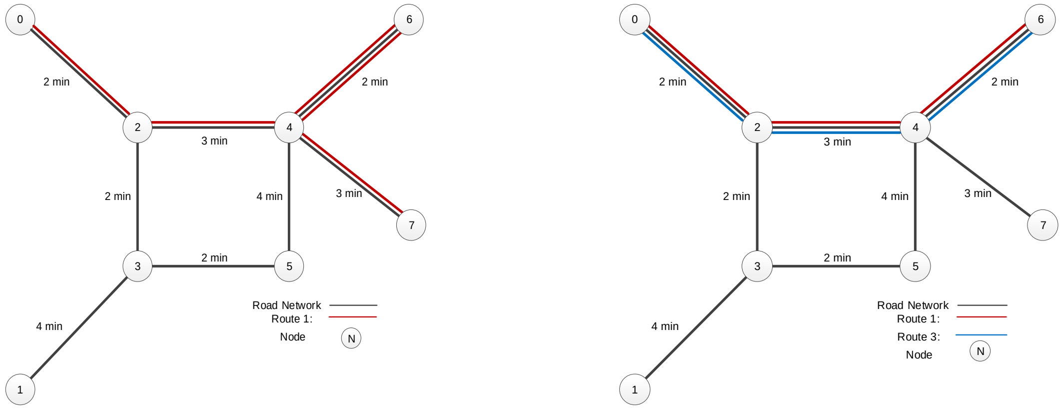
A *node coverage* means that for a given route set, all the nodes in the network must be covered. This is checked by taking the union $\bigcup_{r \in R} r$ of all the nodes of all the routes in the route set R and if the union is equal to the number of nodes in the network N then it means all the nodes are covered in the network. However, node coverage does not mean that the created transit network is also a connected network.

The *network connectivity* means that all routes are connected in such a way that there is always a path for any given OD pair. In other words, a route must share at least one node with any other route and their intersection $r_i \cap r_j \neq \varnothing$ is a non-empty set.

A small example with different conditions for TNF is listed in Fig 4.9. First, the base network is presented (Fig 4.9a), second the concept of node duplication is shown (Fig 4.9b), which is not allowed. Third, the route duplication in a PT network is shown (Fig 4.9c), which is also not allowed. Fourth, node coverage is shown where all the nodes in the PT network are covered (Fig 4.9d), however, it is still disconnected. Fifth, a network connectivity is shown (Fig 4.9e), where the PT network is fully connected. Furthermore, the PT network shown in Fig 4.9e is also called a feasible PT network because it does not have node and route duplication. Moreover, it covers all the nodes along with interconnected routes.

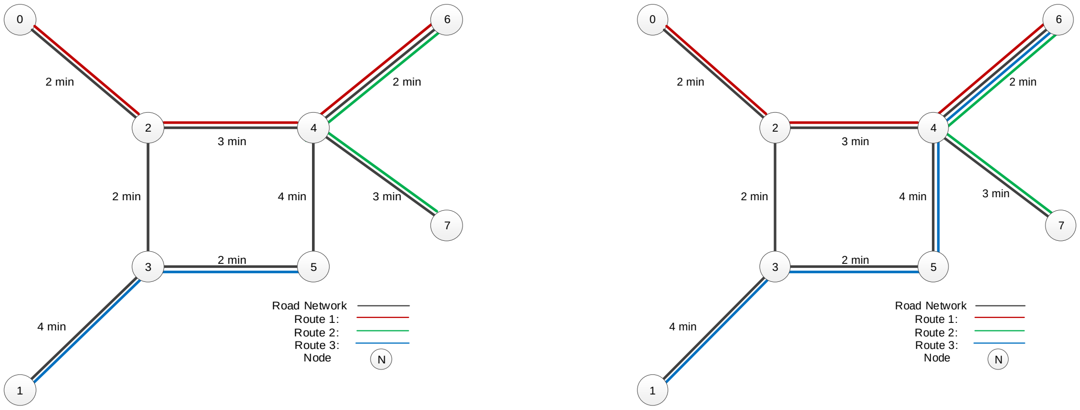


(a) Base network



(b) Node replication

(c) Route duplication



(d) Node coverage

(e) Network connectivity

Fig 4.9 Transit network feasibility conditions

4.1.4 Summary

In the TRNC component, route construction and network construction modules were explained. The route construction is fairly simple, the main idea in TRNC was the concept of ROR-based route insertion and how it is was used to create PT networks. Moreover, with the help of small examples, the potential of the proposed concept was also explained. Different variations of this concept can be used to create customised PT networks. In the next section, the transit demand assignment will be explained as it is one of the core components of the proposed methodology. Th evaluation of PT network heavily dependents on the type of assignment procedure.

4.2 Transit demand assignment and headway derivation

The TDAHD is an important component of the PT network design. There are two basic assignment methods: 1) headway-based; 2) schedule-based. A brief overview and comparison of these assignments is provided in section 2.4.1. In the proposed methodology, a headway-based assignment with optimal strategies as a choice model is used. This is ideal for strategic public transit network planning. The headway-based assignment calculates waiting time, in-vehicle time, transfer waiting time and transfer penalty. In this study, VISUM's PTV [2017] headway-based assignment procedure is used. This is very capable module, it can easily perform the assignment for large network instances.

Once the transit network is TNF (see 4.1.3.3) complaint, the demand is assigned by using TDAHD component. Three modules are used: 1) iterative transit demand assignment (ITDA); 2) headway derivation and mode selection (HDMS); 3) Iterative capacity constrained transit demand assignment (ICCTDA). The first two modules work together, the ITDA generates loads on the PT network whereas, HDMS used those loads to calculate required vehicle type and associated headway for each route. In the third module, the supply capacity of the routes is fixed, and the assignment is performed to get the capacity constrained assignments results. The reason for two different assignment components (ITDA and ICCTDA) is: 1) at first, iterative assignment and headways determination is used alternately, for capturing the general demand trends without any supply restrictions; 2) Later, iterative capacity constrained transit demand assignment is used to get the actual behaviour of transit network, given fixed transit supply. A schematic overview of the TDAHD is listed in Fig 4.10.

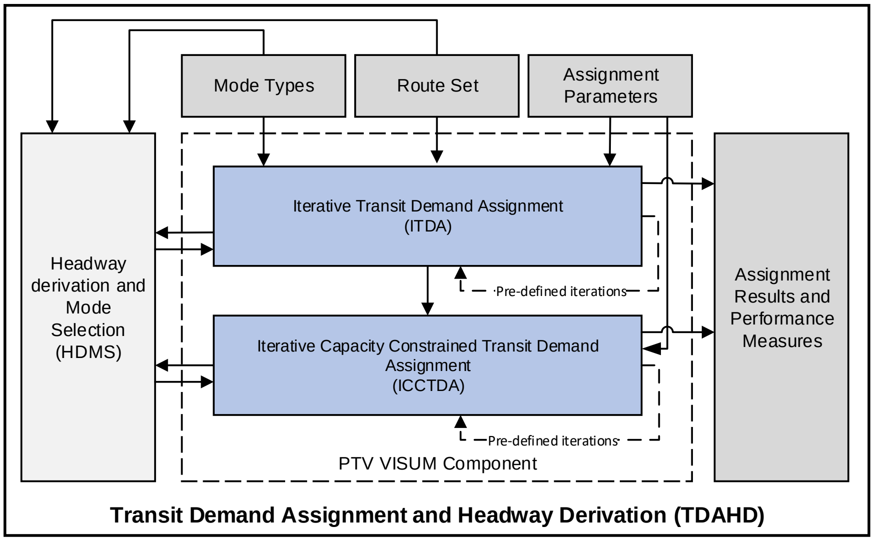


Fig 4.10 Schematic overview of the transit demand assignment

The central idea of TDAHD is that, at first, very short headways are assigned to all the routes. Later capacity free assignment is performed to get the route load profile (RLP, see section 4.2.2.1) and maximum load section (MLS, see section 4.2.2.2). With the help of RLP and MLS, new headways and the suitable mode types (MT, see section 4.2.2.3) for each route are selected. After this, capacity free assignment is performed again, and travel time components are calculated. This iterative assignment process continues, and the average travel time of two consecutive assignments are compared. If both are same, the capacity free assignment stops, otherwise the process continues until predefined number of iterations are reached. The next step is to perform capacity constrained assignment, the headway and the MTs from the last capacity free assignment are used for this purpose. The results of the capacity constrained assignment are used to identify overloaded routes. Later, penalties are assigned to those overloaded routes in terms of increased travel time. The capacity constrained assignment is performed repeatedly until no route overloading is observed or the predefined number of iterations are reached. The results from the last assignment are used for TNE (see section 4.3).

The main procedure for TDAHD is as follows:

1. The TDAHD procedure (see Alg 4.5) starts by performing the capacity free assignment.
2. If it is the first iteration, minimum headways (*Hmin*) are allocated to all the routes in the PT network. Through this assignment (CapFreeDemandAssig), average travel time and RLPs are derived.
3. These RLPs are used to acquire MLS (SelectMaximumLoadSegment), MT (SelectModeType) and new headway (RouteHeadwayDerivation) for each route.
4. After the first iteration, for the consecutive iterations, the headways assigned to the routes are based on the RLPs from previous iteration. The capacity free assignment is performed repeatedly until one of the two conditions is satisfied.
    a. There is no significant change in the average travel time of two consecutive iterations.
    b. The predefined number of iterations are reached.
5. The next step is to lock the attributes of all the routes (such as headway and MTs) and to perform capacity constrained assignment.
6. For each iteration of capacity constrained assignment (CapConDemandAssig), RLPs, travel time costs and supply capacities (RouteSupplyCapacties) of the routes are checked.
7. Based on demand and supply match, penalties (AssignTravelTimePenalty) are assigned to the route sections which are overloaded.
8. If RLPs are lower than the route supply capacities and routes are not overloaded, the capacity constrained assignment stops and the results from the current assignment are saved. Otherwise, the process continues until the predefined number of iterations are reached and the results from the final assignment are used.

Transit demand assignment and headway derivation algorithm

```
Input RouteSet(set of routes), Hmin(minimum allowed headway), Hmax(maximum allowed
headway), MT(mode type lookup table), CapFreeAssigItr(capacity free assignment iteration),
CapConAssigItr(capacity constrained assignment iteration)
RLPRoutes = GenerateArray()
MLSRoutes = GenerateArray()
MTRoutes = GenerateArray()
Hroutes = GenerateArray()
CurrentAvgTravelTime = 0
For each capfreeassigitr in CapFreeAssigItr:
    IF capfreeassigitr == 1:
        RouteLoadProfiles, AvgTravelTime = CapFreeDemandAssig(Hmin)
        CurrentAvgTravelTime = AvgTravelTime
        For each routeloadprofile in RouteLoadProfiles:
            MLSR = SelectMaximumLoadSegment(routeloadprofile)
            MTR = SelectModeType(MLS,MT)
            HR = RouteHeadwayDerivation(MTR, MLS, Hmax)
            Hroutes = UpdateRouteHeadway(HR)
            MTRoutes = UpdateVehicleType(MTR)
    End IF
    IF capfreeassigitr > 1:
        RouteLoadProfiles, AvgTravelTime = CapFreeDemandAssig(Hroutes)
        IF CurrentAvgTravelTime == AvgTravelTime:
            Break
        End IF
        CurrentAvgTravelTime = AvgTravelTime
        For each routeloadprofile in RouteLoadProfiles:
            MLSR = SelectMaximumLoadSegment(routeloadprofile)
            MTR = SelectVehicleType(MT,MLS)
            HR = RouteHeadwayDerivation(MTR, MLS,Hmax)
            Hroutes = UpdateRouteHeadway(HR)
            MTRoutes = UpdateVehicleType(MTR)
    End IF
For each capconassigitr in CapConAssigItr:
    OverLoadRouteExist = False
    RouteLoadProfiles, TravelTimeComponents = CapConDemandAssig(Hroutes)
    SupplyCapacityRoutes = RouteSupplyCapacties(Hroutes, MTRoutes)
    For each routeId, routeloadprofile in RouteLoadProfiles:
        For each stopsectionId, stopsectionload in routeloadprofile:
            IF stopsectionload >= SupplyCapacityRoutes[stopsectionId,routeId]:
                routesection = RouteSet[routeId, stopsectionId]
                AssignTravelTimePenalty(routesection)
                OverLoadRouteExist = True
            End IF
    IF not OverLoadRouteExist:
        Break:
    End IF
Output TravelTimeComponents (Travel time components for a given transit network)
```

Alg 4.5 Transit demand assignment and headway derivation procedure

4.2.1 Iterative transit demand assignment

In ITDA module, the headway-based assignment is used for demand assignment into the transit network. This generates travel time components such as waiting time, in-vehicle time and transfer waiting time as output. Moreover, it also helps in calculating the required headways for each route. Initially, the demand is assigned with low headways for each route. This assignment will generate the RLP for each route in the transit network. With the help of RLP, the MLS is identified. This MLS value is matched against the MT lookup table, to select the most suitable mode and required headways. Once, this is done, the assignment is started again, and this iterative process continues until there is no significant change in the average travel time per passenger in the network. The main purpose for ITDA is to iteratively improve routes and frequencies, starting with a very general frequency assignment, then gradually making the routes and schedules a better fit to reality in terms of travel path choices for the passengers and different PT supply capacities.

4.2.1.1 *Headway-based assignment*

In the proposed ITDA, a headway-based assignment with optimal strategy as a choice model is used. The main reason for choosing headway-based assignment is due to the following:

- It needs less data and information to calculate the assignment results. Moreover, the selected networks used for experiments have limited information such as demand for the whole day and static costs for the network arcs.
- It is computationally inexpensive.
- It is ideal for strategic planning where the planners are only interested in the overall structure of the transit network and its core attributes.
- Ideal for PT network with low headways.

The headway-based assignment suits the proposed methodology very well. The schedule-based assignment can be helpful for more tactical planning such as different frequencies for different time periods and considering the coordination among different routes and getting detailed route loads for different time periods of the day.

The flow chart of the assignment procedure is depicted in Fig 4.11.

1. At iteration zero, all the routes are assigned with 2 min headways and the assignment is performed.
2. The results of first assignment are used to calculate the flow (RLP) in the network and the MLS for each route.
3. In the following iterations, the MLSs obtained from previous assignment are checked against a MT lookup-table (see Tab 4.1) and new headways are assigned based on the offered route capacity and suitable MT. Both the lower bound and upper bounds for the headways are predefined.
4. With new mode types and headways defined, other properties such as route capacities and travel times are also updated. The results of the current assignment are saved.
5. This iterative process continues, and it stops when there is no significant change in the current average travel time with respect to previous iteration’s average travel time.

6. Once the average travel time convergence is established, the headways and MTs for each route are fixed. This information is used for ICCTDA module.

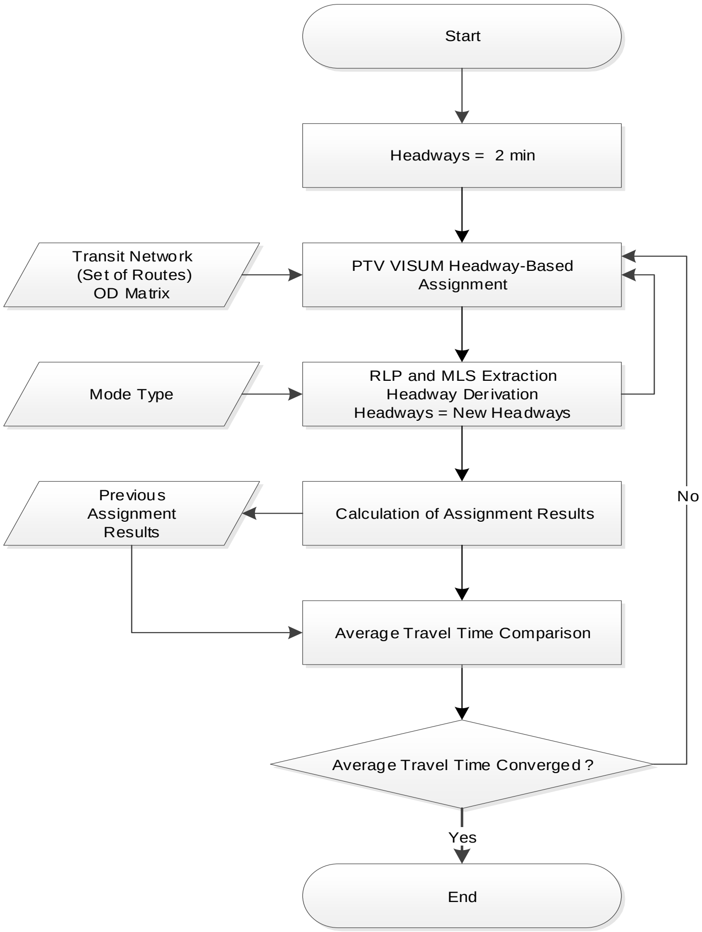


Fig 4.11 Flowchart of iterative transit demand assignment

4.2.1.2 *Optimal strategy as choice model*

The route choice model is part of the demand assignment. The optimal strategy is selected as the choice model. This is the most common route choice model for public transit assignment, proposed by Spiess & Florian [1989] where the passenger opt for a strategy instead of choosing a precise path. A strategy is simply a set of rules, when applied, allowed the passenger to reach its destination with minimum expected travel costs. These strategies can be simple or extremely complicated. The complexity greatly depends upon the available information to the passenger. To understand the concept, let’s take a small example adopted from Spiess & Florian [1989]. In Fig 4.12, there are four routes with stops 1, 2, 3 and 4. The stop sections (arcs) depict travel time in minutes.

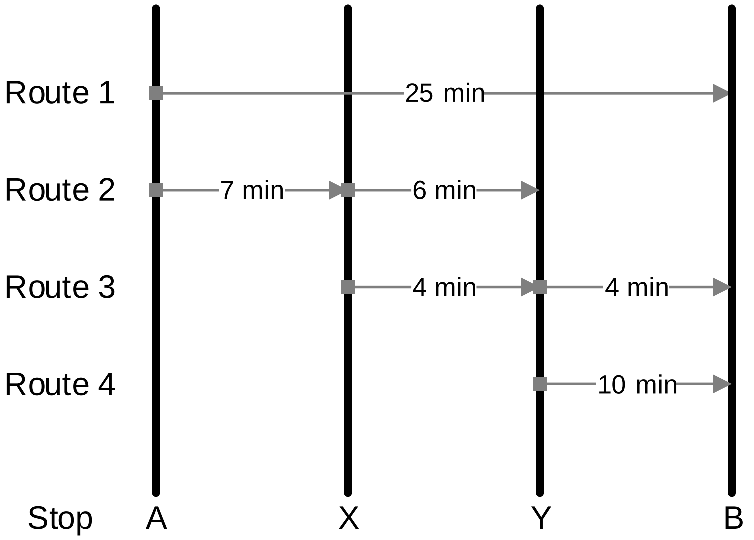


Fig 4.12 Sample transit network adopted from Spiess & Florian [1989]

A simple strategy to go from A to B would be to take which ever vehicle comes first from route 1 or route 2. If the route 1 is selected, then exit at stop B. If the route two is selected, then transfer at stop Y to either route 3 or 4 and exit at B.

A more complex strategy would be, wait for roughly 5 minutes for route 1, otherwise select route 2 and keep an eye at route 3 (express route) at stop 3. If the vehicle is spotted, transfer to route 3 and exit at stop B. Else, transfer at stop Y to either route 3 or 4 and exit at B.

In the proposed choice model, simple strategy is considered where average travel times, waiting times and the offered headways are considered. Moreover, it is also assumed that the only information that a passenger knows is the arrival of the vehicle associated with a route, at the stop while waiting to be served. Based on this information, the passenger will then decide whether to board the arriving vehicle or not.

Provided a strategy is selected the actual trip will be carried out in the following way:

1. Set the node to the origin node.
2. Board the vehicle that arrives first among all the vehicles of the attractive lines at origin node.
3. Alight at the predetermined node.
4. If it is the destination node, the trip is completed else set the node as the current node and move back to setp1.

4.2.2 Headway derivation and mode selection

In HDMS module, the mode type and headways are determined for all the routes. This module works together with IDA module as a pair. Both these modules share information among each other. In HDMS module, initially very low headways (2 minutes) are assigned as default to all the routes. The reason for assigning such headways is to generate RLP, this will show passenger volumes on to different routes. Once the RLPs for all the routes are calculated. The next step is to identify the MLS, for headway derivation. The MLS dictates the headway and the mode type for each individual route. The MLS is matched against a mode type lookup table and the most suitable mode is selected. Moreover, the number of vehicles required per route for one hour of operation are also calculated.

4.2.2.1 *Route load profile*

An RLP is the load on each section of the route for a given period of time, which is usually the peak hour. This load is the demand difference of the accumulated boarding and alighting passengers along the route course. The main objective of RLP is to provide the overview of the passenger load on a given route. Through this information, the most suitable mode and headway can be calculated such that adequate spaces are available for the passengers on the route. The RLP for a sample route is shown in Fig 4.13.

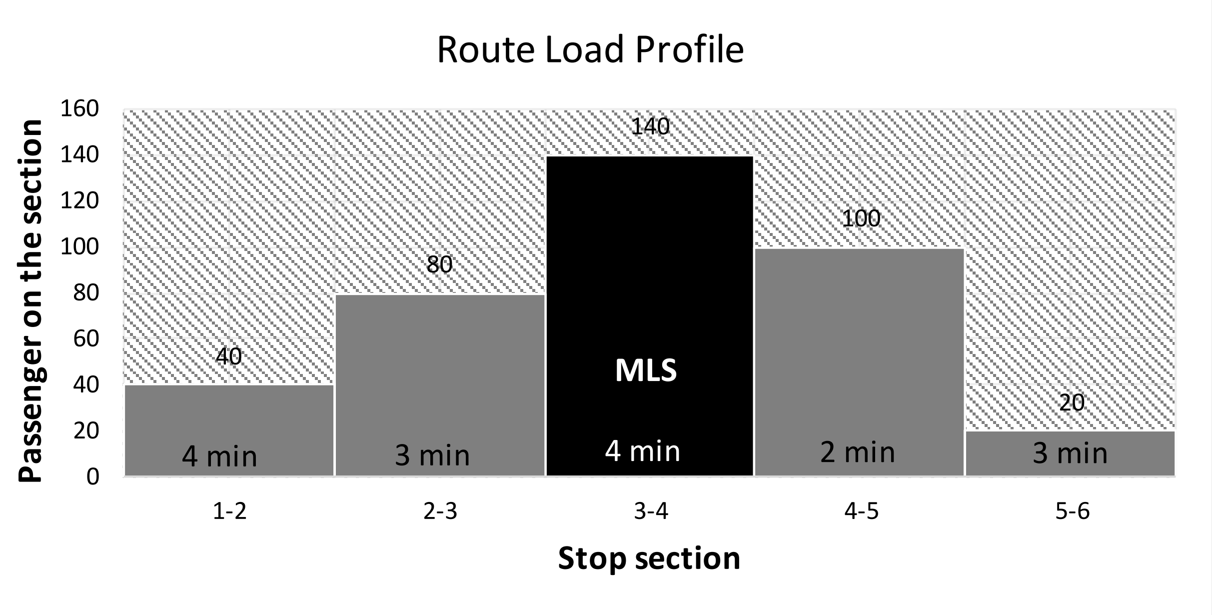
4.2.2.2 *Maximum load section*

The MLS is the stop section within the RLP, that carries the highest number of passengers. This MLS value dictates the required transit mode type and associated headway to satisfy the demand. The frequency of route r for the peak hour period j is:

$$F_r = \frac{\overline{MLS}_j}{d_{0j}}, \qquad d_{0j} = C_v \cdot \delta_i, \;\; 0 \le \delta \le 1 \tag{4.13}$$

where

$\overline{MLS}_j$ is the maximum number of passengers observed in peak hour period j.
d_{0j} is the desired occupancy in the vehicle during the time period j.
C_v is the capacity of the vehicle.
δ_j is the load factor during the time period j.
F_r is the frequency of the route r.



Bus stop	**1**	**2**	**3**	**4**	**5**	**6**
Stop link	**1-2**	**2-3**	**3-4**	**4-5**	**5-6**	
link time	4	3	4	2	3	
Boarding	40	120	130	120	10	0
Alighting	0	80	70	160	90	20
Demand on each section	40	80	140	100	20	0

Fig 4.13 Route load profile with maximum load segment

4.2.2.3 *Mode Type*

The MT is a mode lookup table (see Tab 4.1), which includes all the available modes with their associated properties such as, seating capacity, minimum and maximum headway, operating speed, and hourly supply capacity. The values are derived from Vuchic [2007] and adjusted according to the proposed methodology. For each given MLS value, it is matched against all the MTs and the most suitable mode is selected together with all its properties for the selected route. With the help of MT, the actual number of vehicles N_r^v required to satisfy the demand of a given route r for one hour can be calculated as:

$$N_r^v = \left[\frac{F_r \cdot CT_r}{60} \right]^+ \tag{4.14}$$

$$CT_r = RT_r \left(1+\gamma\right) \tag{4.15}$$

$$RT_r = 2T_r \tag{4.16}$$

where

CT_r is the cycle time of route r
RT_r is the round-trip time of route r
γ is the terminal time coefficient
T_r travel time on route r from start to end

Let's consider the small network in Fig 4.13, the MLS value is 140. Now this can be matched with the values in Tab 4.1. The value is lower than the hourly minibus service which is 1200 and it has a unit capacity of 40, therefore for this route minibus mode is most suitable. The value of F_r is 3.5, which can be rounded up to an integer value of 4. Now the time T_r is 16 min and the round trip CT_r, with γ value of 0.10 is equal to 35.2 min. Finally, the value of N_r^v is 2.3 vehicles which is rounded up to 3 vehicles.

Mode	Vehicle capacity (units*seats)	Minimum headway (seconds)	Maximum headway (seconds)	Maximum offered line capacity (Seats/hour)	Speed (km/h)
Minibus	1* 40	120	1,800	1,200	15
Standard bus high floor	1* 60	120	1,800	1,800	15
Standard bus low floor	1* 80	120	1,800	2,400	15
Articulated bus-1	1*100	120	1,800	3,000	15
Articulated bus-2	1*120	120	1,800	3,600	15
High capacity bus	1*150	120	1,800	4,500	15
Streetcar ROW C	1*200	120	1,800	6,000	22
Bus rapid transit	1*250	120	1,800	7,500	22
Streetcar ROW B-1	3*110	120	1,800	9,900	26.5
Streetcar ROW B-2	4*110	120	1,800	13,200	29
Rapid transit-1	4*150	120	1,800	18,000	38
Rapid transit-2	6*150	120	1,800	27,000	38
Rapid transit-3	8*150	120	1,800	36,000	38

Tab 4.1 Mode type and corresponding characteristics

4.2.3 Iterative capacity constrained transit demand assignment

The ICCTDA module is used to captures the actual loads on the transit network. The ICCTDA is like ITDA, the only difference with ITDA is that the frequencies and MTs are constant. The fixed frequencies and MTs help in determining the actual loads on the routes. A weighting factor on in-vehicle time is used to express the offered capacity limitation. This is done by adding an increased disutility experienced by passengers travelling on a crowded section of a route. To capture this disutility, first the ratio of the volume and offered capacity of each route stop section is calculated and later a volume-delay function is used.

$$\varepsilon_r^{ij} = 1 + ca.\left(\frac{Vol_r^{ij}}{Cap_r^{ij}}\right)^{cb}, \quad ij \in A_r, \ \forall r \in R \tag{4.17}$$

where

Vol_r^{ij} is the passenger volume on the stop section ij of the route r .

Cap_r^{ij} is the offered supply capacity on the stop section ij of the route r .
a,b are the coefficient and the power of the volume to capacity ratio respectively.
ε_r^{ij} is the disutility equivalent to the in-vehicle time.

This iterative process continues until all the route loads are within realistic range (adhere to the offered supply capacity) and there are no fluctuations in the route loads. The results of ICCTDA are used to calculate the travel time components as well as other performance indicators. These components help in evaluating the transit network.

4.2.4 Summary

In the TDAHD component, a two-level iterative demand assignment procedure with multi-modal vehicle selection was proposed. Such assignment procedure provides a realistic demand load on individual routes together with most suitable mode types and their associated properties. The results of these procedures are vital in evaluating the PT network. In the next section, the evaluation performance metrics will be explained from the perspective of passengers, agencies and authorities.

4.3 Transit network evaluation

The evaluation of a transit network must incorporate perspective of the main stakeholders of the system, these include passengers, agencies, and authorities. The proposed transit network evaluation (TNE) component considered all these perspectives and tries to perform quantitative and qualitative analysis of a transit network. A total of nine indicators representative of all these three stakeholders are used for the evaluation. Some are aligned with passengers and agencies while others are aligned with agencies and authorities. These three viewpoints with their associated indicators are listed in Tab 4.2.

Perspectives	Indicators								
	ROI	ESH	ANT	RSC	UDP	TNL	ATT	OC	TND
Passenger	■		■		■		■		■
Agency	■	■		■		■		■	■
Authority	■	■	■	■	■	■	■	■	■

Tab 4.2 Stakeholders perspectives with related performance indicators

4.3.1 Route overlap Index

The route overlap index (ROI) is calculated as the ratio of the sum of the lengths of all the routes to the total network length. A value of one means least amount of overlapping between the routes in a PT network. This is ideal for a transit network with large capacity modes, as the sole objective of such system is to offer direct and reliable services to the passengers. However, there are many high capacity transit systems with some level of route overlapping for providing more direct connections to the passengers and sharing of common sections between different routes to avoid transfers. The ROI represents the viewpoint of the passengers, agency and the authority.

- For passengers, a lower ROI means much simpler PT network.

- For authority and agency, smaller ROI means less redundant transit network, which is considered as cost efficient and operationally viable.

The ROI is adopted from Vuchic [2005], it can be computed as the ratio of the sum of the lengths of all the routes L_r to the length of the transit network TL .The TL is the length of the transit network (sum of all route length) minus the duplicate arcs within the transit network (overlapping routes' stop sections). Moreover, L_a^k is the length of the arc within the transit network which occurs k times.

$$\text{ROI} = \frac{\sum_{r \in R} L_r}{TL} \tag{4.18}$$

$$TL = \sum_{r \in R} L_r - \sum_{a \in TNA} (k-1) \cdot L_a^k \tag{4.19}$$

4.3.2 Required seating capacity

The required seating capacity (RSC) is defined as the number of seats required to satisfy the demand for a given time period. The seats in this case refer to the available seats and/or the standees in a vehicle. To calculate the RSC, first the complete cycle time of a route is calculated, second, the ratio between the cycle time and headway is calculated, and the result is rounded to an integer value to calculate the required number of vehicles. Third, the mode type and number of vehicles are identified through a selection procedure from the lookup table (see Tab 4.1). Fourth, once the total number of vehicles and their capacities are calculated for the whole network, the selected modes are translated into seating capacity.

The reason for converting different mode types and their capacities into a seating capacity is to obtain a single indicator which can represent the required PT supply capacity. The RSC purely represents the perspective of the agencies and authorities.

- For agencies, smaller values for the RSC represent a smaller fleet size and lesser operational costs.
- For authorities, smaller values of RSC mean lesser subsidies for the PT services.

The RSC can be calculated as sum of the product of vehicles N_r^v required for each route r in PT network and its unit capacity C_v (see section 4.2.2.3).

$$\text{RSC} = \sum_{r \in R} N_r^v \, C_v \tag{4.20}$$

4.3.3 Empty seat hours

Empty seat hours (ESH) shows the unutilised offered capacity of the PT system. It refers to the number of seats in PT vehicles that are unoccupied during operation time. This parameter was adopted from Ceder [2016]. It represents the perspective of the agencies and authorities, as both want to have higher utilisation of the offered capacity.

- For the agencies, higher utilisation translates into higher earnings.
- For the authorities, it means less subsidies for the service.

The ESH for a single route is calculated as follows: first, the product of the difference of the offered capacity to the volume on a given arc and arc's travel time is calculated. Later the process is repeated for all the arcs of the route. This process is continued for all the routes present in the transit network. The ESH can be calculated as the sum of the difference of the supplied Cap_r^{ij} and the volume Vol_r^{ij} on the arc ij of route r for the arc time $A_r^{ij,t}$

$$\mathrm{ESH} = \sum_{r \in R} \sum_{ij \in A_r} (Cap_r^{ij} - Vol_r^{ij}) \left(\frac{A_r^{ij,t}}{60} \right) \tag{4.21}$$

4.3.4 Average travel time

Average travel time (ATT) refers to the average time a passenger spent in the PT system for commuting. The ATT is one of the two objectives that are used for the TNDFSP optimisation. It is calculated by dividing the total travel time of all OD pairs by the total demand. It includes waiting time, in-vehicle time, and transfer penalties. This represents the perspective of both the passengers and the authorities.

- For passengers, low ATTs are preferred for reduced travel times.
- From the authorities' perspective, low ATTs mean a more attractive PT services for the passengers.

The ATT can be calculated by using the first objective function (see section 3.3.2 for more details).

$$\mathrm{ATT} = \frac{1}{\sum_{i,j \in N} d_{ij}} \left(\sum_{r \in R} \sum_{i,j \in N_r} d_{ij}^r t_{ij}^r + \sum_{tp \in TP} \sum_{i,j \in N_{tp}} d_{ij}^{tp} t_{ij}^{tp} + \sum_{r \in R} \frac{1}{2F_r} \left(\sum_{i,j \in N_r} d_{ij}^r + \sum_{i,j \in N_{tp}} d_{ij}^{tp} a_{tr}^r \right) + \sum_{tp \in TP} \sum_{i,j \in N_{tp}} d_{ij}^{tp} tr_{ij}^{tp} \mu \right) \tag{4.22}$$

4.3.5 Average number of transfers

Average number of transfers (ANT) is defined as the number of transfers required to complete the trip from the origin stop to the destination stop for all OD pairs. This indicator strictly represents the viewpoints of the passengers and the authorities.

- For passengers, lower values of ANT indicate, on average, a smaller number of transfers per trip, which translates into better service and more direct trips.
- For authorities, lower values of ANT represent the PT-network attractiveness (increase connectivity).

The ANT can be calculated as: at first, the sum of the product of the demand that requires a transfer path to complete the trip d_{ij}^{tp} and the number of transfer tr_{ij}^{tp} involved to complete the trip is calculated. Second, the values are divided by the total demand d_{ij} in the network (inclusive transfer-based and directed trips).

$$\text{ANT} = \frac{1}{\sum_{i,j \in N} d_{ij}} \sum_{tp \in TP} \sum_{i,j \in N_{tp}} d_{ij}^{tp} tr_{ij}^{tp} \tag{4.23}$$

4.3.6 Total network length

Total network length (TNL) refers to the combined length of all the routes in a network. It strictly represents the viewpoint of the agencies.

- A shorter TNL is beneficial for agencies and authorities, as fewer PT units would be required for operations. Therefore, reduction in the fuel costs, mileage of PT units and subsidies.

The TNL is defined as sum of round-length L_r of all the routes R in a transit network.

$$\text{TNL} = \sum_{r \in R} 2.L_r \tag{4.24}$$

4.3.7 Unsatisfied demand percentage

Unsatisfied demand percentage (UDP) refers to the transit demand which is not satisfied by the transit network properly. The main objective of a transit network is to offer fast, reliable service with least number of transfers to the passengers. However, creating such a system is almost impossible due to the costs associated with such system. In the proposed methodology a demand is considered as unsatisfied if it requires three or more transfers. The UDP strictly represent the viewpoints of the passengers and the authorities.

- For passengers, lower UDP values show most of the demand is served adequately.
- For authorities, lower UDP values show the attractiveness of the given transit system.

It is calculated as the percentage of total demand. The UDP can be calculated as the ratio of the demand that is served by transfer paths d_{ij}^{tp} involving more than the maximum number of transfers allowed per trip tr_{ij}^{tp} to the total demand in the transit network d_{ij}. The value of UDP is expressed as percentage and the maximum number of transfers $tr_{\max}$ per trip is predefined (e.g., maximum of two transfers allowed per trip).

$$\text{UDP} = \frac{\sum_{i,j \in N_{tp}} d_{ij}^{tp} tr_{ij}^{tp}}{\sum_{i,j \in N} d_{ij}} * 100; tr_{ij}^{tp} > tr_{\max} \tag{4.25}$$

4.3.8 Operating cost

The operating cost (OC) is the hourly cost associated with a heterogeneous transit system which includes different modes such as buses, BRTs, LRTs and MRTs. Four different cost components are considered: 1) on-vehicle crew cost (C_r^{vhr}); 2) vehicle direct operating cost(C_r^{vkm}); 3) infrastructure maintenance cost (C_r^{mkm}); 4) overhead operating cost (C_r^{ocp}). The on-vehicle crew cost includes wages for the drivers and other on-vehicle crew costs. The vehicle direct operating cost includes vehicle fuel, power and other vehicle maintenance cost. The infrastructure maintenance cost includes track maintenance, right of way and signalling. The overhead operating cost includes the cost which is not considered in the other three costs. These includes scheduling, supervision, non-labour and office costs. The OC represents the viewpoint of the agencies and the authorities.

- For authorities, low OC values mean lesser subsidies for the offered PT services.
- For agencies, low OC values mean lesser operational cost for the offered PT services.

The OC can be calculated by using the second objective function (see section 3.3.2).

$$\mathrm{OC} = \sum_{r \in R} \left\{ \left[(N_r^v \cdot C_r^{vhr}) + (2F_r . L_r . C_r^{vkm}) + (RL_r \cdot C_r^{mkm}) \right] . (1 + C_r^{ocp}) \right\} \tag{4.26}$$

4.3.9 Transit network directness

Transit network directness (TND) relates to the indirectness of all passenger trips in a PT network. There are different types of TND, according to Lee & Vuchic [2005], there are two types: 1) distance-based TND; 2) time-based TND. The TND represents the perspective of all three stakeholders.

- For passengers, a low TND value means more direct service for them.
- For agencies and authorities, least indirectness in the transit systems attracts more passengers and shows a strong image of the transit network.

The distance-based TND is considered in the proposed TNE. The TND can be calculated as: 1) at first, all the paths p_r^{ij} of route r are calculated; 2) later the ratios of these paths $TnPR_r^{ij}$ with their shortest paths sp_{ij} are calculated; 3) these ratios are added together and their sum is divided by the total number of paths TnP_r of the route r ; 4) These calculations are repeated for all the routes, later they are added together, and then divided by total number of routes r . The ideal value of the TND is 1.0, which means whole transit network is based on the shortest path. However, values exceeding 1.0 are considered as indirectness of the offered transit network service.

$$PD_r = \frac{TnPR_r^{ij}}{TnP_r} \tag{4.27}$$

$$TnPR_r^{ij} = \sum_{i,j \in N} \frac{p_{r,ij}}{sp_{ij}}; \qquad \forall sp \in SP \tag{4.28}$$

$$TnP_r = \frac{N_r * (N_r - 1)}{2} \tag{4.29}$$

$$\text{TND} = \frac{\sum_{r \in R} PD_r}{\sum_{r \in R} r} \tag{4.30}$$

4.3.10 Summary

In the TNE component, different performance indicators were briefly explained. These indicators provide a broader perspective of the PT network performance as compared to indicators used in the related TNDFSP literature. These indicators will help in holistically assess the performance of PT networks from different perspectives. In the next section, the metaheuristic search engine used to explore the search space, and to improve the created PT networks, will be explained.

4.4 NSGA-II based metaheuristic search engine

The TNDFSP belongs to a set of NP-hard problems, exact methods cannot solve this problem efficiently especially for large instances. Therefore, heuristic/metaheuristic methods seem more reasonable to solve large instances of this problem with good quality solution in reasonable time. However, the global optimality of the solution is not guaranteed. Over the past three decades, metaheuristics gained a lot of popularity. They are being used in numerous fields such as operational research, computer science, artificial intelligence and management science. The most common metaheuristic methods include GA, LS, SA and TS. Among these, GA is the most popular algorithm to solve the TNDFSP. In the proposed metaheuristic search engine, NSGA-II MSE is used.

In the NSGA-II MSE component, at first route overlap ratio-based initialisation (ROR-INI) is performed to generate the initial population of feasible solutions. Later crowded comparison operator-based selection (CCOS) is performed to select the individuals for reproduction. This reproduction include route overlap-based multipoint crossover (RORMC) and polynomial mutation (PM). There is a high possibility that the newly formed individuals are not feasible. Therefore, once the reproduction is done, the newly created individuals are subjected to the procedures of TNC module. These procedures include: 1) TNI for network improvement; 2) TNF for feasibility check. After these two procedures, the process goes back to CCOS. The whole iterative process stops after reaching predefined number of generations. In NSGA-II MSE component, first the basic concept of GA is explained. Second, NSGA-II is explained followed by the basic components of GA such as representation, initialisation, selection, evaluation and termination. A schematic overview of NSGA-II MSE is depicted in Fig 4.14.

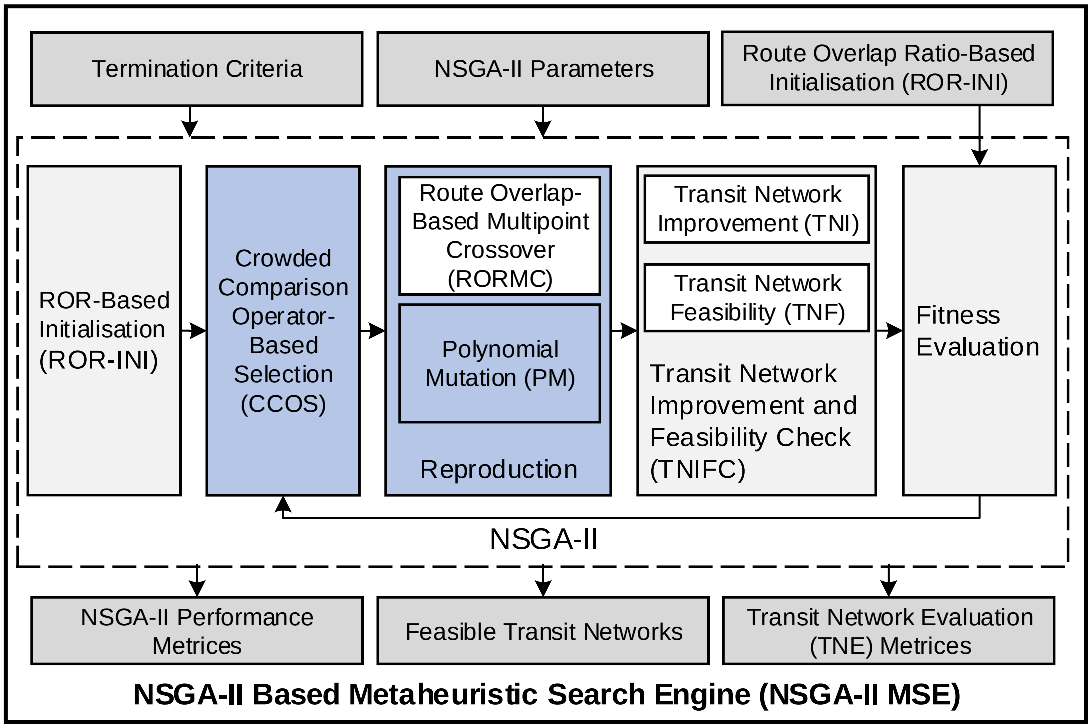


Fig 4.14 A schematic overview of NSGA-II based metaheuristic search engine

4.4.1 Genetic algorithm

The GA is first designed by Holland [1975], it belongs to the class of adaptive search methods based on the natural evolution. Over the past few decades, GA has been employed to solve complex combinatorial optimisation problems. The basic concept of GA is that it works on a set of candidate solutions called population, which evolved under natural selection, reproduction and replacement (see Fig 4.15).

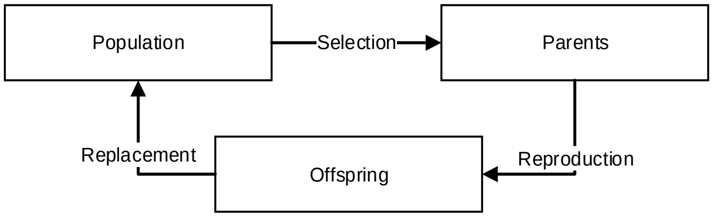


Fig 4.15 Genetic algorithm basic idea

In general, GA works as an iterative procedure, where at first a set of finite candidate solutions called initial population, is generated randomly or via heuristics. Each solution is named as individual, these individuals are evaluated according to a fitness function. Individuals with higher fitness have higher chances of getting selected for the mating process. Two basic GA operators (crossover and mutation) are used in the mating process. The offspring are evaluated just like their parents. The members of new generation with higher fitness value will replace the weaker members from the last generation. This process of evaluation, selection and mating continues until no further improvement is witnessed or given number of generations are reached. The basic GA is depicted in Alg 4.6.

```
Genetic Algorithm
Input $P$ (population size), $T$ (number of generations)
Generate (Pop(0)) ;
t = 0
While $t < T$:
        EvaluatePopulation(Pop(t));
        Pop'(t) = SelecIndividuals(Pop(t));
        Pop'(t) = PerformMating(Pop'(t));
        Pop'(t) = EvaluatePopulation(Pop'(t)) ;
        Pop(t + 1) = Replaceindividuals(Pop(t), Pop'(t)) ;
        t = t + 1 ;
End While
Output (Pop) Population after $T$ iterations
```

Alg 4.6 Generic algorithm of GA

4.4.1.1 *Main components of GA*

Representation of each candidate solution is done by an array of binary numbers. However, real-valued and integer coding are also used.

Population initialisation is used to derive the GA, an initial set of candidate solutions called population is created. This population can be initialised by randomly generating the solutions or by seeding the initial population.

Fitness function is the objective function used to assess the fitness of the candidate solutions.

Selection is the probability of a candidate solution to get selected for mating process. The selection makes sure that candidate solutions with higher fitness values get more chances to mate.

Reproduction is the mating process where crossover and mutation operators are used to create new candidate solutions.

Replacement is employed, once the offspring are generated, both parents and offspring compete to stay in the population. The weak candidate solutions are replaced with stronger ones.

Stopping criteria are based on two conditions whenever one of these conditions is satisfied, the whole process stops. These conditions are: 1) number of generations reached a prespecified number; 2) there is no significant difference in the consecutive generations.

The basic genetic algorithm process is depicted in Fig 4.16.

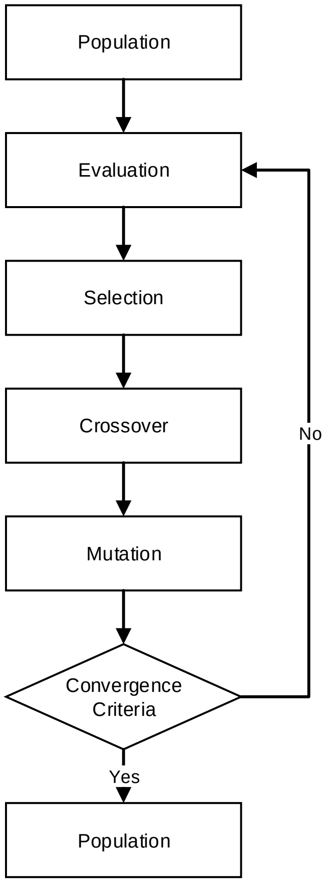


Fig 4.16 Main components of genetic algorithm

4.4.2 Nondominated sorting genetic algorithm-II

The NSGA-II is a fast and elitist MOEA, based on a nondominated sorting algorithm proposed by Deb, Pratap, Agarwal, & Meyarivan [2002]. It is an extension of the GA for solving multi objective optimisation problems. The NSGA-II uses typical GA operators such as selection, crossover and mutation together with nondominated sorting algorithm to create rank based sub-populations. A schematic overview of NSGA-II procedure is depicted in Fig 4.17. The main procedures of NSGA-II include: 1) fast nondominated sort (FNS); 2) crowding distance assignment (CDA); 3) crowded comparison operator (CCO).

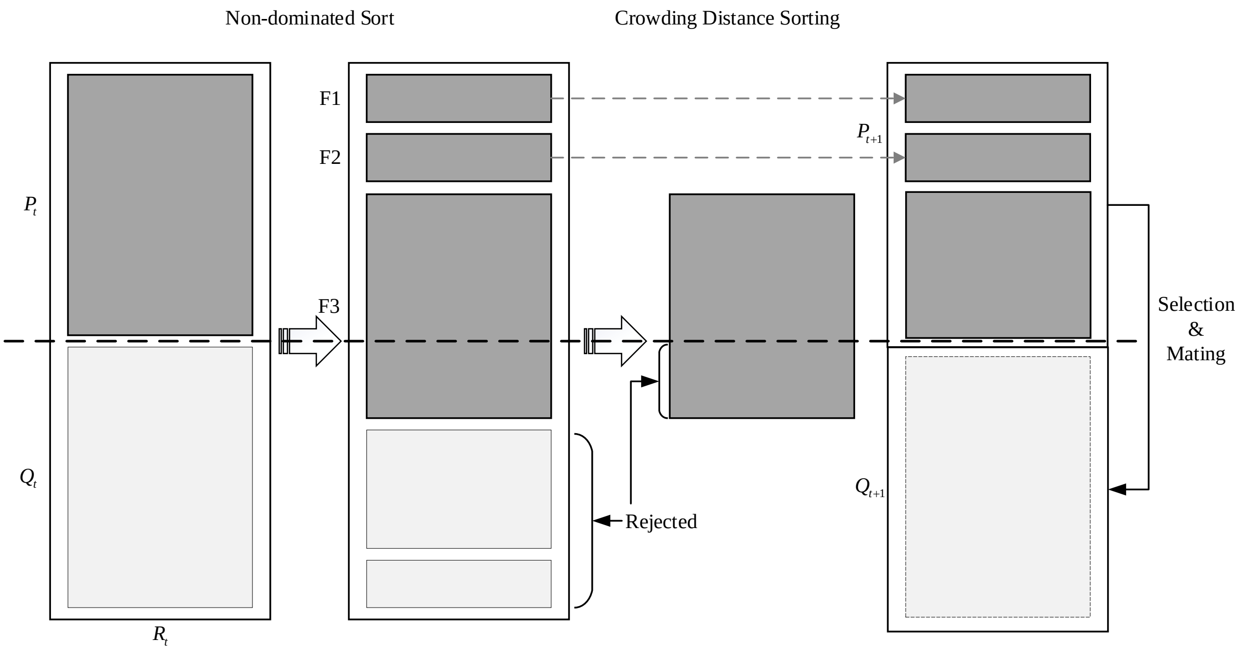


Fig 4.17 NSGA-II procedure adopted from Deb, Pratap, Agarwal, & Meyarivan [2002]

The main idea of NSGA-II is to first create two sets of populations of size N. Later these are combined and sorted based on their nondomination ranks. A new population of size N is created based on the ranking and distance. If two solutions share the same rank, the one which lies in less dense region, is selected. This new population is used to create offspring through selection and mating process. The newly created offspring population of size N and the parent population are combined again to form new population and the process continues until the stopping criteria are reached.

The basic procedure is given below:

1. Initially, a parent population P_t of size N is generated, and offspring population Q_t of the same size is generated from P_t.
2. Both the parent and the offspring are combined to form a combined population R_t of size $2N$.
3. R_t is sorted based on nondomination and new population P_{t+1} of size N is filled by solutions of different nondominated fronts. The filling start with the best nondominated front F_1 and then to second best nondominated front F_2 and so on. Since N slots are available, not all the fronts will be accommodated, only the top fronts will be included. The fronts which are not accommodated will simply be discarded.
4. Now P_{t+1} is subjected to the selection, crossover and mutation to create new population Q_{t+1}. The selection is based on binary tournament, by using the crowded comparison operator.
5. The process continues until the stopping criterion is met.

The basic algorithm of NSGA-II is given in Alg 4.7.

```
NSGA-II
Input $N$ (Pop$_{size}$)
$t = 0$;
Generate ($P_t(0)$) and ($Q_t(0)$);
$R_t = P_t(0) \cup Q_t(0)$;
$F$ = FastNonDominatedSort($R_t$);
$P_{t+1} = \varnothing$ and $I = 1$
While $|P_{t+1}| + |F_i| \leq N$
    CrowdingDistanceAssignment($F_i$);
    $P_{t+1} = P_{t+1} \cup F_i$
    $I = I + 1$
    CrowdedComparisonOperatorSort($F_i$);
    $P_{t+1} = P_{t+1} \cup F_i[1:(N-|P_{t+1}|)]$
    $Q_{t+1}$ = MakeNewPopulation($P_{t+1}$)
    $t = t + 1$
End While
Output $Q_{t+1}$ (All nondominated solutions in the population after final iteration)
```

Alg 4.7 NSGA-II

4.4.2.1 *Fast nondominated sort*

In FNS, the population is sorted based on their domination (see section 4.4.2.2).

The basic, procedure is as follows:

1. For each solution two entities are calculated:
    a. n_i, that is the number of solutions which dominate the solution i.
    b. S_i, a set of solutions that solution i dominates.
2. All the solutions in the first nondominated front will have their $n_i = 0$. With each solution i with $n_i = 0$, each member j of its set S_i is visited and reduce its domination count by one.
3. If the domination count becomes zero, for any member j, it is placed in a separate list P_k.
4. Now these members belong to the second nondominated front.
5. The above procedure is continued to identify all the fronts.

4.4.2.2 *Domination*

The concept of domination in the context of multi-objective optimisation means comparison of two solutions. It checks whether, a solution dominates other solution or not. The domination can be defined as follows:

A solution $x^{(1)}$ is said to dominate other solution $x^{(2)}$, if both conditions 1 and 2 are true:

1. The solution $x^{(1)}$ is no worse than $x^{(2)}$ in all objectives.
2. The solution $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective.

If any of the above-mentioned conditions is violated, the solution $x^{(1)}$ does not dominate the solution $x^{(2)}$. Mathematically it can be denoted as:

$$x^{(1)} \prec x^{(2)} \tag{4.31}$$

4.4.2.3 *Crowding distance assignment*

In order to get an estimate of the concentration of the solutions surrounding a selected solution i in the population. Average distance of two solutions that are immediate neighbours of the solution i along each of the objectives is calculated. This average distance i_{distance} serve as an estimate of the perimeter of a cuboid formed by using the closest neighbours as vertices. This is called as crowding distance.

The procedure of CDA is as follow:

1. The CDA requires first sorting of the population according to each objective in an ascending order.
2. After the sorting, each of the boundary solutions are assigned infinity values.
3. Apart from the two boundary solutions, other intermediate solutions are assigned a distance value, which is equal to the absolute normalised difference in the values of the two adjacent solutions.
4. The total overall crowding distance value is calculated as the sum of individual distance values corresponding to each objective.

4.4.2.4 *Crowded comparison operator*

Once the crowding distance and the front of each solution is calculated. The crowdedness of a solution can be measured. Basically, CCO $\prec_n$ compares two solutions and return the winner. Each solution has two attributes:

1. Nondominated rank i_{rank} in the population.
2. Crowding distance i_{distance} in the population.

A solution i is preferred over solution j:

1. If solution i has better rank than solution j
2. If both solutions have the same rank but solution i has better crowding distance than solution j, that is:

$$i \prec_n j \,\text{if}\, (i_{\text{rank}} < j_{\text{rank}})\,\text{or}\,((i_{\text{rank}} = j_{\text{rank}})\,\text{and}\,(i_{\text{distance}} > j_{\text{distance}})) \tag{4.32}$$

The solution which lies in the better nondominated front is preferred. If this is not the case and both solutions share the same nondominated front. The solution which lies in less dense area or in other words having a larger crowding distance is preferred. Through such CCO, n number of solutions can be selected.

4.4.2.5 *New population and offspring*

The new population P_{t+1} is formed by using the CCO to select N solutions. These selected solutions are used to create offspring population Q_{t+1} of the size N. First, the pair formation of the P_{t+1} with the help of k-combinations where the k is set to 2 and the total number of pairs is set to N.

$$\binom{n}{k} = \frac{n(n-1)....(n-k+1)}{k(k-1)...1} \qquad k = 2 \tag{4.33}$$

These pairs are then subjected to CCO-based binary tournament. Both solutions in a pair compete and only one of them is considered as winner. At the end, N number of solutions are selected. Now these solutions are subjected to typical GA operators such as crossover and mutation. Once the mating process is complete, all the solutions are evaluated. These solutions form the offspring population Q_{t+1}.

4.4.3 Representation

To represent an individual, we used an associative array. It is an abstract data structure composed of a collection of key and value pairs. Such that each key is unique and appears only once. Moreover, a key is encoded as a tuple, where each key is a pair of tuples. The value is encoded as an integer list data structure. For instance, consider the network in Fig 3.1, let's assume the individual S_1 has two routes r_1 and r_2 with nodes N_{r1} and N_{r2}. Based on this information, the representation of the individual route is presented in the Tab 4.3.

Individual-Id	route Id	Route nodes				
S_1	1	0	2	4	7	-
	2	1	3	5	4	6

Tab 4.3 Representation of an individual

4.4.4 Route overlap ratio-based initialisation

Initialisation plays an important role in the population diversification as well as the progression of the search algorithm. According to Talbi [2009], there are five types of initialisation strategies namely: 1) pseudo-random; 2) quasi-random; 3) sequential diversification; 4) parallel diversification; 5) heuristics. These strategies have some pros and cons, some of these strategies are good in providing the initial diversity, some requires lower computational cost, and some provide high quality solutions. In the proposed initialisation, a hybrid of pseudo-random and heuristic strategy along with the concept of ROR is employed.

The procedure for ROR-INI is as follows:

1. Initially, a random sample (RandomSample) of the population size is drawn from the range of the total feasible routes.
2. This sample is used to create initial route sets with each route set having a single route. These routes are derived from the feasible route set based on the random sample.
3. Each route set is filled with the routes from the feasible route set based on the ROR values.

4. After every route insertion, TNF is check. If the route set (transit network) is feasible, then the route insertion for the current route set stops and insertion for next route set starts. Else, the insertion of routes continues (RORInsertion).
5. If the route set fails TNF criteria even after all the routes for insertion are exhausted, TNI procedures (RouteImprovement) such as route extension and node insertion are employed to improve the routes in the route set such that it passes TNF criteria (RouteSetFeasibility).

The procedure for RHIS is shown in Alg 4.8.

```
Pseudo-random heuristic initialisation algorithm
Input N (population size), FRS (feasible route set), ROR (route overlap ratio)
RandomSample = Generate RandomSample (FRS, N)
RouteSet = Generate RouteSetArray (N)
For each rn in RandomSample:
    RouteSet = InsertRoute (FRS[rn])
For each rs in RouteSet:
    RouteSetComplete=False
    While Not RouteSetComplete:
        For each r in FRS:
            RORInsertion(r, rs, ROR)
            IF RouteSetFeasibility(rs):
                RouteSetComplete = True
                Break
            End IF
        IF RouteSetComplete = False:
            RouteImprovement(rs)
        End IF
        IF RouteSetFeasibility(rs):
            RouteSetComplete = True
            Break
        End IF
    End While
Output RouteSet (Initial population of different route sets)
```

Alg 4.8 Pseudo-random heuristic initialisation procedure

4.4.5 CCO-based binary tournament selection strategy

The selection strategy is one of the main components of the GA. The basic idea is to allow the individuals with better fitness to get selected for the sake of mating. This allows the population to move in a solution space where higher quality of the individuals is achieved. However, this is not always true. Sometimes, individuals with lower fitness levels are useful for mating and creating more diverse population. Therefore, a small proportion of low-quality individuals should also be allowed to get selected for mating process. There are many selection strategies such as 1) roulette wheel selection; 2) tournament selection; 3) rank-based selection; 4) stochastic universal sampling. In the proposed selection strategy, a CCO-based binary tournament is used.

The procedure for CCO-binary tournament is as follows:

1. At first, for all the individuals in the population, two attributes are calculated using CCO:
    a. Nondominated rank.
    b. Crowding distance.
2. Later combinations of all the individuals in the population in the form of pairs is performed, of the size Pop * N.
3. For each pair, both individuals compete in a binary tournament. The evaluation is performed based on CCO.
4. The winners of tournament are kept for reproduction process.

4.4.6 Reproduction

Once the selection is performed, the next step is reproduction of the offspring. This is done by using two different operators namely crossover and mutation. The crossover operator is performed on the selected parents for the breeding of offspring. The main purpose of the crossover is to allow the offspring to receive characteristics from the parents. Different kinds of crossover can be used based of the application and the coding type. Some of these include: 1) one-point crossover; 2) two-point crossover; 3) uniform crossover; 4) intermediate crossover; 5) unimodal normal distribution crossover; 6) parent centric crossover. Mutation operators are generally unary in nature and often applied on a single individual. Mutations are used to bring small minor changes into the individuals of the population. Some of the mutation operators include: 1) flip bit mutation; 2) boundary mutation; 3) shrink mutation; 4) uniform mutation. Usually, small values of mutation are preferred.

4.4.6.1 *Route overlap ratio-based multipoint crossover*

In the proposed route overlap ratio-based multipoint crossover (RORMC), at first, for all the selected individuals, combinations in the form of pairs is performed. Later, a sample of the pairs with the same size as the original population is drawn from it. These pairs are later subjected to ROR-based multi-point crossover. At first, a random value is drawn between [0-1], if the number is less than the pre-set crossover rate value. In that case, the RORMC is performed.

The procedure used for RORMC is as follow:

1. At first, a pair of individuals are selected.
2. For each individual (route set), a random sample of elements (routes) of size n is drawn from the individual's sequence.
3. These elements are then removed (ExtractSample) from that individual and placed in an empty array called offspring. With this, two new offspring are created (newly created array) which are partially filled with the elements.
4. The ROR-based route insertion procedure (RORInserion) is applied to the first offspring with route selection from opposite original individual.
5. Only ROR complaint routes will be inserted into the new individual. The same process is applied to the other offspring as well.
6. At the end, two new offspring are created (AddOffspring), which possess some characteristics from their parents.

7. Since it's a ROR-based crossover, there is a possibility that some of the routes are not inserted because of ROR limit and the created individuals are infeasible. This issue is catered in two ways:
    a. The size of the pair list is set much higher than the population size.
    b. The TNI and TNF modules are used to improve and make the individual feasible.

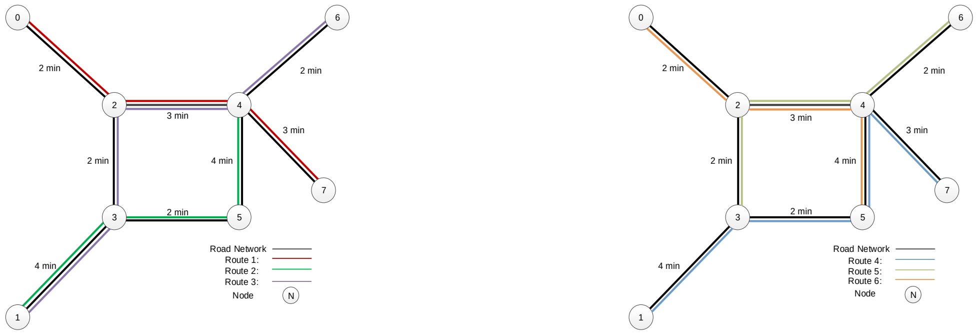
The algorithm for RORC is shown in Alg 4.9.

```
ROR-based multipoint Crossover Algorithm
Input Parents (pair of route sets), Crate (crossover rate), ROR (route overlap ratio), CrPoint
(crossover point)
RN = RandomNumber(0,1)
Offspring = GenerateArray()
For each parent in Parent:
    RN = GenerateRandomNumber([0,1])
    IF RN < Crate:
        offspring1 = ExtractSample (parent[1], CrPoint)
        offspring2 = ExtractSample (parent[2], CrPoint)
        For each p in parent[1]:
            RORInserion(p, offspring2, ROR)
        For each p in parent[2]:
            RORInserion(p, offspring1, ROR)
    Offspring = AddOffspring(offspring1)
    Offspring = AddOffspring(offspring2 )
    Else:
        Offspring = Parents
    End IF
Output Offspring (new route set after crossover)
```

Alg 4.9 Route overlap ratio-based multipoint crossover procedure

A small example is used to demonstrate the RORMC concept (see Fig 4.18). Imagine there are two route sets namely S_1 and S_2 with three routes each. The routes nodes and their structurers are depicted in Fig 4.18a and Fig 4.18b. Due to small size of the network and number of routes, only single point crossover is used with high ROR values. Let's assume that the ROR value is set to 35% and the crossover point is set to 1. At first, from each route set an individual at random is selected and removed from S_1 and S_2. Later, two new route sets namely $S_1^{'}$ and $S_2^{'}$ with the selected routes are created (see Fig 4.18c and Fig 4.18d). Once the new route sets are created, the next step is to perform a ROR-based route insertion with routes from the opposite route sets. For instance, for the $S_1^{'}$, the routes will be drawn from S_2 for ROR-based route insertion and vice versa for $S_2^{'}$. The final route sets are presented in Fig 4.18e and Fig 4.18f with all the routes following 35% route overlap constraint.

In RORMC, there are possibilities of a non-feasible route sets: 1) not all the routes are selected in the new route sets due to ROR values; 2) even after the ROR-based route insertion, there might be a possibility that the newly created routes sets are not feasible. In that case, TNI procedures are employed.

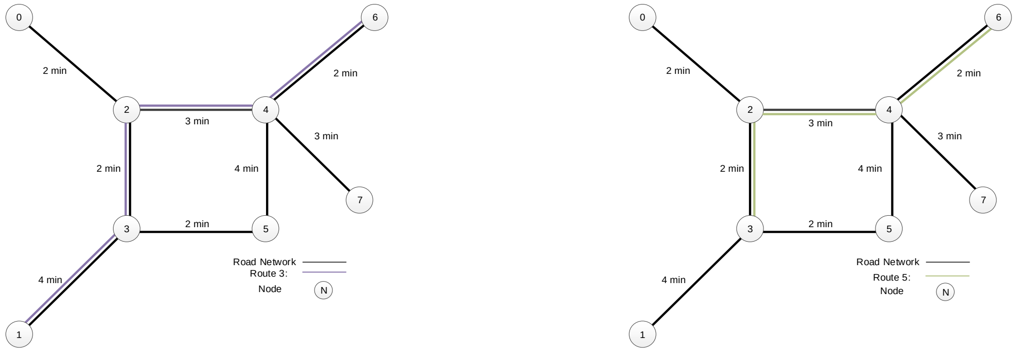


Individual	route	nodes				
S_1	1	0	2	4	7	-
	2	1	3	5	4	-
	3	1	3	2	4	6

(a) Individual 1 before RORMC

Individual	route	nodes				
S_2	4	1	3	5	4	7
	5	3	4	4	6	-
	6	0	2	4	5	-

(b) Individual 2 before RORMC

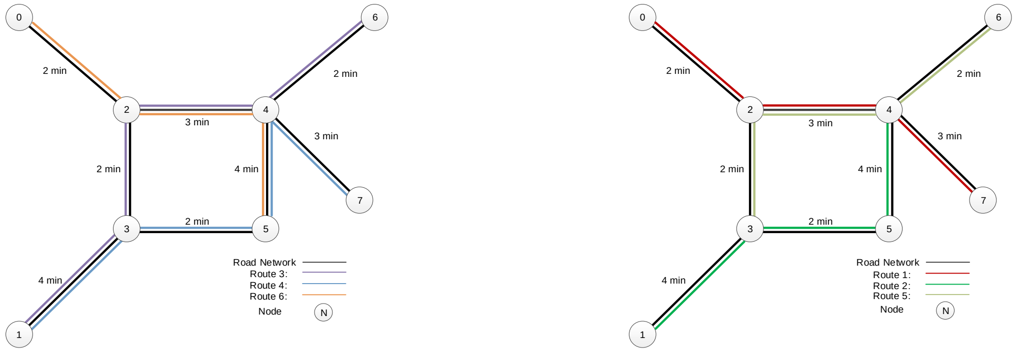


Individual	route	nodes				
S'_1	3	1	3	2	4	6

(c) Individual 1 with a selected route at random

Individual	route	nodes				
S'_2	5	3	4	4	6	-

(d) Individual 2 with a selected route at random



Individual	route	nodes				
S'_1	3	1	3	2	4	6
	4	1	3	5	4	7
	6	0	2	4	5	-

(e) Individual 1 after RORMC

Individual	route	nodes				
S'_2	5	3	4	4	6	-
	2	1	3	5	4	-
	1	0	2	4	7	-

(f) Individual 2 after RORMC

Fig 4.18 Route overlap ratio-based multipoint crossover

4.4.6.2 *Polynomial mutation*

After a successful crossover, the polynomial mutation (PM) is employed to perform small mutation to the newly created offspring. The selected PM procedure is adopted from Deb & Mayank [1996]. For each individual, at first a random value is drawn between [0,1]. If the value is below the pre-set value of mutation rate, PM is performed.

Once the individual is selected for mutation, the individual is mutated such that its mutant form resembles the original individual. The resemblance is controlled by a crowding degree (CD). Higher values of CD lead to high resembles to the original individual whereas, lower values lead to least resemblance to the original individual. The mutation in the individual (route set) is seen as the change of routes.

The procedure for PM is as follow:

1. For each route id in the individual, a random value between [0, 1] is drawn. If the value is lower than the mutation rate, the process of mutation starts. Else, the process moves to next route id in the individual.
2. For each considered route id, a random value is drawn between [0, 1]. Based on the drawn value, and crowding degree, different perturbance factors (Delta) are calculated.
3. Based on the delta factor, the mutated value is calculated, and the considered route id is replaced with the mutated value.

The algorithm for PM is shown in Alg 4.10.

```
Polynomial Mutation Algorithm
Input Individual (route set with route ids), Mrate (crossover rate), CRD (crowding degree), Up
(number of feasible routes)
For each individual in Individual:
    RN = RandomNumber(0, 1)
    IF RN <= Mrate:
        X = Individual[individual]
        Delta1 = (individual – 1) ÷ (Up – 1)
        Delta2 = (Up – individual) ÷ (Up – 1)
        RN = RandomNumber(0,1)
        MutPow = 1 ÷ (CRD + 1)
        IF RN <= 0.5:
            Xy = 1 – Delta1
            Val = 2*RN +(1 – 2*RN) * Xy ^(CRD + 1)
            Delta3 = Val ^(MutPow – 1)
        Else:
            Xy = 1 – Delta2
            Val = 2*(1 – RN) +2 * (RN – 0.5)* Xy ^(CRD+1)
            Delta3 = 1 – Val ^(MutPow)
        End IF
        X = X + Delta3 *(Up – 1)
        Max= GetMaximum(X, 1)
         X = GetMinimum(Max,Up)
        Individual[individual] = X
    End IF
Output Individual (individual after mutation operator)
```

Alg 4.10 Polynomial mutation procedure

4.4.7 Transit network improvement and feasibility check

After the reproduction process, there is a high possibility that the newly created individuals are not feasible, therefore it needs improvement to pass the TNF criteria. At first, all the individuals are subjected to TNI procedure of TNC module. The TNI is used to check: 1) if there are some missing nodes in the individual (transit network); 2) possibility of inserting new routes into the individual.

The transit network improvement and feasibility check (TNIFC) procedure is as follows:

1. For each individual, missing nodes (CheckMissingNodes) are calculated.
2. If there exist missing nodes the process of network improvement starts. Else, the process terminates.
3. The individual is subjected to TNI (ImproveRoutes) and TNF (RouteSetFeasibility) procedure.
4. If the individual becomes feasible, the process terminates, and the feasible individual is saved. Else, the individual is subjected to RORRI procedure (RORInsertion).
5. After every successful route insertion into the individual, the TNF is checked and if the individual is feasible, the process terminates, and the individual is saved.

6. In case, all the routes from FTRS are evaluated and the individual is still infeasible. In that case, the ROR is increased by a predefined increment and the process moves back to step 3.

Note: Even after the TNIFC, there might be many individuals which are not feasible. As mentioned in the CCO-based binary tournament, higher number of individuals are selected for reproduction. This excess of individuals will provide a buffer in selecting the required number of individuals for evaluation.

The procedure for TNIFC is shown in Alg 4.11.

```
Transit network improvement and feasibility check algorithm
Input Individuals (route sets), ROR (route overlap ratio), RORicr (route overlap increment),
Nodes (nodes in the network), FTRS (feasible transit route set), Detour (allowed detour limit)
For each individual in Individual:
    MissingNodes = CheckMissingNodes(individual)
    IF MissingNodes:
        While Not RouteSetComplete:
            ImproveRoutes(individual)
            IF RouteSetFeasibility(individual):
                RouteSetComplete = True
                Break
            End IF
            For each r in FTRS:
                RORInsertion(r, individual, ROR)
                IF RouteSetFeasibility(individual):
                    RouteSetComplete = True
                    Break
                End IF
            IF RouteSetComplete == False:
                ROR += RORicr
            End IF
        End While
    End IF
Output Individuals (feasible route sets)
```

Alg 4.11 Route set correction and feasibility criteria algorithm

4.4.8 Fitness evaluation

The fitness evaluation checks the fitness of the individuals. The fitness here refers to the objective functions (defined in section 3.3.2). Each individual has two fitness values: 1) fitness with respect to passenger costs (see Eq (3.16)); 2) fitness with respect to agency costs (see Eq (3.17)).

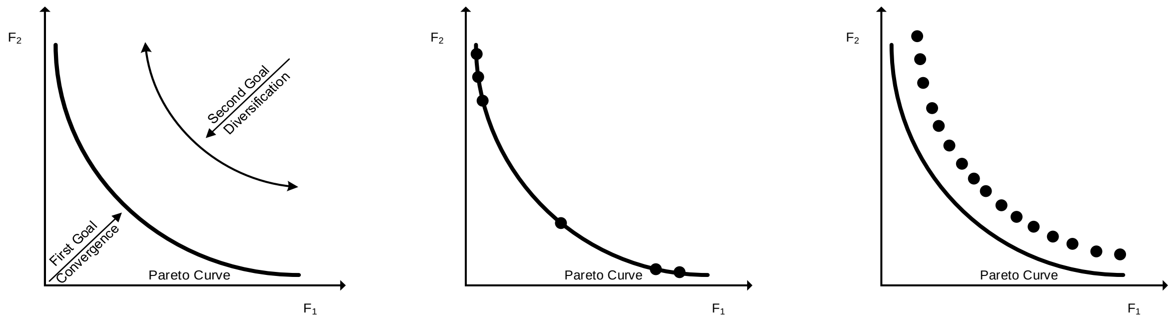
4.4.9 Termination

The termination or stopping criteria deals with the stagnation of the population, meaning it decides when the search should end. There are two variations: 1) static; 2) adaptive. In the static termination, the number of generations are known in advance. For instance, if the maximum number of generations are fixed at 100, the search will stop after 100 iterations. In the adaptive termination, the search stops when there is no improvement in the best achieved optimum value. Often a hybrid of these two procedures is used, where the search stops if either the predefined number of generations are reached, or no improvement is observed for certain generations. In the proposed methodology, the first variation is adopted due to computational cost considerations.

4.4.10 NSGA-II performance metrics

Generally, the performance of the MOEA is demonstrated by showing the nondominated solutions together with true pareto optimal solutions in the objective space. The emphasis is on how closely the obtained nondominated solutions converged to the true optimal front. There are two distinct goals of multi-objective optimisation: 1) to find a solution as close to the true pareto optimal solutions as possible; 2) to find solutions as diverse as possible in the obtained nondominated solutions. Since both these goals are conflicting with each other, as one advocates for convergence and other for diversification, therefore performance measure of the algorithm in absolute sense by a single metric is not possible.

A small example is adopted from Deb [2001], and shown in Fig 4.19. The convergence and diversity goals are depicted visually in Fig 4.19a. In the Fig 4.19b and Fig 4.19c, both goals are depicted individually. In first case, the set of nondominated solution converge on the true pareto optimal front however, with low diversity among the solutions. In the second case, the set of nondominated solution spread evenly however, they lack convergence on true pareto optimal front. This shows, fulfilling both objectives for a given MOEA algorithm is not easy.



(a) Convergence and diversity goals

(b) Convergence

(c) Diversity

Fig 4.19 Goals of multi-objective optimisation adopted from Deb [2001]

There are individual performance metrics for both the convergence and diversity. Deb [2001] compiled a few metrics for both goals with detailed explanation backed by simple examples. A brief information about these metrics is presented below.

Metrics used for measuring convergence:

1. Error ratio, simply takes a union of the solution in the pareto optimal set and the nondominated solution set.
2. Generational distance, calculates the averages distance of the solution of nondominated set to the solution in the pareto optimal set.

3. Maximum pareto optimal front error, check the worst distance among the solutions of nondominated set to the pareto optimal set.

Metrics used for measuring diversity:

1. Spacing, is calculated with the relative distance measure between consecutive solutions in the nondominated solution set.
2. Maximum spread, is calculated by measuring the length of the diagonal of a hyper box, formed by the two extreme boundary solutions values in the nondominated solution set.

Metrics used for measuring convergence and diversity:

1. Hyper volume, is calculated as the sum of hypercubes constructed with a reference point for each solution in the nondominated solution set.
2. Weighted metric, evaluates both goals by combining convergence and diversity measure along with weights for each of the measure.

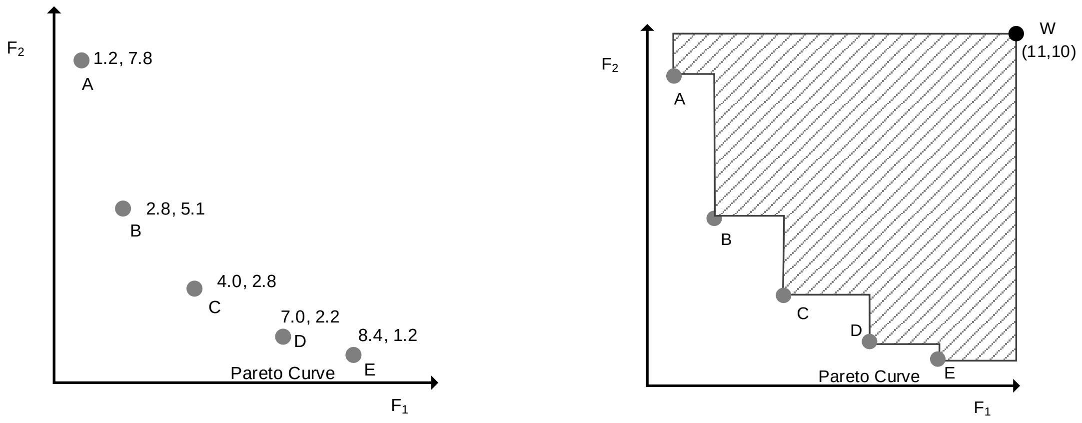
The selection of the most suitable metric/metrics depends upon the available information. Since, the true pareto optimal sets for the selected benchmark networks used in this study are not available, therefore the convergence and diversity-based metrics will not be considered. Instead, hybrid of the two metrics will be considered. Such metrics only provide qualitative measure of the convergence and diversity for given nondominated solution set. In this study, hyper volume metric is selected for the performance evaluation of the NSGA-II. The hypervolume seems most suitable as it is ideal, if the considered objectives are to be minimised.

4.4.10.1 *Normalised hyper volume*

The hyper volume (HV) calculates the area of the objective space covered by the nondominated solution set for problems where all the objectives are to be minimised. This metric can be calculated as for each solution i belong to the nondominated front Q, the 2-dimensioanl hyper cube c_i is constructed with reference point W and the solution i .The reference point must be worst then all the point in the nondominated solution set. Th union of all the hypercubes forms the HV.

$$\mathrm{HV} = Volume\left(\bigcup_{i \in Q} c_i\right) \tag{4.34}$$

A small example of HV is shown in Fig 4.20, adopted from Deb [2001]. The nondominated solution set is depicted in Fig 4.20a. For calculating the HV, first a reference point W is selected with values (11, 10). Later, hypervolume for each of the five solutions is calculated, and then combined to get the aggregated volume. The aggregated volume is depicted as shaded area (see Fig 4.20b).



(a) Nondominated solution set (b) Hyper volume of nondominated solution set

Fig 4.20 Hyper volume for a given nondominated solution set adopted from Deb [2001]

The HV is not free from subjective scaling of objectives therefore, normalised objective function values are preferred. In general, higher values of Normalised HV are desired. The normalisation of the objective function is performed using min-max scaling. For any given set of values, the following calculations are performed to normalise the values. First each value set is sorted in ascending order. Later, a scale range s_r is set with minimum and maximum value. For each value $v_i \in V$ is transformed to nv_i by subtracting the maximum value $V.\max()$ from it , and later divided by the difference of $V.\max()$ and $V.\min()$. The normalised scaled value nsv_i is calculated by multiplying nv_i with the s_r lower and upper bound difference. Finally, the calculated value is added with the s_r lower bound.

$$nv_i = \frac{\left(v_i - V.\max()\right)}{V.\max() - V.\min()} \tag{4.35}$$

$$nsv_i = nv_i(s_{r,upper} - s_{r,lower}) + s_{r,lower} \tag{4.36}$$

4.4.10.2 *Population diversity*

In order to evaluate the diversity of the solutions, different measures can be adopted. However, it depends upon the encoding type. The proposed idea for measuring population diversity is adopted from John [2016]. Moreover, Jaccard distance (JD) is used for measuring the similarity between different solutions in the population. The JD is defined as the dissimilarity between the sample sets C and D. It can be calculated by subtracting Jaccard index (JI) from 1. The JI is simply the ratio of the size of the intersection over size of the union.

$$\mathrm{JI}\left(C, D\right) = \frac{|C \cap D|}{|C \cup D|} \tag{4.37}$$

$$\mathrm{JD} = 1 - \mathrm{JI} \tag{4.38}$$

With the help of a simple example, JD is explained in terms of route set. Let's consider a simple example based on network depicted in Fig 3.1. Let's say, there are two route sets C and D, and their route arcs as CA_{r1}, CA_{r2}, CA_{r3} and DA_{r1}, DA_{r2}, DA_{r3}, DA_{r4} respectively. The JI and JD calculations are shown in (4.46) and (4.47). The value of JD is 0.125, which suggest that both route sets are quite similar. A value of 1 means, both route sets share no common route arc among each other.

$$CA_{r1} = \{(n_0,n_2),(n_2,n_4),(n_4,n_6)\} \tag{4.39}$$

$$CA_{r2} = \{(n_1,n_3),(n_3,n_2),(n_2,n_4),(n_4,n_6)\} \tag{4.40}$$

$$CA_{r3} = \{(n_3,n_5),(n_5,n_4),(n_4,n_7)\} \tag{4.41}$$

$$DA_{r1} = \{(n_1,n_3),(n_3,n_5),(n_5,n_4),(n_4,n_6)\} \tag{4.42}$$

$$DA_{r2} = \{(n_0,n_2),(n_2,n_4),(n_4,n_7)\} \tag{4.43}$$

$$C \cap D = \{(n_0,n_2),(n_4,n_6),(n_2,n_4),(n_1,n_3),(n_3,n_5),(n_5,n_4),(n_4,n_7)\} \tag{4.44}$$

$$C \cup D = \{(n_0,n_2),(n_4,n_6),(n_2,n_4),(n_3,n_2),(n_1,n_3),(n_3,n_5),(n_5,n_4),(n_4,n_7)\} \tag{4.45}$$

$$\mathrm{JI}\left(C,D\right) = \frac{7}{8} \tag{4.46}$$

$$\mathrm{JD} = 1 - \frac{7}{8} = 0.125 \tag{4.47}$$

4.4.10.3 *Route overlap index*

The ROI is already explained in section 4.3.1. It is used to evaluate the solutions that are created and modified during the evolution process. In the context of NSGA-II performance measure, the ROI means average ROI of all the individuals in the population. Ideally, it should maintain the desired route overlap during the whole process with slight increase in its value.

4.4.11 Summary

In the NSGA-II MSE component, the created PT networks using ROR-based concept were evolved to get improved versions of these networks. A new ROR-based crossover operator was proposed, which is consistent with the main ROR-based network construction. Moreover, concept like ROI together with performance metrics such as Normalised-HV and JD were used for the evaluation purposes.

4.5 Summary

This chapter explained different components of the proposed solution methodology with a number of unique features such as ROR-based network construction, iterative capacity constrained demand assignment with multimodal vehicle selection. Moreover, a holistic PT network evaluation, incorporating the perspectives of different stakeholders was also proposed. Lastly, the selection of MOEA-based search engine was used to solve the TNSFSP with custom GA operators. In the next two chapters, the implementation of the proposed methodology on benchmark and the real networks will be presented.

5 Implementation of Proposed Set of Procedures on Benchmark Networks

The proposed methodology is implemented on several benchmark networks. First the parameters and objective settings for each of the benchmark network will be briefly explained. Followed by the results and comparison with other related studies. Finally, sensitivity analysis will be presented for each network instance.

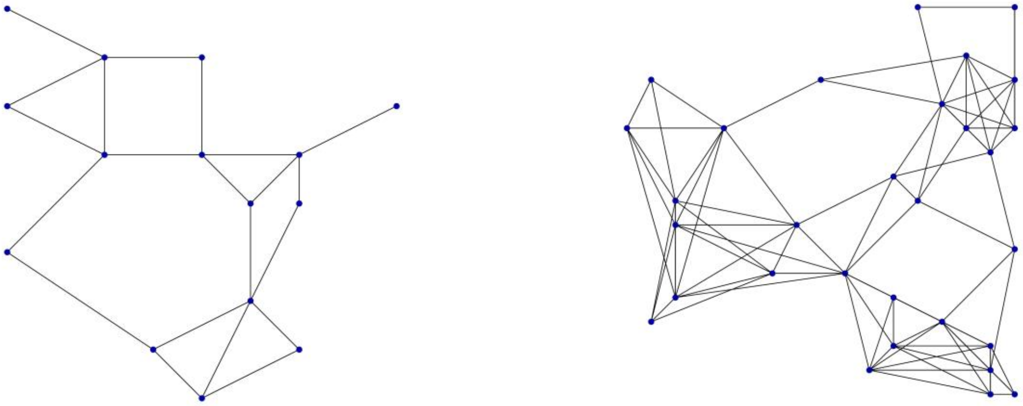
5.1 Network instances

The proposed algorithm is implemented on a five benchmarks networks including one from Mandl (Mandl [1980]) and rest of them from Mumford (Mumford [2013]). Two of the four networks (Mumford2, Mumford3) are derived from the real cities (Hubei, China, Cardiff, United Kingdom (UK)). There are three data sources for each benchmark network: 1) the daily symmetrical transit demand matrix; 2) travel time matrix; 3) coordinates of the nodes. The basic information such as the number of nodes and arcs for all five networks is presented in Tab 5.1.

Benchmark network	Network Properties	
	Number of nodes	Number of links
Mandl	15	21
Mumford0	30	90
Mumford1	70	210
Mumford2	110	385
Mumford3	127	425

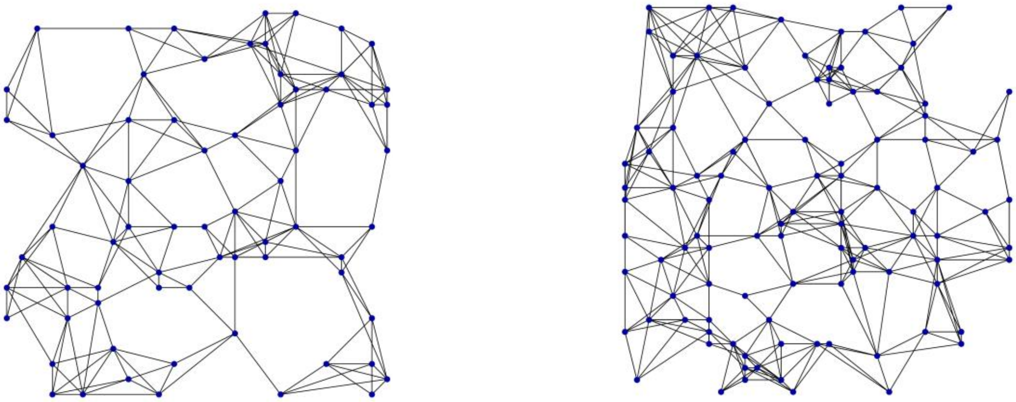
Tab 5.1 Properties of different networks

The network geometries for all five benchmark networks are depicted in Fig 5.1.



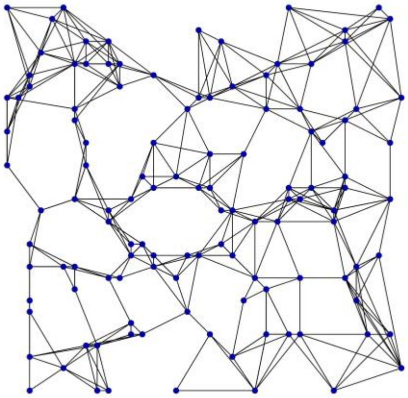
Mandl

Mumford0



Mumford1

Mumford2



Mumford3

Fig 5.1 Network geometries

5.2 Experimental parameters

5.2.1 Basic parameters

The parameters that remain constant for the rest of the experiments such as the population size, ROR rates and feasible route sets are shown in Tab 5.2. However, those parameters which change according to the network type and experiment, will be mentioned explicitly.

Type	Parameters	Value
NSGA-II	Population size	100 individuals
	Generations	100 generations
	Mutation rate	0.005
	Mutation crowding degree	0.5
	Crossover rate	0.9
ROR	ROR initial	10%
	ROR step size	10%
	ROR maximum	50%
Feasible route set	Detour maximum	1.5 times shortest path
	Maximum routes per node pair	5000 per node pair
Route	Minimum headway	2 minutes
	Maximum headway	30 minutes
	Maximum passenger load	1.0
	Terminal time coefficient	0.10%
Objective function parameters passenger costs	Origin wait time coefficient	1.0
	In-vehicle time coefficient	1.0
	Transfer wait time coefficient	1.0
	Transfer penalty	5 minutes
Reproduction buffer	Crossover population	2 times population size
	Feasible solution pool sample	2 times population size
Demand assignment	ITDA	First 50 generations
	ITDA + ICCTDA	Last 50 generations

Tab 5.2 Parameters used in the experiments

5.2.2 Operational cost parameters

The operational cost parameters used for agency costs such as on-vehicle crew cost, vehicle operating cost, infrastructure management and overhead costs for different types of modes are adopted from Australian Transport Council [2006]. The original values are updated to year 2017 with average inflation rate of 2.4% Reserve Bank of Australia [2017] over 11 years. The total change in the cost is 29.4%. These operational cost parameters are presented in Tab 5.3.

Mode	On vehicle crew cost (AUD$/hour)	Vehicle direct operating cost (AUD$/km)	Infrastructure operations and maintenance cost (AUD$/km)	Overhead operating costs (% on other operational costs)	Total operational cost (AUD$)
Minibus	42.7	1.0	0.0	21	52.9
Standard bus high floor	42.7	1.3	0.0	21	53.2
Standard bus low floor	42.7	1.3	0.0	21	53.2
Articulated bus-1	42.7	1.5	0.0	21	53.5
Articulated bus-2	42.7	1.5	0.0	21	53.5
High capacity bus	42.7	1.5	0.0	21	53.5
Streetcar ROW C	77.6	1.9	11.6	17.5	107.0
Bus rapid transit	77.6	1.9	11.6	17.5	107.0
Streetcar ROW B-1	77.6	1.9	11.6	17.5	107.0
Streetcar ROW B-2	77.6	1.9	11.6	17.5	107.0
Rapid transit-1	284.7	3.6	22.0	14	353.8
Rapid transit-2	284.7	3.6	22.0	14	353.8
Rapid transit-3	284.7	3.6	22.0	14	353.8

Tab 5.3 Operational cost parameters considered for different modes for agency costs adopted from Australian Transport Council [2006]

5.2.3 Feasible route set creation parameters

The feasible routes are created for all five-benchmark networks. Two conditions were applied to transit route construction: 1) maximum routes per node pair are restricted to maximum 5,000 routes; 2) the maximum detour allowed was 1.5 times the shortest path for each node pair. With these two conditions the number of routes constructed for each network and time taken are listed in Tab 5.4. The route construction is performed on an Intel Xeon CPU E5-2640 v3 @ 2.6GHz with 128 GB RAM. Moreover, parallel processing is adopted to reduce the computational time.

Benchmark network	Without parallel processing	With parallel processing	Feasible routes
Mandl	1s	1s	773
Mumford0	1,270s	68s	182,996
Mumford1	64,020s	2,520s	3,986,010
Mumford2	343,903s	13,285s	15,520,930
Mumford3	754,731s	23,703s	24,622,947

Tab 5.4 Feasible route set for each benchmark network

5.3 Experimental setup

5.3.1 Experiments

In the normal experiments, first an initialisation (ROR-INI) is performed to create initial population of 100 individuals. Later, NSGA-II is run for 100 generations for each network instance. The crossover and mutation rates used in the experiments are 0.9 and 0.005 respectively. Moreover, the buffer to adjusted non-feasible routes produced during reproduction are set to 2 times the population size for the final selection. In terms of demand assignment, for the first 50 generations, ITDA is used and for the remaining 50 generations ITDA and ICCTDA are used. This means for initial 50 generations, only capacity free assignment is used while in the last 50 generations, capacity free and capacity constrained assignment are used sequentially. This is done to reduce the computation time for the experiments. The whole experiment is repeated for different ROR values, starting with 10% ROR to 50% ROR, with 10% ROR as step size. Therefore, in total five experiments will be performed for each network instance.

5.3.2 Experiments for sensitivity analysis

Sensitivity analysis will be performed to analyse different parameters and their impact on the results. These parameters include population size and generations. Moreover, the ROR value, which provides best results for most performance indicators, for each network from the main experiments will be used for sensitivity analysis. For fixed population size and varied generations. The following parameters are set:

- The population size is set to 100 individuals.
- The generations are varied from 100 to 400 with a step size of 100.
- ITDA is used for first half of the generations, e.g., for 200 generations (first 100 with ITDA).
- ITDA + ICCTDA is used for the last half of the generations (last 100 with ITDA + ICCTDA).

For fixed generations and varied population size. The following parameters are set:

- The generations are set to 100 generations.
- The population size is varied from 100 to 400 with a step size of 100.
- ITDA is used for first 50 generations.
- ITDA + ICCTDA is used for the last 50 generations.

5.3.3 Computation time

The computation time for all the experiments for benchmark networks including the ones used in sensitivity analysis are listed in Tab 5.5 in seconds. For the Mandl, Mumford0, Mumford1 and Mumford2 networks, Intel Core i7-6600U CPU @ 2.60GHz with 24 GB RAM, and for Mumford3 network, Intel Xeon CPU E5-2640 v3 @ 2.6GHz with 128 GB RAM was used. PTV Visum Expert version 17.01.08 64 Bit was used for TDAHD component.

5.3.4 Data Files

The data files include:1) raw input files for networks; 2) results from experiments and sensitivity analysis; 3) comparative results; 4) spreadsheets for results; 5) PTV Visum version files. All these files can be accessed from Ul Abedin [2019].

Benchmark network	Population	Generation	Route overlap Ratio				
			10%	20%	30%	40%	50%
Mandl	100	100	34,394s	33,884s	34,419s	35,681s	36,016s
	100	200	-	-	-	47,729s	-
	100	300	-	-	-	70,269s	-
	100	400	-	-	-	95,190s	-
	200	100	-	-	-	52,003s	-
	300	100	-	-	-	81,720s	-
	400	100	-	-	-	113,847s	-
Mumford0	100	100	29,966s	28,068s	27,727s	31,095s	29,089s
	100	200	-	-	-	58,262s	-
	100	300	-	-	-	86,485s	-
	100	400	-	-	-	293,963s	-
	200	100	-	-	-	63,112s	-
	300	100	-	-	-	99,318s	-
	400	100	-	-	-	134,388s	-
Mumford1	100	100	52,004s	45,956s	47,554s	53,533s	55,033s
	100	200	115,480s	-	-	-	-
	100	300	133,899s	-	-	-	-
	100	400	179,048s	-	-	-	-
	200	100	103,175s	-	-	-	-
	300	100	161,451s	-	-	-	-
	400	100	223,956s	-	-	-	-
Mumford2	100	100	103,437s	99,912s	109,356s	112,349s	116,163s
	100	200	197,150s	-	-	-	-
	100	300	287,333s	-	-	-	-
	100	400	393,636s	-	-	-	-
	200	100	174,536s	-	-	-	-
	300	100	254,761s	-	-	-	-
	400	100	335,573s	-	-	-	-
Mumford3	100	100	216,676s	209,703s	217,771s	227,395s	231,424s
	100	200	418,092s	-	-	-	-
	100	300	383,047s	-	-	-	-
	100	400	179,048s	-	-	-	-
	200	100	463,287s	-	-	-	-
	300	100	713,894s	-	-	-	-
	400	100	335,573s	-	-	-	-

Tab 5.5 Computation time of experiments for benchmark networks in seconds

5.4 Benchmark networks

The proposed methodology is applied on several benchmark networks such as Mandl network and Mumford set of networks (four networks). The details about the networks can be found in Mumford [2013].

5.4.1 Mandl network

The Mandl network is perhaps the most widely used benchmark network for TNDFSP. This network was originally proposed by Mandl [1980], since then, it has been used by more than 20 studies for evaluating their network design methodologies.

For Mandl network, following parameters are set:

- The multipoint for crossover is set to 2 points.
- The number of iterations for ITDA is set to 4 iterations.
- The number of iterations for ICCTDA is set to 4 iterations.

5.4.1.1 *Results*

The results for Mandl network with different ROR values are listed in the Tab 5.6 together with the rest of the TNE indicators. The main objective function values (ATT and OC) are highlighted in grey colour. Moreover, the best values for all indicators are listed in boldface.

The results show that the best passenger optimum solution is achieved with a ROR of 40%, whereas the best agency optimum is achieved with a ROR of 10%. The rest of the indicators such as ESH, ROI and TNL show best values with ROR of 10% and 20% whereas, ANT shows the least values with higher ROR values such as 30% and 40%. Indicators such as TND and RSC show best values for both lower and higher ROR values.

ROR	Perspectives of stakeholders	Performance indicators								
		ROI	ESH	ANT	RSC	UDP	TNL	ATT	OC	TND
	Passenger	■		■		■		■		■
	Agency	■	■		■		■		■	■
	Authority	■	■	■	■	■	■	■	■	■
	Best solutions									
10%	Passenger	1.13	1759	0.14	5000	**0**	49	12.6	5716	1.17
	Agency	1.03	**1194**	0.27	4900	**0**	**32**	15.6	**3647**	1.05
	Compromise	1.06	1196	0.25	4700	**0**	43	14.3	4112	1.08
20%	Passenger	1.31	1804	0.12	4960	**0**	68	12.3	6718	1.17
	Agency	1.03	1723	0.22	5540	**0**	**32**	15.1	3791	1.05
	Compromise	**1**	1330	0.22	**4600**	**0**	42	13.4	4325	**1.02**
30%	Passenger	1.6	1759	**0.08**	5040	**0**	81	12.3	6895	1.15
	Agency	1.07	2327	0.19	6080	**0**	40	14.6	4375	1.2
	Compromise	1.06	1387	0.13	4660	**0**	42	13.3	4554	1.11
40%	Passenger	1.63	1726	**0.08**	4880	**0**	91	**12.1**	7832	1.22
	Agency	1.11	1668	0.22	5240	**0**	35	14.3	3936	1.03
	Compromise	1.07	1271	0.23	**4600**	**0**	39	13.6	4291	**1.02**
50%	Passenger	1.54	1895	0.09	5120	**0**	72	12.4	6118	1.29
	Agency	1.1	1779	0.19	5380	**0**	39	14.2	4088	1.1
	Compromise	1.05	1913	0.14	5320	**0**	40	13.4	4636	1.09

Tab 5.6 Objective function values and performance indicators for Mandl network with different ROR values

In terms of the performance of the NSGA-II, three different indicators are used (see section 4.4.10). Among these, the main unary indicator for measuring the performance of the NSGA-II is NHV. This represents the convergence and diversity of the obtained nondominated front. The performance measure indicators related to NSGA-II for different ROR values along with total feasible solutions and pareto fronts are shown in Fig 5.2. The ROI shows that the overlap remains consistent with different ROR values throughout the evolution process. For the population diversity, the JD distance shows values higher than 0.50, which shows that the diversity among the individuals remains high during the evolution process. The NHV also shows steady high values, during the evolution process.

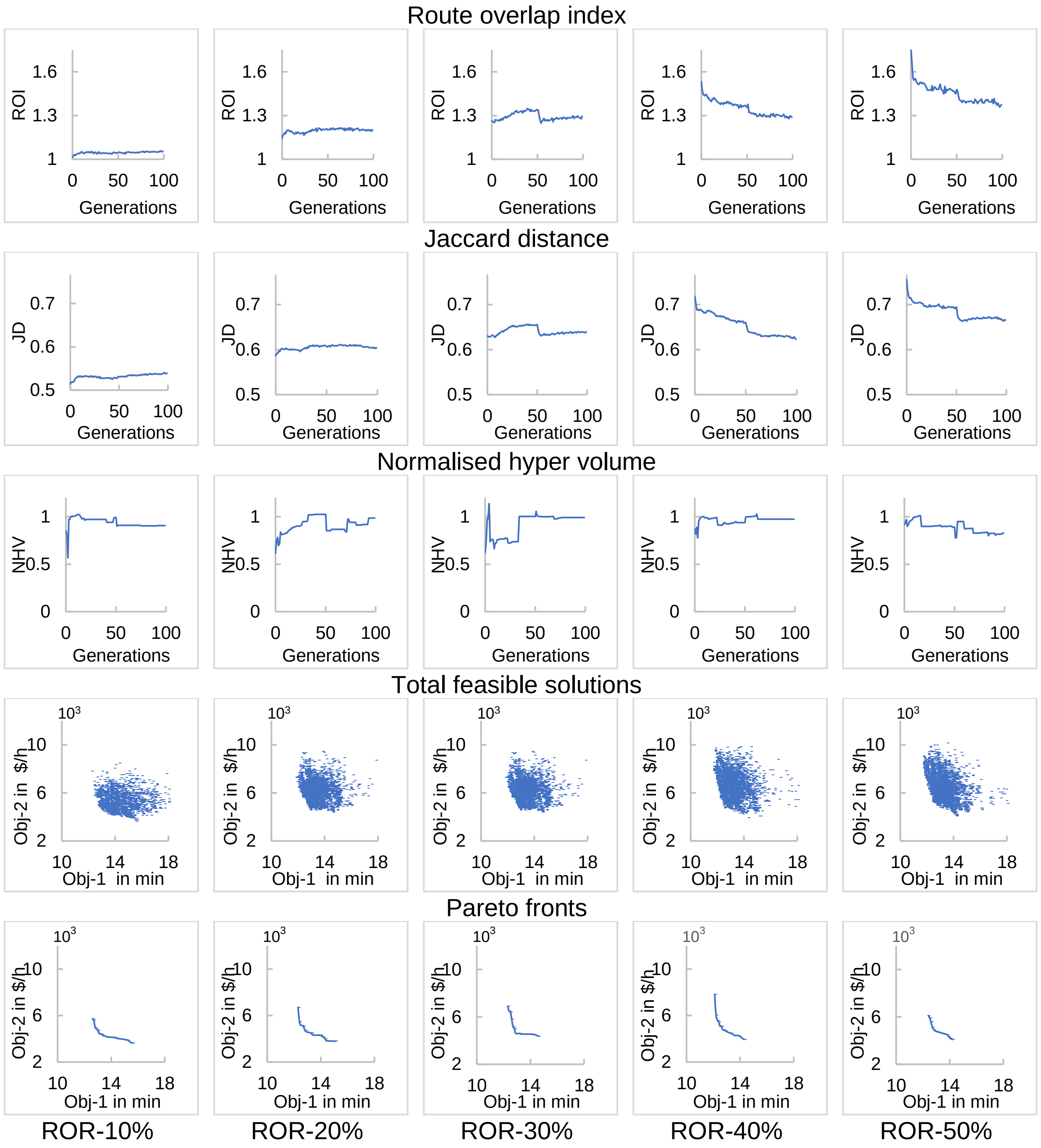


Fig 5.2 Indicators related to NSGA-II performance (ROI, JD and NHV), feasible solutions and pareto fronts for Mandl network with different ROR values

5.4.1.2 *Comparison with other studies*

The proposed algorithm is compared with solutions generated by other studies for Mandl network. For fair comparison, all the solutions (route sets) are subjected to the procedures of the developed TDAHD and TNE components. The results are listed in Tab 5.7. The best solutions generated by proposed methodology for passenger, agency and compromise are listed at the end of the table and highlighted in grey colour. Moreover, the best objective function values among all comparative solutions are listed in boldface. The developed methodology achieves one of the best passenger (12.1, 7832) and agency optimum solutions (15.6, 3647). However, the best passenger optimum solution (12, 7283) comes from Nikolić & Teodorović [2014] and the best agency optimum solution (16.6, 3434) comes from Mumford [2013].

Study	No of Routes	Solutions	ATT	OC
Mandl [1980]	4	Compromise	14.5	4,792
Baaj & Mahmassani [1991]	6	Compromise	14.3	7,241
	7	Compromise	14.3	5,883
	8	Compromise	14.5	5,973
Shih & Mahmassani [1994]	6	Compromise	13.7	5,992
	7	Compromise	14.3	5,883
	8	Compromise	13.6	6,397
Kidwai [1998]	4	Passenger-optimum	14.3	5,047
	6	Passenger-optimum	13.6	6,799
	7	Passenger-optimum	13.6	5,259
	8	Passenger-optimum	13.6	6,397
Gundaliya, Shrivastava & Dhingra [2000]	4	Compromise	13.3	6,396
	5	Compromise	13.3	6,706
	6	Compromise	13.4	6,590
Chakroborty & Wivedi [2002]	4	Passenger-optimum	13.4	7,556
Fan & Mumford [2008]	4	Passenger-optimum	14	8,194
	6	Passenger-optimum	15.2	7,685
	7	Passenger-optimum	12.9	8,648
	8	Passenger-optimum	13.3	9,448
Fan, Mumford, & Evans [2009]	4	Passenger-optimum	12.7	5,759
	4	Agency-optimum	15.7	3,854
	6	Passenger-optimum	12.4	5,975
	6	Agency-optimum	15.6	3,742
	7	Passenger-optimum	12.8	6,495
	7	Agency-optimum	16	3,571
	8	Passenger-optimum	12.8	7,095
	8	Agency-optimum	17.7	3,839
Zhang, Lu, & Fan [2010]	4	Passenger-optimum	13.9	6,368
	4	Agency-optimum	15.7	3,854
	4	Compromise	13	5,793
	6	Passenger-optimum	14.3	9,450

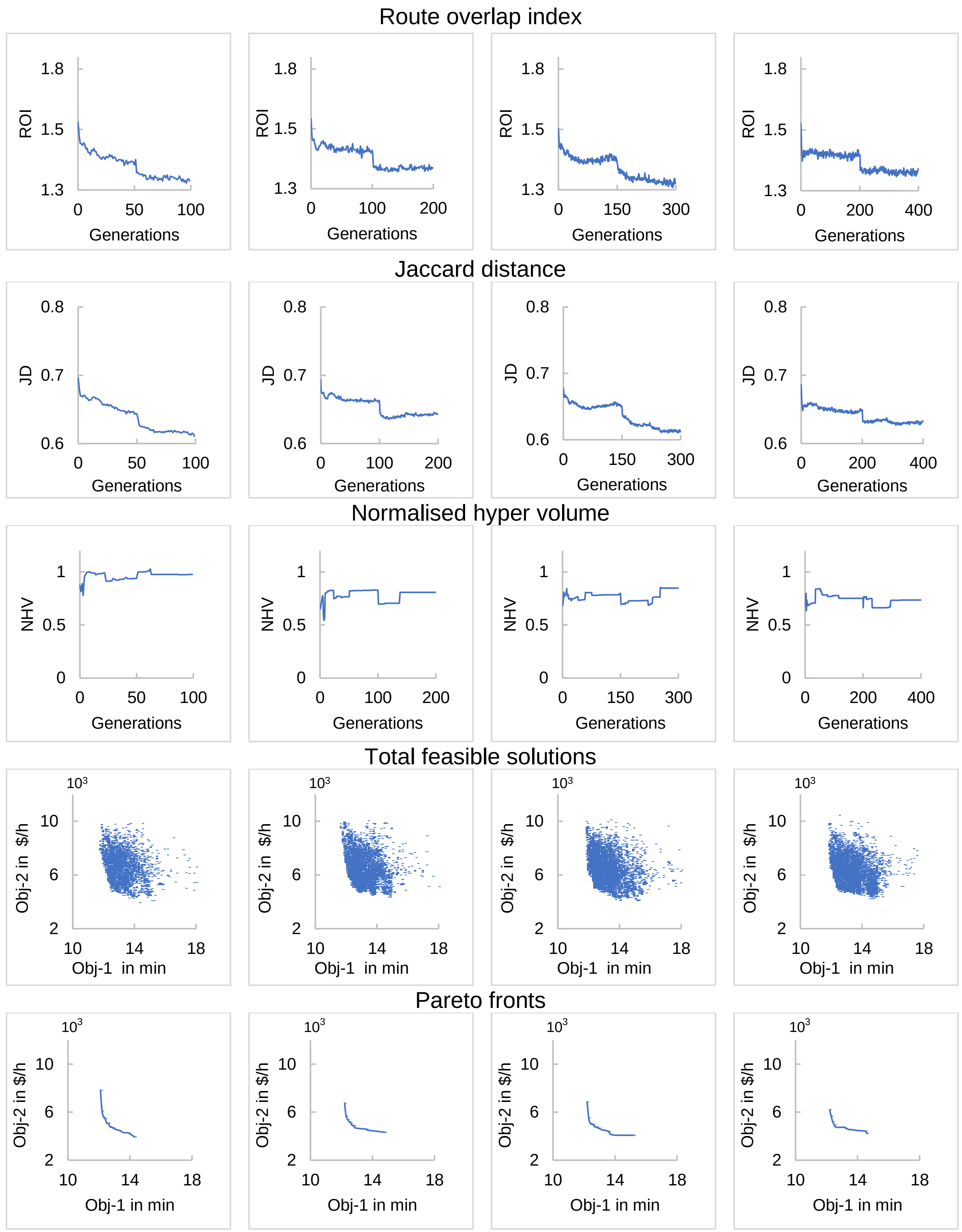
Study	No of Routes	Solutions	ATT	OC
	6	Agency-optimum	15.6	3,742
	6	Compromise	12.9	5,237
	7	Passenger-optimum	12.8	8,410
	7	Agency-optimum	16	3,571
	7	Compromise	13.3	6,169
	8	Passenger-optimum	15.7	9,566
	8	Agency-optimum	17.7	3,839
	8	Compromise	13.5	7,114
Bagloee & Ceder [2011]	12	Passenger-optimum	14.5	6,753
Mumford [2013]	4	Passenger-optimum	12.9	6,345
	4	Agency-optimum	15.7	3,854
	6	Passenger-optimum	12.7	7,492
	6	Agency-optimum	15.3	3,662
	7	Passenger-optimum	13.1	7,812
	7	Agency-optimum	16.2	3,827
	8	Passenger-optimum	13.3	7,223
	8	Agency-optimum	16.6	**3,434**
Nikolic & Teodorovic [2013]	4	Passenger-optimum	14.1	6,279
	6	Passenger-optimum	13.3	7,518
	7	Passenger-optimum	13.3	8,107
	8	Passenger-optimum	12.6	7,771
Chew, Lee, & Seow [2013]	4	Passenger-optimum	12.7	5,759
	4	Agency-optimum	15.7	3,854
	6	Passenger-optimum	12.4	5,975
	6	Agency-optimum	15.6	3,742
	7	Passenger-optimum	12.8	6,495
	7	Agency-optimum	16	3,571
	8	Passenger-optimum	12.8	7,095
	8	Agency-optimum	17.7	3,839
Owais & Moussa [2014]	6	Passenger-optimum	14.5	7,887
Majima, Takadama, Watanabe, & Katuhara [2014]	6	Compromise	13.1	6,306
Kechagiopoulos & Beligiannis [2014]	4	Passenger-optimum	12.9	6,460
	4	compromise	15.1	4,831
	6	Passenger-optimum	12.5	7,337
	7	Passenger-optimum	13.2	7,592
	8	Passenger-optimum	13.3	7,191
Nayeem, Rahman, & Rahman [2014]	4	Passenger-optimum	12.9	9,238
	6	Passenger-optimum	13.5	9,981
	7	Passenger-optimum	12.5	9,305
	8	Passenger-optimum	13.7	8,452
Nikolic & Teodorovic [2014]	4	Passenger-optimum	**12**	7,283
	4	Agency-optimum	12.8	4,934
	6	Passenger-optimum	14	8,657

Study	No of Routes	Solutions	ATT	OC
	6	Agency-optimum	12.9	5,096
	7	Passenger-optimum	12.7	8,438
	7	Agency-optimum	13	8,535
	8	Passenger-optimum	13.4	8,631
	8	Agency-optimum	13.4	6,152
Kilic & Gok [2014]	4	Compromise	12.3	7,185
	6	Compromise	13.7	7,151
	7	Compromise	13.4	7,436
	8	Compromise	13.5	7,800
Arbex & Cunha [2015]	4	Compromise	13.4	7,377
	6	Compromise	12.6	7,209
	7	Compromise	13.5	7,521
	8	Compromise	13.1	6,439
ROR-based best solutions	5	Passenger-optimum	12.1	7,832
	4	Agency-optimum	15.6	3,647
	4	Compromise	13.6	4,291

Tab 5.7 Best proposed solution quality comparison with other solutions for Mandl

5.4.1.3 *Sensitivity analysis*

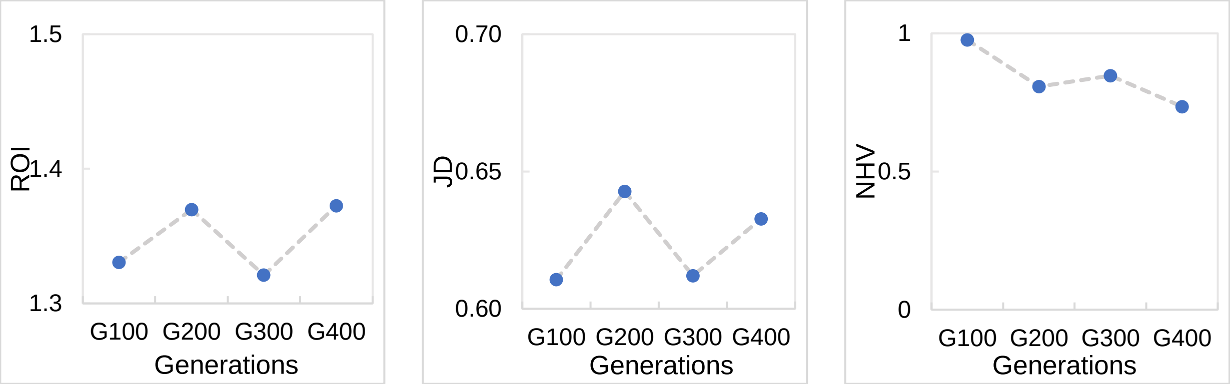
For sensitivity analysis of Mandl network with fixed population size, the ROR value is set to 40%. The ROI, JD, NHV, total feasible solutions and pareto fronts are listed in Fig 5.3. The results are consistent with the main experiment for Mandl network. Similar trends are witnessed with slight drop for all three indicators when, the assignment procedure is shifted from ITDA to ITDA + ICCTDA. Moreover, the number of solutions increases with an increase in the number of generations.



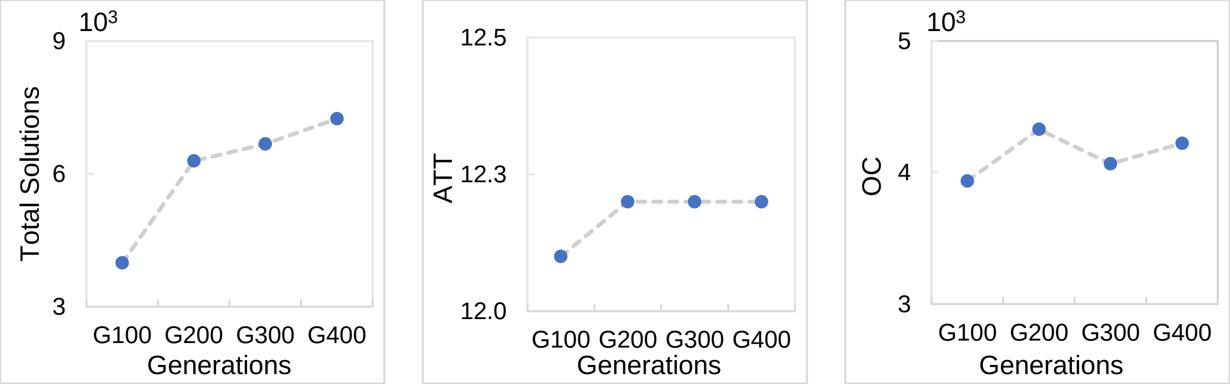
Pop100-Gen100 Pop100-Gen200 Pop100-Gen300 Pop100-G400

Fig 5.3 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mandl network with ROR-40% with fixed population size and varied generations

The trends for different generations are listed in Fig 5.4. The results show that, an increase in the number of generations does not assure improvement in the performance metrics. All performance metrics (ROI, JD and NHV) as well as objective function values (ATT and OC) show a degradation in the values. Only the total number of feasible solutions show a significant increase.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.4 Effect of varied generations with fixed population size for Mandl with ROR-40%

For sensitivity analysis for Mandl network with fixed generations, the ROR is set to 40%.The ROI, JD, NHV, total feasible solutions and pareto fronts are listed in Fig 5.5, The results are also consistent with the main experiment for Mandl network. Similar trends are witnessed with slight drop for all three indicators when the assignment procedure is shifted from ITDA to ITDA + ICCTDA. Moreover, the number of solutions increases with an increase in the number of generations.

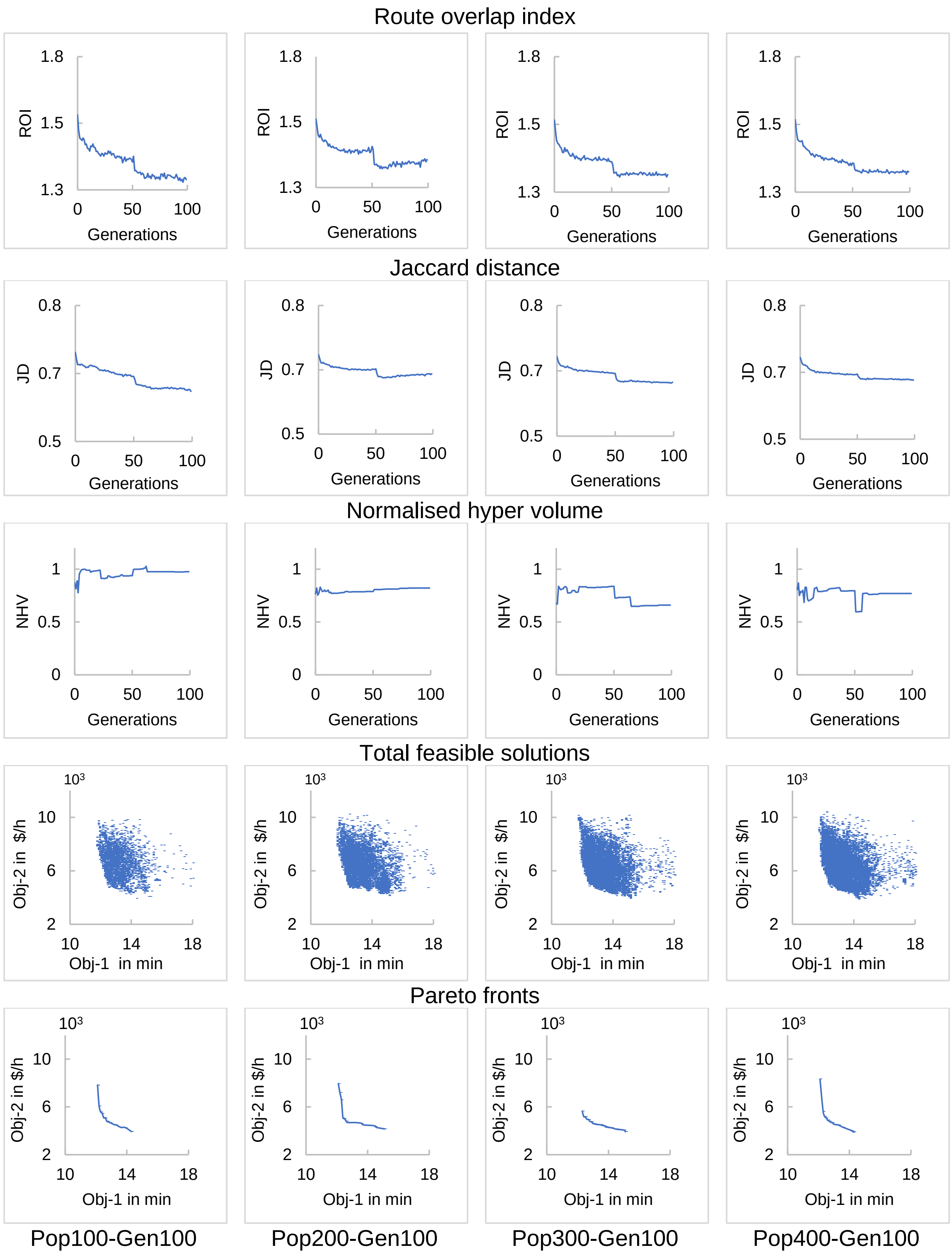
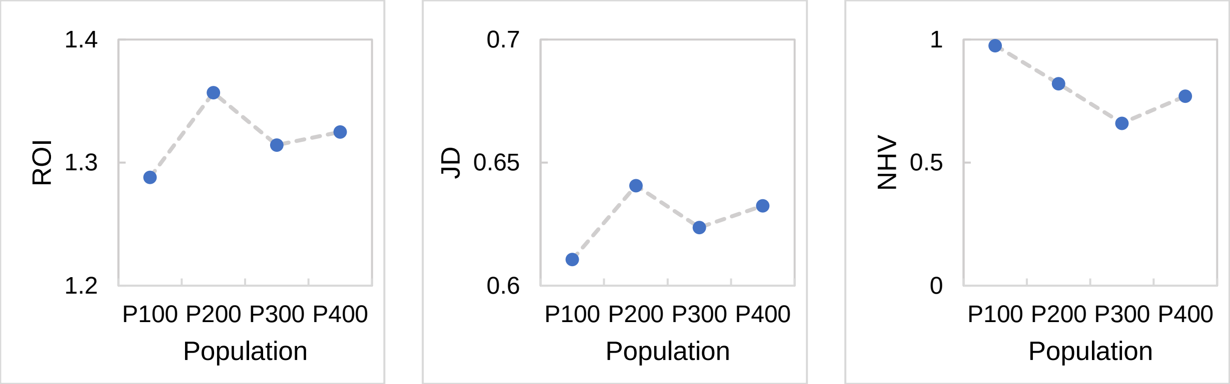
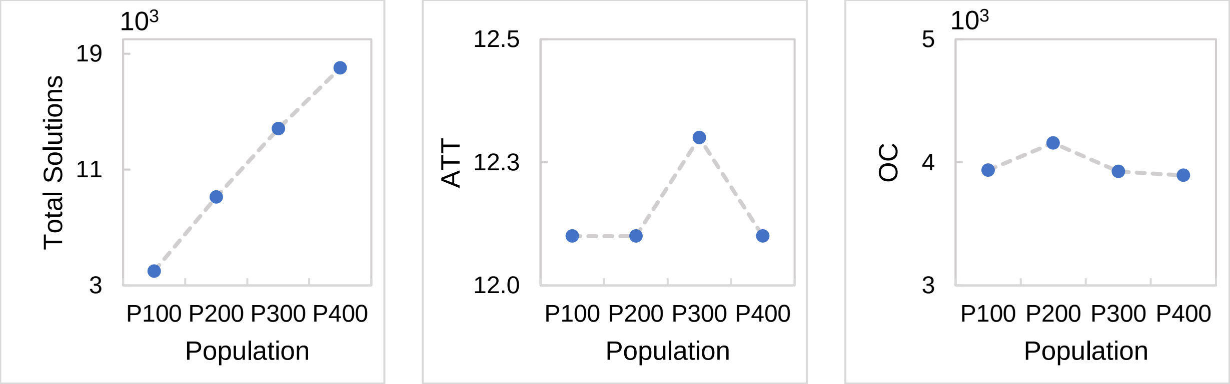


Fig 5.5 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mandl network with ROR-40% with fixed generations and varied population size

The trends for different population sizes are listed in Fig 5.6, The results show that an increase in the size of population leads to a slight improvements in ROI and JD values and a decline in NHV value. The total feasible solutions grew significantly, whereas, the ATT and OC remain consistent except for population size 300, where the ATT shows a slight increase.



Route overlap index   Jaccard distance   Normalised hyper volume



Total feasible solutions   Average travel time   Operating cost

Fig 5.6 Effect of varied population size with fixed generations for Mandl with ROR-40%

5.4.2 Mumford0

The Mumford0 network is the smallest networks among Mumford set of networks.

For Mumford0 network, following parameters are set:

- The multipoint for crossover is set to 2 points.
- The number of iterations for ITDA is set to 5 iterations.
- The number of iterations for ICCTDA is set to 5 iterations.

5.4.2.1 *Results*

The results for Mumford0 network with different ROR values are listed in Tab 5.8 together with the rest of the TNE indicators. The best solution for passenger optimum is acquired with a ROR of 50% and the best solution for agency is acquired with a ROR of 10%.

ROR	Perspectives of stakeholders	Performance indicators								
		ROI	ESH	ANT	RSC	UDP	TNL	ATT	OC	TND
	Passenger	■		■		■		■		■
	Agency	■	■		■		■		■	■
	Authority	■	■	■	■	■	■	■	■	■
	Best solutions									
10%	Passenger	1.13	2531	0.53	6040	**0**	102	19.8	9722	1.36
	Agency	**1.03**	2279	0.69	6200	**0**	66	22.7	**7503**	1.33
	Compromise	1.04	2870	0.55	6620	**0**	85	20.8	8290	1.33
20%	Passenger	1.27	2702	0.44	6200	**0**	133	19.5	9939	1.32
	Agency	**1.03**	2794	0.63	6680	**0**	70	21.9	7829	1.38
	Compromise	1.16	**2078**	0.54	**5520**	**0**	108	20.2	8939	1.28
30%	Passenger	1.42	2800	**0.38**	6520	**0**	123	19.8	10407	1.31
	Agency	1.07	2231	0.66	6260	**0**	**62**	22.9	7522	1.24
	Compromise	1.16	2510	0.54	6360	**0**	85	21.4	8404	1.36
40%	Passenger	1.52	2985	0.47	6520	**0**	140	19.5	10405	1.37
	Agency	1.27	2392	0.54	6340	**0**	83	22.4	7594	1.36
	Compromise	1.22	3066	0.51	6800	**0**	100	20.9	8266	1.39
50%	Passenger	1.2	2676	0.43	6200	**0**	111	**19.4**	10002	**1.22**
	Agency	1.16	2552	0.6	6560	**0**	67	22.4	7695	1.37
	Compromise	1.31	2422	0.49	6020	**0**	108	20.6	8811	1.32

Tab 5.8 Objective function values and performance indicators for Mumford0 network with different ROR values

The performance measure indicators related to NSGA-II for different ROR along with total feasible solutions and pareto fronts are shown in Fig 5.7. The ROI shows that the overlap increases slightly with higher ROR values. The JD distance shows values higher than 0.50 for all ROR values, which shows higher diversity among the individuals. The NHV increases as the evolution progresses in all the experiments, and it hovers around 0.6 during the last iterations.

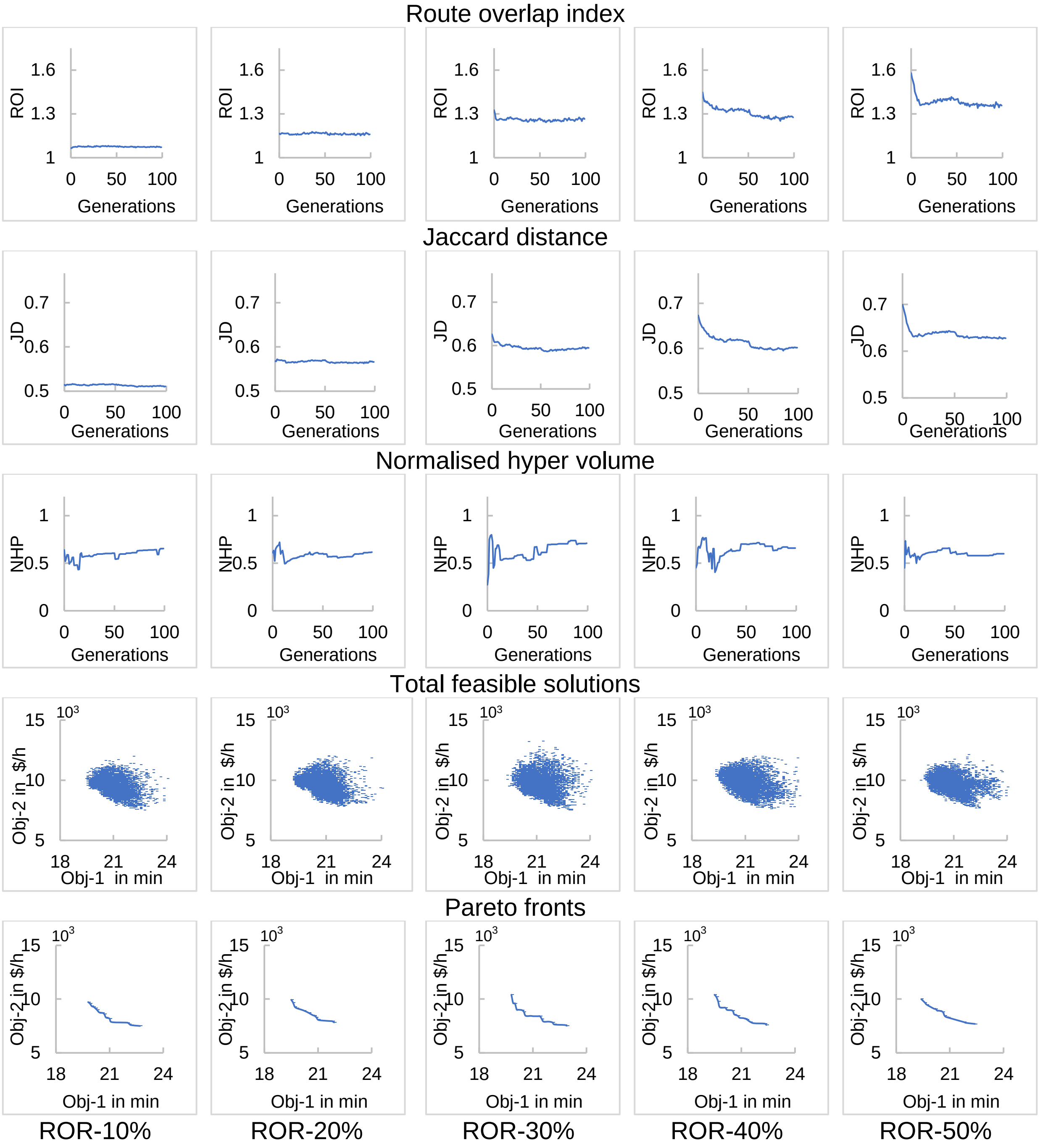


Fig 5.7 Indicators related to NSGA-II performance (ROI, JD and NHV), feasible solutions and pareto fronts for Mumford0 network with different ROR values

5.4.2.2 *Comparisons with other studies*

The solutions created from proposed methodology are compared with solutions generated by other studies for Mumford0 network. The results are listed in Tab 5.9 with the best value for passenger optimum solution (19.3, 12198) is attained by Mumford [2013], whereas the agency optimum solution is attained by the proposed ROR-based solution (22.7, 7503). The best solution for the passenger optimum is by proposed methodology offers just a fraction higher cost (19.4, 10002) compared to the best solution.

Study	No of Routes	Solutions	ATT	OC
Mumford [2013]	12	Passenger-Optimum	21	12,495
	12	Agency-Optimum	34.6	7,543
Kilic & Gok [2014]	12	Compromise	**19.3**	12,198
ROR based solutions	7	Passenger-Optimum	19.4	10,002
	4	Agency-Optimum	22.7	**7,503**
	5	Compromise	20.8	8,290

Tab 5.9 Best proposed solution quality comparison with other solutions for Mumford0

5.4.2.3 *Sensitivity analysis*

For sensitivity analysis of Mumford0 network with fixed population size, the ROR value is set to 40%. The ROI, JD, NHV, feasible solutions and pareto fronts are listed in Fig 5.8. The results are consistent with the main experiment for Mumford0 network. The number of solutions generated for different generations show that the spread of feasible solutions increases with higher generations.

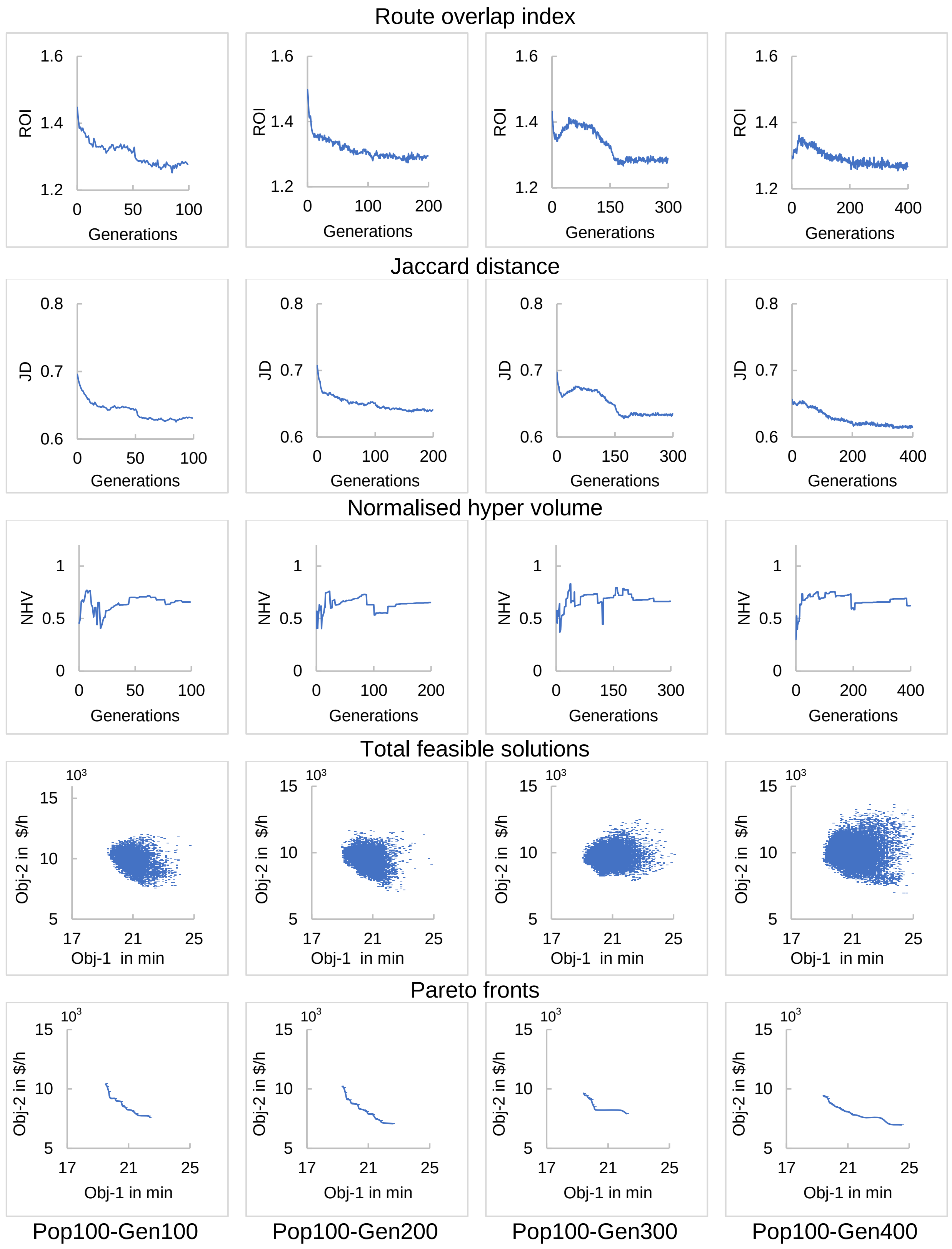
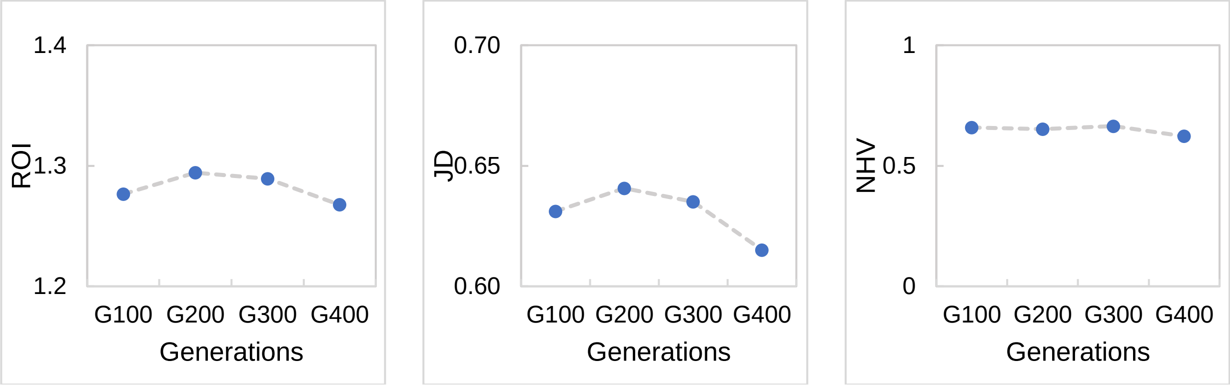
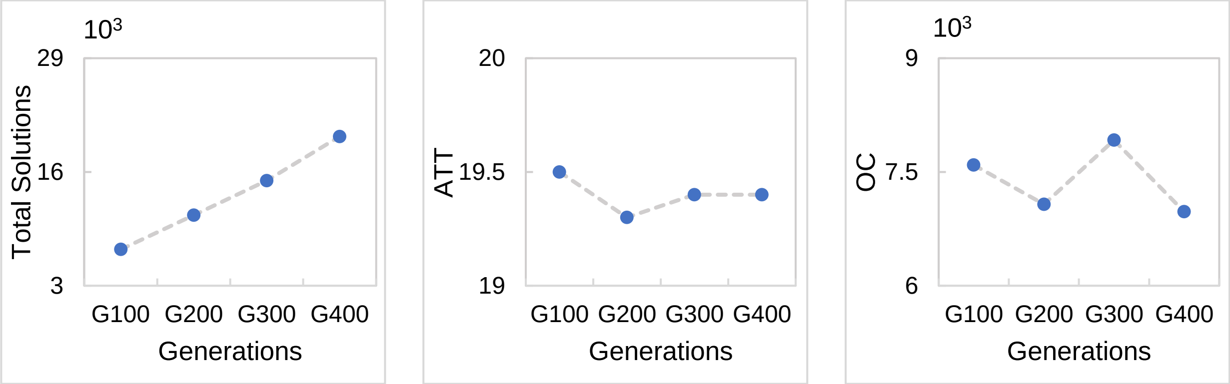


Fig 5.8 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford0 network with ROR-40% with fixed population size and varied generations

The trends for different generations for Mumford0 are listed in Fig 5.9. The ROI and JD show a slight decrease with increase in the generation number. However, the NHV remain stable and doesn't show any significant increase or decrease. The total number of feasible solutions increase significantly for higher generation number. The ATT and OC remain stable with no significant increase or decrease in their values, except for OC for generation 300th.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.9 Effect of varied generations with fixed population size for Mumford0 with ROR-40%

For sensitivity analysis for Mumford0 network with fixed generations, the ROR is set to 40%. The NSGA-II performance measures (ROI, JD, and NHV), total feasible solutions and pareto fronts are listed in Fig 5.10. The results are consistent with the previous experiments for NSGA-II performance measures as well as the feasible solutions and pareto fronts.

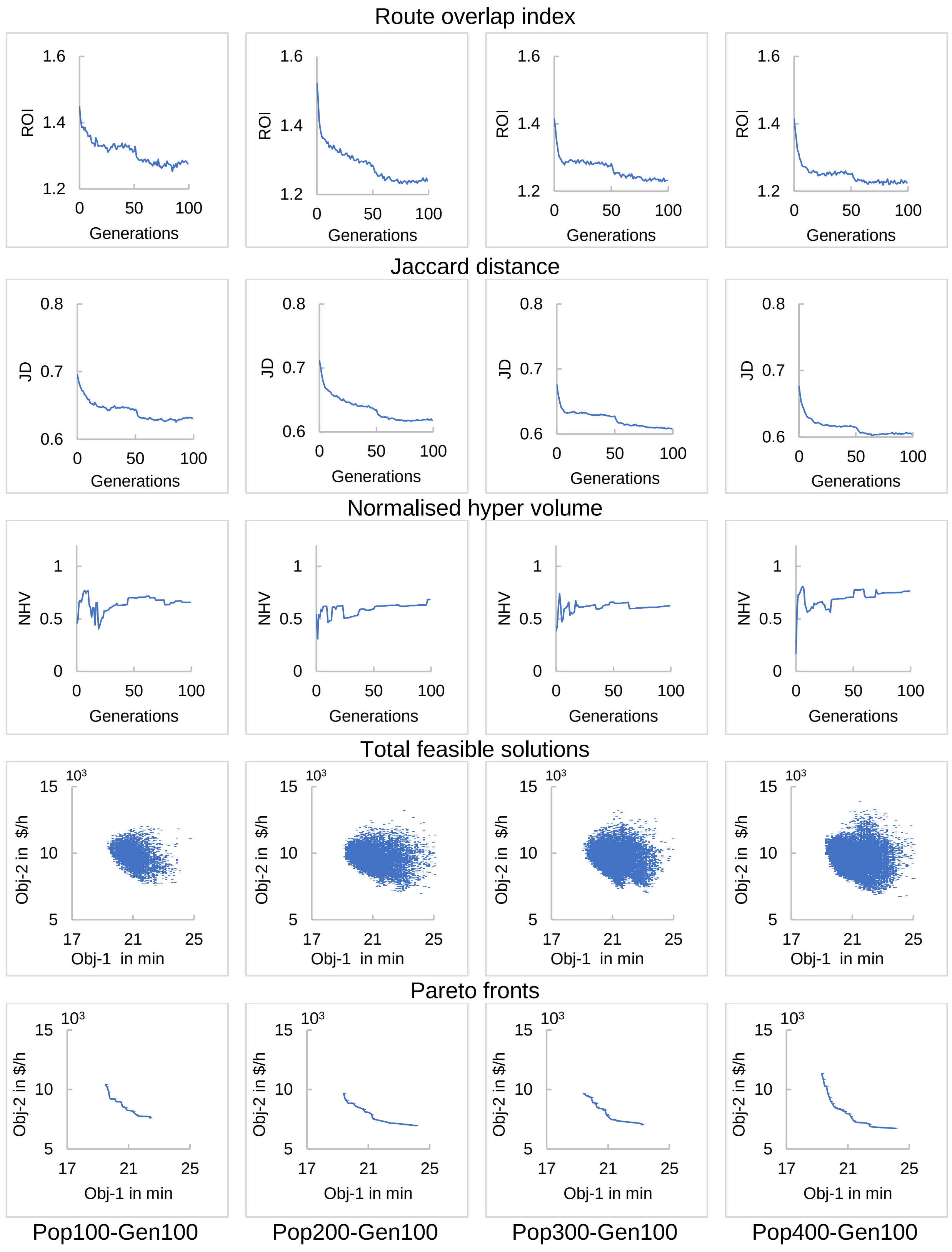
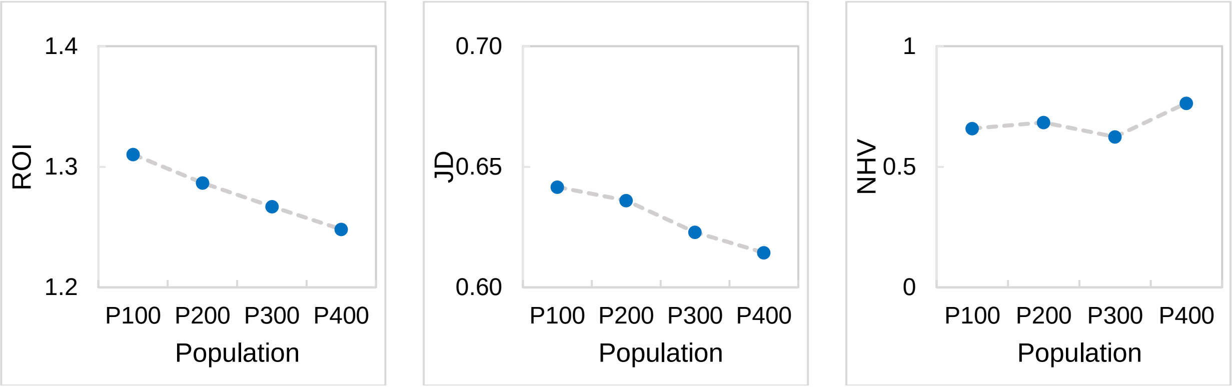
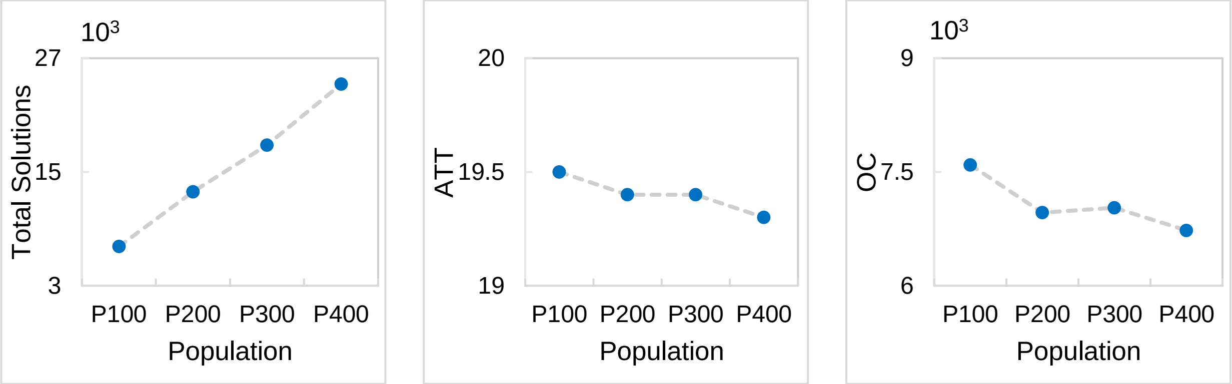


Fig 5.10 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford0 network with ROR-40% with fixed generations and varied population size

The trends for different population sizes for Mumford0 are listed in Fig 5.11. ROI and JD both show a slight decrease with an increase in the population size. However, the NHV shows a slight increase for larger population size. The total number of feasible solutions increases manifolds with an increase in the population size. The ATT and OC, both show a slight decrease with larger population size.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.11 Effect of varied population size with fixed generations for Mumford0 with ROR-40%

5.4.3 Mumford1

The Mumford1 is the mid-level network among the Mumford set of benchmark networks.

For Mumford1 network, following parameters are set:

- The multipoint for crossover is set to 2 points.
- The number of iterations for ITDA is set to 6 iterations.
- The number of iterations for CCITDA is set to 6 iterations.

5.4.3.1 *Results*

The results for Mumford1 network with different ROR values are listed in the Tab 5.10 together with the rest of the TNE indicators. The best solution for passenger and agency optimum is discovered with ROR of 10%. Indicators such as ROI, ESH, RSC show lower values with lower ROR, whereas, ANT, TNL and TND show lower values with higher ROR values.

ROR	Perspectives of stakeholders	Performance indicators								
		ROI	ESH	ANT	RSC	UDP	TNL	ATT	OC	TND
	Passenger	■		■		■		■		■
	Agency	■	■		■		■		■	■
	Authority	■	■	■	■	■	■	■	■	■
	Best solutions									
10%	Passenger	1.1	23122	0.86	52340	**0.01**	183	**24.4**	32066	1.29
	Agency	1.12	24188	0.98	55820	0.02	187	26.9	**25446**	1.27
	Compromise	**1.08**	22231	0.87	51920	**0.01**	178	25.1	27485	1.31
20%	Passenger	1.15	19104	0.85	48160	**0.01**	196	24.8	30859	1.31
	Agency	1.13	19499	0.91	50630	**0.01**	170	26.4	26428	1.27
	Compromise	**1.08**	**18019**	0.91	**47350**	**0.01**	170	25.4	28457	1.29
30%	Passenger	1.19	22323	0.85	52220	0.02	191	25	31753	1.31
	Agency	1.2	19980	0.88	54190	0.02	193	28.3	28635	1.29
	Compromise	1.13	19292	1.07	49800	0.05	180	26.9	29501	**1.25**
40%	Passenger	1.19	21360	**0.79**	51070	**0.01**	205	24.6	32863	1.26
	Agency	1.1	24244	0.99	55630	**0.01**	**169**	26.6	26727	1.32
	Compromise	1.12	21989	0.89	53110	**0.01**	190	26.1	29586	1.26
50%	Passenger	1.47	24290	0.81	54720	**0.01**	250	24.9	37123	1.31
	Agency	1.09	19835	0.9	51920	**0.01**	172	26.8	26481	1.29
	Compromise	1.19	23811	0.89	54450	**0.01**	211	25.5	31881	1.31

Tab 5.10 Objective function values and performance indicators for Mumford1 network with different ROR values

The performance indicators related to the NSGA-II along with total feasible solutions and pareto fronts are depicted in Fig 5.12. The ROI and JD both show slight increase in values with an increase in the ROR values. The NHV shows mixed results for different ROR values.

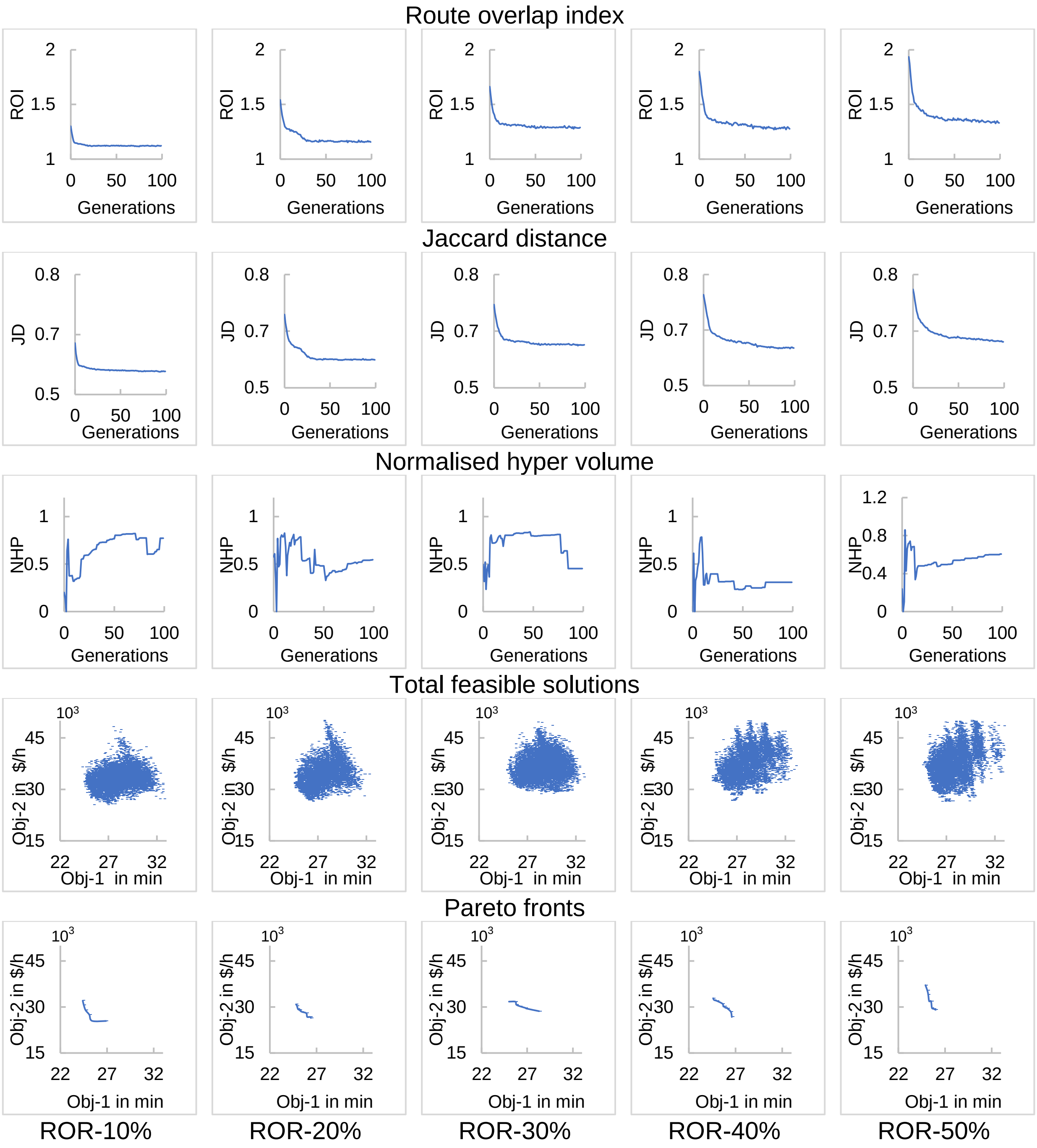


Fig 5.12 Indicators related to NSGA-II performance (ROI, JD and NHV), feasible solutions and pareto fronts for Mumford1 network with different ROR values

5.4.3.2 *Comparisons with other studies*

In terms of comparative results, the proposed solutions are compared with four different solutions from three studies. Moreover, the number of solutions and studies remains same for the next two benchmark networks as well. The results of Mumford1 network are listed in Tab 5.11. The best solutions for passenger (24.4, 32066) and agency (26.9, 25446) are achieved with the proposed methodology.

Study	No of Routes	Solutions	ATT	OC
Mumford [2013]	15	Passenger-optimum	28.7	95,071
	15	Agency-optimum	31.5	42,969
Nayeem, Rahman, & Rahman [2014]	15	Passenger-optimum	27.9	74,685
Kilic & Gok [2014]	15	Compromise	26.9	87,640
ROR-based solutions	6	Passenger-optimum	**24.4**	32,066
	7	Agency-optimum	26.9	**25,446**
	6	Compromise	25.1	27,485

Tab 5.11 Best proposed solution quality comparison with other solutions for Mumford1

5.4.3.3 *Sensitivity analysis*

For sensitivity analysis of Mumford1 network with fixed population size, the ROR value is set to 10%. The ROI, JD, NHV, feasible solutions and pareto fronts are listed in Fig 5.13. The results are consistent with the main experiment for Mumford1 network.

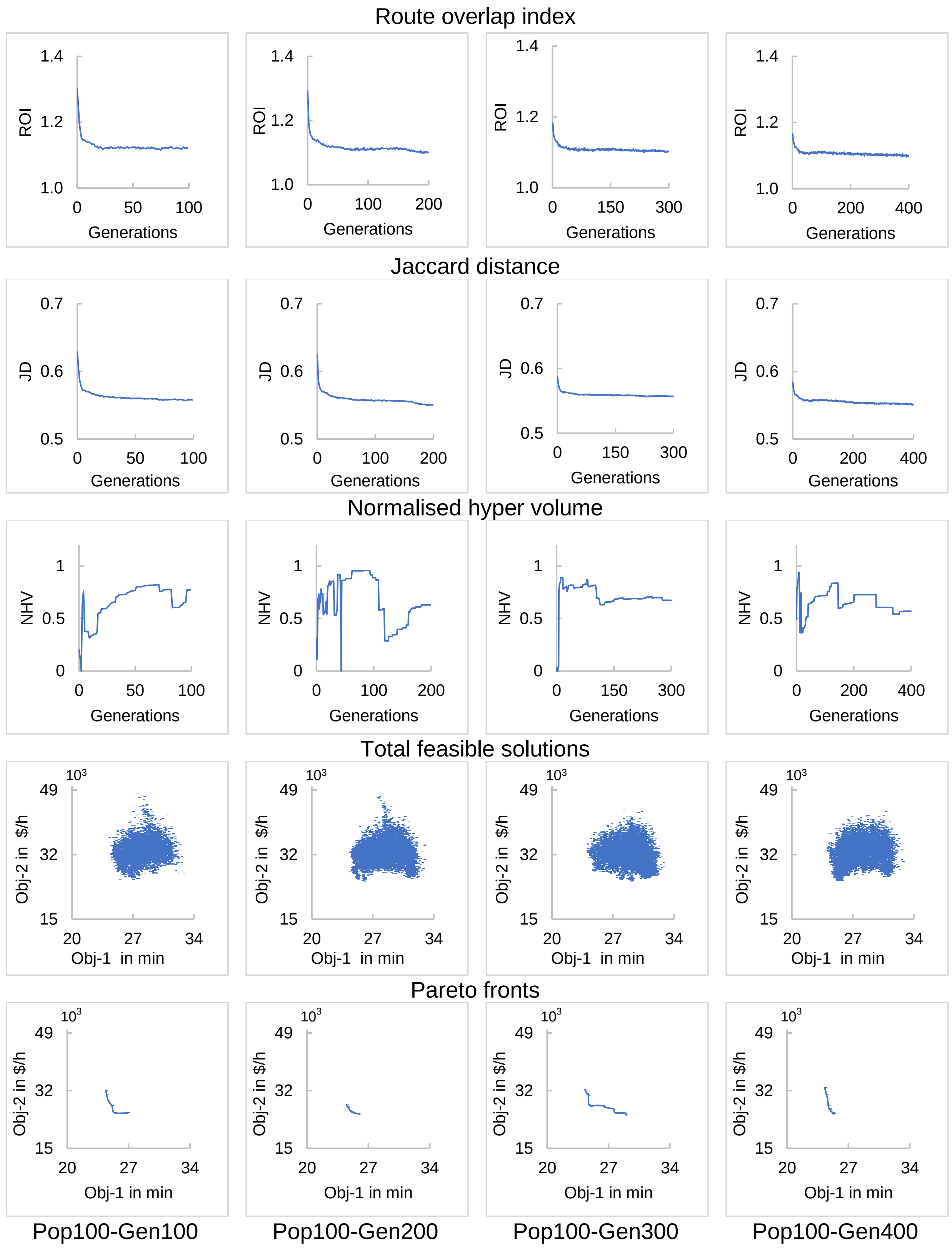
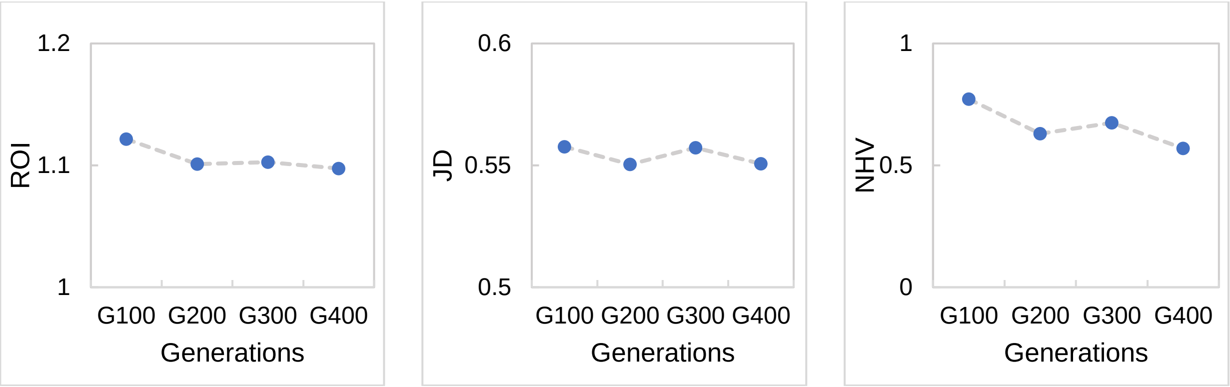
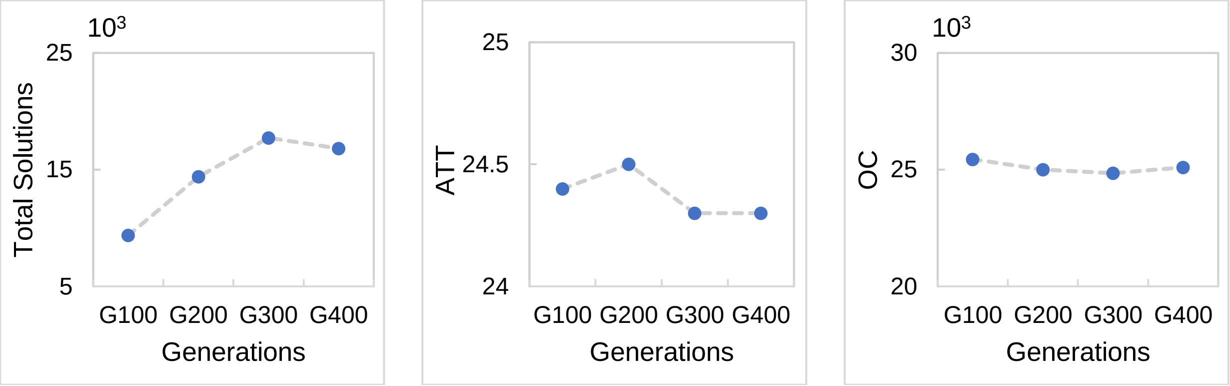


Fig 5.13 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford1 network with ROR-10% with fixed population size and varied generations

The trends for different generations for Mumford1 are listed in Fig 5.14. The results show no significant change in any of the indicators except the total feasible solutions which shows significant increase for higher generation number.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.14 Effect of varied generations with fixed population size for Mumford1 with ROR-10%

For sensitivity analysis of Mumford1 network with fixed generations, the ROR value is set to 10%. The NSGA-II performance measures (ROI, JD, and NHV), total feasible solutions and pareto fronts are listed in Fig 5.15. All the results are similar to previous experiments for Mumford1 network except the NHV, which showsa decline with larger population size.

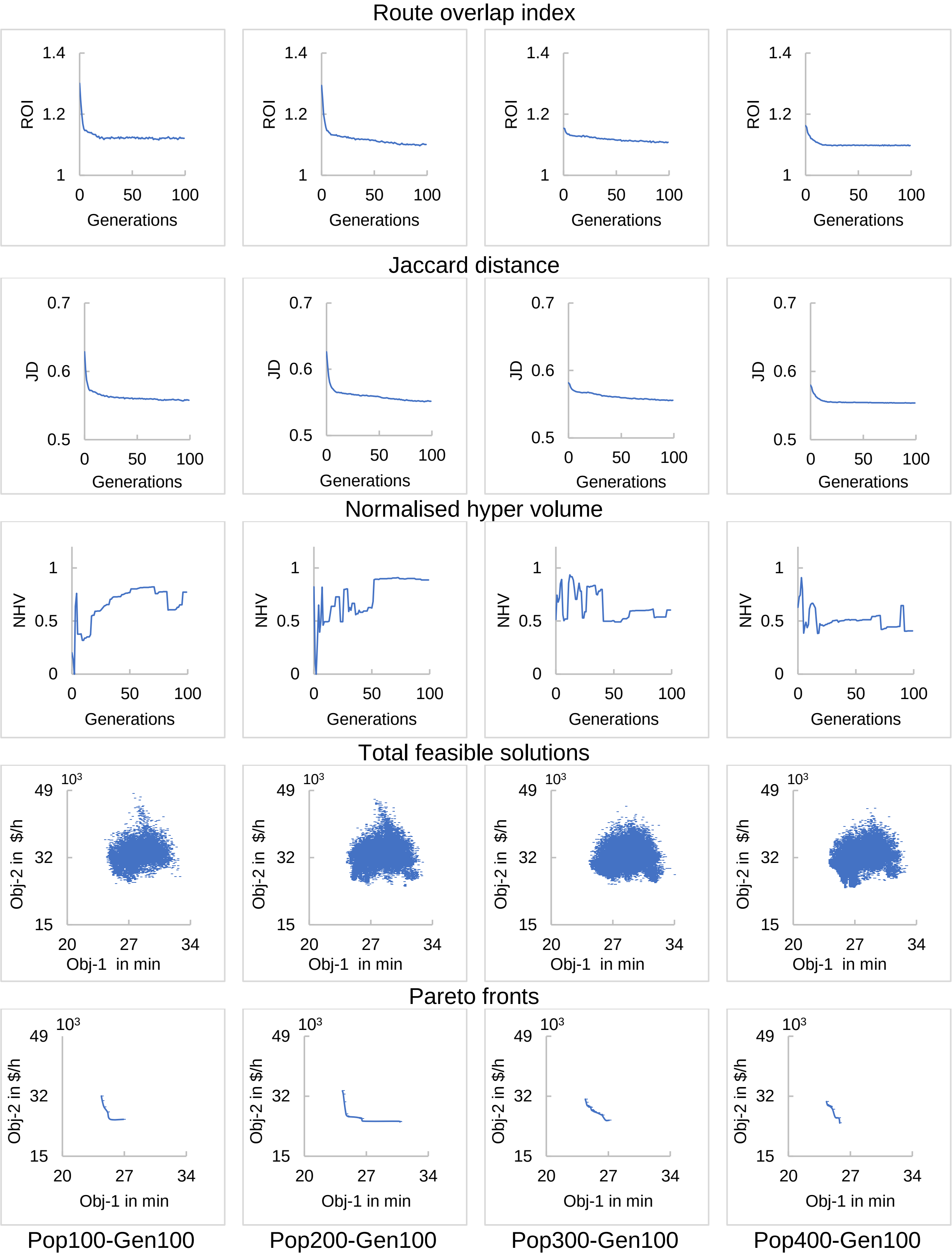
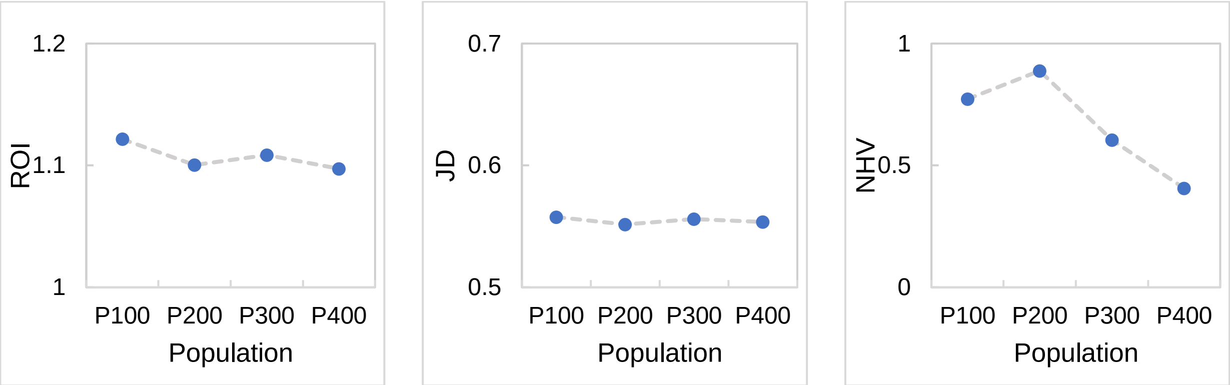
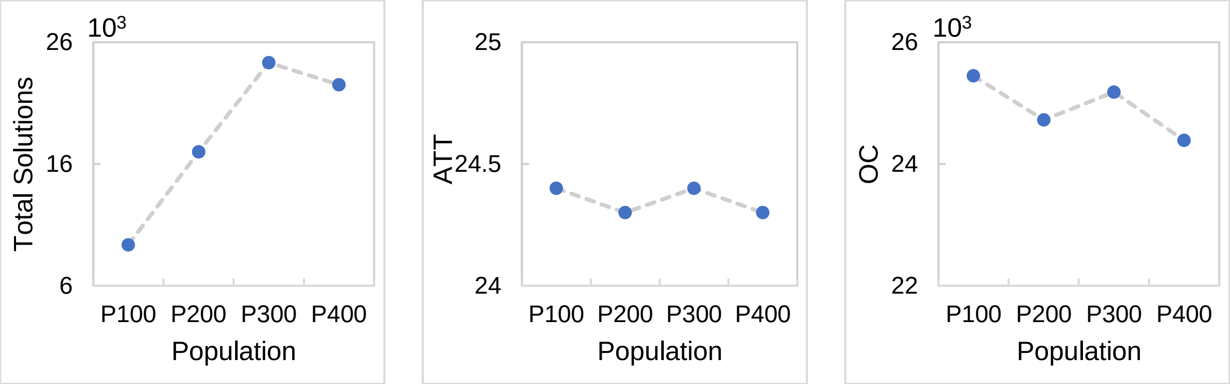


Fig 5.15 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford1 network with ROR-10% with fixed generations and varied population size

The trends for different population sizes for Mumford1 are listed in Fig 5.16. The ROI and JD both remain stable and doesn’t show any change with the increase in population size. However, the NHV shows reduction in its value for larger population sizes. The total number of feasible solutions increases manifolds with an increase in population size. The ATT and OC, both show a slight decrease with larger population size.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.16 Effect of varied population size with fixed generations for Mumford1 with ROR-10%

5.4.4 Mumford2

The Mumford2 network is the second largest network among the Mumford set of benchmark networks. Moreover, it is derived from the city of Hubei, China.

For Mumford2 network, following parameters are set:

- The multipoint for crossover is set to 3 points.
- The number of iterations for ITDA is set to 6 iterations.
- The number of iterations for CCITDA is set to 8 iterations.

5.4.4.1 *Results*

The results for Mumford2 network with different ROR values are listed in the Tab 5.12 together with the rest of the TNE indicators. The best solution for passenger and agency optimum is achieved with ROR of 10%. Moreover, seven out of nine indicators (ROI, ESH, UDP, ATT, OC, TNL and TND) show the least values with minimum route overlap. The results suggest that with minimum route overlap values, a holistic PT network satisfying the needs of different stakeholders can be designed.

ROR	Perspectives of stakeholders	Performance indicators								
		ROI	ESH	ANT	RSC	UDP	TNL	ATT	OC	TND
	Passenger	■		■		■		■		■
	Agency	■	■		■		■		■	■
	Authority	■	■	■	■	■	■	■	■	■
	Best solutions									
10%	Passenger	1.16	51465	0.91	116270	**0.01**	380	**22.5**	78073	**1.29**
	Agency	1.14	49805	0.92	120380	**0.01**	315	24.6	**46834**	1.33
	Compromise	**1.13**	**46624**	0.9	**113760**	**0.01**	**313**	23.4	53328	1.32
20%	Passenger	1.28	56684	0.93	122690	**0.01**	380	22.9	83451	1.38
	Agency	1.26	59645	0.95	133610	0.02	360	25.2	53184	1.38
	Compromise	1.26	52271	0.93	120990	**0.01**	386	23.7	66048	1.4
30%	Passenger	1.3	61225	0.94	129180	**0.01**	421	23.2	94127	1.41
	Agency	1.23	51314	0.97	124650	**0.01**	370	25.3	54440	1.37
	Compromise	1.25	47834	1	115800	0.02	347	24.1	69080	1.36
40%	Passenger	1.53	63455	0.93	133000	0.02	568	23.5	97885	1.4
	Agency	1.32	49675	0.94	122270	**0.01**	403	25	55691	1.37
	Compromise	1.48	55989	0.92	126290	**0.01**	467	24.2	66856	1.37
50%	Passenger	1.68	67973	**0.88**	139510	**0.01**	599	23.7	91534	1.35
	Agency	1.28	59530	0.94	135220	0.02	371	25.5	55854	1.43
	Compromise	1.37	63399	0.95	134870	0.02	408	24.4	69493	1.41

Tab 5.12 Objective function values and performance indicators for Mumford2 network with different ROR values

The performance indicators related to NSGA-II, total feasible solutions and pareto fronts for different ROR values are shown in Fig 5.17. The values and trends for ROI, JD and NHV are similar to that of Mumford1 network with slight increase in ROI and JD values. The NHV values hover around 0.65 in the later stages of the evolution.

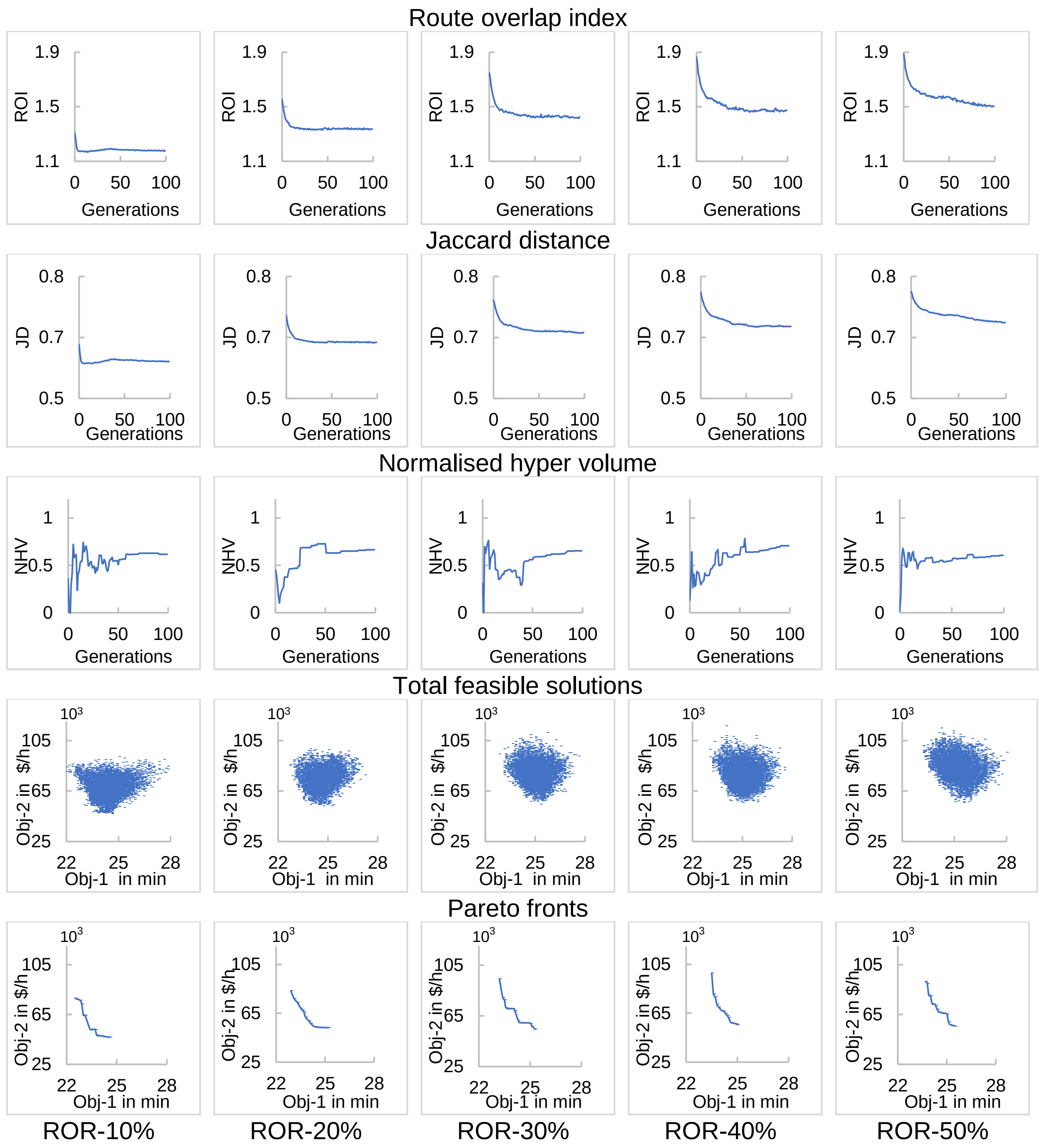


Fig 5.17 Indicators related to NSGA-II performance (ROI, JD and NHV), feasible solutions and pareto fronts for Mumford2 network with different ROR values

5.4.4.2 *Comparison with other studies*

The results of the proposed solutions for passenger, agency and compromise solutions are listed in Tab 5.13 with the rest of the solutions. The proposed solutions offer best passenger optimum solution (22.5, 78073) as well as agency optimum solution (24.6, 46834). This suggests that the proposed methodology can offer a PT network, which is suitable for passengers and agencies.

Study	No of Routes	Solutions	ATT	OC
Mumford [2013]	56	Passenger-optimum	32.5	128,590
	56	Agency-optimum	33.3	265,227
Nayeem, Rahman, & Rahman [2014]	56	Passenger-optimum	30.6	167,030
Kilic & Gok [2014]	56	Compromise	32.2	227,739
ROR-based solutions	11	Passenger-optimum	**22.5**	78,073
	9	Agency-optimum	24.6	**46,834**
	9	Compromise	23.4	53,328

Tab 5.13 Best proposed solution quality comparison with other solutions for Mumford2

5.4.4.3 *Sensitivity analysis*

For sensitivity analysis of Mumford2 network with fixed population size, the ROR value is set to 10%. The ROI, JD and NHV along with total feasible solutions and pareto fronts are listed in Fig 5.18. The results are consistent with the main experiment for Mumford2 network.

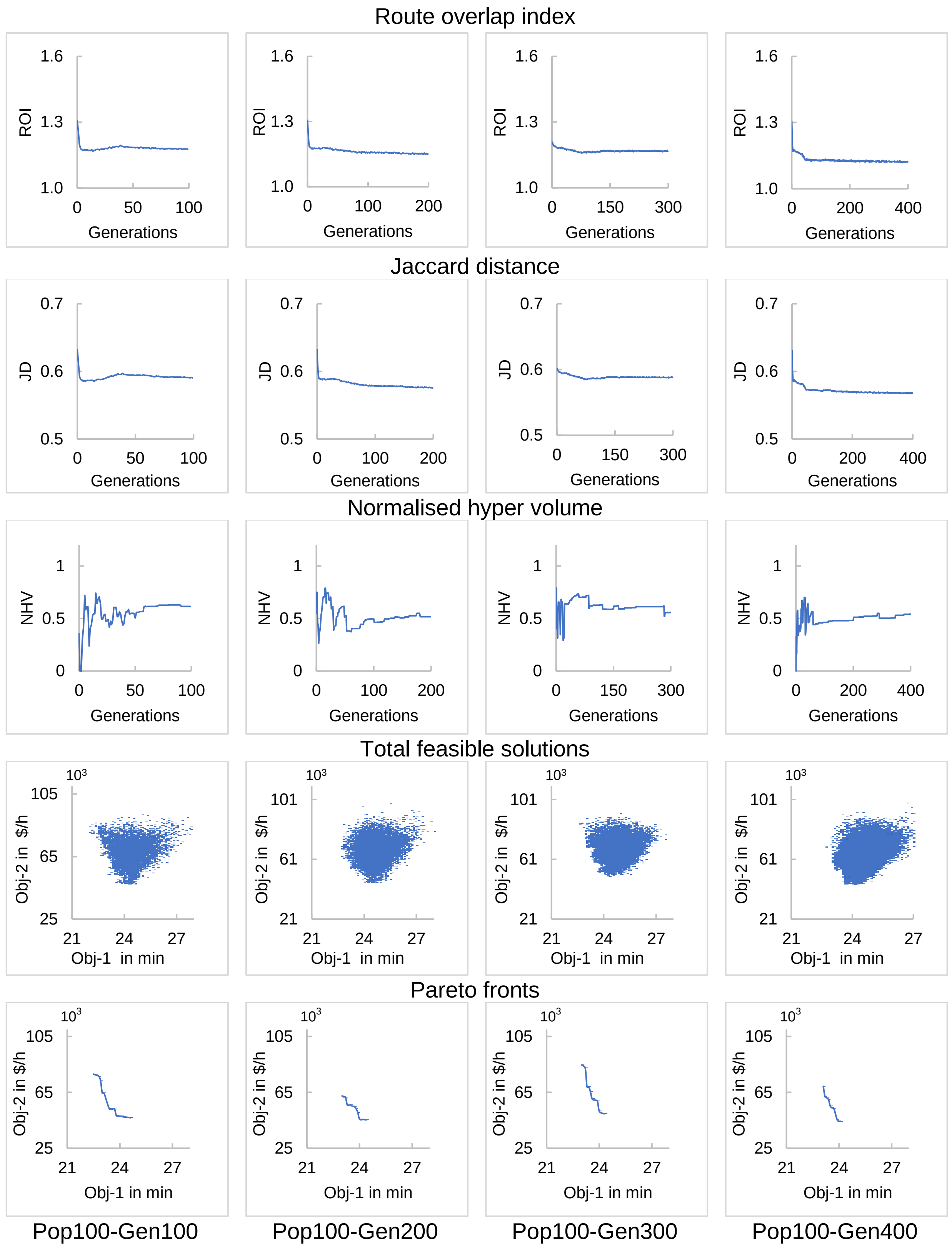
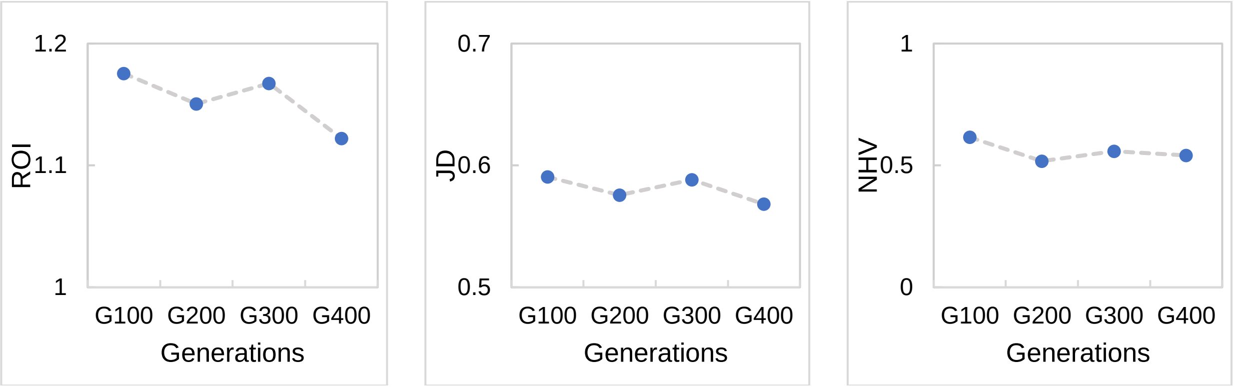
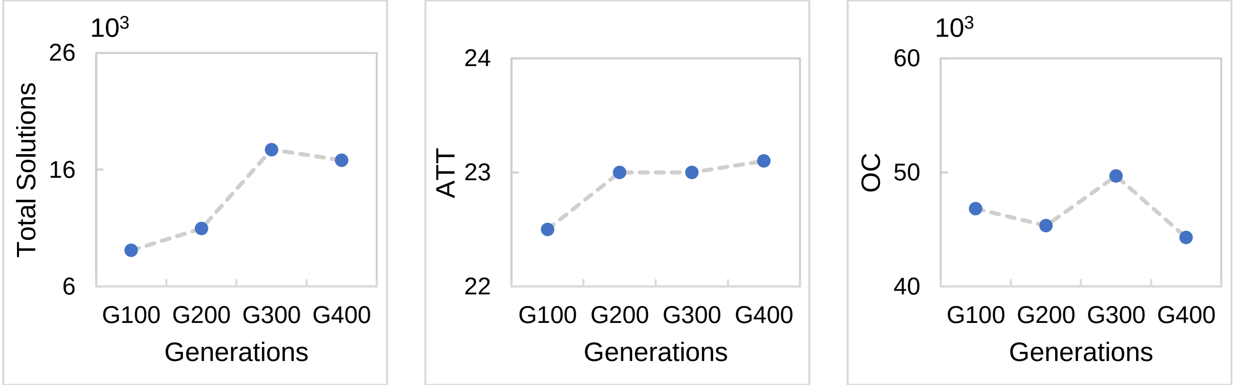


Fig 5.18 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford2 network with ROR-10% with fixed population size and varied generations

The trends for different generations for Mumford2 are listed in Fig 5.19. The results show a decrease in ROI, JD and NHV for higher number of generations. Interestingly, the ATT shows an increase with higher generations, whereas the OC show a decline. The total feasible solutions increase significantly, however, the number drops when switching from generation 300th to generation 400th.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.19 Effect of varied generations with fixed population size for Mumford2 with ROR-10%

For sensitivity analysis of Mumford2 network with fixed generations, the ROR value is set to 10%. The ROI, JD and NHV along with total feasible solutions and pareto fronts are listed in Fig 5.20.

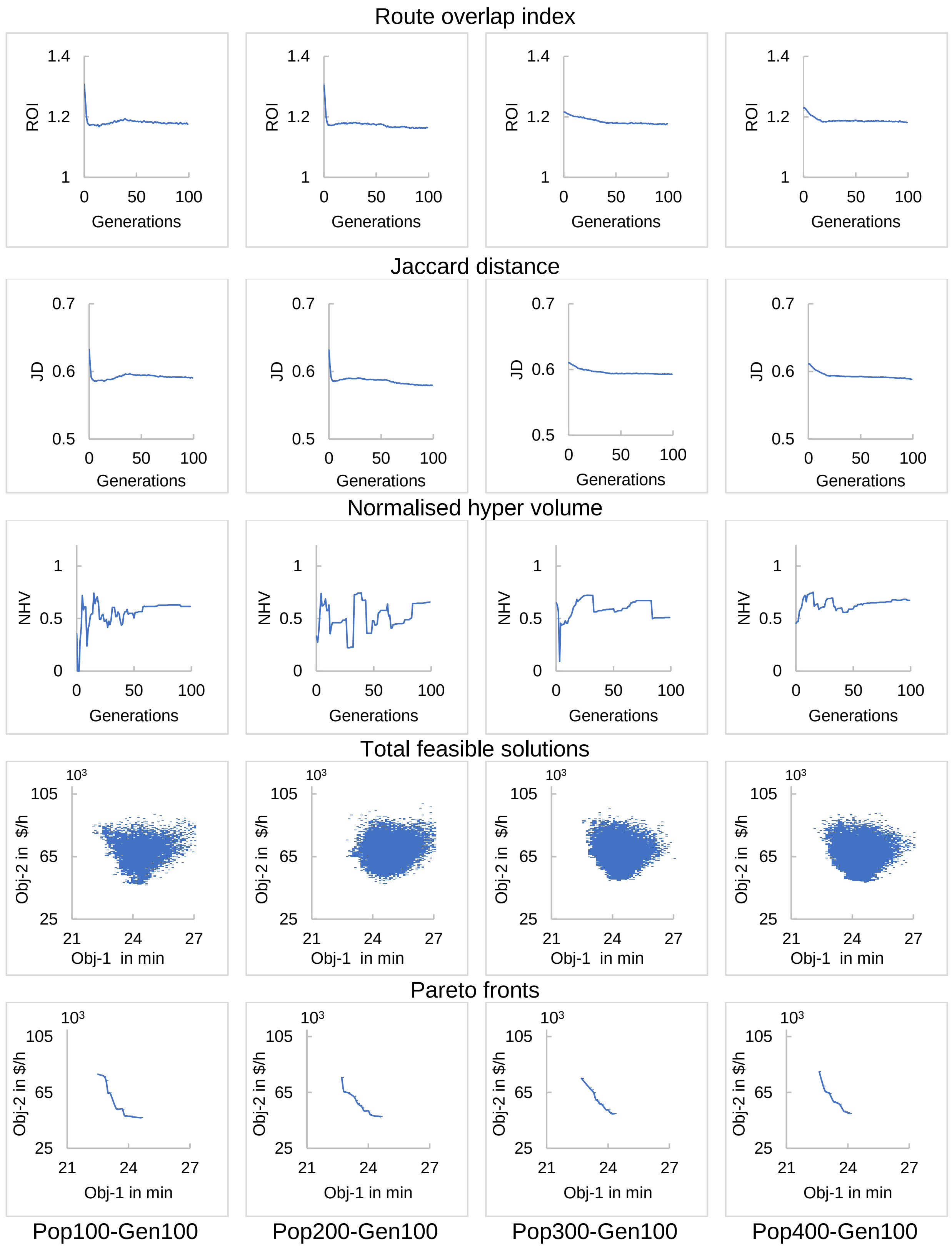
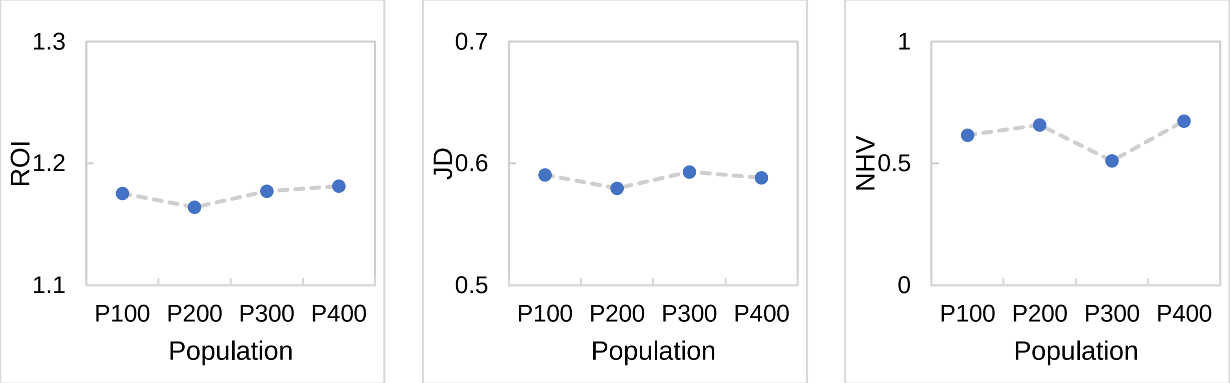
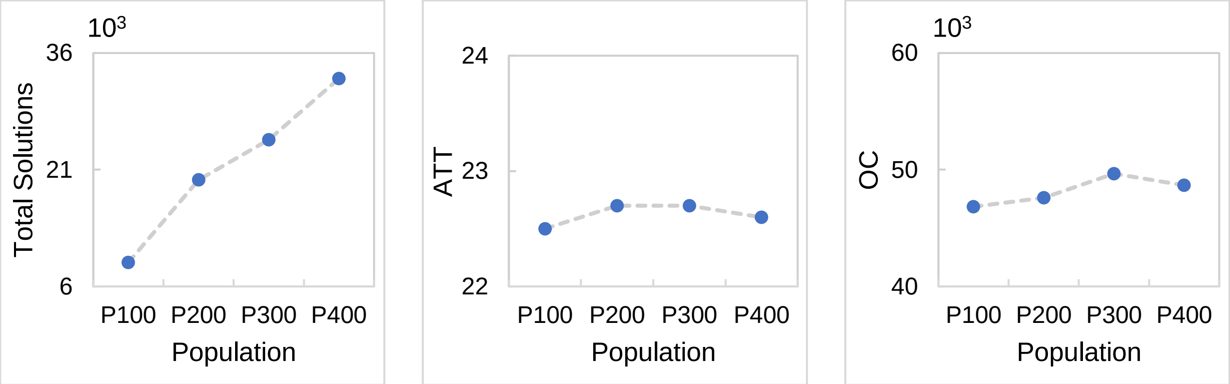


Fig 5.20 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford2 network with ROR-10% with fixed generations and varied population size

The trends for different population sizes for Mumford2 are listed in Fig 5.21. The ROI and JD do not show any change with different population sizes. However, the NHV shows a slight increase for larger population size. The total number of feasible solutions increases significantly with an increase in the population size. However, the ATT and OC don't show any change with larger population size.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.21 Effect of varied population size with fixed generations for Mumford2 with ROR-10%

5.4.5 Mumford3

The Mumford3 network is the largest network among the Mumford set of benchmark networks. Moreover, it is derived from the city of Cardiff, UK.

For Mumford3 network, following parameters are set:

- The multipoint for crossover is set to 4 points.
- The number of iterations for ITDA is set to 6 iterations.
- The number of iterations for CCITDA is set to 10 iterations.

5.4.5.1 *Results*

The results for Mumford3 network with different ROR values are listed in Tab 5.14 together with the rest of the TNE indicators. The best solution for passenger and agency optimum is discovered with ROR of 10%. The results are similar to the Mumford2 network with majority of the indicators show lower values with lower route overlaps.

ROR	Perspectives of stakeholders	Performance indicators								
		ROI	ESH	ANT	RSC	UDP	TNL	ATT	OC	TND
	Passenger	■		■		■		■		■
	Agency	■	■		■		■		■	■
	Authority	■	■	■	■	■	■	■	■	■
	Best solutions									
10%	Passenger	1.2	76714	0.9	**162280**	**0.01**	468	**22.4**	109801	**1.34**
	Agency	1.14	75946	0.94	174150	0.02	**416**	25.3	**62110**	1.36
	Compromise	**1.13**	77111	0.95	166060	**0.01**	432	23.4	77896	**1.34**
20%	Passenger	1.34	92442	0.91	180100	**0.01**	558	22.5	130544	1.38
	Agency	1.27	**75790**	0.88	173630	**0.01**	447	25	66706	**1.34**
	Compromise	1.31	80872	0.88	171600	**0.01**	466	23.4	98733	1.36
30%	Passenger	1.31	90626	0.93	180950	0.02	483	22.9	110959	1.37
	Agency	1.29	83283	**0.85**	183680	**0.01**	437	25.1	64537	1.35
	Compromise	1.27	83208	0.89	173620	**0.01**	430	23.7	85951	1.37
40%	Passenger	1.32	85860	0.92	172870	**0.01**	493	22.5	118223	1.38
	Agency	1.38	81135	0.89	178140	**0.01**	506	24.8	67851	1.39
	Compromise	1.43	90966	0.92	180920	**0.01**	534	23.3	101675	1.37
50%	Passenger	1.54	101045	0.91	190610	**0.01**	647	22.8	127134	1.41
	Agency	1.4	84096	0.92	186210	0.03	479	25.7	69249	1.35
	Compromise	1.44	99753	0.88	196380	**0.01**	571	24	91171	1.38

Tab 5.14 Objective function values and performance indicators for Mumford3 network with different ROR values

The performance indicators of NSGA-II together with total feasible solutions and pareto fronts are listed in Fig 5.22. The values of ROI and JD both show slight increase for higher ROR values. The NHV shows even better values compared to Mumford2 network with average value of 0.85 at the end of evolution.

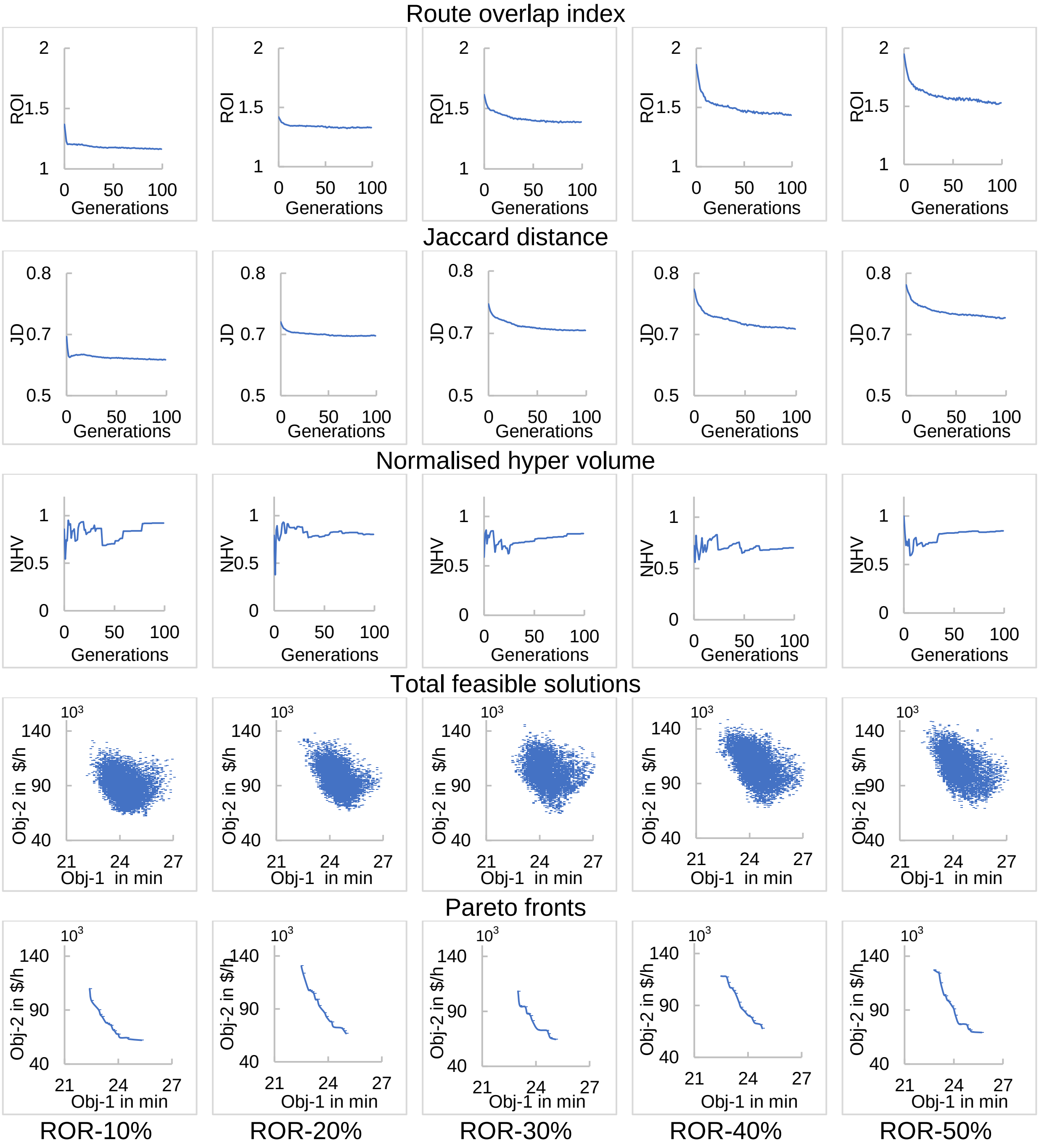


Fig 5.22 Indicators related to NSGA-II performance (ROI, JD and NHV), feasible solutions and pareto fronts for Mumford3 network with different ROR values

5.4.5.2 *Comparisons with other studies*

In terms of comparison with other related studies, the proposed solutions for passenger, agency and compromise perspective are listed in Tab 5.15. The results show that the proposed methodology offer best passenger optimum (22.4, 109801) and agency optimum solution (25.3, 62110).

Study	No of Routes	Solutions	ATT	OC
Mumford [2013]	60	Passenger-optimum	28.5	308,736
	60	Agency-optimum	30.1	181,414
Nayeem, Rahman, & Rahman [2014]	60	Passenger-optimum	32.9	207,594
Kilic & Gok [2014]	60	Compromise	27.7	253,146
ROR-based solutions	12	Passenger-optimum	**22.4**	109,801
	10	Agency-optimum	25.3	**62,110**
	11	Compromise	23.4	77,896

Tab 5.15 Best proposed solution quality comparison with other solutions for Mumford3

5.4.5.3 *Sensitivity analysis*

For sensitivity analysis of Mumford3 network with fixed population size, the ROR value is set to 10%. The ROI, JD and NHV along with total feasible solutions and pareto curves are listed in Fig 5.23, the results are consistent with the main experiment for Mumford3.

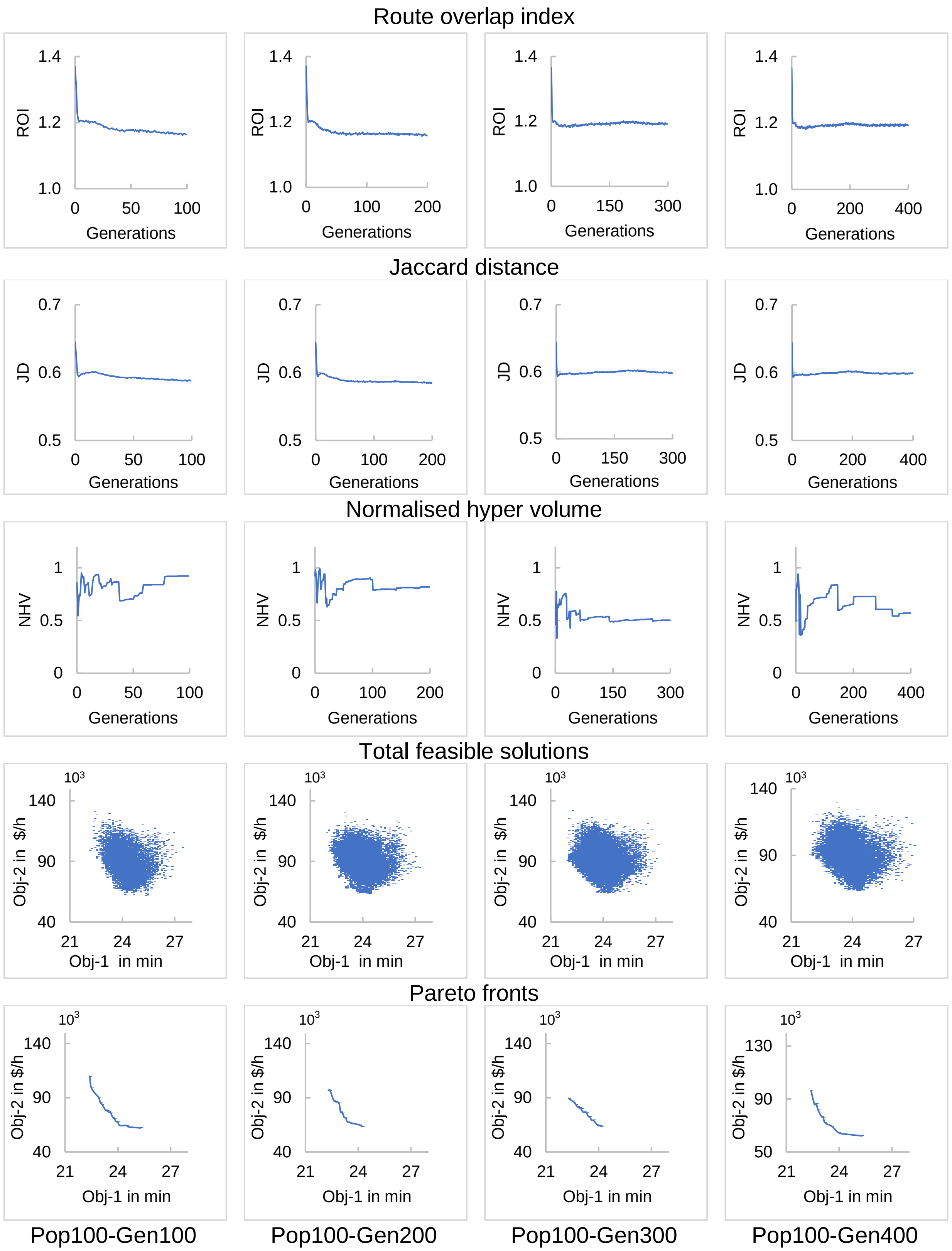
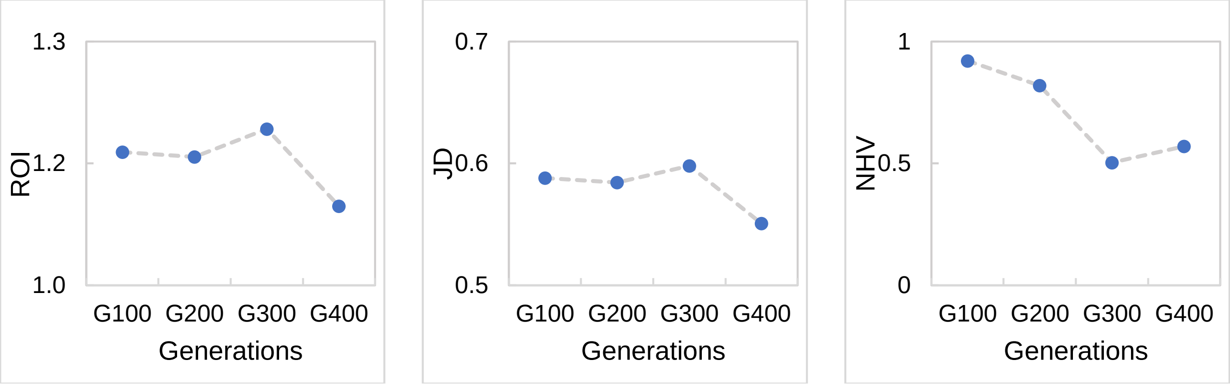
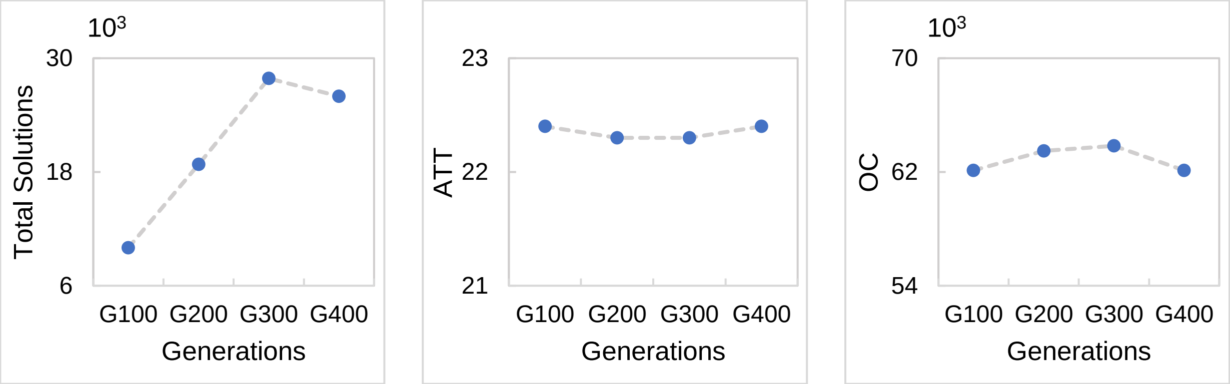


Fig 5.23 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford3 network with ROR-10% with fixed population size and varied generations

The trends for different generations for Mumford3 are listed in Fig 5.24. The results show a decrease in ROI, JD and NHV for higher number of generations, similar to Mumford2 network. The ATT and OC do not show any change with higher generations. Like Mumford2, the total feasible solutions increase significantly, but the number drops slightly from generation 300th to generation 400th.



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.24 Effect of varied generations with fixed population size for Mumford3 with ROR-10%

For sensitivity analysis of Mumford3 network with fixed generations, the ROR value is set to 10%. The ROI, JD and NHV along with total feasible solutions and pareto curves are listed in Fig 5.25. The results are consistent with the previous experiments.

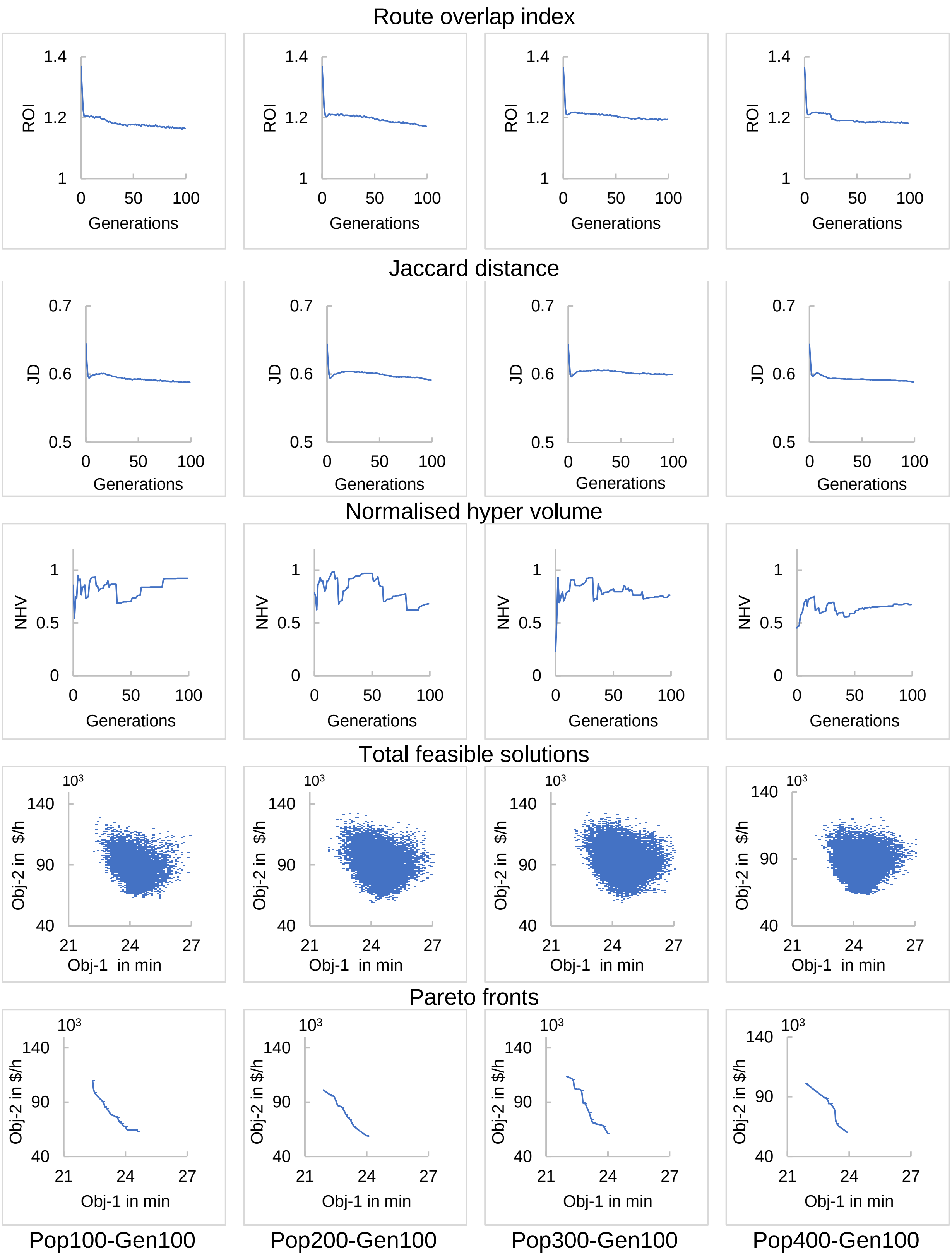
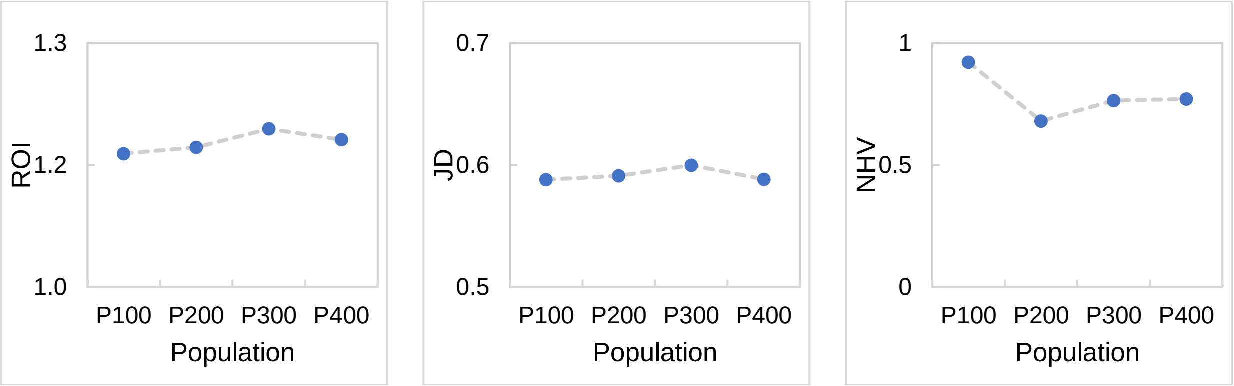
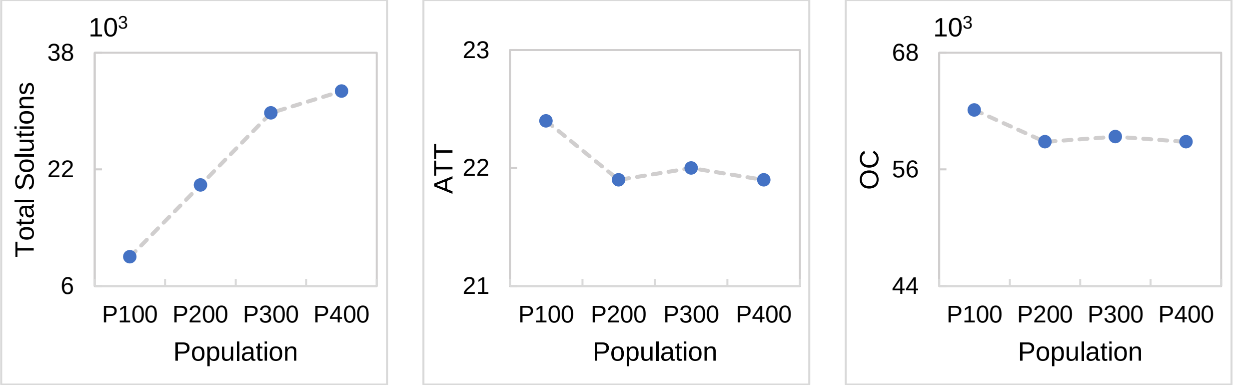


Fig 5.25 Indicators related to NSGA-II performance, feasible solutions and pareto fronts for Mumford3 network with ROR-10% with fixed generations and varied population size

The trends for different population size for Mumford3 are listed in Fig 5.26. The results show that both ROI and JD remain consistent, however, NHV shows a slight decrease with larger population size. The ATT and OC both show a decrease with an increase in the population size. The total feasible solutions increase significantly, from 10,000 (population 100) to 32,000 (population 400).



Route overlap index Jaccard distance Normalised hyper volume



Total feasible solutions Average travel time Operating cost

Fig 5.26 Effect of varied generations with fixed population size for Mumford3 with ROR-10%

5.5 Summary

This chapter focused on the implementation of the developed methodology on a set of benchmark networks. Moreover, benchmark network evaluations along with performance comparison with other related studies were also discussed. The results achieved from the developed methodology are promising and offer significant cost reduction for passenger and agencies. Moreover, different types of PT network configurations and their properties are illustrated via PT network evaluation. The results suggest that for larger networks, lesser route overlapping, and lesser number of PT routes can offer significant cost reduction for passenger and agency. Moreover, such PT networks also show promising results on selected TNE performance indicators. This means a holistic PT network can be designed with a sparse, high frequency PT routes at lower cost for both stakeholders. In the next chapter, the viability of the developed methodology will be tested on the large urban network of Singapore.

6 Implementation of Proposed Set of Procedures on Singapore’s Public Transit Network

The methodology explained in section 4, will be applied to the PT network of Singapore. This is possible because of different data sources available from governmental agencies in Singapore such as Land transport authority (LTA), Singapore land authority (SLA). Moreover, open source data sources such as from OpenStreetMap data and Google Maps. Apart from that, on field surveys also provide valuable information. In this section, firstly different data sources will be briefly explained both from demand and supply side. Secondly, the processing of these data sources will be explained. Finally, the results will be showcased along with comparison with the existing network.

6.1 Singapore public transit data

Singapore transit data is composed of two parts: 1) PT supply data which includes, stop location, and routes layout; 2) PT demand data which includes trips made on PT network. Three different data sources are used: 1) contactless e-purse application (CEPAS) contactless smart card data used for the PT fares in Singapore, acquired from LTA; 2) digitised road and rail information geospatial data, acquired from SLA; 3) bus route and service information data, acquired from DataMall, a database of different datasets and statistics published by LTA.

6.1.1 Public transit supply data

In PT supply data, the first part is the road network and the stop information (bus and rail). The second part is the route layout and the service information data.

6.1.1.1 *Road and stop data*

In road and stop data, the road data is available as geographical information system (GIS) shape files. These shape files have attribute tables associated with them which include information such as name and category of the road section, its alleviation level and direction of the traffic. There are roughly 64,000 arcs and 24,000 nodes in the network. The basic road network is depicted in the Fig 6.1.

(a) Road network arcs

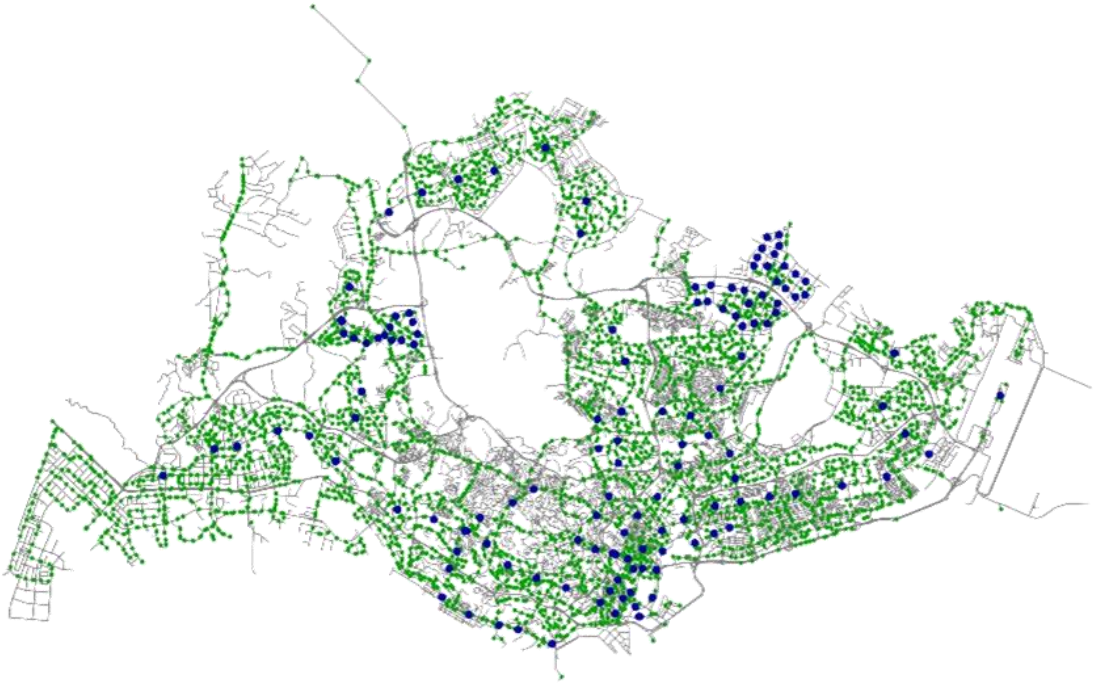
(b) Road network nodes

(c) Combined nodes and arcs combined

Fig 6.1 Road network of Singapore

The bus stop data is available in the form of GIS coordinates. This data is acquired from the DataMall. There are roughly 4,846 bus stops and 119 MRT stops and 42 LRT stops until year 2013. The stops and road network are depicted in the Fig 6.2.

(a) Bus stops (b) MRT and LRT stops



(c) Combined stops and road network

Fig 6.2 Bus, MRT and LRT stops of Singapore public transit network

6.1.1.2 *Public transit routes*

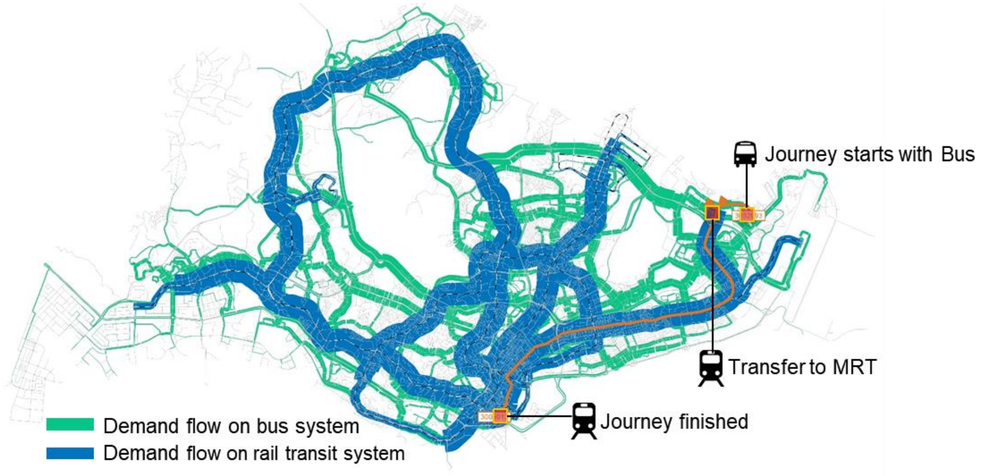
Public transit routes for bus, MRT and LRT are constructed by combing different data sources together. However, since the information regarding routes in the DataMall is from 2014 and the original smart card data is from 2013. Adjustment were made to DataMall and smart card data to match supply and demand data.

There are roughly 380 bus routes including feeder, express and normal services. Although there are more bus routes, but they don't use smart card data for fare payment. There are four MRT lines including east-west, north-south, north-east and circle lines. Moreover, there are three LRT lines including Bukit Panjang, Sengkang and Punggol.

6.1.2 Public transit demand data

A smart card database was provided by the LTA, this data includes information about the public transit trips for 92 days starting from 1st August 2013 till 31st October 2013. One trip means, one tap-in and tap-out of the smart card while using PT service. There was a total of 517,203,124 recorded trips for bus, MRT and LRT in the data. There were roughly, 5.5 million trips on average per day. The available information from such smart card data includes: 1) identification of the smart card type; 2) The ride information of the passenger; 3) the public transit service information. The identification of card means which type of card is used for the public transit ride. For instance, special discounted card for elderly/ students and normal card for ordinary citizens. The ride information includes most of the details such as what type of transit mode was used, at which stop the passenger boards the bus/MRT and at which stop the passenger alights. Moreover, the total ride time and distance is also included in this information. The public transit service information includes, the service name, license plate and vehicle departure sequence. This information is only available for bus services.

If there are multiple consecutive trips with transfers from the same card type, in that case, if the transfer is within 45 min and the combined time of the consecutive trips is within two hours, it is considered as one journey. Otherwise, it will be considered as a new journey. A simple example is sown in Fig 6.3, where a passenger starts its journey from a bus stop with bus service, later the passenger transfers to MRT station and finally alights at the MRT station. Note that the total time and distance are also given.



Trip information

Trip index	Mode	Line	Direction	From stop	Dep. time	To stop	Arr. time	Time	Distance
1	Bus	6	1	Greenview Sec. Sch.	10:02:13	Pasir Ris int.	10:08:47	6min 34s	2.266km
2	Walk	Transfer		Pasir Ris int.	10:08:47	Pasir Ris int.	10:12:51	4min 4s	0.091km
3	MRT	EW	1	Pasir Ris int.	10:14:42	Tanjong Pagar MRT	10:47:38	32min 56s	20.348km

Fig 6.3 Smart card data example of the public transit in Singapore

6.2 Preparation of network

The preparation of the network is necessary to match the proposed methodology described in section 4. There are two things which need to be processed: 1) the number of arcs present in the network; 2) the number of stops in the PT network.

6.2.1 Network processing

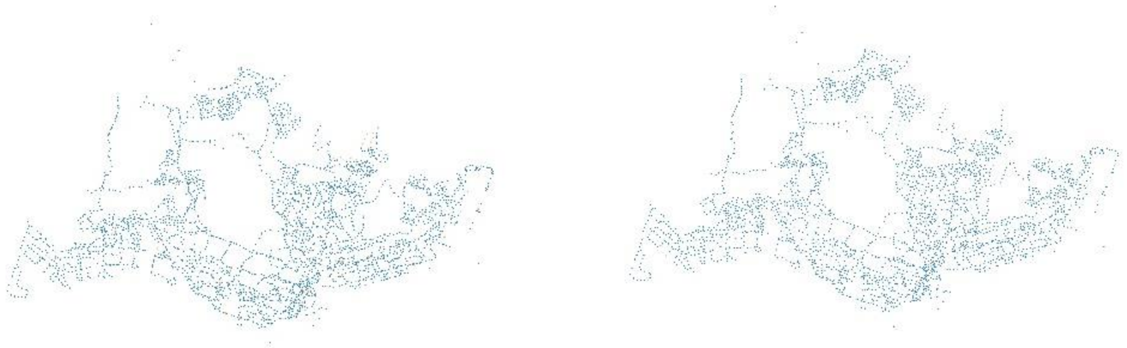
There are 64,000 arcs and 24,000 nodes in the road network of Singapore. These are way too many arcs and nodes in the network. This number can lead to a number of problems in the route construction due to: 1) high computational cost associated with path creation algorithms; 2) some of the arcs and nodes in the network are not required such as restricted areas and industrial zones. Therefore, simplification of the network both at arc and node level is required.

6.2.1.1 *Nodes processing*

In order to simplify the nodes in the network, two assumptions are made: 1) Only the bus stops, MRT and LRT stops are considered as nodes; 2) Opposite bus stops are considered as one bus stop with different directions. With these two assumptions, the nodes in the network are reduced significantly and this results in 2,828 stops. The simplification of the nodes is shown in Fig 6.4a and Fig 6.4b.

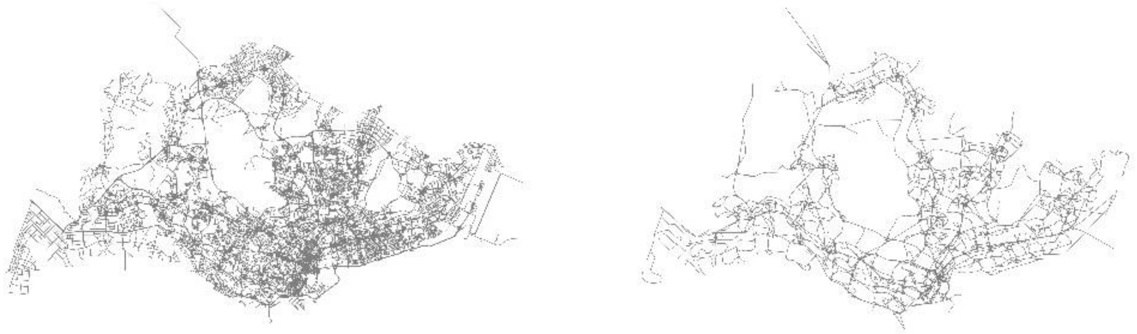
6.2.1.2 *Arc processing*

For the arc processing, instead of the normal arcs in the road network, bus routes' stop sequences are considered as the arcs. These stop sequences are extracted from service information in the smart card data. All the stop sequence pairs in the PT network are considered, and if there is any PT service which serve two consecutive stops, they are considered as arc. The distance between these stops are also extracted from the smart card data. With this simplification, the total arcs are reduced to 9,000 arcs. Please note that, the stops considered for arc extractions are not simplified stops (as explained in section 6.2.1.1) but the real stops from the smart card data. The simplification of the arcs is shown in Fig 6.4c and Fig 6.4d. The combined simplified arcs and nodes are listed in Fig 6.4e.



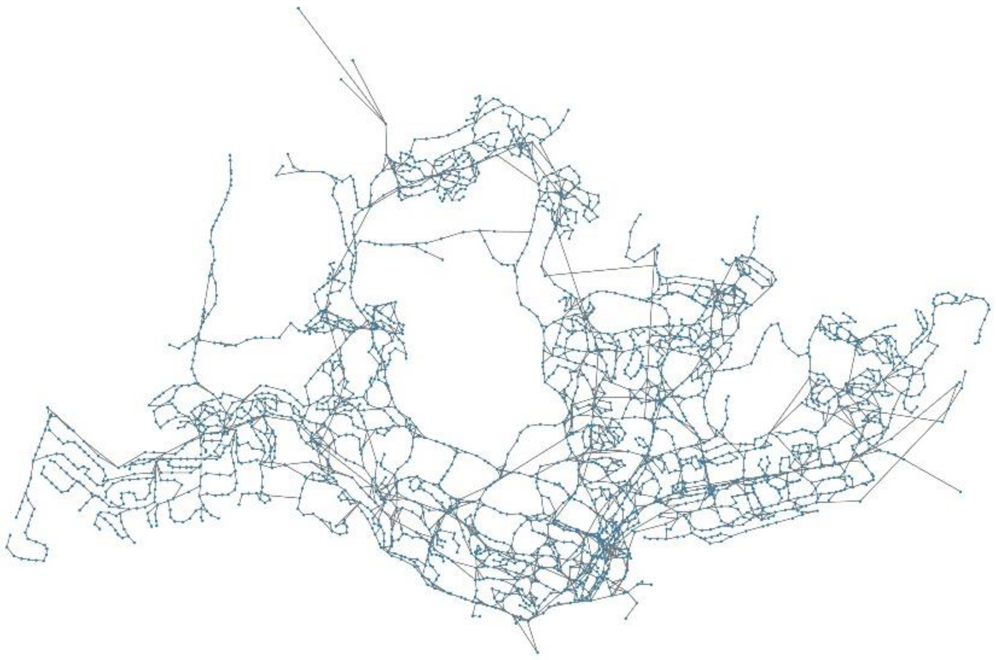
(a) Original bus stops

(b) Reduced bus stops



(c) Original road network

(d) Reduced road network



(e) Reduced road network with reduced stops

Fig 6.4 Node and arc simplification of the public transit network of Singapore

6.2.1.3 *Assumptions and limitations*

The assumptions, simplifications and their implications for nodes and arcs processing are as follow:

- The PT service in Singapore is available in most parts however, there are areas within Singapore where no or limited public transit service is available. This doesn't mean that the PT service is bad in these areas. This can be:
    - There is no demand in that area (reserved area, industrial area).
    - The layout of the streets doesn't allow any PT service operations.
- High number of nodes and arcs increases the computation cost for the route generating algorithms. Therefore, simplification is necessary to obtain a compromise between quality and computation cost.
- Because of model requirement and the strategic level network design, the stops in opposite directions are reduced to single stops with two directions.

Such simplification often leads to reduction in realism. The adopted simplification also has its limitations such as:

- Transfer walking time among transfer stops is not considered and instead a transfer penalty is used.
- Assumption of two opposite stops into one with two directions doesn't account the ease of access to the opposite stops such as there might be an overhead bridge between two stops or the two stops are not facing each other, they might be far apart from each other, and require excessive walking.

6.3 Preparation of demand

The demand data is composed of 92 days. However, different days of the week show different demand patterns. Therefore, only 10-week days are selected with the least fluctuations in the demand patterns. More details can be found in Liu, Zhou, & Rau [2018] regarding selection of the days for demand estimation. The demand for each day is divided into the 15 minutes intervals to get more detailed demand patterns. These 15 minutes intervals are combined to capture the average peak demand for morning and evening peaks.

In the proposed methodology, any demand can be used whether its peak or off-peak hour. Generally, the strategic planning of the network is performed with the peak hour demand data. However, the data reveals the following:

- If the peak demand is considered as many as half of the bus stops show zero demand. This could be because of data aggregation and the nature of the demand patterns in the morning/ evening peak hours.
- If the average daily demand of whole day is considered, majority of the stops show some level of demand values. However, with this approach, the peak demand is underestimated.

In an ideal case, the network should be designed such that at first the network is designed only for peak hours. Later, this network is extended to incorporate average daily demand for the whole

PT network. However, due to time and resource considerations, only the average daily demand is considered for 19 hours of PT service operation. This will provide us a perfect compromise between the spatial and temporal aspect of PT demand. Moreover, the demand between the stops is aggregated and adjusted for the reduced number of stops (see section 6.2.1.1).

6.4 Current public transit network description

The PT network of Singapore consist of different transport layers such as MRT, LRT, Bus, and taxi services. In this section, first a brief description about the PT network is given and later attributes of the PT network are presented.

6.4.1 Network description

Singapore's PT network is composed of intensive bus network and MRT network to serve the major demand corridors in Singapore.

Based on the available data sources, the attributes of Singapore's PT network are as follows:

- There are roughly 380 bus routes which are further divided into feeder, express and normal services.
- There are four MRT routes including east-west, north-south, north-east and circle line.
- There are four LRT routes in different parts of Singapore as well.
- The total network length of Singapore PT network is approximately 11,600 km.
- There are roughly 5,000 bus stops, 120 MRT stops and 40 LRT stops.
- The PT system offers on average 26.4 minutes per trip. This includes in-vehicle time, transfer wait time and transfer walk time. The initial waiting time is difficult to estimate as the data sources don't provide such information.
- In terms of directness, the PT system offers roughly 54% direct trips, 32% with one transfer and 11% with two transfers. The rest 3% of the trips requires more than two transfers. This gives a total of 0.6 transfers on average per trip.

6.5 Implementation of proposed algorithmic procedures

The scalability and viability of the proposed methodology is checked on the public transit network of Singapore. The current MRT and LRT networks are kept unchanged. However, the bus network is removed, and later optimised through developed methodology.

6.5.1 Experimental parameters and setup

The experimental parameters used for Singapore's public transit network are listed in Tab 6.1.

Type	Parameters	Value
NSGA-II	Population size	30 individuals
	Generations	30 generations
	Mutation rate	0.005
	Mutation crowding degree	0.5
	Crossover rate	0.9
ROR	ROR	10%
Feasible route set	Detour maximum	1.5 times shortest path
	Maximum routes per node pair	1,000 per node pair
Route	Minimum headway bus	6 minutes
	Maximum headway bus	30 minutes
	Minimum headway BRT/LRT	5 minutes
	Maximum headway BRT/LRT	30 minutes
	Minimum headway MRT	3 minutes
	Maximum headway MRT	30 minutes
	Maximum passenger load	1.0
	Terminal time coefficient	0.10%
Objective function parameters passenger costs	Origin wait time coefficient	1.0
	In-vehicle time coefficient	1.0
	Transfer wait time coefficient	1.0
	Transfer penalty	5 minutes
Reproduction Buffer	Crossover population	2 times population size
	Feasible solution pool Sample	2 times population size
Demand assignment	ITDA	First 15 generations
	ITDA + ICCTDA	Last 15 generations

Tab 6.1 Parameters used for Singapore public transit network

The experimental setup for Singapore case is as follow:

- 134 terminals are selected between which the feasible routes are created. Among these, there are 40 actual terminals and 94 stops which have a degree 1.
- The total number of feasible routes created between the selected terminals are 9,982,195. An Intel Xeon CPU E5-2640 v3 @ 2.6GHz with 128 GB RAM was used, it took 179,755 seconds with parallel processing.
- First an initialisation (ROR-INI) is performed to create initial population of 30 individuals. Later, NSGA-II is run for 30 generations.
- The crossover and mutation rates used in the experiments are 0.9 and 0.005 respectively. Moreover, the buffer to adjust non-feasible routes, produced during reproduction, are set to 2 times the population size for the final selection.
- In terms of demand assignment, for the first 15 generations, ITDA is used and for the remaining 15 generations ITDA and ICCTDA are used.
- The value of ROR is set to 10%.
- The multipoint for crossover is set to 30 points.
- The number of iterations for ITDA is set to 4 iterations.

- The Singapore's experiment took 712,765 seconds with the Intel Xeon CPU E5-2640 v3 @ 2.6GHz with 128 GB RAM.
- Due to lack of information about the operational cost of PT services in Singapore, the operational cost parameters presented in Tab 5.3 are considered.
- The considered mode types and their characteristics are listed in Tab 6.2.

Mode	Vehicle capacity (units*seats)	Minimum headway (seconds)	Maximum headway (seconds)	Maximum offered line capacity (Seats/hour)	Speed (km/h)
Minibus	1*40	360	1,800	400	19
Standard bus high floor	1*60	360	1,800	600	19
Standard bus low floor	1*80	360	1,800	800	19
Articulated bus-1	1*100	360	1,800	1,000	19
Articulated bus-2	1*120	360	1,800	1,200	19
High capacity bus	1*150	360	1,800	1,500	19
Streetcar ROW C	1*200	300	1,800	2,400	22.2
Bus rapid transit	1*250	300	1,800	3,000	22.2
Streetcar ROW B-1	3*110	300	1,800	3,960	22.2
Streetcar ROW B-2	4*110	300	1,800	5,280	22.2
Rapid transit-1	4*150	180	1,800	6,000	34
Rapid transit-2	6*150	180	1,800	13,500	34
Rapid transit-3	8*150	180	1,800	24,000	34
Rapid transit-4	6*300	180	1,800	36,000	34
Rapid transit-5	8*300	180	1,800	48,000	34
Rapid transit-6	10*300	180	1,800	60,000	34

Tab 6.2 Mode type and corresponding characteristics for Singapore's public transit network

Due to high stop density surrounding terminals, a small proportion (10%) of stops are not covered with given ROR value and TNI procedure. Therefore, the process of RORRI and TNI are repeated continuously with different values of ROR until all the stops are covered and the created PT network is feasible (see section 4.1.3.3). The selected ROR value is increased from 10% up to 50% with a step size of 10%.

6.5.2 Results and evaluation

The results from Singapore network with ROR value of 10% are listed in the Tab 6.3 together with the rest of the TNE indicators. The results suggest high ROI values in the proposed PT network, this is due to: 1) TNF criteria, where all four conditions are required to be satisfied; 2) high stop density in the vicinity of terminals and city centre. All three PT network configurations show similarity, which suggests that for given inputs and considered constraints, there aren't many varied network configurations possible.

ROR	Perspectives of stakeholders	Performance indicators								
		ROI	ESH	ANT	RSC	UDP	TNL	ATT	OC	TND
	Passenger	■		■		■		■		■
	Agency	■	■		■		■		■	■
	Authority	■	■	■	■	■	■	■	■	■
	Best solutions									
10%	Passenger	1.85	249830	**1.01**	338640	**0.1**	5867	**32.1**	183359	**1.18**
	Agency	**1.64**	**244697**	1.07	333280	0.12	**5166**	32.8	**169951**	1.19
	Compromise	1.65	245313	1.06	**333240**	0.11	**5166**	32.4	175005	1.2

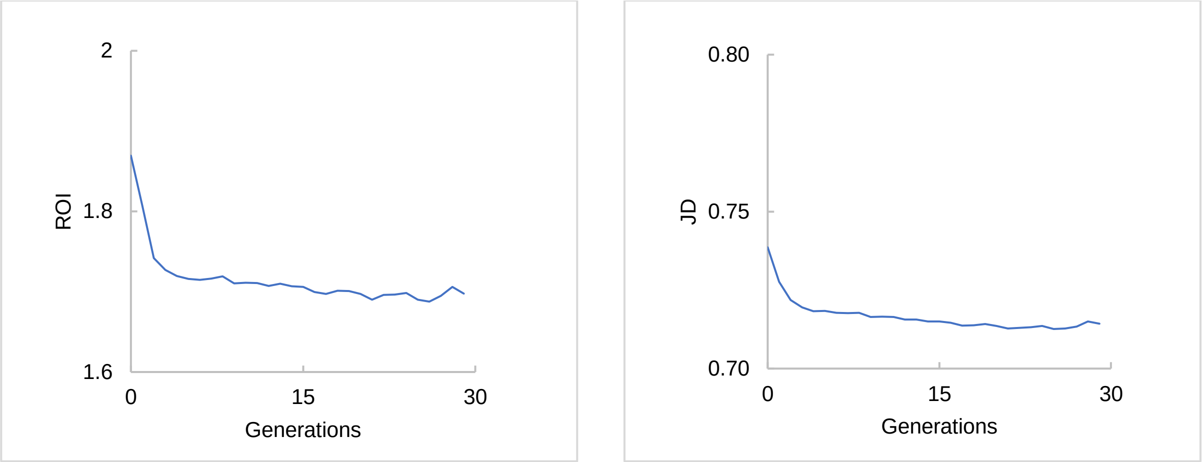
Tab 6.3 Objective function values and performance indicators for Singapore network with ROR value of 10%

A direct comparison between the current Singapore’s PT network and the proposed PT networks is not possible due to a number of reasons such as: 1) data availability; 2) land use factors; 3) historical PT network evolution. However, some of the indicators can be compared. Therefore, the best PT network representing passenger’s perspective is selected with following indicators:

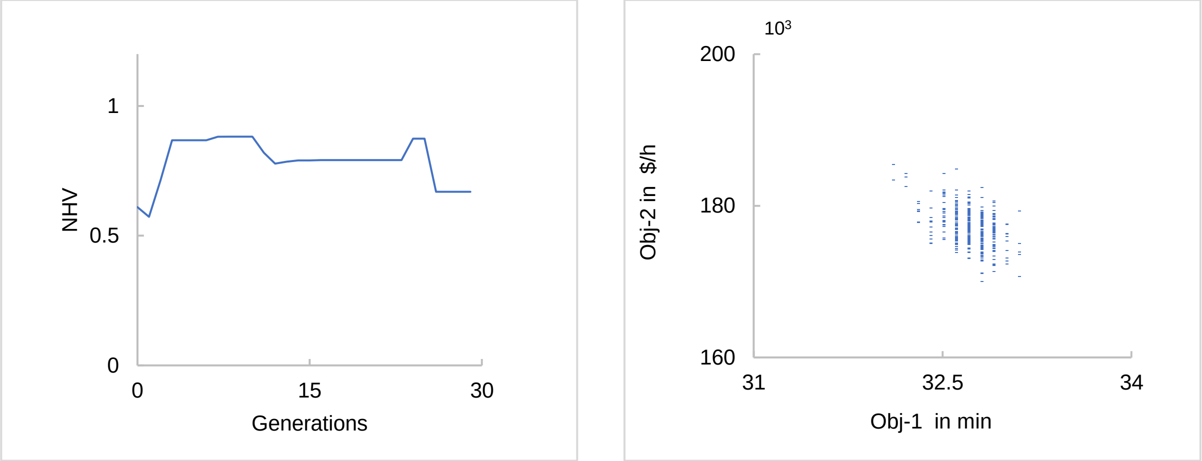
- ATT, since not all travel time components are available only in-vehicle time and transfer waiting time in minutes can be compared. The proposed PT network offers on average 20.03 in-vehicle time and 3.69 transfer wait time compared to 21.9 in-vehicle time and 2.03 transfer wait time. This sums up to total travel time of 23.72 minutes for the proposed model and 23.93 minutes from the current PT network.
- ANT, the proposed PT network offers on average 1.01, transfers per trip compared to 0.6 transfers per trip. This shows, the proposed network requires more transfers to complete the trips.
- TNL, the length of proposed PT network is 5,867 km compared to 11,600 km length of current PT network. This shows significant reduction in the PT network length compared to existing one, but at the expense of higher number of transfers.
- The proposed PT network is composed of 303 bus routes compared to 380 routes for existing PT network. This suggest the existing PT network can be improved further yet maintaining similar service levels.

The performance indicators of NSGA-II along with total feasible solutions and pareto front with a ROR value of 10% are listed in Fig 6.5. The ROI shows higher values, which hovers around 1.65. This is partly due to the TNF criteria (see section 4.1.3.3) and limited number of terminals. The JD Indicator shows higher values, which hover around 0.725. This shows higher diversity among the generated PT networks. The NHV, also hovers around 0.7 at the end of evolution. A total of 380 feasible solutions are generated during the evolution period. Moreover, the final pareto front includes 8 solutions.

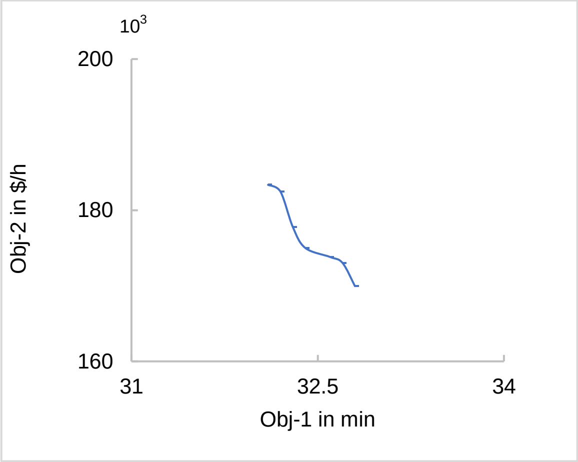
A selected PT network created by the proposed methodology representing passenger perspective is listed in Fig 6.6. The route layout is depicted in Fig 6.6a, which also includes existing MRTs and LRTs. The route layers are depicted in Fig 6.6b which includes bus network, BRT/LRT network and MRT network. The demand flow for one hour is depicted in Fig 6.6c, which is similar to existing demand flow pattern.



Route overlap index Jaccard distance

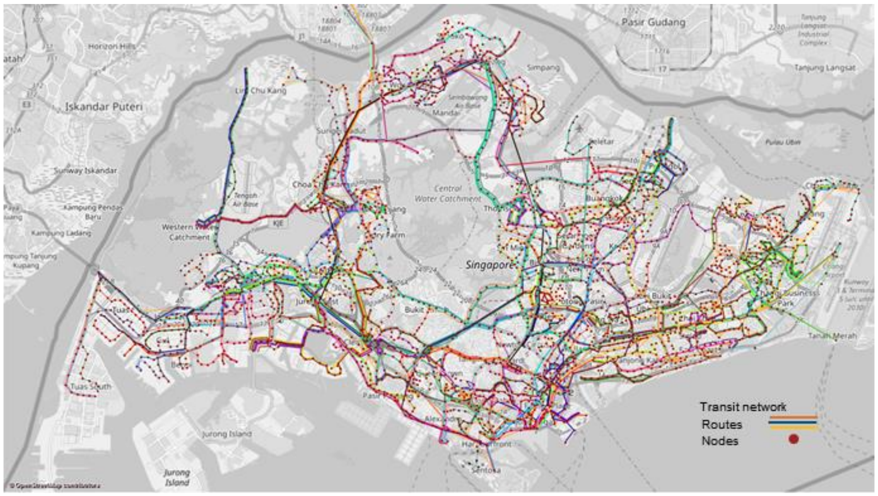


Normalised hyper volume Total feasible solutions

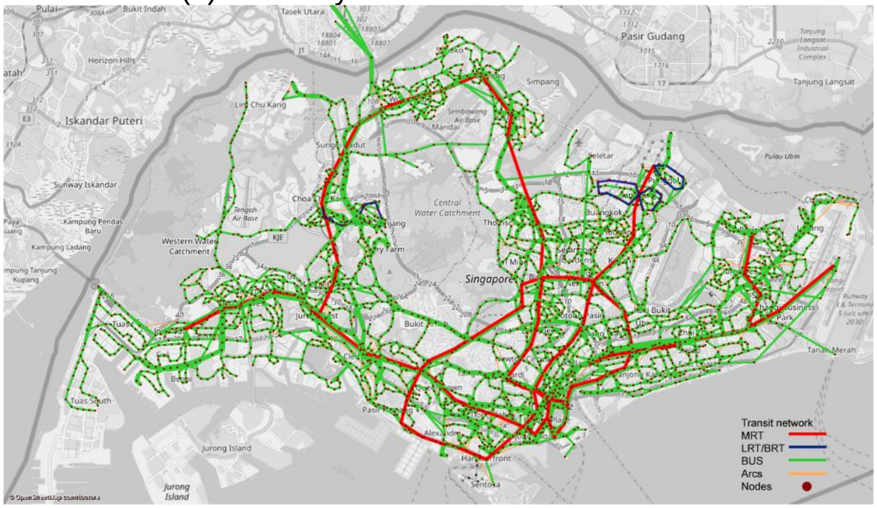


Pareto front

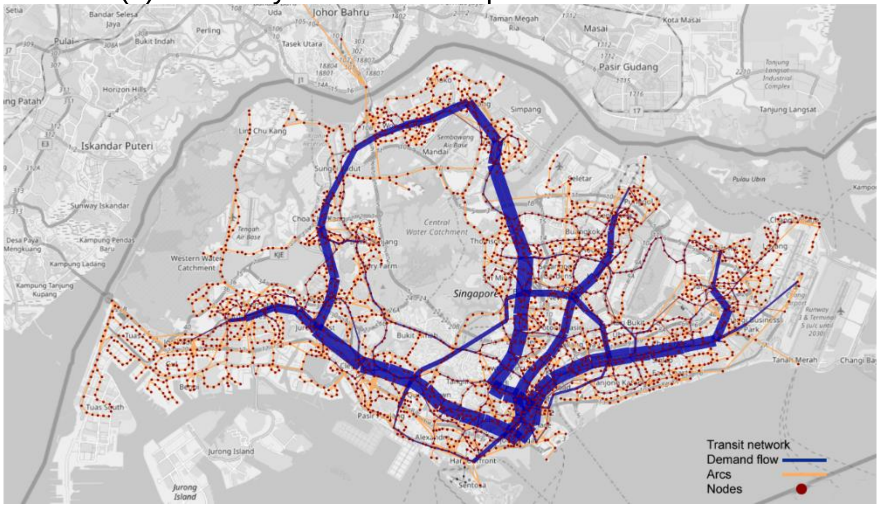
Fig 6.5 Indicators related to NSGA-II performance (ROI, JD and NHV), feasible solutions and pareto front for Singapore network with ROR value of 10%



(a) Route layout of selected PT network



(b) Route layers of selected public transit network



(c) Relative demand flow on selected public transit network

Fig 6.6 Proposed public transit network for passengers along with route layouts, route layers and demand flow patterns

6.6 Discussion and summary

The whole exercise of the implementation of developed methodology on Singapore's PT network is purely from the route-design perspective and for testing and evaluation of the developed methodology. The results show:

- The developed methodology is capable of handling large size actual networks. In considered network, there are roughly 3000 nodes and 9000 arcs.
- The concept of route overlap-based PT network creation shows promising prospects. A 20% reduction in the number of routes and 49% reduction in total network length compared to current PT network. The only downside is the number of transfers, which doubled from 0.6 to 1.0 per average trip.
- The network configurations generated by proposed methodology are comparable to the current PT network. This is demonstrated by similarity in selected travel time components.

This is a strong validation of proposed methodology on actual size network. Nonetheless there are other factors, not considered in the study, that may be an obstacle for implementation such as land use, local laws, labour policies and transport planning guidelines. Overall, the methodology can be used to assist transport planning agencies in developing new PT networks or restructuring current networks. Moreover, the developed methodology can serve as a planning tool for negotiation with other stakeholders (passengers, agencies and authorities), involved with the creation of a new network or necessary changes in the existing one.

7 Summary, Conclusions and Outlook

7.1 Summary

This thesis presented a route overlap ratio-based network design methodology for creating public transit networks, considering the perspectives of the two main stakeholders i.e., passengers and agencies. The importance of PT system is highlighted in the beginning of the thesis where the motivation for this thesis was briefly mentioned. Moreover, the goals and the contribution of this thesis were also emphasised.

After the introduction, the literature review of TNDFSP is presented where different approaches and methodologies were discussed. Within these methodologies, different combinations of their procedures were also explained. Other important ingredients of TNDFSP such as the type of transit demand assignment and PT network evaluation were also examined. At the end, some of the key limitations of the current state of the art literature were highlighted. These include: 1) disregard for the multiclass modes for network design; 2) simplistic demand assignment procedures instead of realistic assignment procedures with limited PT supply capacity; 3) the atomistic evaluation of the created PT networks instead of holistic evaluation; 4) lack of quality adherence with variation in network instances; 5) lack of real-world implementation of the proposed methodologies. These limitations act as the starting point for the proposed PT network design methodology. Almost all of these gaps were bridged in some form in the proposed methodology.

The problem definition and the model formulation for TNDFSP include the basic graph structure and the related terms such as network, path, shortest path and route were explained. The two objectives considered in this thesis, one belongs to passengers and other belongs to agency. Both considered objectives give a realistic viewpoint of the stakeholders and their costs. Passenger's cost includes all the typical travel time components such as waiting time and in-vehicle time. For agency, operational costs such as on-vehicle crew cost, operating cost and infrastructure maintenance cost are considered.

Perhaps the most important highlight of this thesis is the proposed metaheuristic-based approach to TNDFSP using ROR concept with NSGA-II. The main idea is to come up with a set of procedures working in a sequential order to solve TNDFSP. The basic concept used to create PT network is based on route overlap concept. Though a simple concept but highly capable of creating different types of PT network configurations such as a sparse network or a dense network. Furthermore, to fully tap the potential of this concept, demand based multiclass modes are introduced. In order to guarantee feasibly of the PT network, simple network improvement procedures are used. To assess the quality of created PT network, a holistic evaluation must be performed. However, before that, selection of the most suitable transit demand assignment is necessary. In this thesis, headway-based transit demand assignment with optimal strategies as route choice model is adopted with both unlimited and limited PT supply capacities. This assignment procedures provides more realistic behaviour of the passengers in the PT network and the results are used to performed holistic evaluation, considering all three perspectives of passengers, agencies and authorities. The whole set of procedures along with NSGA-II is used to search for better PT networks, and to provide a variety of options for planners and policy makers to select different PT network configurations (pareto fronts).

Once the methodology is developed the next step is to test it, for that, a number of benchmark networks were selected. The methodology was implemented on five different benchmark networks with different network configurations and demand patterns. To establish the credibility of the proposed methodology, the results must be compared with other related studies. Therefore, the results were compared with other studies which used the same benchmark networks. Moreover, to perform a fair comparison, the created PT networks by other studies were subjected to the newly created PT demand assignment procedures and transit network evaluation. These benchmark networks provide a fair playground to test the newly developed methodologies. However, the real test for any methodology is its ability whether it can be implemented on a real case study or not. Thanks to the availability of the required data and information of Singapore's PT network, the methodology was tested on the Singapore's PT network. The results are promising, and the generated PT network configurations are comparable to the existing PT network.

7.2 Conclusions

This thesis presented a PT network design methodology, which belongs to the strategic planning level within PT system planning. The main intention of this thesis is to provide a systematic set of heuristic and metaheuristic sequential procedures to design, evaluate and improve large urban PT networks.

The proposed PT network design methodology consist of four main components: 1) route and network creation; 2) demand assignment; 3) network evaluation; 4) metaheuristic-based search engine. Although, these components are not new, several research studies have already used similar terms. In many ways, the developed methodology heavily overlaps with previous methodologies. However, some of the major features which differentiate the proposed methodology from the existing methodologies are as follows:

1. The PT network creation is not a simple selection of routes from feasible route set. Instead, it uses a novel concept of ROR. This concept allows different PT network configurations such as with lower values of ROR, a simple high capacity PT network can be created. With higher values of ROR, a more complex, direct, low capacity PT network can be created.
2. The concept of ROR is further strengthened by introducing demand-based multiclass modes, such that the demand on a route will decide what type of mode is most suitable. Therefore, providing a window to create/restructure PT service concepts such as hub-and-spoke structure or to provide direct connections to low density sub centres.
3. The concept of demand-based multiclass modes is appropriate, only if one can estimate the demand flow over the PT network. This can be achieved by following: 1) the selection of a realistic demand assignment and route choice model; 2) unlimited and limited PT supply capacity. In the developed methodology, both are considered and implemented in this thesis. An iterative PT capacity free and capacity constrained headway-based transit demand assignment with optimal strategy as route choice model is used.

4. The demand assignment, route mode type and headway provide a valuable information which can be used to evaluate the PT network. However, atomistic evaluation is not enough, instead a holistic evaluation is required. A set of nine different performance indicators representing the viewpoints of all three stakeholders i.e., passengers, agencies and authorities are used to evaluate the created PT network.
5. Given multi-objective nature of the TNDFSP, no one solution can satisfy passengers and agencies together. However, a set of pareto optimal solutions seems more suitable for planners and policy makers to choose the most suitable PT network. Therefore, NSGA-II is adopted to get a set of pareto optimal solutions for the PT networks having different ROR values.
6. Real world implementations were very rare, mainly due to lack to data availability and/or data access from different transit authorities and agencies. Recently, data availability provides the opportunity to test whether the developed methodology works on a realistic PT network or not. The developed network design methodology was implemented on the PT network of Singapore.

The main conclusions drawn from this thesis includes the use of ROR concept to create variety of PT network configurations suitable for both passengers and agencies for different sizes of road networks. For small networks with sparse demand, the networks for passengers and agencies are quite different. The PT networks which suit passengers, requires more routes with increased lengths to offer attractive travel times. Whereas, the networks which suits agencies require least redundant routes with increased headways, but they are not attractive to the passengers. For large networks, the least redundant PT networks with high capacity and low headways are suitable for both passengers and agencies. Such PT networks require least number of routes along with multiple PT layers including bus routes, BRTs/LRTs and MRTs, all operating at lower headways.

In term of solution quality (PT network) comparison with the rest of the studies, the results are compared with other related studies. Different benchmark network shows varying results. The comparative results for smaller networks show that the developed methodology offers comparable results to that of other studies. However, it shows superior performance in the considered objectives, especially in the larger instances of the PT network. This is mainly due to the low demand requirements for the small networks, thus the concept of multiclass mode is not used to its fullest potential. However, it is totally opposite for large networks. Moreover, one of the main limitations of the current methodologies was lack of consistency in solution quality when network size is varied. The results from the developed methodology show that the solution quality remain consistent for small and large network.

The real-world case study also suggests the practical viability of the proposed methodology. The results show that significant reduction in the number of routes can be achieved yet offering similar or in many cases better performance. Since, the PT network design and planning is a complex task. A lot of other factors should also be considered such as land use, socioeconomic situation, local policies and passenger behaviour during design process. Nevertheless, the results show that such methodology can aide the PT network planning agencies and authorities to plan, design and evaluate the PT networks.

7.3 Outlook

The developed methodology and the results suggest that it can be of great help for PT network planning authorities. However, it is not perfect, and several areas are worth investigating for further research. The proposed methodology is data driven and as other data driven model, the result heavily dependent upon the quality of the data. Implementation on benchmark network is one thing and real-world implementation another. The developed methodology provides the basic building blocks and procedures for the PT network planning. However, a lot of parametric values and data need to be acquired by using other methods. Moreover, their processing, calibration, authenticity and applicability must be established before use.

Starting from the transit route construction, there exist plenty of route creation methodologies, but none of them were able to establish their superiority on others. One can’t assess the quality of the routes from typical PT network evaluation and it doesn’t reveal any insights into the route creation. However, route creation is of great importance, and if PT planning guidelines are considered while planning the routes. A significant reduction in: 1) the number of total feasible routes; 2) computation cost involved in route selection, can be achieved. Such considerations can significantly reduce the computation cost for both route creation and route selection

The concept of ROR can be strengthened further by incorporating new concepts such as: 1) layer-based PT network with different overlaps; 2) coordinated in-vehicle transfers. A PT network with minimum overlap for the MRTs can be created, this will serve as the backbone network. Later, a second network with higher route overlap can be created to complement the main network, which act as feeder. Through such layer-based networks, different network configurations can be created and integrated together. The coordinated in-vehicle transfers can assist in eliminating the need for transfers. This can be done when two routes overlap for certain segments and provide a time window for the transfers. Vehicle coupling technology can be used for seamless transfer. When the overlap ends, the vehicles separate and on to their route course. Through such concept, minimum routes would be required for the service and the focus would be more on reliability and seamless transfers. Moreover, such service can increase PT service attractiveness.

Due to the conflicting interests of different stakeholders, MOEAs are perhaps the most suitable algorithms to solve TNDFSP to get acceptable quality solutions, and with reasonable computation time. However, the question, which one of these algorithms is most suitable for TNDFSP is still unanswered. The selection and performance of the available state of the art MOEAs should be investigated along with which type of genetic operators such as mutation and crossover should be used.

The assumption that the stops in the opposite direction always face each other is fairly simple but not realistic. Often the case, opposite stops are away from each other thus required excessive walking. This happens when there is a separation between the two road, and the access to/from stops either done via next signalised junction or via overhead bridge. From a passenger perspective, transferring from one route to another is perhaps one of the most important decision. If the transfers are not easy and fast, such PT network may lose its attractiveness quickly. Therefore, stop locations and transfer behaviour must be researched thoroughly and integrated into network design methodology.

The demand data used for transit demand assignment and route choice model perhaps play the most important role in the whole process. Typical consideration for demand data is assumed to have a whole day demand data or peak hour demand data. However, neither peak period nor whole day demand matrix are portraying the true picture of demand. The peak period generally provides high volumes directed towards the city centres and sub centres. The spread of such peak demand is very narrow, and it doesn't show demand in other areas of the considered region. Similarly, average demand for the whole day has widespread, and it is inclusive of all the demand patterns. However, it underestimates the demand in areas where, there is a heavy load in the peak period and overestimating the demand in areas, with low demand. One must use multiple demand matrices for designing a PT network with spatio-temporal consideration. Such as, first the network design for the peak period demand and associated regions. Later, the PT network is extended to incorporate the normal/low demand regions. Such considerations can provide a PT network which can assist in the optimal use of resources, and it can cater peak and off-peak demand appropriately.

The preferences of the passengers for path choice within the PT network for route choice are heavily influence by local conditions and culture. Moreover, with the availability of the PT trips data, weights related to travel time components such as in-vehicle time, waiting time, transfer time and transfers can be estimated and incorporated into developed methodology. This can help in improving the credibility and authenticity of the results generated by the developed methodology.

The computation cost is one of the major issues that can act as a bottleneck especially if very large network is considered. Moreover, interaction with external software packages and standalone programs also lead to increased computation time. The execution time can be reduced to certain extent by incorporation of concepts such as parallelism and distributed computing. Furthermore, parallel-MOEA (pMOEA) versions of MOEA are available, they can help in reducing the overall computation time.

The future trends and technologies will leave a great impact on traditional PT networks and services. The emergence of new PT services and their structures will radically change the PT services. With the advancements in the autonomous driving capabilities, and technologies such as smart infrastructure, Car2Car and Car2Infrastructure communications and personal mobility devices (PMD). Moreover, growing interest in micro transit and establishment of shared mobility services will play a major role in reshaping PT network and act as a catalyst for PT transformation.

In near future, the main focus would be on the seamless integration of a variety of PT services. Whether, it is on demand modular micro transit service, use of shared PMDs or shared door-to-door mobilities. The integration would be such that, a passenger will require minimum transfers, or seamless in-vehicle transfer while availing door to door service. Either, the main part of the trip would be served with optimum sized vehicles and service level with complementing first-last mile shared autonomous PMDs, or the autonomous modular vehicles would combine to serve the main leg and split to individual modules to serve the individual destinations. In both cases, the passenger would be on ease, and will be able to significantly reduce the total travel time cost as compared to travel times offered by current PT services.

In terms of PT network, only the MRT networks will remain active, because of high capacity support, dedicated ROW and high capital investment. The typical bus routes will see a drastic change, instead of fixed routes operating on fixed schedules with fixed vehicle size, the routes will be dynamic and depending upon the spatio-temporal demand, the routing, schedule and vehicle type will be calculated on the fly. On top of that, the concept of stop to stop service will be replaced by door to door service. There won't be any stops, instead the pickups and drop-offs locations will be known in advanced. The backbone of the service would be a comprehensive, routing, scheduling, dispatching and booking system.

The question of how to serve varied demand across different time periods, can be answered by autonomous modular multi-size vehicles-based PT service. Such concept can capture better of both worlds, with capacity matching that of a BRT with door-to-door comfort of a private car. For instance, the individual vehicles gather all the passengers from different regions and later stack up to create required capacity service. Now the comfort of car is achieved together with the desired capacity with stacking up of vehicles together. Next thing is to achieve the exclusive ROW, this can be achieved by utilising the availability of the smart infrastructure and high penetration of autonomous vehicles. Through this, high operational speed and maximum utilisation of the road capacity can be achieved. Later, based on the destinations, different vehicles split from the stack and serving the door to door service again. The same concept with different varied capacities can be used all day long. If the demand patterns change, the PT service also readapt itself to cope with such changes. The size of the vehicle can be of different sizes with smallest being just a single person. If such concept is implemented uniformly across whole city, a paradigm shift in the private car ownership can be achieved. This means, no one will own a personal vehicle rather people will share different form of mobilities and use more PT services.

References

Arbex, R. O., & Cunha, C. B. [2015]: Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. Transportation Research Part B: Methodological, 81, 355–376.

Australian Transport Council. [2006]: National Guidelines for Transport System Management in Australia. transportinfrastructurecouncil.gov.au.

Baaj, M. H., & Mahmassani, H. S. [1991]: An AI-based approach for transit route system planning and design. Journal of Advanced Transportation, 25(2), 187–209.

Baaj, M., & Mahmassani, H. [1995]: Hybrid route generation heuristic algorithm for the design of transit networks. Transportation Research Part C: Emerging Technologies, 3(1), 31–50.

Bagloee, S. A., & Ceder, A. [2011]: Transit-network design methodology for actual-size road networks. Transportation Research Part B: Methodological, 45(10), 1787–1804.

Buba, A. T., & Lee, L. S. [2016]: Differential Evolution for Urban Transit Routing Problem. Journal of Computer and Communications, 04(14), 11–25.

Bussieck, M. [1998]: Optimal lines in public rail transport.

Cancela, H., Mauttone, A., & Urquhart, M. E. [2015]: Mathematical programming formulations for transit network design. Transportation Research Part B: Methodological, 77, 17–37.

Carrese, S., & Gori, S. [2002]: An Urban Bus Network Design Procedure. Transportation Planning SE - 11, 64(1995), 177–195.

Ceder, A. [2016]: Public Transit Planning and Operation: Modeling, Practice and Behavior (Second). CRC Press.

Ceder, A., & Wilson, N. [1986]: Bus network design. Transportation Research Part B: Methodological, 208(4), 331–344.

Chakroborty, P., & Wivedi, T. [2002]: Optimal Route Network Design for Transit Systems Using Genetic Algorithms. Engineering Optimization, 34(1), 83–100.

Chew, J. S. C., Lee, L. S., & Seow, H. V. [2013]: Genetic Algorithm for Biobjective Urban Transit Routing Problem. Journal of Applied Mathematics, 2013(2013), 1–15.

Chu, J. C. [2018]: Mixed-integer programming model and branch-and-price-and-cut algorithm for urban bus network design and timetabling. Transportation Research Part B: Methodological, 108, 188–216.

Cipriani, E., Gori, S., & Petrelli, M. [2012]: Transit network design: A procedure and an application to a large urban area. Transportation Research Part C: Emerging Technologies, 20(1), 3–14.

Constantin, I., & Florian, M. [1995]: Optimizing Frequencies in a Transit Network: a Nonlinear Bi-level Programming Approach. International Transactions in Operational Research, 2(2), 149–164.

Deb, K. [2001]: Multi-objective optimization using evolutionary algorithms. John Wiley & Sons.

Deb, K., & Mayank, G. [1996]: A combined genetic adaptive search (GeneAS) for engineering design. Computer Science and Informatics, 26, 30–45.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. [2002]: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182–197.

Dijkstra, E. [1959]: A note on two problems in connexion with graphs. Numerische Mathematik, 1(1), 269–271.

Fan, L., & Mumford, C. L. [2008]: A metaheuristic approach to the urban transit routing problem. Journal of Heuristics, 16(3), 353–372.

Fan, L., Mumford, C. L., & Evans, D. [2009]: A simple multi-objective optimization algorithm for the urban transit routing problem. 2009 IEEE Congress on Evolutionary Computation, 1–7.

Fan, W., & Machemehl, R. B. [2006]: Optimal tranist Route Network Design Problem with Variable Transit Demand : Genetic Algorithm Approach. Journal of Transportation Engineering, 132(1), 40–51.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., & Rashidi, H. [2013]:A review of urban transportation network design problems. European Journal of Operational Research, 229(2), 281–302.

Floyd, R. W. [1962]: Algorithm 97: Shortest Path. Commun. ACM, 5(6), 344–348.

Furth, P. G., & Wilson, N. H. [1981]: Setting frequencies on bus routes: Theory and practice. Transportation Research Record, 818, 1–7.

Guan, J. F., Yang, H., & Wirasinghe, S. C. [2006]: Simultaneous optimization of transit line configuration and passenger line assignment. Transportation Research Part B: Methodological, 40(10), 885–902.

Guihaire, V., & Hao, J.-K. [2008]: Transit network design and scheduling: A global review. Transportation Research Part A: Policy and Practice, 42(10), 1251–1273.

Hart, P., Nilsson, N., & Raphael, B. [1968]: A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Transactions on Systems Science and Cybernetics.

Holland, J. H. [1975]: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press Books.

Hosapujari, A. B., & Verma, A. [2013]: Development of a Hub and Spoke Model for Bus Transit Route Network Design. Procedia - Social and Behavioral Sciences, 104, 835–844.

Huang, D., Liu, Z., Fu, X., & Blythe, P. T. [2018]: Multimodal transit network design in a hub-and-spoke network framework. Transportmetrica A: Transport Science, 14(8), 706–735.

Ibarra-Rojas, O. J., Delgado, F., Giesen, R., & Muñoz, J. C. [2015]: Planning, operation, and control of bus transport systems: A literature review. Transportation Research Part B: Methodological, 77, 38–75.

Israeli, Y., & Ceder, A. [1989]: Designing transit routes at the network level. Conference Record of Papers Presented at the First Vehicle Navigation and Information Systems Conference (VNIS ’89), 310–316.

John, M. P. [2016]: Metaheuristics for Designing Efficient Routes & Schedules For Urban Transportation Networks.

Kechagiopoulos, P. N., & Beligiannis, G. N. [2014]: Solving the Urban Transit Routing Problem using a particle swarm optimization based algorithm. Applied Soft Computing Journal, 21, 654–676.

Kepaptsoglou, K., & Karlaftis, M. [2009]: Transit route network design problem: review. Journal of Transportation Engineering, 135(8), 491–505.

Kiliç, F., & Gök, M. [2014]: A demand based route generation algorithm for public transit network design. Computers and Operations Research, 51, 21–29.

Kim, H.-S., Kim, D.-K., Kho, S.-Y., & Lee, Y.-G. [2016]: Integrated decision model of mode, line, and frequency for a new transit line to improve the performance of the transportation network. KSCE Journal of Civil Engineering, 20(1), 393–400.

Kuo, Y. [2013]: Design method using hybrid of line-type and circular-type routes for transit network system optimization. TOP, Spanish Society of Statistics and Operations Research, (84), 1–14.

Lam, W. H. K., & Bell, M. G. H. [2002]: Advanced Modeling for Transit Operations and Service Planning. Emerald Group Publishing Limited.

Lampkin, W., & Saalmans, P. D. [1967]: The Design of Routes, Service Frequencies, and Schedules for a Municipal Bus Undertaking: A Case Study. Journal of the Operational Research Society, 18(4), 375–397.

Lee, Y., & Vuchic, V. [2005]: Transit network design with variable demand. Journal of Transportation Engineering, 131(1), 1–10.

Liu, X., Zhou, Y., & Rau, A. [2018]: Smart card data-centric replication of the multi-modal public transport system in Singapore. Journal of Transport Geography.

Majima, T., Takadama, K., Watanabe, D., & Katuhara, M. [2014]: Application of Community Detection Method to Generating Public Transport Network. In Proceedings of the 8th International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS) (pp. 243–250). ACM.

Majima, T., Takadama, K., Watanabe, D., & Katuhara, M. [2015]: Characteristic of Passenger's Route Selection and Generation of Public Transport Network. SICE Journal of Control, Measurement, and System Integration, 8(1), 67–73.

Mandl, C. [1980]: Evaluation and optimization of urban public transportation networks. European Journal of Operational Research, 5, 396–404.

Mauttone, A., & Urquhart, M. E. [2009]: A route set construction algorithm for the transit network design problem. Computers & Operations Research, 36(8), 2440–2449.

Mumford, C. L. [2013]: New heuristic and evolutionary operators for the multi-objective urban transit routing problem. 2013 IEEE Congress on Evolutionary Computation, CEC 2013, 939–946.

Nayeem, M. A., Rahman, M. K., & Rahman, M. S. [2014]: Transit network design by genetic algorithm with elitism. Transportation Research Part C: Emerging Technologies, 46, 30–45.

Nes, R. van, Hamerslag, R., & Immers, B. [1988]: Design of public transport networks. Transportation Research Record, 1202, 74–83.

Ngamchai, S., & Lovell, D. J. [2003]: Optimal Time Transfer in Bus Transit Route Network Design Using a Genetic Algorithm. Journal of Transportation Engineering, 129(5), 510–521.

Nikolić, M., & Teodorović, D. [2013]: Transit network design by Bee Colony Optimization. Expert Systems with Applications, 40(15), 5945–5955.

Nikolić, M., & Teodorović, D. [2014]: A simultaneous transit network design and frequency setting: Computing with bees. Expert Systems with Applications, 41(16), 7200–7209.

Ortuzar, J., & Willumsen, L. [1994]: Modelling Transport (4th ed.). JohnWiley & Sons, Ltd.

Pattnaik, S. B., Mohan, S., & Tom, V. M. [1998]: Urban Bus Transit Route Network Design Using Genetic Algorithm. Journal of Transportation Engineering, 124(4), 368–375.

Petrelli, M. [2004]: A transit network design model for urban areas. Publication of: WIT Press.

PTV. [2017]: PTV Visum 17ptv – Manual.

Rea, J. C. [1971]: Designing Urban Transit System: An Approach to the Route Technology Selection Problem.

Reserve Bank of Australia. [2017]: RBA: Inflation Calculator.

Schéele, S. [1980]: A supply model for public transit services. Transportation Research Part B: Methodological, 14(1–2), 133–146.

Schöbel, A. [2012]: Line planning in public transportation: Models and methods. OR Spectrum.

Schöbel, A., & Scholl, S. [2005]: Line planning with minimal traveling time. In L. G. Kroon & R. H. Möhring (Eds.), 5th Workshop on Algorithmic Methods and Models for Optimization of Railways (Vol. 1).

Shih, M.-C., Mahmassani, H., & Baaj, M. [1997]: Trip Assignment Model for Timed-Transfer Transit Systems. Transportation Research Record, 1571(1), 24–30.

Shih, M. C., & Mahmassani, H. S. [1994]: Design methodology for bus transit networks with coordinated operations. Research report. Texas Univ., Austin, TX (United States). Center for Transportation Research.

Silman, L. A., Barzily, Z., & Passy, U. [1974]: Planning the route system for urban buses. Computers & Operations Research, 1(2), 201–211.

Spiess, H., & Florian, M. [1989]: Optimal strategies: A new assignment model for transit networks. Transportation Research Part B: Methodological.

Szeto, W. Y., & Wu, Y. [2011]: A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. European Journal of Operational Research, 209(2), 141–155.

Talbi, E. [2009]: Metaheuristics: from Design to Implementation. John Wiley & Sons Ltd.

Tom, V., & Mohan, S. [2003]: Transit route network design using frequency coded genetic algorithm. Journal of Transportation Engineering, 129(2), 186–195.

Ul Abedin, Z., Busch, F., Wang, D. Z. W., Rau, A., & Du, B. [2018]: Comparison of Public Transport Network Design Methodologies Using Solution-Quality Evaluation. Journal of Transportation Engineering, Part A: Systems, 144(8), 04018036.

Ul Abedin, Z. [2019]. Experiments. [Data files]. Retrieved from https://www.dropbox.com/sh/pl21bj1zf5411c2/AABLeOYIO05Oia2j7Jwg6HvFa?dl=0

Vuchic, V. R. [2005]: Urban Transit: Operations, Planning and Economics. Wiley.

Vuchic, V. R. [2007]: Urban Transit Systems and Technology. John Wiley & Sons, Inc. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Wan, Q. K., & Lo, H. K. [2003]: A Mixed Integer Formulation for Multiple-Route Transit Network Design. Journal of Mathematical Modelling and Algorithms, 2(4), 299–308.

Yen, J. Y. [1971]: Finding the K Shortest Loopless Paths in a Network. Management Science, 17(11), 712–716.

Zhang, J., Lu, H., & Fan, L. [2010]: The multi-objective optimization algorithm to a simple model of urban transit routing problem. 2010 Sixth International Conference on Natural Computation, (Icnc), 2812–2815.

Zhao, F., & Zeng, X. [2006]: Optimization of transit network layout and headway with a combined genetic algorithm and simulated annealing method. Engineering Optimization, 38(6), 701–722.

Zhao, H., Xu, W. (Ato), & Jiang, R. [2015]: The Memetic algorithm for the optimization of urban transit network. Expert Systems with Applications, 42(7), 3760–3773.

Abbreviations

AB	Available budget
ACO	Ant-colony optimisation
ANT	Average number of transfers
AON	All or nothing
ATT	Average travel time
BCO	Bee colony optimisation
BM	Benchmark
BRT	Bus rapid transit
CAM	Custom assignment model
CC	Cloud computing
CCHWAM	Capacity constrained headway-based assignment model
CCO	Crowded comparison operator
CCOS	Crowded comparison operator-based selection
CD	Crowding degree
CDA	Crowding distance assignment
CDAM	Combined distribution assignment model
CEPAS	Contactless e-purse application
DS	Demand satisfaction
DS 0/1/2	Demand satisfaction through 0, 1, 2 transfers
DT	Direct trips
ESH	Empty seat hours
FiC	Fictious
FNS	Fast nondominated sort
FS	Fleet size
FTRS	Feasible transit route set
GA	Genetic algorithms
GIS	Geographical information system
H	Heuristics
HDMS	Headway derivation and mode selection
HPC	High-performance computing
HV	Hyper volume
HWAM	Headway-based assignment model
ICCTDA	Iterative capacity constrained transit demand assignment
ITDA	Iterative transit demand assignment
JD	Jaccard distance
JI	Jaccard index
LF	Load factor
LP	Linear programming
LRT	Light rapid transit
LTA	Land transport authority
M	Mathematical

MAS	Multi agent system
MH	Metaheuristics
MILP	Mixed integer linear programming
MIP	Mixed-integer programming
MLS	Maximum load section
MOEA	Multi-objective evolutionary algorithm
MRT	Mass rapid transit
MSE	Metaheuristics search engine
MT	Mode type
NC	Network construction
ND	Network directness
NHV	Normalised hyper volume
NI	Network improvement
NoR	Number of routes
NoS	Number of stops
NoT	Number of transfers
NP-hard	Non-deterministic polynomial-time
NSGA-II	Nondominated sorting genetic algorithm-II
OC	Operator cost
P&OC	Passenger and operator cost
PAM	Probabilistic assignment model
PC	Passenger cost
PD	Path deviation
PM	Polynomial mutation
PMD	Personal mobility devices
pMOEA	Parallel-multi-objective evolutionary algorithm
PSO	Particle swarm optimisation
PT	Public transit
RBT	Route backtracking
RC	Route creation
RCO	Route configuration
RE	Route efficiency
RH	Route headway
RL	Route length
RLP	Route load profile
ROI	Route overlap index
ROR	Route overlap ratio
ROR-INI	Route overlap ratio-based initialisation
RORMC	Route overlap-based multipoint crossover
ROW	Right-of-way
RS	Route set
RsC	Route set connectivity
RSC	Required seating capacity

RSL	Route selection
SA	Simulated annealing
SLA	Singapore land authority
SPAM	Shortest path-based assignment model
TDAHD	Transit demand assignment and headway derivation
TND	Transit network directness
TNDFSP	Transit network design and frequency setting problem
TNDP	Transit network design problem
TNE	Transit network evaluation
TNF	Tranist network feasibility
TNI	Tranist network improvement
TNIFC	Transit network improvement and feasibility check
TNL	Total network length
TNRC	Transit network route container
TRE	Transit route extension
TRNC	Transit route and network construction
TRNI	Transit route node insertion
TS	Tabu search
TTT	Total travel time
UDP	Unsatisfied demand percentage
USD	Unsatisfied demand
VC	Vehicle capacity
WT	Waiting time

List of Figures

List of Tables

List of Algorithms

List of Publications

- Z. Ul Abedin and W. Rashid Ahmed. "Modelling Inductive Charging of Battery Electric Vehicles using an Agent-Based Approach." Journal of Sustainable Development of Energy, Water and Environment Systems 2.3 (2014): 219-233.
- Z. Ul Abedin and A. Rau, “A genetic algorithm based robust public transport network design model,” in Proceedings of the 5th Symposium arranged by the European Association for Research in Transportation (hEART), Delft, Netherlands, 2016.
- Z. Ul Abedin, A. Rau, F. Busch, and D. Z. Wang, “A hierarchical public transit network design method for large networks,” in Proceedings of the Transportation Research Board (TRB) 97th Annual Meeting, Washington D.C., USA, Jan. 2018.
- Z. Ul Abedin, F. Busch, D. Z. Wang, A. Rau and B. Du, “Comparison between public transport network design methodologies using solution quality evaluation,” ASCE's Journal of Transportation Engineering, Part A: Systems, Feb. 2018, ISSN: 2473-2907.
- Z. Ul Abedin, A. Rau and F. Busch, “A Route Overlap Ratio-Based Transit Network Design and Frequency Setting Model Using NSGA-II Algorithm,” in Proceedings of the Transportation Research Board (TRB) 98th Annual Meeting, Washington D.C., USA, Jan. 2019.