Application of Graph-Theoretic Indicators in Multi-Modal Public Transportation Networks

by

Bryan Whited

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Transportation Engineering

Department of Civil and Environmental Engineering
University of Alberta

ABSTRACT

This thesis extends the current research on graph theory-based techniques to meet the needs of planners of multimodal transit networks. Graph theory was originally developed in the 1700s to study transportation systems although it was not adapted to study public transportation networks until the 1980s. Since then it has been primarily used to study rail and metro transit networks, resulting in numerous metrics being developed. Bus networks much larger and operate very differently from rail networks and so while many researchers have applied those metrics to bus networks, there remains a notable gap between the needs of bus transit planners and the graph theory-based tools seen in the current literature. This gap is bridged using graph representations of Edmonton's multimodal transit network to understand the limitations of the current methodology and develop new methods specific to bus transit. The public transportation network operated by Edmonton Transit Service (ETS) is analyzed in five time periods, each of which is represented using four graph types, four edge types, both with and without pedestrian links for a total of thirty-two graph configurations. A common battery of indicators was calculated using each of these to investigate the effect of graph type, edge type, and pedestrian links, as well as changes in the network between time periods. The results indicate the metrics commonly used in the current literature provide more information about the choice in graph representation than in the underlying transit service. The needs of transit planners are identified, and four new metrics are developed specifically to meet those needs. Each of these metrics uses shortest paths through Time-Expanded graph representations of the Edmonton Transit Service (ETS) transit network. The simplest new metric, "Directional Speed" combines travel times with the distance travelled toward the downtown core. As stops are not equally important, the remaining three indicators are also weighted by ridership. "Weighted Potential" (WP) is a connectivity indicator which combines weighting by ridership with schedule availability between each stop-stop O-D pair. "Potential Travel Time" (PTT) is the average travel time between stop-stop O-D pairs, which is also weighted by ridership. The final metrics, "Effective Travel Time" (ETT), combines Weighted Potential and Potential Travel Time to provide a single indicator that reflects travel times and schedule availability. These metrics are used to quantify the quality of service in Edmonton's public transit network to both demonstrate the utility of the indicators and provide information to planners who are redesigning the bus network.

Acknowledgements

I could never have done this without the help and support from a number of individuals, to whom I owe many thanks. More than anyone, I want to thank my academic supervisor Dr. Karim El-Basyouny, who has provided guidance and endless patience since first convincing me to take on a thesis. His knowledge and advice have been critical to developing this thesis.

I also want to thank my coworkers at ETS, who provided data, inspiration, insight, and support over the past few years. Most notably, I want to thank my supervisor Andrew Gregory, and Directors Bill Sabey and Sarah Feldman, who have been flexible and supportive. Thanks also to Brian Korthius for countless conversations about the nuances and quirks of the GTFS and ridership data for the ETS network. Brian, along with Richard Leclerc, introduced me to KNIME and helped me obtain and clean numerous ridership data sets.

In addition, I want to thank Dr. Tae Kwon for his time, effort and willingness to serve on my examination committee, and Dr. Manish Shirgoakar for providing a very useful compilation of GIS data, in addition to serving on my examination committee.

Finally, I want to give special thanks to my wife Christy for her unwavering encouragement, and my friends and family who have been supportive and understanding over these past few years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
BTN	Bus Transit Network
CBD	Central Business District
CSV	Comma Separated Value text format
ETS	Edmonton Transit Service
FOSS	Free, Open Source Software
GTFS	General Transit Feed Specification
LOS	Level of Service
O-D or OD	Origin-Destination pairs
RM-C or RM-Complex	Route Map Complex graph type
RM-S or RM-Simple	Route Map Simple graph type
SLIM	Spatial Land Inventory Management
TAZ	Traffic Analysis Zone
TCSQM	Transit Capacity and Quality of Service Manual
TCRP	Transit Cooperative Research Program
TE	Time Expanded graph type
TM	Trip Map graph type
TT	Travel Time edge weighting type
TRB	Transportation Research Board
UD	Unweighted, Directed edge weighting type
UU	Unweighted, undirected edge weighting type
VoT	Value of Time edge weighting type

LIST OF VARIABLES & METRIC VALUES

Symbol	Variable / Indicator
α	Alpha-index, also called cyclicity
β	Beta index, also called complexity
δ	Maximum number of transfers to cross the network
ξ	Transfer Efficiency
γ	Gamma Index, also called complexity
ρ	Structural Connectivity
λ	Line Overlap
τ	Directness
ψ	Weighted Average Shortest Path Distance
A_j	Alightings at stop j
APL	Average Path Length
A_{tot}	Total Alightings at all stops in the time period
B_j	Boardings at stop i
B_{tot}	Total boardings at all stops in the time period
DI_j	Importance of stop j as a destination
$\underline{E}$	Total Number of edges
$\underline{E^m}$	Total multiple edges in Network
E_{max}	Maximum Possible Edges
$\underline{E^{st}}$	Total single Edges in network
ETT	Effective Travel Time
$ETT^D{}_i$	Effective average travel time departing from stop i to all other stops
$ETT^A{}_j$	Effective average travel time arriving at stop j from all other stops
GCC	Global clustering coefficient
iss	Average Inter-Stations Spacing
$\underline{K}$	Degree
L	Total Number of Lines
LCC	Local Clustering Coefficient
L_{no}	Total Non-Overlapping Length of Lines
L_t	Total Length of Lines
N	number of active stops in the time period
NCC	Node Clustering Coefficient

NDD	Node Degree Distribution
N_{iss}	Total Number of Inter-station spacings
OD_{ALL}	All Possible O-D paths
ODSTP	Origin-Destination Stop-Transfer Potential
PTT	Potential Travel Time
PTD_i	Potential Travel Times Departing from Stop *i* to all reachable stops
PTA_j	Potential Travel Times Arriving from all stops to stop *j*
Q	Modularity
S	Total Number of Stops/Stations
SA_{ij}	Schedule availability from stop *i* to stop *j*
SI_i	Importance of stop *i* as a source
SPL	Shortest Path Length (Shortest Path distance)
S_t	Total Number of Transfer Stops/Stations
S^t_c	Possible Transfers at all transfer stations/stops
Stop *i*	(Source/Origin)
Stop *j*	(Destination)
TR_{ij}	Minutes when stop j is reachable from stop i in the time period
TR_{tot}	Total minutes in the time period
TT_{ij}	Travel Time from stop *i* to stop *j*
V	Total Number of Vertices/Nodes
WP	Weighted Potential
WPD_i	Weighted Potential Departures from Stop *i*
WPA_j	Weighted Potential Arrivals at Stop *j*

1 INTRODUCTION

1.1 Background

Throughout the world, city planners are working to address the challenges of growing urban populations and economies in ways that are environmentally, socially and economically responsible. Accomplishing this monumental task requires urban forms that efficiently utilize infrastructure while supporting diverse industries and an equally diverse population. These needs are in direct opposition to the reality of automobile-oriented cities where buildings must be separated by wide roads and vast parking lots. Automobile-oriented cities require infrastructure to serve larger areas, penalize those who cannot drive, and the increasing number of automobiles, inevitably, causes congestion and pollution. The only way this can be overcome is through increased use of public transportation. Such a system moves large volumes of passengers, requires no parking lots and can be made universally accessible (i.e., users need not own automobiles).

Unfortunately, luring drivers out of their cars and onto buses has proven very difficult, partially due to poorly funded transit systems which fail to meet the needs of potential customers. This is made more complicated by both citizens and civic leaders who are hesitant to increase transit spending and question if current budgets are being used efficiently. To address this concern, an increasing number of cities are conducting system-wide overhauls of their transit networks. However, the success of these projects has been mixed, due to the limited tools available to transit planners.

Transit planners are tasked with providing the best service with the available funding, which requires understanding what makes one possible network “better” or “worse” than another. (Kittelson & Associates, Inc. 2013) To this end, researchers have developed numerous ways of modeling public transportation systems. These include supply-side and demand-side econometric models, discrete choice models, and data mining methods. These models provide insight into the likely impact from changes to aspects of service, such as frequency, vehicle crowding and amenities at stops. These models can then be used to simulate different networks or combined with the cost of providing service and then optimized. Such models are useful to evaluate policies, set fares and justify investments, as well as estimate the impact of large capital projects such as new rail lines. (dell'Ohlio et al. 2017)

However, these models are less useful to those planning and operating large bus networks, which might connect thousands of locations. It is not possible to provide high quality service from every location to every other location, and so planners need tools to identify the most important connections. In addition, transit is a public service that must meet the needs of a diverse population scattered throughout a very heterogeneous service area. Further complicating this situation is that the needs of users also change over the course of the day, day of the week as well as over larger time scales. This level of complexity is well beyond the ability of simulations or econometric modeling and so transit planners must rely on ridership, customer feedback and personal experience to estimate the likely impact of possible changes to the bus network.

Since the 1980's, a number of researchers have drawn from the fields of network science and graph theory to develop tools to help planners evaluate service changes. This led to several graph theory-based tools being developed for rail and metro networks in the 1990s. However, bus networks operate very differently from rail lines, as they have many more lines of service, and stop at many more locations, and transfers often require walking a short distance to a nearby stop. This challenge has been noted by many researchers who have devised a variety of ways to adapt these metrics to use with bus transit networks. Unfortunately, while these efforts have produced a wide range of graph theory-based tools, they are not used by those planning bus networks today. Instead of meeting the needs of transit planners, this variety of methods has created an additional barrier to transit planners who are now faced with a plethora of options to choose between.

Recent advances in computer technology and network science have provided new tools to study networks of computers and of social connections, which many theorists are applying to transit networks. Unfortunately, these efforts have not yielded useful tools for transit planners and, in some cases, have further complicated the situation.

1.2 Objectives and Expected Contributions

While graph theory has tremendous potential to help planning and operating bus transit networks, the current research shows a significant gap between the tools available and the needs of transit planners. This thesis bridges this gap and develops graph theory based methods that provide the information transit planners need. Accomplishing this task will result in four major contributions.

The first contribution is to investigate how choices in graph representation impact the results obtained. Studies in the current literature represent transit networks using a variety of graph configurations, with no agreement upon the applicability of each. Using methods commonly seen in the literature thirty-two (32) graph configurations are constructed from a common transit network and then used to calculate a battery of indicator values. As the underlying transit service is constant, comparing the results provides insight into how the choice of graph configuration influences the results. This is repeated for five (5) time periods to identify trends with regard to changes in the underlying transit service.

The second contribution is to develop graph theory-based tools that meet the needs of transit planners. This is accomplished by identifying the needs of transit planners and then developing tools to meet those needs. These methods clearly define the appropriate graph configuration and the resulting metrics combine shortest path calculations with ridership data to reflect the actual experiences of passengers as they travel through the network.

The third contribution is to overcome a number of technical challenges which have prevented previous researchers from developing indicators that meet the needs of transit planners. The software used, and methods required accomplish this are described so that they may be used by other researchers in future studies.

The final contribution is to use the newly developed indicators and those from the current literature to analyze Edmonton's public transit network. This is provided as quantitative analysis, map plots and qualitative descriptions of the service in each time period. In addition to demonstrating the utility of the newly developed tools, the results are directly useful to the transit planners who are currently redesigning Edmonton's transit network which will be implemented in 2020.

1.3 Thesis Structure

The remainder of this thesis is organized into the following sections. Chapter 2 provides a literature review of how graph theory is currently used to analyze public transportation networks. The first three sections focus on key concepts of both transit quality of service and graph theory. The following four sections focus on methods commonly used to represent transit networks as graphs and metrics used to analyze them and the final section discusses limitations of the current literature.

Chapter 3 discusses the methodology used to construct and analyze graph representations of Edmonton's transit network in this thesis. The first three sections describe the data and software used and discusses the primary technical challenges. This is followed by a section discussing the metrics chosen and another describing graph construction. The final four sections describe the specific methodology used to calculate the metrics in each of the four phases of the investigation.

Chapter 4 provides the results of the analysis, which are presented in four phases. In the first phase, a common battery of tests is applied to thirty-two (32) graph configurations to investigate how the choice of graph configuration is reflected in the resulting indicator values. The second phase uses a similar methodology to investigate how changes in the transit service between time periods are reflected in common indicator values. The third phase diverges from the current literature to develop new indicators specifically based on the needs of transit planners and the fourth phase combines the previous information with maps of the data to characterize transit service in each time period.

Finally, Chapter 5 concludes this thesis by providing a summary of the work done, followed by a discussion of both the strengths and limitations of the newly developed tools followed by an outlook for potential future research.

2 LITERATURE REVIEW

2.1 Transit Quality of Service

Public transportation service has a variety of goals, which can be at odds with one another, and compete for the same limited funding. The competing goals of “serving the largest number of individuals” and “serving those who need it the most”, is a common example. This is because those most dependent on public transportation often make up only a small portion of the ridership and have different needs from the majority, who primarily use public transportation to commute to work or school. In addition, short-term goals, such as attracting high ridership and supporting the current economy, can be at odds with future-oriented goals, such as encouraging specific types of development, changing travel behaviours or mode split. The relative priorities of such goals are further complicated by the unique geography and economy of each service area, along with available funding, culture, regulations, and policies of the operating agency.

As the goals, along with shape and size of public transit systems, are widely varied it should come as little surprise that few industry standards are available to measure “performance”. Instead many public transportation authorities have developed ad-hoc methods, based on intuition along with the goals and conditions specific to the service area. Such methods are both limited in scope and difficult to transfer between agencies.

While such site-specific metrics might be used for local planning activities, transit agencies usually receive funding from public sources which require standardized methods be used for reporting. To fill this need, industry associations such as Canadian Urban Transit Association (CUTA) (Canadian Urban Transit Association 2015) Federal Transit Administration (Federal Transit Administration, U.S. Department of Transportation 2017) American Public Transport Association (APTA) (American Public Transportation Association 2017) have developed a number of standardized operational and financial indicators. These indicators are intended to be simple for all transit agencies to collect, regardless of an agency’s size, types of service provided, or technology used. This limits those metrics to only system-wide totals, such as total ridership, operating costs, fleet size and amount of service provided which are useful only to identify broad trends and meet accountability requirements, and of limited usefulness to planners.

The Transportation Research Board’s (TRB’s) Transit Cooperative Research Program (TCRP) has taken a different approach (Kittelson & Associates, Inc. 2013) in the “Transit Capacity and Quality of Service Manual” (TCSQM). This provides a methodology to quantify the service provided by a public transportation network to understand both operation and the experience of passengers. The Level of Service (LOS) for public transportation is described in terms of six items divided into two categories. The category of "Availability" consists of frequency, service span, and access while the category of "Comfort and Convenience" is made up of passenger loads, reliability and travel time. Several metrics are provided for each item, to provide insight into numerous aspects of service. Unfortunately, these metrics require a large amount of data which is expensive to collect, assemble and process. In addition, establishing targets requires first establishing a baseline calculated from several years of data, which must be collected in a consistent manner. As such, any changes in the method or technology will require establishing new baselines and targets.

Measuring transit performance has also been studied in the academic literature for many years and has played as much of a role in the industry standards previously mentioned as practical experience of planners and engineers. A variety of methods are commonly used, including passenger surveys, analysis of ridership and those based on graph theory. The latter type, which uses graph theory, are currently a vibrant area of study due to advances in networks science and increasing power of computers.

Graph theory, like most technical fields, uses unique terminology and the current methodology has evolved over many years with simple methods forming the basis of more sophisticated ones. As such, an overview of the history and terminology is helpful in understanding the methods used in the current literature. Section 2.2 provides a brief summary of the history of graph theory as it has been used in public transportation followed by a summary of common terminology used in graph theory in section 2.3

2.2 History of Public Transportation as a Graph

Graph theory was first created by Leonhard Euler in 1741 to solve a transportation problem known as the seven bridges of Konigsberg. Euler developed a simplified map where land masses were represented as points and the bridges shown as lines connecting those points, as shown in Figure 2-1 below. This allowed him to prove that it was not possible to cross all seven of the bridges without crossing at least one of them twice. This simple methodology has been generalized to represent roads between locations, cables connecting computers, stages of a chemical process or relationships between ideas (Bondy and Murty 2008). Regardless of what they represent, all graphs use the same framework, so methods originally developed to study networks of computers, relationships on social networks or biological systems can be easily adapted to use on graphs representing public transportation systems.

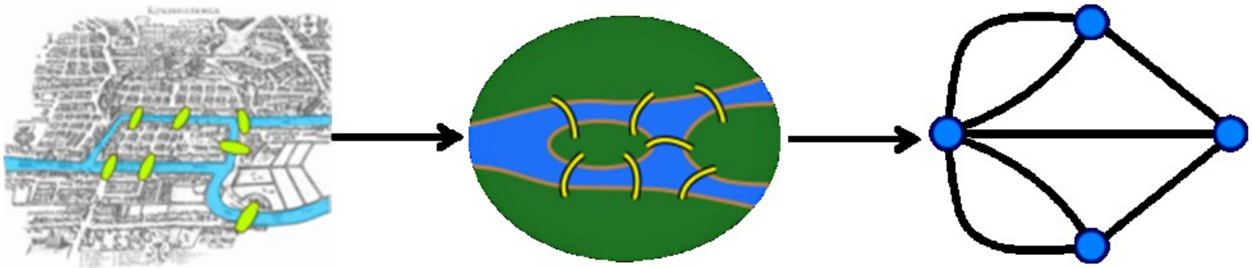


Figure 2.1 Seven Bridges of Konigsberg

(Wikimedia Foundation Inc. 2017)

Despite having been used for centuries in the transportation field, graph theory was not applied to public transportation until the 1980s (Derrible and Kennedy 2011). Public transportation networks are very different from other types of transportation systems as the connections are inherently time-dependent, due to each vehicle only being present at a given stop for a short time after which travellers must wait for the next vehicle to arrive. In addition, transit service is made up of separate lines that only stop at specific locations along their routes and which passengers must often transfer between.

The first study known to use graph theory in public transportation networks was done by Lam and Schuler in 1981 (Lam and Schuler 1981) where they introduced a trip time indicator (RT) which is a ratio of the "ideal" travel time to the actual travel time in the network is evaluated. RT requires calculating the harmonic means for both the ideal and actual travel time between all O-D pairs in the network. Unfortunately, this indicator is onerous to calculate for very large transit systems and treats all O-D pairs as being equally important.

Many important advances were brought to widespread use in the "trilogy" of textbooks written by Vukan Vuchic. "Urban Public Transportation Systems and Technology" (Vuchic 1981/2007) in 1981 (updated in 2007), Urban Transit: Operations, Planning, and Economics in 2005 (Vuchic 2005), and "Transportation of Livable Cities" (Vuchic 1999) in 1999. These books drew from work he had done with Musso 1990s (Musso and Vuchic 1988) (Vuchic and Musso 1991) and countless predecessors (Derrible and Kennedy 2011). Vuchic expanded graph theory methods to accommodate a variety of transit-specific characteristics such as frequency and passenger transfers between separate lines as well as distinguishing between different types of stations and different types of service. He also introduced terminology to distinguish between named or numbered service "lines" and "routes", the sequence of stops they follow, described lines by function and geometry and categorized overall network form.

Research using graph theory on transit networks continued to progress during the 1990s and 2000's culminating in 2009 when Derrible and Kennedy introduced seven measures intended to provide planners with a better understanding of metro networks (Derrible and Kennedy 2009) (Derrible and

Kennedy 2010) (Derrible and Kennedy 2010). The authors grouped these into three categories, i.e., “State” indicators described the complexity and connectivity of the network, “Form” indicators described line lengths and spacing of stations, and “Structure” indicators described passenger transfers as they moved through the network. The authors were attempting to identify the phases of development that metro networks undergo. To do this, they represented metro systems as unweighted and undirected graphs with nodes representing terminal and transfer stations and with parallel edges representing service lines and transfers.

The current literature on transportation networks as graphs is heavily influenced by developments in network science. The field of network science has existed since the late 1950’s (Erdos and Renyi 1959), but came into popular use in the 1990’s do to studies on “Small-World” networks by (Watts, and Strogatz 1998), and “Scale-Free” networks by (Barabasi and Albert 1999). Both of these, which are discussed in more detail in section 2.7, was found to occur in computers, social interactions, biology, and transportation. This has led to a dramatic increase in the study of public transportation networks, and adoption of abstract graph types, such as L-Space and P-Space which are discussed in section 2.4. Unfortunately, the scientists conducting these studies are focused on identifying small-world and scale-free properties, with little interest in contributing to the field of public transportation. As such, this increase in study has yielded limited value to those planning or operating transit systems (Derrible and Kennedy 2011).

Recent advances in technology have contributed to public transportation in other ways. One important advance came in 2006 when Google introduced the “General Transit Feed Specification” (GTFS), also called the “Google Transit Feed Specification”. GTFS, which was developed to allow Google Maps to display transit information, provided an open, text-based standard that was supported by a large internet company and, so it quickly became a de-facto industry standard (Roush 2012). This allows agencies to provide schedule data to researchers in a consistent format that is easily understood and can be processed without proprietary software (Google 2017). Another advance came from the increasing power of computers and of open source analytical tools like KNIME and the Python programing language. These allow researchers to create and study Time-Expanded networks in ways previously not possible and has led to increasing use of these networks which can now be partially or fully enumerated.

2.3 Graphs and Their Components

Graph theory is the study of “Graphs”, which are made up of “Nodes” (also called “Vertices”), that are connected by “Edges” (also called “Links”). Figure 2.2 shows the most common way to visualize graphs with edges beings represented as line segments and nodes being circles at each end of the edges. This configuration requires that every edge connect exactly two nodes and that edges only intersect at nodes. When used to represent transportation networks, the nodes typically represent locations and edges represent ways to travel between them, typically roads (Boeing 2017) (Porta et al. 2006), rail lines (Akse 2014) (Derrible and Kennedy 2011), or bus routes (Chen et al. 2007) (Zhang et al. 2013)).

Edges have three important characteristics that dramatically impact the structure of the graph. The first is directionality where links are either “directed” or “undirected”. Undirected links are equivalent to a two-way road where travel is symmetric in both directions. Such links are appropriate along rail lines as travel in either direction is possible with a similar time and distance. A directed edge, shown in Figure 2.3 represents where travel is only possible in one direction between two nodes, which is often described as travelling from the “source” node to the “target” node. Directed edges are required for bus routes as buses arrive at stops following an ordered sequence, and many stops are along roads and only accessible to a single direction of traffic (Feng et al. 2016).

The second important characteristic is “weight”, which represents the time, distance or cost required to travel along the edge. For transportation networks, the total distance, time or cost is the sum of the weights of all edges traversed (Biggs et al. 1986) (Bondy and Murty 2008). In some studies where identifying the number of possible paths or the number of transfers required, edge weights are irrelevant and so “unweighted” edges are used instead (Zhang et al. 2015) (Tsekeris and Souliotou 2014) (Hadas et al. 2014) (Zhang 2016). Unweighted edges are equivalent to simply setting all edge weights to one, which dramatically simplifies many calculations (Diestel 2010).

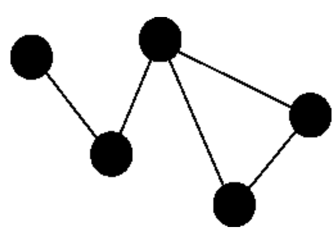


Figure 2.2 Graph with Undirected Edges

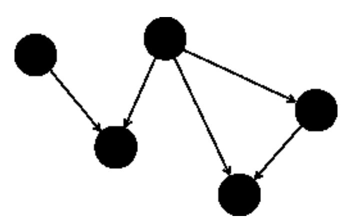


Figure 2.3 Graph with Directed Edges

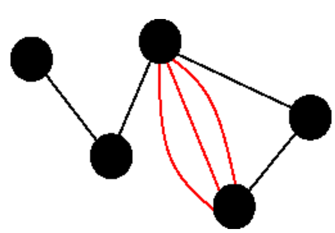


Figure 2.4 Graph with Parallel Edges

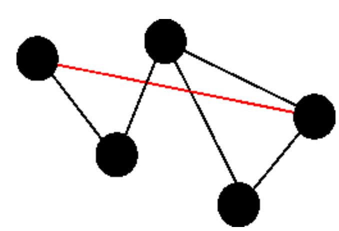


Figure 2.5 Multidimensional Graph

The third characteristic reflects if the edges are “simple”, where only one edge connects each pair of nodes, or “complex”, where two nodes can be connected by several “parallel” edges, as shown in Figure 2.4. In transportation networks, complex edges are used to represent multiple roads or bus lines connecting the same two locations and can also be used to represent individual trips. However, if this level of detail is not required, a single edge can be used to dramatically simplify the calculation process.

Another important aspect of the graph representation is dimension. The simplest form is a "planar" graph where edges that cross one another must do so at a node. For most local streets, this type of network is an accurate representation because most local roads cross at an intersection (Derrible and Kennedy 2011). However, this is not always suitable for bus networks as bus lines need not cross at a stop, and multiple lines may use a common corridor and only share a subset of stops. To accommodate this type of connectivity, graphs must have multiple layers or "dimensions" to correctly represent the transit network, as shown in Figure 2.4. Such "multidimensional" (or multilayer) graphs are very powerful tools, however, they tend to be more complicated as the number of possible edges and nodes increases with each dimension (Gross and Yellen 2004).

By contrast, the only inherent property that nodes have is a unique location, which must correspond to edge dimension. Each node has two other important properties which are derived from edges, i.e., degree and adjacency. The degree of a node is defined as the number of edges that are connected to it while adjacency is a list of nodes that are joined to it by at least one edge. Nodes that are adjacent are often called "neighbours" and all neighbours of a node makes up its "neighbourhood". The situation becomes more complicated when directed edges are used where degree and neighbours are both described as "in" or "out" based on the direction of the link

These simple properties allow nodes and edges to be combined to form graphs, which have properties as well. These graphs have properties, such as the number of nodes and average degree, which are simply the aggregate of the components, as well as emergent properties, which are not apparent in the individual components but only emerge from the entire system (Gignoux et al. 2017). Since 1981, researchers have worked to identify these properties and understand how they relate to the underlying transit network. However, this work has proceeded piecemeal by researchers with diverse backgrounds, developing equally diverse methods to create and study transit networks as graphs. Regardless of the methods used, the first step is to convert the transit network, or networks, being studied into a graph, or graphs. This single step can be accomplished in a variety of ways, which can be described in terms of three choices which researchers make when creating the graphs to be studied.

2.4 Decisions on Graph Configurations

Graph theory does not dictate how a transit network must be represented as a graph. This flexibility allows researchers to include only the properties of the transit service they wish to study but also limits the applicability of the findings to the specific graph configuration used. While a vast number of graph configurations are possible, only a small number are commonly seen in the literature, and they can be defined in terms of three choices the researchers make. The first choice is the Graph Type, which describes what the edges and nodes represent. The second choice is the edge type which defines movement through the network, and the third choice is whether or not to include pedestrian links between nearby stops.

2.4.1 Graph Type

Graph Type is the result of decisions regarding how transit links and transfers are represented. Before considering the actual transit system, whether transfers are to be considered must first be determined. If transfers are not considered, the graph can use "simple stops" where each stop is represented as a single node allowing properties of that node to be directly translated to the properties of the actual stop. However, as all transit links at that stop connect to the same node, there is no distinction between passengers remaining on the bus and those transferring at the location.

If transfers are to be considered, "complex" stops are required, where each stop is represented as an array of nodes, denoted by both the stop location and the trip or line of service. In addition, a "base" node may also be used to allow the stop location to be specified as the origin or destination of a trip without also having to specify a specific transit line or trip. The nodes at each stop are then connected by links representing possible transfers. Complex stops can accurately model travel through the network and so are particularly useful for estimating travel times between locations and studying transfers. However, this level of detail dramatically increases the complexity of the resulting graph.

Next, how transit links are represented must be determined. The most detailed representation is to model each trip separately, thus each line is represented as an array of individual directed links. More commonly, individual trips are ignored, and each link represents a specific line of service. Alternately, each link might only indicate if the connection is possible, regardless of the number of lines or trips connecting the two stops. Finally, abstract representations are possible, with the most common being to represent each line as a clique, or completely connected subgraph.

Taken together the decisions regarding the use of complex or simple stops and which of the four types of links are chosen, allow for eight potential configurations, although only six are commonly used. Where complex stops are used; transit links can represent individual trips to create a "Time-Expanded" (TE) graph, shown in Figure 2.8, or edges can represent individual lines to create a "Route Map-Complex" (RM-Complex) graph, shown in Figure 2.7. Where simple stops are used; edges can represent trips to create a "Trip Map" (TM) graph, or they can represent lines to create "Route Map-Simple" (RM-Simple) graph, both shown in Figure 2.6. Where each link represents only that a connection is possible, transfers are not modelled, and so only simple stops are appropriate, resulting in an "L-Space" graph, shown in Figure 2.6. The final type represents each line as a clique, is call “P-Space”, which are only created with simple stops in the literature.

Link Travel Times				
Graph	Start	End	Type	Travel Time
L-Space	X	Y	N/A	2
L-Space	Y	X	N/A	5
L-Space	Y	Z	N/A	1
L-Space	Z	Y	N/A	1
L-Space	Z	X	N/A	4
RM-Simple	X	Y	GTFS	2
RM-Simple	X	Y	Pedestrian	5
RM-Simple	Y	X	Pedestrian	5
RM-Simple	Y	Z	GTFS	2
RM-Simple	Y	Z	Pedestrian	1
RM-Simple	Z	X	GTFS	4
Trip Map	X	Y	GTFS (x2)	2
Trip Map	X	Y	Pedestrian	5
Trip Map	Y	X	Pedestrian	5
Trip Map	Y	Z	GTFS (x2)	2
Trip Map	Y	Z	Pedestrian	1
Trip Map	Z	X	GTFS	4

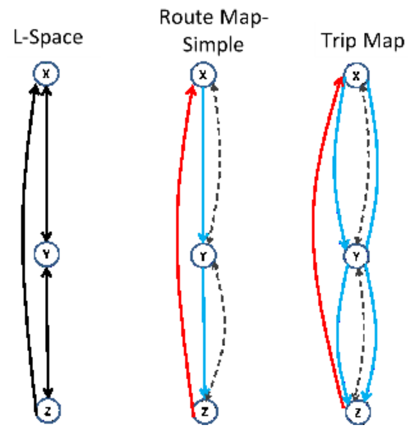


Figure 2.6 Graph Types with Simple Stops

Link Travel Times			
Start	End	Type	Travel Time
X	Y	GTFS	2
X	Y	Transfer/Ped	5
X	Y	Transfer/Ped	5
Y	X	Transfer/Ped	5
Y	Z	GTFS	0
Y	Z	Transfer/Ped	1
Y	Z	Transfer/Ped	1
Z	Y	Transfer/Ped	1
Z	Y	Transfer/Ped	1
Z	X	GTFS	4
Z	Z	Transfer/Ped	3

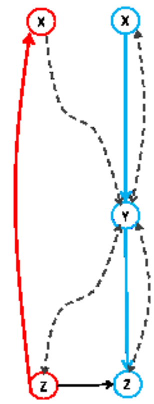


Figure 2.7 RM-Complex Graph

Link Travel Times			
Start	End	Type	Travel Time
X^a_3	X^d_3	GTFS (layover)	0
X_0	X^d_1	Boarding	1
X^a_3 / X^a_5	X_4 / X_6	Alighting	1
X_0 / X_1	Y_5 / Y_6	Pedestrian	5
Y_0 / Y_1	X_5 / X_6	Pedestrian	5
X^d_1 / X^d_3	Y^a_3 / Y^a_5	GTFS (Travel)	2
Y^a_3 / Y^a_5	Y_4 / Y_6	Alighting	1
Y^a_3 / Y^a_5	Y^d_3 / Y^d_5	GTFS (layover)	0
Y_2 / Y_5	Y^d_3 / Y^d_5	Boarding	1
Y_0-Y_5	Z_1-Z_6	Pedestrian	1
Y^d_3 / Y^d_5	Z^a_3 / Z^a_5	GTFS (Travel)	0
Z^a_3 / Z^a_5	Z_4 / Z_6	Alighting	1
Z_0	Z^d_1	Boarding	1
Z^d_1	X^a_5	GTFS (Travel)	4

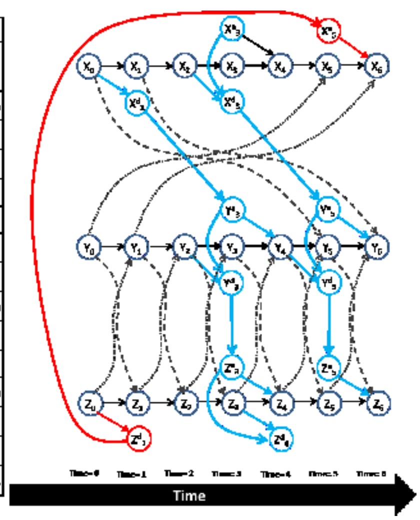


Figure 2.8 Time-Expanded Graph

The five graphs shown in Figures 2.6-2.8 show the same transit service: Two trips on the blue route and one trip on the red route

2.4.1.1 Time-Expanded: Complex Stops with Links for each Trip

Time-Expanded (TE) graphs, which represent transfers between individual trips at each stop, allow for travel through the network to be modelled for specific departure and arrival times with detailed information for the time and location of transfers. However, this level of detail results in a phenomenally complicated graph with potentially millions or even tens of millions of edges and links. In Time-Expanded Graphs, each stop is represented as a collection of Stop-Trip-Time nodes, each of which represents a unique trip at that stop. Two other types of nodes may also be included. If included, the Base node is connected by directional links from each Stop-Trip-Time node, with no links in the reverse direction. Transfers can then be modelled by directed links from each Stop-Trip-Time node to all nodes that occur at a later time at the same stop. Also, a collection of “Stop-Time” nodes may be used to model waiting at the stop with each Stop-Trip-Time node connected to a specific Stop-Time node.

Time-Expanded graphs are primarily used to calculate very realistic travel times between locations. Time-Expanded graphs are less commonly used in the literature as such graphs are seldom used for studies other than calculating shortest travel times between locations. In (Geisberger 2011), (Akse 2014), and (Fortin et al. 2016) the authors use Time-Expanded graphs to model passenger waiting times and explore methods of schedule optimization. Another use of Time-Expanded graphs was in (Fortin et al. 2016), where the authors calculated the number “Opportunities” or locations reachable from stops in the downtown.

2.4.1.2 Route Map-Complex: Complex Stops with Links for each Line

Route Map-Complex (RM-Complex) graphs, which show transfers between lines but not individual trips, are commonly used in the literature to study a range of properties. These have been used by several authors to study travel times and transfers, through both bus and multimodal networks (Hadas et al. 2014), (Zhang 2016) (Alessandretti et al. 2016). They are commonly used to study the structure of the underlying service, including the “state”, “form” and “structure” (Quintero-Cano et al. 2014) (Derrible and Kennedy 2010) (Zhang et al. 2015) and abstract measures such as community structure (Zhang et al. 2015) (De Bona et al. 2016), small-world, and scale-free properties (De Bona et al. 2016), all of which will be discussed in section 2.7

2.4.1.3 Trip Map: Simple Stops with Links for each Trip

Trip Map (TM) graphs show each line as a collection of parallel edges between stops, but do not show transfers. The parallel edges represent the frequency of service between stops, and while these may be useful to investigate connectivity and community structure, the large number of edges adds complexity and they rarely appear in the literature.

2.4.1.4 Route Map-Simple: Simple Stops with Links for each Line

Route Map-Simple (RM-Simple) graphs, which show lines but not transfers between them, are intuitive, versatile and very simple to work with. Such graphs have far fewer edges than Trip Maps but still have many parallel edges, reflecting overlapping routes. Route Map-Simple graphs are very versatile, being able to capture network connectivity and structure approximate travel through the network. Directed Route Map-Simple graphs were used in (Dimitrov and Ceder 2016) to study the small-world and scale-free characteristics of the multimodal public transportation network in Auckland. Unlike bus lines, which can follow different routes and run at different frequencies for each direction, rail lines are nearly always symmetric and so can be modelled using undirected edges, as was done in (King and Shalaby 2016), when analyzing the robustness of the Toronto's Rapid Rail system. However, the lack of complex stops makes them poorly suited for studies related to transfers.

2.4.1.5 L-Space: Connections between stops only

L-Space graphs are simplified, abstract representations that do not reflect routes, trips or transfers, which limits their validity in studying the underlying transit service. Instead, they are commonly used, along with P-Space graphs to study structural properties of the graph, such centrality (Derrible 2012), community structure (Sun et al. 2016) (Tsekeris and Souliotou 2014) along with Small-world and Scale-free properties (Tsekeris and Souliotou 2014) which are not directly linked to the transit service.

2.4.1.6 P-Space: Lines as Cliques

The P-Space graph type is an abstract form that was developed to study interconnections between lines and transferring in the network. Each line is shown as a fully connected subgraph, called a "clique", such that each stop is only a single link away from all others which share a line with it. Transfer possibilities are shown as nodes which are members of multiple cliques and so the number of edges traversed when travelling between points in the network is equal to the number of lines that were used (Yang et al. 2014). However, this structure dramatically increases the number of edges and interpreting calculated metrics in the underlying transit service is very challenging.

2.4.2 Edge Type

With the graph type is chosen, the interpretation of both edges and node is defined. However, before the network can be converted into a graph, the edge type must also be decided. The edge type is defined by the choice of directionality and weighting, which together, define how travel along each edge occurs. As mentioned earlier, "directionality" describes if the edges are directed, allowing travel in a single direction only, or undirected, allowing travel in both directions, and "weighting" describes how the "shortest" path through the network is calculated. While many any numbers of edge types are possible, only five types are commonly seen in the literature: Unweighted-Undirected (UU), Unweighted-Directed (UD), Directed-Travel Time (TT) and Directed-Value of Time (VoT).

Undirected edges are nearly always unweighted as well, and the resulting (UU), graphs provide a simplified representation of the underlying service, which is particularly useful to explore abstract aspects of the graph, such as centrality (Derrible 2012), community structure (De Bona et al. 2016) (Sun et al. 2016) (Zhang et al. 2015) (Tsekeris and Souliotou 2014) along with small-world and scale-free characteristics (Dimitrov and Ceder 2016) (Tsekeris and Souliotou 2014) (De Bona et al. 2016). While less commonly seen, UU graphs can be used to study the operation of the underlying service (Derrible and Kennedy 2010) (Zhang et al. 2015) and prevalence of transfers (Hadas et al. 2014).

UU links provide an adequate representation of rail lines as the routes and frequency are the same in both directions and stop spacing is often reasonably consistent (Vuchic 2005). However, the same is not true for bus lines which frequently run asymmetric frequencies and routes. Even lines which travel along the same road in each direction will use different stops in each direction due to the stops only accessible from vehicles travelling in a single direction.

Connectivity of asymmetric lines can be modelled using Unweighted-Directed (UD) edges, however, weighted links are required to represent inconsistent stop spacing and asymmetric aspects of service quality, such as frequencies, speed, distance, or travel time. Weighted graphs are better able to represent the actual service being provided, and so they are frequently used to study travel times (Akse 2014) (Zhang 2016) (Alessandretti et al. 2016) but are also used to study the state, form (Quintero-Cano et al. 2014) (Zhang et al. 2015)and structure of the transit network. Edge weighting is irrelevant for studying small-world or scale-free properties, but weighted graphs can be used to study centrality (Akse 2014) (King and Shalaby 2016) and community structure (Zhang et al. 2015).

Numerous edge weighting schemes are possible including distance (King and Shalaby 2016) and frequency (Quintero-Cano et al. 2014) along with travel time (Geisberger 2011) (Akse 2014) (Zhang et al. 2015) (Fortin et al. 2016) (Fielbaum et al. 2016) (Zhang 2016) (Alessandretti et al. 2016) and composite weighting schemes intended to reflect different "Values of Time" (VoT) placed on different portions of the trip. Travel Time is the most commonly used as it is easily calculated from posted schedules and can reflect service frequency if waiting times are included. However, research has shown that passengers consider waiting at a stop and transferring to be a greater hardship than time spent travelling on a vehicle (Kittelson & Associates, Inc. 2003). Including this requires different values of time be used for each portion of the trip (waiting, boarding, riding, alighting, transferring, walking, etc.). Several methods of accomplishing this have been proposed, (Yang et al. 2014) (Kittelson & Associates, Inc. 2013) (Bunker 2015) but these require calibration for each location due to differences in amenities and passenger preferences.

With the decision of an edge type, most aspects of how the graph representation will model the underlying transit system are clearly defined. In addition to having an unambiguous interpretation of what is represented by nodes and edges, the rules governing connectivity and movement through the graph are firmly established. However, one critical item must be decided before the graph can be precisely defined, treatment of pedestrian connections.

2.4.3 Pedestrian Links

Pedestrian connections are seldom mentioned in the literature regarding transit system as graphs, which is understandable as they are not included in the service information provided by the transit agencies (Google 2017), and so might be considered separate from the transit service. Furthermore, pedestrian connections have only a minimal impact when studying most rail systems, as stops/stations are spaced far from one another, so walking between stations to make a connection is very rare (Akse 2014). However, bus systems are very different as stops are often spaced closely and many transfers require passengers to walk a short distance between stops, for example crossing a street or walking around a corner (Chatterjee et al. 2016). This can have a dramatic impact on estimating travel through the network, as shown by (Hadas et al. 2014) where longer walking distances were shown to dramatically improve connectivity and (Yang et al. 2014) where pedestrian links reduced average distance traveled by 7-16%, and average number of transfers required by between 29-34%.

Several studies note that including pedestrian introduces challenges (Quintero-Cano et al. 2014) (Mishra et al. 2012) (Hadas et al. 2014). One challenge is that unweighted graphs do not distinguish between walking links and transit links. While this can be overcome by using links weighted by TT or VoT, identifying and weighting pedestrian links may be problematic due incomplete data on pedestrian infrastructure. Another challenge is that pedestrian links can outnumber transit links, increases the complexity of the graph and, potentially mask the structure of the transit network. This is shown in (Yang et al. 2014) where the authors found including pedestrian links changed the node degree distribution from following an exponential to a Gaussian distribution, and resulted in a six to ten-fold increase in node clustering. These changes indicate including pedestrian links results in a very different graph structure (Chatterjee et al. 2016), which is explored in sections 3.6 and 4.1.

2.5 Analyzing Transit Networks with Graph Theory

With six graph types, four edge types and the choice of either including or excluding pedestrian links; this allows for 48 ways a public transportation network can be represented as a graph. While each of these is mathematically valid, the degree to which they vary in the aspects of the underlying transit service is substantial. Furthermore, there is, currently, no consensus regarding the relative value of each graph configuration to guide transit planners.

Regardless of the choice of graph representation, this only defines how graphs will be constructed. Researchers must also decide what aspects of the graph or underlying service they are interested in studying. This is no simple tasks, as a wide range of possible methods and metrics have been developed, by researchers in diverse fields, to study a spectrum of different properties. Categorizing these methods and identifying the most appropriate metrics or for planners is further frustrated by the piecemeal fashion in which they were developed. Two ways of categorizing these methods are by whether the methods focus on the actual service or on the graph structure and the level of sophistication. Those which focus on the actual transit service are presented in section 2.6 while those developed to study the graph structure described in section 2.7.

Within each of those sections, the methods are presented in order of increasing sophistication and complexity. In both cases, the simplest and least sophisticated methods are those that calculate single value “Global” properties as if the entire transit network as if it were homogeneous. Simplest among these are total values, which are either counts of elements (e.g. the number of stops or nodes) or sums of values (e.g. the total length of all trips or all edges) which are primarily used to describe the total size of the system. Following the totals are global properties which are calculated from those totals, such as ratios of properties, which provide insight into the general structure of the network.

Following global properties are methods that address the actual heterogeneity of both the transit service and the graph representation of it, which is accomplished in several ways. For studies focusing on the transit service, frequency of service is presented, followed by connection between stops and finally shortest paths through the transit system. A very different approach is taken by studies which focus on the graph structure, where researchers first developed methods of partitioning the graph into subgraphs. When taken to the logical extreme, this results in each node being treated separately and then considering the distribution of properties, which is followed by identification emergent properties which arise when specific combinations of characteristics are present.

2.6 Analysis of the Underlying Transit System

The first indicators to be discussed are those that use graph theory to study the underlying transit system such that the resulting values are directly linked to the transit system, and thus provide planners with directly useful tools. This can be accomplished in four ways. The simplest method, identifying the global properties of the service, provides simplified information regarding the size and complexity of the network. A more sophisticated method is to model the frequency, which varies between locations in the network. However, recent increases in computer power have seen an increase in, computationally expensive, the use of shortest path calculations which can be used to study the connectivity between locations or travel times.

2.6.1 Global System Properties

The simplest way to consider a transit system is to treat it as being homogeneous to gain insight into the size and complexity of the system. These measures, shown in Tables 2-1 and 2-2, are inherent to the underlying service, without regard to the graph representation, and are determined directly from the network itself. The most basic are system totals which include counts of the number of stops (S), lines (L) and a total of lengths of all lines (L_t). Other system total metrics are more difficult to obtain, including the number of stops where transfers can take place (S_t) and the number of such transfer possibilities (S^t_c), which must be counted for each stop and then totalled. Similar metrics for lines include the number of inter-stop spacings, which must be done for each line along with the number of overlapping segments (E^m) and the non-overlap length of all lines (L_{no}), which require identifying portions of overlapping lines. The most difficult global total is the maximum number of transfers required to cross the network (δ), which requires full enumeration, although (Dong and Chen 2005) demonstrated graph structure to simplify the task.

Table 2.1 Global Properties of Transit System

Variable	Symbol
Total Number of Stops/Stations	S
Total Number of Transfer Stops/Stations	S_t
Total Number of Lines	L
Total Length of Lines	L_t
Total Non-Overlapping Length of Lines	L_{no}
All Possible O-D paths	OD_{ALL}
Maximum number of transfers to cross the network	δ
Total Number of Inter-station spacings	N_{iss}
Overlapping Segments in Network	E^m
Possible Transfers at all transfer stations/stops	S^t_c
Shortest Path Length (Shortest Path distance)	SPL

Table 2.2 Formulas for Global Properties of the Transit Service

Indicator	Symbol	Formula	
Line Overlap	λ	$\lambda = \frac{(L_t - L_{no})}{L_t}$	2.1
Average Inter-Stations Spacing	iss	$Iss = \frac{L}{N_{iss}}$	2.2
Directness	τ	$\tau = \frac{L_t}{\delta}$	2.3
Structural Connectivity	ρ	$\rho = \frac{v^t_c - e^m}{v^t}$	2.4

One indicator developed by Vuchic and Musso was the degree of overlap (λ) which is simply the ratio of the total distance of overlapping segments to the total length of all lines in the entire network.

This indicator is useful to distinguish between systems with a branching structure and those with many overlapping routes (Vuchic 2005).

The number of stations, average line length and average station spacing are used to define what Derrible and Kennedy called the "Form" of a metro network. Using these measures, the authors found that metro networks tended to fall into one of three categories based on the degree to which they serve the local area and the larger region. Metros with fewer than 130 stops, with an average line length less than 20 kilometres and average stop spacing of less than 1 kilometre were focused on "local coverage". Those with more than 130 stops and average line lengths in excess of 20 kilometres were categorized as providing "regional coverage: Systems with fewer than 130 stops with average stop spacing in excess of 1.7 were described as providing "regional accessibility" regardless of the average spacing between stations. (Derrible and Kennedy 2010)

Two other indicators are Directness (τ) and Structural connectivity (ρ), which Derrible and Kennedy developed to describe the "Structure" of metro networks based on transfers. Directness (τ) is a measure of how likely transfers are to be required, using the total number of lines in the system and a maximum number of transfers required to traverse the network (δ). Structural connectivity (ρ) is a measure of the affluence of transfers based on the total number of transfer possibilities, parallel links and transfer stations in the network. The authors found that these two indicators were correlated with one another and provide useful information about the geography of the service area and travel needs of customers. They grouped these into three categorize separated based on the relationship between τ and ρ. Directness oriented networks have relatively few transfer options for the number of lines they offer but also require fewer transfers and are characterized by having a directedness above. Connectivity oriented systems, have many more opportunities to transfer but also require that passengers transfer more often and are characterized by having directness less than 3.15. They called the remaining group, "integrated" as they focused on neither directedness nor connectivity and represent a hybrid of these two extremes (Derrible and Kennedy 2010).

2.6.2 Frequency of Service

Some authors take a different approach, using graph theory tools to study the frequency of service, which is an average of the number of trips on each line over a given period of time. This can be used instead of actual trip departure and arrival times to estimate waiting times which varies widely between locations within the network. As such, a single global value would not be used and so these studies divide the network into smaller parts which can be investigated separately.

Frequencies are useful when estimating travel times and transfers, particularly when using the actual time schedule is impractical due to complexity or the level of granularity. (Hadas et al. 2014)

Calculated “Destination-direct stop level indicator” for each stop using the frequency of service from that stop as an origin to a set of destination stops along that route and “Destination-indirect stop level indicator” which allows for a single transfer. These indicators were calculated using individual trips but presented as frequencies to represent the impact on connectivity that resulted from allowing longer transfer times. Another use of frequency was in (Quintero-Cano et al. 2014) where the authors weighted all links in their networks by frequency to model connectivity of the public transit system within and between traffic analysis zones (TAZ) in the Greater Vancouver Metropolitan Area.

With the increasing power of computers, frequencies are less commonly used as the actual timetables provide more accurate results when used to model transfers and travel times. However, investigations on the connectivity between stops are more likely to use frequency as using the detailed schedule information is computationally expensive.

2.6.3 Connectivity between Stops

Some authors have chosen to take a very different approach by studying the connectivity between stops in the network using “Shortest Path Lengths” (SPL). Determining the shortest path between nodes is a very commonly used application of graph theory, being used to calculate directions in map programs, schedule parcel delivery and route data through the internet (Diestel 2010). Typically, calculating the shortest path between a single “Source” or “Origin” node to all other nodes in the network is done using Dijkstra’s algorithm (Bast et al. 2014), although several modified versions exist which perform better in specific types of graphs (Bherkassky et al. 1996).

In addition to working directly with SPL, it can be used to investigate connectivity between stop locations, such as the number of possible connections and the number of transfers required. These metrics provide useful insight to planners and overcome the challenges that arise from working with the full SPL data, which are discussed in section 3.3. Two types of metrics for connectivity between stops are seen in the literature, those related to transfers and those related to the number of stops reachable stops.

Transfers, where passenger change from one line to another over the course of their trip, add time to the trip and passengers risk missing a connection. While passengers prefer to avoid transferring, constructing a network with many opportunities to transfer allows for travel between more locations, resilience in the event of a failure and can reduce travel times. While SPL must still be calculated, the actual distance need not be retained, thus potentially simplifying calculation and processing the data.

In (Hadas et al. 2014) the authors constructed a directed, unweighted, Time-Expanded graph representing a portion of the regional transit system near Dolo in Italy with pedestrian links included. The authors constructed the graph to model transfers that take place at each stop, as well as those within a short walk of one another, where the maximum walking time ranges between 5 and 15 minutes. This was then used to calculate connections between nearby municipalities and the central hospital. The authors define "Origin-Destination Stop-Transfer Potential" (ODSTP) as the number of stops accessible from a given stop via trips that use it and routes that visit stops within the defined walking distance. They find that increasing the maximum walking time (and thus distance) causes an increase in both ODSTP and "Destination-Indirect Stop Level Indicator", although the specific relationship is not defined.

Using directed graphs with edges weighted by distance, the authors of (Zhang 2016) were able to model bus networks in four Chinese cities. They then calculate "Weighted Average Shortest Path Distance" (ψ) which uses the time spent transferring as the weighting and "Transfer Efficiency" (ξ) which relates transfer times to the number of routes available at the origin and destination stops, using equations 2.5 and 2.6.

$$\Psi = \sum_{O}^{V}\sum_{D}^{V} APL_{OD} * \frac{1 + transfer\ time_{OD}}{V(V-1)} \tag{2.5}$$

$$\xi = \frac{\left(\sum_{O}^{V}\sum_{D}^{V} \frac{1 + transfer\ time_{OD}}{(k_O^2 + k_D^2)}\right)}{V(V-1)} \tag{2.6}$$

Another method of studying connectivity between bus stops to determine the number of stops which can be reached from, or which can reach specific stops. These studies calculate shortest paths using Time-Expanded graphs with directed, time-weighted edges. Additional limitations are then imposed to determine the number of stop locations that can be reached from a given location within a reasonable amount of time and without requiring excessive transfers. These studies can be particularly useful when combined with GIS software to study geographic variation.

An example of this type of study is (Fortin et al. 2016), where the authors used the Conseil Intermunicipal de Transport de Chambly-Richelieu-Carignan public transportation network. Two connectivity indicators were introduced using the number of stops reachable with a maximum travel time of 120 minutes (the validity of a ticket). The first indicator, "Dynamic Connectivity" is a system-wide indicator that shows how connectivity varies with time using the number of reachable stops from each stop at each departure time within a one hour window. These are then summed for the entire network and normalized by the maximum possible dynamic connectivity. The authors then demonstrated that the dynamic connectivity increased dramatically in peak times but never exceeded 45%, although due to the complexity of the network, the authors only calculated dynamic connectivity for a subset of locations.

The second connectivity indicator, "Opportunities" instead shows how the connectivity varies by location. To reflect that some stops are important destinations while others are important origins, both "departure opportunities" and "arrival opportunities" were calculated. To obtain the Departure opportunities, the number of reachable stops was calculated for each stop for all departure times in the entire time period. Likewise, the Arrival Opportunities for each stop was calculated as the number of locations that could reach that stop location within two hours, for all arrival times within the time period. Once again, the authors were only able to calculate opportunities for a small subset of stop locations. The number of opportunities for each stop was then combined with the location data and plotted on a map to show the relative importance of each stop location. The maps for both arrival and departure opportunities were very similar, although this may be due to the chosen stops all being located in the downtown area.

Similar indicators were developed in (Hadas et al. 2014), although transfers were also considered, and a distinction was made between connections that required at transfer and direct ones that did not. The intent of the study was to investigate only if transfers were required and so unweighted, undirected L-Space graphs. As might be expected, the authors found that allowing transfers increased the number of possible connections by several times.

2.6.4 Shortest Path Length (SPL) and Average Path Length (APL)

It is also commonly seen in the literature for authors to calculate shortest paths and then work with that data directly. When this is done, the shortest path length (SPL) is the sum of the weights of all edges traversed, which can be output as either a single value describing the distance between the Origin and one specific Destination (O-D) or a table of values between the origin and all other nodes. The length will depend on the relative locations of the origin and destination nodes, available connections as well as the directionality and weighting of edges. As such, SPL can be used to represent a range of items, such as travel time, distance, number of transfers or number of links traversed. Alternately, edge weights and thus SPL can be a composite "Cost" value calculated from multiple items, such as was used in (Yang et al. 2014). Direct use of SPL data is typically considered either as a single "Average Path Length" (APL) or as a distribution of travel times.

Conceptually, the simplest method of analyzing the shortest paths through a network is a simple average of all possible O-D pairs. The impact of possible changes to the network can then be investigated by comparing the APL before and after the changes is made. An example of this method is the "Weighted average Shortest path distance" (ψ) (Zhang 2016), and the study by (Yang et al. 2014), on how walking links impact APL, However, the most common use of APL in the current literature is in identification of Small-world characteristics that will be discussed in section 2.7.

While the APL is used to evaluate large-scale changes, investigating the network in detailed requires consideration variation in SPL between O-D pairs within the network. A simple example of this is demonstrated in (Feng et al. 2016), where the authors demonstrate that travel speeds between well connected "Rich Club" nodes are significantly higher than connections with regular nodes. A similar method was used in (Huang and Levinson 2015), where the authors created several simulated cities with idealized land use patterns serviced by idealized network configurations. Estimated demand and travel times were then used to evaluate how well each network configuration met the needs of the city.

Another use was demonstrated in (Huang and Levinson 2015), where the authors used GIS software to calculate "Trip circuity" between 300 random O-D pairs for both the public transportation network and by personal vehicle. The travel time, travel distance and geodesic distances were calculated for each trip and trip circuity calculated as a ratio of the travelled distance to the geodesic distance. This method was used to investigate the relative circuity in 36 cities in the United States to identify that circuity of both auto and transit are correlated, due to using the same roads, and are strongly correlated to transit mode share.

Shortest path lengths can also be used to investigate the relationship between travel time and distance for various O-D pairs in the network as was demonstrated in by (Alessandretti et al. 2016). The authors modified Dijkstra's algorithm Directed, to calculate SPL based on travel time but also record the distance travelled through weighted Route Map graphs of five European public transportation networks. By comparing the travel time and distance, the authors identified connections with very low travel times for the distance that was travelled. The authors found that these "privileged" connections were often along corridors used by many commuters.

Instead of analyzing the individual path lengths of billions of individual O-D pairs, the authors of (Fielbaum et al. 2016), grouped stops by land use and geography into zones labeled as either "Central Business District" (CBD) or "periphery" (P) and then compared the quantity and travel time of links between CBD-CBD, CBD-P, P-CBD, and P-P. This was then compared to land use and population statistics to evaluate how well the network matched land uses.

Individual shortest paths can also be used to optimize schedules by minimizing the daily aggregate waiting time of all passengers. This method was demonstrated on metro networks in (Akse 2014), Using a Time-Expanded network with travel-time-weighted edges. The number of passengers flowing between each O-D pair was available and so by weighting the SPL by the passenger demand, the actual travel time for all passengers could be estimated for various possible schedules. Unfortunately, the method demonstrated requires very complete, precise ridership data and becomes impractically complicated to calculate for larger networks

2.7 Analysis of the Graph

The previously mentioned methods were all intended to study the underlying transit service with a minimal level of abstraction. While this maintains an intuitive understanding of the graph and calculated results, it also limits the study to obvious properties and expected relationships. However, since the late

1990's there have also been a growing number of studies which take a fundamentally different approach. Instead of focusing on properties of the transit system using graph theory, they study the structural properties of the abstract graph representation. This approach allows for a much broader range of properties to be considered, and for those studies to be done using graph representations with much greater abstraction from the underlying transit service.

A diverse range of such studies appears in the literature, although they fall into four broad types. The simplest approaches use system-wide or global, totals or averages to model the network as single, homogeneous structure. The second, more sophisticated, approach is to divide the network into partitions and then analyze the structure within and between these partitions. The third approach is to calculate properties of each node and investigate the distribution they follow. The final method is used to study the graph structure to investigate characteristics related to "Small-worlds" and being "Scale-free"

2.7.1 Global Graph Properties

The simplest way to consider a graph is to assume it is homogeneous, with similar connectivity and structure throughout and calculate a single metric based on system-wide totals. These measures are simple to calculate and useful to understand the large-scale properties of the network. The most commonly used Global characteristics related to graphs are the Graph Clustering Coefficient (GCC) (Derrible and Kennedy 2011), which are dependent on the choice of graph representation.

These total values, however, are commonly used as part of more sophisticated metrics, with GCC being used to identify small-worlds and the total number of edges and nodes being used to calculate the Alpha (α), Beta (β) and Gamma (γ) indices. α, β and γ were originally introduced into transportation networks by William Garrison and Duane Marble in the 1960s (Garrison and Marble 1962) (Garrison and Marble 1964) (Garrison and Marble 1965) and then adapted to public transportation in the work by Vuchic and Musso between 1988 and 1991. (Musso and Vuchic 1988) (Vuchic and Musso 1991).

The α index, also called Degree of Cyclicity, is based on the number of cycles in the network. A cycle is formed when a node is reachable from itself after travelling through the network. Cycles increase the number of possible paths between points in the network and so are useful in metrics related to connectivity (Watkins and Mesner 1967). α is a ratio of the number of cycles in the graph to the maximum possible number of cycles, both of which can be determined from the total number of vertices and edges in the graph. The number of cycles is equal to the number of edges in the graph, minus the number of vertices in the tree graph (which is one less than the number of vertices in the graph) (Berge 1962). The maximum number of edges will depend on whether the graph is planar or not, but cannot be defined where parallel edges are used. Vuchic renamed this indicator Circle Availability and used it as a means of describing the topology, with higher values indicating more choices are available to passengers and thus are less impact from congestion and higher resilience to failures.

β and γ are both simple ratios with the number of edges as the numerator with the β index, also called Complexity, using the number of nodes as the denominator while Connectivity (γ) using the maximum possible edges as the denominator. While initially adapted to use with public transportation networks by (Vuchic 1981/2007), they were repurposed in 2010 by (Derrible and Kennedy 2010) to describe the "State" of a network. They used these indicators to study metro networks and found that both indicators tend to increase as the number of stations in the network increase, progressing through three distinct phases. In the first phase, the networks are small and typically add new stations almost as often as the links between them, which results in complexity values near 1 and γ values near 0.33. The second phase, the network transitions to improving the connections between existing stations. As fewer new stations are needed resulting in links being added about twice as often as new stations. The stage occurs after many years of growth where β has increased to slightly below 2 and γ has increased to values near 0.66. (Derrible and Kennedy 2010)

While complexity and connectivity were originally developed for metro networks, they have also been used to study bus networks. In (Zhang et al. 2015), the authors demonstrated that connectivity could be used to measure connectivity on bus networks. They studied four bus networks in China and found all four networks had a complexity values between 1.32 and 1.46 with connectivity between 0.44 and 0.49 although they did not establish if they follow a similar evolution to those of metro lines. Another method was used in (Quintero-Cano et al. 2014) to analyze the Greater Vancouver Regional District bus network. The authors noted that in addition to having many more stops and lines of service, bus service is not limited to the same route or frequency in both directions and passengers often walk a short distance between stops to make transfers. To accommodate these characteristics, the authors combined stops within walking distance and used directed edges, weighted by frequency. They then derived variants of complexity and connectivity based on what the authors describe as “Normalized Frequency”. Instead of studying the entire network the authors chose to divide it into partitions and so the findings of this paper are further discussed next.

Table 2.3 Graph Global Properties

Variable	Symbol
Total Number of Lines	$\underline{L}$
Total Length of Lines	$\underline{L_t}$
Total Non-Overlapping Length of Lines	$\underline{L_{no}}$
Total Number of edges	$\underline{E}$

Total Number of Vertices/Nodes	$\underline{V}$
Total single Edges in network	$\underline{E^{st}}$
Total multiple edges in Network	$\underline{E^{m}}$
Degree	$\underline{K}$
Shortest Path Length (Shortest Path distance)	$\underline{SPL}$

Table 2.4 Formulas for Graph Global Properties

Measure Name	**Symbol**	**Formula**	
Network complexity Beta index	β	$\beta = \frac{E}{V}$	2.7
Degree of Cyclicity Alpha-index	α	$\alpha = \frac{E - V + 1}{E_{Max} - V + 1}$	2.8
Connectivity Gamma Index	γ	$\gamma = \frac{E}{E_{Max}}$	2.9
Global clustering coefficient	GCC	$GCC = \frac{3X\ number\ of\ triangles\ \ in\ the\ graph}{Connected\ triplets\ in\ the\ graph}$	2.10
Maximum Possible Edges	E_{max}	$\begin{cases} IF\ Planar = 3(V-2) \\ If\ non-planar = \frac{1}{2}V(V-1) \\ If\ multiple\ edges\ are\ used = \infty \end{cases}$	2.11

2.7.2 Partition Analysis

While global indicators are used to identify the global properties of the network, they use a very simplified model of the network, treating it as a single, homogeneous entity. However, most cities are heterogeneous due to geography, patterns of development and changes to planning and design standards as each city grows. To better understand the heterogeneity of the network, it is useful to divide the network into partitions which can then be investigated to understand the connectivity within and between these partitions. Dividing the network and studying the components can be done in several ways using either geographic areas or the graph structure itself.

As public transportation networks provide service throughout a geographic area, typically a municipality, several authors chose to study it by first partitioning it into smaller geographic. Two methods of dividing the geographic area found in the literature are: Using traffic analysis zones (TAZ) and segregating areas into either being in the central business district (CBD) or the periphery.

It is common practice for large municipalities to analyze the transportation network and predict traffic flow in terms of traffic analysis zones (TAZ). City Planners define each TAZ to correspond to areas within the city that are, relatively, internally consistent in terms of road patterns and land use. As such, TAZ is an ideal choice when dividing a city geographically to study individual areas.

For this reason, the authors in (Quintero-Cano et al. 2014) chose to divide the bus transit network for the Greater Vancouver Regional District into partitions based on the TAZ. As was mentioned previously, their graph representation had consolidated stops and represented links between them as weighted, directed edges. Instead of the distance or time, the edges were weighted using frequency of service that was first normalized by the link with the highest frequency in each zone. The authors found they could further reduce the network into an L-Space graph by adding the frequencies of overlapping routes, and then again normalizing the frequencies by the highest combined frequency in each TAZ. Using these indicators, they then derived β and γ indicators that used edges weighted by normalized frequency. This allowed them to analyze each TAZ separately and thus gain a more detailed understanding of the network structure, and how it varied between locations. While they only investigated seven adjacent TAZ subnetworks, they found dramatic variation between them which the authors speculate likely represents heterogeneity in geography, land use, and socioeconomic characteristics.

Another approach is to divide the network into geographic zones that are then each categorized as either being part of the Central Business District (CBD) or as part of the Periphery. In (Fielbaum et al. 2016) this methodology was used to evaluate the ideal network structure for several demand patterns. The authors identified three archetypal demand patterns (monocentric, polycentric & dispersed). They then model four idealized transit network structures (Direct Service, Trunk & Feeder, Hub & Spoke and Shuttle Service), as a directed L-space network with nodes representing zones and links weighted by estimated average travel times. The authors also included parameters representing differential costs for time spent waiting, travelling and transferring. Using these idealized models, the author found that hub and spoke networks can minimize travel times for nearly all demand structures, so long as passengers are willing to transfer multiple times. Feeder and Trunk networks are ideal only where demand is highly dispersed and direct lines are ideal where demand is mostly radial. While shuttle service may be preferable to passengers it requires a large fleet of small buses which is only economical if total demand is very high.

In contrast to using the physical shape of the underlying transit network, many authors chose to partition the graph based on its own features. The most commonly used methods seen in the literature are the identification of community structure and investigation of the distribution of properties between individual nodes. Community structure exists when nodes can be partitioned into more than one "communities" such that nodes are much more strongly connected to nodes within their community than to nodes outside of it (Peixoto 2015). This provides a means of investigating the connectivity between nodes in the network, regardless of their physical locations. While several methods have been developed to identify communities, only modularity maximization is regularly seen in the literature on public transportation networks.

The modularity (Q) is a measure of the number of connections within a community as compared to the connections between communities and is calculated using Formula 2.12 below, where e_{ii} is the percent of edges in community i and a_i is a percentage of edge ends in community i (Peixoto 2015). Various algorithms are used to generate numerous partition schemes and then compare modularity values of each until a maximal value is achieved. Graph modularity ranges from 0 (no community structure) to 1 (maximum community structure) with a small number of large communities tending to have much smaller values than a larger number of moderately sized ones (Sun et al. 2016).

$$Q = \sum_{i=1}^{c} (e_{ii} - a_i^2) \tag{2.12}$$

(Zhang et al. 2015) is typical of community identification on public transportation. The authors created undirected, distance-weighted, route-map graphs of bus networks in four Chinese cities which were then subjected to community identification using modularity maximization. They found that these networks all had very strong community structure with between 19 and 23 communities being identified and modularity values between 0.838 and 0.869. The authors provide a figure showing show the topological structure of the communities and note that the modularity indicators appear to be positively correlated with efficiency indicators. However, this was only one of several indicators is studied and so the authors did not explore the topic in further detail.

A similar method was used to study the impact of potential extensions and additional lines in the multimodal public transportation network in Athens (Tsekeris and Souliotou 2014). In this study, modularity was calculated for the current network and several potential future scenarios, all of which were represented as unweighted, undirected, L-Space graphs. The authors claim that increasing modularity indicates *"...the emergence of local network effects, through the creation of transfer stations in the core city and a few peripheral areas*". They found that the including the extensions increases the modularity, but that adding new lines can either increase or decrease the modularity. From this, they conclude that *"increased modularity is found to be more appropriate for the system at its early stages of development than the later ones, when it approaches a complete (and more hierarchical) graph structure".* In (Chatterjee et al. 2016) six bus networks in India were analyzed using an L-Space graph, finding between 4 and 6 communities in each although the modularity was not provided

A more abstract approach was taken in (De Bona et al. 2016), where the community structure was studied using an unweighted, undirected P-space representation of the public transportation network of the Brazilian city of Curitiba. Once again communities were identified using a modularity maximization method, although communities were permitted to overlap one another. Using this method, they identified 187 communities and calculated a modularity of 0.676, which indicates a strong community structure. The authors then describe how a strong community structure in a P-Space graph indicates a small number of nodes being responsible for all inter-community connectivity. As with scale-free networks, the loss of only a few of these hubs could dramatically reduce connectivity throughout the entire network.

An attempt to connect communities to the real-world operation of the public transit system was provided in (Sun et al. 2016). The authors study the multimodal public transportation network of Dublin Ireland, as an unweighted, undirected Route Map network. The unique contribution of this study was to then attempt to investigate the geography and structure of each of the communities. The first observation they note is that when these communities are shown on a map, they clearly follow geographic and geopolitical divides within the city. The role of nodes within each community is further explored using a "z" score to identify "hubs" that provide significant connectivity and a "p" score to identify importance in connecting communities to one another. The results show that only 4% of nodes provide significant connectivity with only 1% provide connectivity between communities and the most important hubs being located near the city center. While these findings may not be particularly surprising, they do indicate that community structure may have significance in the real world.

Unfortunately, the validity of these findings is questionable due to the very unusual community structure that was found. The best partition scheme that was found had a modularity of only 0.191, largely due to more than 75% of the nodes and 90% of edges being allocated to the largest single community. 34 other communities were identified, but only 4 were large enough to be significant. This community structure is contrary to most other studies in the literature and the low modularity may indicate the community structure is either not present or not significant.

2.7.3 Node Properties and Distributions

While partition analysis allows for heterogeneity between partitions, each individual partition is treated as being internally homogeneous which limits the "resolution" of these methods to large-scale structures (Fortuanto and Barthelemy 2006). Investigation of features smaller than this resolution requires using smaller subgraphs (Karypiu et al. 1999), of which the smallest possible example is a single node (Diestel 2010). Using individual nodes also removes the need to define internally homogeneous subgraphs and simplifies determining properties, but also limits analysis to properties which can be easily obtained for individual nodes. The properties most commonly seen in the literature are node degree, betweenness centrality, and clustering which are considered as a distribution. Node degree distribution is nearly always related to identifying scale-free characteristics and so is described in section 2.7.4, while node betweenness and clustering are mentioned here.

2.7.3.1 Node Betweenness Centrality

Graph theory does not mandate any particular method of laying out nodes or edges and so the same graph can be drawn in a myriad of different ways that would each have a different geometric centre. Instead of identifying a geometric centre, betweenness is often used to describe the centrality which is the number of shortest paths through the network that passes through a given node (Gross and Yellen 2004). As most passengers are assumed to use the shortest path from their origin to destination, betweenness is an intuitive way to identify nodes that are important, although this treats all O-D pairs as equally important (Derrible 2012). As the number of nodes increases so too does the number of possible O-D pairs and thus the total betweenness of the network. To highlight the relative importance of nodes, betweenness of individual nodes is often normalized by the total betweenness.

Using Route Map – Simple graph representations of 28 metro systems using UU links and only including termini and transfer stations the authors found that average betweenness tends to increase with network size according to a second-degree polynomial relationship, which the authors believe is due to metros being relatively planar.

A study done on six bus networks in L-Space in India found that betweenness centrality tended often followed an exponential law and while nodes with a high degree also had high betweenness, the opposite was not always true. They found that the cumulative distribution of betweenness also tended to follow a quadratic relationship. More interestingly, they found that as the number of stations increased the total betweenness of all nodes increased, but not uniformly. When betweenness was normalized, they found that the relative importance hubs tended to decline which the authors referred to as "democratization". This was seen in a reduction in the quadratic coefficient in the cumulative distribution, which the authors found tended to be proportional to the number of stations in the network.

2.7.3.2 Node Clustering

Another node property which is occasionally studied is the Local Clustering Coefficient (LCC). Unlike betweenness, which relies on shortest path calculations, LCC, is calculated in the same way as cyclicity (α), although the calculation is done for every node using that node's "neighbourhood" which is made up of the connections between "neighbours" which share an edge with the node being considered. Thus, a high LCC indicates that the nodes within the neighbourhood are well connected (Gross and Yellen 2004). The value of studying LCC is questionable, and the authors of (Xu et al. 2007) even stated that it was meaningless unless used in P-Space. In that paper, the authors did note that for the three bus systems studied in China LCC tended to follow a power law in P-Space. The authors in (Chatterjee et al. 2016) found the same results in bus networks in India when modelled in P-Space. However, they also studied CC in L-Space and found that it was inversely correlated to node degree, which they believe is due to those stops being part of many more bus routes which share few stops and thus neighbouring stops to be less well connected.

2.7.4 Small-world and Scale-Free Networks

Distributions of node properties are used to describe the heterogeneity of graph properties; however, there is a noticeable gap in knowledge between identifying these properties and interpreting them. The final methodology attempts to fill this gap by looking for a specific combination of properties that result in a well-understood structure when they occur together. Since they were first discovered at the end of the 1990s, there has been a noticeable increase in the analysis of public transit networks using the network science concepts of "Small-worlds" and "Scale-Free" networks. The vast majority of these are published in journals related to mathematics, network science and physics, and not those related to transit or public transportation. The authors of (Derrible and Kennedy 2011) believe this trend is due to mathematicians and network scientists using public transportation networks as real-world networks where theories developed in fields unrelated to transportation can be tested. They note that while not intentional, such studies may, none-the-less, benefit the transportation community.

2.7.4.1 Small-worlds

Small-world networks are structured so that only a "small" number of links must be traversed to travel between nodes in the network (Latora and Marchiori 2001). One famous example of a "small-world" is the parlour game "six degrees of Bacon", where nearly any actor or movie can be linked to Kevin Bacon through movie casts within 6 or fewer steps. This can be represented as a graph with nodes representing cast members and links connecting the cast members who worked together. The number of "degrees of Bacon" is the shortest path between Kevin Bacon and the chosen individual in the network.

Formally, small-world networks are defined by two characteristics, when compared to random graphs with the same number of edges and vertices (and thus the same average degree). The first is that global clustering coefficient (GCC), previously discussed, is found to be much larger than what is found in a random graph and the second is that the average shortest path (APL) is much smaller than the random graph. This was formally introduced into graph theory Watts and Strogatz in 1998 and was initially used in the study of social networks. However, researchers quickly found small-world properties were common to many unrelated fields that could be represented as graphs (Latora and Marchiori 2001).

Both APL and GCC are aggregated values that must be calculated iteratively for the entire network single scalar values used to represent the entire network. While both can be calculated for either weighted or unweighted networks, only unweighted graphs are used in the literature. APL is a measure of the distance to travel through the networks which require calculating the shortest path between all nodes in the network and then averaging them. GCC is a measure of how edges are distributed through the network, which requires identifying all "triples" and triangles in the graph as shown in formula 2.10 below.

$$GCC = \frac{3X\ number\ of\ triangles\ \ in\ the\ graph}{Connected\ triplets\ in\ the\ graph} \qquad 2.10$$

Small-world properties have been explored for undirected and unweighted L-space graph representations of many metro networks. A typical example of such studies is (Tsekeris and Souliotou 2014), where the authors model the current public transportation network in Athens along with scenarios where various extensions and planned lines are installed. They find that each of these networks demonstrates small-world characteristics with clustering coefficients between 0.353 and 0.574 and average path lengths between 10.195 and 6.891, which the authors state are much larger than equivalent random graphs. While the significance of the small-world property itself is unclear, the authors do claim that increasing the graph clustering coefficient (GCC) improves connectivity and robustness and shorter APL will result in shorter travel times.

In (Chatterjee et al. 2016), six bus networks in various locations in India were represented in L-Space graphs and found to have GCC between .08 and .26, which were all substantially larger than those of random graphs which were between 0.10 and 0.17. While they found that the APL, which ranged from 3.87 to 10.02, was larger than the random graphs, which ranged from 3.51 to 5.67, they concluded that those with APL very close to the random graphs (5.59 vs 4.09 & 3.87 vs 3.48) did have small-world characteristics. They also found those with small-world characteristics were systems with more geographic spacing between stop locations.

The authors in (Dimitrov and Ceder 2016) claim to be the first to investigate a directed graph for small-world properties. GTFS data for Auckland to generate an L-Space graph. The authors note that since bus stops are only available in a single lane of travel, the directionality is critical to accurately represent the bus network. Contrary to previous studies, they found that the directed graph did not have small-world characteristics. When compared to a random network of the same size they found that the global clustering coefficient of the Auckland PTN was 131 times larger than the random network (5.23x 10^{-2} vs. 3.97 10^{-4}) and the average shortest path was 68% longer than the random graph (29.13 vs. 17.25). The authors speculate this is likely the result of the directed edges and peculiarities of stop locations. For example, travelling between stops that are on opposite sides of a road requires travelling to stop used by routes travelling in both directions. While not mentioned by the authors, the introduction of walking links may have alleviated this.

2.7.4.2 Scale-Free Networks

Scale-free networks were first identified by Barabasi and Albert in 1999, as having the majority of connectivity provided by a small number of highly connected "hubs" while the vast majority of nodes are poorly connected. Formally, such networks are defined by having a node degree distribution (NDD) that follows a power law, which is described by the exponential coefficient, also called a "Scale Factor" (SF) which typically ranges from 2-3. Scale factor values larger than 4, however, indicate the distribution can also be fitted to an exponential distribution, which is characteristic of a random network. In addition, most scale-free networks also demonstrate small-world properties.

In most cases, hubs are not equally important but instead form a hierarchical structure resembling a tree or the familiar "hub-and-spoke" commonly seen in transportation systems. These networks typically arise when new links are preferentially attached to vertices that are already well connected, a phenomenon, sometimes referred to as "rich-get-richer". In public transportation networks, this is often due to a conscious choice by planners as focusing the majority of transfers at a small number of locations simplifies operations and provides travellers with more connections to important locations (Derrible and Kennedy 2011).

Scale-free transportation networks operate differently from other types of networks in two important ways. The first is that transfers are incredibly important because very few nodes are directly connected, and thus many trips will require transferring at one or more hubs. However, placing hubs at important destinations can dramatically reduce the number of transferring passengers. In addition, scale-free networks are very resilient to "random failures" but very susceptible to "targeted failures". Failures that occur at random will most often be at poorly connected nodes resulting in minimal loss of connectivity. However, a catastrophic loss of connectivity can occur if important hubs are intentionally targeted (Hadas et al. 2014).

The first study on scale-free characteristics in a public transportation network was done in 2004 by (Wu et al. 2004) which used an L-Space graph of the Beijing urban transit network, where only transfer and terminus nodes were included. The authors confirmed small-world characteristics and found it to have an exponent parameter of 2.24. In In 2007, the same method was used to confirm the public transportation networks of two other Chinese cities (Xu et al. 2007) and 14 major cities throughout the world (Von Ferber et al. 2008) were scale-free with scale factors between 1.24 and 4.99. The public transportation networks of 22 Polish cities were studied and all found to be scale-free with scale factors ranging from 2.4 to 4.1, although only seven were larger than 3. While this study also used L-Space graphs, it considered all stops. (Sienkiewicz, and Holyst 2005) Scale factors larger than 4 could be random or scale-free which might indicate those networks are more disperse or at different stages of network evolution, however, this does not appear to have been studied further.

Directed L-Space representations of can also be used to investigate scale-free characteristics (Dimitrov and Ceder 2016). In (Dimitrov and Ceder 2016) the authors studied both the entire public transportation network and the bus network only, and the while the NDD of both networks could be fitted to a power law with SF ~2.5 with an R^2 of ~0.86, they found that an exponential curve fit the data better with an R^2 of .99. However, because the network has hubs which are not present in other network types, they believe the Auckland PTN to be a hybrid network that is not entirely scale-free or random. In (Chatterjee et al. 2016), a similar evaluation was done for several types of bus networks in India and the authors found scale factors between 2.47 and 4.96, with most being between 3.0 and 3.6. The authors also compared the degree distribution to the number of parallel lines, equivalent to a Route Map-Simple graph, and found that the correlation varied dramatically between systems.

However, this contrasts with (Yang et al. 2014), where authors use directed, weighted L-Space graphs of bus networks in three large Chinese cities to investigate the node degree distribution (NDD) and average node clustering coefficient (NCC). They find that the node degree distributions (NDD) follow exponential distributions indicating *".... that new stations and routes are more likely to be distributed randomly during the evolution of [Bus Transit Networks] BTNs, probably aimed to obtain a wider coverage or to disperse the traffic flow evenly".* Similar results have been found in other Chinese (Chen et al. 2007) (Sui et al. 2012) and European (Von Ferber et al. 2009) cities.

NDD has also be used to study P-Space representations of numerous public transportation networks in China (Wu et al. 2004) (Huapu and Ye 2007) (Feng et al. 2016), Poland (Sienkiewicz, and Holyst 2005), India (Chatterjee et al. 2016) and Brazil (De Bona et al. 2016). While many of these studies found the NDD could be fitted to a power law, the accepted consensus is that an exponent curve is more appropriate (Feng et al. 2016). P-Space is an abstract representation of the service where node degree indicates the number of stops reachable without transferring. This is dependent on the number of routes that serve the stop as well as how many other stops are severed on each of those routes, making the interpretation of scale-free and small-world characteristics very challenging.

For both Small-world and Scale-free properties, the current state of the art research is providing inconsistent results with many transit systems being found to have both, others found to only have one or the other and still others showing neither. Some of the inconsistency is due to choices in graph configuration, particularly between P-Space and L-Space. However, this occurs even when limited to only studies in L-Space and there is little study on how the networks with these properties are different from those without.

2.8 Limitations in Current Research

It may seem surprising that despite being a powerful tool that was specifically developed to study transportation networks, graph theory is not commonly used to analyze actual transit networks by those who plan or operate them. As a mathematic tool, graph theory is an abstraction which represents the underlying transit service as a model. Those developing are given the conflicting mandates of including as many important characteristics as possible while also keeping the model as simple as possible. This is a particularly difficult task when both the needs of transit planners and available data vary widely between transit systems and are often poorly documented as well. The task is further complicated by the size and complexity of bus public transportation systems. In this context, it is not surprising that there is a gap between the tools researchers have developed and the actual needs of transit planners.

This gap is shrinking, thanks to the efforts of countless researchers, coupled with increasing computation power and improvements in the available data. However, the fundamental challenge is that most of the graph theoretic tools were intended to study metro networks. Metro networks typically operate a modest number of lines, which provide symmetric, frequent service to a limited number of important locations, with reasonably consistent spacing between them, running along a small number of established corridors. In addition, questions of connectivity are particularly important as metro networks offer a very limited number of possible paths between locations. L-Space and P-Space graphs using unweighted, undirected edges provide a reasonably accurate approximation of the service and are well suited for studying questions of connectivity.

However, in many cities, the majority of public transit service is provided by the bus network which both operates differently from and faces different challenges when compared to metro networks. Bus service connects a large number of diverse locations with variable demand spread throughout a large service area. In addition, bus service usually operates on city streets with inconsistent stop spacing and travel speed. To accommodate this, bus networks have many more lines and stops/stations with service that changes dramatically between time periods. As a result of the much larger number of stops and lines, it is often possible to transfer between lines by walking a short distance and many paths are often possible between locations in the network. In addition, lines of service in a bus network can be easily adapted to changing demand over time, or between time periods.

This shortcoming has been noted by several authors who have added elements such as directed edges, weighting by travel times, use of route map graphs and inclusion of pedestrian links. While these efforts have dramatically reduced the before mentioned gap, it is still far from being closed. The resulting graphs are much more complicated with little resemblance to the metro systems the indicators were developed for, the indicator values are more challenging to calculate, and authors are at a loss as to how planners should use the calculated values.

This thesis bridges the gap between the current graph theory techniques and the needs of planners by taking a different approach than that typically taken in the literature. As speculated by (Derrible and Kennedy 2011), starting with existing metrics and attempting to connect them to the needs of planners may not be the most fruitful approach. This insight suggests that a more successful method may be to first ascertain the needs of planners and then devise new metrics, or revise existing ones, to meet those needs. Unfortunately, formulating the needs of bus transit planners in terms of graph theory problems appears to remain within the before mentioned gap between planners and researchers.

Fortunately, insight into the needs of planners can be found by considering the networks they are tasked with. As mentioned previously, bus networks have many lines of service and transfers can be made at many locations, although a short walk is often required. From this, it is likely that multiple paths are available to travel between locations within the network. In addition, the service area is large and heterogeneous with a demand that varies both geographically but between time periods. Finally, bus service can be radically changed very quickly, which reduces its ability to influence travel patterns than a permanent metro line and instead is forced to constantly adapt to changes in demand.

From this it can be implied that questions of network structure and resilience to failures are less critical to bus planners who are more likely to be concerned with adjusting link availability, routing and frequencies to maximize the number of customers served and minimizing travel times. To accomplish this task, planners need tools to analyze the service provided and compare this to the needs of passengers. As such, graph models must accurately reflect the service provided in terms of the experience of travellers and metrics must be linked to the needs of travellers and expressed in terms of demand.

For the graph models to meet these requirements several critical aspects must be included. To reflect asymmetric service between inconsistently spaced stops, edges must be directed and weighted. To represent the experience of travellers, edge weights should be related to time. As transfers are a critical part of bus service, links must be provided to represent not only transfers between lines at a location, but also between nearby locations. Finally, some accommodation must be made to reflect time spent waiting at the origin, either through boarding links or frequency based edge weights.

Establishing metrics is more challenging because, while demand is usually synonymous with ridership, the needs of travellers are more ambiguous. Insight into the needs of travellers can be found in the literature which tends to focus on travel times, transfers and frequency. Another source of insight is TCSQM which includes those items, along with walking distance and various measures of comfort that are less suitable for graph theory analysis. Fortunately, all three of these items can be converted to units of time as frequency can be thought of as the amount of time spent waiting at the origin and transferring can be thought of as the time spent alighting, waiting and boarding at the transfer location. Time spent walking can be approximated by distance and an estimated walking speed and time spent on the vehicle can either be obtained from the schedule or distance and speed. This method might be further improved by applying a multiplier to each activity that represents the relative hardship; however, this requires obtaining and verifying those multipliers.

This thesis applies these insights to extend the current research in graph theory to better meet the needs of those planning and analyzing bus transit networks. First, the strengths and limitations of methods seen in the current literature are investigated by using them to analyze a transit network. Following this, new indicators are developed which specifically provide the information planners need, by combining ridership information and shortest path calculations through graphs with the highest level of accuracy and detail possible.

3 METHODOLOGY

Bus service is reactive to the needs of riders and as those needs change over time, so too must the service. Planners are tasked with adjusting the network to meet those needs, which can be expressed as a combination of maximizing ridership and connections while minimizing travel times. This task becomes increasingly difficult as the size and complexity of the networks grow, and so a graph theory based tool is needed. However, the graph must use weighted, directed edges, to reflect travel on the bus, walking, waiting and transferring. Finally, the resulting metrics from this tool should reflect travel times and ridership.

To this end, methods will be developed that define appropriate graph structures and metrics in four phases. In the first phase, the impact of using directed, weighted edges and pedestrian links will be investigated using several common graph representations. In the second phase, commonly used graph theory methods will be used to investigate the variation between time periods. The third phase will demonstrate new metrics developed to meet the actual needs of planners and the fourth phase will provide additional geographic context and characterize the service in each of the five time periods.

This challenging task could only be accomplished using detailed information for an actual multimodal public transportation system. The public transportation system in Edmonton, Alberta was chosen for several reasons. It was initially considered due to a previous relationship with the University of Alberta which allowed for detailed ridership information to be easily obtained. Moreover, Edmonton is an appropriate choice as it is a moderately large city with approximately one million residents that is facing many common problems such as; inconsistent road and land use patterns, urban sprawl, rapid growth and a public transportation system dominated by bus service. In addition, the entire schedule was available in the Google Transit Feed Standard (GTFS) format and the transit agency is currently undergoing a large-scale redesign of the bus network which might benefit from this study (City of Edmonton Budget Office 2017) (City of Edmonton 2017).

3.1 Data Sources

3.1.1 Google Transit Feed Specification (GTFS)

The GTFS data for the Edmonton Transit Service (ETS) public transportation network is freely available and was obtained from the City of Edmonton's "Open Data Catalogue" at https://data.edmonton.ca.

GTFS data is available through a number of comma-separated value (CSV) formatted text files with shared references to allow for the data to be interconnected. The majority of the data is stored in the "stop-times.txt" file which is divided into a number of specific "trips" each of which is represented as a sequence of stop-time events, each of which is assigned a specific minute. Information from "trips.txt", routes.txt and "calendar.txt" was also used to sort process this data into the graph links. Geographic information for every active stop in the network was obtained in the "stops.txt" file, which was used to calculate distances and group stops into geographic areas (Google 2017).

3.1.2 GIS Data

Additional geographic information was required to construct pedestrian links, to group stops by geographic area and used to provide visual context. Constructing the pedestrian links was accomplished using street data and aerial photographs obtained from Google Maps, OpenStreetMap.org and the Edmonton "Spatial Land Inventory Management" (SLIM). The transportation analysis zone (TAZ) data was also provided by Edmonton "Spatial Land Inventory Management" (SLIM). Data from the SLIM database was available through the City of Edmonton "Open Data Portal" at https://data.edmonton.ca/, although much of this data was compiled, formatted, organized and provided by Dr. Manish Shirgaokar.

3.1.3 Ridership Data

Ridership data was obtained from Edmonton Transit Service for the months of February, March and April of 2016. This information was provided in the form of CSV files with average boardings and alightings for each stop that was disaggregated to show the activity for each time that each route visited each stop, with separate data provided for Weekday, Saturday and Sunday service.

3.2 Software Packages Used

3.2.1 Microsoft Excel

Several versions of Microsoft Excel were used for initial inspection of the GTFS data, along with processing and visualizing input, output and intermediate data. In addition, most charts and tables were produced in Microsoft Excel.

3.2.2 Esri ArcGIS

Esri ArcGIS was used to construct the pedestrian links, group stops into TAZ and to visualize, verify and debug geographic information.

3.2.3 Konstanz Information Miner (KNIME)

Konstanz Information Miner (KNIME) is a free open source software (FOSS) graphical analytic platform that allows workflows to be constructed using hundreds of specialized nodes each of which must be configured. This program was used for many tasks, including; processing the GTFS data into links and nodes, generating maps, regression analysis, and processing data files that were beyond the capabilities of MS Excel.

3.2.4 Python 3

The majority of data processing, including analysis of the graphs, was done using the Python 3.5 and 3.6 programing language largely due to the very helpful developer community on stackoverflow.com and having countless free online resources. In addition, Python 3 has very intuitive syntax with comparable versatility and power with other languages such as R and N and is the chosen platform for the graph-tool library.

3.2.5 Graph-tool Python Library

To ensure each type of graph representation was analyzed in the same way, a single software platform was used on all graphs. While many graph analysis software packages are available, very few are able to efficiently process the Time-Expanded graphs which had as many as 1.2 million nodes and 3.9 million labelled, directed and weighted edges. This was found to be well beyond the capabilities of graphical programs such as Gelphi, Cytoscape, and Neo4j. While other free tools exist using programming languages such as R and C++ and Python, the graph-tool python library created by Tiago Peixoto was chosen as it uses the simple Python language but with much greater performance as most algorithms are implemented in C++. NetworkX was mentioned in several papers (King and Shalaby 2016) (Tsekeris and Souliotou 2014), but being written largely in python, it was found unable to provide the needed performance on the available computer hardware. Unfortunately, the graph-tool library is only available for Unix-like operating systems, and so the data processing environment was installed on a computer running Lubutnu, Linux.

3.3 Shortest Path Calculations

One of the greatest challenges in analyzing public transportation networks is efficiently calculating and processing shortest paths. For example, the ETS network has approximately 7,200 stops, which results in 51,840,000 possible O-D pairs. As mentioned earlier, increasing how accurately the graph represents the underlying network requires increasing the number of edges and nodes by several orders of magnitude, such that the Time-Expanded networks had several trillion possible O-D pairs. This challenge has been mentioned by several authors who resolved to study only a small portion of the network (Fortin et al. 2016) (Mishra et al. 2012). That approach was not considered adequate for this thesis and so much effort was spent identifying methods of calculating shortest paths between all stops in each Time-Expanded graph.

The graph-tool python library implementation of Dijkstra's algorithm is able to, very efficiently, calculate the distance from one point all other points in the network and returns the output as a two-dimensional array of scalar values representing either the total weight of all edges or the number of edges traversed if no weights are used (Peixoto 2014). This array of values can then easily be aggregated in several ways or matched to each location and output directly. While this is a very powerful tool, it is also limited because gathering any other information about the path requires iterating over each path individually which increases computer time a data storage exponentially. This can be partially circumvented by using edge weights that reflect the composite of multiple criteria. However, edge weights are limited to static, scalar, numeric values (not formulas, text labels, or vector/arrays).

In addition to the difficulty in obtaining data for the large numbers of O-D pairs results in hundreds of gigabytes of data or, tens of gigabytes of compressed data. Because of these limitations, two simplifications were required. The first was that when using the Time-Expanded network, trip origins were limited to stop-time nodes and final destinations were limited to "Base" nodes so that a separate trip was calculated between each stop in the network for each possible departure time. This method excluded meaningless paths between intermediate nodes, which accounted for the majority of possible O-D pairs. The second was that the code included some means of aggregating the calculated values in each iteration and only the aggregated values were retained. This slightly increased the process complexity but reduced the memory and data storage required by several orders of magnitude and sped computation by reducing the von Neumann bottleneck between CPU and RAM (Backus 1978).

3.4 Approach

The approach taken throughout this thesis was to treat each time period as if it were a separate network and then analyze it in four phases. In the first phase, the impact of graph type, edge weighting and including pedestrian links was investigated. This was done by representing each time period as four graph types (L-Space, Route Map-Simple, Route Map-Complex and Trip Map), each of which was constructed with four edge types (unweighted and undirected, unweighted and directed, travel time-weighted and directed, along with value of time-weighted and directed), both with and without pedestrian links. These graphs were then subjected to a battery of tests to assess the impact of the chosen graph type, edge type and inclusion of pedestrian links.

The second phase was to investigate how the network appeared to change in each time period, using common indicators and methodology seen in the current literature. This phase used many of the same indicators calculated in the first phase, as well as several indicators which directly measure the underlying service.

The third phase was to develop new metrics to meet the inferred needs of transit planners. These metrics required calculating travel times between all stops in the network in each of the five time periods. These values were then combined with distance, schedule availability and ridership to create four new metrics which reflect the experience of passengers.

The fourth phase combined information gained in the previous three phases with additional geographic information to characterize the service in each time period. The additional information was largely qualitative descriptions obtained from plotting the calculated values for each stop on a map.

3.4.1 Time Periods

As mentioned previously, graphs were created for several time periods to reflect that the bus network changes considerably over the course of a weekday and during the weekend. The city of Edmonton's Policy C539 (City of Edmonton Transportation Department 2009) defines 11 separate time periods within the Edmonton Transit Public Transportation Network, described in Table 3-1 below. However, as six of these time periods have very low usage, only the five time periods shown bold were analyzed.

Table 3.1 ETS Time Periods

Label Used	Time Period	Start Time	End Time	Span of Service
AM Peak	**Weekday AM Peak**	**6:00 A.M.***	**9:00 A.M**	**3 Hours**
PM Peak	**Weekday PM Peak**	**3:00 P.M.**	**6:00 P.M.**	**3 Hours**
Midday	**Weekday Midday**	**9:00 A.M.**	**3:00 P.M.**	**6 Hours**
N/A	Weekday Early Evening	6:00 P.M.	10:00 P.M.	4 Hours
N/A	Weekday Night	10:00 P.M.	2:00 A.M.**	4 Hours
N/A	Saturday Morning	6:00 A.M.	8:00 A.M.	2 Hours
Saturday	**Saturday Midday**	**8:00 A.M.**	**7:00 P.M.**	**11 Hours**
N/A	Saturday Night	7:00 P.M.	2:00 A.M. **	7 Hours
N/A	Sunday Morning	6:00 A.M.	10:00 A.M.	4 Hours
Sunday	**Sunday Midday**	**10:00 A.M.**	**7:00 P.M.**	**9 Hours**
N/A	Sunday Night	7:00 P.M.	2:00 A.M. **	7 Hours

N/A indicates time periods not considered for analysis in this thesis

* ETS planners use 6:00 a.m. for analysis although a very small amount of service begins earlier

** A small amount of service continues several hours later

The transit planners working for Edmonton Transit Service have a number of assumed characteristics for each of these time periods, which are typically grouped into the AM Peak, PM Peak and Off-Peak times.

AM Peak is known to have very high ridership, which is very focused in a narrow time period called the "peak of the peak", where the highest number of buses are in service at once. As such, service frequency is higher and nearly every active stop will have service. However, due to road congestion travel along some corridors is noticeably slower than during other times. The demand during this time is high and very directional with many passengers travelling from the periphery toward either the University of Alberta or the downtown core. Because of this, there is a significantly higher amount of express service and many local routes have "peak extensions" providing direct service to the downtown.

PM Peak is very similar to the AM peak, although demand is more dispersed temporally and geographically. This is due to workers leaving work at different times and passengers run errands or make personal trips instead of travelling straight home. This results in a noticeably higher total ridership as each trip a passenger makes is counted separately. Compared to the AM peak, the number of buses in service is slightly less than the AM peak and many routes run with slightly reduced frequencies. However, the number of active stops is slightly higher due to a few locations that are not open during the AM peak. Ridership is directional, from the University and Downtown to the periphery with express service and "peak extensions" in place to reflect this.

The off-peak (Weekday-Midday, Saturday and Sunday) all have similar characteristics, with both demand and service being less directional and dispersed both temporally and geographically than the AM or PM peaks. In addition, there is also a reduced amount of demand and service. The primary differences between the three are the span of service and the magnitude of both demand and service. Weekday-Midday has the shortest span of service but higher frequencies due to more demand per hour. Saturday has highest total demand due to having the longest span of service and has both higher demand and more service than Sunday

3.4.2 Graph Configurations

To investigate the behaviour of a range of different graph configurations, four graph representations were chosen, and four edge weighting schemes were used. As all graphs considered both with and without pedestrian links, a total of thirty-two (32) different graph configurations for each of the five time periods were developed. The graph types that were chosen were L-Space, Route Map-Simple, Route Map-Complex, and Trip Map. When combined with the chosen edge weighting schemes Unweighted-Undirected (UU), Unweighted-Directed (UD), Directed-Travel Time (TT), and Directed-Value of Time (VoT) these provide a wide range of complexity, abstraction, and precision. P-Space graphs were not included due to the level of abstraction and undirected weighted edges were not used because they are rarely seen in the literature.

3.4.3 Chosen Indicators

Several indicators are used throughout this thesis. These include both simple and sophisticated indicators from the current literature as well as several new indicators. As each phase investigates different characteristics, different indicators were deemed appropriate.

3.4.3.1 Phase I: Impact of Graph Configuration

Phase I used a common battery of indicators that were calculated for each graph type. For this phase, only indicators that describe the properties of the graph, as described in section 2.7, were used. This is because of the properties of the underlying service, such as the number of active stops and lines of service, are independent of the graph type and so remain constant for each time period were appropriate for this investigation. The chosen indicators, which are shown in Table 3.2, range in complexity from simple system totals to more sophisticated node property distributions and abstract metrics related to small-world and scale-free properties.

Table 3.2 Phase I Indicators

Type	Indicators
System Totals	Total Edges & Nodes, Cyclicity(α), Complexity (β), Connectivity (γ)
Node Property Distributions	Node Local Clustering Coefficients (LCC), Betweenness Centrality
Small-world & Scale-free	Identify Small-worlds (Average Path Length, Global Clustering Coefficient), Identify Scale-free (Node Degree Distributions)

3.4.3.2 Phase II: Changes between Time Periods

The second phase investigated how current metrics represented the variation of service between time periods. This was done using the indicators from phase I as well as those related to the underlying service and ridership, as shown in Table 3.3. Once again, these metrics range in complexity and sophistication, as well as provide information about both the graph structure and the underlying service.

Table 3.3 Phase II Indicators

Type	Indicators
System Totals	Total Edges & Nodes, Non-overlapping length, Number of lines, Transfer Nodes, Cyclicity(α), Complexity (β), Connectivity (γ), Directness (τ), Structural connectivity (ρ) and Maximum number of transfers required to traverse the network (δ)
Node Property Distributions	Node Clustering Coefficients, and Betweenness Centrality
Small-world & Scale-free	Identify Small-worlds (Average Path Length, Global Clustering Coefficient), Identify Scale-free (Node Degree Distributions).

Transit Service	Average and Distribution of Frequency, Ridership, Opportunities and Reachable O-D pairs

3.4.3.3 Phase III: New Indicators and Methods

In this phase, only indicators derived from the shortest path length (SPL) through the network are used. Those used in previous research include APL and the distribution of travel time in relation to the distance travelled. In addition, several adapted indicators were developed to reflect the relative importance of locations and travel through different parts of the city. To ensure the resulting metrics were accurate, they were all calculated using Time-Expanded graphs with edges weighted by travel times and which included pedestrian links.

One important limitation in the current literature is that each stop location is typically treated as equal to all others, despite some locations being used by thousands of passengers and others not being used at all. To address this, each O-D pair was weighted based on ridership by normalizing each origin by total boardings and each destination by total alightings. This method risks over-representing transit centres as they have many boardings and alightings despite not being either the origin or destination for many trips. Again, to account for this, transit centres were entirely removed from the calculation. This weighting scheme was then used to develop “weighted” and “unweighted” versions of several indicators.

The first method “Directed Speed” simply combined the travel time between locations with the change in geometric distance to an assumed destination location. This was first developed with the assumed destination being the nearest transit centre and then revised to use the centroid of the Edmonton city limits. This simple indicator provides a more meaningful method to quantify travel speed and was further refined by weighting locations based on ridership.

The remaining indicators were developed together using all stop locations which are given weights based on ridership. Weighted Potential (WP) is a connectivity indicator based on the schedule availability, Potential Travel Time (PTT) is based on the travel time between locations and Effective Travel Time (ETT) is a combination of WP and PTT. All of these were then developed using the average travel times between all stops in the entire network for each time period, as calculated using Time-Expanded graphs

3.4.3.4 Phase IV: Geographical Context

In the final phase, several indicators were calculated for individual stop locations and then plotted on a map using KNIME. This allowed for the disparity in various aspects of service quality to be visualized to provide additional context to the information generated in the previous sections. In addition, the ridership and frequency distributions were investigated in terms of distance from the core. The indicators used in this phase included ridership, frequency and the newly developed indicators Potential Travel Time and Effective Travel Time at each stop location. In addition, node clustering, betweenness and node degree, which were calculated for L-Space, Route Map-Simple, Route Map-Complex and Trip Map graph types.

3.5 Graph Construction

Before constructing graphs, complete, consistent sets of schedule and ridership data had to be obtained. While the Google Transit Feed Specification provides a simple system for providing scheduling information it does not dictate time periods, nor does it provide ridership data. As such, ridership data first had to be obtained and then the GTFS data for the same time period was matched.

Due to collective agreements and operation concerns, Edmonton Transit only makes significant changes to transit service five (5) times each year. These are referred to as “signups” because operators signup for specific blocks of work. The range of dates chosen for each roughly correspond to the seasonal changes in travel behaviour and typically run from February-April, April-June, June-September, September-December, and December-January. While planners use all five signups, they consider the signup from September to December to be the most representative of service as schools are in session, the weather is usually not overly severe and there are few major holidays. If service is abnormal or data quality suspect, the signup from February to April can be used instead as both schedules and ridership are typically very similar. The signups from April to June and June to September are impacted by the summer break and the December-January signup is impacted by the holiday season and often has unpredictable weather as well.

Initially, the data from the September-December signup was chosen for use, however, both the GTFS and ridership data was found to contain numerous inconsistencies and anomalies. Discussion with ETS planners and schedulers revealed this was due to sudden changes in travel patterns and unpredictable traffic congestion caused by the opening of the Metro LRT. As such, data from the February-April signup was used instead.

In addition to transit schedules and ridership data, the pedestrian links between stops were required. The pedestrian network was constructed using GIS data from SLIM and aerial photography provided by Google and OpenStreetMap.org. All sidewalks, trails, alleys, and roads in SLIM were combined to generate an initial pedestrian network. However, this network did not include many facilities that had been recently constructed or open areas such as parks, fields and parking lots that are regularly traversed by pedestrians. As such the pedestrian network was overlaid on aerial photographs and several hundred missing links were manually added. The OD cost matrix function in ArcGIS was then used to calculate the distance between all stops using this network with the maximum walking distance set at 400 meters as this is the maximum distance pedestrians are expected to walk to bus service in City Policy C539. These distances were then converted into travel times using a travel speed of 1.3 m/s, which is the value recommended by (APTA Standards Development Program 2009) and corresponds to the 5- and 10-minute walking distances stated in the Transit Oriented Development Guidelines adopted by the City of Edmonton (Sustainable Development and Transportation Services Departments 2012). As the temporal resolution of the Time-Expanded graph is limited to 1-minute increments, these values were then rounded to the nearest minute, with no values below 1-minute permitted.

3.5.1 Parsing the Data

To minimize opportunities for inconsistencies between the different graph representations, only the Time-Expanded graphs were derived directly from the GTFS data. Instead, the other four graph types were constructed systematically by combining specific nodes and links in the Time-Expanded graphs. This simplified debugging because any systematic errors were obvious with consistent impacts on either multiple graphs in a single time period, in multiple time periods or in multiple indicators.

Constructing the links began with first combining the arrival and departure times in stop-times.txt with the route number and day of week information stored in trips.txt. In addition to the regular service, many specialty routes, such as school specials, and regional service are included. Fortunately, these were easily removed as they have numbers greater than 399, with the exception of the LRT Routes 501 & 502 which were retained.

Information for each stop, including a location description along with the latitude and longitude, was then added from “stops.txt”. ETS uses a number of “virtual stops” that do not physically exist but are required to properly represent some locations that are under construction or have complicated geometry. In addition, many of the trips include duplicate events where the same trip-stop-time-route event is shown as multiple events in the sequence. These are largely due to minor errors in location data or software settings used to generate the GTFS data. Regardless of the cause, including these links would introduce errors into the graph and so they had to be identified and removed.

The next step was to properly represent transit centres as they are made up of numerous closely spaced stops but function as a single facility. As such, all stops at each transit centre were combined into a single stop that was given the name of the facility and given the average latitude and longitude coordinates of the stops that made it up. Walking links between the stops within each transit centre were removed while the distance and travel time for walking links connecting the transit centre to other stops was averaged and then rounded to the nearest 1 minute.

With this done, the GTFS data was divided into the 11 time periods based on the time of day and day of the week and grouped into separate batches of data. Only data for the 5 time periods being investigated was retained. For each time period, all stops with service were identified and all pedestrian links between those "active" stops were then added. For each of the five-time periods, five types of graphs were then created: L-Space, Route Map-Simple, Route Map-Complex, Trip Map and Time-Expanded.

3.5.2 Time-Expanded Graphs

Construction of the Time-Expanded graphs was performed by KNIME which used the combined GTFS and pedestrian link data. The first step was to generate three types of nodes: GTFS, "Stop-Time" and "Base". GTFS nodes were created for each stop-time-route-trip "event". While this created a large number of GTFS nodes, it allowed for each transfer to be modelled separately. Stop-Time nodes were then generated to represent each stop at each minute of the time period. Stop-time nodes accounted for the vast majority of nodes but were critical to represent transfers and pedestrian links. Base nodes represent the stop location without a temporal element or being part of any trip or route. These were used to simplify calculation of shortest travel times as they could be specified as a destination without requiring an arrival time to be specified.

The next steps involved creating the six types of directed links required to link all of the nodes into a network: GTFS, Boarding, Alighting, Waiting, "End" and Pedestrian. To allow for analysis, each link was given several characteristics including the source and target nodes, link type, time period, travel time, distance and "Value of Time" (VoT) as described in (Kittelson & Associates, Inc. 2013).

The sequence of stops in each trip was used to generate GTFS links by linking each sequence to the next in each trip. This required correcting the sequences for each trip to account for the removal of virtual stops, duplicate events, and aggregation of transit centres. For each link, the distance was calculated by the geometric distance between the stop locations and the travel time was calculated from the schedule data. Many GTFS links had only a temporal element which represented a bus "laying over" at a stop location for more than one minute. Unlike the other types of links, many GTFS links had travel times of zero due to the timestamps being limited to a resolution of 1-minute. To prevent potential cycles from forming in the graph, each GTFS node was connected to stop-time nodes by boarding and alighting links, which all had travel times of 1 minute. As such, transferring between two routes at a stop required at least 2-minutes, one to alight and another to board the next trip.

For each stop location, a number of waiting links and end links were constructed. Waiting links connected each stop-time node to the stop-time node at the same location one minute later and thus each had a distance of zero and a travel time of one minute. "End" links were constructed between every Stop-Time event and the base node at that stop. These links provided a one-way connection that allowed the end node to be chosen as a destination for shortest path calculations and had both distance and travel time equal to zero.

The final and most abundant type of link were the pedestrian links which were derived from the walking links previously mentioned. The first step was to remove two types of extraneous links that pedestrians would not actually use. The first type removed were links with at least one end connecting to a node that was not active in the time period. The second type of link was those that did not provide access to any additional trips, such as consecutive stops served by the same routes. With these extraneous pedestrian links removed, each stop-time node was connected to nearby stop-time nodes via pedestrian links using the distance and travel times previously calculated. As each stop location has separate Stop-Time nodes for each minute of the time period, a separate set of pedestrian links was created for each minute, thus resulting in millions of pedestrian links in each time period.

The completed graphs were then run through several filters and integrity checks to ensure that no duplicate links or loops had been created. The final result for each time period was then exported as comma separated value (CSV) text files which contained all the necessary information for each node and link to be constructed in the graph.

3.5.3 Trip Map Graphs

The simple trip map was created from the links previously created in the Time-Expanded graphs. There are two primary differences between the simple trip map and the Time-Expanded graph. Most fundamentally is that all nodes at each stop location are consolidated into a single "base" node representing that stop location without temporal information or being tied to any specific trip or route. As such, waiting, alighting, boarding and end links were no longer required and could be removed. The remaining GTFS and pedestrian links were then modified by converting the stop-time nodes they connected into the equivalent base nodes, resulting in many duplicate links with the same source and target nodes. Duplicate GTFS nodes were retained as each was tied to a unique trip, but pedestrian links were consolidated so that any two stops were connected by only a single pedestrian link.

3.5.4 Route Map-Complex Graphs

The complex route map was also created from the links in the Time-Expanded graphs. The complex route map continues to represent each route by a separate collection of route-stops so that transfers can be represented, but does not include a time element. The Time-Expanded links used individual trips, each of which was part of a specific route. The trip identifier of each node was replaced by the route number which resulted in many duplicates that could be removed. Each GTFS link was also updated to reference the consolidated nodes which created a great deal of edge duplication, although with some variation in travel times, distances and VoT. To capture as much information as possible when consolidating the links, Travel time, distance and VoT were averaged and the number of links being consolidated were counted as well.

Unfortunately, transfer links could not be generated as easily as the Time-Expanded graph relied heavily on waiting links that were no longer present and so all transfers would average to exactly two minutes (one minute for alighting and one for boarding), which was deemed unacceptable. Instead, the time for each possible transfer between routes was calculated for each stop using the minimum possible total alighting, waiting and boarding times. This transfer time was then averaged for each combination of routes at each stop and the VoT calculated from this new value. Pedestrian links were generated in a similar way with the walking link also included which resulted in all transfers being represented as route-stops directly linked to one another.

3.5.5 Route Map-Simple and L-Space Graphs

The simple route map and L-space map were both very similar to the trip map with the only difference being whether the GTFS edges represent individual trips, routes or simply the presence of a connection. This similarity allowed for the L-Space and Route Map-Simple graphs to be generated simultaneously from the previously created simple trip map.

The simple route map was generated by replacing each trip ID with the associated route ID and then averaging the travel time, distance and VoT for each link connecting the same nodes via the same route. The number of trip-links that were consolidated into each route-link were also counted. No further changes were required as pedestrian links were already consolidated and transfers at each stop were not represented.

The L-space graph was created by entirely removing the trip and route information on each link leaving each with only the starting and ending stop locations. Links connecting the same nodes were then counted and travel time, distance and VoT averaged.

3.5.6 Verify and Debug

In order to identify any potential errors, all KNIME workflows were carefully debugged which included inspection of intermediate and final outputs. Once the graphs were imported into graph-tool, they were each verified with a number of diagnostics to check for systematic problems or obvious errors by checking for loops, directionality and calculating shortest paths between predetermined O-D pairs with known travel times and distances. In addition, each type of node and link was systematically tested, both separately and in various combinations, using built-in functions of the graph-tool library.

To minimize the risk of errors in the analysis, a common set of python scripts were used and modified as little as possible to only reflect the inherent differences in the graph structures. For each indicator, and for each graph type, a number of checks were performed to confirm that the graphs were correctly represented in the graph-tool platform and verify the methods and parameters being used.

3.6 Phase I: Impact of Graph Configuration

The first phase of analysis was intended to identify the impact of choices made when representing a network as a graph. While countless choices are made regarding the platform, methodology and functional details, three items were thought particularly likely to significantly impact the resulting graph properties. The most obvious choice is the graph representation, which defines how the underlying service and stop locations will be represented in the graph. The choice of edge weighting is also critical as it defines what movements are possible along with what is used to determine the units of "distance" used when calculating shortest paths. The third choice is whether or not to include pedestrian links, which provide many critical transfer options but also dramatically increase the number of links in the graph.

To investigate the impact these three choices, a common battery of tests was calculated on thirty-two (32) graphs, which included four graph representations, four edge weighting schemes, with and without pedestrian links in each time period. Each was generated with four edge weighting schemes; unweighted, undirected (UU), Unweighted, directed (UD), Travel Time (TT) and Value of Time (VoT). While each test was run separately for each edge weighting scheme, only a subset of the tests actually reflected edge weighting.

Calculating the five global was done in three steps. The total number of edges and nodes was generated automatically by both KNIME and graph-tool during construction of each graph. This information was input into Microsoft Excel. Complexity (β) was then directly calculated as the ratio of edges to nodes, using formula 2.7, while the cyclicity (α) and connectivity (γ) were calculated from formulas 2.8 and 2.9, but required first calculating the maximum number of possible edges which was calculated from the number of edges and nodes using formula 2.11.

Python scripts were written to calculate the graph clustering coefficient (GCC), as well as local clustering coefficient (LCC), degree and betweenness centrality of each node in all graphs using built-in functions of the graph-tool python library. All of these scripts concluded by outputting the results as CSV text files which were then imported into KNINE where the average was calculated, and a scatterplot was inspected for obvious patterns. These were then fitted to exponential, logarithmic, linear, power and polynomial curves using the linear regression nodes in KNIME.

Calculating APL was more challenging as the built-in function for calculating the shortest paths between all O-D pairs would return error values if the path was not possible. To ensure these were properly accounted for, the "Counter" python function was used to count the number of times each path length occurred. As mentioned previously, this could not be performed on the Time-Expanded graphs, but

CSV files were generated for the other four graph types which were then imported into KNIME and Microsoft Excel for inspection and averaging of the non-error values.

In order to inspect for small-world characteristics, the GCC and APL had to be compared to random graphs with the same number of edges and nodes. A python script was written to generate ten random graphs and then the APL and GCC calculated as described above. The values calculated for each of the ten graphs were then averaged and the values compared to those of the actual graphs to identify if the GCC was larger and the APL smaller. This process was repeated for each graph type, in each time period, both with and without pedestrian. These random graphs were generated using UU and UD edges, although TT and VoT edges were not used, as those weighting schemes are not used for identification of small-world or scale-free properties in the literature.

3.7 Phase II: Changes between Time Periods

The second phase was intended to identify how the graph structure changed to reflect changes in the underlying service between time periods. This was accomplished largely by reusing the indicators calculated in the previous section, along with several additional indicators directly related to the underlying service instead of the graph representation as described in Table 3.3.

The number of lines of service was an intermediate output from directly from the GTFS data using summary data produced by KNIME when constructing the graphs. The number of transfer nodes was obtained as the number of nodes with service from more than one route and the maximum possible number of transfers was calculated for each stop as one less than the number of routes, which was then summed for the entire network. The number of multiple edges was obtained using KNIME to identify the number of edges connected each pair of nodes. Non-overlapping system length was simply equal to the total length of the L-Space graph without pedestrian lengths. Structural Connectivity (ρ) was then calculated from the number of transfer nodes, maximum possible transfers and number of multiple edges using equation 2.4.

Directness (τ) was more challenging as it required first calculating the maximum number of transfers on shortest paths through the network (δ) which can only be estimated by calculating the shortest paths between all stops (Derrible and Kennedy 2010). This was calculated as the shortest path using the Time-Expanded graphs with a customized edge weighting scheme. For this calculation, all pedestrian links were removed, boarding edges given a weight of 100,000 and all other edges given a weight of 1. As usual, base nodes were used as destinations, but GTFS nodes were used as sources instead of stop-time nodes to remove the initial boarding link in each trip. The number of transfers was then obtained by calculating all SPL from each stop and only retaining the longest SPL value. The number of transfers was thus simply found by simply truncating the total distance to the nearest 100,000 and then dividing by 100,000. Once δ was calculated, τ was then calculated using formula 2.3.

The only other indicators not already calculated in phase I was the frequency and opportunities. The frequency was calculated using KNIME as the number of GTFS events at each stop during the time period, divided by the number of hours in the time period. Calculating the number of opportunities was done using the time-expanded graphs, with the travel time limited to 90 minutes, the validity of an ETS ticket. To simplify the calculation, only the number of opportunities for each stop location was retained. These metrics were then matched to each stop in each time period, along with location, TAZ and distance from the centroid of the city using KNIME and then exported to Excel for inspection.

3.8 Phase III: New Indicators and Methods

In this phase, all indicators were derived from the shortest path length calculations between all stops in each network. However, these indicators were designed to reflect that locations are not equally important by investigating the direction of travel and incorporating ridership data.

The first such indicator developed was devised to avoid processing all O-D pairs by only calculating the travel times between each stop and the nearest transit centre, and the reverse trip from the nearest transit centre to each stop. The average travel times between each stop and the nearest transit centre was calculated using a python script that first identified all stop-time nodes as sources and all Base nodes as destinations and then removed all destination nodes that were not designated as being transit centres. As there are far fewer base nodes than stop-time nodes, the most efficient method was to calculate the travel times by reversing the graph and then calculating the travel time from each destination node to all possible origin nodes. The travel times were then aggregated recording only the stop number of both the origin and destination and then using the Counter function to record the number of occurrences each distance was calculated for each stop-stop pair. This process was then repeated for each destination node until all O-D pairs had been calculated and aggregated into a single list that recorded only the source node, destination node, and the number of times that each SPL value was calculated, which was then output into a compressed CSV file. Calculating the reverse trips was done in the same way, using only stop-time source nodes at transit centres.

These files were then imported into KNIME which discarded the errors resulting from O-D pairs for which no path was possible and then used the remaining values, along with how often they occurred to obtain average travel times. In addition, the distance geometric distance between the source and destination was calculated using KNIME to produce a “Directed Speed”. This dramatically reduced the number of shortest paths calculated, but travel time to the nearest transit centre is of limited value. A more useful metric was developed by extending the work by (Quintero-Cano et al. 2014), who had divided a transit network by Traffic Analysis Zones (TAZ) and (Fielbaum et al. 2016), who had studied travel between the core and periphery.

Average travel times between each pair of Traffic Analysis Zones (TAZ) was calculated using a python script that first identified all stop-time nodes as sources and all Base nodes as destinations. It then sorted lists of both origins and destinations by TAZ as doing so reduced the memory required and sped calculation significantly. As there are far fewer base nodes than stop-time nodes, the most efficient method was to calculate the travel times by reverse the graph and then calculate the travel time from each destination to all possible origins. The travel times were then aggregated recording only the TAZ of both the origin and destination and then using the Counter function to record the number of occurrences each distance was calculated for each TAZ-TAZ pair. This process was then repeated for each destination node until all O-D pairs had been calculated and aggregated into a single list that recorded only the SPL, and the number of times that each value was calculated for each TAZ-TAZ pair. This was then output into a compressed CSV file.

Each stop in the network was assigned to a specific TAZ using ArcGIS to combine the latitude and longitude information from the GTFS data and the geographic location and shape data for each TAZ from the SLIM data. Each TAZ was then further designated as either being part of the core outside of the core using the "Neighborhood Classification Map" as defined in the 2017 Annual Growth Monitoring Report released (City Planning 2017)by the City of Edmonton. To account for the different distances between each TAZ-TAZ pair, an average distance was calculated using the geometric distance of the centroid of each TAZ. Each of the 203,401 TAZ-TAZ pairs was then grouped as either representing a trip from Outer to Core, Core to Outer, Outer to Outer or Core to Core to provide a means of quantifying the "directionality" of service. This method was also used to calculate TAZ-TAZ travel times for each of the other four graph types to allow for comparison by graph type.

This information was then combined with the travel time data from the previous step in KNIME which calculated an average travel time based on the number of occurrences of each travel time for each TAZ-TAZ pair. Speed was calculated from the geometric distances and average travel times and then averaged by each direction and for the entire system.

After a significant upgrade to the computer hardware and several code optimizations suggested by individuals at stackoverflow.com, it was possible to calculate and process the full travel time matrix between all stops within the system using the Time-Expanded graphs. This method used similar methodology but instead of aggregating SPL by TAZ-TAZ, they were aggregated by the origin and destination stop. However, the compiled data for each destination stop had to be exported as a separate compressed CSV file and the list of travel times cleared after each stop to free RAM for the next iteration. In addition, the process required several times the RAM and processing time to complete compared to the previous TAZ-TAZ method. However, once this had been completed for the Time-Expanded networks, adapting it to the other graphs was trivial as they were much less complicated.

In addition, a new weighting scheme was developed where the importance of each location as a destination was estimated based on the total number of passenger alightings at that location over the course of the time period. To avoid counting transferring passengers twice, the activity at transit centres was entirely removed. The number of alightings at each stop, or TAZ, was then normalized using the total number of alightings in the time period to establish the relative importance as a destination. Passenger trips were then estimated between each O-D pair using the number of boardings recorded at each origin location multiplied by the relative importance of each destination. This was calculated for each stop-stop and TAZ-TAZ pair using KNIME along with the ridership data provided by ETS. This weighting matrix was then combined with the O-D travel time information to calculate "Passenger Weighted" versions of the TAZ-TAZ and Stop-Stop travel times and directionality.

While the two methods above provided a method to study travel between regions in the network, additional information is required to fully understand the level of service. As mentioned previously, the items considered critical to transit service include transfers, frequency, span of service, and number of locations reachable, as well as travel times. While transfers and frequency are partially reflected in travel times, the available connections and span of service must be added. In addition, the level of service must reflect both the quality of the inbound service and outbound service for each location (Fortin et al. 2016).

The first step to accomplish this was to formalize how each stop was weighted for trips to and from that location. Two ratios were created: "Source Importance" (SI), Equation 3.1, is a ratio of the boardings at that location to the total boardings in the time period and "Destination Importance" (DI), Equation 3.2, is the number of alightings at that location divided by the total number of alightings in the time period. These two values were then included in several additional indicators.

Two metrics similar to "opportunities" introduced by (Fortin et al. 2016) were developed to quantify the levels of inbound and outbound service for each O-D pair called "Potential Travel Times". In each O-D pair, the source has a "Potential Travel Time Departing" (PTD), weighted by the DI of the destination while the destination has a "Potential Travel Time Arriving" (PTA), weighted by the SI of the origin. The total PTD and PTA for each stop, shown in Equations 3.3 and 3.4, is then found by summing over all other stops in the network to get a measure of travel times to and from each stop. However, the potential travel times have an inherent ambiguity as a small value could be the result of either short travel times or few available connections and, so they must be normalized by a connectivity indicator.

The connectivity indicator should reflect the number of connections available, the importance of those connections and their temporal availability. DI and SI provide the relative importance of each stop, and the temporal availability was represented by the percentage of minutes in the time period when the connection is available or "Schedule Availability" (SA), calculated using Equation 3.5. SA was then combined with SI, using Equation 3.7, to determine the "Weighted Potential Arrivals" (WPA), and with DI to create the "Weighted Potential Departures" (WPD), using Equation 3.6.

The final metrics used to quantify the quality of service were named "Effective Travel Time" (ETT) and represent an average travel time, normalized by the importance of each connection and adjusted to reflect the importance of unavailable connections. As before, each location has separate values for service from important locations to that stop (ETT^A), found by Equation 3.9, and service from that stop to all important destinations (ETT^D), using Equation 3.8.

Table 3.4 Variables for New Indicators

N = number of active stops in the time period

Stop i = (Source/Origin)

Stop j = (Destination)

TT_{ij} = Travel Time from stop i to stop j

A_j = Alightings at stop j

B_j = Boardings at stop i

A_{tot} = Total Alightings at all stops in the time period

B_{tot} = Total boardings at all stops in the time period

TR_{ij} = Minutes when stop j is reachable from stop i in the time period.

TR_{tot} = Total minutes in the time period

Table 3.5 New Indicator Formulas

SI_i = Importance of stop i as a source	$SI_i = \frac{B_i}{B_{tot}}$	3.1
DI_j = Importance of stop j as a destination	$Di_j = \frac{A_j}{A_{tot}}$	3.2
PTD_i = Potential Travel Times Departing from Stop i to all reachable stops	$PTD_i = \sum_j^N TT_{ij} * DI_j$	3.3
PTA_j = Potential Travel Times Arriving from all stops to stop j	$PTA_j = \sum_i^N TT_{ij} * SI_i$	3.4
SA_{ij} = Schedule availability from stop i to stop j	$SA_{ij} = \frac{TR_{ij}}{TR_{tot}}$	3.5
WPD_i = Weighted Potential Departures from Stop i	$WPD_i = \sum_j^N SA_{ij} * DI_j$	3.6

WPA_j = Weighted Potential Arrivals at Stop j

$$WPA_j = \sum_{I}^{N} SA_{ij} * SI_i \qquad 3.7$$

$ETT^D{}_i$ = Effective average travel time departing from stop i to all other stops

$$ETT_i^D = \frac{PTD_i}{WPD_i} = \frac{\sum_j^N TT_{ij} * DI_j}{\sum_j^N SA_{ij} * DI_j} = \frac{TR_{tot} * \sum_j^N TT_{ij} * A_j}{\sum_j^N TR_{ij} * A_j} \qquad 3.8$$

$ETT^A{}_j$ = Effective average travel time arriving at stop j from all other stops

$$ETT_j^A = \frac{PTA_j}{WPA_j} = \frac{\sum_j^N TT_{ij} * SI_i}{\sum_j^N SA_{ij} * SI_i} = \frac{TR_{tot} * \sum_j^N TT_{ij} * B_i}{\sum_j^N TR_{ij} * B_i} \qquad 3.9$$

The values of each of these metrics were calculated based on the Stop-Stop travel time calculations performed on the Time-Expanded networks and ridership data for each stop in each of the five-time periods using KNIME. These values were then plotted onto maps where the colour of each stop was based on the ETT and the size of each stop reflected the importance. The colour scale was maintained, although the scale used for size was recalculated for each time period.

In addition, these indicators were investigated in several ways. The first was to investigate how WPA, WPD, $ETT^{A,}$ and ETT^D were distributed by plotting each as an "unweighted" cumulative percentage of the number of active stops in the time period and as a "weighted" the cumulative percentage of ridership activity. System averages of these were also calculated with "unweighted" values giving all stops equal weighting while the "weighted" values instead used SI or DI as appropriate. A third metric was investigated called "Effective Speed". For this calculation, PT-Distance was calculated in the same way PTA & PTD were calculated, although using the geometric distance between O-D instead of the travel time. The effective speed was then just the ratio of the PT-Distance to the Effective Travel Time.

3.9 Phase IV: Geographic Context

Few new indicator values were calculated in Phase IV; instead several of the indicators previously calculated were considered together and mapped to gain additional insights. The coverage data was found using ArcGIS with a 400-meter radius buffer assigned to each stop. The buffers were then merged to remove any overlapping area.

The centroid of the Edmonton City boundary was calculated using ArcGIS and found to be very near to Churchill Square in the downtown. The geometric distance between each stop location and the centroid of the Edmonton city boundary was calculated using the latitude and longitude information from the GTFS data using KNIME. The frequency and ridership in each time period were then combined in KNIME and then the data were plotted using Microsoft Excel and trend lines calculated.

The maps were generated using KNIME by linking each calculated indicator value to the appropriate stop and then assigning colours to each location based on that value. Each map was then produced as both a low resolution "overview" map and a higher resolution "detail" map which were then output as portable network graphics ".png" files and visually inspected.

4 RESULTS AND DISCUSSION

Following the methodology presented in section 3, the graphs constructed for each time period were subjected to a range of tests intended to investigate aspects of the graphs as well as the underlying transit service. These were grouped into four phases based on the purpose of the tests being performed. In Phase I, graph indicators was used to investigate the impacts of the chosen graph representation while in Phase II indicators for both the graphs and underlying service were used to investigate changes between time periods. Phases III and IV took a different approach, with Phase III presenting newly developed indicators, intended to meet the needs of transit planners and Phase IV providing additional geographic information to form a subjective characterization of service in each time period.

4.1 Phase I: Impact of Graph Configuration

As mentioned earlier, the impact of each of the three "choices" made when determining a graph configuration (i.e. graph type, edge type, and pedestrian links) was investigated. To identify the impact of each on the graph structure and resulting indicators, a battery of tests was conducted on up to 40 graph configurations for each of the five time periods, although not all configurations were possible for all tests. The resulting indicators for each time periods were compared to investigate the impact of each of the three choices. As the changes in underlying service remain constant in each time period, only tests which depend on the graph representation described in section 2.7 were used.

Several of the graph configurations were excluded from specific tests, due to limitations of the tools used to analyze the network. The most apparent is the Time-Expanded graphs which were excluded from community detection, modularity, APL, betweenness, and which only had useful NDD and LCC values when pedestrian links were included. The extremely sparse configuration resulted in too few values for NDD and LCC to obtain a distribution, while the size and complexity of these graphs were beyond the capability of the hardware and software. In addition, the very large number of parallel edges in the Trip Map graphs, and several of the Route Map-Simple graphs caused errors when calculating betweenness, and so they were excluded as well.

Several graph configurations were excluded from specific tests due to limitations of the methodology. The first example of this is that Cyclicity (α) and Connectivity (γ) both require calculation of the maximum possible edges which is not defined for the RM-Simple or Trip Map graphs. In addition, edge weighting is either irrelevant or only partially considered in several metrics, which will be discussed in more detail in section 4.1.2.

The findings are presented first by the choice (i.e. graph type, edge type and pedestrian links) and within each element, the tests are presented in the same order shown in section 2.7, with global properties presented first, followed by partition analysis, distributions of node properties and finally the presence of small-world and scale-free properties.

4.1.1 Impact of Graph Type

The first "choice" investigated was the graph type, which defines the interpretation of edges and nodes in the graph. Five graph types were considered with the simplest being L-Space graphs, which represented all connections between each pair of nodes as a single edge, and transfers at each stop are not shown at all. The Route Map-Simple graphs represented each route separately while Trip Map graphs showed each trip, both of which used simple stops and so could not represent transfers at stops. The Route Map-Complex and Time-Expanded graphs were analogous to the RM-Simple and Trip Map graphs, although using complex stops so that transfers could be represented correctly. As will be shown in the following sections, the choice of graph type had a dramatic impact on nearly all properties studied, although several graph types show clear relationships.

4.1.1.1 Global Indicators

For this investigation five global indicator values were considered: the number of edges and nodes, as well as the Degree of Cyclicity (α), Complexity (β), and Connectivity (γ) indices.

As shown in Table 4.1 and Figures 4.1 and 4.2 below, the three graph types with simple stops: L-Space, RM-Simple, and Trip Map all have the same number of nodes, which is equal to the number of active stops while the RM-Complex graphs used 2.5 and 2.8 times as many nodes and the Time-Expanded graphs further increased the number nodes by several orders of magnitude. This property was the expected result of the RM-Complex and Time-Expanded graphs using complex stops, where each stop is represented as a collection of nodes.

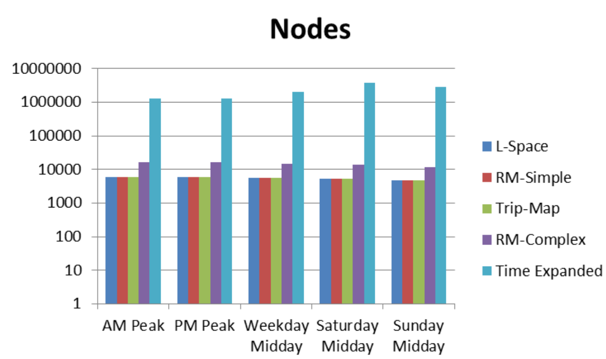


Figure 4.1 Nodes by Graph Type

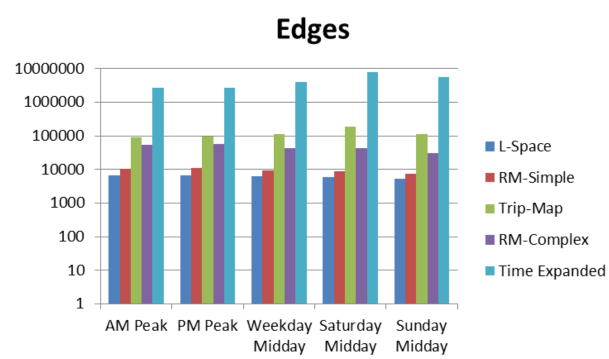


Figure 4.2 Edges by Graph Type

The number of edges changed between each type of graph with the RM-Simple being slightly larger than L-Space, and the Trip Map graphs having between 8 and 20 times as many links as the RM-Simple. This trend is another expected result of the graph type, as most of the links in the L-Space represent multiple routes, which are shown in the RM-Simple graph and most routes are made up of multiple trips, shown in the Trip Map graphs. The RM-Complex graphs had approximately 5 times as many edges as the RM-Simple graph, due to including transfer and waiting links, but much fewer links than the Trip Map, indicating that there are fewer transfer links between routes than there are trips on those routes. Finally, the Time-Expanded graphs hade several orders of magnitude more links than any of the other graph types, which is due to including transfers between every trip and time spent waiting in 1-minute increments.

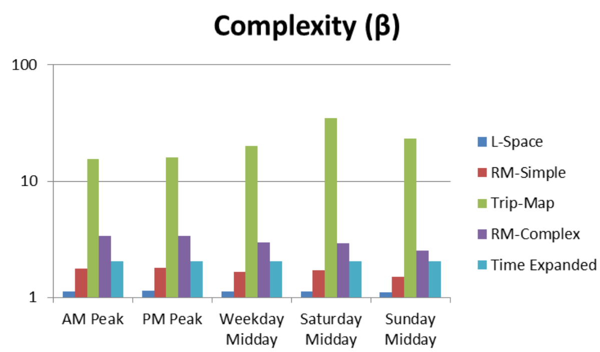


Figure 4.3 Complexity by Graph Type

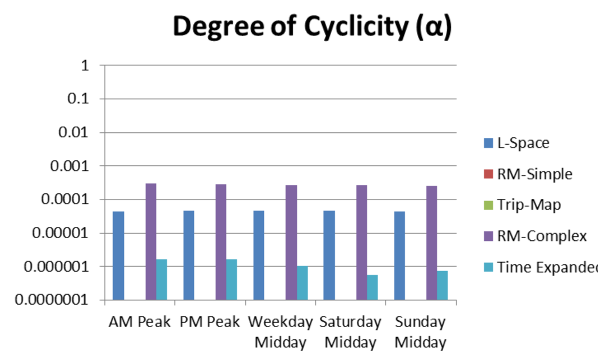


Figure 4.4 Cyclicity by Graph Type

Complexity (β) shown in Figure 4.3, which is a simple ratio of the edges to vertices, saw a substantial increase between L-Space and RM-Simple, followed by a dramatic increase between RM-Simple and Trip Map, due to the increase in the number of edges while the number of nodes remained constant. This was a dramatic decline between Trip Map and RM-Complex, due to the fewer edges and increase in nodes, and a modest decline between RM-Complex and Time-Expanded due to dramatic increases in both the number of edges and nodes.

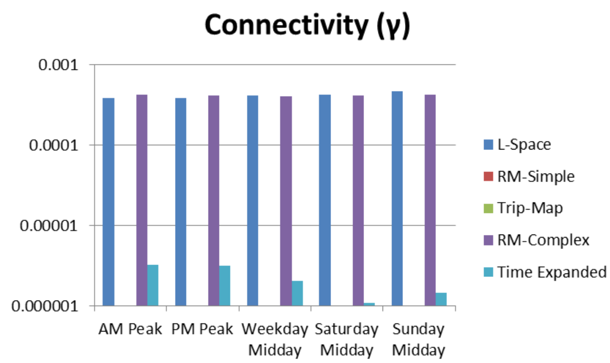


Figure 4.5 Connectivity by Graph Type

Table 4.1 System Totals

Graph Type	Ped Links	Time Period	Edges					Nodes				Basic Indicators			
			Total	GTFS	Ped	Transfer	End	Total	Base	GTFS	StopTime	β	E_{max}	α	γ
L-Space	With Ped Links	AM Peak	**25,635**	6,604	19,031			**5,844**	5,844			4.39	1.707E+07	1.159E-03	1.501E-03
		PM Peak	**26,254**	6,747	19,507			**5,930**	5,930			4.43	1.758E+07	1.156E-03	1.493E-03
		Midday	**23,990**	6,186	17,804			**5,485**	5,485			4.37	1.504E+07	1.230E-03	1.595E-03
		Saturday	**23,086**	5,907	17,179			**5,265**	5,265			4.38	1.386E+07	1.286E-03	1.666E-03
		Sunday	**19,943**	5,256	14,687			**4,753**	4,753			4.20	1.129E+07	1.345E-03	1.766E-03
	No Ped Links	AM Peak	**6,604**	6,604				**5,844**	5,844			1.13	1.707E+07	4.457E-05	3.868E-04
		PM Peak	**6,747**	6,747				**5,930**	5,930			1.14	1.758E+07	4.653E-05	3.838E-04
		Weekday Midday	**6,186**	6,186				**5,485**	5,485			1.13	1.504E+07	4.668E-05	4.113E-04
		Saturday Midday	**5,907**	5,907				**5,265**	5,265			1.12	1.386E+07	4.640E-05	4.263E-04
		Sunday Midday	**5,256**	5,256				**4,753**	4,753			1.11	1.129E+07	4.463E-05	4.654E-04
Route Map Simple	With Ped Links	AM Peak	**29,457**	10,426	19,031			**5,844**	5,844			5.04			
		PM Peak	**30,219**	10,712	19,507			**5,930**	5,930			5.10			
		Weekday Midday	**26,955**	9,151	17,804			**5,485**	5,485			4.91			
		Saturday Midday	**26,171**	8,992	17,179			**5,265**	5,265			4.97			
		Sunday Midday	**21,895**	7,208	14,687			**4,753**	4,753			4.61			
	No Ped Links	AM Peak	**10,426**	10,426				**5,844**	5,844			1.78			
		PM Peak	**10,712**	10,712				**5,930**	5,930			1.81			
		Weekday Midday	**9,151**	9,151				**5,485**	5,485			1.67			
		Saturday Midday	**8,992**	8,992				**5,265**	5,265			1.71			
		Sunday Midday	**7,208**	7,208				**4,753**	4,753			1.52			
Trip Map	With Ped Links	AM Peak	**109,319**	90,242	19,077			**5,844**	5,844			18.71			
		PM Peak	**114,835**	95,262	19,573			**5,930**	5,930			19.37			
		Weekday Midday	**127,274**	109,230	18,044			**5,485**	5,485			23.20			
		Saturday Midday	**199,082**	181,704	17,378			**5,265**	5,265			37.81			
		Sunday Midday	**125,171**	110,102	15,069			**4,753**	4,753			26.34			
	No Ped Links	AM Peak	**90,242**	90,242				**5,844**	5,844			15.44			
		PM Peak	**95,262**	95,262				**5,930**	5,930			16.06			
		Weekday Midday	**109,230**	109,230				**5,485**	5,485			19.91			
		Saturday Midday	**181,704**	181,704				**5,265**	5,265			34.51			
		Sunday Midday	**110,102**	110,102				**4,753**	4,753			23.16			
Route Map Complex	With Ped Links	AM Peak	**170,593**	10,540	115,439	34,244	10,370	**16,214**	5,844	10,370		10.52	1.314E+08	1.175E-03	1.298E-03
		PM Peak	**178,633**	10,826	122,130	35,024	10,653	**16,583**	5,930	10,653		10.77	1.375E+08	1.179E-03	1.299E-03
		Weekday Midday	**126,560**	9,245	83,432	24,799	9,084	**14,569**	5,485	9,084		8.69	1.061E+08	1.055E-03	1.193E-03
		Saturday Midday	**119,495**	9,126	77,839	23,623	8,907	**14,172**	5,265	8,907		8.43	1.004E+08	1.049E-03	1.190E-03
		Sunday Midday	**78,465**	7,312	48,414	15,593	7,146	**11,899**	4,753	7,146		6.59	7.079E+07	9.404E-04	1.108E-03
	No Ped Links	AM Peak	**55,154**	10,540		34,244	10,370	**16,214**	5,844	10,370		3.40	1.314E+08	2.963E-04	4.196E-04
		PM Peak	**56,503**	10,826		35,024	10,653	**16,583**	5,930	10,653		3.41	1.375E+08	2.904E-04	4.110E-04
		Weekday Midday	**43,128**	9,245		24,799	9,084	**14,569**	5,485	9,084		2.96	1.061E+08	2.691E-04	4.064E-04
		Saturday Midday	**41,656**	9,126		23,623	8,907	**14,172**	5,265	8,907		2.94	1.004E+08	2.737E-04	4.148E-04
		Sunday Midday	**30,051**	7,312		15,593	7,146	**11,899**	4,753	7,146		2.53	7.079E+07	2.564E-04	4.245E-04
Time Expanded	With Ped Links	AM Peak	**6,406,710**	90,464	3,787,677	1,353,925	1,174,644	**1,273,353**	5,844	92,865	1,174,644	5.03	8.107E+11	6.332E-06	7.903E-06
		PM Peak	**6,554,946**	95,545	3,886,023	1,381,448	1,191,930	**1,295,967**	5,930	98,107	1,191,930	5.06	8.398E+11	6.262E-06	7.806E-06
		Weekday Midday	**10,179,250**	109,856	6,108,707	2,090,302	1,870,385	**1,988,571**	5,485	112,701	1,870,385	5.12	1.977E+12	4.143E-06	5.148E-06
		Saturday Midday	**19,516,982**	184,449	11,791,974	3,955,094	3,585,465	**3,778,338**	5,265	187,608	3,585,465	5.17	7.138E+12	2.205E-06	2.734E-06
		Sunday Midday	**14,084,300**	111,782	8,417,431	2,888,654	2,666,433	**2,784,841**	4,753	113,655	2,666,433	5.06	3.878E+12	2.914E-06	3.632E-06
	No Ped Links	AM Peak	**2,619,033**	90,464		1,353,925	1,174,644	**1,273,353**	5,844	92,865	1,174,644	2.06	8.107E+11	1.660E-06	3.231E-06
		PM Peak	**2,668,923**	95,545		1,381,448	1,191,930	**1,295,967**	5,930	98,107	1,191,930	2.06	8.398E+11	1.635E-06	3.178E-06
		Weekday Midday	**4,070,543**	109,856		2,090,302	1,870,385	**1,988,571**	5,485	112,701	1,870,385	2.05	1.977E+12	1.053E-06	2.059E-06
		Saturday Midday	**7,725,008**	184,449		3,955,094	3,585,465	**3,778,338**	5,265	187,608	3,585,465	2.04	7.138E+12	5.529E-07	1.082E-06
		Sunday Midday	**5,666,869**	111,782		2,888,654	2,666,433	**2,784,841**	4,753	113,655	2,666,433	2.03	3.878E+12	7.432E-07	1.461E-06

Cyclicity (α) and Connectivity (γ), shown in Figures 4.4 and 4.5, both require calculation of the maximum possible edges which is not defined for graphs with parallel links and so could not be calculated for the RM-Simple or Trip Map graphs. For the remaining three graph types, RM-Complex showed the highest in both metrics, although L-Space was much closer in terms of γ than α, and Time-Expanded had very small values for both. These trends indicate that RM-Complex has more possible paths through it and thus is more robust to failures than the L-Space. The very low value for the Time-Expanded graphs is the result of the number of possible edges having a quadratic relationship with the number of nodes, resulting in the number of possible edges increasing much faster than the number of actual edges. Furthermore, the Time-Expanded graphs were acyclic, and so any values of α other than 0 are entirely due to the formula not correctly accounting for directed links.

When taken together the global indicators demonstrate that the choice of graph type results in dramatic changes in the resulting values, so much so that a logarithmic scale is required for them to appear on the same graph. In addition, the commonly used α and γ indicators are only meaningful for three of the five graph types.

4.1.1.2 Node Distributions

Both the distribution of local clustering coefficient (LCC) and betweenness were investigated to identify the impact of graph type. As LCC depends on the directionality, both the distributions for the directed and undirected LCC were calculated. As shown in Table 4.2, both the directed and undirected local clustering coefficients followed similar trends. It was not possible to fit the Time-Expanded graphs, without pedestrian links, to a distribution and the RM-Complex graph only fit the logarithmic curve. However, the L-Space, RM-Simple, and Trip Map graphs all fit very well to a power law but had a slightly higher R^2 value when fit to a logarithmic curve. This is consistent with the previous literature (Chatterjee et al. 2016) (Xu et al. 2007) as only exponential and power-law distributions were usually considered. However, as shown in Figure 4.6, the scale coefficient of those distributions follows very different trends between the graph types, with L-Space having the highest values, RM-Simple having small positive values, Trip Map having very small negative values and RM-Complex having large negative values.

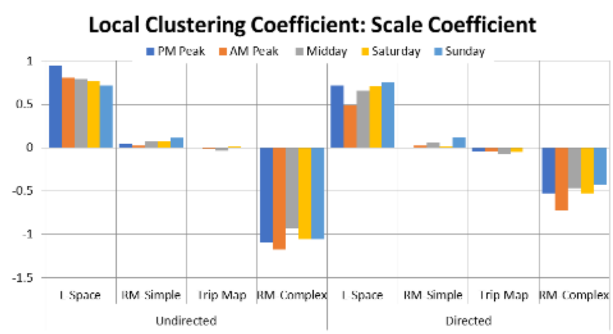


Figure 4.6 Local Clustering Scale Coefficient

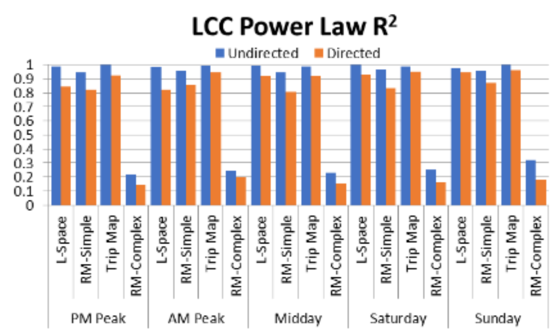


Figure 4.7LCC Power Law R^2

Moreover, the trends between time periods vary with the graph type as well. Average LCC values vary so dramatically that they can only be shown together using a logarithmic scale as shown in Figure 4.7. L-Space and RM-Simple showed very similar values, which were highest for Midday and Saturday. The trip map graphs followed a similar trend, for peaks and Sunday but showed notably higher undirected and lower directed clustering for Midday and Saturday. The Route Map-Complex graphs had a substantially higher average and also followed a different trend with the peaks having higher values than off-peak. This appears to indicate that the number of links has little impact on average clustering while the use of complex stops dramatically increases clustering. Interpretation of this finding is unclear; however, it is clear that graph type has a substantial impact on how node clustering is distributed.

Betweenness could not be calculated for the Time-Expanded, Trip Map graphs, and several of the Route Map-Simple graphs. As shown Figure 4.8, the remaining graphs fit the logarithmic curve best, for both directed and undirected betweenness, although the directed value for the RM-Simple fit almost as well to the linear curve. Despite this similarity, there are clearly differences between the graph types, as the magnitude of the R^2 for all four curves varies, as does the average betweenness values shown in Figure 4.9. This variation is found regardless of the edge type and inclusion of pedestrian links, indicating that travel through these graphs is fundamentally different, despite being used to represent the same underlying transit service.

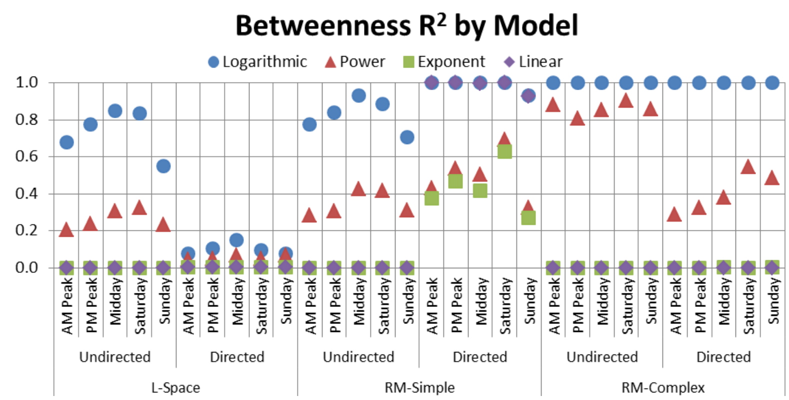


Figure 4.8 Curve Fits for Node Betweenness

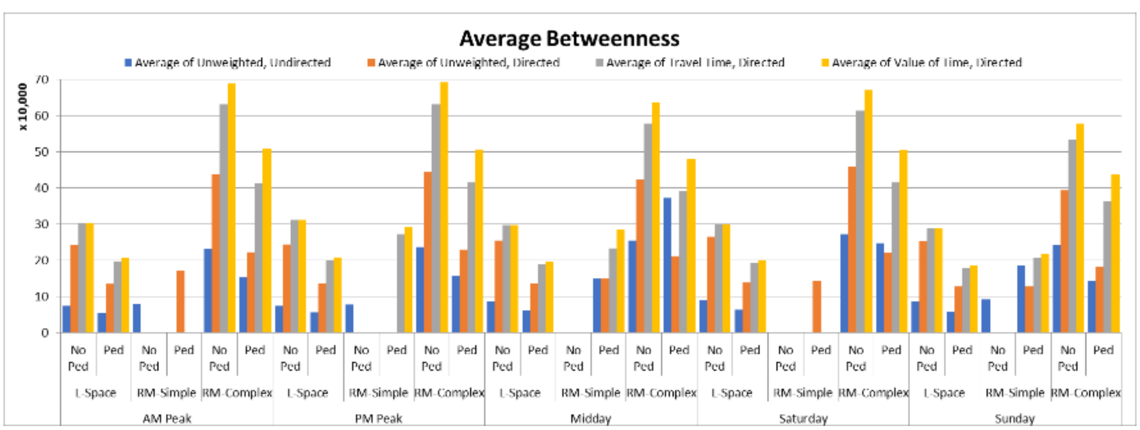


Figure 4.9 Average Betweenness

Table 4.2 Local Clustering Distributions

Graph Type	Time Period	Ped/No Ped	Clustering- Directed									Clustering- Undirected								
			Unique Values	Power		Exponent		Logarithmic		Linear		Unique Values	Power		Exponent		Logarithmic		Linear	
				Formula	R^2	Formula	R^2	Formula	R^2	Formula	R^2		Formula	R^2	Formula	R^2	Formula	R^2	Formula	R^2
L-Space	AM Peak	No Ped	10	20.26x^.804	0.983	4.04e^(1.403x)	0.006	**3.92ln(x)+14.50**	**1.000**	-2,781.98x+854.25	0.051	18	8.73x^.487	0.820	3.47e^(.517x)	0.004	**2.42ln(x)+11.29**	**1.000**	-855.28x+457.47	0.024
		Ped	189	4.61x^.068	0.050	5.57e^(-.399x)	0.002	**-0.61ln(x)+26.81**	**0.201**	-56.42x+60.82	0.011	433	4.53x^.199	0.033	4.38e^(-.146x)	0.001	**1.33ln(x)+13.38**	**0.237**	-12.41x+20.65	0.005
	PM Peak	No Ped	8	29.92x^.949	0.989	4.93e^(1.618x)	0.008	**5.19ln(x)+18.88**	**1.000**	-3,260.10x+1,095.79	0.068	19	14.50x^.719	0.842	2.60e^(1.579x)	0.027	**3.98ln(x)+15.64**	**1.000**	-810.97x+416.98	0.019
		Ped	193	4.72x^.145	0.049	5.18e^(-.327x)	0.002	**0.17ln(x)+27.31**	**0.194**	-51.02x+57.82	0.010	450	4.24x^.154	0.031	4.48e^(-.246x)	0.002	**0.94ln(x)+12.84**	**0.234**	-12.61x+20.52	0.006
	Midday	No Ped	6	20.40x^.795	0.993	10.57e^(-.723x)	0.002	**3.88ln(x)+14.67**	**1.000**	-3,948.61x+1,435.38	0.114	11	23.50x^.659	0.920	6.61e^(.787x)	0.009	**3.83ln(x)+18.03**	**1.000**	-1,026.84x+668.44	0.034
		Ped	183	4.03x^-.114	0.050	7.14e^(-.849x)	0.012	**-3.62ln(x)+23.80**	**0.213**	-59.27x+62.39	0.015	426	3.90x^.002	0.030	5.30e^(-.506x)	0.009	**0.11ln(x)+11.98**	**0.244**	-14.43x+21.37	0.008
	Sunday	No Ped	6	17.94x^.770	0.978	10.25e^(-.950x)	0.022	**3.24ln(x)+12.44**	**1.000**	-3,796.11x+1,378.59	0.175	12	20.95x^.708	0.946	4.65e^(1.197x)	0.000	**3.79ln(x)+16.48**	**1.000**	-919.50x+582.80	0.053
		Ped	178	4.35x^-.009	0.055	6.92e^(-.787x)	0.005	**-2.85ln(x)+24.33**	**0.194**	-53.19x+58.71	0.016	417	3.87x^.015	0.031	5.03e^(-.442x)	0.008	**0.10ln(x)+11.90**	**0.167**	-12.84x+20.19	0.007
	Saturday	No Ped	5	11.13x^.719	0.997	15.23e^(-2.572x)	0.003	**3.02ln(x)+10.78**	**1.000**	-4,329.53x+1,615.62	0.114	9	19.39x^.756	0.927	7.36e^(.145x)	0.019	**3.71ln(x)+14.82**	**1.000**	-1,128.70x+749.59	0.030
		Ped	159	4.56x^.042	0.048	6.13e^(-.547x)	0.011	**-3.33ln(x)+23.94**	**0.173**	-63.83x+64.33	0.013	394	3.48x^-.012	0.028	4.71e^(-.477x)	0.006	**-0.41ln(x)+11.03**	**0.178**	-13.77x+20.20	0.006
RM-Simple	AM Peak	No Ped	25	1.23x^.025	0.953	2.13e^(-.827x)	0.031	**0.07ln(x)+1.43**	**1.000**	-553.72x+409.49	0.031	57	1.39x^.021	0.856	1.59e^(-.146x)	0.007	**0.04ln(x)+1.59**	**1.000**	-60.70x+118.96	0.003
		Ped	324	3.98x^.247	0.061	3.13e^(.021x)	0.000	**4.30ln(x)+20.17**	**0.277**	-9.38x+22.44	0.001	569	4.12x^.186	0.036	3.95e^(-.179x)	0.002	**2.27ln(x)+11.31**	**0.321**	-4.94x+12.91	0.002
	PM Peak	No Ped	25	1.26x^.050	0.943	2.00e^(-.498x)	0.024	**0.13ln(x)+1.54**	**1.000**	-331.00x+373.43	0.024	55	1.35x^-.008	0.820	1.82e^(-.485x)	0.023	**-0.02ln(x)+1.60**	**1.000**	-189.74x+157.27	0.009
		Ped	330	3.89x^.248	0.059	3.04e^(.036x)	0.000	**4.33ln(x)+20.16**	**0.263**	-9.48x+22.40	0.001	602	4.07x^.235	0.041	3.56e^(-.084x)	0.000	**2.33ln(x)+11.00**	**0.314**	-4.21x+12.10	0.002
	Midday	No Ped	17	1.42x^.074	0.944	3.47e^(-1.703x)	0.060	**0.19ln(x)+1.83**	**1.000**	-1,169.67x+683.87	0.070	41	1.66x^.061	0.807	1.85e^(-.094x)	0.007	**0.14ln(x)+2.21**	**1.000**	-35.55x+148.73	0.003
		Ped	307	3.69x^.183	0.054	3.41e^(-.156x)	0.001	**3.56ln(x)+18.99**	**0.280**	-13.10x+24.16	0.002	566	3.68x^.154	0.034	3.73e^(-.224x)	0.003	**1.90ln(x)+10.39**	**0.316**	-5.50x+12.70	0.003
	Sunday	No Ped	17	1.36x^.073	0.955	3.03e^(-1.529x)	0.101	**0.15ln(x)+1.58**	**1.000**	-1,024.82x+605.82	0.122	37	1.56x^.017	0.871	2.00e^(-.155x)	0.012	**0.08ln(x)+2.03**	**1.000**	-56.87x+165.31	0.024
		Ped	290	4.33x^.284	0.063	3.11e^(.150x)	0.000	**4.43ln(x)+20.48**	**0.255**	-9.30x+22.49	0.003	551	3.90x^.221	0.035	3.47e^(-.084x)	0.001	**2.24ln(x)+10.67**	**0.213**	-4.14x+11.78	0.003
	Saturday	No Ped	11	1.51x^.111	0.967	6.33e^(-2.474x)	0.050	**0.29ln(x)+2.02**	**1.000**	-1,526.23x+956.98	0.060	25	2.05x^.124	0.833	2.50e^(-.588x)	0.012	**0.41ln(x)+2.96**	**1.000**	-452.58x+299.68	0.005
		Ped	231	4.38x^.231	0.065	3.72e^(-.054x)	0.001	**3.84ln(x)+21.56**	**0.248**	-17.37x+29.01	0.001	483	3.62x^.164	0.038	3.56e^(-.164x)	0.000	**1.81ln(x)+10.50**	**0.244**	-5.82x+13.04	0.002
RM-Complex	AM Peak	No Ped	169	2.61x^-1.171	0.248	55.15e^(-3.906x)	0.224	**-175.02ln(x)+-67.68**	**0.932**	-1,215.93x+808.67	0.095	338	2.54x^-.723	0.200	6.86e^(-.843x)	0.085	**-47.53ln(x)+7.33**	**0.838**	-86.87x+118.07	0.010
		Ped	566	6.29x^-.393	0.071	18.89e^(-1.523x)	0.053	**-24.15ln(x)+.09**	**0.988**	-226.28x+148.78	0.024	833	4.62x^-.684	0.137	17.51e^(-1.388x)	0.191	**-1.84ln(x)+13.76**	**0.808**	-17.30x+31.75	0.002
	PM Peak	No Ped	153	2.87x^-1.096	0.222	48.14e^(-3.579x)	0.178	**-183.49ln(x)+-68.37**	**0.933**	-1,263.01x+846.66	0.098	325	3.07x^-.524	0.146	6.64e^(-.651x)	0.053	**-43.27ln(x)+9.30**	**0.834**	-79.31x+113.58	0.009
		Ped	592	5.97x^-.516	0.088	22.01e^(-1.729x)	0.077	**-25.22ln(x)+-.28**	**0.988**	-213.84x+144.07	0.024	863	4.50x^-.748	0.153	18.24e^(-1.429x)	0.205	**-2.43ln(x)+13.47**	**0.804**	-17.61x+32.05	0.002
	Midday	No Ped	128	2.37x^-.925	0.231	34.21e^(-3.646x)	0.186	**-119.49ln(x)+-47.52**	**0.952**	-1,154.67x+722.70	0.082	286	2.65x^-.471	0.156	5.76e^(-.665x)	0.053	**-28.39ln(x)+10.29**	**0.896**	-83.56x+114.06	0.009
		Ped	555	5.18x^-.443	0.083	18.49e^(-1.795x)	0.076	**-20.72ln(x)+-.04**	**0.988**	-211.81x+136.40	0.023	820	4.02x^-.694	0.147	15.53e^(-1.416x)	0.190	**-2.08ln(x)+11.71**	**0.876**	-21.14x+32.52	0.003
	Sunday	No Ped	133	2.13x^-1.053	0.319	33.82e^(-3.736x)	0.206	**-119.57ln(x)+-55.59**	**0.967**	-1,140.92x+680.06	0.093	312	2.64x^-.528	0.180	5.84e^(-.643x)	0.052	**-25.85ln(x)+8.10**	**0.939**	-71.37x+96.25	0.009
		Ped	567	4.92x^-.451	0.128	17.73e^(-1.825x)	0.107	**-19.33ln(x)+-.20**	**0.985**	-196.68x+124.54	0.029	827	3.95x^-.707	0.172	15.04e^(-1.400x)	0.168	**-2.47ln(x)+11.31**	**0.923**	-21.18x+31.73	0.004
	Saturday	No Ped	99	1.75x^-1.053	0.251	32.84e^(-4.217x)	0.184	**-104.39ln(x)+-58.05**	**0.953**	-1,434.60x+771.02	0.079	196	2.80x^-.428	0.160	6.12e^(-.659x)	0.048	**-21.88ln(x)+11.19**	**0.909**	-91.35x+123.50	0.008
		Ped	485	3.12x^-.723	0.086	18.22e^(-2.422x)	0.081	**-25.77ln(x)+-7.41**	**0.987**	-251.59x+145.54	0.022	705	3.48x^-.743	0.146	13.55e^(-1.437x)	0.175	**-3.15ln(x)+9.75**	**0.890**	-26.40x+33.54	0.003
Trip Map	AM Peak	No Ped	30	1.00x^-.011	0.994	1.68e^(-.775x)	0.026	**-0.02ln(x)+1.00**	**1.000**	-485.81x+323.35	0.023	74	1.00x^-.043	0.942	1.27e^(-.065x)	0.004	**-0.06ln(x)+1.01**	**1.000**	-25.01x+87.61	0.002
		Ped	708	1.66x^-.681	0.471	6.70e^(-1.141x)	0.314	**-3.92ln(x)+2.73**	**0.908**	-7.83x+12.38	0.026	870	1.77x^-.696	0.465	7.41e^(-1.297x)	0.406	**-3.47ln(x)+2.76**	**0.877**	-7.07x+10.69	0.058
	PM Peak	No Ped	33	1.00x^.000	**1.000**	1.39e^(-.157x)	0.008	**0.00ln(x)+1.00**	**1.000**	-106.38x+223.91	0.008	77	1.00x^-.042	0.923	1.25e^(-.031x)	0.002	**-0.08ln(x)+.97**	**1.000**	-12.08x+81.16	0.001
		Ped	728	1.60x^-.703	0.501	5.73e^(-.844x)	0.256	**-4.04ln(x)+2.70**	**0.907**	-5.73x+11.45	0.023	891	1.87x^-.627	0.425	5.44e^(-.714x)	0.263	**-3.17ln(x)+3.10**	**0.870**	-3.84x+9.09	0.038
	Midday	No Ped	23	0.99x^-.035	0.988	2.46e^(-2.226x)	0.063	**-0.05ln(x)+.98**	**1.000**	-1,276.40x+506.08	0.051	60	1.02x^-.073	0.916	1.41e^(-.044x)	0.005	**-0.12ln(x)+1.04**	**1.000**	-13.80x+98.97	0.001
		Ped	672	1.40x^-.711	0.536	5.26e^(-.888x)	0.256	**-4.45ln(x)+2.14**	**0.785**	-6.48x+11.82	0.025	840	1.64x^-.666	0.478	5.58e^(-.913x)	0.308	**-3.57ln(x)+2.49**	**0.753**	-5.13x+9.59	0.050
	Sunday	No Ped	20	1.13x^.016	**1.000**	2.11e^(-1.564x)	0.081	**0.02ln(x)+1.17**	**1.000**	-1,043.55x+429.54	0.081	58	1.01x^-.049	0.960	1.33e^(-.027x)	0.014	**-0.08ln(x)+1.02**	**1.000**	-9.59x+98.60	0.012
		Ped	600	1.26x^-.778	0.559	4.53e^(-.589x)	0.233	**-6.05ln(x)+.87**	**0.851**	-4.73x+11.84	0.027	807	1.40x^-.738	0.530	4.83e^(-.753x)	0.344	**-4.67ln(x)+1.44**	**0.820**	-4.53x+9.50	0.068
	Saturday	No Ped	17	1.00x^.000	0.988	3.14e^(-2.304x)	0.028	**0.00ln(x)+1.00**	**1.000**	-1,289.36x+640.91	0.033	44	1.11x^-.006	0.949	1.51e^(-.476x)	0.004	**-0.01ln(x)+1.14**	**1.000**	-243.31x+162.46	0.002
		Ped	629	1.36x^-.696	0.656	4.88e^(-.791x)	0.188	**-4.26ln(x)+1.64**	**0.842**	-5.47x+10.62	0.028	801	1.48x^-.678	0.616	5.42e^(-.981x)	0.276	**-3.48ln(x)+1.98**	**0.794**	-5.15x+9.08	0.062
Time Expanded	AM Peak	No Ped	1	Too Few Values		Too Few Values		Too Few Values		Too Few Values		1	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	143	**8,549.98x^1.221**	**0.169**	96.62e^(12.589x)	0.074	5,107.78ln(x)+23,511.39	0.074	58,108.57x+4,864.60	0.012	505	155.08x^.321	0.018	78.05e^(-.972x)	0.001	**1,141.58ln(x)+5,200.85**	**0.114**	5,970.48x+1,896.74	0.002
	PM Peak	No Ped	1	Too Few Values		Too Few Values		Too Few Values		Too Few Values		1	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	147	**8,129.29x^1.187**	**0.166**	105.21e^(12.182x)	0.073	5,013.68ln(x)+23,075.31	0.079	55,928.57x+4,914.61	0.011	497	115.61x^.156	0.014	96.13e^(-1.791x)	0.003	**1,089.32ln(x)+5,123.32**	**0.125**	5,556.84x+2,020.71	0.001
	Midday	No Ped	1	Too Few Values		Too Few Values		Too Few Values		Too Few Values		1	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	133	**13,395.03x^1.239**	**0.148**	154.70e^(11.847x)	0.059	7,635.49ln(x)+36,551.75	0.045	88,062.01x+8,500.76	0.011	469	125.30x^.053	0.011	141.26e^(-2.242x)	0.004	**1,941.74ln(x)+8,831.41**	**0.071**	11,209.17x+3,057.13	0.003
	Sunday	No Ped	1	Too Few Values		Too Few Values		Too Few Values		Too Few Values		1	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	116	**15,617.05x^.948**	**0.094**	547.96e^(8.732x)	0.045	16,767.58ln(x)+79,659.62	0.027	166,981.79x+20,070.41	0.011	421	266.44x^.040	0.018	308.96e^(-2.204x)	0.011	**4,149.96ln(x)+18,851.83**	**0.033**	23,234.19x+6,531.75	0.002
	Saturday	No Ped	1	Too Few Values		Too Few Values		Too Few Values		Too Few Values		1	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	17	**17,043.93x^.954**	**0.089**	608.10e^(8.778x)	0.034	16,065.76ln(x)+71,991.20	0.034	145,150.64x+16,217.27	0.010	379	65.93x^-.480	0.009	328.58e^(-3.915x)	0.003	**3,248.27ln(x)+14,918.35**	**0.054**	16,855.80x+5,496.85	0.003

As with the global properties, the distributions of node properties demonstrated that the choice of graph type resulted in fundamental changes to the values of properties and how they are distributed throughout the graph. These impacts were further investigated by considering how small-world and scale-free properties varied by graph type.

4.1.1.3 Small-world and Scale-free

Finally, the impacts of graph type on the presence both small-world and scale-free characteristics were investigated by identifying the properties that define them, GCC and APL for Small-worlds and NDD for scale-free properties. When investigated for small-world characteristics, all graph types chosen had very similar results with each having both GCC and APL several times larger than the random graphs, as shown in Figures 4.10 and 4.11. This indicates that they do not meet the criteria for small-worlds, which is in agreement with (Dimitrov and Ceder 2016) and is likely due to including all stops, many of which are very distant from useful transfer locations. However, the ratio between the actual and random networks did appear to vary between graph types, and so APL was investigated further as shown in Figure 4.12 below.

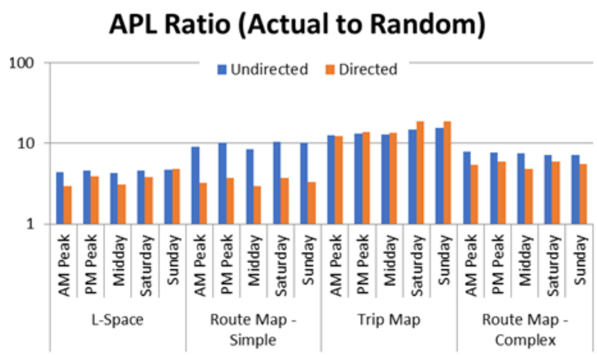


Figure 4.10 APL Ratios

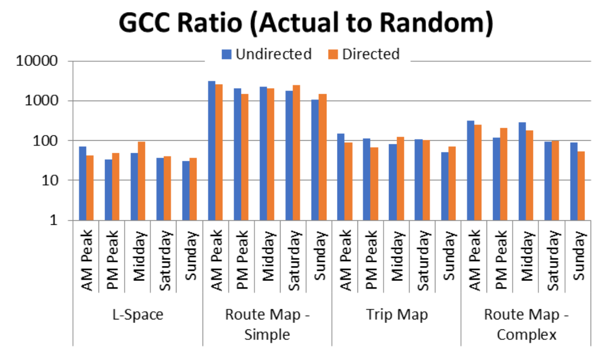


Figure 4.11 GCC Ratios

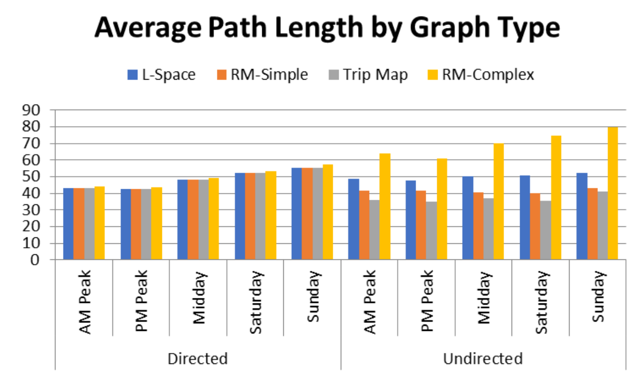


Figure 4.12 APL by Graph Type

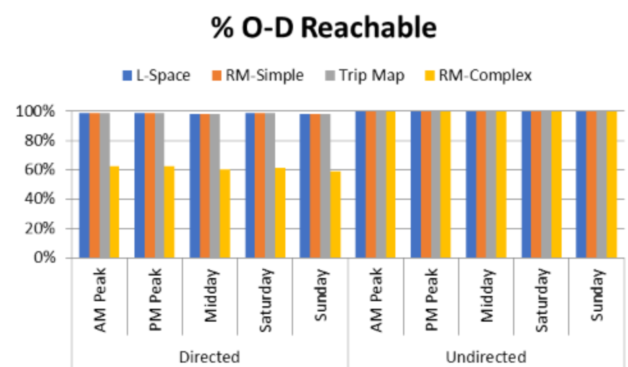


Figure 4.13 Reachable O-D Pairs

While all graphs had a very similar APL when directed links were used, a very different pattern emerged when undirected links were used. L-Space showed slightly higher travel times than RM-Simple, and Trip Map had the shortest travel times, while RM-Complex was much higher than any of the others. Furthermore, the RM-Complex showed higher APL when undirected links were used, as did the L-Space in weekday time periods. This trend was found to be caused by the relative travel time to between O-D pairs that were not reachable when directed links were used. As shown in Figure 4.12, the RM-Complex graphs had far more O-D pairs which were only reachable when undirected links were used, many of which required circuitous routing.

The impact of graph type on the ratio of GCC between the actual and random graphs was less profound, however as shown in Figure 4.14, the impact on the actual value of the GCC was substantial with RM-Complex having values two orders of magnitude higher than the others three types. GCC also varied between the other three, although the pattern changed between time periods.

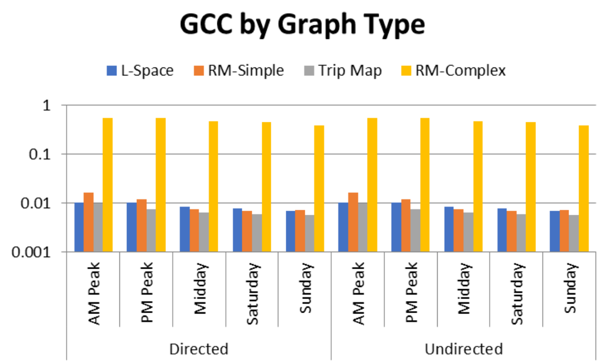


Figure 4.14 GCC by Graph Type

As shown in Figures 4.15 and 4.16, the graph type also was found to play a significant role in both the how closely the NDD followed a power law and the scale factor of the resulting power law. As shown in Table 4.3, despite not being small-worlds, the Route Map-Complex and Trip Map graphs might be considered to have scale-free characteristics as the power law provided the best fit, although the exponential distribution had only a slightly lower R^2 value. For the T-Space and RM-Simple graphs, the exponential curve fits noticeably better and the Time-Expanded graphs did not fit well to any of the curves. The scale factors were also impacted by the graph type with L-Space graph and RM-Complex with L-Space being higher in weekday time periods but lower in the weekend time periods. RM-Simple was consistently the third highest with Trip Map having the smallest values.

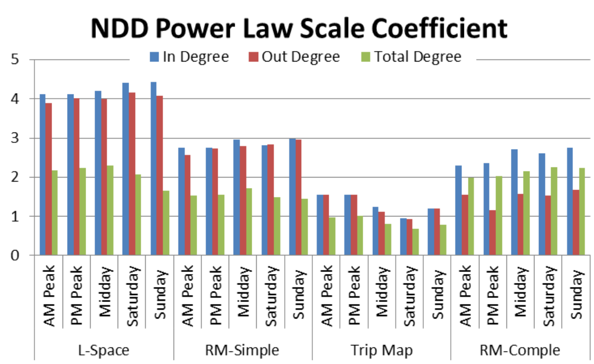


Figure 4.15 NDD Scale Coefficient

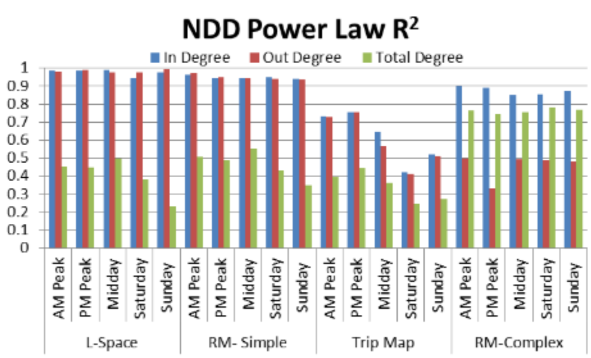


Figure 4.16 NDD Power Law R^2

Despite not being small-worlds, the Route Map-Complex and Trip Map graphs might be considered to have scale-free characteristics as the power law provided the best fit. However, the exponential distribution had only a slightly lower R^2 value, and when pedestrian links were not included, the R^2 was less than 0.4, indicating a poor fit.

4.1.1.4 Summary

As was demonstrated, nearly every test that was performed showed a noticeable impact from the choice of graph type. Global indicators such as β fluctuated dramatically while α and γ were not even defined for two graph types. While node degree and local clustering continued to follow similar distributions, betweenness, the coefficients of those distributions were impacted. Metrics related travelling through the graph were inconsistent with betweenness and the number of destinations reachable from each origin showing a clear impact while APL was only impacted significantly when combined with undirected links. Finally, the impact to small-world and scale-free properties was unclear as all failed to be small-worlds and scale-free properties were found in all but those with complex nodes.

Within the variation between graph types, there are also similarities between the graphs which share simple stops (L-Space, RM-Simple and Trip Map). The RM-Complex graphs, which use complex stops, are noticeably separated from the others in number of nodes, tests related to clustering, both local and global, as well as tests related to travel through the network, betweenness, APL, O-D Reachable. However, there is also a clear interplay between the graph type and the edge type. This is most obvious in the betweenness values, but also appears in several other metrics, and this is the subject of the next section.

Table 4.3 Node Degree Distributions

Graph Type	Time Period	Ped/No Ped	Node Degree								
			Unique Values	Power		Exponent		Logarithmic		Linear	
				Formula	R^2	Formula	R^2	Formula	R^2	Formula	R^2
L-Space	AM Peak	No Ped	15	617.07x^-2.171	0.451	**282.21e^(-.394x)**	**0.564**	-699.33ln(x)+1,715.05	0.209	-102.54x+1,257.76	0.171
		Ped	38	1,873.50x^-1.538	0.364	**701.89e^(-.161x)**	**0.719**	-137.28ln(x)+527.38	0.289	-12.52x+403.43	0.435
	PM Peak	No Ped	15	689.12x^-2.235	0.448	**362.83e^(-.439x)**	**0.567**	-741.66ln(x)+1,774.79	0.217	-121.24x+1,365.22	0.189
		Ped	38	31,110.63x^-2.479	0.703	**1,349.71e^(-.184x)**	**0.911**	-206.80ln(x)+738.75	0.496	-13.84x+445.93	0.524
	Midday	No Ped	14	780.16x^-2.303	0.497	**326.44e^(-.422x)**	**0.595**	-677.46ln(x)+1,631.86	0.205	-102.72x+1,206.19	0.168
		Ped	38	2,143.28x^-1.604	0.398	**727.74e^(-.165x)**	**0.753**	-130.78ln(x)+499.95	0.301	-11.94x+381.50	0.448
	Sunday	No Ped	14	408.82x^-2.075	0.233	**238.07e^(-.411x)**	**0.397**	-655.19ln(x)+1,577.00	0.192	-102.88x+1,191.74	0.181
		Ped	36	25,951.29x^-2.464	0.696	**1,229.18e^(-.189x)**	**0.889**	-186.03ln(x)+662.17	0.446	-12.98x+406.30	0.478
	Saturday	No Ped	12	197.12x^-1.651	0.380	**213.65e^(-.425x)**	**0.514**	-653.98ln(x)+1,499.65	0.205	-125.23x+1,241.38	0.173
		Ped	35	20,918.59x^-2.430	0.679	**1,083.06e^(-.192x)**	**0.887**	-179.86ln(x)+629.19	0.460	-13.01x+387.13	0.498
RM-Simple	AM Peak	No Ped	36	**546.61x^-1.531**	**0.508**	58.07e^(-.088x)	0.500	-297.21ln(x)+985.97	0.211	-11.16x+417.19	0.089
		Ped	58	3,628.14x^-1.766	0.529	394.21e^(-.109x)	**0.789**	-119.47ln(x)+474.19	0.395	-6.35x+293.18	0.437
	PM Peak	No Ped	36	**545.53x^-1.549**	**0.490**	56.90e^(-.090x)	0.468	-303.23ln(x)+1,003.57	0.215	-11.97x+434.48	0.095
		Ped	58	33,296.98x^-2.420	0.782	520.94e^(-.113x)	**0.872**	-157.93ln(x)+607.98	0.560	-6.44x+306.15	0.472
	Midday	No Ped	32	**690.69x^-1.717**	**0.553**	66.23e^(-.109x)	0.523	-308.34ln(x)+995.39	0.207	-13.74x+453.59	0.096
		Ped	52	3,935.04x^-1.817	0.516	**522.68e^(-.128x)**	**0.807**	-118.01ln(x)+461.52	0.401	-7.22x+300.63	0.472
	Sunday	No Ped	28	408.72x^-1.486	0.351	**65.91e^(-.104x)**	**0.378**	-298.93ln(x)+962.49	0.202	-13.82x+457.46	0.102
		Ped	49	29,349.62x^-2.432	0.732	**727.03e^(-.139x)**	**0.905**	-153.11ln(x)+573.35	0.518	-7.72x+313.32	0.510
	Saturday	No Ped	26	224.54x^-1.449	0.428	**44.63e^(-.118x)**	**0.471**	-313.57ln(x)+958.85	0.209	-17.47x+475.06	0.101
		Ped	43	23,966.35x^-2.422	0.740	**838.38e^(-.160x)**	**0.891**	-156.58ln(x)+567.95	0.550	-9.21x+325.41	0.516
RM-Complex	AM Peak	No Ped	61	**15,974.64x^-1.992**	**0.763**	664.49e^(-.100x)	0.763	-559.65ln(x)+2,035.72	0.463	-20.33x+899.37	0.242
		Ped	203	**18,207.97x^-1.649**	**0.808**	117.64e^(-.019x)	0.668	-173.26ln(x)+844.87	0.387	-1.16x+218.09	0.110
	PM Peak	No Ped	60	**17,433.92x^-2.022**	**0.748**	684.21e^(-.102x)	0.726	-576.36ln(x)+2,089.31	0.480	-21.51x+934.62	0.255
		Ped	227	**22,252.82x^-1.703**	**0.824**	106.77e^(-.018x)	0.681	-164.60ln(x)+818.34	0.386	-1.02x+208.08	0.111
	Midday	No Ped	48	**16,816.83x^-2.146**	**0.754**	709.17e^(-.125x)	0.726	-596.76ln(x)+2,065.00	0.469	-25.75x+958.57	0.250
		Ped	143	**18,540.19x^-1.698**	**0.797**	211.37e^(-.029x)	0.748	-206.16ln(x)+940.10	0.400	-2.06x+273.23	0.126
	Sunday	No Ped	55	**22,093.07x^-2.265**	**0.768**	631.97e^(-.119x)	0.684	-539.58ln(x)+1,918.84	0.458	-21.28x+870.05	0.229
		Ped	168	31,093.09x^-1.898	0.811	**230.95e^(-.034x)**	**0.853**	-196.90ln(x)+903.49	0.425	-2.13x+270.08	0.163
	Saturday	No Ped	42	**13,657.45x^-2.228**	**0.783**	519.02e^(-.134x)	0.736	-561.19ln(x)+1,871.23	0.491	-25.33x+856.31	0.261
		Ped	116	**17,929.05x^-1.829**	**0.824**	342.31e^(-.050x)	0.812	-230.81ln(x)+978.67	0.430	-3.80x+330.22	0.152
Trip Map	AM Peak	No Ped	196	**349.84x^-.966**	**0.396**	14.11e^(-.008x)	0.344	-32.62ln(x)+173.48	0.150	-0.21x+56.12	0.073
		Ped	211	**2,536.03x^-1.328**	**0.593**	24.73e^(-.010x)	0.485	-31.95ln(x)+171.68	0.351	-0.20x+54.01	0.187
	PM Peak	No Ped	202	**461.34x^-1.006**	**0.446**	15.64e^(-.008x)	0.380	-32.09ln(x)+171.79	0.154	-0.20x+55.27	0.074
		Ped	225	**2,262.34x^-1.303**	**0.606**	24.71e^(-.010x)	0.503	-28.96ln(x)+157.86	0.340	-0.19x+52.22	0.195
	Midday	No Ped	221	**175.18x^-.814**	**0.363**	10.00e^(-.006x)	0.308	-27.00ln(x)+148.73	0.041	-0.15x+46.90	0.020
		Ped	252	**772.90x^-1.063**	**0.544**	19.45e^(-.008x)	0.500	-20.34ln(x)+115.92	0.146	-0.13x+41.93	0.097
	Sunday	No Ped	273	**115.37x^-.689**	**0.273**	7.57e^(-.003x)	0.243	-18.12ln(x)+109.15	0.082	-0.07x+33.70	0.041
		Ped	332	**375.86x^-.878**	**0.464**	11.95e^(-.004x)	0.443	-13.50ln(x)+82.81	0.213	-0.05x+27.54	0.134
	Saturday	No Ped	204	**155.46x^-.790**	**0.246**	9.06e^(-.005x)	0.213	-24.86ln(x)+138.86	0.061	-0.12x+42.79	0.033
		Ped	251	**669.89x^-1.030**	**0.432**	17.15e^(-.007x)	0.396	-18.12ln(x)+103.32	0.114	-0.11x+35.15	0.074
Time Expanded	AM Peak	No Ped	3	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	48	159,200.40x^-1.493	0.204	5,042.27e^(-.035x)	0.106	**-25,603.10ln(x)+102,327.23**	**0.296**	-570.70x+42,329.55	0.139
	PM Peak	No Ped	3	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	48	156,449.15x^-1.448	0.212	5,471.51e^(-.034x)	0.109	**-25,922.70ln(x)+103,744.55**	**0.298**	-579.09x+43,032.78	0.141
	Midday	No Ped	3	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	47	**3,586,907.19x^-2.336**	**0.412**	4,770.06e^(-.015x)	0.050	-50,456.65ln(x)+195,434.79	0.399	-393.34x+54,595.68	0.073
	Sunday	No Ped	3	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	47	226,906.15x^-1.316	0.108	5,196.74e^(-.004x)	0.013	**-64,493.85ln(x)+272,151.29**	**0.208**	-240.11x+89,452.99	0.032
	Saturday	No Ped	3	Too Few Values		Too Few Values		Too Few Values		Too Few Values	
		Ped	43	86,067.57x^-.931	0.152	6,694.68e^(-.004x)	0.010	**-48,225.76ln(x)+204,185.00**	**0.240**	-231.33x+72,688.27	0.030

4.1.2 Impact of Edge Type

While the graph type defines the interpretation of the nodes and edges, the structure of the network is also dependent on the directionality and weighting of the edges. In order to investigate this, four types of edges were used for each of the graph types and for each time period. The most commonly seen edge types in the literature are unweighted, undirected (UU) and unweighted directed (UD). While weighted edges can be undirected, these are seldom seen in the literature, and so the two weighting methods used, Travel Time (TT) and Value of Time (VoT), were only constructed as directed.

Unfortunately, most of the metrics commonly used in the literature fail to take into account the edge type or only consider directionality. Global properties only consider quantities of edges and nodes, while community identification, clustering, small-worlds, and scale-free properties are only impacted by the directionality of the edges. This leaves only Betweenness and Average Path Length which are able to demonstrate the impact of both edge weighting and directionality. Despite these limitations, a clear impact from the edge type was identified for every metric that was considered.

4.1.2.1 Global Indicators

As mentioned previously, the global properties which were considered did not reflect either the directionality or weighting of edges. This is likely due to these indicators being developed for rail systems which follow symmetric routes with consistent station spacing.

4.1.2.2 Node Property Distributions

While only the use of directed or undirected links could be investigated for node properties, edge type played an important role in both of the node property distributions investigated. For local clustering, edge weight is not considered, however, there is a clear impact from directionality. As shown in Figure 4.17, when fit to a power law the directed LCC had a lower R^2 for all cases. Figures 4.19 and 4.20 show that while there was an impact to the scale coefficient the nature of that impact varied with graph type and time period. L-Space and RM-Complex graphs saw coefficients closer to zero, while the graph types with parallel edges (RM-Simple and Trip Map) saw increases in some time periods and decreases in others. The average local clustering, shown in Figure 4.18, was between 31% and 92% lower, or between 14% and 58% reduction when pedestrian links were included. In addition, the directed and undirected clustering follow different trends between time periods for all but the RM-Complex graphs. The undirected have higher values for midday and Saturday than peaks or Sunday, while the directed tends to show the highest values in the PM peak and lower values in the off-peak times. These findings indicate that directed and undirected graphs have very different values for clustering with undirected tending to be highest when the directed values are lowest.

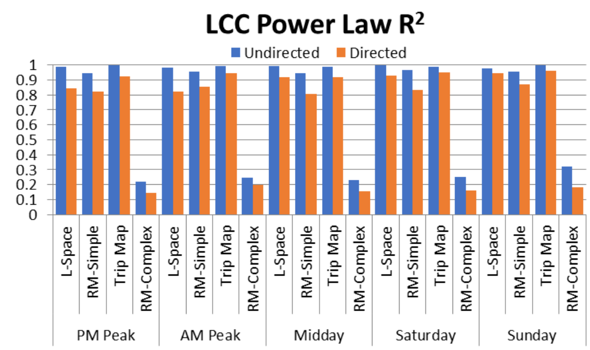


Figure 4.17 LCC Power Law R^2

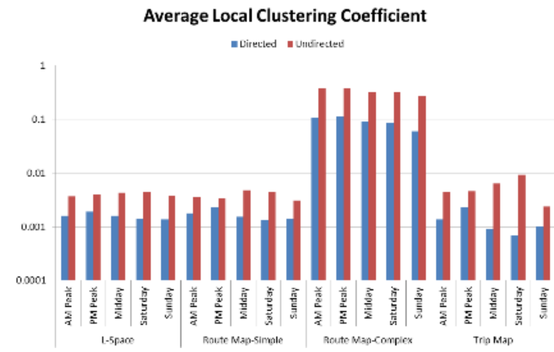


Figure 4.18 Average LCC

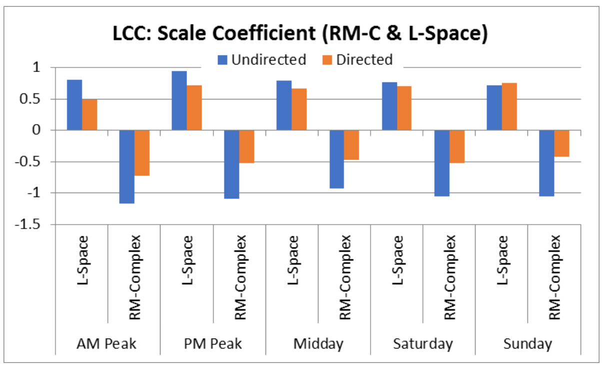


Figure 4.19 LCC Scale Coefficient RM-Complex & L-Space

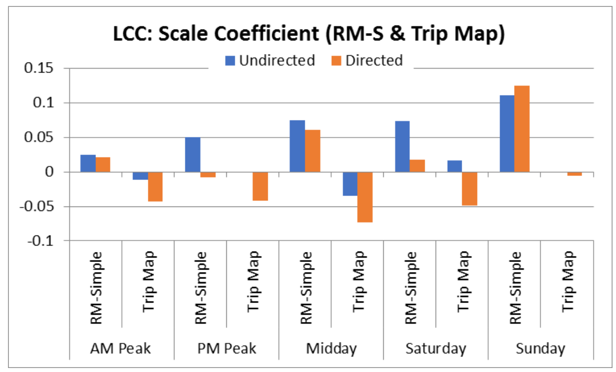


Figure 4.20 LCC Scale Coefficient RM-S & Trip Map

The impact to betweenness indicates a complex relationship with graph type and edge type, as shown in Figure 4.21. Travel Time and Value of Time tended to follow very similar trends to the Unweighted Directed links, indicating that directionality had a larger impact than weight. In general, the undirected links had much lower R^2 for all models, shown in Figure 4.22, in L-Space, an improvement for all models in RM-Simple, although the linear model showed the greatest change, while only fit to the power law curve was impacted in the RM-Complex graphs. There was also a noticeable impact on average betweenness with Value of Time and Travel Time edges tending to be higher, followed by UD and UU having the lowest average values, although there were some exceptions to this trend when pedestrian links were introduced.

The impact on both local clustering and betweenness is unmistakable, albeit inconsistent between graph types. This demonstrated that the edge type plays an important role in the fundamental structure and properties of graphs. To further investigate this, the impact to small-world and scale-free properties were considered.

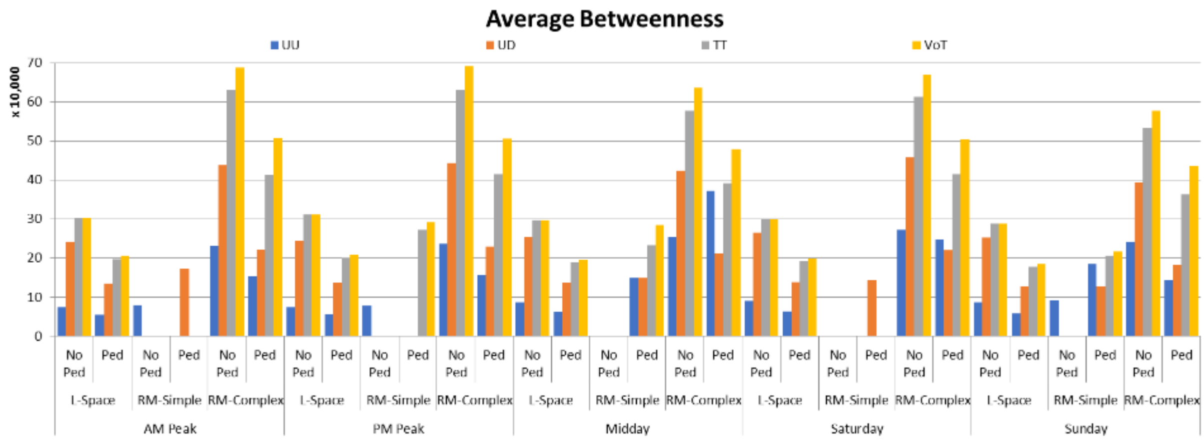


Figure 4.21 Average Betweenness by Edge type

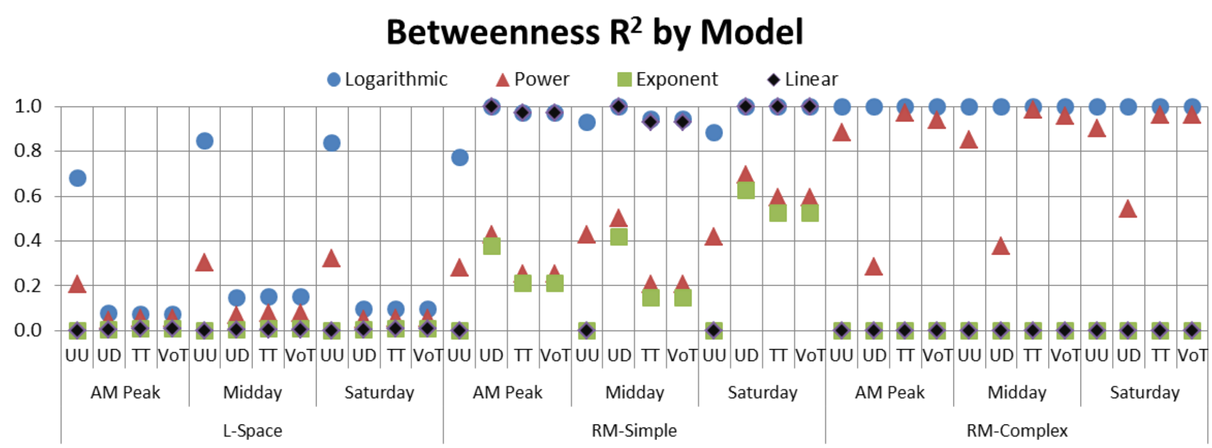


Figure 4.22 Betweenness R2 by Edge Type

4.1.2.3 Small-world and Scale-free

As with local clustering, Small-world and Scale-free characteristics only consider directionality and not the actual weights of edges. As such, GCC, APL and NDD were all calculated using both UU and UD edges. While edge type did have a noticeable impact on the relative APL and GCC between the actual and random graphs, it was neither consistent nor sufficient for any of the graphs to have small-world properties. As shown in Figure 4.14, the impact of edge type on GCC was less pronounced than that of graph type and varied with both time period and graph type. However, as is shown in Figure 4.23, directed links increased the APL ratio in the Trip Map graphs while reducing it for all others, most significantly in the RM-Simple graphs. Closer inspection of the data reveals that the APL is less dramatically impacted by edge weighting than the ratio of the actual to the random graphs. This unexpected finding indicates the edge type had a more dramatic impact on the random graphs, which was confirmed by inspection of the data.

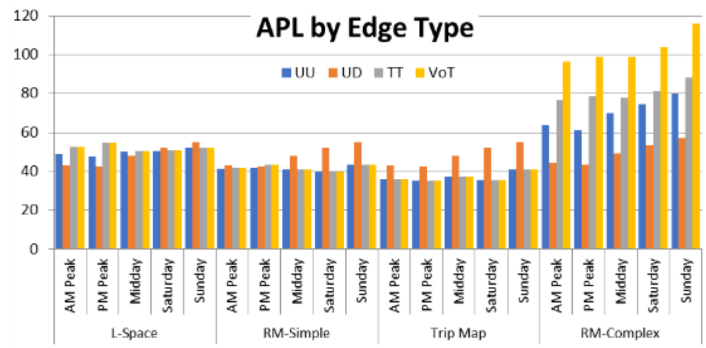


Figure 4.23 APL by Edge Type

The impact of edge weighting was further investigated by calculating APL using the TT and VoT edges, where it was found that use of the UD edges had the highest APL for most scenarios. This finding was also unexpected and shows that the impact of links with weights of zero. This event is an unavoidable consequence of the close stop spacing on the bus network, where buses are often scheduled to visit multiple stops in less than a minute. When combined with simple stops, which allow for transfers at stops with no cost, it is possible to dramatically reduce the travel cost between many locations. As a result, use of the TT and VoT links caused a reduction in the APL, compared to the UD, for many of the L-Space, RM-Simple, and Trip Map graphs, bringing it close to the values seen for when UU links were used. However, this is not seen on the RM-Complex graphs, where transferring is more accurately modelled. Instead of increasing APL, the use of UD links actually reduces the APL, due to many locations no longer being reachable. TT and VoT then increase APL due to a cost now being associated with transferring. Furthermore, a different value was found for VoT than TT, due to the transfer links which, unlike GTFS links, are given higher weights. In addition to demonstrating the impact of edge weights, this indicates the impact of correctly representing transfers, which is discussed in more detail in phase II below.

The impact of edge type on scale-free characteristics is limited to the use of directed links, which divides the total node degree into an "in" degree and an "out" degree, based on the direction of the edge. As is shown in Figures 4.10 and 4.11directed edges improved the fit of and increased the scale coefficient for all but the RM-Complex graphs. As a result, RM-Simple and L-Space graphs might be considered to have scale-free characteristics, although the scale coefficient of several of the L-Space graphs does exceed 4 which might exclude them.

4.1.2.4 Summary

The impact of edge types was investigated by constructing each of the graph types with four types of edges. The Unweighted, Undirected (UU) and Unweighted Directed (UD), were particularly useful as it allowed for inspection of edge type on metrics, such as clustering and NDD, which are affected by directed links but not by edge weights. While results are muddled by the choice of graph type, an impact from edge type is also found in each of the tests. The use of directed edges had a subtle effect on local clustering, which tended to reduce how well it fit a power-law curve while inflicting the scale factor, albeit inconsistently. Both average and distributions of betweenness showed impacts from both the edge weights and betweenness, although once again the impact followed different trends for the various graph types. While none of the graphs were found to have small-world characteristics, regardless of edge type, directed edges did tend to increase APL, while the TT and VoT weighting schemes largely offset that increase. Directed edges also showed a significant impact on NDD and such that L-Space and RM-Simple graphs now appear to show scale-free properties and the scale coefficients tended to increase.

The interplay between edge type and graph type is also apparent in every metric, with each graph type following different patterns in terms of the distribution of betweenness and LCC scale coefficient. In addition, the complex stops used in the RM-Complex graphs resulted in it being affected very differently than the others in terms of APL. As will be seen in the next section, there is an even more dramatic interplay between the inclusion of pedestrian links and both graph type and edge type.

4.1.3 Impact of Pedestrian Links

The impact of including pedestrian links cannot be overstated, as they dramatically impacted every metric that was investigated. This was the expected result of the sheer number of pedestrian links combined with those links being totally independent of the transit network itself. This introduces two challenges with using pedestrian links. The first, and most obvious, is that the dramatic increase in links increases the complexity of the network. The second is that the pedestrian links fundamentally change the network such that for some graph types, calculated metrics no longer describe the transit system so much as the network made by pedestrian connections. This is seen in each of the following sections on global properties, partition analysis, node property distributions and finally, small-world & scale-free characteristics.

4.1.3.1 Global Properties

While the number of nodes was not impacted by pedestrian links, pedestrian links accounted for more than any other type of link in all graphs other than the Trip Maps. Pedestrian links accounted for ~74% of links in the L-Space Graphs, 65-67% of links in the RM-Simple, 62-68% in the RM-Complex, 9-17% in the Trip Map and 59-60% in Time-Expanded graphs. This dramatic increase in edges led to dramatic increase in β, as well as α and γ in graphs where they were defined. As shown in Figures 4.24, 4.25 and 4.26 the impact appeared to follow the relative percentage of edges that were pedestrian links with the L-Space having the greatest impact and the Trip map having the lowest.

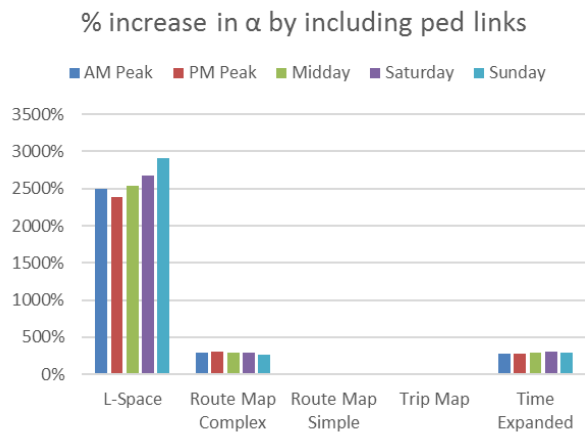


Figure 4.24 Circle Availability with Ped Links

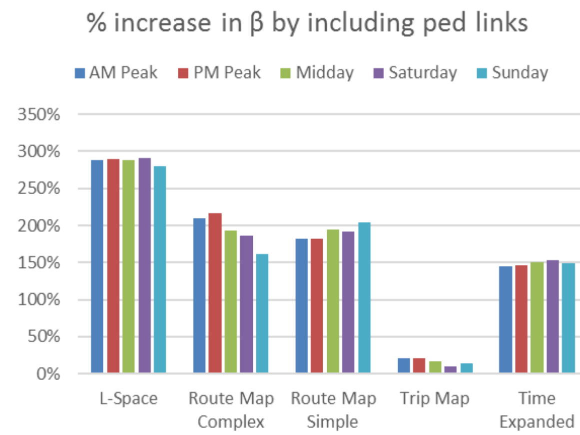


Figure 4.25 Complexity with Ped Links

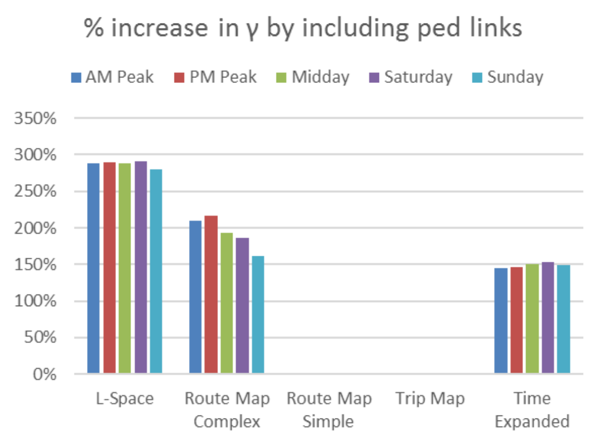


Figure 4.26 Complexity with Ped Links

4.1.3.2 Node Distributions

As is shown in Figure 4.27, pedestrian links increased both directed and undirected average local clustering by several orders of magnitude in all graphs, other than the Route Map Complex. This is likely due to the dramatic increase in links between adjacent stops, which are also connected to each other via pedestrian connections. The less dramatic impact on the RM-Complex graphs is likely the result of the complex nodes which led to much higher clustering than the other graphs before pedestrian links were included. In addition, pedestrian links had a dramatic impact on the distribution of LCC, shown in Figures 4.28 and 4.29, which no longer fit the power law curve and reduced the fit with the logarithmic curve in all except for the RM-Complex with directed links. This change indicates that the pedestrian network follows a different distribution, which is overshadowing the GTFS links.

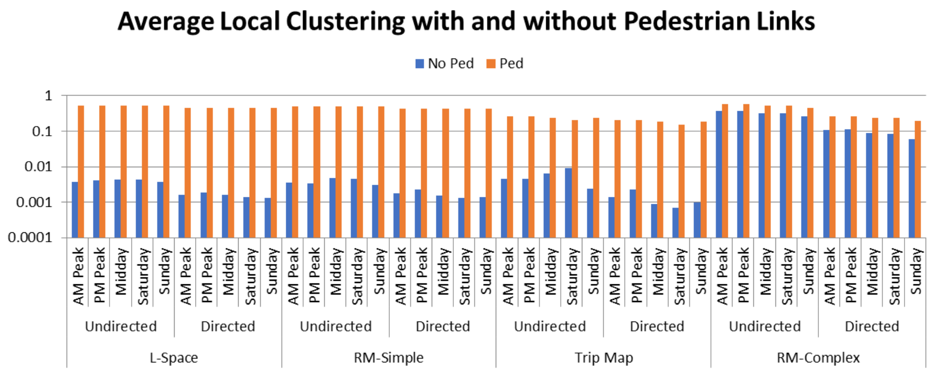


Figure 4.27 Average Local Clustering with and without Ped Links

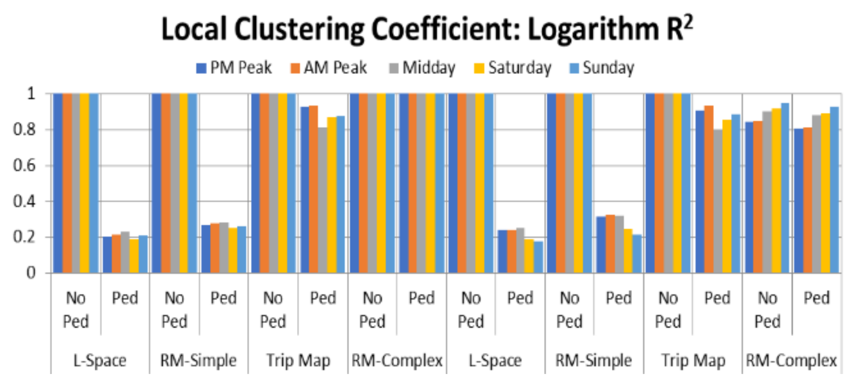


Figure 4.28 Local Clustering Log Curve R^2 with and without Ped Links

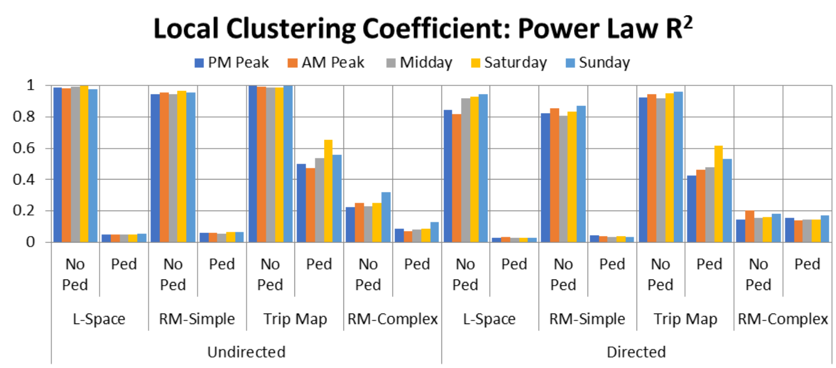


Figure 4.29 Local Clustering Power Law R^2 with and without Ped Links

A less dramatic impact was seen on the betweenness values, which saw reduced average values in all cases, shown in Figure 4.9, and inconsistent changes in the distribution, shown in Figure 4.30. These changes indicate that travel through the network is fundamentally changed, with many shortest paths being substantially reduced. While some impact was expected, the magnitude of the change for both the LCC and betweenness supports earlier assumptions regarding the importance pedestrian links play in the transit network.

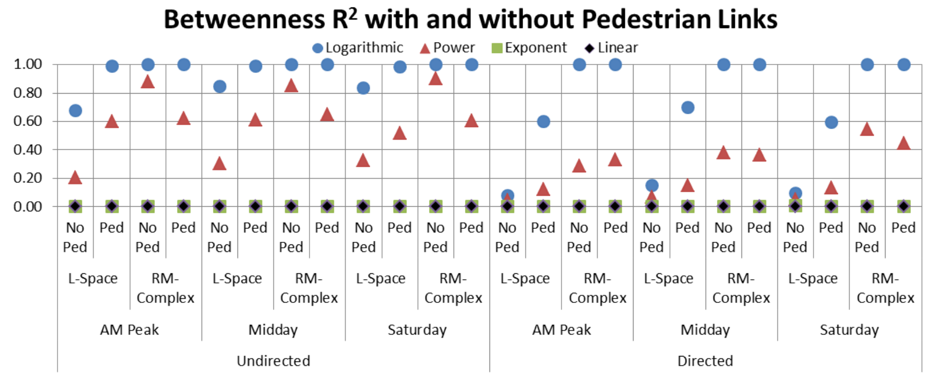


Figure 4.30 Betweenness R^2 with and without Ped Links

4.1.3.3 Small-worlds & Scale-free

Small-worlds are defined as having a global clustering coefficient (GCC) larger than a random graph and an average path length (APL) significantly shorter than a random graph. As is shown in Figures 4.31 and 4.32, including pedestrian links tended to increase the GCC by several orders of magnitude, while reducing APL by 20-50% for most graphs. This trend was identified across all weighting schemes and time periods and in most graph types, although the RM-Complex graphs saw only a modest increase in GCC. While adding the pedestrian links did move the GCC and APL toward having small-world characteristics, the additional edges had a similar impact on random graphs. The combined effect, shown in Figures 4.33 and 4.34, was only a modest change to the ratios of GCC and APL between the actual and random graphs. As a result, the inclusion of pedestrian links does not appear to improve the overall "Small-world" characteristics.

When investigating the impact on APL itself, including pedestrian links significantly reduced travel times, although the magnitude depended on the weighting scheme used. An interesting finding was that the impact was most notable when UU edges were used with TT and VoT having very similar impacts that were slightly less than that seen with UD edges. APL in UU and UD is the number of links traversed, and so this indicates that walking connections allowed for the total number of links to be reduced more than the actual travel time. This may seem contradictory, except that many GTFS links have a travel time of zero (where multiple nearby stops are visited within one minute) while pedestrian links always have travel times of at least one minute, and often much longer. This is most noticeable in UD due to many stops being only accessible from a small number of routes and only from travel in a single direction. With undirected edges, connections can be made by travelling in the opposite direction while directed edges require a more circuitous trip would be needed. The pedestrian links allow travellers to walk across the street to travel in the opposite direction.

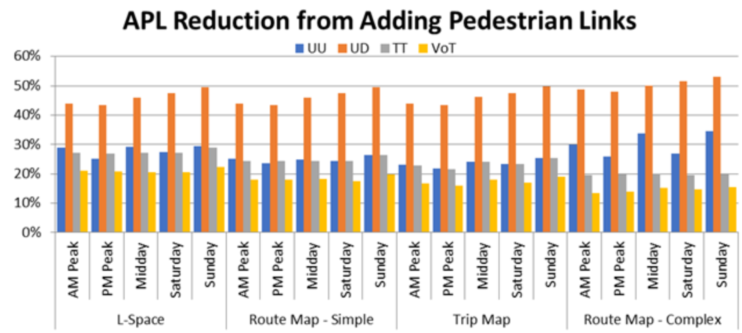


Figure 4.31 APL Reduction from Ped Links

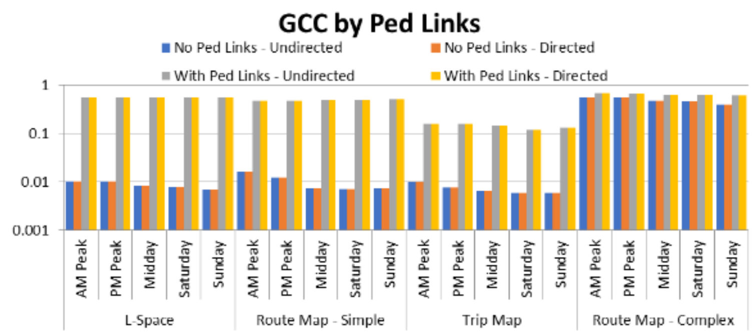


Figure 4.32 GCC Impact from Ped Links

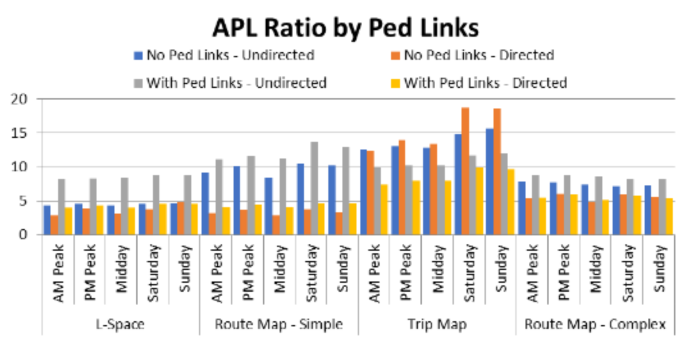


Figure 4.33 APL Ratio Impact from Ped Links

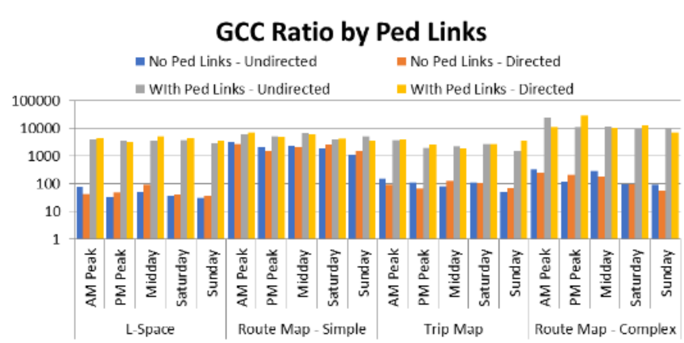


Figure 4.34 GCC Ratio Impact from Ped Links

In regard to scale-free properties, the inclusion of pedestrian links was found to consistently improve the R^2 value when fitting the NDD with a power law. As shown in Figures 4.35, 4.36 and 4.37, in all but three of the twenty graphs, the inclusion of the pedestrian links improved the R^2 value although the increase ranged from only 4% in five graphs to almost 200%. The L-Space models saw both the highest and lowest impacts followed by the Route Map- Simple models. The Trip Map graphs saw consistently higher R^2 values with improvements between 36% and 76%, while the Route Map Complex graphs saw modest improvements of only 6-11%.

This finding is likely due to the relative increase in links provided by pedestrian connections. L-Space and Route Map-Simple Graphs are much less complicated, and so including pedestrian links increases the total number of links by 279-289% and 182-204%, respectively. Trip maps have many more links and so the total increase is only 10%-21%. The Route Map complex graphs also a substantial increase in the number of links (161%-216%); however, this may have been offset using the complex nodes.

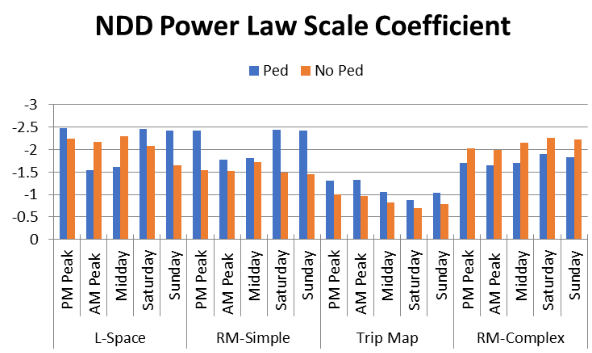


Figure 4.35 Impact on NDD Scale Coefficient from Ped Links

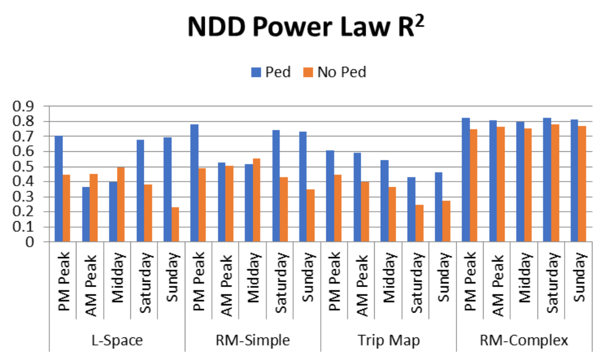


Figure 4.36 Impact to NDD Power Law R^2 from Ped Links

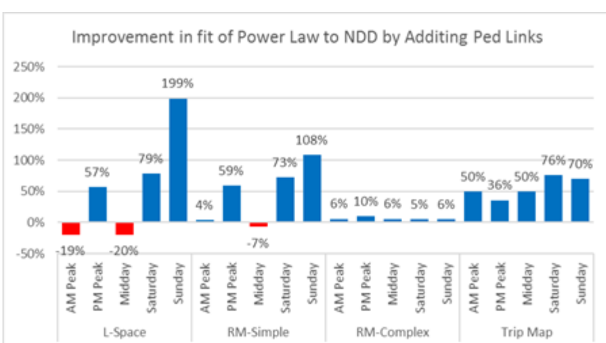


Figure 4.37 Improvement to NDD Power Law fit from Ped Links

4.1.3.4 Summary

As mentioned earlier, the impact of pedestrian links cannot be overstated as it influenced every aspect that was tested. The increase in total links showed a notable increase in the global properties α, β, and γ, relative to the proportion of additional links. The additional links also had the impact of increasing local clustering and betweenness dramatically, while reducing APL and improving how well the NDD fit a power law curve. This fundamental change from pedestrian links highlights how important they are to the operation of the transit system, but the huge increase in links also overshadows the actual transit system.

4.1.4 Phase I - Summary

From this investigation, each of the three identified choices (i.e. graph representation, pedestrian links and edge weighting) has a substantial impact on the graph structure, and the chosen indicators. While several trends are apparent, few of them hold true for all the tested graphs. For example, including pedestrian links improved how well the power law curves fit the NDD in most, but not all cases and the trip map graph had a scale factor significantly smaller than the graph representations for all time periods except for Sunday Midday. Due to the very different methods used to construct the graphs, the impact of pedestrian links and time period would be expected to vary. More surprising was the dramatic differences seen in some graphs between time periods where the service is relatively similar, such as in AM and PM peaks and between Midday and Saturday.

Another interesting finding was that while several of the graphs were found to have scale-free characteristics (NDD follows a power law), none were found to have small-world characteristics, as all had APL significantly larger than the random graphs. This contrast (Wu et al. 2004) and (Tsekeris and Souliotou 2014), however, agrees with (Dimitrov and Ceder 2016) and may be due to all stops being included instead of just transfer and terminal stops. However, (Wu et al. 2004), only included transfer stops and terminals which, along with neglecting pedestrian links, dramatically reduces the number of edges and nodes in the system and likely reduces APL considerably. In (Dimitrov and Ceder 2016), the author encounters a similar finding in Auckland, although the anomaly is assumed to be due to directional edges. Both directed and undirected edges were considered in this study and the same result was found.

While the importance of each of these three choices has been identified, there remains a question regarding which graph configuration is preferable. This is a challenging question that might best be approached by first excluding graph configurations that were problematic. When inspecting by graph type, the Time-Expanded graphs had to be excluded from many of the metrics due to size and complexity, and the many parallel edges were problematic for several global indicators, and betweenness centrality. This leaves only the L-Space graphs, which only indicate available connections and the RM-Complex graphs, which are largely dominated by transfer links between routes at stops.

When considering edge types, none were problematic, although several authors have suggested excluding UD links to better reflect that bus service is asymmetric. In metrics such as NDD and LCC, where edge weight is not used, TT and VoT functioned identically to UD, and VoT links functioned identically to TT unless transfers or pedestrian links are included. As such, the VoT edge type has the most information and so should best reflect the actual operation of the transit network.

This reduces the choices to two graph types (L-Space and RM-Complex) and a single edge type (VoT), which results in four graph configurations when combined with the option of either including or excluding pedestrian links. The choice between these four options will depend on the purpose of the study being conducted. Use of L-Space with pedestrian links might be useful to study the available connections for passengers, such as to identify poorly connected locations or to study resilience of the network to failures at specific locations or along specific corridors. L-Space without pedestrian links might be useful for similar studies from the perspective of individuals with mobility challenges or who are unfamiliar with the network. This configuration is also best able to represent items such as clustering and NDD when only the transit service itself is of concern. RM-Complex with pedestrian connections is used to study movement through the network, as transfers are represented with some degree of detail. However, this might not be appropriate for studies of the structure of the transit service as the combination of complex nodes and pedestrian links will overshadow the GTFS links. RM-Complex without pedestrian links is dominated by transfer links between the complex nodes but lacks many very common transfers between nearby stops. Once again, this might be useful when considering individuals with mobility challenges.

4.2 Phase II: Changes between time periods

4.2.1 Global Properties

As shown in Tables 4.4 and 4.5, the service varied quite considerably between the time periods with the AM and PM peak times having many more lines, edges, active stops and system length followed by weekday-midday with Saturday and Sunday having the fewest active stops and routes. These patterns follow a trend similar to the expected travel demand typically found in most transportation systems.

However, the number of trips and the total length of trips followed a different trend with Saturday having the most, followed by Midday, PM Peak and AM Peak with Sunday, once again, being last. This pattern is due to a combination of the demand and the number of hours in the time period. Ridership also followed an unusual trend with Saturday having the highest ridership, followed by PM Peak and Weekday Midday. The low values for ridership in the AM peak are surprising as they would be expected to mirror the PM peak. This is likely due to some combination of factors including passengers running errands on the way home in the PM, workers who leave for work before 6 a.m. or arrive after 9 a.m.

As was described above, the global indicators, number of nodes, number of edges, Cyclicity (α), Complexity (β) and Connectivity (γ) are influenced more by the choice of graph configuration than by changes in the underlying service. However, within each graph configuration, there are trends which clearly separate the peak times from off-peak times. The relative values of complexity indicate that individual stops tended to have more routes serving them during peak times, as shown by the RM-Simple and RM-Complex graphs, although the number of available trips at each stop was higher during off-peak times due to the increased span of service, as shown by the Trip Maps. However, this impact was slight, as is shown by the relative stability of the TE network. The relative stability is also seen in the relative sparsity of the network, as seen in both cyclicity and connectivity, which only change substantially in the Time-Expanded networks due to the longer time off-peak time periods.

Table 4.4 Time Period Summary System Totals

Time Period Summary															
Time Period	Span of Service			System Totals									Ridership		
	Start	End	Hours	Stops	Stations (N_S)	Lines (N_L)	Trips	Line Length (L)	Trip Length	Single Edges (e^s)	Multiple Edges (e^m)	Total Edges (e)	Boarding	Alightings	Usage
AM Peak	6:00	9:00	3	6,074	5,844	173	2,464	3,395	28,514	6,604	3,822	10,426	73,598	68,780	77,515
PM Peak	15:00	18:00	3	6,158	5,930	176	2,624	3,455	29,829	6,747	3,965	10,712	97,191	94,887	102,064
Weekday Midday	9:00	15:00	6	5,687	5,485	147	2,908	2,782	32,978	6,186	2,965	9,151	96,491	94,614	114,406
Saturday Midday	8:00	19:00	11	5,454	5,265	134	3,159	2,689	55,055	5,907	3,085	8,992	111,473	108,686	117,496
Sunday Midday	10:00	19:00	9	4,916	4,753	113	1,873	2,140	33,359	5,256	1,952	7,208	75,295	72,268	79,907

Table 4.5 Time Period Summary Calculated Values

Time Period Summary - Calculated Values													
Time Period	System Averages							Transfers				Other	
	Routes / Stop	Trips / Stop	Stop Spacing	Station Spacing	Line Length	Non-Overlap	Trip Length	δ	v^t	v^t_c	ρ	τ	Opportunities
AM Peak	1.72	15.47	0.56	0.58	19.63	12.49	11.57	8	2,172	24,880	9.70	21.63	355,466,684
PM Peak	1.74	16.06	0.56	0.58	19.63	12.45	11.37	9	2,279	25,652	9.52	19.56	382,905,457
Weekday Midday	1.61	19.85	0.49	0.51	18.92	12.89	11.34	6	1,786	17,010	7.86	24.50	455,175,874
Saturday Midday	1.65	33.01	0.49	0.51	20.07	13.23	17.43	6	1,844	16,708	7.39	22.33	795,370,378
Sunday Midday	1.46	22.16	0.44	0.45	18.94	14.02	17.81	6	1,355	9,680	5.70	18.83	400,110,848

Table 4.4 summarizes a number of simple indicators for each time period. Once again there is a clear distinction between peak and off-peak times in nearly all indicators. Peak times have higher values for routes per stop, and structural connectivity due to the larger number of routes and active stops. They also tend to have lower values for trips/stop, and opportunities as the off-peak times have many more hours of service and trips. The measures of distance are more difficult to characterize or explain. Average

stop spacing increases during the peaks due to the much longer system length with only a slight increase in the number of active stops. This is likely due to an increasing amount of overlapping lines and service to remote areas during peak times. The peak times also show nearly identical distances for both average line length and average-non-overlapping line length, as would be expected due to PM routes being largely the opposite of the AM routes. The off-peak non-overlapping line length shows a steady increase from peak to midday, followed by Saturday and Sunday; however, the total line length follows no clear trend due to the specialized routes that operate in the different off-peak times. Another interesting finding was that the number of required transfers was higher in the peaks than the off-peak, which resulted in Midday and Saturday having higher values for "directness". This phenomenon is likely caused by the increasing connectivity during the peaks as the most remote locations are only reachable in the peaks.

4.2.2 Node Distributions

As with the global properties, the changes in the distribution of node properties were more dramatically impacted by the choice of graph type. This is seen in the relative fit between the LCC and betweenness values and each distribution type, which indicates that the general structure of the network and distribution of service at stops is largely the same between time periods. While there are trends with regard to the average betweenness, this can indicate either a change in number of possible O-D pairs or the length of shortest paths. Normalizing the average betweenness by the number of possible O-D pairs indicates the relative number of nodes in shortest paths. As shown in Figures 4.38 and 4.39, the normalized betweenness is lowest in the peaks with values increasing in the off-peak times, indicating that shortest paths tend to be longest on Sunday and shortest during the peaks with Midday and Saturday in between the two. This trend was apparent for both directed and undirected links in both RM-Complex and L-Space, the only two graph types with valid betweenness values in all time periods, which indicates it is likely indicative of the actual operation of the transit service.

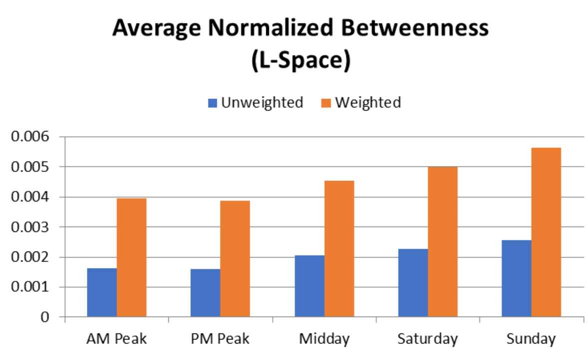


Figure 4.38 Average Normalized Betweenness (L-Space)

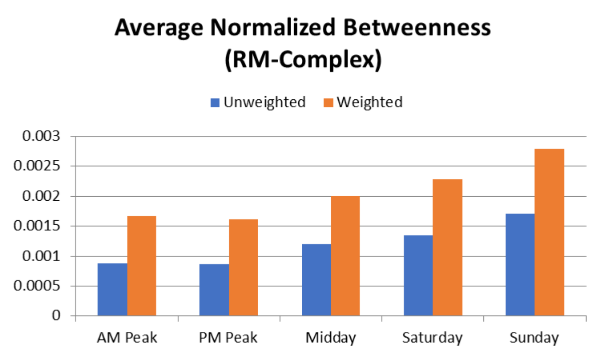


Figure 4.39 Average Normalized Betweenness (RM-Complex)

Average local clustering also fluctuates between time periods, although the trend is less consistent. Undirected values tend to be highest in Midday and Saturday and slightly lower in the peaks and an on Sunday, while the directed local clustering follows much the opposite trend. The exception to this trend is the RM-Complex graphs which have the highest average local clustering during peaks and a slight decline to midday, Saturday and finally Sunday with the lowest. Local clustering is a normalized value and so the implication of higher clustering values is that nodes within the neighbourhoods around each node tend to be more interconnected. The opposite trends between the directed and undirected links indicate that service in off-peak times is more interconnected although a higher proportion of those connections are asymmetric.

4.2.3 Small-worlds and Scale-free Networks

Neither small-world or scale-free characteristics follow a consistent pattern between time periods but instead appear to be entirely dependent on the graph representation. However, there were some apparent trends in both GCC and APL in regard to the time period. For both directed and undirected edges, GCC tended to be higher in the peaks and lower in the off-peak while APL tended to follow the opposite trend with shortest paths tending to be slightly longer in the off-peak times. This indicates that the graphs are more interconnected in the peak times as well as being slightly faster to travel through.

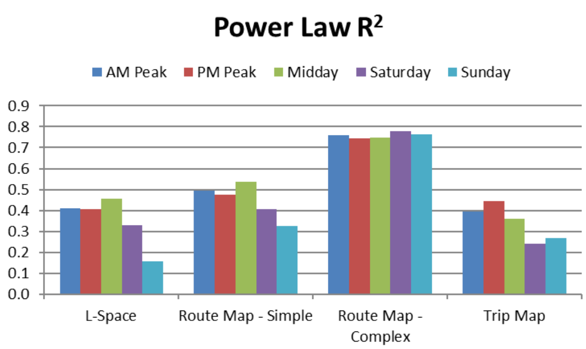


Figure 4.40 NDD Power Law R^2 by Time Period

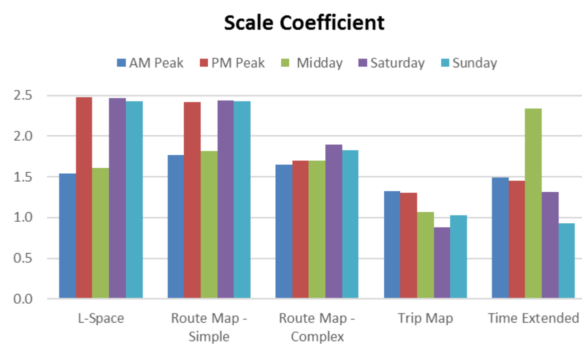


Figure 4.41 NDD Scale Coefficient by Time Period

4.2.4 Transit Service

As shown in Table 4.6 and Figures 4.42 and 4.43, the service and ridership changes dramatically between time periods with much more activity during the peak times, followed by Weekday, midday, Saturday and then finally Sunday. The AM shows the lowest total usage and GTFS events, of any time period which is surprising, particularly in light of the much higher values seen in the PM peak. This is likely due to individuals running errands or other activates on their way home, which results in multiple trips. The higher values for both Saturday and Midday are not surprising as those time periods have a much longer span of service.

Table 4.6 Transit Service by Time Period

Time Period	Active Stops	GTFS Events	Span of Service	Ave Events Per Stop	Ave Frequency	Ave Wait Time	Boardings	Alightings	Board /Hr	Alighting /Hr
AM Peak	5,844	96,065	3.0 hrs	14.8	4.9 bus/hr	12.2 min	73,345	68,052	24,448	22,684
PM Peak	5,930	101,272	3.0 hrs	15.4	5.1 bus/hr	11.7 min	96,888	92,445	32,296	30,815
Midday	5,485	116,036	6.0 hrs	22.5	3.8 bus/hr	16.0 min	96,491	94,614	16,082	15,769
Saturday	5,265	188,563	11.0 hrs	34.8	3.2 bus/hr	19.0 min	111,229	108,425	10,112	9,857
Sunday	4,753	114,198	9.0 hrs	23.2	2.6 bus/hr	23.3 min	75,046	72,077	8,338	8,009

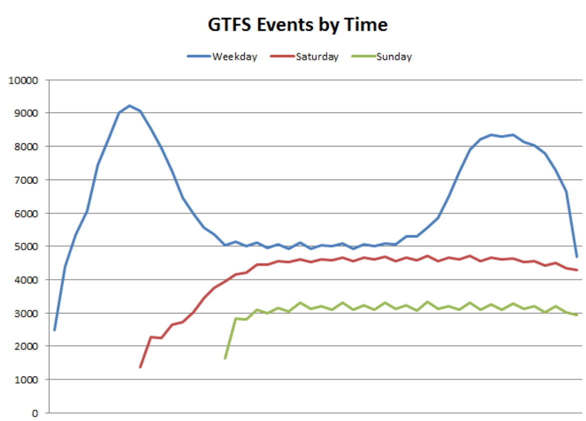


Figure 4.42 GTFS Events by Time of Day

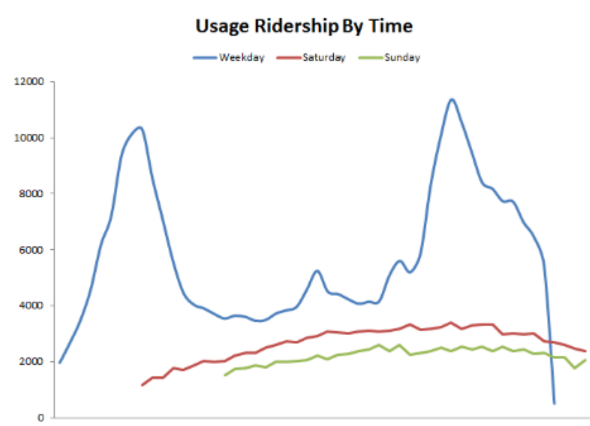


Figure 4.43 Ridership by Time of Day

The temporal distributions of both GTFS events and ridership during weekdays follow a bi-modal pattern very typical of transportation systems with very steep increases in the peaks and ridership also has a slight jump over the lunch hour. On the other hand, weekends have a single, gradual arch with the high point in the afternoon. Undulations are clearly visible in the off-peak GTFS data are due to most off-peak service running with 30 minute, or greater, headways while the resolution of the figure is in 15-minute increments. These undulations are most apparent in Sunday service where a larger portion of the service runs infrequently.

When investigated further, both stop usage and frequency of service were found to follow power laws, indicating that the majority of stops have both low frequencies with only a small number having either high frequency or high ridership. As both followed a similar distribution the correlation between boardings and frequency at each stop location was also investigated. As is shown in Table 4.7, ridership appeared to closely follow a second order polynomial relationship with the number of events that occurred during the time period. However, this relationship breaks down when transit centres are removed. This finding is interesting, as it highlights the importance transit centres play by providing very high levels of service to a large number of passengers, which are likely responsible for the power law distribution.

Table 4.7 Ridership and Frequency Distributions

Time Period	Stations	Hours	Ridership Totals			Stop Usage			Total	Frequency at Stops		
			Boardings	Alightings	Usage	Average	Formula	R^2	Departures	Average	Formula	R^2
AM - Peak	5844	3	24,621	40,668	55,030	9.42	$7332.8x^{-1.778}$	0.920	96,065	5.48	$2696.6x^{-1.749}$	0.790
PM - Peak	5930	3	47,479	34,264	62,207	10.49	$7659.5x^{-1.783}$	0.901	101,272	5.69	$3095.1x^{-1.762}$	0.816
Midday	5485	6	39,627	43,685	58,292	10.63	$3987.8x^{-1.615}$	0.892	116,036	3.53	$3180.5x^{-1.951}$	0.845
Saturday	5265	11	44,762	45,403	62,375	11.85	$4866x^{-1.653}$	0.918	188,563	3.26	$2997.1x^{-2.034}$	0.846
Sunday	4753	9	30,962	29,925	41,898	8.82	$4594.3x^{-1.711}$	0.917	114,198	2.67	$2934.7x^{-2.132}$	0.845

All Values Include Transit Centres

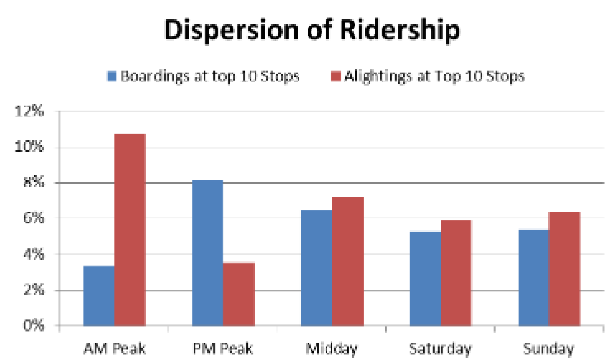


Figure 4.44 Ridership at Ten Busiest Stops

As shown in Figure 4.44, ridership is both more dispersed and more focused on the peaks than in off-peak times. This further illustrates the directionality of travel, where the busiest ten stops account for more than 10% of all alightings in the AM and more than 8% of all boardings in the PM peak. However, in the opposite direction, the demand is much more dispersed with only 3.4% of boardings occurring at the busiest 10 stops in the AM and only 3.5% of alightings occurring at the top ten stops in the PM peak. The off-peak time periods shown are more dispersed with only a slight bias for alightings at the busiest stops.

Table 4.8 Reachable Stop-Stop Pairs

Time Period	Active Stops	Stop - Stop Pairs		
		Possible	Reachable	%
AM Peak	5,844	34,140,649	32,873,143	96.3%
PM Peak	5,930	35,153,041	33,032,595	94.0%
Midday	5,485	30,074,256	28,237,753	93.9%
Saturday	5,265	27,709,696	25,902,882	93.5%
Sunday	4,753	22,581,504	20,391,733	90.3%

In order to understand the connectivity of the network, the number of possible stop-stop O-D pairs that could be reached within 90 minutes in each time period was also investigated. As with maximum transfers (δ), this could only be done using a full enumeration of the Time-Expanded networks for each time period. As shown in Table 4.8, while the PM Peak had the highest number of stop-stop pairs that could be reached, the AM Peak had a higher percentage of the maximum possible. Also, while there is a notable decline of possible O-D pairs when comparing Saturday to Midday and Midday to the PM Peak, there is only a 0.5% decline in the relative percentage, from 94.0 to 93.5. Furthermore, even the lowest level of connectivity, seen on Sunday, allows travel between 90% of all possible O-D pairs. This further supports the previous assertions that while the coverage area and frequency of service decline in off-peak, the distribution of the service among the active stops changes very little.

4.2.5 Phase II: Summary

In this phase, the transit network was characterized for each of the five-time periods, using both graph indicators and those which are more directly linked to the underlying service. Among the graph indicators were global properties, node distributions as well as both small-world and scale-free properties. An ever-present challenge with the graph indicators was in excluding properties due to the graph configuration, which was only possible after the findings from Phase I. In addition, a number of metrics related to the underlying service were used as well, including structural properties of the network, bus stop activity, usage, and available connections. Only with all of these combined was it then possible to characterize the service.

The graph metrics provided several interesting insights. The global graph properties demonstrated that the AM and PM peaks are very similar to one another with many active routes and high-frequency service which results in high values for the structural connectivity. However, they have short spans of service and many overlapping lines which also extend to poorly connected locations not active in other time periods. As a result, travel between distant locations may require more transfers and the total amount of service is higher in off-peak times. Further insight was provided through investigation of node properties, along with small-world and scale-free characteristics, which indicated that during peak times the network is more interconnected and travel through it is slightly faster. However, while the amount of service changes between time periods; the general structure of the network and distribution of service at stops is largely the same between time periods.

Investigation of the GTFS and usage data indicates that they follow very similar trends, both temporally and in how they are distributed among stops. During peak times, both service and ridership are very high and temporally focused, while during off-peak times both are lower and more uniform. Peak times were also found to have higher connectivity in terms of both the total number of Stop-Stop connections reachable and the relative percentage of the maximum possible number.

Prompted by investigations of node property distributions, the distribution of ridership and service were investigated, which had not been seen in the literature previously. The result was that both fit very well to power law curves with parameters that followed interesting trends. For stop usage, the curves were found to have very similar exponent coefficients between 1.615 and 1.783 with constant coefficients which are very similar between peak times (7332 and 7659) and weekend times (4866 and 4594). The frequency of service at stops had exponential coefficients that, while more dispersed (between 1.749 and 2.132), appeared to increase between time periods while average frequency decreased.

This further confirmed that while the temporal distribution and total amount of service did vary between time periods, it is distributed between stops consistently. While these findings are novel, their usefulness to planners is unclear. The insight provided by (Derrible and Kennedy 2011) inspired an alternate approach which might leverage the power and sophistication of graph theory to assist transit planners. This resulted in several new indicators and methods which are presented in the following section.

4.3 Phase III: New Indicators and Methods

The previous two sections demonstrated a serious challenge as many of the commonly used metrics and methods are influenced more by the choice of graph configuration than the underlying transit service. This challenge would be alleviated by establishing a standardized graph configuration, methods, and metrics. This shifts the challenge to identifying the ideal graph configuration and most useful metrics. To that end, this section demonstrates the importance of accurately representing the transit system in the graph and then introduces several new metrics intended to meet the needs of transit planners.

4.3.1 Representing the Underlying Transit Service

As graph-theoretic methods have advanced, the methods have allowed for increasingly complex graphs, which allow for them to more accurately represent the underlying transit service. The Time-Expanded configuration is used to help passengers plan their trips through transit systems and so can be assumed to represent a very close approximation to the actual experience of travellers. However, the Time-Expanded graph is difficult to use due to being large and very complicated, and so it is important to evaluate the validity of other graph representation for modelling travel through the network. While variations of the Route Map-Simple and Route Map-Complex graphs are commonly used in the literature, and transit planning software such as Remix, the performance of these models is seldom tested. To provide a simple comparison, travel times between all locations in each network were calculated and grouped into TAZ O-D pairs. The distributions of average shortest path lengths were then compared to one another using KNIME. As is shown in Figures 4.45 and 4.46, there is a clear similarity between the graphs with simple stops and those with complex stops. However, Figure 4.47 shows that even the RM-Complex graphs, follow different distributions of path lengths, indicating that the TE graph configuration should be used to model all travel through the network.

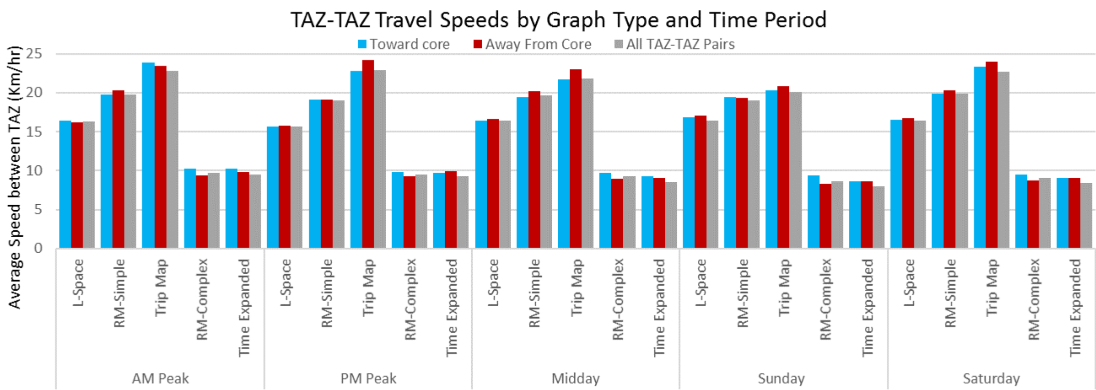


Figure 4.45 Travel Times between TAZ by Graph Type

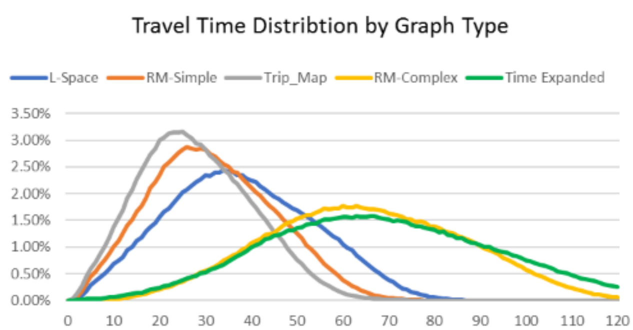


Figure 4.46 TAZ-TAZ Travel Time Distributions by Graph Type

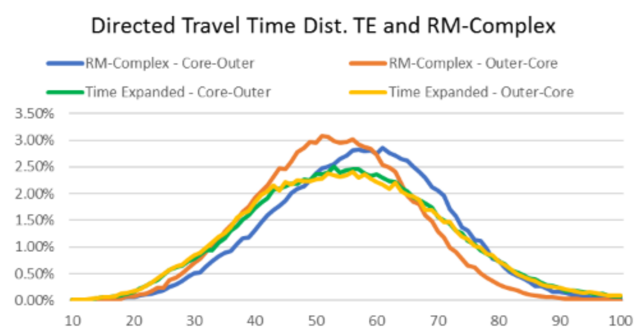


Figure 4.47 Directed TAZ-TAZ TT for TE and RM-C Graphs

4.3.2 Demand Directionality

As with service, the directionality of travel can be defined in a variety of ways. The ETS system is largely designed with the assumed travel demand to and from the downtown core area, particularly during the peaks. Ideally, this flow would be verified using the O-D information for individual passenger trips; however, this information is not readily available at this time. As an approximation, the number of boardings and alightings were used with alightings assumed to be the demand to the location and boardings the demand from that location. Using this methodology, it was found that stop activity in the AM strongly favoured boarding in the non-core areas and alighting in the core areas and this trend was reversed in the PM peak, as shown in Figure 4.48. Off-peak times showed some bias, although the trend was less pronounced. However, this finding must be tempered by the relative usage in the core and non-sore areas. As shown in Figure 4.49, only a small fraction of the total boardings and alightings actually occurred in the core during any time period. This indicates that while travel to and from the core is highly directional, the majority of travel throughout the city is far less so.

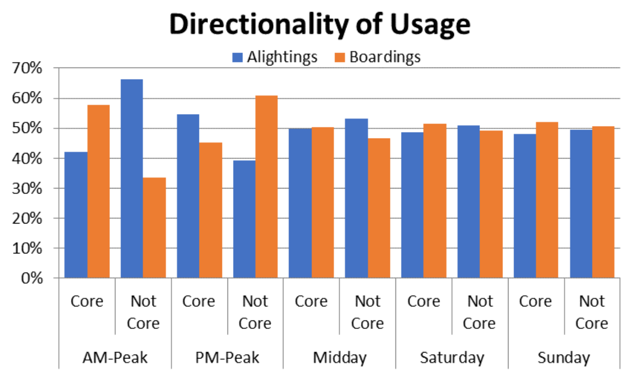


Figure 4.48 Directionality of Stop Usage

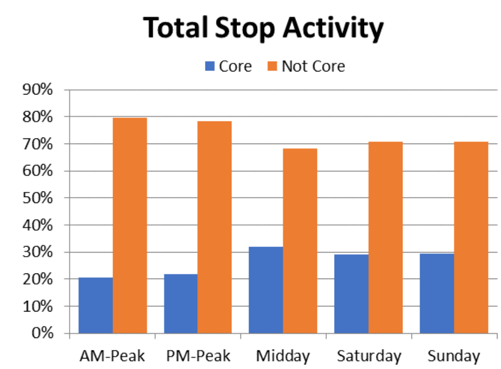


Figure 4.49 Relative Stop Activity in Core

4.3.3 Directional Travel to Nearest Transit Centre

As mentioned previously, with the advent of GTFS data and increasingly powerful computers, there has been an increase in the complexity and sophistication of metrics and methods used to analyze transit networks. One aspect of this is a gradual increase in methods that study actual travel through the network, using Time-Expanded models. However, the state-of-the-art currently treats all locations as equally important and is limited to studies of only a small portion of the network. Both issues are due to the sheer complexity of working with a full O-D matrix of travel between all stops in a large transit network. The simplest method by which this technical limitation could be overcome was to reduce the number of O-D pairs by calculating travel times from each stop to a small number of destinations, or from a small number of origins to each stop.

Transit Centres form a critical part of the ETS network and so directional service was examined between each stop and the nearest transit centre (TC). As travel is asymmetric, the direction of travel had to be defined as either from the stop to the TC, which was examined in terms of the travel time to and from the nearest TC and system-wide average travel times. Figures 4.50 and 4.51 show that while the unweighted values have only a slight directional bias, the weighted method resulted in a clear bias with shorter travel times from a TC in all time periods except for the PM Peak where a slight bias was seen in the opposite direction. Another interesting finding was that while the unweighted system-wide average showed the AM and PM peaks having nearly identical travel times that were faster than the Midday, the weighted method a 1.5 minute shorter travel time in the PM with AM and midday having nearly the same travel time.

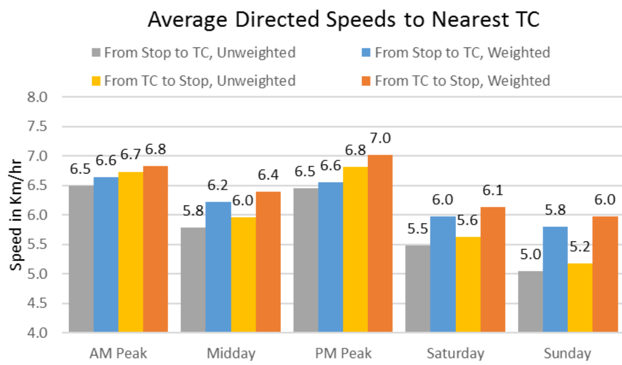


Figure 4.50 Travel Speed to and From Nearest Transit Centre

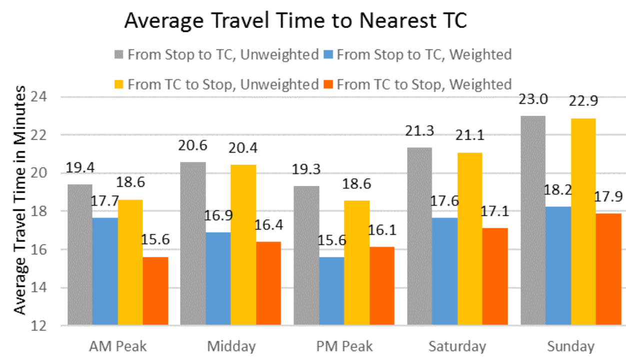


Figure 4.51 Travel Times to and From Nearest Transit Centre

The transit system provides far more connections than travel to the nearest transit centre, and so this simple method has limited actual use to planners. However, this metric does demonstrate the value of investigating directional travel and weighting travel by ridership, which was then used in more comprehensive metrics which are discussed in the following sections.

4.3.4 Directional Travel to Core

As with travel to the nearest transit centre, the speed of travel increases in the peaks with unweighted travel being only slightly favoured in the direction of demand, which can be seen in figures 4.52, 4.54, 4.56 and 4.58. However, this assumes that all locations are equally important as both origins and destinations, which is not an accurate representation of the actual experience of travellers. When the service is weighted based on actual usage, as shown in figures 4.53, 4.55, 4.57 and 4.59, the directionality of travel is much more pronounced, which confirms the assumed usage of the system. Both findings are true for both the TAZ-TAZ model and the Stop-Stop model. The finer detail provided by the Stop-Stop models, in Figures 4.58 and 4.59, allows for directional travel to be shown within the core and between locations in the periphery which is too small to be shown in the TAZ-TAZ (Figures 4.54 and 4.55). While the system average speeds are nearly identical between the TAZ-TAZ and Stop-Stop models, the directionality of travel is slightly more pronounced in the Stop-Stop model, particularly in the ridership where weighting by ridership is used.

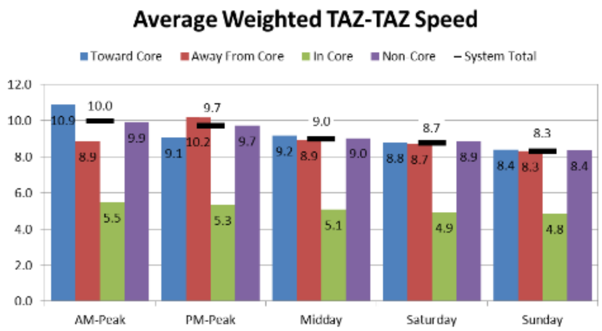


Figure 4.52 TAZ-TAZ Speeds weighted by ridership

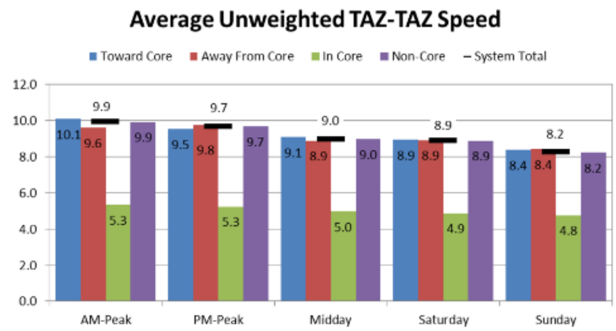


Figure 4.53 Unweighted TAZ-TAZ Speeds

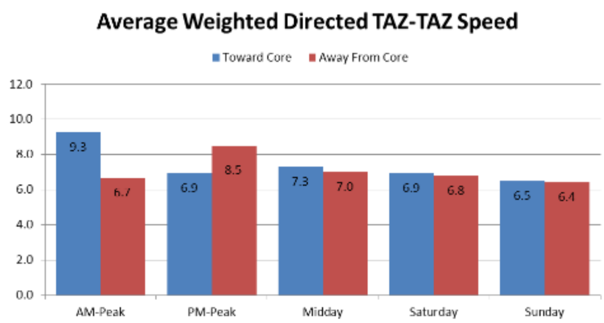


Figure 4.54 Weighted TAZ-TAZ Directed Speed

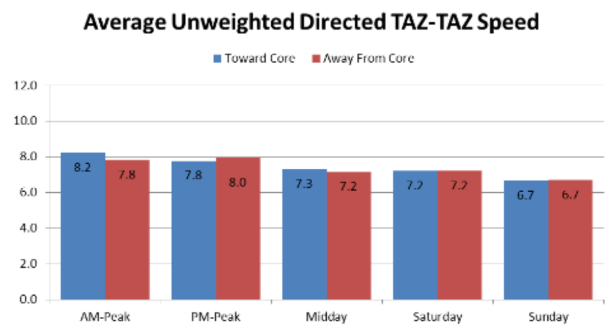


Figure 4.55 Unweighted TAZ-TAZ Directed Speed

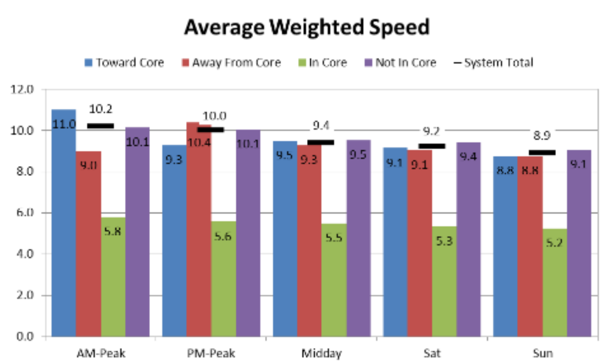


Figure 4.56 Weighted Stop-Stop Speeds

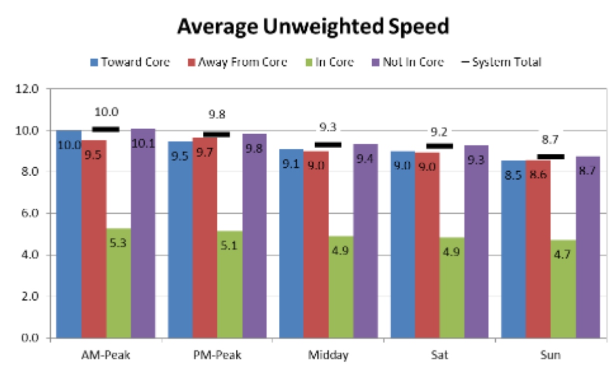


Figure 4.57 Unweighted Stop-Stop Speeds

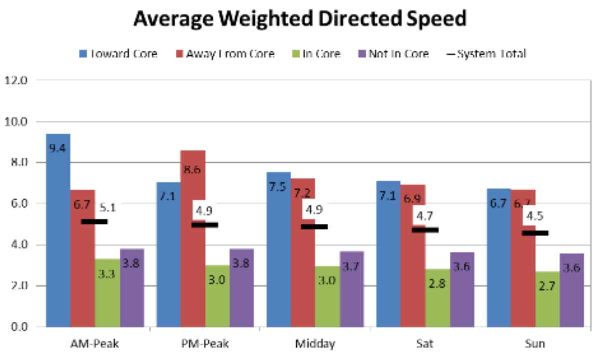


Figure 4.58 Weighted Stop-Stop Directed Speed

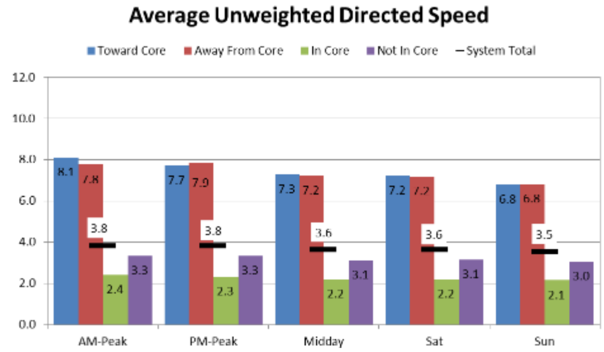


Figure 4.59 Unweighted Stop-Stop Directed Speed

This service directionality is partially obscured by another trend, whereby the speed of travel which tends to increase as the distance between the origin and destination increases. This trend, which is likely due to long initial waiting times, combined with the use of express service, appears to be largely independent of the direction of travel as is shown in Figures 4.60-4.63. As directional speed only considers the change in distance from the core, both travel within the core and between outlying areas is

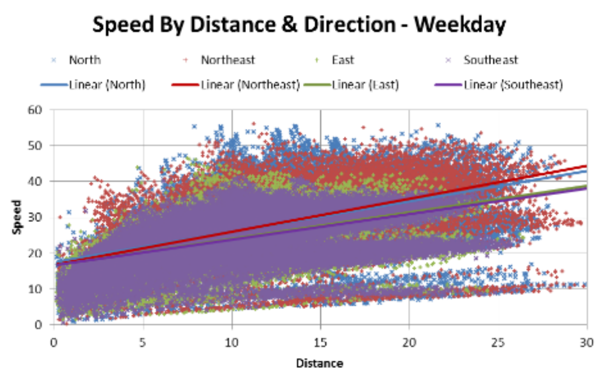


Figure 4.60 Weekday Speed by Distance

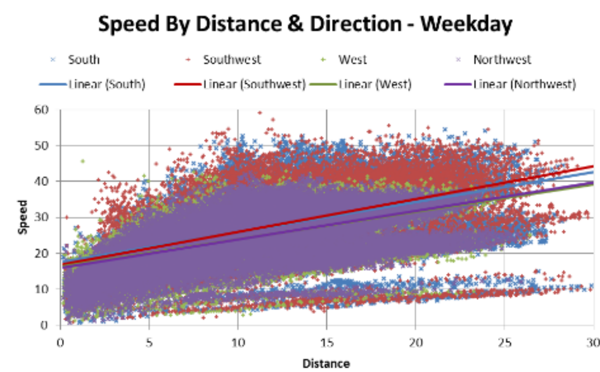


Figure 4.61 Weekday Speed by Distance

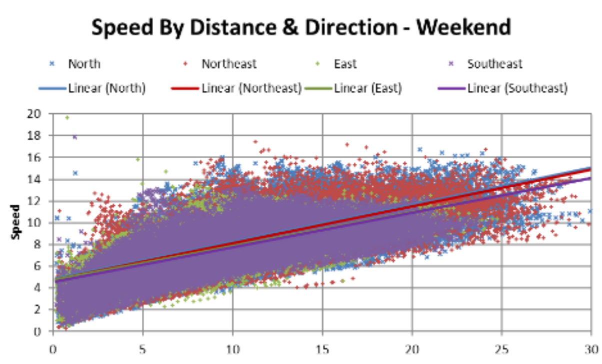


Figure 4.62 Weekend Speed by Distance

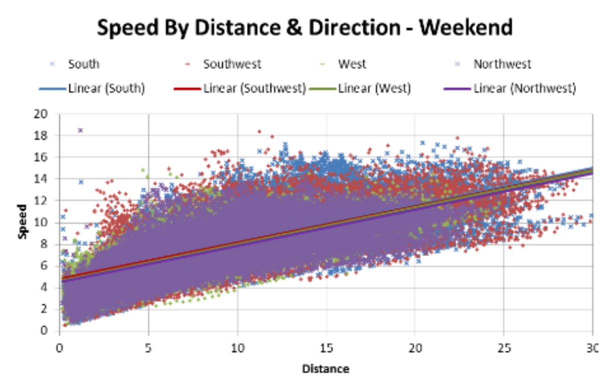


Figure 4.63 Weekend Speed by Distance

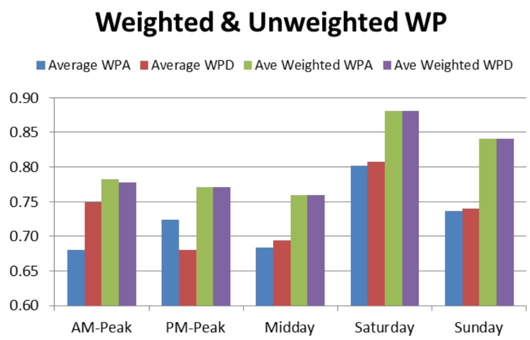
largely removed and so the impact of directionality is more apparent.

However, accurate travel times are not sufficient without a means of weighting trips based on the importance of each O-D pair. Unfortunately, the actual desirability via all modes of transportation is seldom available with any accuracy. However, the number of passenger boardings and alightings is commonly assumed by transit planners to be an adequate approximation, and so this information is collected by many transit agencies.

4.3.5 Effective Travel Times (ETT)

As would be expected, Potential Travel Time to and from each stop tended to increase with the distance from the city centre, as did the resulting ETT. Another expected result was that AM and PM peak times showed a more pronounced difference between ETT^A and ETT^D as well as between WPA and WPD than the off-peak times. However, there were also several interesting findings when comparing the values between different time periods.

First, connectivity, as indicated by WP, was substantially higher in both Saturday and Sunday than any of the Weekday time periods. Likewise, travel times, as indicated by ETT, were found to be notably smaller in the off-peak, particularly Saturday, than the peak times. These findings, shown in Figures 4.64 and 4.65, are counterintuitive as a higher level of service is provided during peaks with more frequent service along more routes and to more locations. However, upon further reflection, this finding serves to highlight a limitation of these metrics. The lower connectivity, as indicated by WPA and WPD, is the result of three phenomena that occur during peak times. One is that hundreds of stops which only have service during peak times are in periphery areas with poor connectivity. In addition, the temporal availability in peaks is reduced as a significant portion of the service is only available during a portion of the peak times. This is exacerbated by the much smaller time period during the peaks. As each trip must both start and end within the time period, many trips arriving within 90 minutes of the beginning of the time period or departing within 90 minutes of the end of the time period are excluded. As the peak time periods are much shorter than the others, this effect will be much more substantial. The third factor is the slower speed of travel during peak times due to traffic congestion.



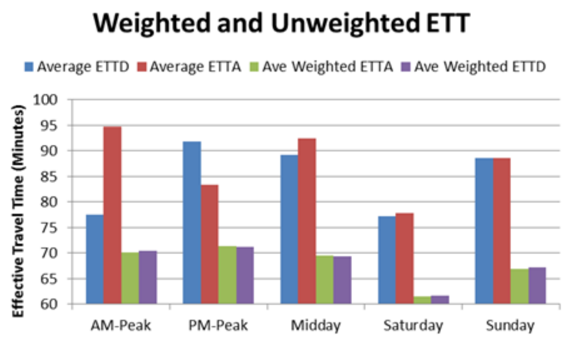


Figure 4.64 Average Weighted Potentials **Figure 4.65 Average Effective Travel Times**

Another interesting observation was the severity of the impact of weighting ETT and WP by usage. In addition to both ETT^A and ETT^D being substantially reduced in all time periods difference seen in the peak times is almost entirely removed. A similar impact is seen in WPA and WPD, although the weighted values are higher, indicating a higher degree of connectivity. While the weighted values were expected to show higher service, the weighting was expected to exaggerate the disparity between arriving and departing service during peak times. The unexpected convergence is likely the result of normalizing the arriving service by alightings and the departing service by boardings as the differences in the patterns of alightings and boardings are what led to the differing values in the first place.

When comparing the ETT, PTT and Average Travel Times, shown in Figure 4.66, several interesting observations were found. The first is that in Average Travel Time (ATT) is always higher than Potential Travel Time (PTT) and that Effective Travel Time (ETT) is always higher. The difference between ATT and PTT shows the impact of weighting by ridership and thus removing many unused stops in the periphery. The much higher values for ETT show the impact of considering connectivity. The second is that the difference between Arriving and Departing is greater for PTT than ATT and even larger for ETT. This shows the importance of ridership and the availability of the connections within the time period. The final observation is that PTT and ATT show a linear relationship while PTT and ETT follow a power law relationship. In both cases, the coefficient and R^2 vary between time periods and the relationships appear to break down as values increase. However, the relationship appears to be much stronger between ATT and PTT than between PTT and ETT as shown by the R^2 in Table 4.9 below. The fact that these relationships exist indicates how closely linked these variables are, while the complexity of the relationship and goodness of fit indicate the additional information in the PTT and ETT metrics.

Table 4.9 Relationships Between TT, PTT, and ETT

	PTT vs TT Arriving		PTT vs TT Departing		ETT vs PTT Arriving		ETT vs PTT Departing	
	Formula	R^2	Formula	R^2	Formula	R^2	Formula	R^2
AM Peak	$y = 0.790x + 9.2486$	0.727	$y = 1.302x - 26.436$	0.912	$y = 11.509e^{0.0346x}$	0.463	$y = 11.206e^{0.034x}$	0.942
PM Peak	$y = 1.237x - 22.384$	0.867	$y = 0.799x + 7.3768$	0.726	$y = 14.112e^{0.0305x}$	0.582	$y = 10.813e^{0.0358x}$	0.532
Midday	$y = 1.097x - 13.27$	0.870	$y = 1.163x - 18.247$	0.880	$y = 11.617e^{0.0343x}$	0.642	$y = 10.576e^{0.036x}$	0.663
Saturday	$y = 1.085x - 12.468$	0.827	$y = 1.107x - 14.185$	0.825	$y = 10.894e^{0.0326x}$	0.681	$y = 10.239e^{0.0336x}$	0.603
Sunday	$y = 1.079x - 13.365$	0.777	$y = 1.048x - 11.573$	0.715	$y = 8.985e^{0.0373x}$	0.568	$y = 11.690e^{0.0328x}$	0.493
	x= Average Travel Time				x= Potential Travel Time (PTT)			
	y= Potential Travel Time (PTT)				y= Effective Travel Time (ETT)			

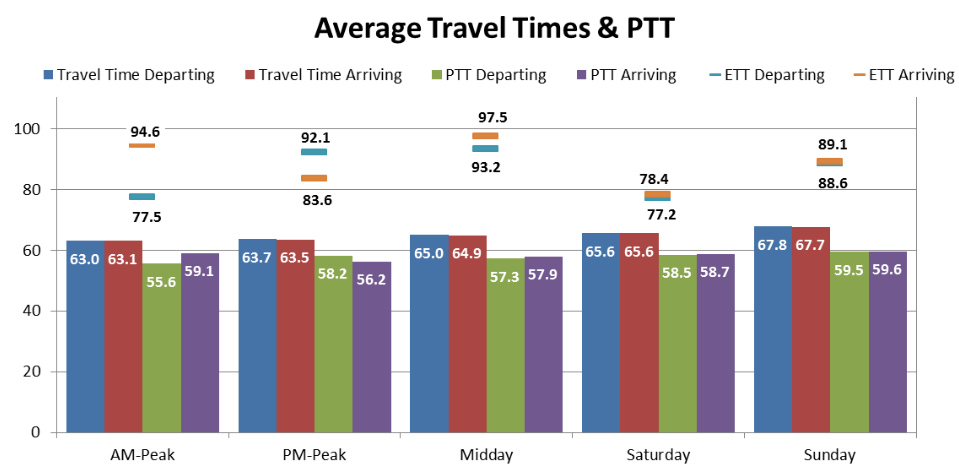


Figure 4.66 Average TT, PTT, and ETT

Another interesting finding was the trends of SI and DI between the time periods, which are shown in Figures 4.67 and 4.68. The AM and PM peak times show opposite distributions where SI is most focused in the PM Peak and least focused in the AM Peak while DI is the opposite. The other time periods are largely similar between the two time periods with Midday being more focused than the weekend time periods which are largely the same. This further demonstrates the degree to which ridership is focused in the peaks with AM showing a “many-to-few” pattern as riders travel from dispersed neighbourhoods to work, while PM shows the opposite “few-to-many” pattern as workers travel back to their dispersed residences.

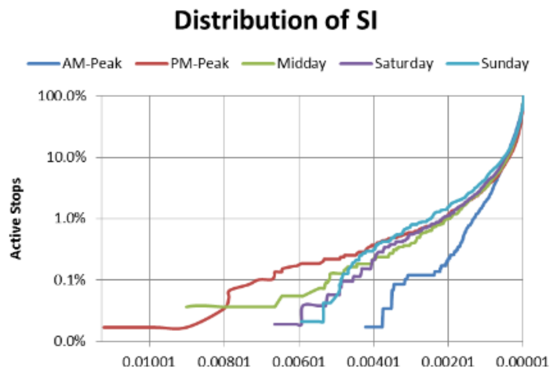


Figure 4.67 Distribution of Source Importance (SI)

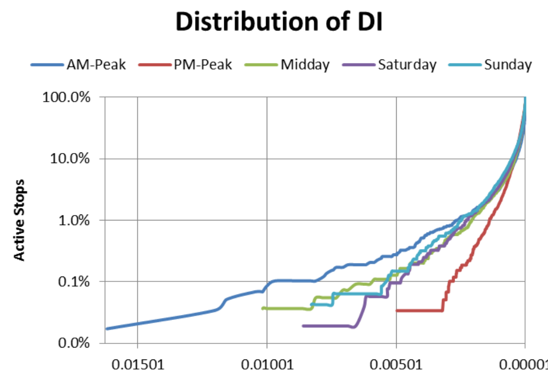


Figure 4.68 Distribution of Destination Importance (DI)

4.3.6 Phase III: Summary

The methods demonstrated in this phase represent a departure from the current state of the art, to overcome certain gaps in the existing literature. The approach taken was to first identify the needs of transit planners and develop metrics to address those needs. With the needs inferred in section 2.8, the next step was to determine the impact of commonly used graph simplifications, which confirmed that the Time-Expanded network operated very differently from the others, including the Route Map-Complex. Then, the assumed directionality of demand was confirmed by demonstrating a clear bias for travel to the city core in the AM, and away from the core in the PM peak, while off-peak times were more evenly balanced between the two.

With this information, the first new metric was developed, which demonstrated that travel times from stops to the nearest transit centre were longer during off-peak times than during peaks. While this finding was not surprising, it provided a directional travel metric which could then be used to demonstrate the impact of weighting travel times by ridership.

A new method was then developed to investigate travel speeds between Traffic Analysis Zones (TAZ), which were then used to demonstrate a directional bias toward the core during the AM and away from it in the PM. When a ridership weighted variant of this metric was used, the result had no impact on the system average travel speed but showed that ridership strongly favoured the directional travel. This simple metric now provides a tool for planners to assess how well the directional service being provided aligns with the desire of passengers.

Finally, five new metrics were developed to combine ridership weighting with the shortest paths between all stops in Time-Expanded graphs representing the ETS network. The first metric is "Importance" which is a simple ratio of the ridership activity at a stop to the total ridership activity in the system. The second is "Schedule Availability" which is the portion of the time period when it is possible to travel between the two stops being considered. When these two are combined, the result "Weighted Potential" is a useful measure of connectivity. When importance is combined with the travel times, the results is a ridership weighted travel time called "Potential Travel Time" which can then be combined with the weighted potential to obtain a composite metric called "Effective Travel Time", which can then be used to systematically quantify the level of transit service at every stop in the entire network.

4.4 Phase IV: Geographical Context

This intent of this phase is to provide a characterization of the service and demand in each time period. In order to accomplish this, additional information regarding the geographic distribution of service and ridership is needed. This information is provided through a combination of objective relationships and subjective descriptions of the data when plotted on maps.

4.4.1 Coverage

The first aspect of service that was inspected geographically was the coverage provided by the transit system between various time periods. As shown in Figure 4.69, even during the highest level of service, 44% of the area of the city is beyond 400 meters of transit service and increased to 57% during the lowest level of service on Sunday. This is largely due to much of the land inside the city boundary being reserved for future use or sparsely developed, particularly in the periphery areas. In addition, transit cannot penetrate into much of the land dedicated to the Saskatchewan River Valley parks or the Rights of Way for the Yellowhead highway, Anthony Henday Drive, and several rail lines. In this context, it is not surprising that large portions of the city are without transit access.

Also shown in Table 4.10 is the number of stops active in each time period which highlights two interesting facts. The first is that all the active stops used by ETS are never active in a single time period. Instead, each time period has a small number of stops that are only active during that time period. This is largely due to express service which runs in one direction in the AM peak and another in the PM peak with the remaining stops being served by the specialized school or community routes. The second observation is that 1,388 stops have been installed but are not used in any time period. Once again, some unused stops were expected but that they account for more than 18% of all stops is surprising and implies a number of areas are underserved. These unused stops are mostly located in periphery neighbourhoods where stops are installed as part of road construction, long before transit service is even possible. This is mandated by City of Edmonton design guidelines because stop installation is much less expensive if done during the initial construction. The remaining stops are either in locations where service has been moved to run along different roads or removed entirely.

Table 4.10 Coverage by Time Period

	Installed Stops			Non-Overlapping Area		
	Total	% Edmonton	% ETS	Total KM^2	% Edmonton	% ETS
City of Edmonton	7,524	100%	123%	699.8	100%	177%
Entire ETS System	6,136	82%	100%	394.6	56%	100%
AM Peak	5,844	78%	95%	391.6	56%	99%
PM Peak	5,930	79%	97%	392.7	56%	100%
Midday	5,485	73%	89%	364.7	52%	92%
Saturday	5,265	70%	86%	352.1	50%	89%
Sunday	4,753	63%	77%	331.4	47%	84%

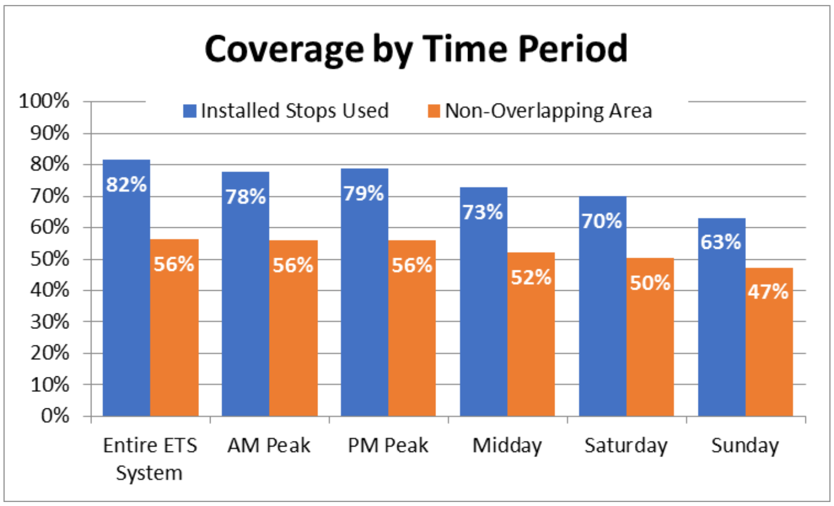


Figure 4.69 Coverage by Time Period

4.4.2 Ridership and Service by Distance from City Centre

The relationships between frequency, ridership, and distance from the city centre were also investigated. To investigate this relationship, the geometric centroid of the city boundary was found using ArcGIS and is located close to the intersection of 102A Avenue and 100 Street (approximately 120m South West of City Hall). The geometric distance from this point to each stop was then calculated and compared with frequency and ridership using KNIME. As shown in Tables 4.11 and 4.12 the number of events, active stops, boardings alightings, and usage all followed second order polynomial relationship with distance from this centroid point. The only item investigated that did not follow this relationship was frequency, which followed a power law distribution. This confirms the assumption that frequency and ridership tend to decrease with distance from the downtown core. When inspecting the formulas, the AM-Peak and PM Peak follow very similar distributions, particularly for the total ridership, number of events, and active stops. The formula for boardings in the AM Peak is very similar to the formula for alightings in the PM Peak, although the same is not true for the alightings in the AM Peak and Boardings in the PM Peak. The formulas in the off-peak times bear some similarities, but what is more interesting is that the formulas for boardings and alightings are less dramatically different than in peak times. Also, the formula for usage is dominated by Boardings in the AM Peak and by Alightings in the PM Peak, but in the off-peak times is different from both.

Table 4.11 GTFS Activity by Distance from Centroid of City

Time Period	GTFS Stop Activity by Distance From Core					
	Frequency		Number of Stop-Time Events		Active Stops	
	Formula	R^2	Formula	R^2	Formula	R^2
AM - Peak	$15.884x^{-0.499}$	0.873	$-94.904x^2 + 1659.3x + 931.52$	0.8093	$-7.727x^2 + 151.18x - 202.85$	0.861
PM - Peak	$18.17x^{-0.551}$	0.900	$-96.504x^2 + 1654.5x + 1491.2$	0.8266	$-7.749x^2 + 151.36x - 196.93$	0.863
Midday	$9.762x^{-0.467}$	0.780	$-122.28x^2 + 2012.5x + 1863.9$	0.8402	$-8.341x^2 + 157.3x - 216.01$	0.881
Saturday	$9.558x^{-0.525}$	0.932	$-226.63x^2 + 3876.3x + 430.76$	0.8293	$-7.900x^2 + 149.88x - 209.48$	0.887
Sunday	$7.414x^{-0.478}$	0.858	$-123.35x^2 + 2021.2x + 1801.3$	0.844	$-7.121x^2 + 134.13x - 178.65$	0.898

All Values Include Transit Centres

Note: These values are not for individual stops but for all stops at a given distance from the geometric centroid of the Edmonton City Boundaries

Table 4.12 Ridership by Distance from Centroid of City

Time Period	Ridership by Distance From Core					
	Boardings		Alightings		Total Ridership	
	Formula	R^2	Formula	R^2	Formula	R^2
AM - Peak	$-47.56x^2 + 853.51x - 245.41$	0.736	$-6.351x^2 - 67.067x + 2906.5$	0.848	$-49.389x^2 + 768.42x + 1677$	0.816
PM - Peak	$-13.522x^2 + 23.724x + 3455$	0.854	$-47.23x^2 + 783.18x + 817.6$	0.738	$-50.854x^2 + 737.6x + 2589.2$	0.81
Midday	$-17.873x^2 + 66.512x + 4120.9$	0.81	$-10.509x^2 - 75.688x + 4390.9$	0.833	$-24.664x^2 + 108.5x + 5402.3$	0.823
Saturday	$-21.189x^2 + 128.22x + 4016$	0.73	$-17.931x^2 + 55.934x + 4297.2$	0.761	$-30.554x^2 + 212.91x + 5328.7$	0.747
Sunday	$-12.305x^2 + 43.802x + 2848.1$	0.716	$-12.615x^2 + 43.1x + 2954.5$	0.752	$-19.978x^2 + 127.97x + 3662.3$	0.738

All Values Include Transit Centres

Note: These values are not for individual stops but for all stops at a given distance from the geometric centroid of the Edmonton City Boundaries

4.4.3 Inspection of Graph Indicator Maps

4.4.3.1 Betweenness Centrality

For the L-Space, Route Map-Simple, Route Map-Complex, and Trip Map graphs, the calculated values of betweenness centrality, local clustering and node degree were matched to the related stop in each of the five time periods, both with and without pedestrian links. These were then plotted on maps using a colour code with light values being the lowest and darker values being higher. For betweenness, this was done for three weighting schemes (UU, UD, and TT). For local clustering, only UU and UD were used as edge weight is not relevant, and for degree, only UU was used as only total degree was used, which is irrelevant of either weight or directionally.

A total of 60 maps were generated for betweenness values, although the only notable variation between time periods was in areas where no service was present, as shown in Figures 4.70 and 4.71. The L-Space graphs and Route Map-Complex graphs both followed very similar trends where the majority of nodes had very low values with the only notable “hotspots” appearing in three types of locations. The most noticeable was at transit centres and along nearby portions of the connecting arterial roads. The second most noticeable area was the downtown core, particularly between Churchill Square, near LRT stations along Jasper Avenue and near Government Centre Transit Centre. The third location was the Bonnie Doon mall, which is an important transfer point between many routes running north-south from Mill Woods to Downtown and those running East-West to University of Alberta and West Edmonton Mall (WEM). Including pedestrian, links increased the number of locations that were noticeable, although they remained clustered in the same locations (Near Transit Centres, Downtown and near Bonnie Doon Mall). There was minimal impact from edge weighting or directionality.

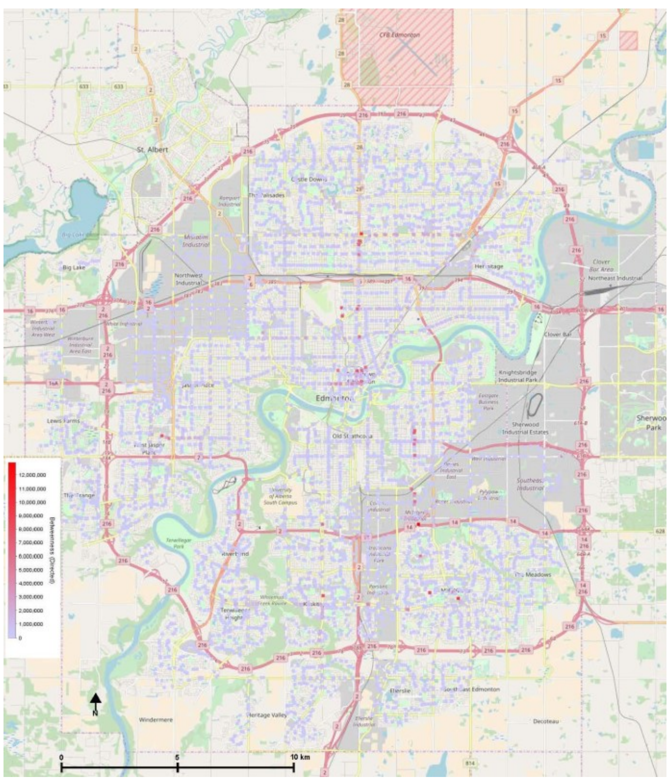


Figure 4.70 Map of Betweenness in PM Peak

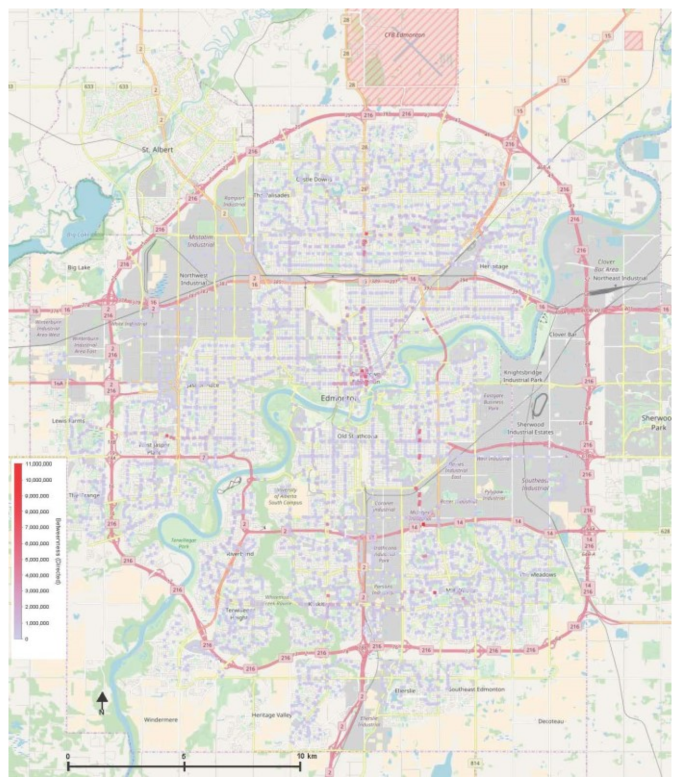


Figure 4.71 Map of Betweenness on Saturday

Typical betweenness maps, where the most notable difference between the Peak and Saturday is the lack of nodes along some peripheral roads (Ellerslie and 167 ave).

The Route Map – Simple maps appeared nearly uniform in peak times while in off-peak times a small number of corridors were noticeable. The corridors appeared to be random, such as along 129 Avenue which has a number of overlapping routes. When pedestrian links were included, the graphs in all time periods appeared much more like the L-Space and RM-C where the locations which lit up were either downtown, near transit centres or along the arterials adjacent to transit centres. Once again, directionality and edge weights had no noticeable impact on the pattern.

The trip map graphs were discarded due to numerous errors in the calculations as described in section 4.1.

4.4.3.2 Local Clustering Coefficients

Local clustering, L-Space, and RM-S graphs only showed the transit centres and intersections along transit corridors as having higher than normal values. In the RM-C maps, high-frequency routes appeared much more noticeable, with intersections among them being even more so, shown in Figures 4.72 and 4.73. Including pedestrian links in these graph types always had the same effect, with all nodes having nearly uniformly high values, which tapered off slightly near the outer edges of the city. This demonstrates that the GTFS links were largely overshadowed by the pedestrian links. The trip map graphs showed very few nodes of noticeably high values, with only transit centres being noticeably different and no pedestrian links having only a minimal impact. No graphs showed noticeable differences as a result of including directed links compared to the undirected ones.

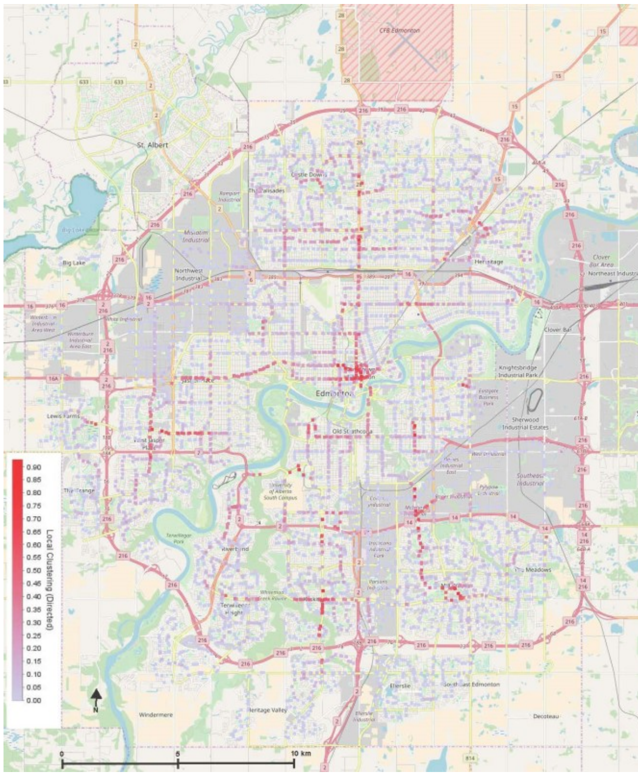


Figure 4.72 AM Peak Local Clustering in RM-Complex

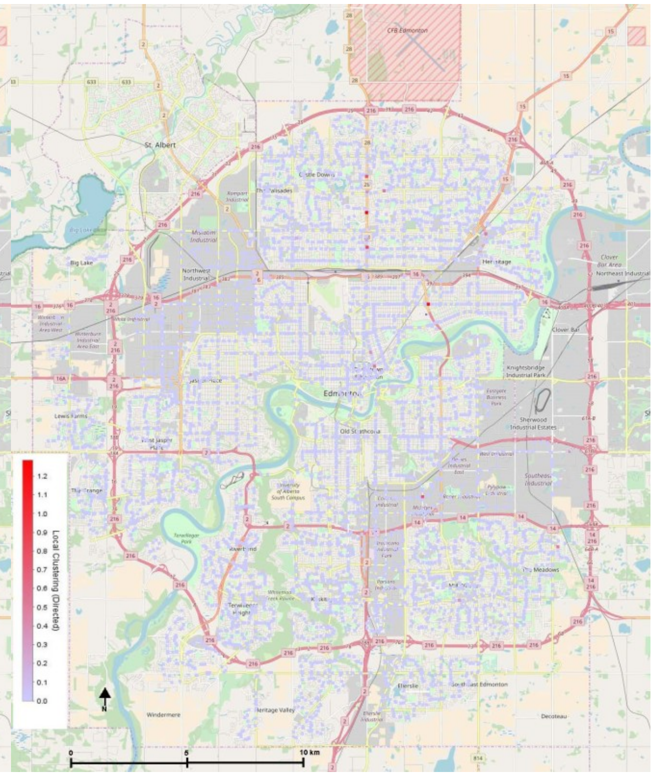


Figure 4.73 AM Peak Local Clustering in RM-Simple

Typical examples of clustering maps. These are both showing the AM peak, the RM-Complex is on the left while the RM-Simple is on the right.

4.4.3.3 Node Degree

The node degree maps for L-Space and RM-S were very similar, with important corridors showing up faintly in off-peak and slightly more apparent during peak times. Including the pedestrian, links had the largest impact on the downtown core, which is likely due to the very close spacing of stops.

The RM-C and Trip Map plots accentuated the downtown core, major corridors like as Stony Plain Road, and arterials near Transit Centres and LRT stations. In RM-C the pedestrian links only served to further accentuate intersections between corridors while the trip map graphs showed no obvious impact from the pedestrian links.

4.4.4 Inspection of Ridership Maps

For each of the time periods, alightings and boardings were plotted separately with both the colour and size of nodes being linked to the ridership. This resulted in very clear patterns that changed between AM and PM Peaks as well as between peak and off-peak times.

4.4.4.1 AM Peak

Alightings in the AM Peak showed very high concentrations in three primary locations: Kingsway/NAIT, Downtown and at Century Park. The other transit centres and LRT stations were prominent, albeit less so as were several other locations including the Wagner School (south of Argyle Road), Bonnie Doon and busy intersections like 66 Street & 142 Avenue, and Whitemud Drive & 53 Avenue. With the exception of the Century Park LRT station, the prominent locations were closer together in the core, with the number declining and distance between increasing as distance from the core increased.

The map of boardings shows a very different pattern with individual locations showing much greater uniformity, although larger areas of higher demand were apparent. The only particularly prominent individual locations were near the Coliseum & Southgate TC/LRT stations, and busy transfer locations near Bonnie Doon and the intersection of Whyte Avenue & 99 Street. Larger areas with high activity include the downtown, and Whyte Avenue, as well as portions of the Palisades, Jasper Place, and Millwoods areas.

This indicates the directional nature of travel in the AM-peak, which collects riders scattered throughout the dispersed stops and then transports them to a small number of very important locations. The activity near the transit centres and transfer locations show the degree to which travellers transfer from local routes to either express routes or the LRT. The exception to this is seen along Whyte Avenue and near Bonnie Doon mall which shows passengers travelling to the University of Alberta which originate in either the dense development near Whyte Avenue or transfer from lines running north-south to the lines running East-West along Whyte Avenue.

4.4.4.2 PM Peak Ridership

The PM Peak alightings were quite dispersed with the only individual stops appearing prominent being near Northgate Mall and at near Bonnie Doon Mall. However, the downtown and corridors along Whyte Avenue 118 Avenue and 87 Avenue near WEM showed high activity as well. Otherwise, the activity was dispersed.

PM Boardings were much more focused on a small number of individual stops near Transit Centres/LRT stations and in the downtown. In addition, several intersections appeared very busy, including the intersection of 66 Street & 142 Avenue, 178 Street & 69 Avenue, and 97 Street & 132 Avenue. The only corridors or larger areas with particularly heavy ridership are along Whyte Avenue and 87 Avenue near West Edmonton Mall.

This further demonstrates the directional nature of the travel in the PM Peak, which is nearly the opposite of that during the AM Peak. While the AM-Peak collects riders throughout the network and delivers them to the core areas, the PM Peak collects them in the core areas and then delivers them to their homes scattered throughout the city. The activity along 66 Street and south of Northgate is largely due to several high schools and junior high schools closely spaced along 132 Avenue and 144 Avenue which dismiss near the beginning of the PM Peak.

4.4.4.3 Off-peak Ridership

The map of midday alightings showed concentrated activity near Kingsway/NAIT, Northgate Mall and at locations downtown. Less prominent locations included most transit centres, as well as transfer locations near Bonnie Doon Mall, as well as locations along Whyte Avenue, Stoney Plain Road in the west, and the intersection of 66 Street and 142 Avenue. In addition, several areas had higher ridership including downtown, Whyte Avenue, 118 Avenue and 87 Avenue East of WEM. Boardings appeared much the same, although Northgate was far less prominent while locations near LRT and TC as well as the intersections of Whitemud Drive & 53 Avenue and Whitemud Drive and 122 Street were much more so.

Saturday Alightings appear very solar to the midday alightings, with only three notable differences. The first being that NAIT/Kingsway and downtown had less focused activity, the second being that The University and Whyte Avenue had slightly higher activity and the third being that service as not present at all in parts of the North West industrial and in areas north of 167 Avenue.

Boardings followed a similar trend compared to Midday, although the overall amount of activity also appeared notably lower.

Sunday followed a pattern almost identical to Saturday, although with activity reduced throughout and more areas without service including all the North West Industrial and individual roads in portions of the Terwillegar Heights and Heritage Valley areas.

Two interesting observations can be made from the off-peak ridership data. The first is that the maps of boardings and alightings are much the same, indicating that travel is less directionally focused. The higher activity at Kingsway/ NAIT than during the other times likely shows the activity of students travelling to either NAIT or the University of Alberta, which is absent during the weekend. The activity south of Northgate is somewhat surprising, particularly because it is seen only in alightings and not boardings. While this location is used to transfer when travelling downtown, however, transfers should appear as both alightings and boardings. The only explanation available is that this intersection has several large commercial centres and passengers likely alight close to their destination but when leaving they chose to walk to the nearby Northgate Transit Centre where service is much more frequent to wait for departing service.

4.4.5 Inspection of Frequency Maps

AM and PM peaks appear nearly identical, showing a very clear radial pattern with the downtown core having very high frequency and moderate frequency along several corridors which then connect to high-frequency service at transit centres. These corridors include the East-West corridors of Jasper Avenue, Stoney Plain Road, and 118 Avenue as well as the North-South corridors of 97 Street, 127 Street, and 83 Street/Connors Road. Whyte Avenue and 87 Avenue near WEM also show high-frequency corridors that serve the University instead of the downtown. Another busy corridor runs south from Century Park along 111 Street, which reflects the converging service to the LRT.

Off-peak Times follow much the same pattern, although with lower frequencies in general and voids where service does not run. Midday shows higher frequencies to more locations than Saturday, while Sunday has the lowest level of service.

This pattern shows that the service frequency, whether due to high-frequency routes or overlapping routes, tends to form a radial pattern along several important corridors. These then diverge at key transit centre locations to provide less frequent local service. It also demonstrates the clear difference in service between peak and off-peak times with much more service in peak times, supplemented by both dedicated express routes and express extensions of local routes to provide direct service to the downtown core. The LRT is not prominent despite being the single most frequent route in the ETS network because many bus stops are served by several routes with a very high combined frequency.

When comparing the frequency to the ridership, several interesting observations can be made. The first is that they both show high activity near transit centres, LRT stations, in the downtown, near WEM, and along Whyte Avenue in all time periods. This shows the degree to which the service has been matched with the ridership demand.

The second is that many areas which show very high-frequency service do not ever show high ridership activity, such as portions of Stony Plain Road east of 156 Street, parts of 75th Street in Millwoods and the areas south of Century Park on 111 Street. These areas are the result of several routes converging as they approach the downtown or Century Park LRT. While this may at first appear to be a miss-match between frequency and demand, it is the result of the road networks and the excessive frequency is likely benign as very few trips will actually need to stop at those locations. However, there may be slight efficiency gains if several of those overlapping routes were to skip those stops.

The third is the degree to which each type of map changes between the time periods. Ridership tends to appear focused at a small number of individual locations, although there is considerable diversity between those locations and the degree to which demand is focused. By contrast, the frequency maps are focused along corridors or larger areas instead of at individual locations with similar patterns in each time period. This is likely the result of differences in the underlying forces that shape them. Ridership is the aggregate behaviour of countless individuals as they go about their lives. As the activities of those individuals change between time periods, so too will their travel patterns. Frequencies are the aggregate result of the service, which is designed by a small number of planners who are working to provide the needed service, while also minimizing costs and passenger confusion. While entirely customizing the service in each time period may improve service, using similar routes and frequencies between time periods simplifies operations and use by customers.

4.4.6 Inspection of Effective Travel Times Maps

The effective Travel Time Maps resemble aspects of both the ridership and frequency maps, with additional information not present in either. A similar pattern was observed for both arriving and departing ETT in all time periods, which resembled a "bullseye" pattern with the downtown core having the highest level of service (lowest ETT) with LOS declining with distance from the downtown core. However, there were notable trends within this general appearance that varied between the direction of travel and time period:

4.4.6.1 Peak Times

The AM –Peak Arriving ETT showed the highest levels of service only near city hall, along Jasper Avenue and at the University of Alberta. The level of service to the rest of the downtown core and even along Whyte Avenue was noticeably reduced and ETT increased quickly with distance from these locations. As a result, Kingsway /NAIT and other locations with high usage had longer ETT than might have been expected. The AM –peak departing map showed an even higher level of service near City Hall and along Jasper Avenue, but ETT increased at a reduced rate with distance. As a result, the ETT to many locations within the downtown and along Whyte Avenue was noticeably lower with a less dramatic improvement visible in most peripheral locations.

This pattern reflects that service to the downtown core is slightly faster than the other way, combined with passenger travel being focused toward the core. ETT arriving is higher because boardings are more dispersed throughout the city, and so many important destinations are on opposing sides of the city, while ETT departing is lower because of the higher concentration of passenger alightings near the core. However, this indicates the system provides poor service for those attempting to travel to locations outside the core. Another interesting observation is the limited impact of express service which reduces the ETT-departing from some areas much more than nearby areas without the express service. This is only noticeable on a small number of peripheral locations in the north and in Millwoods, despite a disparity in express service within peripheral locations. The limited impact of the express service could be due to several factors, including the limited amount of time the service runs and transfers to the express service at transit centres.

The maps during the PM Peak are much the opposite, with ETT-Departing declining more rapidly with distance from the core, which follows the expected pattern caused by the directional travel of passengers which is supported by the underlying service. However, a small number of locations have much higher values for ETT than the surrounding stops. Inspection of the PTT maps shows most of these locations have very low travel time, which indicates these locations have fast connections to only a small number of other locations and few connections beyond those.

4.4.6.2 Off-peak times

During off-peak times, the maps of ETT-Arriving are nearly identical to the ETT-Departing, other than that the departing travel times are slightly lower along Whyte Avenue, Bonnie Doon and near the coliseum transit LRT station. However, an interesting trend appears when comparing the service maps between time periods. As would be expected the Midday shows lower levels of service to peripheral areas, particularly in the northwest industrial than peak times. Also, the number of active stops in outlying areas decreases noticeably between Peak and Midday, Midday and Saturday and Sunday have the fewest stops. However, many locations show a higher level of service (lower ETT) during Saturday than during Midday or even peak times. This is most apparent in areas outside the core where many of the stops have poor service in the Midday. This is most likely due to a combination of a more dispersed demand pattern, removal of many peripheral stops and more consistent schedules throughout the time period. However, the service on Sunday is noticeably lower than the Saturday, as would be expected from the reduction in service frequency and connectivity.

4.4.6.3 Comparison with Potential Travel Time (PTT) maps

The maps of PTT followed trends similar to those described above for ETT, with three notable differences. The most obvious difference being that the PTT was significantly lower which required a different scale be used to differentiate locations. This had a secondary effect of reducing the differentiation between the highest and poorest levels of service resulting in maps dominated by lime green, yellow and orange, even using a refined colour scale. The second notable difference is a number of locations with much lower PTT than the surrounding stops, the most obvious being aberrant individual stops. Less obvious, but still present were roads near the extremities, which had PTT values only slightly lower than the surrounding locations, in most cases. When compared to the ETT maps, those same locations showed up with very high ETT indicating that these locations are poorly connected.

4.4.7 Characterization of Individual Time Periods

4.4.7.1 AM Peaks

The AM Peak is characterized by having very frequent service and very high coverage, combined with express connections to provide fast service from busy locations in the periphery to the university and downtown core. This results in the service being very directional with much better connectivity and speed of travel from the periphery to the core than in the other direction. While the number of possible origins and destinations is high, connectivity is limited due to the short time window within the time period, many peripheral areas being poorly connected, and much of the service operating for only a portion of the time period. In addition, the directionality of service shows comparatively poor connectivity when travelling from the core to the periphery.

The service is also concentrated both geographically and temporally as shown by the very steep peak in GTFS events by time. The geographic focus is less prominent as many trips start in a peripheral area and travel to the core. However, the temporal concentration is apparent in dramatically reduced service before and after the peak hour, resulting in infrequent service to many stops. When compared to the PM peak, the number of trips and routes are reduced, and so average frequency is lower.

The demand follows a similar trend with boardings dispersed throughout the city while alightings are highly concentrated at stops in the downtown, NAIT, University of Alberta and at LRT stations connected to them. Activity is also highlight focused temporally as shown by the very steep curve when plotting demand over time.

4.4.7.2 PM Peaks

In many ways, the PM peak is simply the reverse of the AM peak with higher levels of service away from the core. However, the directionality is less dramatic and the amount of service, while still very peaked, is more consistent throughout the time period. In addition, the total amount of service is slightly higher than the AM Peak with more active stops, more trips on more routes, and higher average frequencies. The average speed is comparable to the AM Peak, with some metrics showing it slightly higher and others showing it slightly lower. Likewise, the connectivity is similar to the AM Peak with better service to the periphery, but many periphery locations are poorly connected.

4.4.7.3 Midday

The Midday peak is the most active of the off-peak times with more active stops, higher frequency service and faster travel speeds that either Saturday or Sunday. However, the coverage and frequency are lower than peak times. The downtown and several important corridors have much higher frequency service, although temporally, service is nearly uniform throughout the time period. The service is largely symmetric, with only a slight bias for travel to the core. However, connectivity (as measured by WP) is the lowest of any time period, which may be the result of stops in industrial areas, periphery neighbourhoods and near schools that only have service for a small portion of the time period.

Demand is much more consistent temporally than the peaks with only a slight increase for lunch. Geographically, ridership is largely symmetric with boardings and alightings having similar distributions, and most of the focused activity taking place downtown or along Whyte Avenue.

4.4.7.4 Weekend Service

Saturday and Sunday both follow virtually identical trends on most measures, albeit with Sunday service performing significantly lower than Saturday. The amount of service gradually increases in the morning and then remains nearly constant throughout the time period. Travel times are slightly biased for travel toward the core, while non-core travel is slightly below both. Both are geographically dispersed with stop frequency following nearly identical trends when plotted on a map or through linear regression. Both Saturday and Sunday have much higher connectivity (as shown by WP) than peak times, which is likely due to a combination of factors including service not being provided to many poorly connected stops in periphery areas, more dispersed usage and more consistent service throughout the time periods.

Demand for Saturday and Sunday time periods follow a simple arch-shaped temporal progression with slightly higher activity in the afternoon when compared to the start or end of the time period. Geographically, alightings and boardings are nearly identical and largely consistent between Midday, Saturday, and Sunday where demand is more dispersed and high activity is only noticeable downtown, on Whyte Avenue and near LRT stations.

As mentioned, the primary difference between Saturday and Sunday is in the magnitude of service where Saturday has a higher level of service on almost every metric than Sunday. As a result, Saturday has faster travel time, the greatest number of opportunities, with both directionality and structural connectivity very similar to Midday service. Sunday service is characterized as only providing the most basic service with the lowest values on nearly all performance measures.

4.4.8 Phase IV: Summary

The purpose of this phase was to supplement the findings of the previous three phases with additional insights related to the geographic distribution of ridership, service and calculated indicator values. As expected, the coverage was found to be highest in the peaks, followed by the Midday and Saturday time periods with Sunday having the lowest coverage. In addition, it was also found that approximately 56% of the total area within the city has access to transit service and a small number of stops are served only in off-peak times.

Geographic investigation of the stop activity and ridership revealed that both fit well to a quadratic curve with respect to the distance from the centroid of the city boundary, although this is partially explained by the distribution of active stops, which also follows a quadratic relationship. This finding does serve to confirm the assumption that both service and ridership is most heavily concentrated in the core areas and becomes more dispersed with distance from the core. In addition, the coefficients of those formulas further support the directional bias of ridership and similar levels of service in the AM and PM peak times.

The remaining findings were provided through qualitative descriptions of metric values when plotted on maps. While this methodology was less rigorous and sophisticated, it provided practical insight that is often overlooked. Maps of the various graph indicators, described in phases I and II, were more affected by the graph configuration than changes in the service between time periods, although they also highlighted areas with high frequency of service. Mapping service frequency and stop usage were more insightful, clearly indicating important corridors and locations. Ridership maps confirmed the directional bias in the peaks, although they also demonstrated that locations along the LRT, downtown and along Whyte Avenue remained important in all time periods. Maps of service frequency confirmed that peaks had more activity and service to more locations than other times. However, they also showed locations where the service did not match demand, and highlighted the radial network pattern, along with the degree to which local service declines in off-peak times.

Maps of the effective travel times demonstrated both the validity and usefulness of this new indicator. The quality of service showed the expected bullseye pattern with higher quality service in the core which gradually declined with distance from the downtown. They also demonstrated the variations in quality and directional bias of service between time periods. However, combining travel times, connectivity and ridership, provided maps that appeared very different from either ridership or frequency maps. For example, there was a noticeable improvement in ETT where buses provided express service or had direct links to the LRT. Also, several locations in the periphery which had both high ridership and frequent service showed as having poor service. This indicates where additional express service is likely to benefit a large number of ETS patrons.

5 CONCLUSIONS

5.1 Research Summary

To bridge the gap between the current use of graph theory with public transportation systems, existing methods were evaluated, and new methods developed. Both were accomplished by generating graph representations of the public transportation network operated by Edmonton Transit Service (ETS). Publicly available GTFS data was used to generate a total of fifty graph representations of this single network. As the service provided changes between the time of day and day of the week, the network was represented in five-time periods: AM Peak, PM Peak, Weekday- Midday, Saturday, and Sunday. Each of these time periods was treated as a separate transit network, which was then converted to five types of graphs, L-Space, Route Map-Simple, Route Map-Complex, Trip Map and Time-Expanded, both with and without pedestrian links. Each of these was then constructed with four edge types, Unweighted and Undirected, Unweighted and Directed, Travel Time Weighted and directed, and finally Value of Time-weighted and directed.

These graphs were then subjected to a number of tests to identify the impacts graph configuration along with actual changes in the network between time periods. This analysis was done in four phases. In the first phase, commonly used methods in the current literature were investigated, by identifying the numerous ways researchers construct graph representations of transit networks and then subjecting them to a battery of tests. It was demonstrated that many of the commonly used indicators are heavily dependent on how the graph is constructed, including the choice of graph representation, edge weighting and if pedestrian links were included or not. This was expanded in the second phase where the indicator values were used to attempt to understand how the underlying service changed between time periods. The most sophisticated indicators were found to describe the graph structure without providing insights into the underlying service that would be useful to planners. Several of the global indicators and those based on travelling through the network were found to provide useful information. Unfortunately, the global indicators were unable to provide adequate detail to be particularly useful to planners. However, those based on travelling through the network showed very strong potential.

The third phase developed new indicators from those demonstrated in the previous two phases. This was done by first demonstrating that only the Time-Expanded network is suitable for metrics related to travelling through the network. The Time-Expanded network was then used to expand the currently available indicators to represent the travel between different areas within the network. These metrics were then used to demonstrate how O-D pairs could be weighted using current ridership. Through the use of freely available graph theory and data processing tools, a method was developed to calculate the travel time between all possible pairs of stops in the network. These travel times were then combined with the weighting by ridership for each time period to create three new metrics, which apply separately for inbound and outbound service to each stop. The first “Potential Travel Times” (PTT) provides an average travel time between the stop being considered and all other stops in the network, weighted by usage. However, the values of PTT need to be adjusted to reflect the availability of connections to these other stops. As such, “Weighted Potential” was calculated for each O-D pair to reflect how much of the time period the connection was actually possible. The third metric “Effective Travel Times” (ETT) combined these two metrics to provide an indicator of the level of service at each stop in the network.

The fourth phase was an investigation of how graph properties, service, and ridership varied throughout the city of Edmonton. This provided both additional information and served as a “Sanity Check” to ensure the newly developed metrics displayed characteristics of the underlying transit service. The first test was to quantify coverage in terms of the area where service was available in each time period. Next was an investigation of relationships between properties identified for each stop and that stop’s distance from the city centre. Finally, several properties were plotted on maps so that the values could be visualized, and trends identified qualitatively. Collectively, these demonstrated that both the amount of service and demand varies between time periods; both tend to follow predictable patterns. The amount of service always follows a pattern where the highest frequency is in the downtown core and along several corridors that radiate out from the downtown and University of Alberta. While demand is also highest in the core areas, it is more directional in the peaks while largely symmetrical in off-peak times. Maps of PTT and ETT demonstrated the expected trend where travel times tend to increase with distance from the downtown core. However, the rate of change varied between time periods and each time period showed a unique pattern. The patterns were particularly pronounced in the ETT, which showed a much broader range of values reflecting both travel times and connection availability.

While these findings are largely consistent with what might be expected, they demonstrate new methods of objectively quantifying the quality of service provided by the transit network. This information can then be used in several ways. In addition to simply inspecting the maps of ETT to visualize the quality of service, planners can combine the quality of service with the relative importance of locations (SI or DI) to identify locations where service improvements are likely to benefit a large number of customers, as shown in Figure 5.1. Alternately, planners might combine ETT with population, demographic or land use information to identify underserved communities or developments.

In addition to investigating the current network, planners could use ETT to evaluate several possible networks, with a common set of SI and DI values at stops. This would allow planners to visualize and quantify which locations would see a net increase in service and which would see a net service loss. These locations could then be further weighted as mentioned previously to quantify the net improvement or decline.

Figure 5.1: Locations for potential service improvements

Size indicates SI and color indicates ETT^D, so the large red and orange circles indicate locations where improvements in ETT might benefit a large number of passengers.

5.2 Research Contributions

The initial contribution was to identify limitations in the current literature which prevented graph theory tools from being widely used by transit planners. The findings indicated that numerous graph configurations are commonly used, and most metrics seen in the literature are more influenced by the choice of graph configuration than by the underlying transit service. Furthermore, there is a substantial gap between the needs of transit planners and the information which can be obtained from the indicators seen in the current literature.

This led to the second contribution, which was to develop new indicators based on the actual needs of transit planners. Four new metrics were then developed which combined travel times between stops, the relative importance of those stops, and the geospatial information regarding their location. These new metrics, Directed Speed, Potential Travel Times, Weighted Potential and Effective Travel Time, can be used by planners to quantify the level of service provided compared to the actual needs of customers.

These metrics were all based on accurate shortest path travel times between locations, and developing a method to obtain this information was the third contribution. Investigation of shortest path calculations indicated that adequately accurate travel times could only be found using Time-Expanded graphs which were also extended to include pedestrian connections between stops and waiting at stops. While Time-Expanded graphs are seen in the literature, pedestrian links are excluded, and full enumeration is seldom possible. As such, construction, enumeration and processing the output proved a notable technical challenge.

The final contribution was to use these newly developed indicators, along with those used in the literature to evaluate the current Edmonton Transit Service, public transportation network. This contribution both demonstrated the newly adapted indicators and provided input for the planners of Edmonton's Transit Service who are currently working on a large-scale redesign of the bus network.

5.3 Limitations

ETT^A and ETT^D are used to quantify the quality of service for individual stops or areas in terms of how well the service meets the needs of the current users. ETT can then be aggregated by either an unweighted average by stop or by weighting each stop by the equivalent importance indicator (DI for Effective Arrival Travel Times and SI for Equivalent departure travel times). It can also be used to evaluate the net effect from routing modifications by comparing the ETT for each scenario.

However, ETT has three important limitations that planners must be aware of. The first, and most obvious, is that it is closely tied to the O-D data used to weight it and so is only useful when comparing the current service to that O-D data. Where the current ridership is used to estimate the O-D travel demand, ETT will only show how well the service matches the current demand at the stops currently active in that time period. It cannot be directly used to estimate how service changes might change demand, such as potential increases from expanding service to a new area. In addition, ETT cannot distinguish between situations where the service is tailored to the underlying demand or where the observed demand is made up only of those who the system happens to work for.

The second limitation of ETT is that travel times do not entirely reflect the hardship passengers experience when travelling through the network. It is well known that time spent waiting for a bus is valued higher than time spent riding, and several studies indicate an additional cost should be imposed for each required due to the potential of missing a connection. This can be partially overcome by weighting the edges to reflect the relative hardship, although this requires those values be established. Also, decimal numbers require more time and memory to process and so problems may arise if non-integer edge weights are required. However, items such as crowding on buses and shelters at stops simply cannot be reflected in ETT.

The third limitation is that ETT cannot be directly used to compare different time periods or different networks. This limitation is because SI and DI only reflect the relative importance of stops during the time period. Also, stops in remote locations or with poor connectivity will reduce the ETT for all other stops in the entire network. In order for a connection to be considered possible, it must start and end within the time period and have total travel time less than 90 minutes. The impacts of infrequent service and long travel times are exaggerated as they both increase PTA & PTD and reduce WPA &

WPD. As a result, ETT is inherently biased against service in periphery areas the bias is made more severe where periphery locations have high ridership.

As a result of these limitations, ETT, PT, and WP are primarily useful for studying a single network and have limited validity when used to compare different networks or different time periods. However, these metrics can be expanded in utility by providing alternate O-D weighting data. For example, O-D demand could be estimated from land uses population density or multi-modal travel studies. The resulting SI, DI, ETT, PT, and WP would then show the level of service in terms of those items.

Both “ridership weighted” values and “Directed Speeds” share the limitations mentioned above, although they do not include connectivity and, so it simply cannot account for poorly connected locations. Directed speed has a greater limitation in that it is only useful when analyzing travel toward a specific location which must be defined manually.

5.4 Future Research

While this thesis developed powerful tools to quantify the level of service provided in a transit network from both travel times and ridership, these tools fail to fully meet the needs of planners. Topics for future research include identifying actual travel needs, integrating modes other than fixed route transit, calibrating the new metrics and using them to estimate changes to service and understand how transit service impacts both ridership and development.

The largest unmet needs are related to identifying current travel patterns and preferences of travellers as well as predicting how they will change in the future. This would improve accuracy by providing a calibrated value of time (VoT) and weighting travel by actual travel needs instead of current ridership. Such information would be valuable for many uses and so there is active research in numerous fields including transportation engineering, urban planning, and computer modelling.

Another challenge is integrating modes other than fixed route transit into travel models and trip planners. Fixed route transit is often not economical areas with low demand or sparse populations. Instead of cancelling service, it may be possible to implement alternative services such as flexible routes, dial-a-bus, taxi, ride-source, and ride-sharing. However, these services lack a pre-determined fixed schedule and so are very difficult to integrate into the transit networks. Currently, several cities in the United States are developing tools to address this need, through grant funding provided by the Federal Transit Administration’s “Mobility on Demand (MOD) Sandbox Program” (Federal Transit Administration 2017).

The newly developed metrics ETT, PTT, WP and Directed Speeds must be calibrated before being used for decision making. This calibration should include factors to justify expanding service to new or growing areas, and interpretation of calculated values which may change between time periods and transit systems. Such calibration will require comparison of calculated values to the actual experiences of riders, obtained through surveys and ridership data. An initial calibration can be performed using historical records, and then refined with a continued collection of data.

These newly developed tools, also allow transit agencies to quantify the impacts of changes to routes, frequencies, and schedules. Due to the complexity of transit systems, the impact of items such as transfer timing and directly connected locations are often based on anecdotes or intuition. The actual value of these can now be quantified by comparing ETT of the current network with hypothetical networks. These findings can then be used to justify pilot projects, and the actual results can be used to calibrate the ETT metrics as a feedback loop.

These tools also allow transit planners to perform quantitative analysis on how that service impacts ridership and the surrounding land uses. A cross-sectional study can be done using the current network to identify obvious correlations, and several years of data could be used to investigate correlations between changes in service quality to changes in usage and development. However, as transit service is only one factor and is itself dependent on external factors, such studies will need to account for such correlations.

6 REFERENCES

Akse, Frederique A.E. Aggregate waiting time reduction on public transportation Networks: An application of multilayer networks. MSc Thesis, Linacre College, University of Oxford, University of Oxford, 2014.

Alessandretti, Laura, Marton Karsai, and Laetitia Gauvin. "User-based representation of time-resolved multimodal public transportation networks." Royal Society Open Science, 2016.

American Public Transportation Association. Ridership Report. February 17, 2017. http://www.apta.com/resources/statistics/Pages/ridershipreport.aspx.

APTA Standards Development Program. Defining Transit Areas of Influence. Recommended Practice, Washington D.C.: American Public Transportation Association, Standards Development Urban Design Working Group, 2009.

Backus, John. "Can programming be liberated from the von Neumann style?: a functional style and its algebra of programs." Communications of the ACM 21, no. 8 (1978): 613-641.

Barabasi, A., and R. Albert. "Emergence of scaling in random networks." Science 286, no. 5439 (1999): 509-512.

Bast, Hannah, et al. Route Planning in Transportation Networks (MSR-TR-2014-4. Microsoft Research, 2014.

Berge, Claude. The Theory of Graphs and Its Applications. London: Methuen & Co, 1962.

Bherkassky, Boris, Andrew V Goldberg, and Tomasz Razik. "Shortest paths algorithms: theory and experimental evaluation." Mathematical Programming. Ser A 2, no. 73 (1996): 129-174.

Biggs, N, E Lloyd, and R Wilson. Graph Theory, 1736–1936. Oxford University Press, 1986.

Boeing, Geoff. "OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks." Computers Environment and Urban Systems 65, 2017: 126-139.

Bondy, Adrian, and M. Ram Murty. Graph Theory. London: Springer, 2008.

Bunker, Jonathan M. "Assessment of Transit Quality of Service with Occupancy Load Factor and Passenger Travel Time Measures." Transportation Research Record: Journal of the Transportation Research Board (Journal of the Transportation Research Board), no. 2535 (2015): 45-54.

Canadian Urban Transit Association. Canadian Transit Fact Book 2014 Operating Data. Toronto, Ontario: Canadian Urban Transit Association, 2015.

Chatterjee, Atanu, Manju Manohar, and Gitakrishnan Ramadurai. "Statistical Analysis of Bus Networks in India." PLoS ONE 11, no. 12 (December 2016).

Chen, Yong-Zhou, Nan Li, and Da-Ren He. "A study on some urban bus transport networks." Physica A (Elsevier) 376 (2007): 747-754.

City of Edmonton Budget Office. 2016-2018 Operating Budget Fall 2017 Supplemental Operating Budget Adjustment. Budget Adjustment, City of Edmonton Budget Office, Edmonton, Alberta: City of Edmonton Budget Office, 2017.

The City of Edmonton. City Council Minutes December 6-7, 2017. 2017. http://sirepub.edmonton.ca/sirepub/mtgviewer.aspx?meetid=2171&doctype=MINUTES (accessed 12 10, 2017).

City of Edmonton Transportation Department. "City Policy C539: Transit Service Standard and Planning Guidelines." Edmonton, Alberta, 2009.

City Planning. Annual Growth Monitoring Report 2017, Our Growing City. City Planning, City of Edmonton, Edmonton, Alberta: City of Edmonton, 2017.

De Bona, A.A, K.V.O Fonseca, M.O. Rosa, R Lunders, and M.R.B.S. Delgado. "Analysis of Public Bus Transportation of a Brazilian City Based on the Theory of Complex Networks Using the P-Space." Mathematical Problems in Engineering (Hindawi Publishing Corporation) 2016 (2016).

Dell'Olio, L., Ibeas, A., Oña, J. de, & Oña, R. de. Public transportation quality of service : factors, models, and applications. Amsterdam, Netherlands: Elsevier, 2017.

Derrible, S, and C Kennedy. "Evaluating, comparing, and improving metro networks: an application." Transportation Research Record, 2146, 2010: 43-51.

Derrible, S, and C Kennedy. "network analysis of subway systems in the world using updated." Transportation Research Record, (2112), 2009: 17-25.

Derrible, Sybil. "Network Centrality of Metro Systems." PLoS ONE 7(7), 2012.

Derrible, Sybil, and Christopher Kennedy. "Applications of Graph Theory and Network Science to Transit Network Design." Transport Reviews 31, no. 4 (July 2011): 495-519.

Derrible, Sybil, and Christopher Kennedy. "Characterizing metro networks: state, form, and structure." Transportation (Springer Science & Business Media B.V.), no. 37 (2010): 275-297.

Diestel, Reinhard. Graph Theory. 5th. Heidelberg: Springer-Verlag, 2010.

Dimitrov, Stavri Dimitri, and Avishi Ceder. "A method of examining the structure and topological properties of public-transport networks." Physica A (Elsevier), no. 451 (2016): 373-387.

Dong, Jiyang, and Luzhuo Chen. "Algorithm for the Optimal Riding Scheme Problem in Public traffic." 2005 International Conference on Neural Networks and Brain. Beijing: Institute of Electrical and Electronics Engineers (IEEE), 2005. 62-66.

Erdos, P., and A. Renyi. "On Random Graphs I." Publicationes Mathematicae, 6, 1959: 290–297.

Federal Transit Administration, U.S. Department of Transportation. National Transit Database: Policy Manual 2017. Washington DC: U.S. Department of Transportation, 2017.

Federal Transit Administration, U.S. Department of Transportation. Mobility on Demand (MOD) Sandbox Program. https://www.transit.dot.gov/research-innovation/mobility-demand-mod-sandbox-program.html (accessed 12 10, 2017)

Feng, Shumin, Bayou Hu, Cen Nie, and Xianghao Shen. "Empirical study on a directed and weighted bus transport network in China." Physica A, no. 441 (2016): 85-92.

Fielbaum, Andres, Sergio Jara-Diaz, and Antonio Gschwender. "Optimal public transport networks for a general urban structure." Transportation Research Part B 94 (2016): 298-313.

Fortin, Philippe, Catherine Morency, and Martin Trepanier. "Innovative GTFS Data Application for Transit Network Analysis Using a Graph-Oriented Method." Journal of Public Transportation 19, no. 4 (2016): 18-37.

Fortuanto, Santo, and Marc Barthelemy. "Resolution limit in community detection." Proceedings of the National Academy of Sciences 104, no. 1 (2006): 36-41.

Garrison, W. L., and Duane F. Marble. "A Factor Analytic Study of the Connectivity of a Transportation Network." Regional Science 12(1), 1964: 231-238.

Garrison, W. L., and Duane Francis Marble. The Structure of Transportation Networks. Arlington Virginia: U.S. Army Transportation Command, 1962, 100.

Garrison, W.L., and D.F Marble. A Prolegomenon to the Forecasting of Transportation Development. Fort Eustis, Virginia: U.S. Army Aviation Materiel Laboratories, 1965.

Geisberger, Robert. "Advanced Route Planning in Transportation Networks (Dissertation for Ph.D., Fakultät für Informatik, Karlsruher Instituts für Technologie, 2011.

Gignoux, Jacques, Guillaume Cherel, Ian D Davies, Shayne R Flint, and Eric Lateltin. "Emergence and complex systems: The contribution of dynamic graph theory." Ecological Complexity (Elsevier) 31 (2017): 34-49.

Google. Overview General Transit Feed Specification Reference. 2017. https://developers.google.com/transit/gtfs/reference/ (accessed 12 17, 2017).

Gross, Jonathan L, and Jay Yellen. Handbook of Graph Theory. Boca Raton: CRC Press, 2004.

Hadas, Yuval, Riccardo Rossi, Massimiliano Gastaldi, and Gergorio Gecchele. "Public Transport Systems' Connectivity: Spatiotemporal Analysis and Failure Detection." 17th Meeting of the EURO Working Group on Transportation. Sevilla, Spain: Elsevier, 2014. 309-318.

Huang, Jie, and David M Levinson. "Circuity in urban transit networks." Journal of Transport Geography 45 (2015): 145-153.

Huapu, Lu, and Shi Ye. "Complexity of Public Transport Networks." Tsinghua Science & Technology 12, no. 2 (2007): 204-213.

Karypiu, G Karypis, R. Aggarwal, and V. Kumar. "Multilevel hypergraph partitioning: applications in VLSI domain." IEEE Transactions on Very Large Scale Integration (VLSI) Systems 7, no. 1 (1999): 69-79.

King, David, and Amer Shalaby. "Performance Metrics and Analysis of Transit Network Resilience in Toronto." Presented at 95th Annual Meeting of the Transportation Research Board. Washington D.C., 2016.

Kittelson & Associates, Inc TCRP Report 88 A Guidebook for Developing a Transit Performance-Measuring System. Washington D.C.: Transportation Research Board of the National Academies, 2003.

Kittelson & Associates, Inc. TCRP Report 165 Transit Capacity and Quality of Service Manual Third Edition. Washington DC: Transportation Research Board, 2013.

Lam, T.M, and H.J. Schuler. Report DMT-084. Irvine California: Institute of Transport Studies, University of California, 1981.

Latora, V., and M. Marchiori. "Efficient behavior of small-world networks." Physical Review Letters 87, no. 19 (2001).

Mishra, Sabyasachee, Timothy F Welch, and Manoj K Jha. "Performance indicators for public transit connectivity in multi-modal transportation networks." Transportation Research Part A, no. 46 (2012): 1066-1085.

Musso, A, and V.R. Vuchic. "Characteristics of metro networks and methodology for their evaluation." Transportation Research Record, no. 1162 (1988): 22–33.

Nesetril, Jaroslav, and Patrice Ossona de Mendez. Sparsity: Graphs, Structures, and Algorithm. Heidelberg: Springer, 2012.

Peixoto, Tiago P. "Model Selection and Hypothesis Testing for Large-Scale Network Models with Overlapping Groups." PHYSICAL REVIEW X 5, no. 1 (2015).

Peixoto, Tiago P. The graph-tool python library. 09 10, 2014.

Porta, Sergio, Paolo Crucitti, and Vito Latora. "The network analysis of urban streets: A Primal Approach." Environment and Planning B: Urban Analytics and City Science (Sage Journals), 2006: 705-725.

Quintero-Cano, Liliana, Mohamed Wahba, and Tarek Sayed. "Bus networks as graphs: new connectivity indicators with operational characteristics." Canadian Journal of Civil Engineering (NRC Research Press), no. 41 (2014): 788-799.

Roush, Wade. Google Transit: How (and Why) the Search Giant is Remapping Public Transportation. 2012, https://www.xconomy.com/san-francisco/2012/02/21/google-transit-a-search-giant-remaps-public-transportation/ (accessed 12 17, 2017).

Sienkiewicz, Julian, and Janusz A Holyst. "Statistical analysis of 22 public transport networks in Poland." Physics and Society, 2005.

Sui, Yi, Feng-jing Shao, Ren-cheng Sun, and Shu-jing LI. "Space evolution model and empirical analysis of an urban public transport network ." Physica A: Statistical Mechanics and its Applications 391, no. 14 (2012): 3708-3717.

Sun, Yeran, Lucy Mburur, and Shaohua Wang. "Analysis of community properties and node properties to understand the structure of the bus transport network." Physica A (Elsevier), no. 451 (2016): 523-530.

Sustainable Development and Transportation Services Departments. Transit Oriented Development Guidelines. Policy, Edmonton, Alberta: City of Edmonton, 2012.

Tsekeris, Theodore, and Anastasia-Zoi Souliotou. "Graph-theoretic Evaluation Support Tool for Fixed-Route Transport Development in Metropolitan Areas." Transport Policy (Elsevier), no. 32 (2014): 88-95.

Von Ferber, C, Yu Holovatch, and V. Palchykov. "Scaling in public transport networks." Condensed Matter Physics, 2008: 1-9.

Von Ferber, C., T. Holovatch, Yu Holovatch, and V. Palchoykov. "Public transport networks: empirical analysis and modeling." The European Physical Journal B 68, no. 2 (2009): 261-275.

Vuchic, V.R. Transportation for Livable Cities. New Brunswick, NJ: Center for Urban Policy Research (CUPR Press), 1999.

Vuchic, V.R. Urban Transit Systems and Technology. Hoboken, NJ: Wiley, 1981/2007.

Vuchic, V.R. Urban Transit: Operations, Planning, and Economics. Hoboken, N.J.: Wiley, 2005.

Vuchic, V.R., and A. Musso. "Theory and practice of metro network design." Public Transport 3, no. 40 (1991): 298.

Watkins, M.E., and D.M. Mesner. "Cycles and connectivity in graphs." Canadian Journal of Mathematics 19 (1967): 1319-1328.

Watts, D.J., and S.H. Strogatz. "Collective dynamics of ‘small-world’ networks." Nature 396, no. 6684 (1998): 440-442.

Wikimedia Foundation Inc., Seven Bridges of Königsberg. 2017. https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Wu, Jianjun, Ziyou Gao, Huijun Sun, and Hai-Jun Huang. "Urban Transit System as a Scale-Free Network." Modern Physics Letters B 18, no. 19 & 20 (2004): 1043-1049.

Xu, Xinping, Junhui Hu, Feng Liu, and Lianshou Liu. "Scaling and correlations in three bus-transport networks of China." Physica A 374 (2007): 441-448.

Yang, Xu-Hua, Guang Chen, Sheng-Yong Chen, Wan-Liang Wang, and Lei Wang. "Study on some bus transport networks in China with considering spatial characteristics." Transportation Research Part A (Elsevier), no. 69 (2014): 1-10.

Zhang, Hui. "Structural Analysis of Bus Networks Using Indicators of Graph Theory and Complex Network Theory." The Open Civil Engineering Journal (Bentham Open) 11 (2016): 92-100.

Zhang, Hui, Peng Zhao, Jian Gao, and Xiang-ming Yao. "The Analysis of the Properties of Bus Network Topology in Beijing Basing on a Complex Networks." Mathematical Problems in Engineering, 2013.

Zhang, Hui, Peng Zhao, Yinhai Wang, Xianming Yao, and Chengjiang Zhuge. "Evaluation of Bus Networks in China: From Topology and Transfer Perspectives." Discrete Dynamics in Nature and Society (Hindawi Publishing Corporation), 2015.

7 APPENDICES

7.1 Suggested methodology

The second and third contributions described above resulted in both new indicators and methods by which to obtain and implement them. As these methods were developed for research purposes, many of the technical details are connected to the specific dataset used in this research. This section provides a summary of the portions which can be readily transferred to other datasets for use in other transit agencies.

The suggested methodology for evaluating transit network using graph theory is as follows:

The first step is to use the GTFS standard to convert a transit network into a collection of Time-Expanded graphs. Each such graph should be constructed to reflect a specific level of service, with a time window of at least 3 hours. This graph must reflect the actual arrival and departure times at each stop, not average frequencies. It must also allow for transfers between each trip at each stop and the subsequent trips that visit that stop and to stops within a reasonable walking distance during that time period. All links must be directed and weighted using either time or an estimated equivalent value of time. In any case, care must be taken to prevent any cycles or negative edge weights.

With this completed, the agency must determine if the full enumeration is possible and desirable for the analysis being performed. While other methods are surely possible, the method of full enumeration done on the ETS network required a computer with four CPU cores, 32 gigabytes of RAM memory, a 400-gigabyte solid state hard drive for cache and intermediate data as well as a 1 terabyte storage drive, running Lubutnu Linux. This computer setup took approximately 2 weeks to create the time-weighted directional graphs, calculate the full O-D matrix and then process the data to calculate and map ETT, PTT, WP as was previously described. The Python scripts and KNIME workflows used can be provided although they must be adapted to use with another different data set. If a full enumeration is possible and desirable, the following method should be used:

For each graph, the travel time between each pair of stops must be calculated for each minute in the time period using Dijkstra's algorithm. A maximum travel time must be established, based on the validity of the ticket (unless a more accurate of estimate for passenger willingness to travel is available.) Travel times beyond this maximum must be discarded, as must any values indicating the O-D pair is not reachable. The remaining values should then be used to calculate an average travel time between those stops, which must also note the number of minutes when the connection is possible. Care must be taken to ensure that only travel between actual stop locations is retained and that trips between intermediate nodes are excluded. The output will be an N x N matrix where N is the number of active stops active in that time period with each O-D pair having an average travel time and the number of minutes the connection is available.

To account for differences between the relative importance of each O-D pair, SI and DI must be obtained for each stop. This could be done using current boardings and alightings, recorded on buses or could be estimated from other sources such as travel surveys, population or land use information data. Regardless, this value should be used to calculate PTT, WP, and ETT for both the service arriving and departing from each stop in the network for each time period.

This information can then be used to quantify the current level of service at all locations in the service area. Ideally, poor service will correspond to locations with low ridership while locations with high ridership will have good service. To investigate the degree to which this is occurring, the ETT arriving should be compared to the DI and the ETT Departing compared to the SI for each location. Locations, where potential improvements can be made, will be those where SI and ETT Departing are both high or where DI and ETT Arriving are both high.

If a full O-D enumeration is either not practical or not desirable, the process can be dramatically simplified in two ways. If service in only a portion of the network is of concern, then only a subset of the stop-stop O-D pairs need be calculated. O-D pairs where neither the origin or destination are being considered can be ignored (or simply not calculated). If a network-wide evaluation is required it can be divided into geographic subnetworks, such as Traffic Analysis Zones or neighbourhoods. However, this method is far less detailed and of less utility. This is done by first determining the size, shape, and location of each subnetwork. Using this information, the geometric centroid can then be calculated, and each stop assigned to a specific subnetwork. The centroids can then be used to calculate the distance between each subnetwork and if desired other important locations within the coverage area. The Time-Expanded networks are then used to calculate the travel time between each stop and all other stops. However, instead of aggregating this for each stop-stop pair, it is instead aggregated for each subnetwork which allows for optimizations that can dramatically simplify the entire process. The output from this is an S x S matrix of travel times where S is the number of subnetworks. The travel times can then be combined with the distance between centroids to estimate travel speeds. SI and DI can then be calculated for each subnetwork and used to calculate weighted average speeds between them. If travel to other locations (such as a central business district) is also required, then “directed speed” is calculated as shown in equation A1 below:

$$TT_{ij} = Travel\ time\ from\ i\ to\ j$$

$$D_{it} = Distance\ from\ the\ centroid\ of\ origin\ subnetwork\ i\ to\ location\ t$$

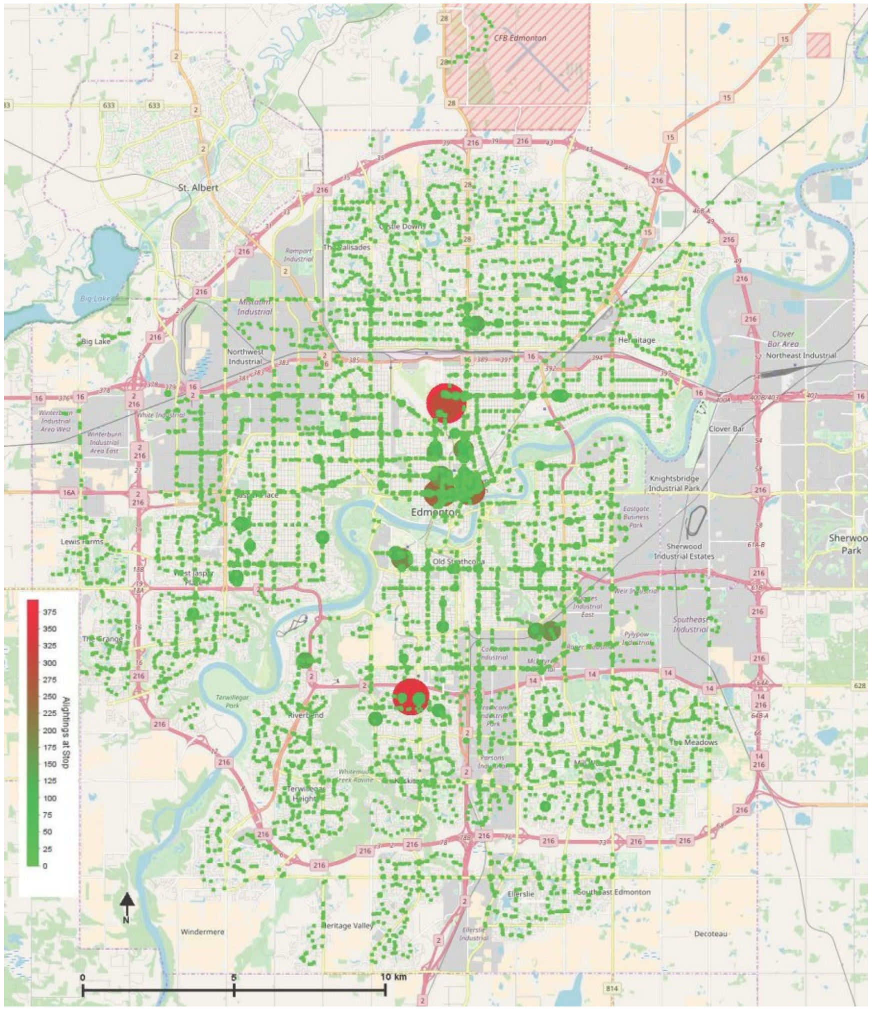
$$D_{jt} = Distance\ from\ the\ centroid\ of\ destination\ subnetwork\ \ j\ to\ location\ t$$

$$DS_{ij}^{t} = Directed\ Speed\ between\ i\ and\ j\ with\ respect\ to\ location\ t$$

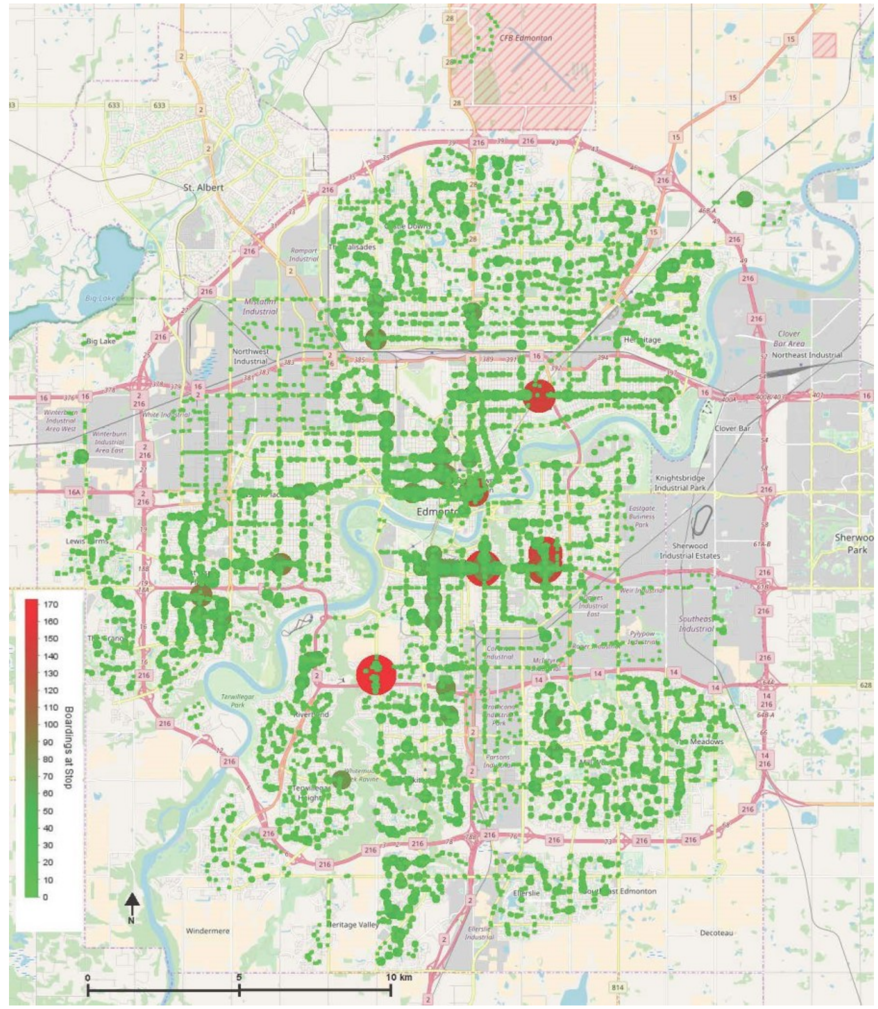
$$DS_{ij}^{t} = \frac{D_{jt} - D_{it}}{TT_{ij}}$$

This process can also be used to evaluate the impact of potential changes to the network by creating Time-Expanded Graphs representing the network with the proposed changes, performing the calculations and comparing the results with those from the current network.

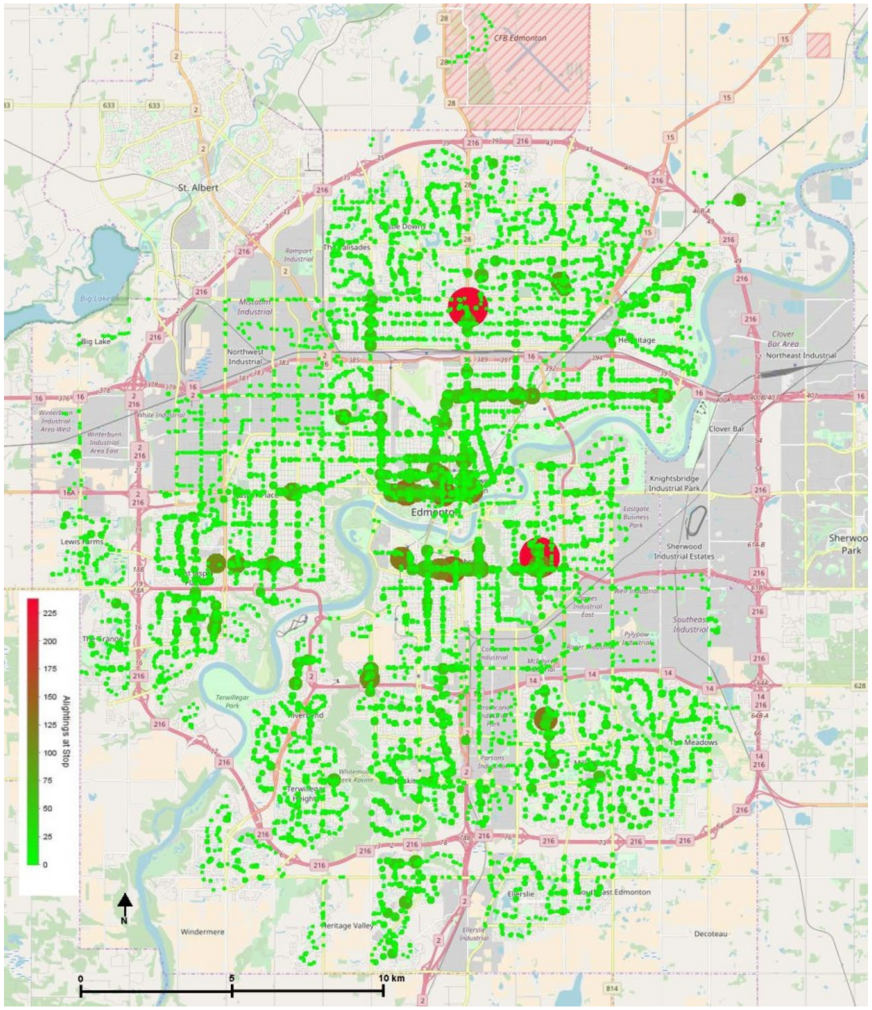
7.2 Ridership Maps



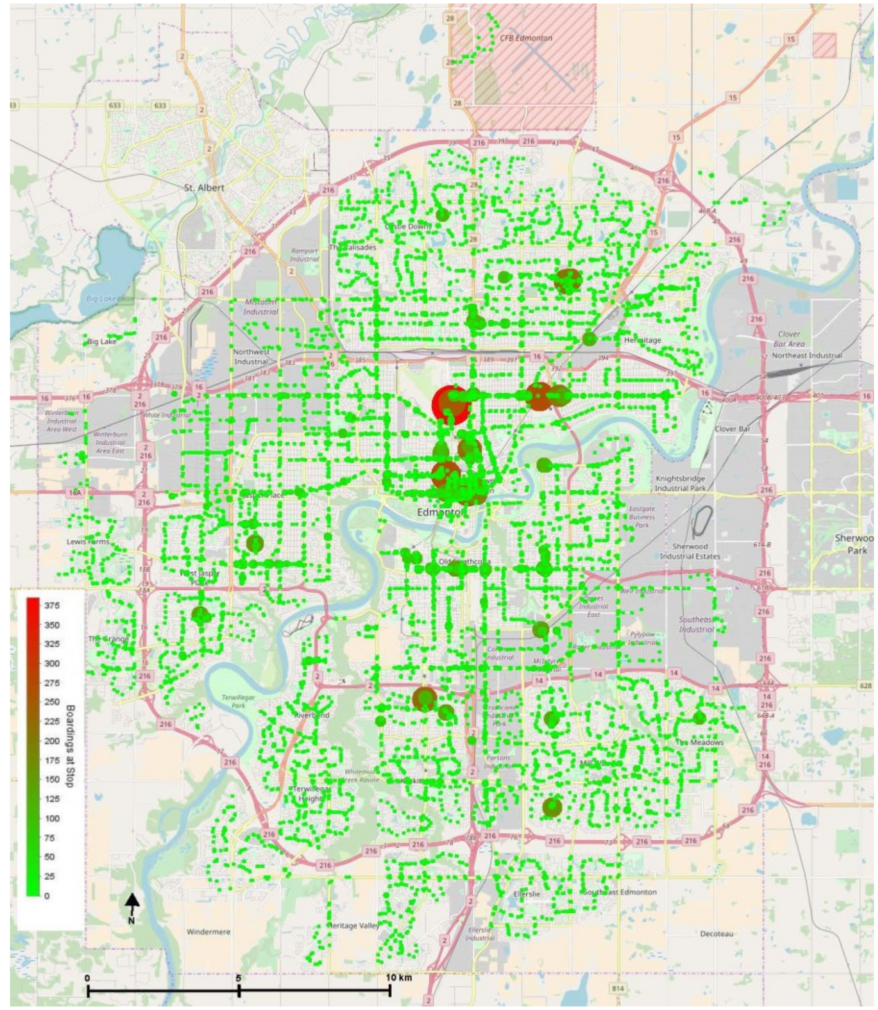
AM Peak Alightings



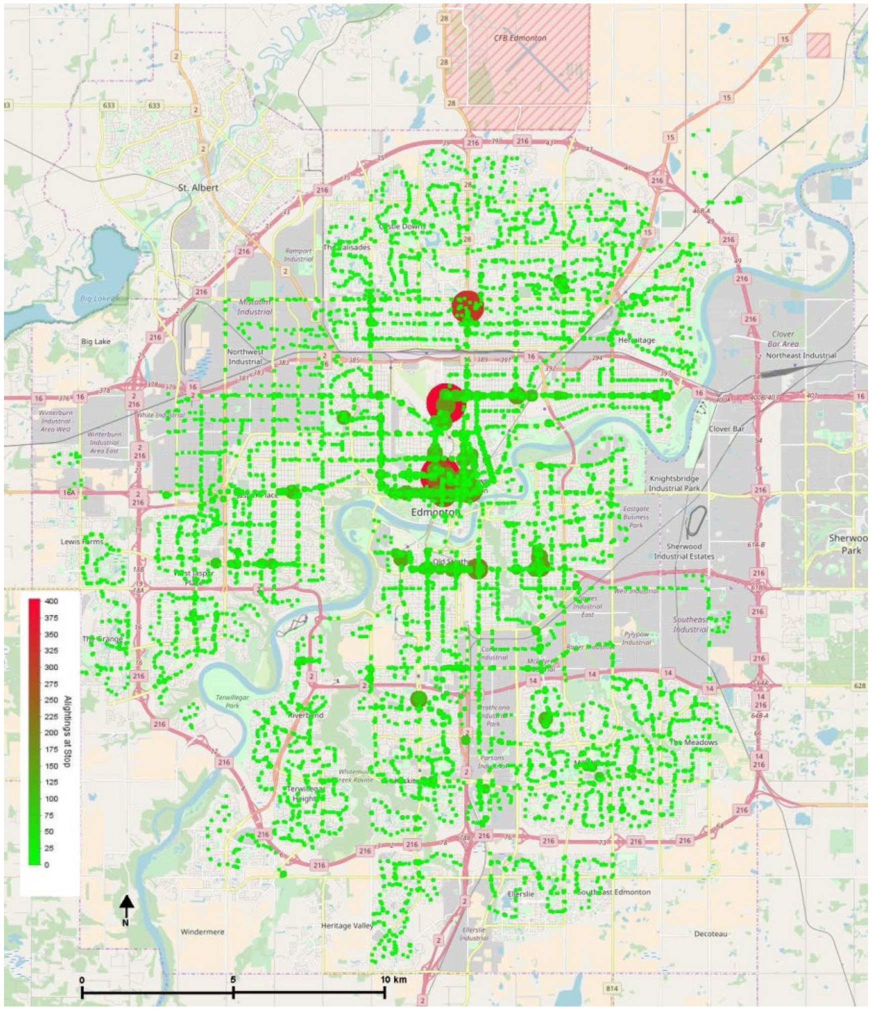
AM Peak Boardings



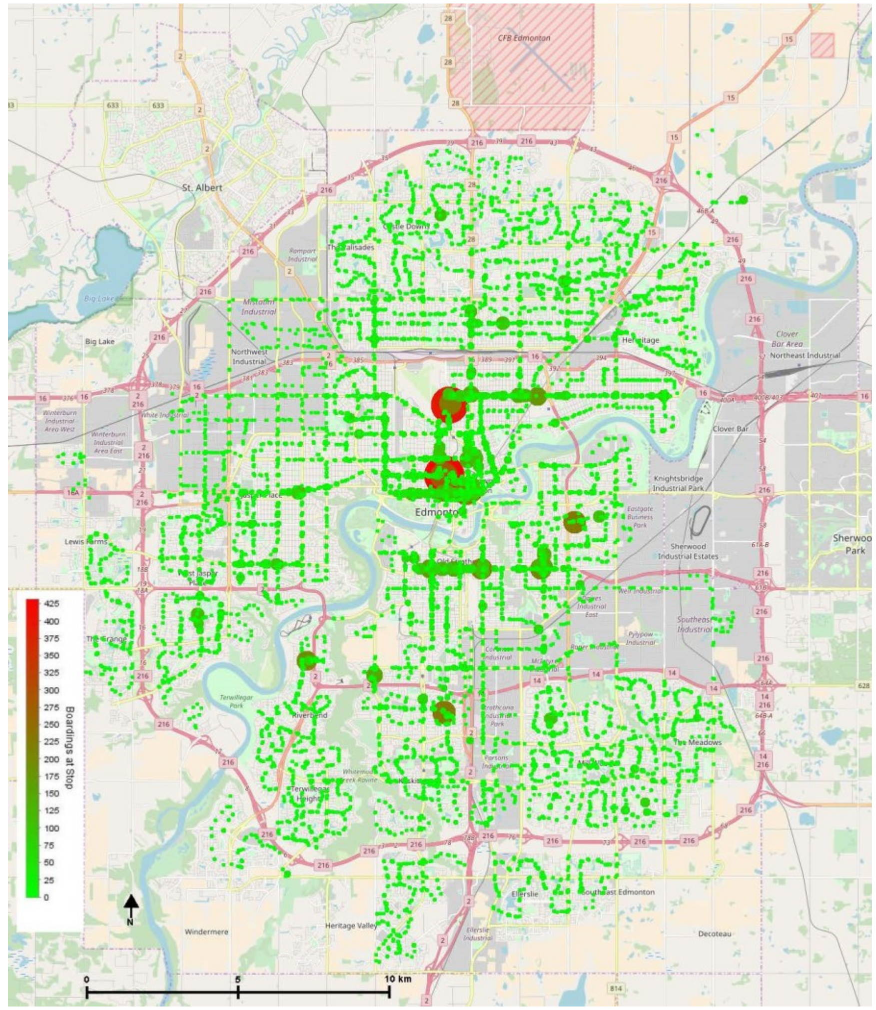
PM Peak Alightings



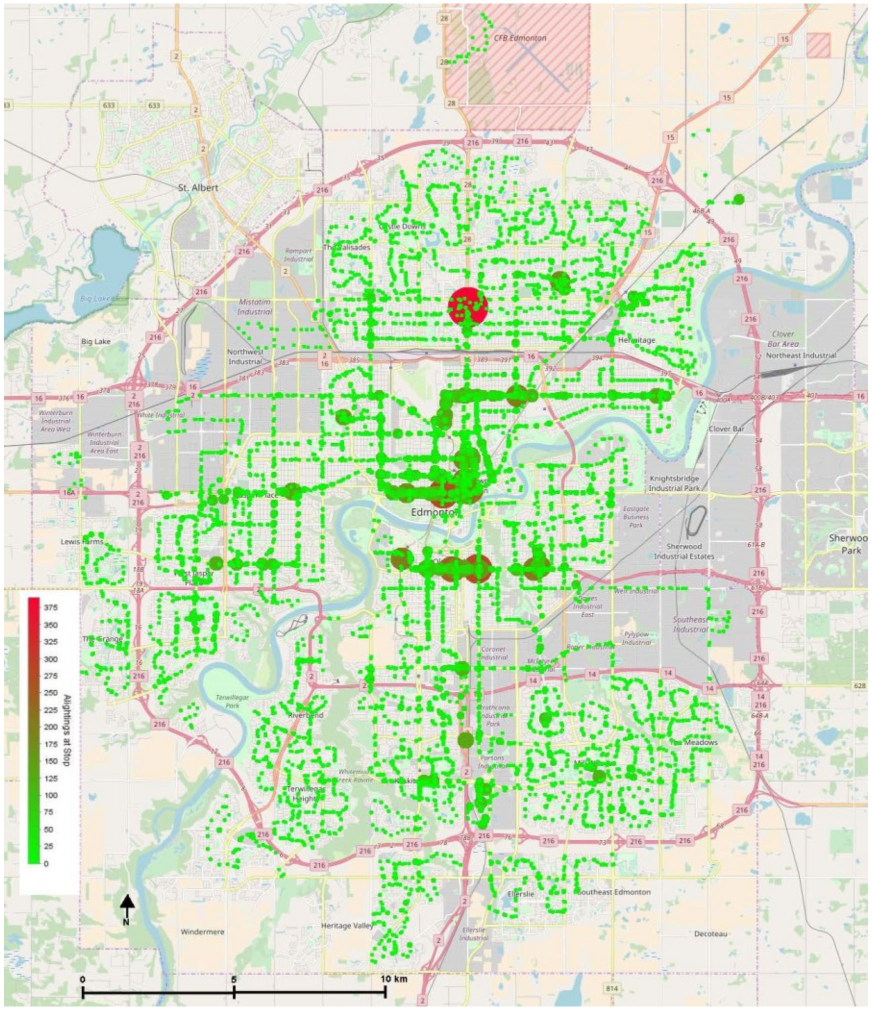
PM Peak Boardings



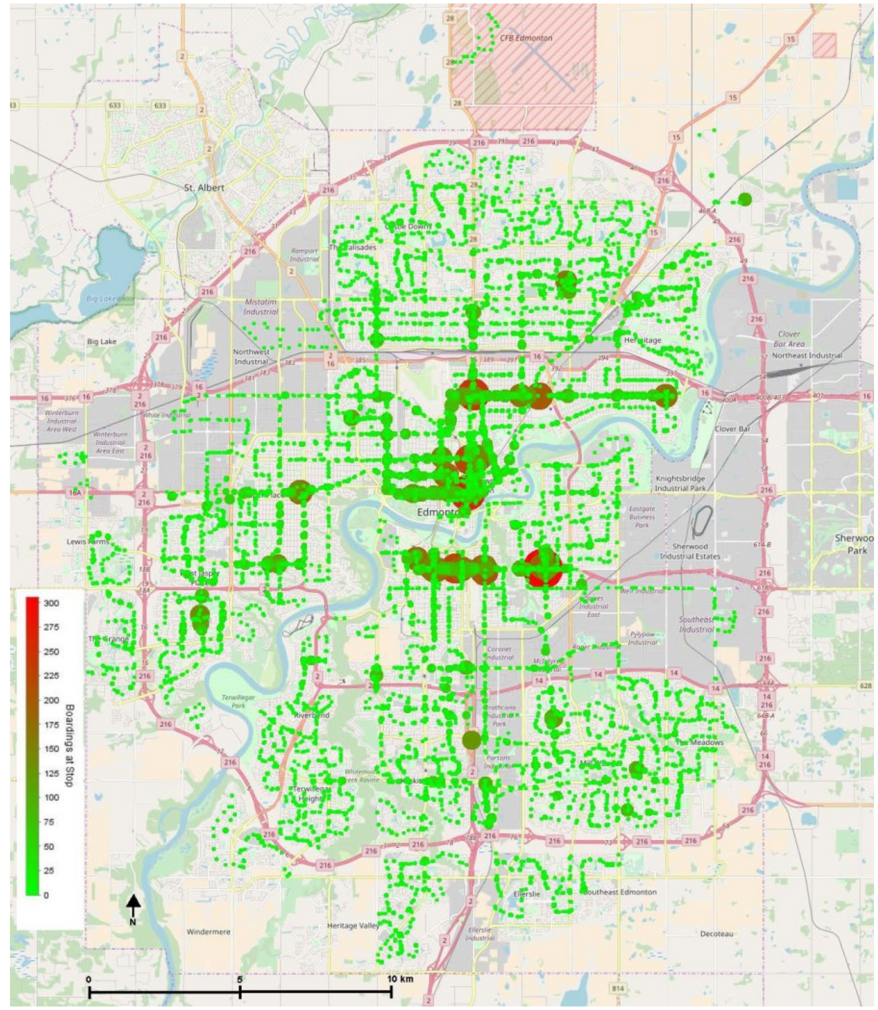
Weekday- Midday Alightings



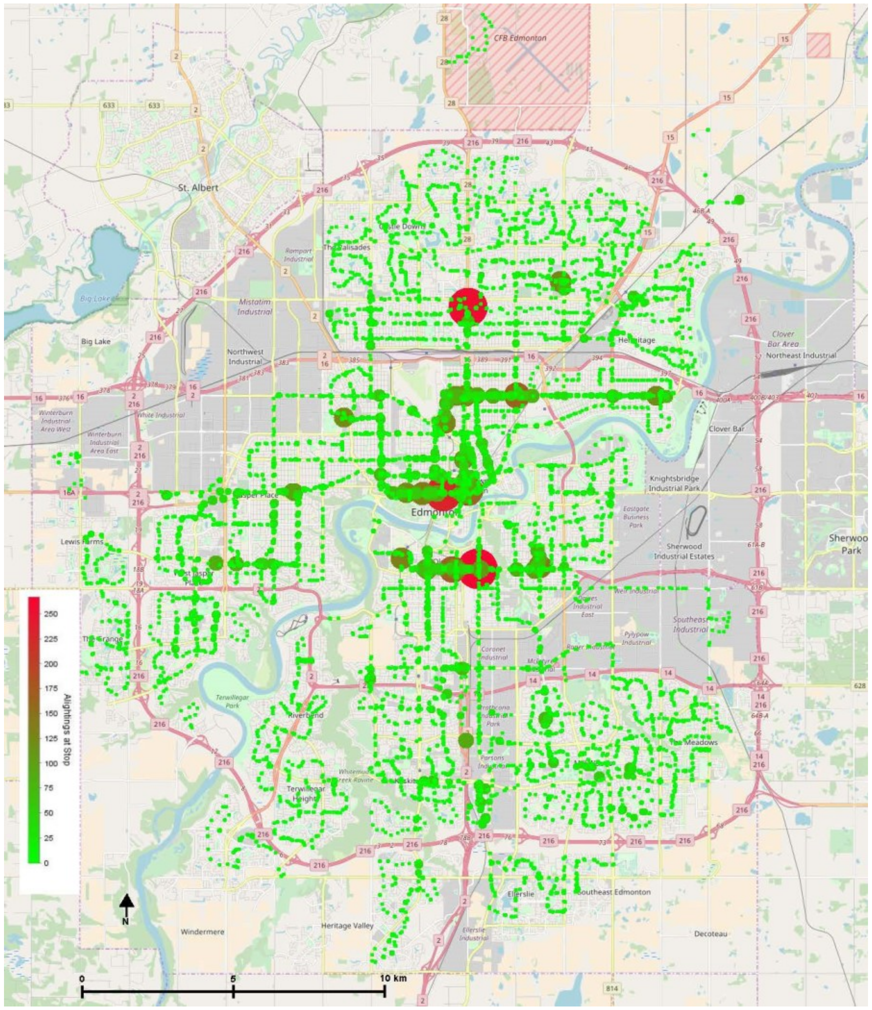
Weekday-Midday Boardings



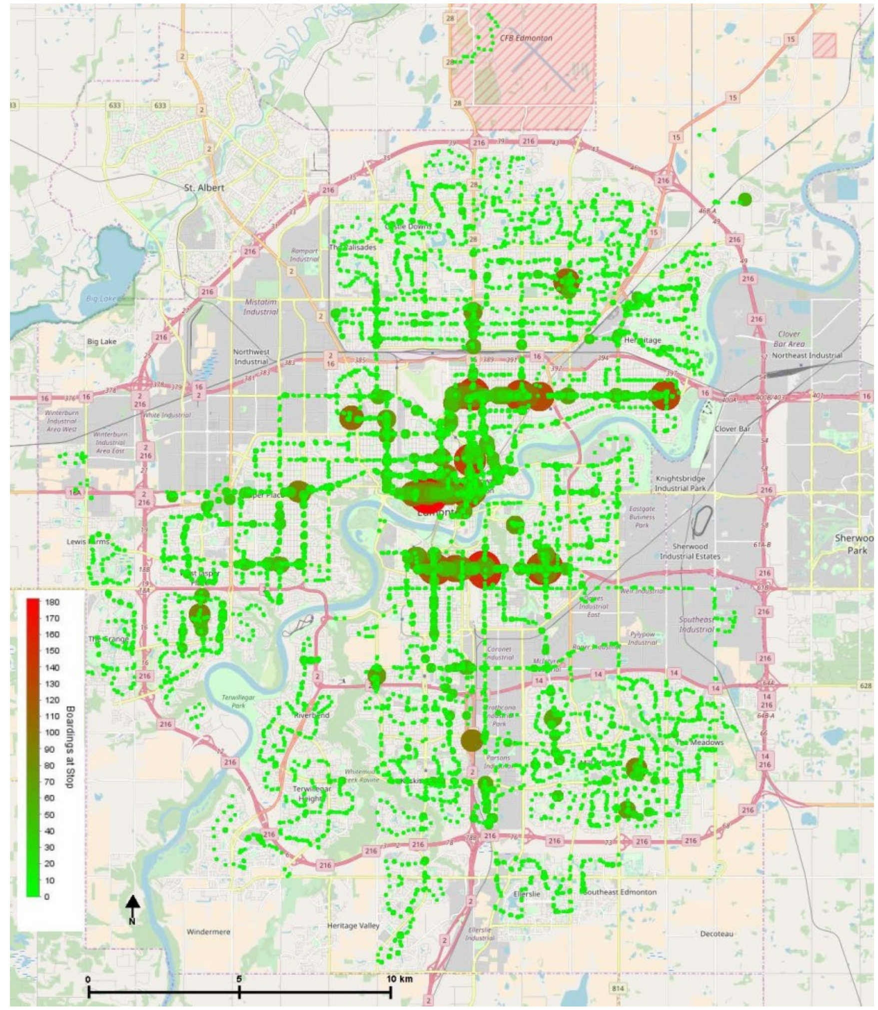
Saturday Alightings



Saturday Boardings

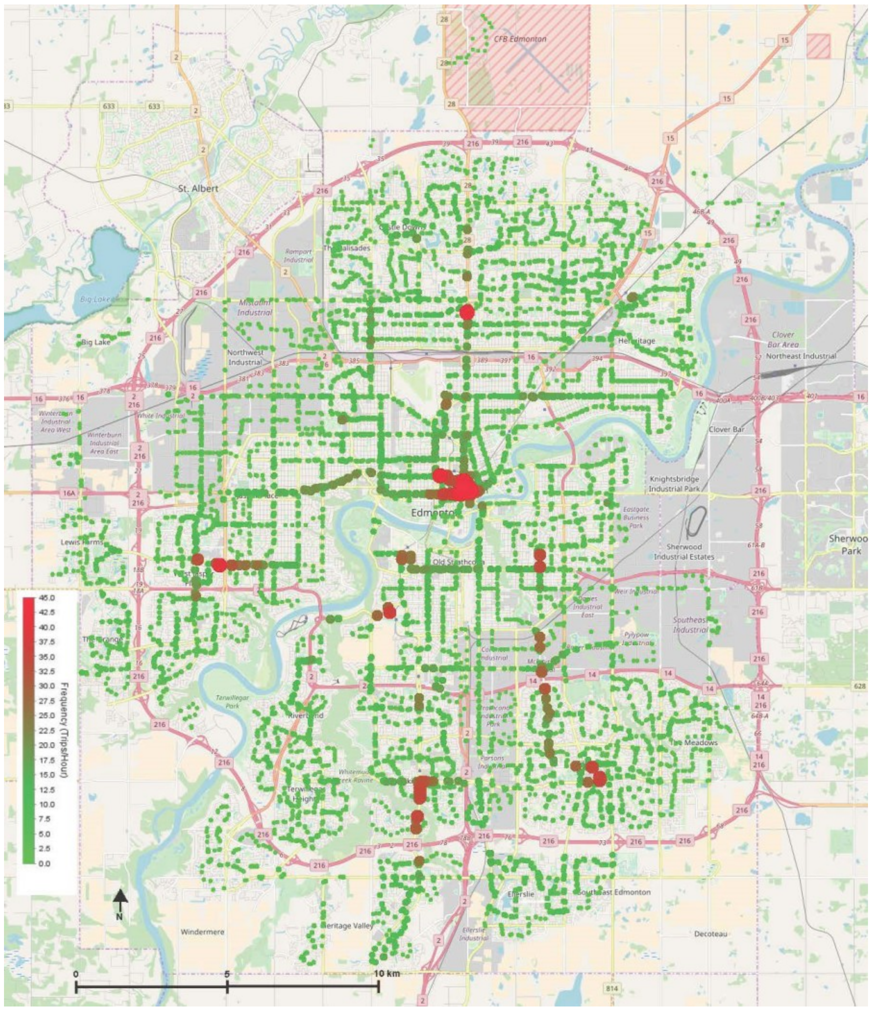


Sunday Alightings

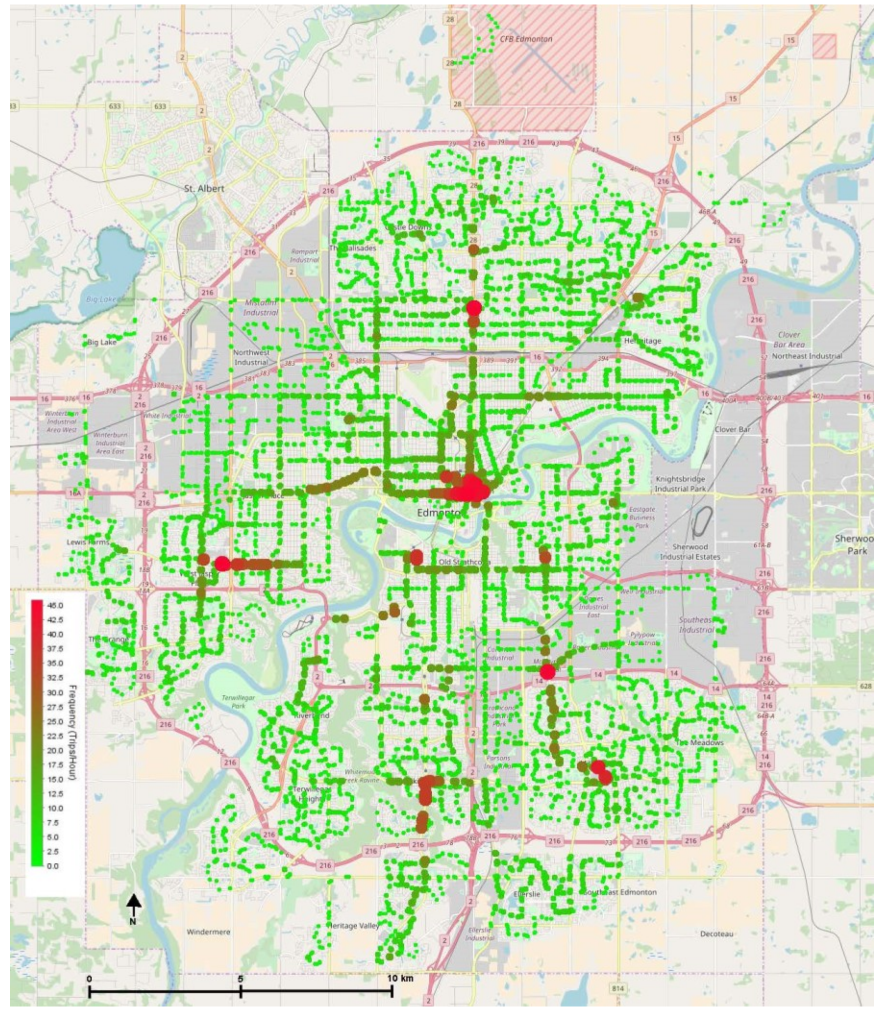


Sunday Boardings

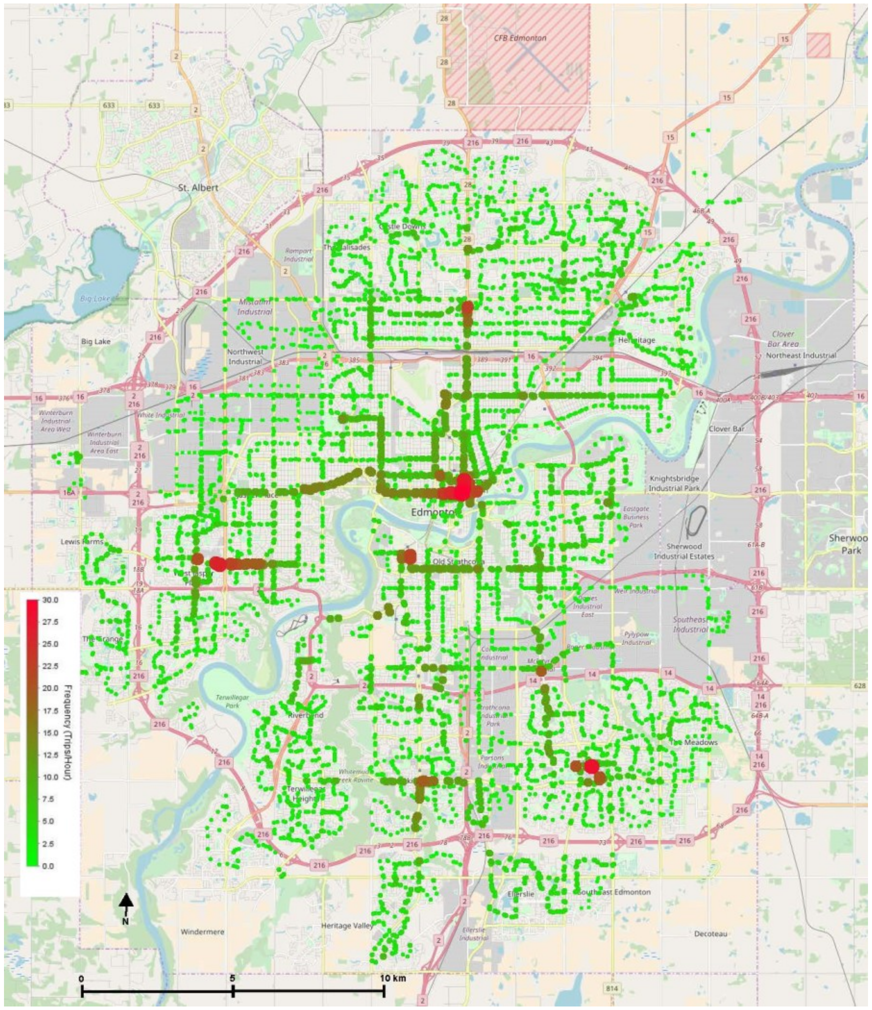
7.3 Frequency Maps



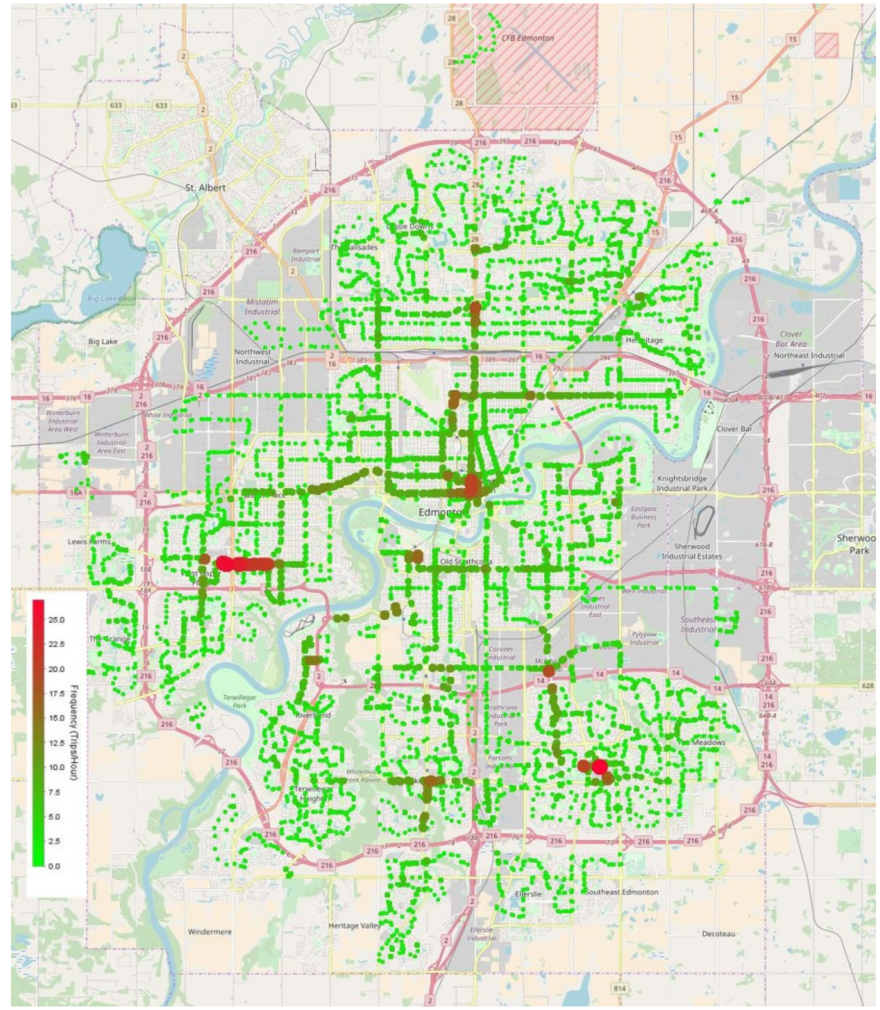
AM Peak Frequencies



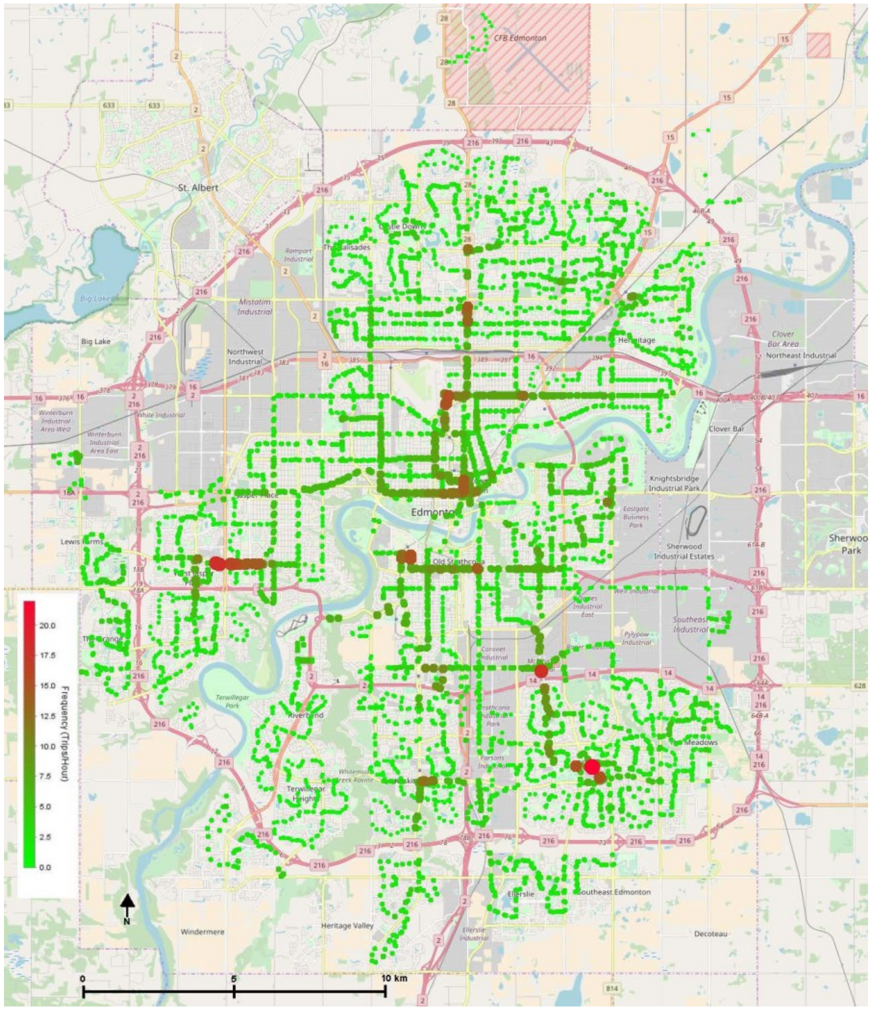
PM Peak Frequencies



Weekday Midday Frequencies

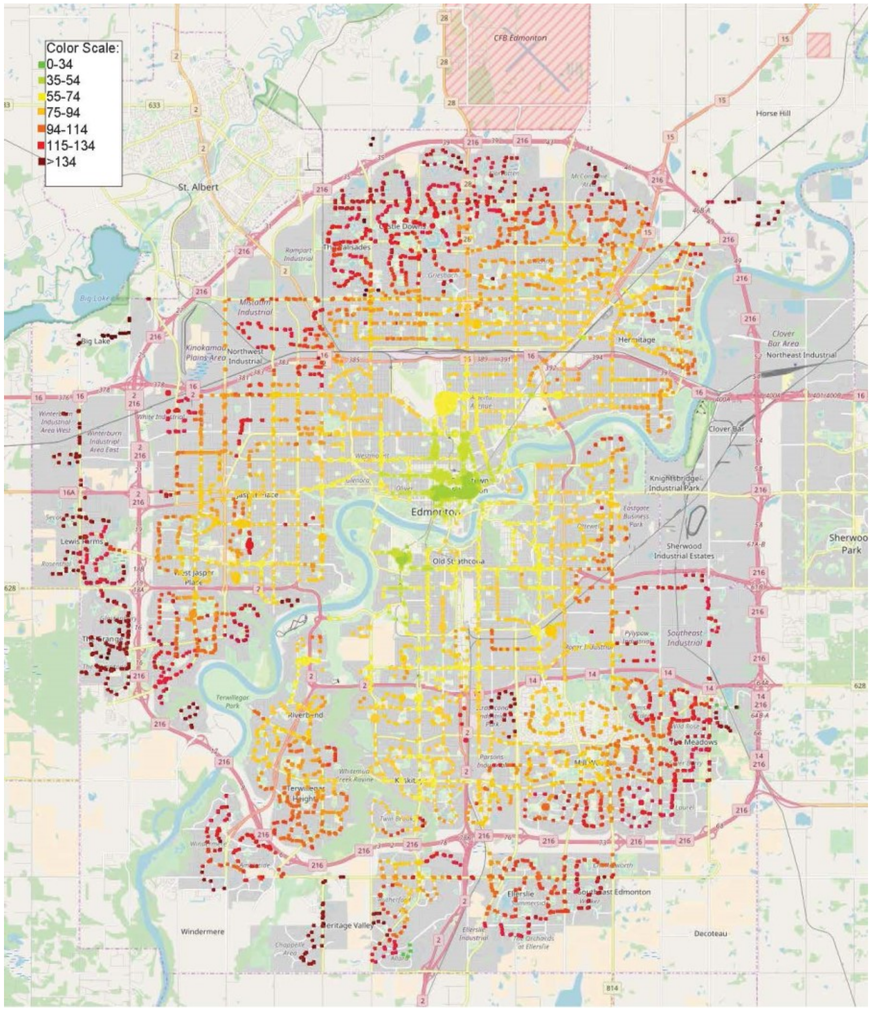


Saturday Frequencies

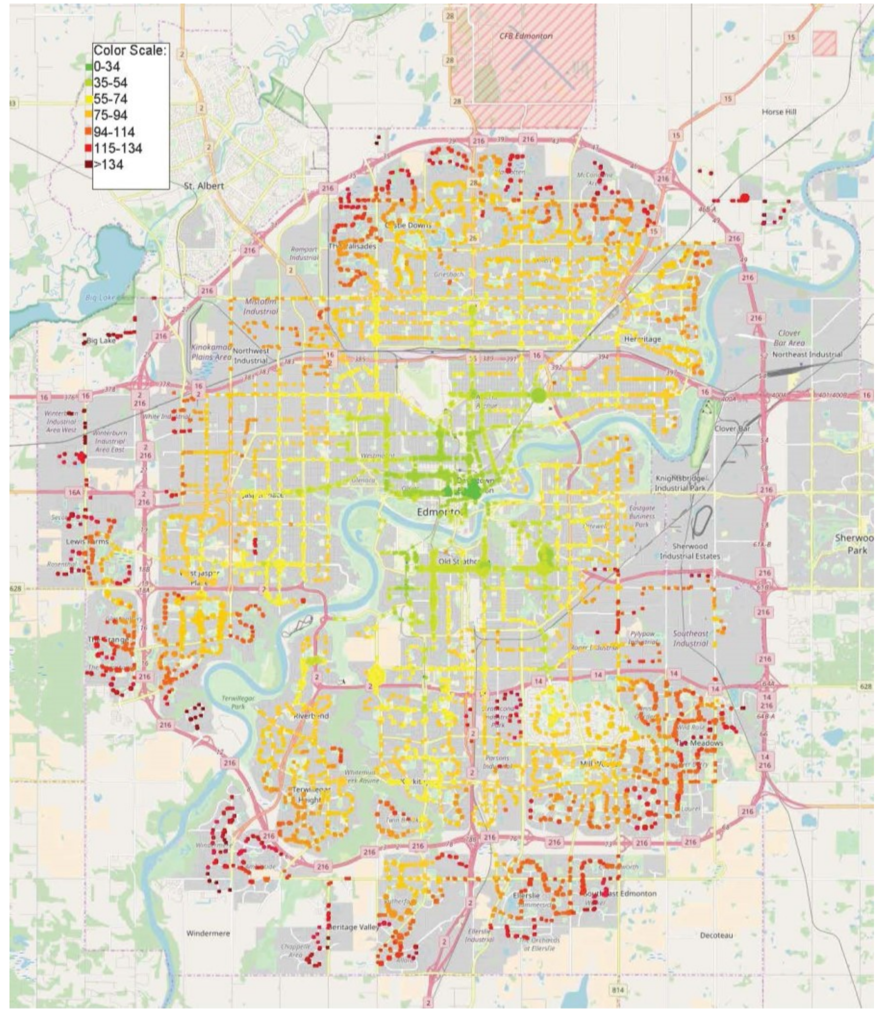


Sunday Frequencies

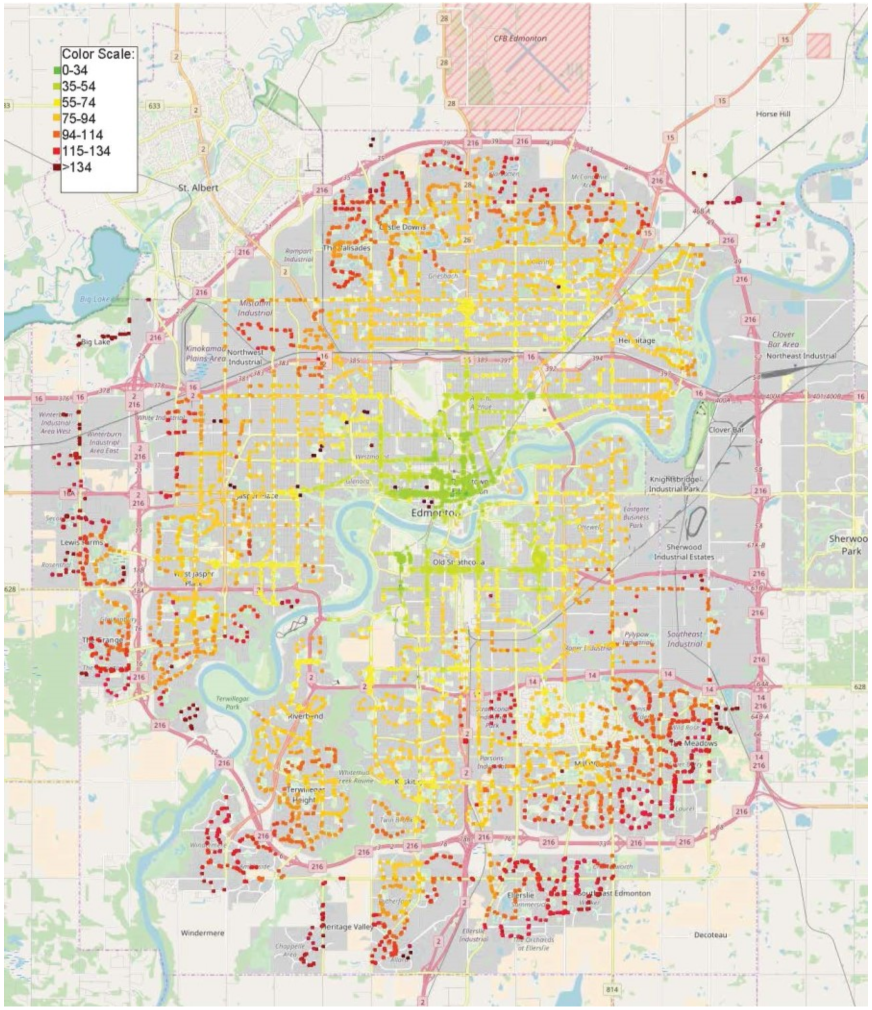
7.4 Effective Travel Time Maps



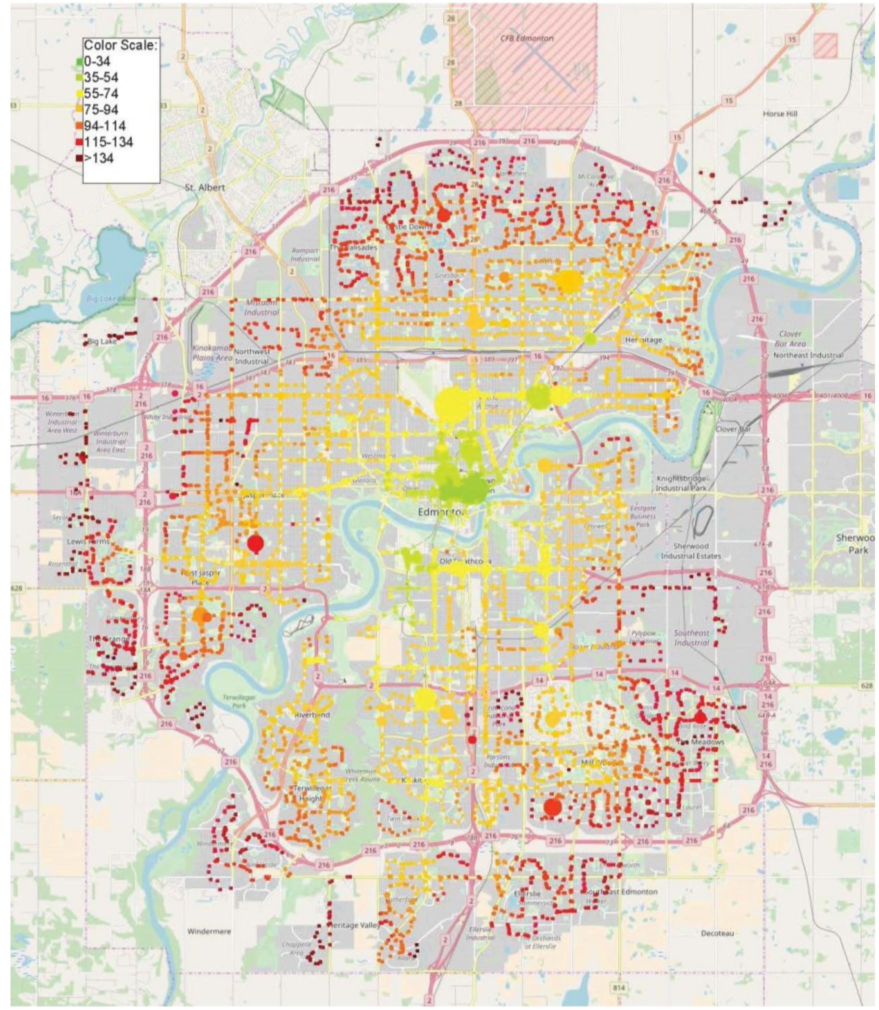
AM Peak Effective Travel Times-Arriving



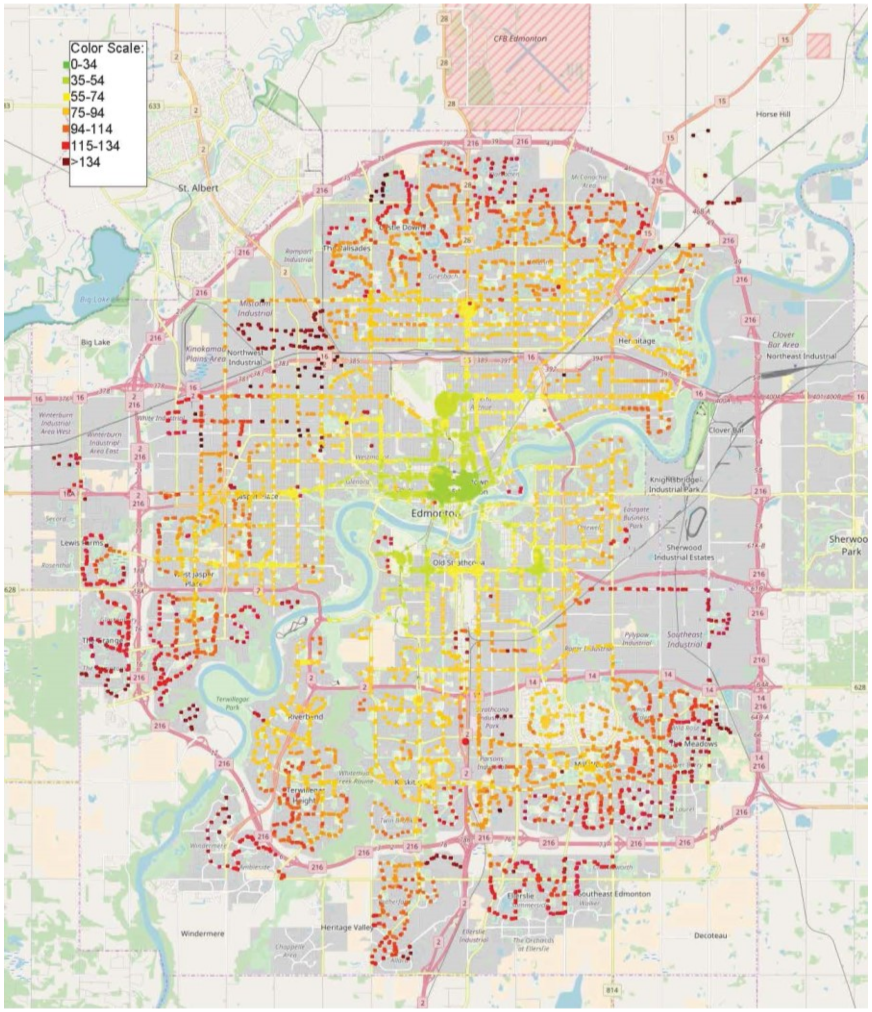
AM Peak Effective Travel Times-Departing



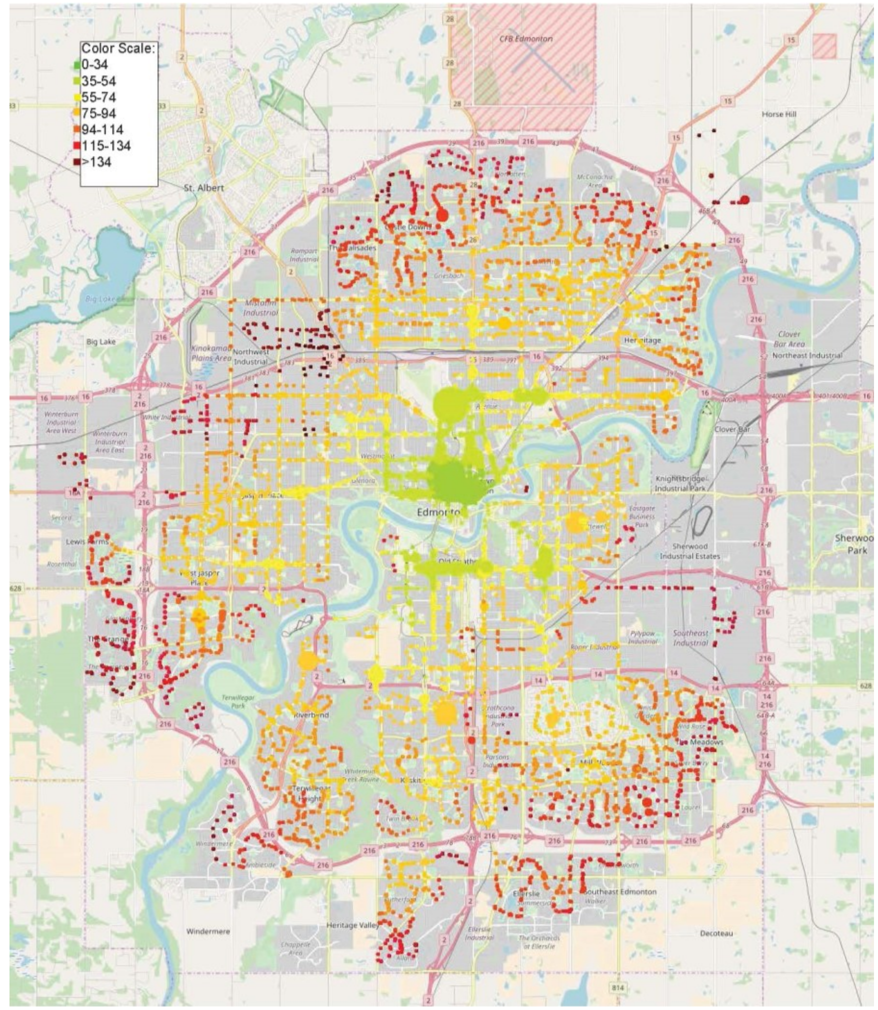
PM Peak Effective Travel Times-Arriving



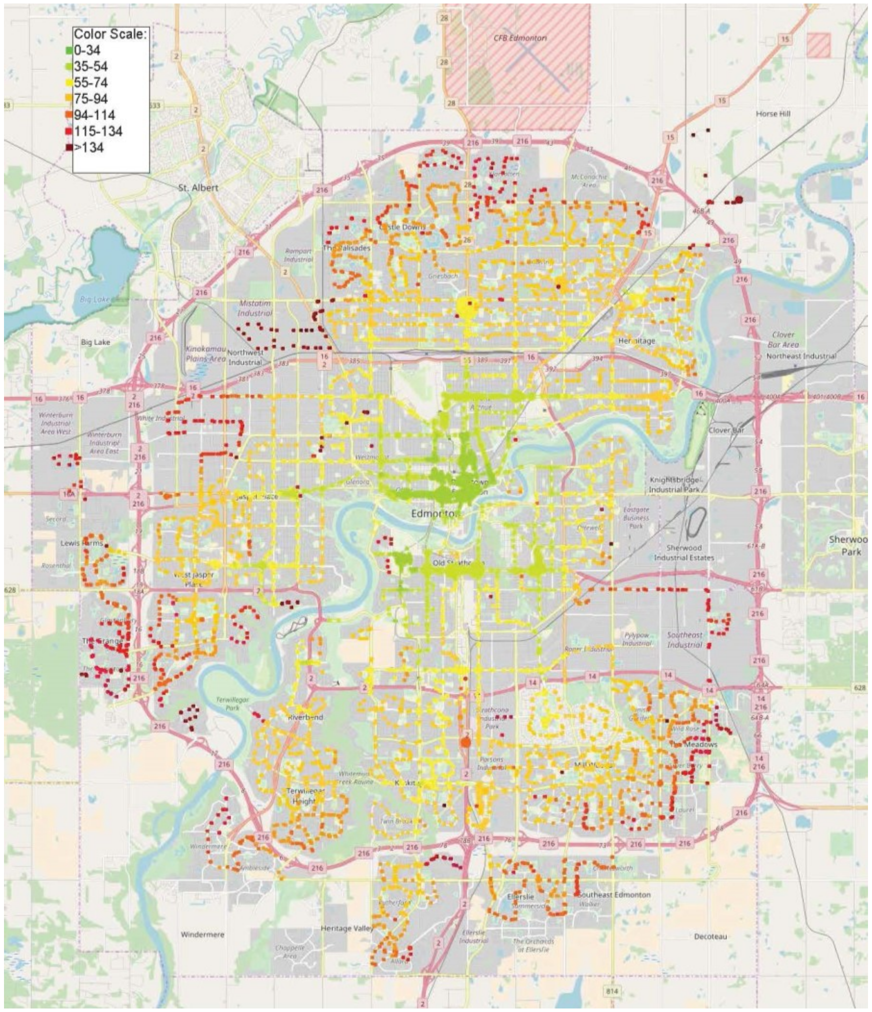
PM Peak Effective Travel Times-Departing



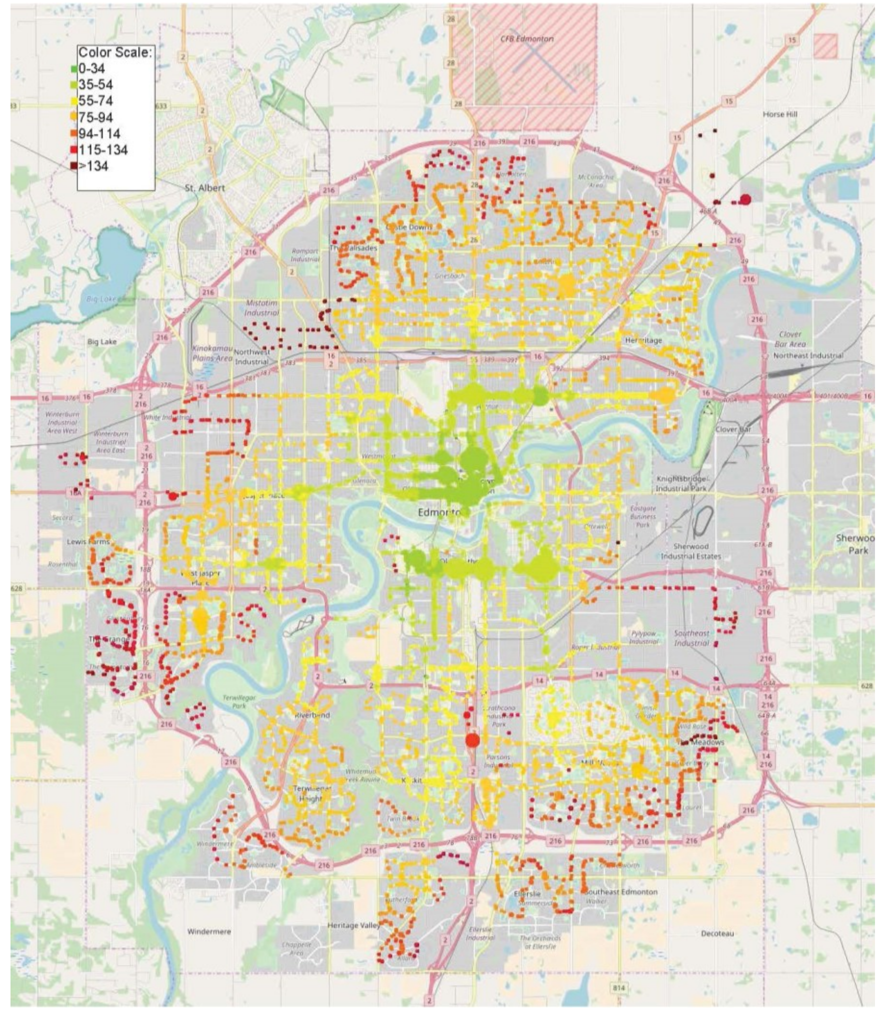
Weekday- Midday Effective Travel Times-Arriving



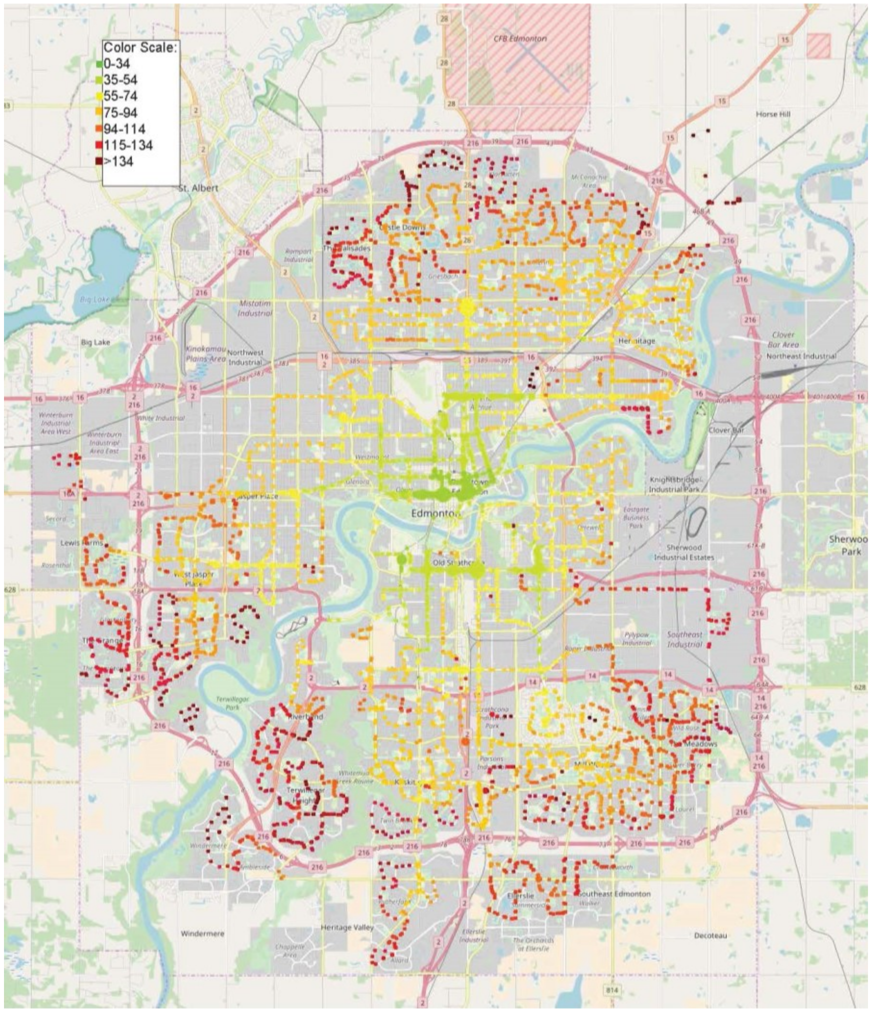
Weekday- Midday Effective Travel Times-Departing



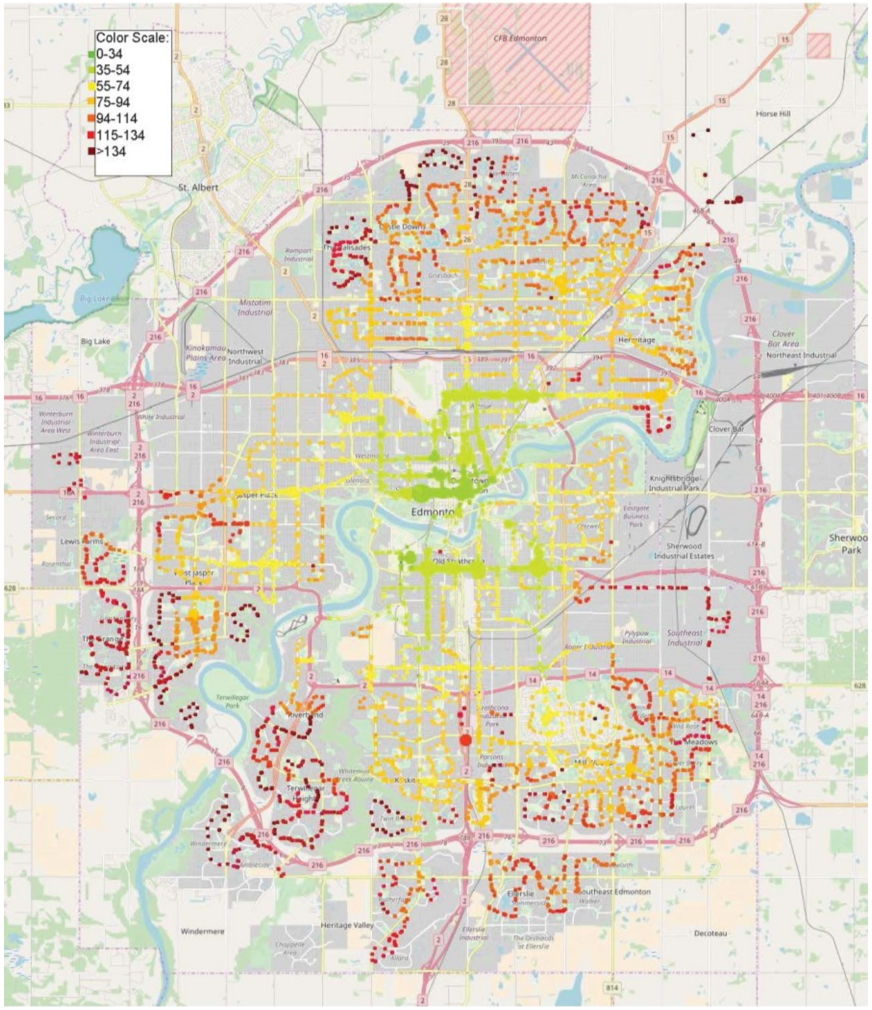
Saturday Effective Travel Times-Arriving



Saturday Effective Travel Times-Departing



Sunday Effective Travel Times-Arriving



Sunday Effective Travel Times-Departing