University of Tartu

Faculty of Science and Technology

Institute of Ecology and Earth Sciences

Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Using public bus data to assess urban seasonal rhythm of commuters in Tartu County

William Gyamfi Kumi

`

Supervisor:

Anto Aasa, PhD

Tartu 2022

Abstract

Using public bus data to assess urban seasonal rhythm of commuters in Tartu County

Urban system is an important concept for spatial planning. Knowledge of the underlying and existing spatial structure of transport networks is vital for transport planners and policy makers to understand commuters' behaviour over time to make future informed decisions.

This master's thesis sought to characterise the spatial structure and hierarchy of public bus transportation, assess variations in the seasonal rhythm of commuters and how it influences travel distance in Tartu County. To achieve this, bus data was sourced from MTÜ Tartumaa Ühistranspordikeskus. Tartu County settlement units were used as spatial units for the analysis which was sourced from the Estonian Land Board Geoportal. Population data for 2019 was also sourced from estat.stat.ee. The analysis utilised the spatial interaction theory (by Limtanakool et. al., 2007) to describe the spatial structure and hierarchy of public transportation in Tartu County. Also, an assessment was carried out on the variations in the seasonal rhythm and how it influences travel distance in Tartu County using exploratory factor analysis.

The results showed that Tartu County exhibited a central settlement unit which is Tartu city (dominant) while the other settlement units acted as peripheral areas (non-dominant). Seasonal rhythm was observed and a distinct difference in increased commuter trips from the peripheries to Tartu city was evident during the summer period from June to August while commuters from Tartu city to peripheral areas decreased. The results also showed that maximum inflows from the periphery to the centre occurred at short to relatively shorter distances whiles the least number of inflows were seen at distances further away from the centre with the exception of some settlement units. Likewise, a similar phenomenon was seen from the outflows. This can be attributed to round trips throughout the year.

Keywords: Urban system, centre, periphery, exploratory factor analysis, S-three spatial interaction, seasonal rhythm

CERCS code: S230-Social geography

Annotatsioon

Busside avalike andmete kasutamine Tartumaa pendelrändajate hooajalise rütmi hindamiseks

Linnasüsteem on ruumilise planeerimise oluline kontseptsioon. Teadmised transpordivõrkude peamistest ja olemasolevastest ruumilistest struktuuridest on olulised, et transpordiplaneerijad ja poliitikakujundajad mõistaksid pendelrändajate käitumist läbi aja, et teha tulevikus teadlikke otsuseid.

Käesoleva magistritöö eesmärk oli iseloomustada ühistranspordi bussiliikluse ruumilist struktuuri ja hierarhiat, hinnata pendelrändajate sesoonse rütmi varieeruvust ja selle mõju sõidukaugustele Tartumaal. Eesmärgi saavutamiseks saadi bussiandmed Tartumaa Ühistranspordikeskuselt. Ruumiliste üksustena kasutati Tartumaa asutusüksusi, mis saadi Maa-ameti geoportaalist. Analüüsis kasutati Tartumaa ühistranspordi ruumilise struktuuri ja hierarhia kirjeldamiseks ruumilise interaktsiooni teooriat (Limtanakool *et al.,* 2007). Kasutades uurivat faktoranalüüsi hinnati sesoonse rütmi kõikumist ja seda, kuidas see mõjutab reisikaugust Tartumaal.

Tulemused näitasid, et Tartumaal on keskne asustusüksus, milleks on Tartu linn (dominantne), samas kui teised asustusüksused käitusid äärealadena (mittedomineerivad). Jälgiti hooajalist rütmi ja täheldati selget erinevust, et äärealadelt Tartu linna sõitjate suurem arv ilmnes suveperioodil juunist augustini, samal ajal kui Tartu linnast äärealadele sõitjate arv vähenes. Tulemused näitasid ka, et enim tuldi äärealadest keskusesse lühikese või suhteliselt lühikese vahemaa tagant, samas kui kõige vähem tuldi keskusesse kaugemal asuvatest asutusüksustest mõne üksiku erandiga. Sarnane nähtus esines ka keskusest välja minemise puhul. Seda saab omistada edasi-tagasi reisidele aastaringselt.

Märksõnad: linnasüsteem, keskus, perifeeria, uuriv faktoranalüüs, S-kolme ruumiline interaktsioon, sesoonsus.

CERCS-kood: S230 – Sotsiaalne geograafia

Introduction **4**
1. Theoretical overview **6**
1.1 Importance of public transport 6
1.2 Bus connectivity and its importance 6
1.3 Related works 7
1.4 Urban systems 10
1.5 Urban System Concepts and Theories 10
1.5.1 Central place theory 10
1.5.2 Distance decay 12
1.5.3 Lorenz curve 14
1.6 Mobility as a characteristic of an urban system 15
1.7 Background - the three spatial dimensions of spatial interactions 16
1.8 The three-S dimensions of spatial interactions 16
1.8.1 Strength of interaction 16
1.8.2 Symmetry level 16
1.8.3 Structure of the System 16
1.9 Factor Analysis 19
1.10 Public transport data collection methods 19
1.10.1 Travel surveys 19
1.10.2 Ride checks 20
1.10.3 Automatic Passenger Counting Systems (APC) 21
1.10.4 Automatic Vehicle Location Systems (AVL) 22
1.10.5 Automatic Fare Collection Systems (AFC) 22
1.10.6 Electronic Ticketing system 23
2. Data and Methodology **25**
2.1 Study area 25
2.2 Geography of Tartu County 26
2.3 Data 26
2.5 Exploratory factor analysis procedure 33
3. Results **37**
3.1 Hierarchical structure of Tartu County 37
3.2 Seasonal rhythm 38
3.3 Seasonal rhythm in relation to distance 42

4. Discussion **45**
5. Conclusion and Recommendation **50**
5.1 Conclusion 50
5.2 Recommendations 51
Summary **52**
References **57**

List of figures

Figure 1. Central place model… 12
Figure 2. Distance decay curve.. 14
Figure 3.Lorenz curve 15
Figure 4 Archetypal networks: directional and non-directional interaction 18
Figure 5. APC collection systems 22
Figure 6. Tartu GoBus AFC system 23
Figure 7.Sample mobile e-ticket 24
Figure 8.Map of Tartu County settlement units 25
Figure 9. Workflow of analysis 29
Figure 10.Factor analysis flow chart... 33
Figure 11. Map of settlement hierarchy and number of commuters of Tartu County 38
Figure 12. Factor 1 showing commuting outflows variation from Tartu city to peripheral area in summer period 39
Figure 13. Map showing summer outgoing trips 40
Figure 14. Factor 2 showing incoming variation flows from the peripheral regions to Tartu City during summer period 41
Figure 15. Map of incoming trips from peripheries to Tartu city 41
Figure 16. Graph showing number of commuters in relation to distance 43
Figure 17.Map of Tartu County showing settlement units without bus connectivity in 2019 44

List of Tables

Table 1.Summary statistics of current eight municipalities 26
Table 2.Sample bus data 27
Table 3. Sample spatially joined data 30
Table 4.Sample aggregated origin-destination data for spatial structure and hierarchy settlement 31
Table 5. Sample communality and uniqueness values 35
Table 6. Sample factor 1 loadings from dataset 39
Table 7. Sample factor 2 loadings from dataset 42
Table 8. Sample summer trips from Tartu city to peripheral areas and peripheral areas to Tartu city 43

Introduction

The ease with which people and goods move about has been made possible due to transport, providing wide range accessibility to services. People's everyday activities are linked to certain locations, which they move in between on a regular basis and for specific reasons at specified times The transportation system, which makes possible the movement of people from one place to the other owing to its ability to provide links between places or different destinations, has a considerable impact on movement of people (OÜ Stratum, 2011).

Public bus operations play a vital role in providing transportation for travellers all over the world. Buses now supply most of the public transportation, although it might not be the case in the largest city in Estonia due to multiple means of public transportation. Despite this, the transit bus remains the most popular transport mode which transports most passengers each year (Dockendorf et al., 2000). In Tartu County, public bus transport is the main source of transportation for most of the dwellers.

Over the last few decades, urban development has shifted away from the monocentric model created by von Thünen (Roca Cladera et al., 2009) with cities having many trip-generating patterns which are clustered between the non-dominant nodes outside the conventional dominant nodes.

The initial monocentric structure of major metropolises tends to disintegrate progressively over time into a polycentric structure as they expand in size. The central business district (CBD) loses its prominence, and clusters of travels are spread around the city. Most of the cities function in a hybrid fashion, with some of the journeys being radial and conform to the monocentric model, while others have random origins and destinations and fit to the polycentric model (Bertaud, 2002).

A change in urban spatial configuration has occurred and has been seen at various geographical scales in the last few decades. The reality of urban growth has shown significant changes in the organization of metropolitan regions that the conventional model cannot predict (Roca Cladera et al., 2009). In recent times, the dimensions of urban nodes are usually defined and studied by network interactions and flows; these attributes are a result of what moves through cities than what might be fixed inside of them (Smith & Timberlake, 1995).

Though there are many types of flows which could be used to study the relationship between spatial units (for example, information, money), this research focuses on human interaction because face-to-face interactions remain significant for the growth of urban systems (Smith & Timberlake, 2001). Furthermore, the decision to focus on human flows was driven by the fact that human interaction faces a different set of constraints (for example, physical distance) than other types of flows such as information and money, which are considered instant (Castells, 1996).

Consequently, problem of connectivity arises from human interaction and rapid urban sprawl. For instance, in Tartu County, it is generally impossible to provide acceptable public transportation schedules for everyone due to low demand and sparse population (OptiTrans,2018). This problem continues to persist all year round and affects the mobility of inhabitants in these areas of the County. Additionally, there is also the problem of changes in commuter behaviour during different times of the year due to seasonal changes in demand.

In this regard, it is important for transport planners and policies makers to understand the underlying and existing spatial structure of transport networks to make future informed decisions.

Limited studies have been conducted on transport networks at the County level therefore, this research takes a different approach in that origin-destination data from public bus transport in Tartu County will be analysed using data from GoBus AS, an authority responsible for public bus transport within and outside settlement units in Tartu County. The structure of Tartu County on the other hand will be described using Limtanakool et al (2007) spatial interaction theory.

In light of the above, the main aim of this research is to characterise the spatial structure and hierarchy of public bus transportation, identify variations in distinct seasonal rhythms of commuter trips as well as the influence of seasonal rhythm on distance trips.

Knowledge of these phenomena is beneficial for urban transport planning in the county due to its ability to provide relevant information, needed to make informed decisions by authorities in the event of the revision of existing transport plans and policies or the formulation of new urban policies. This study will also prove beneficial in coping with seasonal demands in the transport needs of people in the county. The findings and methods of the research can be applied at different spatial scales such as inter-metropolitan levels. To achieve this, the following research questions will help to arrive at the main aim of the study:

1. What is the spatial structure and hierarchy of public transportation in Tartu County?
2. What is the seasonal rhythm of commuters in and out of Tartu city?
3. How does seasonal rhythm affect travel distance?

In chapter two of the thesis, an in-depth overview of urban systems, public transportation and factor analysis concepts are explained. The third chapter of the research consists of the data and methods used in analysing the dataset and the last part consists of discussion and conclusion.

1. Theoretical overview

1.1 Importance of public transport

Public transportation helps to strengthen the labour market and the economy while also lessening the isolation of rural and peri-urban communities (European Commission, 1998). It is critical in ensuring that persons from low-income and other disadvantaged groups have access to jobs and services. Buses now supply most of the public transportation, although it might not be the case in the largest city in Estonia due to multiple means of public transportation. Despite this, the transit bus remains the most popular transport mode which transports most passengers each year (Dockendorf et al., 2000).

The lack of access to transportation in rural areas, as well as in peripheral metropolitan areas, can lead to exclusion from public services, recreational activities, work, and education. Affordability can be a substantial barrier to use, in addition to the practical availability of services supplied by public transportation systems. People who do not have access to public transportation are unable to: go to another location; arrive at a specified time; and participate in social activities, possess the financial means to travel; make use of existing mobility options (DETR/TRaC 2000; Titheridge, 2004).

As a result, major cities around the world have faced a shared problem: connectivity. The bus service system is a key means of public transportation that draws a huge number of people due to its ease and low cost. To increase the service efficiency of public transportation systems, urban policymakers must grasp the spatial design of urban bus systems when optimizing bus stops and lines (Wu et al., 2018).

1.2 Bus connectivity and its importance

Cities, which serve as centres connecting markets such as labour, investment, education, commerce, recreation, and health care, among others, rely on the accessibility of these markets to thrive. The availability of an efficient and effective transportation system, in turn, influences accessibility. Effective transportation is one that can meet the many and varied needs of urban mobility, such as reducing travel time between different areas while also internalizing externalities to improve the general state and quality of life of the residents who benefit from it (UNEC, 2015). Since people and cargo must be carried from one location to another, spatial interaction activity is a typical occurrence in transportation systems. As a result, several forms of transportation network systems, such as urban bus or metro networks, train networks, maritime networks, and aviation network systems, are created.

In human geography, therefore, the study of bus connections in space is a common topic. When looking at the spatial structure of a bus network system, selecting acceptable investment locations to dramatically increase less attractive areas is a crucial challenge. Many scholars, on the other hand, have succeeded in conducting extensive studies and discoveries to explain transit connectedness through spatial interaction. In view of this, various network models have shown to be a useful tool for expressing and retrieving latent topological features of public transportation networks. Places and interaction activities are, therefore, represented as nodes and edges in a network system, in most studies, respectively for effective analysis of bus connectivity or network (UNEC, 2015).

1.3 Related works

It is proposed that a more significant regional geography might be built on the growing literature on urban systems, which focuses on nodal concentrations and their interconnection. The relevance of city size and functional variety is highlighted by urban geography (Simmons, 1981). Regional geography, conventional macro-scale urban studies, and regional economics all contribute to the concept of an "urban system." Duncan et al., (1960) coined the phrase as part of their attempt to establish a functional organisation on the United States urban areas. In his research of empirical regularities in regional and national urban areas, Berry (1964) pioneered the General Systems Theory to urban geography, and consolidation of US data to expand the theories further accompanied by the study on nationwide journey-to-work patterns by Berry et al. (1968). In order to assess particular links inside the system, one can also utilise its hierarchical structure (Simmons, 1981). The method concentrates on the function of urban regions within the urban system in totality including their specialities, important links, and response to development and modification once the urban system has been delineated. Every city's connection with other cities is significant, and researchers are interested in the nature of each link, its significance, and the quantity and spatial expanse of all interactions (Simmons, 1981).

Several researchers have used different methods to address issues or describe some phenomena in various urban systems. Several studies exist in the literature on various aspects of urban planning, including transportation and traffic issues (Abrate et al. 2009; Ubbels and Verhoef, 2008; Boyce, 2007) as well as urban growth and land development (Sims and Schuetz, 2009; Wang et al. 2009; Wu, 2009; Xu et al. 2009; Jenkins, 2001; Yeh and Wu, 1996). The data on urban planning that is generally accessible is multivariate data. As a result, multivariate approaches to analysis are feasible. Principal Component Analysis (PCA), Cluster Analysis (CA), and Factor Analysis are the most often utilised methodologies for urban planning.

In their analysis, Hamidi and Ewing (2014) employed cross-sectional data from major cities in the United States in 2010 (162 urban areas). As a criteria, a population of more than 200,000 people in 2010 was used to identify significant urban centres. They calculated the extent of sprawl in these locations by measuring 15 variables from four categories (development density, land use composition, centralization of activity, and access to roadways). The structural accuracy and categorisation of four components were established using Principal Component Analysis (PCA) on these 15 variables (Zerbadast & Ghanooni, 2019). Finally, the urban sprawl index was calculated by combining factor scores. They then transformed the 2010 figures into 2000 figures to get comparable figures for the year 2000. The findings revealed a general rise (minor) in urban sprawl in significant urban centres across the United States (Zerbadast & Ghanooni, 2019).

In another study, out of 14 urban transport variables, Joly (2004) used PCA to create 5 principal components which were employed to describe the relationships between urban structures and the transit time budget, which measures the accessibility of space and time. Steg (2005) proposed 33 car-attractiveness metrics based on three main components that reflect car-use intentions. Another common use of PCA in travel behaviour analysis is market segmentation, which ensures that a transportation strategy is tailored to the demands of specific travellers. In their article Wedel and Kamakura (2000), examined various market segmentation approaches in depth. In principle, market segments (meaningful user groups) are created based on essential consumer attributes. Contrary to a priority segmentation, such as expert-based segmentation, PCA uses a systematic, data-driven technique to discover essential travel behaviour traits. In one of such studies, Anable (2005) studied the relationship between personal attitudes and travel behaviour, focusing on the method of transportation. PCA was used to analyse 105 attitudinal statements, yielding 17 significant components. As a result, six unique market categories representing the respondent's car-orientation could be created. Wittwer (2014) employed 15 travel activity factors such as frequency, timing, and length to create 8 principal components that enabled young adults in Germany to be segmented into six groups. In a study by Kandt et al. (2015) 63 attitudinal assertions about the participant's relationship to sustainable and information technologies were turned into 13 primary components, allowing for the creation of six mobility profiles for each traveller in the studied regions. The preceding scenarios demonstrate that factor analysis is a useful and significant statistical tool for studying travel behaviour.

Limited research is available on the assessment of urban seasonal rhythm of commuters. However, similar studies on urban patterns, mobility and seasonality explain some works which are related to this study.

The creation of spaces that stimulate outdoor activities in both the winter and summer is a major urban planning problem in winter cities. Understanding how changes in weather due to climate change may affect people's soft mobility choices is a closely connected challenge (Chapman, et, al., 2019). The need for understanding how the interaction between urban form, weather, seasonal fluctuations, and climate change affects human outdoor activity was the rationale for their research. However, the emphasis on outdoor activity was the problem in their study because of the concern that people invest a low percentage of their time outside in the winter. Their paper therefore delved into connectivity for soft mobility in the winter. Collaborating with residents of Luleå, Sweden's subarctic city, the study investigated how the combination of the built environment and the winter season influenced people's utilization of the outdoor environment. Document studies, questionnaires, mental mapping, and picture elicitation were among the study tools used to arrive at their objectives. These were employed to learn more about people's perceptions of soft mobility in the winter. The research was conducted on three scientific levels. The study's primary purpose was to determine the link between the built environment and people's soft outdoor mobility in the winter. The second objective was to learn how the season, climate, and weather affect connectivity for soft mobility in the winter. The third purpose was to come up with new innovative ideas on how the urban form could be planned and structured to improve outdoor soft mobility during the winter season. The results and discussion established the fact that the interaction between the urban form and the winter season is a feature of urban morphology in winter settlements and could be an integral process in constructing the public space and its connectedness for soft mobility in winter. (Chapman, et, al., 2019).

Gan, et, al., (2020) posited that resident movement could be investigated from a macro perspective using smart card data collected from public transit's automatic fare collection (AFC) systems. Varied land uses generate different traffic rhythms, reflecting changes in human activity patterns. As a result, comprehending daily ridership and mobility patterns necessitates comprehending the relationship between daily ridership patterns, station characteristics, and their immediate environment. The goal of their research was to develop a framework for recognising urban mobility patterns and dynamics from a spatio-temporal viewpoint, as well as to highlight the connections between mobility and land cover/land use (LCLU). Based on one month's transactions data from Nanjing metro's AFC system, 110 metro stations were divided into seven clusters: employment-oriented stations, residential-oriented stations, spatial mismatched stations, just to mention a few, each with its own ridership pattern (combining boarding and alighting). To test if the clustering results were plausible, a comparison of the peak hourly ridership of the seven clusters was made. Ultimately, the link between local environment characteristics and cluster membership was estimated using a multinomial logit model. Their findings showed that clustering based on ridership patterns produces meaningful interpretable clusters, and that there are substantial correlations between local LCLU characteristics, distance from the city centre, and cluster membership. Their findings and analytical methodology could help improve the efficiency of public transportation and urban planning (Gan, et, al.,2020).

1.4 Urban systems

Berry (1964) and Pred (1977) were the pioneers who introduced the concept of urban systems as a set of functionally interdependent cities. In this study, an urban system refers to a set of interdependent geographical settlement units (SU's) that are connected to each other through flows. However, the form of these urban systems can range from entirely monocentric to fully polycentric (Batten, 1995). The main concept of the monocentric city is that majority of economic activities take effect in the urban centre, while the periphery solely provides a residential purpose. As a result, under the monocentric model, the connection between the urban centre and its suburbs is hierarchical–nodal or centralized in view that most commute flows are oriented from the suburbs into the major cities(de Goei et al., 2010).

On the other hand, research shows that polycentric cities have a multiscale and multi-layered structure (Meeteren, 2018), however, it typically refers to a more balanced urban structure with numerous autonomous urban "centres" of relatively similar "importance"(Kloosterman, R. C. and Musterd, 2001). As nodes participate more actively in the system of interaction, the variations in interaction among nodes reduce over time, and the system evolves toward a completely polycentric structure(Limtanakool et al., 2007a).

1.5 Urban System Concepts and Theories

1.5.1 Central place theory

Walter Christaller (1933) is credited with pioneering the theory by making comprehensive observations of urban hierarchies and then attempting to model them. The core assumptions are that consumers are evenly spread across the country, while businesses are concentrated in cities. Cities constitute a hierarchy in which higher-ranking cities generate all of the items produced by cities one level down in the hierarchy, and one further (Berliant, 2005). A central place, according to Walter Christaller, exists primarily to supply goods and services to its nearby population. In effect, the city serves as a distribution centre. Any activity carried out in the urban area that receives at least some of its support from persons living in the surrounding rural areas is referred to as a central place function. The concept of functional interdependence between a town and its surrounding rural area was central to Walter Christaller's theory (King, 2020).

Walter Christaller had to construct a set of assumptions in order to focus on the economic components of his theory. First, Christaller assumed that the area in the locations he was conducting the research would be flat, with no obstacles obstructing people's mobility. That is, there was a limitless and homogeneous plain with the same soil fertility and other natural resources throughout. This plain was equally inhabited, and farmers had the equal levels of income and demand for goods and services throughout. Travel across the plain was accessible in all directions, and the cost of travel and freight transportation was solely determined by the distance travelled (King, 2020).

Furthermore, Christaller assumed that both farmers as customers and businesspeople in cities as producers of products and services were rational individuals who would aim to reduce their costs (whether transportation or production costs) and maximise their profits. In the consumer's perspective, this would mean that they would only travel to the nearest central location that offered the goods and services they required. Businesspeople understood this to mean that if a profit could not be made, a product or service would not be manufactured or sold. It was expected that if there was enough demand, for instance, for them to at least break even, they would not provide the service or manufacture the product. Christaller also assumed that all of the settled plain would be equally well served by central areas, which is tied in part to the assumption of rational behaviour and also to the notion that new enterprises may show up wherever and whenever they desired (King, 2020).

In addition, two assumptions regarding human behaviour were made. The first is that humans will always buy items from the closest shops that sells them and secondly, when there is a great demand for a certain good, it will be made available in close proximity to the people. When demand for a product decreases, so does its availability (Christaller, 1933).

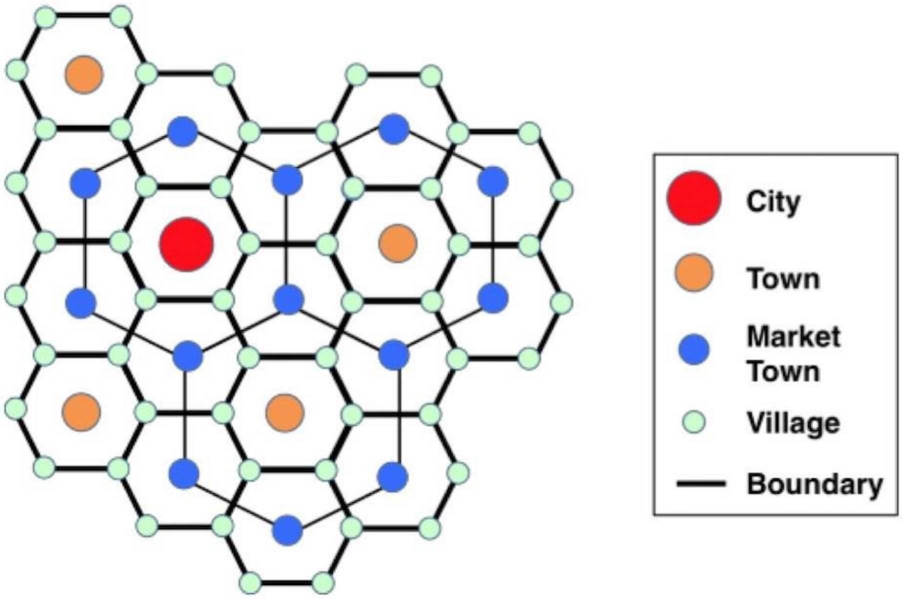


Figure 1. Central place model Source: (Haseeb, 2017)

1.5.2 Distance decay

In geography, the term "distance decay" has long been used to characterize the impact of distance on cultural or geographical relationships between locations (Hanks, 2011). Although the phrase can be applied to ecology or the habitat of plant and animal species, it is used to describe human activities in the disciplines of urban, economic, and cultural geography, in other words, a concept that connects geography to the occurrence or frequency of a pattern of activity (Pun-Cheng, 2016). Distance decay is a key concept in spatial analysis, particularly for spatial interaction models and conceptions of cultural diffusion (Hanks, 2011). The effect of mere physical distance has been decreasing with the development of transportation and communication technology, and a new perspective for the fundamental variables influencing the decay of spatial relationships is thus required (Pun-Cheng, 2016).

The concept is connected to Waldo R. Tobler's first law of geography, "All things are related but near things are more related than far things." *Distance* has traditionally been defined as the distance between two points on a physical map or, more accurately, the distance between two points on a topographic surface (Pun-Cheng, 2016). Considering a flat surface with equal accessibility in all directions from a centre point, lines of equidistance would form a circular pattern around that point. This was a major presupposition in the von Thünen rural land-use model and Christaller's central place theory at a period when the road network was still in its infancy and animal transportation still reigned supreme (Hanks, 2011). When two locations are far apart, the phrase *decay* refers to the fading effect of a phenomena, attribute, or activity. Because accessibility and land value decline as one moves away from a centre, there is less interaction between remote regions (Pun-Cheng, 2016).

Distance decay is commonly expressed graphically by a curving line that swoops concavely downward as distance along the *x*-axis grows. It can be expressed formally by the expression $I = 1/d^2$, whereby I is a measure of spatial interaction and d is distance (The Stands4 Network, 2013). Distance decay has a significant link to the gravity model in that a precise mathematical relationship exists between the strength of a spatially expressed phenomena and the distance over which it is dispersed (Hanks, 2011). Geographers who use gravity theory in their work try to build a similar model to analyse labor or trade flows, transportation, population movements, and language, religion, and technology diffusion (Hanks, 2011).

Traditionally, the rate of decay was displayed inversely with one variable distance, with the gradients varying depending on the phenomena and area under study. Until the 1960s, consumer behaviour was regarded as primarily homogeneous and deterministic, following a tight normative route outlined by the initial retail gravity models in the 1930s (Pun-Cheng, 2016). For towns with less developed transportation or more challenging terrain, land value has a steeper downward sloping curve (Pun-Cheng, 2016). For greater distances, distance decay may still be valid today. The effect of distance on human interactions has become less with the advent of transportation and communication technologies, the increasing diversity of modes of transportation, the introduction of various fare structures, and the increasing demand for better living conditions, and thus the curve deviates significantly from the traditional trend (Hanks, 2011). Models that describe human activity between locations should include not only traditional normative criteria, but also other environmental and people-based elements (Pun-Cheng, 2016).

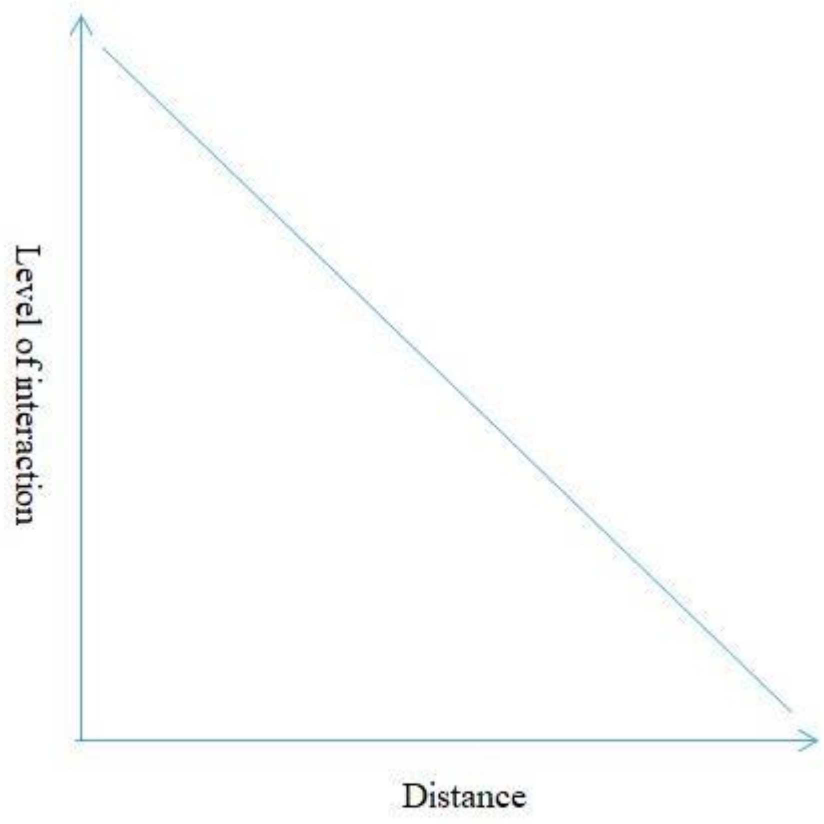


Figure 2. Distance decay curve Source: Author

1.5.3 Lorenz curve

Lorenz curve is a graphical illustration tool proposed by Lorenz (1905) which indicates what proportion of total income is held by a given percentage of the population. As shown in Figure 3 below: the percentage of the population is on the x-axis and the percentage of the income is on the y-axis. The higher the amount of inequality, the further the curve deviates from the baseline, which is represented by the straight diagonal line.

The Lorenz curve always falls below the equidistributional line which is line of equal distribution throughout the population. The Lorenz distribution has demonstrated to be a universal or fractal: it is a 'natural' scale invariant property observed in any wealth or income distribution of any population – cultural or not (Damgaard & Weiner, 2000).

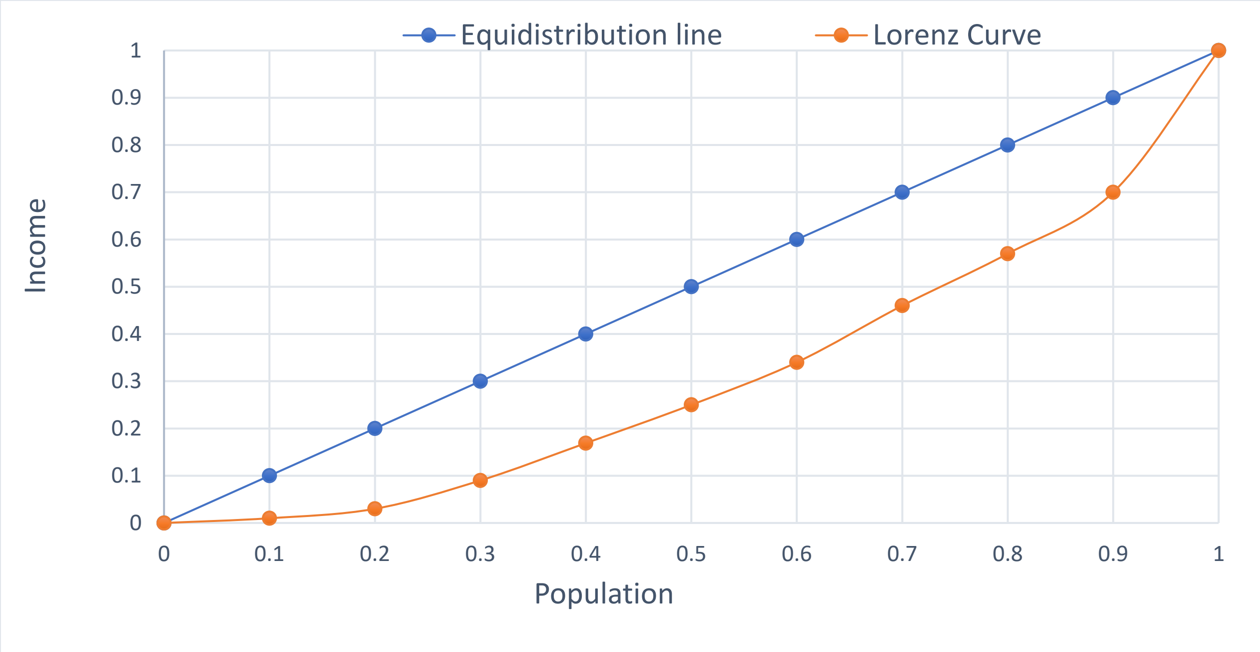


Figure 3.Lorenz curve Source: Author

1.6 Mobility as a characteristic of an urban system

Based on this premise, this research will adopt the theoretical framework for characterising an urban system based on flows which were proposed by Limtanakool et al., (2007), the theory of S-three spatial dimensions. Although their framework concerns spatial interactions with regard to urban systems at the metropolitan level, this theory will be used in relation to the settlement units in Tartu County as the theory can be applied to other spatial scales. Limtanakool et al., (2007) mentioned that according to Pred (1977), spatial interactions occurred in various ways including flows of people, goods, and information just to mention a few. In that regard, their focus was on the flow of people who travelled between distinct functional urban regions as less frequent journeys (daily commuting) occurring over greater distances were important in urban system development (Dieleman & Faludi, 1998). It is in such travels that flows play a role in characterising urban systems and hence the use of this framework to characterise the urban system of Tartu County based on distinct flows between settlement units from the existing bus network. This study will therefore use the theory of three-S spatial dimensions to explain the urban system of Tartu County.

The three-S dimensions used to characterise urban systems are Strength of interaction, Level of Symmetry and Structure of the System. Urban systems can be classified as monocentric or polycentric (Avid & Batten, 1995). Monocentric systems are one or a few nodes being dominant based on the concentration of certain specialised functions. Fully polycentric systems, on the other hand, refer to a situation where there is a lack of dominant nodes due to the distribution of specialised functions across urban centres (Kloosterman & Musterd, 2001).

1.7 Background - the three spatial dimensions of spatial interactions

The methodology of estimating flows between places is relevant to transportation. Such fluxes referred to as spatial interactions, are used to assess the demand for transportation services (current or projected). They include things like commutes to work, migrations, tourism, access to public facilities, the transfer of knowledge or money, retail operations, international trade, and distribution of goods. Spatial models are one of the steps used in transport models due to their ability to help in the estimation of how trips are generated and distributed. Many spatial interaction models start with the assumption that flows are a function of the characteristics of the origin destinations, destination locations, and the friction of distance between them (Rodrigue, 2020).

1.8 The three-S dimensions of spatial interactions

1.8.1 Strength of interaction

In the first dimension, which is the strength of interaction, it is assumed that when nodes have a strong interaction, it is easier to move between each other (Limtanakool et al., 2007b). Hence, the level of integration is dependent on the aggregation of all flows within the whole system (Limtanakool et al., 2007b). Bourne & Simmons (1978) posit that the building blocks of urban systems rely on the availability of strong interactions among cities and regions. In the same vein, the frequency of movements between the settlement units determines the strength of interaction between them.

1.8.2 Symmetry level

In terms of the level of symmetry, flows could be fully asymmetrical (uni-directional) or symmetrical (bi-directional). Typically, asymmetric interactions are associated with fully monocentric systems. This is because as fully asymmetrical interactions are unidirectional, it means flow will be geared towards just one important node (monocentric) where all specialised functions are concentrated. In that case, this node will be at the receiving end due to its importance with flows originating from less important cities. This is similar to settlement units which serve as the frequent destination for less important peripheral areas due to their location and jurisdiction functionality. Most symmetric interactions, on the contrary, are alluded to as polycentric systems since specialised functions are distributed across the urban area and so all nodes are interdependent (Limtanakool et al., 2007b).

1.8.3 Structure of the System

In terms of structure, an urban system can be hierarchical or non-hierarchical. Fully monocentric systems have hierarchical structures with flows moving downward from one or few major nodes to less important ones. In polycentric systems, the non-hierarchical structure is evident because there are varied nodes with specialised functions; spatial flows are seen in varied directions. They could be horizontal, diagonal, reciprocal, or in other directions. The hierarchical structure of the urban system is considered to decline as the urban system evolves toward a completely polycentric system. (Sinclair, 1983).

Dominance is a concept in the third dimension (structure), which could either factor or not factor in the direction of flows. Direction of flow is considered when spatial flows regarding the access to facilities, products, or services are connected to the geographical locations at the nodes where arrivals occur, in other words, the destination of incoming flows. This phenomenon highlights the importance of such a location in meeting the needs of people and thus attracting a lot of inflows (Alderson & Beckfield, 2004). This might not always be the case and in such instances, the direction of flow is not considered.

This study will combine the use of the three S-dimensions and the level of dominancy to explain the urban structure and seasonal rhythm of Tartu County with regard to flows that occur or exists between the distinct settlement units. A clear depiction of the various types of urban systems for four node networks in an ideal case, which spans across a fully monocentric network, defined by fully asymmetrical interactions between the dominant and non-dominant nodes, to a no dominance polycentric network with equally strong nodal relationships is seen in Fig.1. To distinguish between directional and non-directional urban systems, the relevance of the directionality of flows will be used.

The one-directional nature of monocentric networks is seen in fig.1 where A1/B1 is acting as the main node. The figure also shows a symmetrical increase for directional networks from the left to the right along the horizontal axis with the hierarchical extent of the dominant node getting weaker along the vertical axis. Interactions between non-dominant nodes are possible at the onset, but smaller than the magnitude of dominant nodes (Network A2). More symmetrical flows then emerge between dominant and non-dominant nodes (Network A3) resulting in a less pronounced strength of flow between the two (Network A4), evidence of a weakened hierarchical structure. In the end, a flat network is created because of equally large and fully symmetrical flows with all nodes gaining equal importance (Network A5), otherwise referred to as a fully polycentric network. However, (Limtanakool et al., 2007)) argue that, in reality, networks tend to fall within the extremes of a monocentric and polycentric system, but few urban systems would look like A1/B1 or A5/B4 as seen in Figure 4.

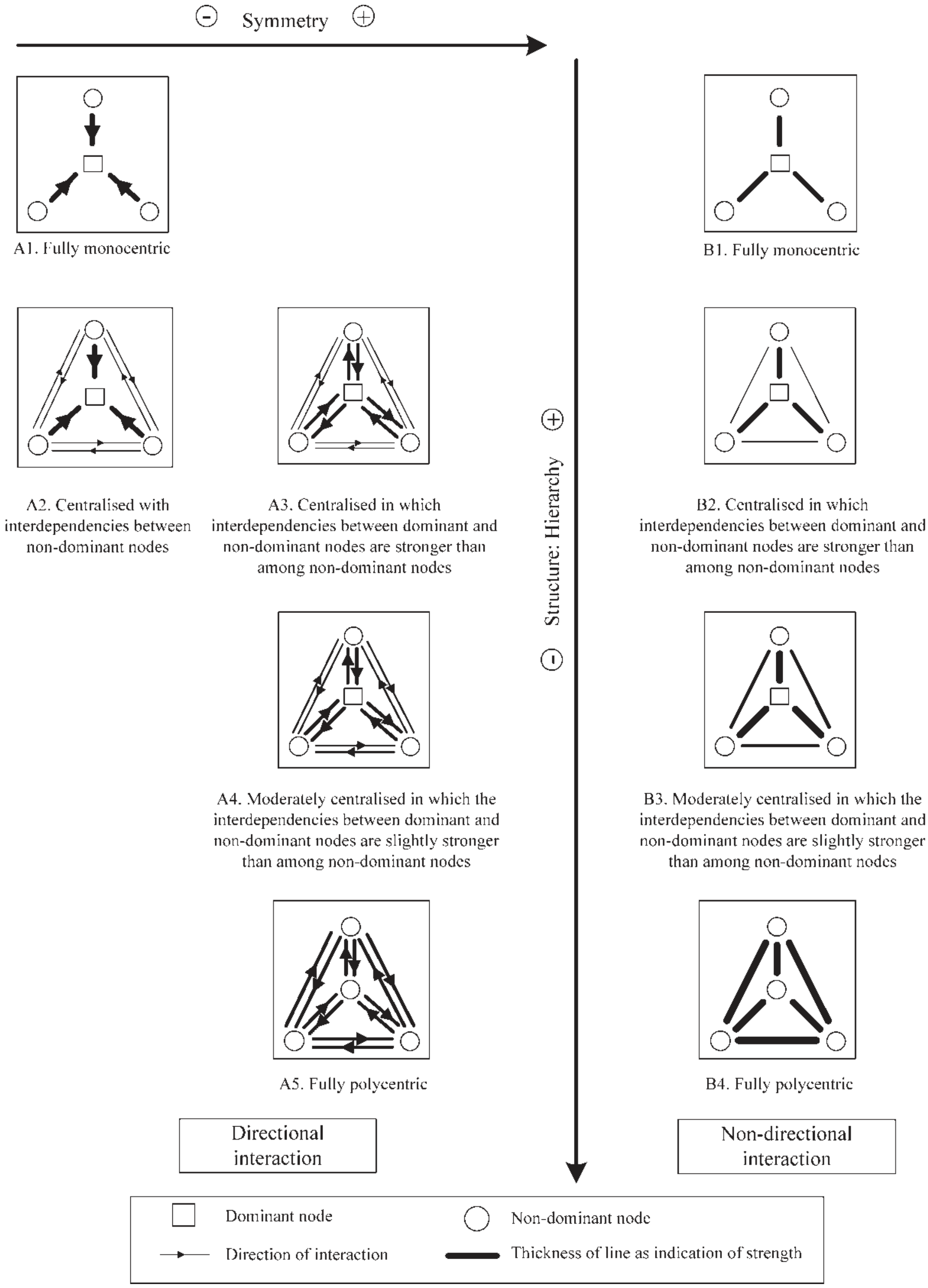


Figure 4 Archetypal networks: directional and non-directional interaction

Source: Limtanakool, et al. (2007)

1.9 Factor Analysis

The emergence of factor analysis can be attributed to the work of Pearson and Spearman over a century ago. Concisely, factor analysis can be defined as a statistical procedure that is extensively used in the fields of information systems, education, and psychology. In addition, it is regarded as the ideal method for interpreting self-reporting surveys (Bryant et al., 1999). Through factor analysis, a vast number of variables can be compressed into a manageable quantity. This also creates underlying dimensions between measurable parameters and latent constructs, enabling for the development and improvement of theory (Gorsuch, R., 1983).

Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are defined as the two main types of factor analyses (Williams et al., 2010). With regards to EFA, as the name implies, it allows the researcher to investigate the major factors in order to develop a theory from a large number of implicit variables, which are often represented by a series of items (Pett et al., 2003). In comparison to exploratory factor analysis, confirmatory factor analysis has a clearly marked out hypothesis and thus expectations about the number of constructs and which construct theories or best match based on priori model and theory (Williams et al., 2010).
To summarise, EFA is best suited for scale development and is used when there is limited theoretical foundation for identifying the quantity and patterns of common components. However, it has a limitation in that judgments about the number of elements and rotational scheme are made on the basis of pragmatic rather than theoretical grounds (Hayton et al., 2004).

1.10 Public transport data collection methods

1.10.1 Travel surveys

Travel surveys have been in existence since the 1950s. The goal is to gather data that properly depict the travel behaviour of an area's population and to analyze travel trends to guide transportation and land-use planning policies. In the 19th century, travel surveys were mostly done in bigger cities such as the United States and certain significant cities such as Melbourne in Australia (MMBW, 1951). These surveys were normally conducted as 'home interviews,' in which interviewers paid unannounced visits to families and asked them about their trip the previous day as well as demographic data (Stopher et al., 2003). This implied, the interviewer paid a personal visit to the participant's house and performed the survey there. This sort of survey allows for direct interaction between the interviewer and the participant, allowing for the chance to clarify questions, verify replies and accurately understand, and confirm answers right away. A key drawback of this survey type is that it involves a significant amount of time and employees, as well as expensive expenditures (Henk et al., 2006). Another challenge is determining the suitability and safety of a house visit. In the United States, for example, house interviews were displaced by other survey formats because it was considered that conducting interviews in particular regions of the study areas was unsafe (Stopher et al., 2003).

Face-to-face house surveys have mostly been displaced in many countries by alternative forms of surveys, such as personal delivery/pick up; mailout/telephone retrieval; mail-out/mail-back; and telephone-only surveys. Recently, certain nations and localities have begun to utilize a combination of mail-out, phone, and online surveys (Henk et al., 2006; Stopher et al., 2003). The main causes for adjustments in a survey and recruiting procedures were concerns with response rates, data quality, coding, and survey expenses. Unexpected home interviews, for example, frequently yielded untrustworthy data since participants did not recall their travels precisely. In addition, because the number of responses was too low in some areas, the mail-back option was replaced with a telephone data collection. Moreover, telephone surveys are often less expensive than home interviews or mail-back surveys, for example, due to higher response rates, which suggest lesser working hours to gather data. Not only are participants rates crucial, but so is the 'correct' mix of participants, i.e., their representativeness of the research region. Recruitment is sometimes one of the most challenging components of surveys since, in many countries, there is no publicly (or privately) accessible list of residential addresses, making random sampling problematic(Inbakaran & Kroen, 2018).

1.10.2 Ride checks

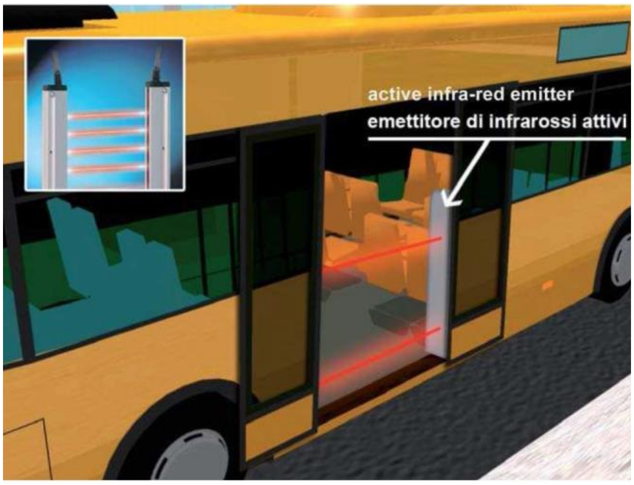
Historically, transportation agencies have relied on human ride checks to gather data onboarding and alighting operations. Manual ride checks include dispatching ride checkers to sit in public transport vehicles while they are in revenue service to watch and record passenger boardings and alighting at each stop for a random sample of one-way vehicle rides over the course of a year (Xuehao Chu, 2010). They show (1) overall boardings on route and route divisions, (2) passenger-miles, (3) peak load point location, (4) maximum load per trip, and (5) average load at the peak load point or any other location. Ride checks are useful for route and schedule design especially on lengthy or high-volume lines that are suitable for scheduling alternatives like short-turning, alternating deadheading, zoning, and giving restricted stop service. Unfortunately, ride checks are costly to do and, as a result, are performed occasionally. A typical route or schedule planner's ride-check data typically consists of a single day's data which might be several years old.(Furth & Day, 1985).

Cities are investing in technology to monitor and save data from real-world activities. With the advancement of technology in recent decades, urban transportation operators began to have the ability to collect and store reliable data on the geographic location of buses, journey times, stopping times, and passenger counts. Among these are Global Positioning System (GPS), farebox, and Automatic Fare Collection (AFC), Automatic Passenger Counting Systems (APC). Through these methods, operators enhance real-time control techniques, including the dynamic management of unexpected occurrences in operations, as well as the planning phase, which included the development of solutions tailored to the unique characteristics of each system(Hora et al., 2017).

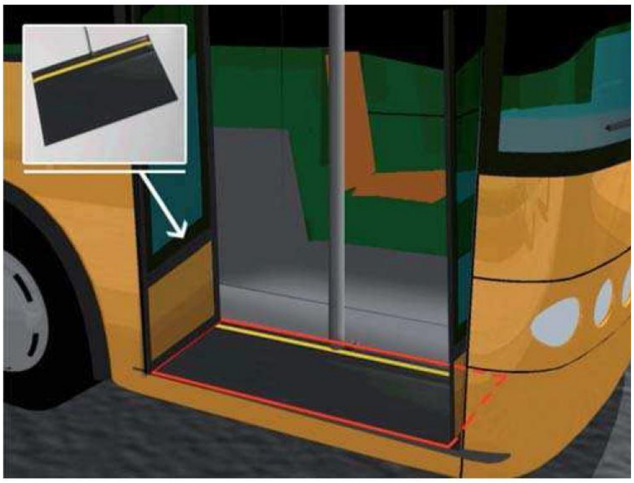
1.10.3 Automatic Passenger Counting Systems (APC)

APC devices are capable of counting passengers as they get on board and disembark from a bus, as well as record timings at each stop, and therefore give disaggregate data that is useful for service operation and scheduling (Townes & Hunter-zaworski, 1998). A myriad of competing APC systems has emerged over time such as passive infrared detectors, stereoscopic video cameras, infrared cameras, ultrasonic detectors, microwave radars, electronic weighing equipment, infrared light beam cells, piezoelectric mats, switching mats, laser scanners are a few examples of detection technologies (Kuutti, 2012). The infrared beam break or light barrier technology is the most often employed APC technology by transportation agencies (Boyle, 2008). The infrared technology employs various infrared beams that span the doors at the front and rear of the bus to provide orientation and tally of riders loading and offloading. The way the beams are broken determines the orientation of the passenger's movement (Mathews & Poigné, 2008). When a person enters, the beam closest to the entrance is broken first, and when a passenger exits, the beam farthest from the door is broken first. The data is saved during operation time until the bus returns to the garage where the day’s data is transferred to a central database. This data is kept for several years and is freely accessible (Kotz et al., 2015).

More novel APC technologies that are also capable of counting passengers such as the treadle mat sensors have emerged in recent times. Treadle mats count passengers by physical touch, utilizing integrated sensors that identify passenger entry and departures. Smart mats have optical fibres that detect deflection and record foot placement (Reuter, G, 2003). Typically, mat sensors placed on the two steps of each bus entrance assess the body mass of a single individual. When a minimum design load is applied to the treadle, these mats activate (Pinna & Chiara, 2016).



a) Infrared beam system



b) Mat sensor system

Figure 5. APC collection systems Source: (Pinna & Chiara, 2016)

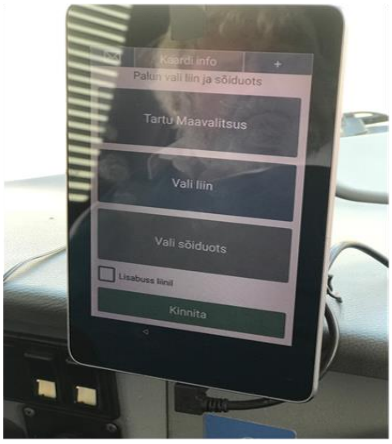
1.10.4 Automatic Vehicle Location Systems (AVL)

Recent advancements in automatic vehicle location (AVL) systems based on global positioning systems (GPS) have offered the transit sector and public transportation corporation's tools to monitor and regulate vehicle operations and maintain their fleet in an effective and economical manner. It has also enabled the potential to provide consumers with trustworthy, updates of transit with real-time information via traveller information services (TIS). A key feature of such a system is the journey time and location information, thus the time a vehicle will potentially arrive at a certain location or the location in which a vehicle would be at a specific time (Predić et al., 2007). The output of an AVL is a series of points in time and two dimensions. This type of location data is ideal for visualizing a vehicle's location on a map or determining if the vehicle is following the designated path (Jeong, 2004).

1.10.5 Automatic Fare Collection Systems (AFC)

AFC systems are now often employed in urban transit networks. Buses with entry-only AFC systems require commuters to only tap a smart card when entering the system. On the one hand, passengers in entry-exit AFC systems must tap a smart card when boarding and alighting the bus. Operators have been investing in AFC systems to maximize the efficiency of ticketing and revenue collection procedures. Nonetheless, these systems have several advantages, such as easy passenger usage, increased revenue management efficiency, easy integration of intramodality, easy cooperation between multiple operators, and systematic data collecting and gathering tools(Barry et al., 2008; Nunes, A et al., 2016). The data generated through these methods are more accurate, less expensive, and more readily available than data gathered via manual surveys (Barry, J et al., 2002). The data capture each passenger's journey data, such as entry and exit stations, as well as the times they swiped their cards (Zhang & Yao, 2015).

a) Single-entry bus



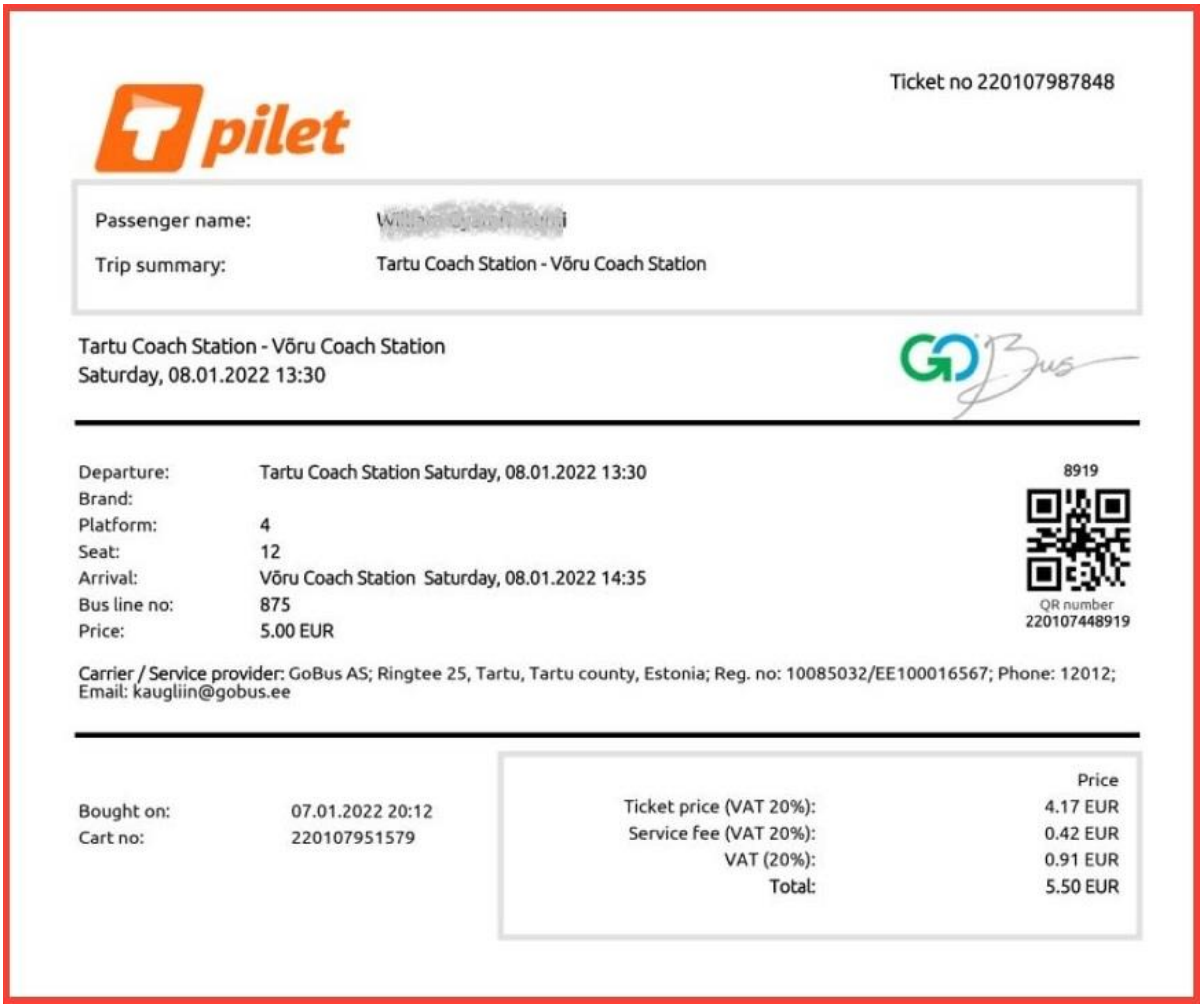
b) Ticket validator device

Figure 6. Tartu GoBus AFC system Source: Author

1.10.6 Electronic Ticketing system

An E-ticket, often known as an electronic ticket, is a digital version of a paper ticket. An e-ticket system is a more effective and dependable form of ticket entry, processing, and marketing that is used by airlines, trains, and other transportation and entertainment firms. It is sometimes known as a travel card or a transit card (Sanam et al., 2018). In public transportation, e-ticketing systems are not only a means of payment, but they also handle a great quantity of data, opening a wide variety of options for making public transportation simpler to use, administer, and regulate. They also provide possibilities to construct integrated pricing structures that are difficult to achieve with standard payment methods. Electronic ticketing systems are classed based on how they are paid for (Oloyede et al., 2014).

The system employs a multi-service approach and maintains information electronically. Furthermore, the acquired data gives accurate data on passenger flows, which may be utilized for planning purposes (Puhe et al., 2014). Smart cards or mobile ticketing are the two most prevalent technical solutions used in e-ticket systems. In the context of smart cards, two types may be distinguished depending on how they connect with other devices: contact-based smart cards, which need close contact with the reading sensor, and contactless smart cards, which are read by moving them in close contact to the sensor device (Mikulski, 2014). Mostly in the case of mobile ticketing, smart devices such as phone, the cell phone, tablet, or personal digital assistant serve as the ticket (PDA). Mobile ticketing has three primary functionalities: optical character recognition (OCR), contactless (Near Field Communication – NFC or Radio Frequency Identifier – RFID) and premium SMS-based transactional payments (Juniper Research, 2011).



pilet

Ticket no 220107987848

Passenger name:	[redacted]
Trip summary:	Tartu Coach Station - Võru Coach Station

Tartu Coach Station - Võru Coach Station
Saturday, 08.01.2022 13:30

GoBus

Departure:	Tartu Coach Station Saturday, 08.01.2022 13:30
Brand:	
Platform:	4
Seat:	12
Arrival:	Võru Coach Station Saturday, 08.01.2022 14:35
Bus line no:	875
Price:	5.00 EUR

8919
QR number
220107448919

Carrier / Service provider: GoBus AS; Ringtee 25, Tartu, Tartu county, Estonia; Reg. no: 10085032/EE100016567; Phone: 12012; Email: kaugliin@gobus.ee

Bought on:	07.01.2022 20:12
Cart no:	220107951579

	Price
Ticket price (VAT 20%):	4.17 EUR
Service fee (VAT 20%):	0.42 EUR
VAT (20%):	0.91 EUR
Total:	5.50 EUR

Figure 7.Sample mobile e-ticket Source: Author

2. Data and Methodology

2.1 Study area

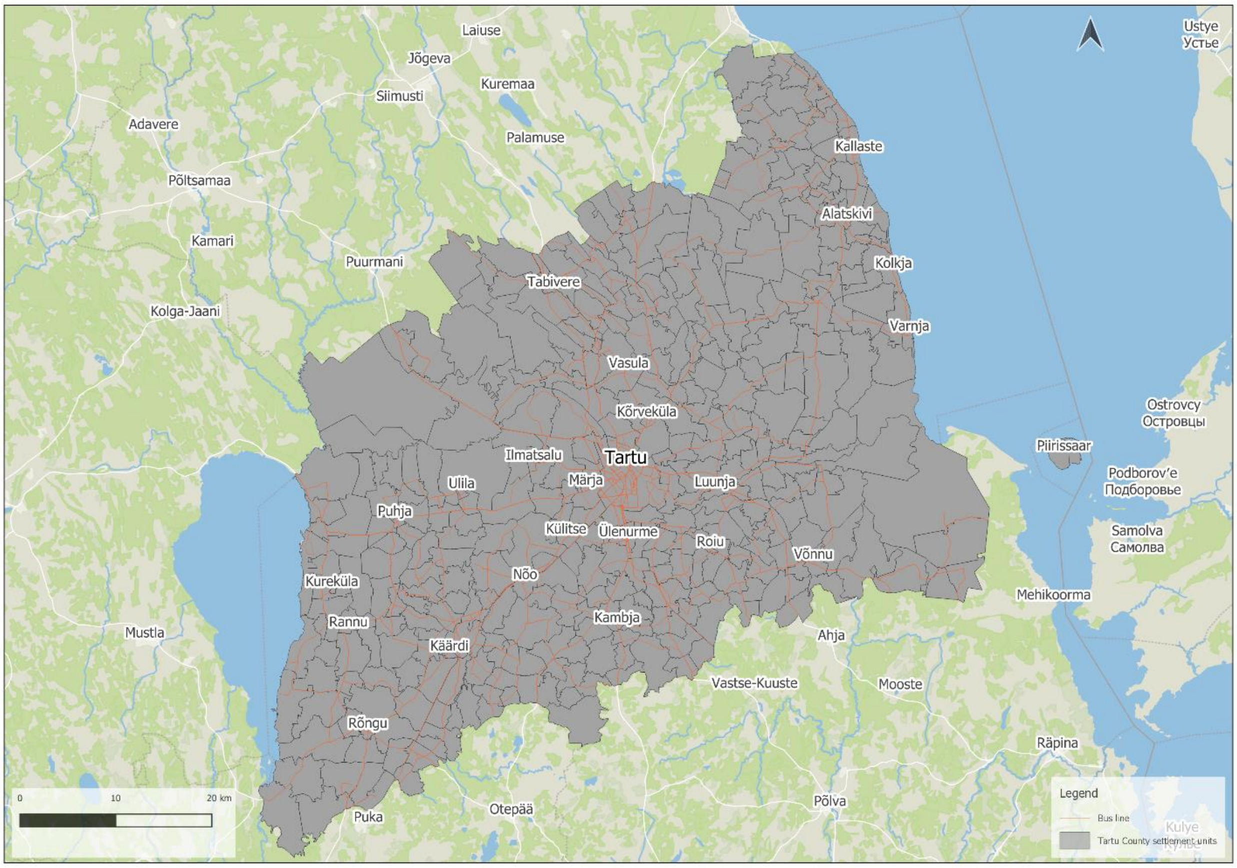


Figure 8.Map of Tartu County settlement units

The Tartu County is one of the fifteen (15) counties in Estonia located in the eastern part of the country and bothers other four counties namely, Põlva, Valga, Viljandi and Jõgeva Counties. It has an area of 2,993 km^2 which makes up 6.9% of Estonia's territory. As of 2017, the population of Tartu County was 145,550 making up about 11% of the total population in Estonia (Statistics Estonia, 2017). There are 408 geographical settlement units in Tartu County.

Tartu County had 22 municipalities before the administrative reform in 2017. These municipalities were Alatskivi vald, Elva city, Haaslava vald, Kallaste vald, Kambja vald, Konguta vald, Laeva vald, Luunja vald, Meeksi vald, Mäksa vald, Nõo vald, Peipsiääre vald, Piirissaare vald, Puhja vald, Rannu vald, Rõngu vald, Tartu city, Tartu vald, Tähtvere vald, Võnnu vald and Ülenurme vald.

Currently, there are Eight (8) municipalities that make up Tartu County due to the reformation. Elva vald which consists of Rannu vald, Rõngu vald, Elva city, Puhja vald, Rõngu vald. Nõo vald and Luunja vald remained as Nõo vald and Luunja vald respectively. Kambja consists of Kambja vald and Ülenurme vald but maintained name the Kambja vald for the region. Tartu city is also consists of Tartu city and Tähtvere vald but also maintatined the name Tartu city for the region. Peipsiääre vald on the other hand is made up of Vara vald, Kallaste vald, Alatsklivi vald, Peipsiääre vald but maintained the name Peipsiääre vald for the region. Finally Kastre vald is made up of Haaslava vald Luunja vald, Meeksi vald,Võnnu vald,Piirissaare vald and Mäksa vald.

As of 2017, there is 1 urban municipality (Tartu city) and 7 other rural municipalities in the county. Its capital and largest city is also called Tartu, located at a distance of 186 km (116 mi) from Tallinn city. Below are summary statistics of the current eight municipalities.

Table 1.Summary statistics of current eight municipalities

Municipality	Type	Population 2021	Area km^2	Density
Elva	Rural	14,635	732.27	20
Kambja	Rural	10,511	276	38.1
Kastre	Rural	5478	472	11.6
Luunja	Rural	5393	131.8	40.9
Nõo	Rural	4345	170	25.6
Peipsiaare	Rural	5462	652	8.4
Tartu city	Urban	94,750	38.8	2442
Tartu	Rural	11,969	742	16.1

Source: Tartu County Official Site (2021)

2.2 Geography of Tartu County

Tartu is the second largest city of Estonia as well as the centre of southern Estonia. It lies 186 kilometres south of Tallinn which is the capital city of Estonia (Tartu County Official Site, 2021). It lies between Lake Võrtsjärv and Lake Peipsi. River Emajõgi (100 km long), Estonia's only navigable river, is known to flow through the Tartu County, thus serving as a link between Lake Peipsi and Lake Võrtsjärv. The landscape of the county is also characterised by wavy plains with one-third being covered by forests. Wetlands at the headwaters and lower course of the Emajõgi also make up a quarter of the landscape. It has drumlin fields and lakes in its northern part as well about nature reserves with the latter constituting 10% of the county's territory.

2.3 Data

The buses used in Tartu County are run by an authority named GoBus AS which is responsible for the supply of buses to the competent authorities in charge of transport management in Tartu County.

The primary data was sourced from *MTÜ Tartumaa Ühistranspordikeskus* which contains 1,280,693 rows and 10 fields for the year 2019 in csv format.

The AFC system used in the buses allows riders to make contactless payments onboard the bus. Also, there are priced designated bus cards which also serve as an alternative mode of trip payment. The AFC system can record the number of users who swipe with the bus cards or make contactless payments at every designated bus station and bus stops in the transport network of Tartu County.

The GPS system in the bus records the coordinates of the buses at the bus stops and bus stations on their designated trajectory.

The data in csv format contained fields such as line, trip, start, departure_stop, destination, stop_longitude, stop_latitude, amount, destination_longitude, destination_latitude Summary of the fields is below:

- Line: contains the route of the buses as they operate from their origin to destination
- Trip : gives information about the start and end of a particular bus journey on a route
- Start: contains information about the date of the trips in the UNIX system
- Departure_stop: consists of names of all departed bus stops within the bus network
- Destination : column for the names of all trip destination
- Amount: refers to the number of passengers who journeyed from a departure stop to a destination
- Destination_longitude: contains coordinates of the destination longitude
- Destination_latitude: contains coordinates of the destination latitude
- Stop_longitude: contains longitude coordinates of the departure_stop
- Stop_latitude: contains latitude coordinates of the departure_stop

Shapefile of the settlement units in Tartu County (last update: 01/02/2022) was sourced from Estonian Land Board Geoportal. The shapefile contains fields such as *animi* which was specifically used to spatial join the csv data. Population data for 2019 was also sourced from estat.stat.ee.

Table 2.Sample bus data

Line	Trip	Departure_stop	Stop_longitude	Stop_latitude	Amount
313	Tartu-Kambja-Otepaa-Sangaste-Antsla	Tartu Bus station	26.73214	58.37796	1
99G	Paulva- Taevaskoja-Vaste-Kuuste-Tartu	Teemeistri	26.7342	58.27663	2

193	Jogeva-Kaarepere-Tabivere-Tartu	Vasula tee	26.67375815	58.45415	1

Start	**Destination**	**Destination_longitude**	**Destination_latitude**
1970-01-01 17:30:00	Otepaa	26.49666	58.05937
1970-01-01 08:10:00	Tartu Bus station	26.73214	58.37796
1970-01-01 13:30:00	Tartu Bus station	26.73214	58.37796

2.4 Method

The flow chart below gives an overview of the processes involved in analysis and results

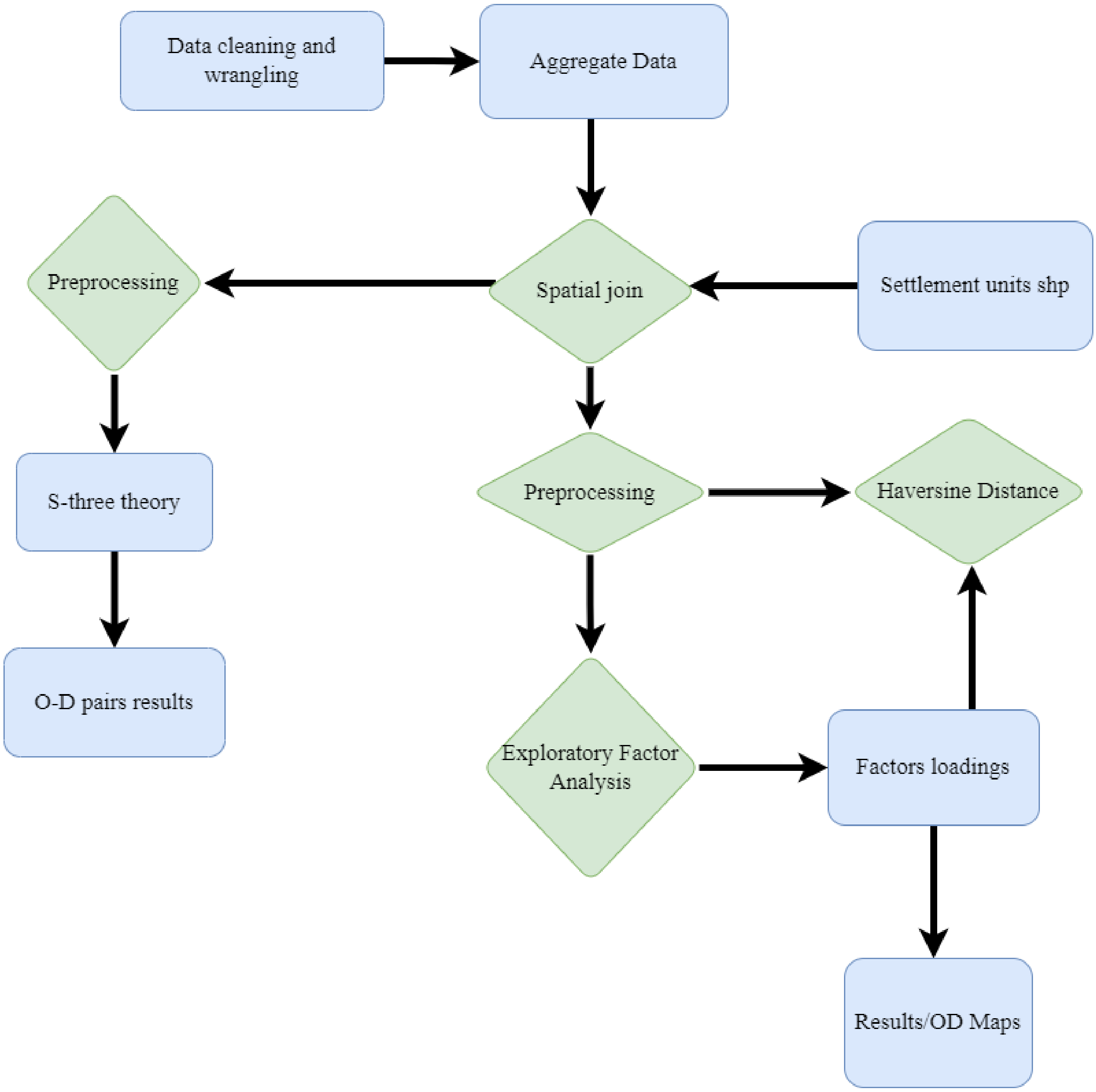


Figure 9. Workflow of analysis

Firstly, the data which is made up of 1 280 693 individual observations had some discrepancies. Using R software, data columns such as *Start* which represents the date and time in the UNIX system had to be converted and mutated into meaningful columns and formats. Columns such as time, hour, and minute, were created and converted to their rightful data types. However, there were some missing observations without destination coordinates and were removed from the data. These null values summed up to 53 331 observations and made up about 4% of the data.

The data was then aggregated to the number of trips(n) and unique ids were created for all observations for the dataset. The next step was to convert the data into a spatial data. The columns with coordinates namely destination_longitude, destination_latitude, stop_longitude, and stop_latitude were converted to WGS 84 and then converted to the Estonian projected coordinated system with EPSG code 3301. In order to spatially join the data to the settlement unit's shapefile, it was necessary to create centroids of the settlement unit shapefile. These centroids represent the distinct settlement units that were used as the basis for the spatial analysis. The centroids of the shapefile were joined to the data where all points which fall within the boundaries of the settlement units were captured. Unnecessary columns were removed, and column names were renamed for better understanding. The table below shows a summary of the spatially joined data.

Table 3. Sample spatially joined data

id	origin_name	origin_x	origin_y	n	time	destination_name
191818	Võsivere town	642511.1	6468698	3	17:30:00	Puhja town
600638	Tartu city	659627.8	6473200	5	06:40:00	Tila town
32292	Kõdu town	635005.5	6446892	1	13:40:00	Elva city

destination_x	destination_y	hour	minute
635530.9	6468786.7	17	30
661760.0	6477724.6	6	40
642136.2	6456562.7	13	40

New column meanings:

Id: Unique id for each row observation

origin_name: Origin name of settlement unit

origin_x: Latitude coordinate for origin_animi

origin_y: Longitude coordinate for origin_animi

n: Number of trips

destination_name: Destination name of settlement unit

destination_x: Latitude coordinate for dest_animi

destination_y: Longitude coordinate for dest_animi

The next step in the methodology was to pre-process the data to characterise the urban system in Tartu according to Limtanakool et., al (2007) theory. As a result, the data was further aggregated into settlement unit's level based on the number of trips from the origin of settlement units to the destination of settlement units. Since the transportation network does not cover all the settlement units in Tartu County, the settlement units with no corresponding destination settlement and coordinates were removed from the data as they became redundant in the analysis. Settlement unit's trips from Kapsta town to Undi town, Lilu town to Tilga town were among some of the settlement units that were removed. Also since we wanted to analyse trips between distinct settlements, settlement units with the same origins and destinations were removed from the dataset as well. As a result of this, the data was reduced to 3,885 observations after aggregation and filtering. This helps reduce the reduncy and noise as well as focus on the primary aim of distinct trips between settlements. The table below shows a summary of aggregated trips between settlement units.

Table 4.Sample aggregated origin-destination data for spatial structure and hierarchy settlement

origin_name	origin_x	origin_y	destination_name	destination_x	destination_y	n
Tartu city	659627.8	6473200	Elva city	642136.2	6456563	159179
Elva city	642136.2	6456563	Tartu city	659627.8	6473200	158211
Nõo town	648774.7	6462166	Tartu city	659627.8	6473200	64867
Tartu city	659627.8	6473200	Nõo town	648774.7	6462166	62546

Factor analysis

Exploratory factor analysis and confirmatory factor analysis (CFA) are defined as the two main types of factor analyses (Williams et al., 2010). With regards to EFA, as the name implies, it allows the researcher to investigate the major factors in order to develop a theory from a large number of implicit variables, which are often represented by a series of items (Pett et al., 2003). In comparison to exploratory factor analysis, confirmatory factor analysis has a clearly marked out hypothesis and thus expectations about the number of constructs and which construct theories or best match based on priori model and theory (Williams et al., 2010). The formula for EFA are as follows:

Take p manifest variables, indicated by $x1,x2,...,xp$, as well as the variables mean, denoted by $\mu1,\mu2,...,\mu p$, the covariance matrix of X, denoted by Σ.

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \mu_{p\times 1} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}, \mathbf{\Sigma}_{p\times p} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix},$$

Also consider m factors which is denoted by F_1, F_2, F_3,……, Fm

$$\mathbf{F}_{m\times 1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}$$

The rationale behind factor analysis is that of a regression prediction which implies that we can describe each manifest variables as a linear combination of latent variables or factors. Consequently, *p* manifest variables and *m* factors can also be rewritten as:

$$x_j = \mu_j + \lambda_{j1}F_{j1} + \lambda_{j2}F_{j2} + \cdots + \lambda_{jm}F_{jm} + e_j j = 1,2,\ldots,p$$

Where the Factors F have zero means and a unit standard deviation. The error term ej is assumed to also have a mean and standard deviation, σj. The equation can be expressed in a matrix form,

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \ldots & \lambda_{pp} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix},$$

and can be as a matrix notation as:

$$\mathbf{X}_{p\times 1} = \mu_{p\times 1} + \Lambda_{p\times m}F_{m\times 1} + \mathbf{e}_{p\times 1}$$

The equation can be rewritten as $\mathbf{X} - \mu = \Lambda F + \mathbf{e}$ which represents the exploratory factor model.

After we have defined the factor model, we will like to know how much of the variability in X, denoted by the covariance matrix Σ, where $\Sigma = \text{Cov}\ (\mathbf{X}) = (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T$ is explained by the factor model.

Factor analysis breaks down the original *p x p* variance-covariance matrix Σ of the initial variable **X** into a loadings matrix of *p x m* matrix Λ , where Λ= Cov (XF) and a *p x p* diagonal matrix of the number of unexplained deviations per original variable ,Ψ, where Ψ = Cov(e), where $\hat{\Sigma} = \Lambda\Lambda^T + \Psi$. The loadings matrix Λ, in the first term provides the coefficients *(jm)* that tie the factors *(Fjm)* with each single observation *(xj).* The term $\Lambda\Lambda^T$ relates to the variability which can be explained by the factors. This is overall variability which is explained to be a linear combination of the factors is referred to uniqueness as denoted by Ψ (Hartmann et al., 2018) . The variability is given by:

$$\hat{\Sigma} = \underbrace{\Lambda\Lambda^T}_{\text{communality}} + \underbrace{\Psi}_{\text{uniqueness}}$$

2.5 Exploratory factor analysis procedure

The diagram below summarise the procedures used in achieving the factor analysis of the research.

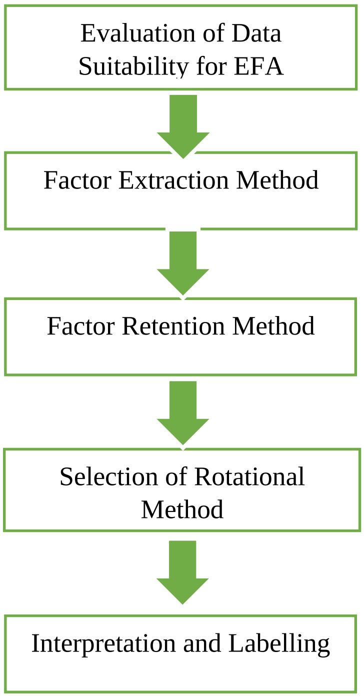


Figure 10.Factor analysis flow chart Source: Author

Evaluation of Data suitability for EFA

Even though exploratory factor analysis appears to be a complicated statistical method, the analysis involved is systematic and linear, with several alternatives (Thompson, 2004). Among other things the reduction of number of variables, as well as the evaluation of construct validity in a survey as (Fabrigar et al., 1999) state in their works, there are five methodological difficulties that researchers should consider while using exploratory factor analysis. First, the researcher must assess whether the EFA is the most suitable statistical method for achieving the study's goal

As a result, for finding the latent factors in the dataset, exploratory factor analysis (EFA) on principal component analysis was performed on the dataset. For every settlement unit the daily number of trips were aggregated for all 365 days in 2019 from the dataset. Using R statistical software, a matrix of 3,845 entries by 366 total columns were created to perform the factor analysis. Values in the cell which had missing values were replaced with 0's as factor analysis deals with only numerical values. The means and standard deviations of the numerical variables were also computed.

Kaiser-Meyer-Olkin (KMO) factor adequacy were tested on the dataset to examine the adequacy and suitability of the dataset for factor analysis. The result was a MSA value of 1 and according to (Netemeyer et al., 2003) a KMO correlation greater than 0.60 is appropriate for assessing EFA output. In addition, a Bartlett test of sphericity was conducted to further prove suitability of the data. The p-value obtained from the test was 0 which also indicates a better chance of EFA.

Factor extraction

Factor analysis generates linear combinations of factors in order to extract the fundamental communality of the variable. Fewer variables represent the majority of the variance in the data set to the extent that the variables have an underlying communality. This enables us to aggregate a large number of observable variables in a model to represent an underlying concept, making the data easier to comprehend.

Principal component analysis was applied on the dataset to get an initial number of factors to use for the analysis. Two factors were suitable for the analysis. According to (Gorsuch, 1997), there are no significant differences when factors have high reliability or there are thirty or more factors.

Factor retention methods

The scree test is an algorithmic graphic method that involves plotting the eigenvalues (y-axis) against the components (x-axis) and inspecting the shape of the resulting curve to detect the point at which the curve abruptly fluctuates (and the "scree on the hill slope" begins). This point on the curve represents the maximum number of components that can be retained (Zwick & Velicer, 1982). The scree plot indicated 2 factors as the suitable numbers for this analysis. Using the *factanal()* built-in function in R, with two factors, the factor analysis on the dataset was computed.

Interpretation of factor results

Factor analysis generates linear combinations of factors in order to extract the fundamental communality of the variables. Fewer variables represent the majority of the variance in the data set to the extent that the variables have an underlying communality. This enables us to aggregate a large number of observable variables in a model to represent an underlying concept, making the data easier to comprehend. Given Σ, and its estimate $\hat{\Sigma}$ the variability in the dataset can be explained by a linear combination of communalities and the variability that cannot be described by a linear combination of the uniqueness.

Table 5. Sample communality and uniqueness values

Days	01/01/2019	27/02/2019	29/12/2019	22/04/2019	20/06/2019	11/06/2019	26/01/2019
Communality values	0.9502277	0.9697124	0.9703714	0.9839261	0.979862	0.9787352	0.9757163
Uniqueness values	0.04977231	0.0302876	0.02962861	0.9839261	0.979862	0.9787352	0.9757163

Selection of rotational method

Rotation aids in the production of a more understandable and simplified solution by boosting high item loadings and decreasing low item loadings. Rotation techniques include oblique, orthogonal rotations and promax. A promax rotation was used on the factor loadings which aids in explaining and understanding of how the factors relate to each other. From the analysis factor 1 and factor 2 have common and uncommon variables that relate and do not relate with each other. This reflects the common pattern found within travelling days of the year for incoming and outgoing trips.

Interpretation and labelling

The process of selecting variables that are relevant to a construct and assigning a name to that construct is known as interpretation. Construct labelling is a theoretical, subjective, and inductive process (Pett et al., 2003). According to (Henson et al., 2006), at least two or three variables must load on a factor in order to provide a meaningful interpretation. As a result of this factor 1 represented outflows of summer commuters while factor two represents inflow of commuters during the summer.

Distance measurement

The Haversine formula is a method for estimating any distance between two points on the earth's surface, by connecting the points contingent on their longitude and latitude. To compute the two distances, four variables must be prepared. The formula is crucial in navigational issues, as it can offer a big circle distance between two points on the sphere's surface, irrespective of the altitude and depth of the given points (Chopde et al., 2013). It does so by taking into account that the earth is not a plane, but rather a curved plane with a radius of 6,367.45 km (Siahaan et al., 2016).

As a result, distance in kilometres was computed using the Haversine method in *R statistics* which is the shortest distance between two points *as the crow flies* or *great-circle-distance.* This method ignores the ellipsoidal effects of the earth and as a result the actual road distance is not reflected or used in this research. The calculation of distance in Haversine is the difference or magnitude of changes in longitude (Δlongitude) and latitudes (Δlatitude) of two coordinate points in radians. The Δlonitude = latitude2 – latitude1 and Δlatitude = longitude2 – longitude1

The formula can be calculated as follows where R is the radius of the earth 6371 (Rezania & Darnis, 2020)

$$\text{distance } = 2 \cdot R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta lat}{2}\right) + \cos(\text{ latitude 2}) \cdot \cos(\text{ latitudel }) \cdot \sin^2\left(\frac{\Delta long}{2}\right)}\right)$$

3. Results

This section discusses the results and findings of the trips between settlement units, seasonal rhythm and distance trips in the Tartu County.

3.1 Hierarchical structure of Tartu County

Figure 11 below shows a flow map between settlement hierarchies of Tartu County from commuter trips.This emphasises the variation of inflows and outflows of trips between the settlement hierarchies. To begin with, the summarised aggregated data for origin-destination trips between settlement units provided in Table 4 above provides only a brief description of trips between settlement units. However, for the entire analysis, Tartu city which is located in the central part of the map has the highest number of inflows from the peripheral areas. This makes Tartu city the most dominant city among other cities such as Elva (south-west of Tartu city) and Kallaste (north-east of Tartu city) which exhibit a relative dominance yet act as peripheral areas in relation to Tartu city. About 158 211 trips emanated from Elva city to Tartu city. Also trips from Tartu city to Elva city which accounted for the highest number of settlement trips in 2019 were 159 179. Trips between Tartu city and Nõo village accounted for about 62 546 trips. city.

Nonetheless, there are evidence of other trips between peripheral areas or settlements but not as strong as trips between a dominant city and non-dominant peripheral areas. A typical example in the analysis is the trips between Elva city and Annikoru village which account for 6,315 trips. Another example is also evident between settlement units from Lähte village to Äksi town accounting for 6 026. It can also be said that the strength of interaction between Tartu city and the peripheral areas is the strongest. This is because Tartu city reciprocates most of the trips which emanate from the peripheral areas. For instance, round trips between Tartu city to Elva city are the highest in the public bus network structure. peripheral areas peripheral settlement.

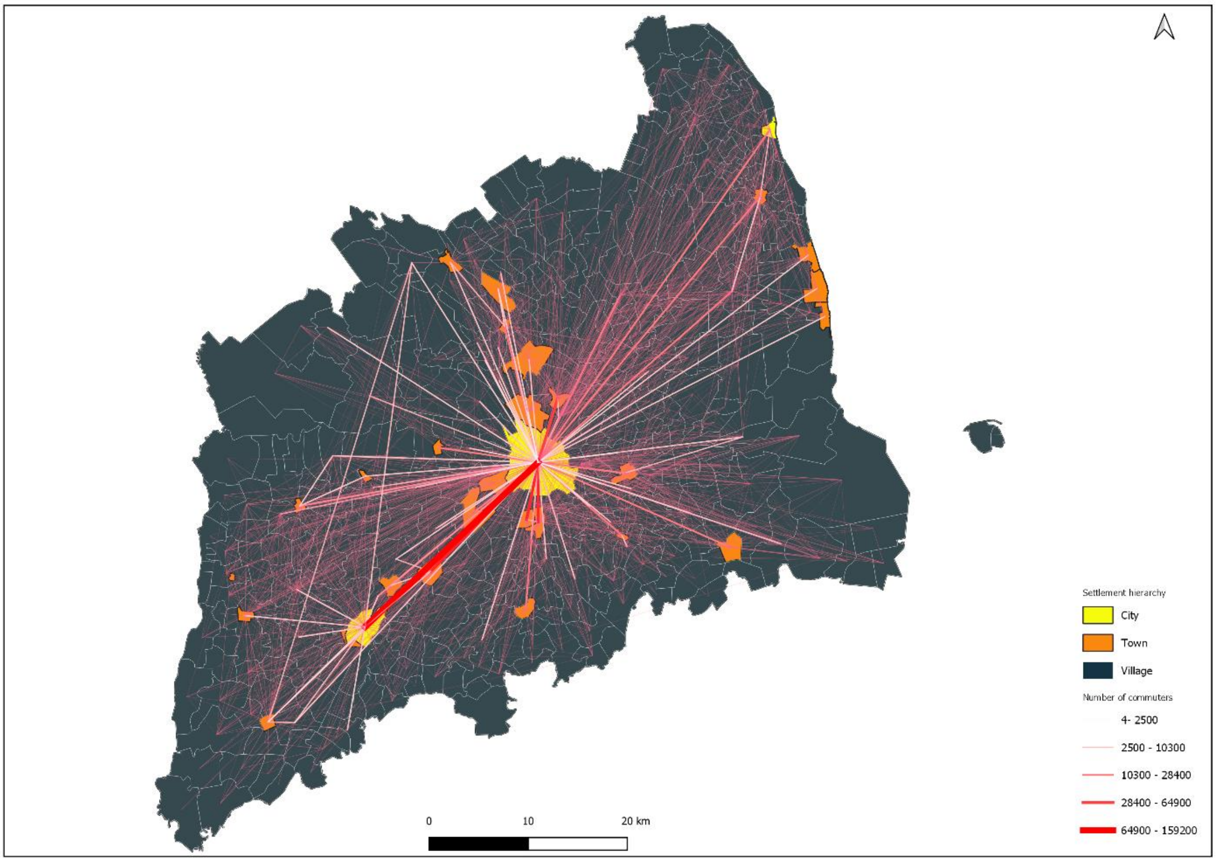


Figure 11. Map of settlement hierarchy and number of commuters of Tartu County

3.2 Seasonal rhythm

From the figure below there is a high factor loading from the days in the months of January to April but begins to drop steadily until it becomes more pronounced from late May till August. It then begins to rise till the end of the year. The highest factor loadings are around 0.83 and are seen in the days of the months between January to April. The latter days in May till August experience a fluctuation in factor loadings with a value ranging between 0.56 to 0.57. However, the lowest factor loading is recorded in the latter days of the month of December with a loading of 0.55.

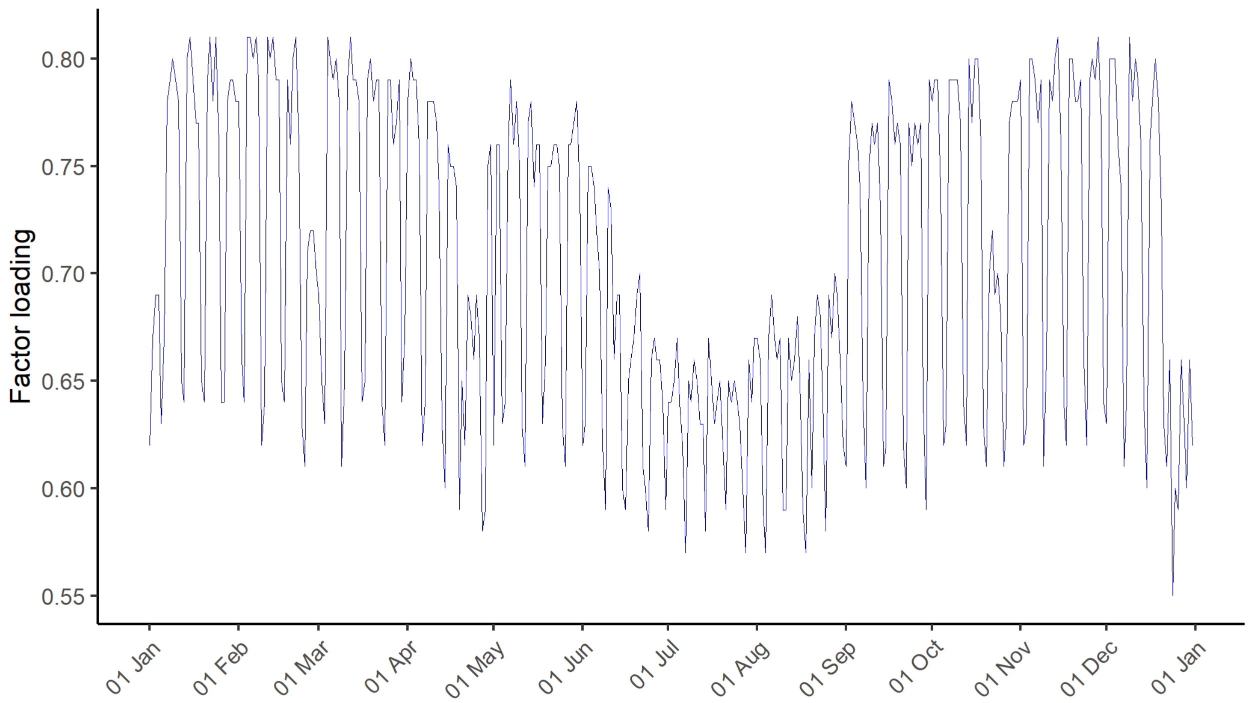


Figure 12. Factor 1 showing commuting outflows variation from Tartu city to peripheral area in summer period

Table 6. Sample factor 1 loadings from dataset

Origin me	Destination name	Factor 1 loadings	Date
Tartu city	Luunja town	0.78	20/03/2019
Tartu city	Aadami village	0.59	10/08/2019
Tartu city	Lemmatsi village	0.8	19/03/2019
Tartu city	Aadami village	0.78	08/03/2019
Tartu city	Luunja town	0.78	20/03/2019
Tartu city	Naelavere village	0.78	08/03/2019
Tartu city	Kallaste city	0.77	25/09/2019
Tartu city	Pangodi village	0.59	09/06/2019

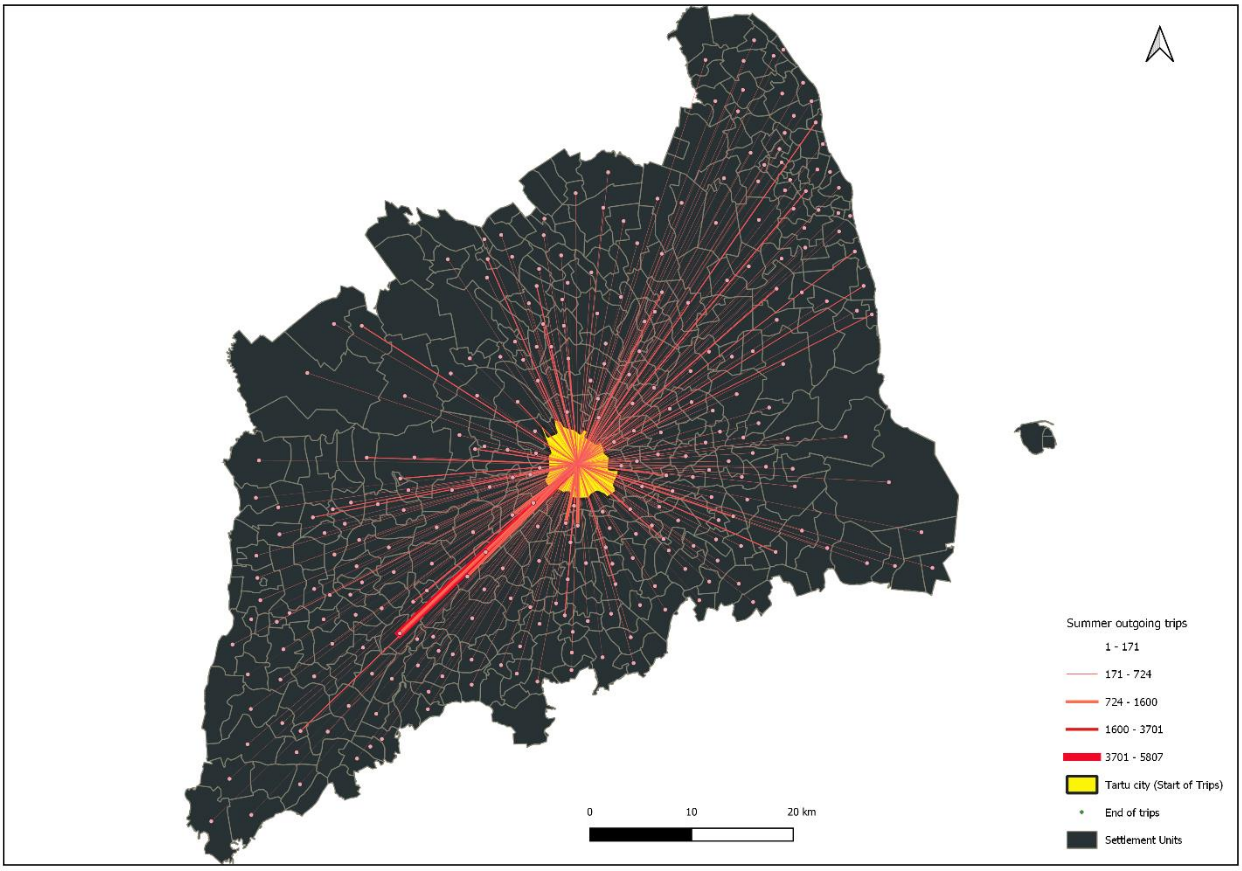


Figure 13. Map showing summer outgoing trips from Tartu city to peripheries

The figure below shows the incoming variation of flows from the peripheral areas to Tartu City during the summer period. There are high factor loadings in the days of the months of June until August. Other variations can be seen in the days of the months of January until April which is similar to the days in the months of October to December. The highest factor loadings are evident in the days in August and September with values around 0.80. High factor loadings can also be noticed in the days of December with a loading of 0.80. In the days of January through April there are low factor loading values ranging from 0.56 to 0.57. Also, low factor loadings are evident in the days of the months October till December with values ranging from 0.57 to 0.58.

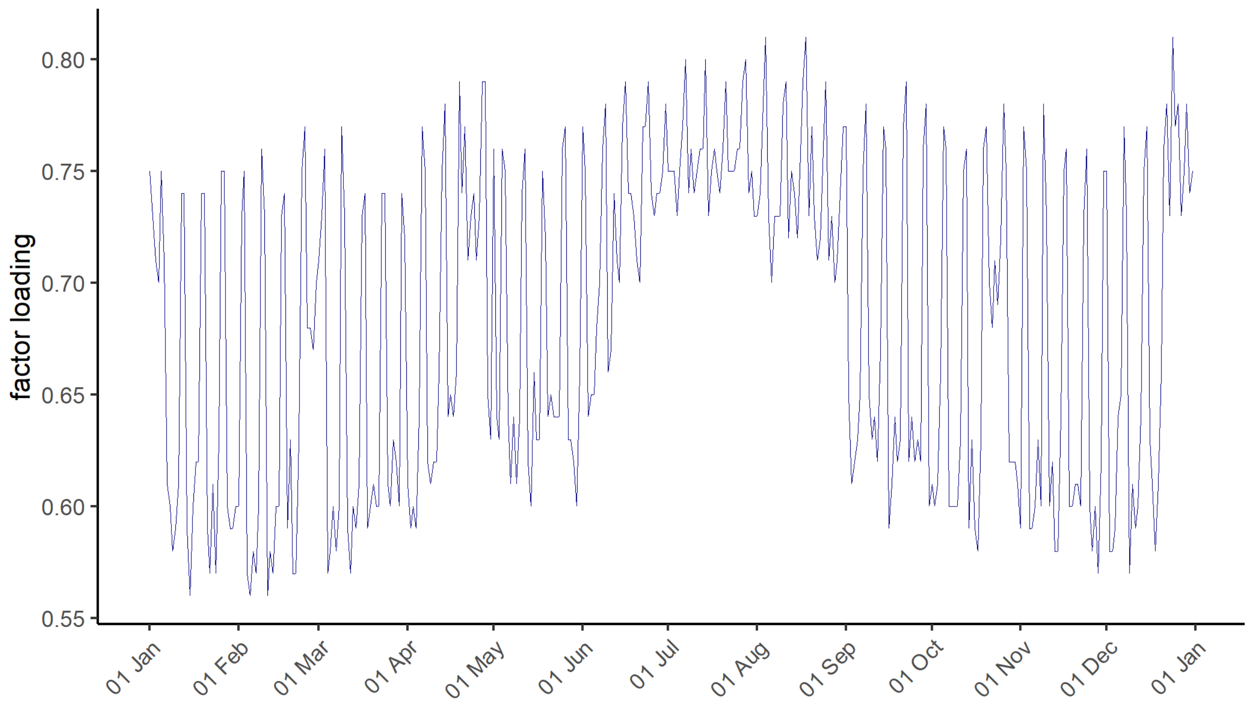


Figure 14. Factor 2 showing incoming variation flows from the peripheral regions to Tartu City during summer period

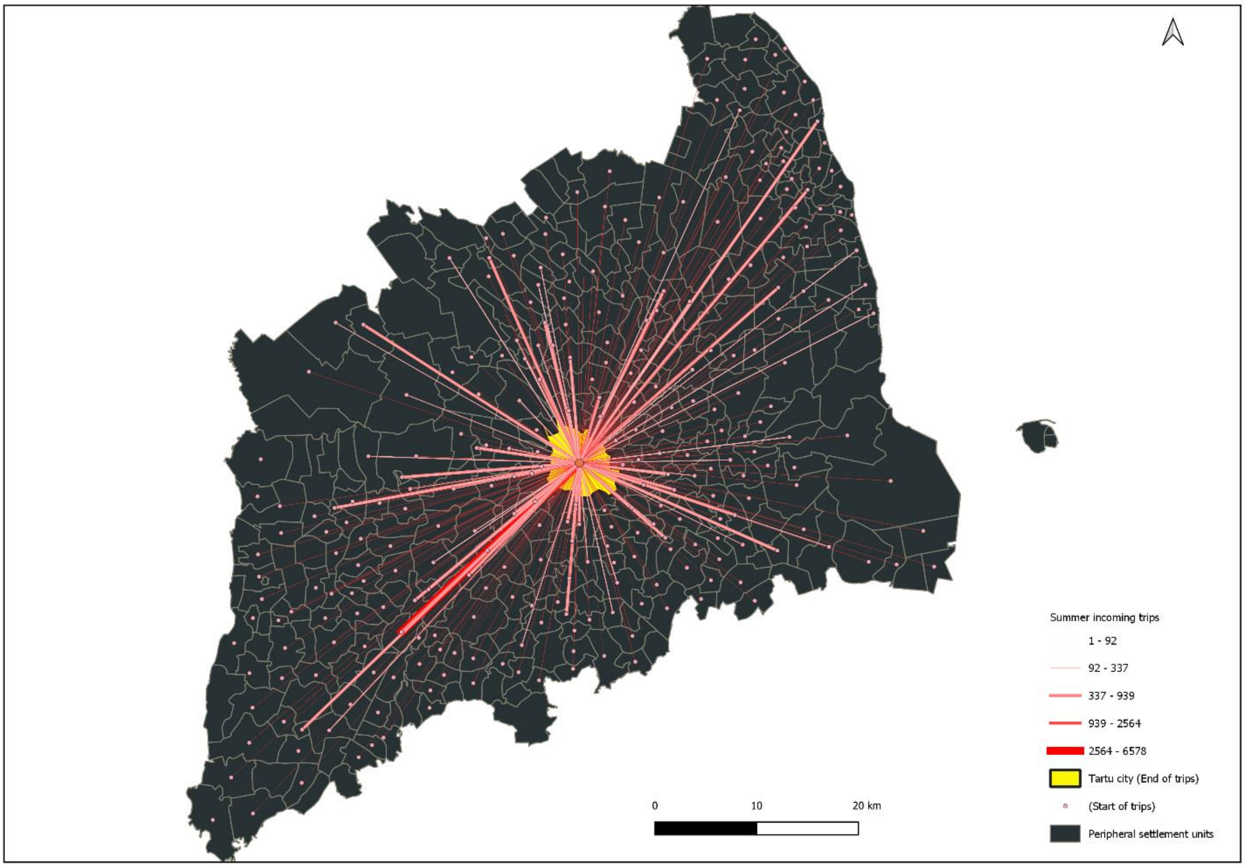


Figure 15. Map of incoming trips from peripheries to Tartu city

Table 7. Sample factor 2 loadings from dataset

Origin name	Destination name	Factor 2 loadings	Date
Kaagvere village	Tartu city	0.68	26/02/2019
Elva city	Tartu city	0.81	07/07/2019
Juula village	Tartu city	0.61	01/10/2019
Lahepera village	Tartu city	0.61	19/03/2019
Aadami village	Tartu city	0.56	15/01/2019
Kaagvere village	Tartu city	0.68	26/02/2019
Sipe village	Tartu city	0.56	11/02/2019
Rahinge village	Tartu city	0.76	19/10/2019

3.3 Seasonal rhythm in relation to distance

The figure above shows the disparities in commuter trips in summer (June to July) from the periphery to Tartu city and vice versa by distance. It can be seen from the diagram that there are considerable similarities in both inflows and out flows to and from Tartu city. However, a distinct observation is seen in the number of trips at a distance coverage between 37.2km to 37.5km with the highest peak seen at 37.2km. This highest peak in the number of commuter trips accounts for more than 40 000 trips in the summer season both for outgoing and incoming trips. Furthermore, among the two peaks with the same distance of 37.2km, incoming trips to Tartu city has the highest number of commuter trips at 37.2km is from Elva city with a total commuter trips of 40693.

The second highest peak is seen at 23.4km for both incoming and outgoing trips representing a total trip of 12 234 and 12 097 respectively. The third highest peak is seen at 5.4km and 5.7km which also makes about 9 777 and 8 085 trips for outgoing trips respectively and for incoming trips with the same distances, the total number of trips were 10 763 and 7 307 respectively which indicates more incoming trips than outgoing trips. At 7.09km for both Kõrve-village town and Reola village the number of commuter trips are 3 662 and 2 262 for outgoing trips respectively. For incoming trips at the same distance, the total number of trips were 5 585 and 2 263 respectively. This implies that there were more incoming trips from Kõrve-village town than Reola village as seen in the graph denoted by the pronounced red line. There is also a significant observation between Koosa village and Tartu city at a distance of 41.33km between them. Trips from Koosa village to Tartu city accounted for 4 803 whiles trips from Tartu city to Koosa village accounted for 4 737. Again, another variation in commuter trips is seen at a distance of around 46.18km. Here, the curve sows a more pronounced increase in incoming trips from the peripheral areas to Tartu City than outgoing trips. With most of the trips coming from Puhja town the total number of incoming trips was 3 071 while that of outgoing trips was 1 219.

For long distance trips between 59km and 80km people rarely travel to and from Tartu city. For instance, trips from Purtsi village to Tartu city were only 4 throughout the year with a distance of 76.8km between them.

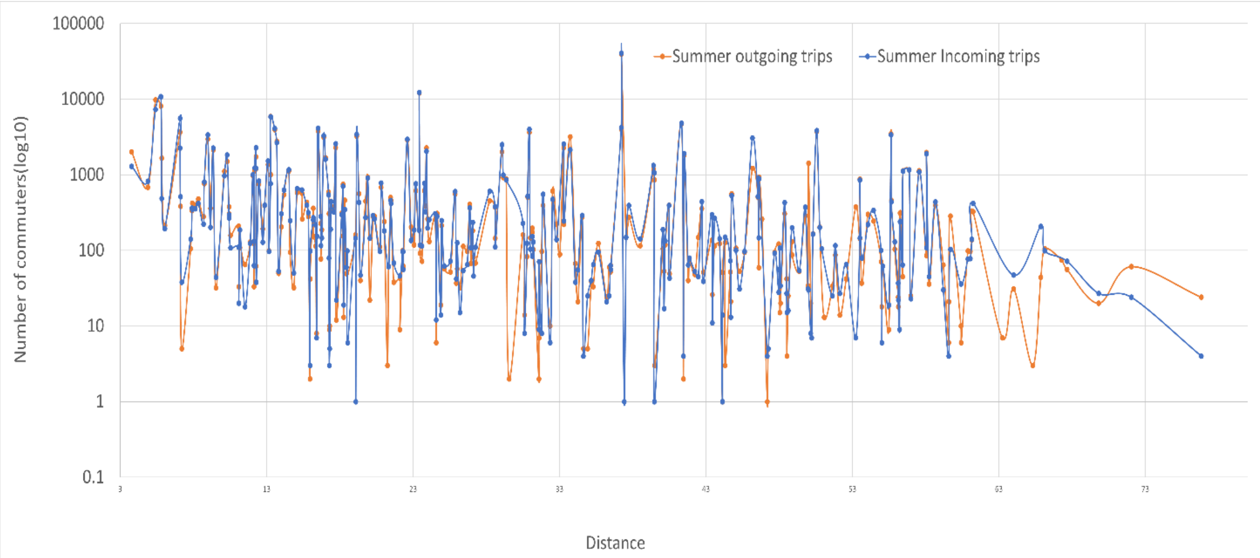


Figure 16. Graph showing number of commuters in relation to distance

Table 8. Sample summer trips from Tartu city to peripheral areas and peripheral areas to Tartu city

Origin name	Destination name	Distance(km)	n
Sirvaku village	Tartu city	17.59	323
Tartu city	Elva city	37.21	39 524
Elva city	Tartu city	37.21	40 693
Tartu city	Reola village	7.09	2 262
Tartu city	Kõrve-village town	7.09	3662
Tartu city	Koosa village	41.33	4 737
Koosa village	Tartu city	41.33	4 803
Purtsi village	Tartu city	76.8	4
Tartu city	Ülenurme town	5.41	9 777
Vissi village	Tartu city	33.73	2 134

Puhja town	Tartu city	46.18	3 071
Tartu city	Puhja town	46.18	1 219

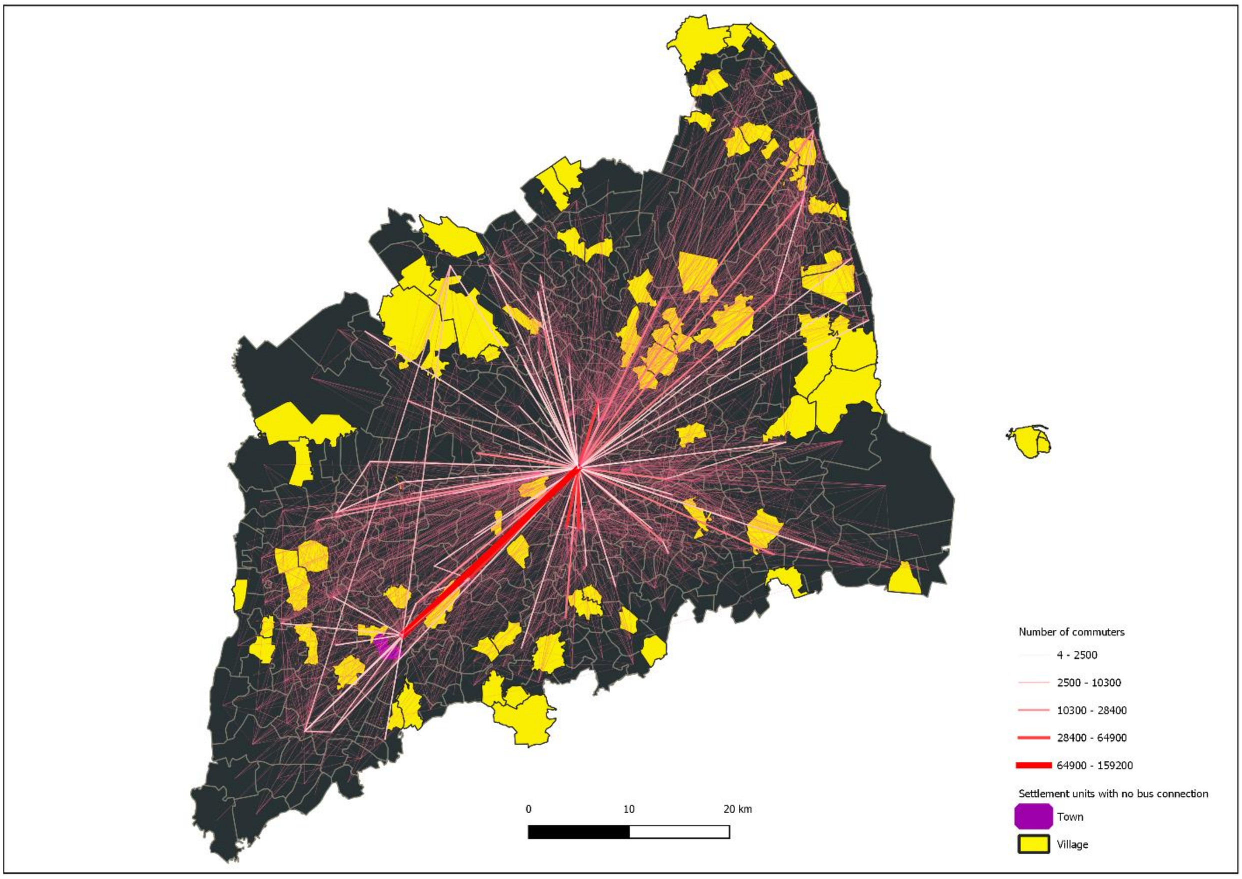


Figure 17.Map of Tartu County showing settlement units without bus connectivity in 2019

4. Discussion

This section discusses the results presented in the previous chapter. In view of this, relationships between peripheral areas and incoming and outgoing trips in and out of Tartu City will be discussed as well their seasonal rhythm, particularly in summer. In so doing, some factors that could account for these variations or disparities will be highlighted. In addition, this section will also discuss the effects of distance on commuter trips in and out of Tartu City with respect to its peripheral areas.

First of all, it was found that Tartu County can be classified into three settlement units which are cities, towns and villages. Of the 403 settlement units in the Tartu County, only three of them are cities, while the rest represent towns and villages, with villages being the highest number of settlement units. The three cities in Tartu County are Tartu city, which also serves as the capital of the county, Elva and Kallaste. Out of these three cities, Tartu is the major one.

To answer the first question which is the determination of the spatial structure and hierarchy of Tartu County with regards to public transportation, the theory by Limtanakool, et, al., (2007) was used. In this case, the commuter trips, representing human flows through the use of public transportation (bus) was used. Tartu County was therefore assessed using the theory's three spatial dimensions which are symmetry, structure, and strength. First of all, the strength, according to the aforementioned authors assumes that when nodes have a strong interaction, it is easier to move between each other (Limtanakool et al., 2007). In this vein, it was seen from the map (Figure 11) that strong interactions were evident between Tartu City and the other settlement units. However, the strengths of interaction was not equal for every settlement unit. Whereas some of the settlement units exhibited strong interactions, others exhibited very weak interactions. For instance, Elva exhibited the strongest interactions in term as of both inflows and outflows to Tartu city and this was replicated on the map (Figure 11). However, its counterpart, which is Kallaste city did not exhibit similar strong interactions. On the other hand, towns such as Körve-village town, Ulenurme town and Törvandi town had the next strongest strength of interaction, next to Elva. Aside Tartu city, significant commuter trips was also seen between other settlement units. These included trips between Elva and Annikoru as well as between Lähte and Äksi. Bourne & Simmons (1978) posit that the frequency of movements between two set of nodes (origin and destination) determines the strength of interaction between them. Based on this premise, it can be concluded that settlement units with high number of commuter trips representing the aggregate inflows and outflows through public bus transportation, corroborates the strength of interactions between these places and Tartu City.

Another phenomenon worthy of mention is the structure of Tartu County. According to Limtanakool, et. al., the structure of an urban system can be defined as fully monocentric or fully polycentric. A fully monocentric node represents flows from one or few major nodes to less important ones. This connotes a hierarchical structure. However, in polycentric systems, the non-hierarchical structure is evident because there are varied nodes with specialised functions where spatial flows are seen in varied directions, whether horizontal, diagonal, reciprocal. In the case of Tartu County, flows were not directed to just one settlement unit. There were varied flows to different settlement units although the flow sizes differed, such as the in the case of Tartu City where flows were highest for incoming and outgoing trips. Again, as stated, in the previously, there were interactions between cities, towns and villages with different origins and destinations. Again, in terms of direction, it was also noted that some of the settlement units exhibited some form of reciprocity. This means that there were exchange of flows between some of the settlements. An example is that of Tartu city and Elva city. Inasmuch as people moved from Elva to Tartu City, commuters also moved from Tartu City to Elva. In light of this phenomena, it can be concluded that Tartu County has a non-hierarchical structure because it is polycentric, that is, it has varied nodes with specific functions. However, in most cases a system can never reach a fully polycentric structure (Limtanakool, et, al., 2007). Tartu County is therefore not fully polycentric but has a form of polycentricity connoted by Type A4, according to Limtanakool, et, al., 2007. This is characterised by interdependencies between both dominant and non-dominant nodes, strong interactions existing between dominant nodes and weaker interactions existing between non-dominant nodes. This is typical of urban systems due to the fact that several needs and functions spread to different areas with time, reducing the need to obtain particular services only from one node. As emphasised by Sinclair (1983), the hierarchical structure of the urban system is considered to decline as the urban system evolves toward a completely polycentric system.

In addition to structure, another concept which characterises the structure of an urban system is dominance, according to Limtanakool, et, al., (2007) which could, or could not influence the direction of flows. Direction of flow is considered when spatial flows regarding the access to facilities, products, or services are connected to the geographical locations at the nodes where arrivals occur, in other words, the destination of incoming flows. This phenomenon highlights the importance of such a location in meeting the needs of people and thus attracting a lot of inflows (Alderson & Beckfield, 2004). In the case of Tartu County, based on Alderson & Beckfield (2004), Tartu city exhibits the highest dominance among the other settlement units making them act as peripheries, in relation to Tartu City. The reason for this is that as stated above, Tartu City is the capital of Tartu County, which makes its the centre for most of the county's specialised functions as well as the provision of goods and services. This phenomenon can be explained by Christaller Walter's Central Place Theory 1933 which explains that people are drawn to places where they can easily assess goods and services, classifying such places as the central place owing to its centralised functions. This means that most of the flow direction from the other settlement units (peripheries) will be directed towards this city.

Based on the structure of an urban system, the symmetry, which is the last spatial dimension can be determined. In most cases asymmetric interactions are associated with fully monocentric systems because they are unidirectional in nature. This means flows are geared towards just one important node (monocentric) where all specialised functions are concentrated. This is not the case with Tartu County because although Tartu City receives the most inflows and outflows due to its function as the capital of the county and the situation of most specialised functions in the city, other cities like Elva also receive both inflows and outflows. Flow is therefore multi-directional as opposed to being uni-directional, if it were asymmetrical. From this, it can be said that flows in Tartu County is symmetrical. Most symmetric interactions are alluded to polycentric systems since specialised functions are distributed across the urban area and so all nodes are interdependent (Limtanakool et al., 2007b).

The second research question focused on the seasonal rhythm with regards to commuter trips. From the results, it was established that there were variations in commuter trips with regards to seasons. Estonia, as with most European countries have four seasons in year, that is, winter (late November to March), summer (June to August), spring (March to May) and autumn (September to November). it was found that commuter trip patterns experience acute changes from June to August, which falls under the summer season. In this case, commuter trips increase for both inflows and outflows, to Tartu City. This represents the peak of most commuter trips in the year. A reason for such increase in commuter trips is as a result of the various activities that occur during summer and in this case, attracts more people than usual to Tartu city. Some of these pull factors that attract people during the summer season include climate in that most summer seasons are characterised by warm temperatures and so outdoor activities increase. People therefore go to tourist sites, participate in carnivals or fairs such as the Hanseatic fair, Emajõgi summer concerts and also some go on vacation. The peak or increase in movement is equally embedded in the seasonal nature of these activities. Owing to the fact that these activities are mostly organised during the summer period annually, people capitalise on this opportunity to take part in them or even travel as this might not be possible in other seasons such as winter where the weather might be cold or unfavourable. An exception is seen in December and January. Although it is also in winter season, there is an increase in commuter trips similar to that of the summer season. The reason accounting for this is that this period represents the festive seasons (Christmas and New Year) where people travel to be with family and other planned activities which are equally scheduled for winter. In sum, two seasonal variations were seen in Tartu county with regards to commuter trips and they are evident in summer and winter, although that of summer spans across more months (June to August) than winter (middle of December to the early parts of January).

Again, this study also sought to find out the extent to which the seasonal rhythm of affects distance trips. In this case, the summer period was used for this analysis owing to the fact that it has a pronounced rhythm as opposed to the winter season where the festivities or activities spans for a maximum of one month (middle of December to the first few weeks in January). Also due to the fact that the rhythm was more pronounced from the factor analysis. The graph shows various distance relationships between the various settlements and the number of commuters. To better explain this phenomenon, the distance decay theory which is further explained in Tobler's (1970) first law of geography will be used as the basis. The distance decay theory states that the further apart things are, the less likely it is that they will interact very much. This is evident in Tobler's first law of geography which states that "everything is related to everything else, but near things are more related than distant things". Based on this theory, it is expected that settlement units closer to Tartu City, which is the dominant node will have higher commuter trips than those who are distant. However, this was not the case. The graph showed a mixed phenomenon with some defying the distance decay theory. Some settlements units which were closer to Tartu had a lower number of inflows than those further away from Tartu city. For instance, it was noticed that the highest peak of commuter trips was achieved at a distance of 37.2 km, by Elva city and this represented about 6.21% of the incoming summer trips to Tartu and 6.05% of outgoing trips from Tartu city of total trips.

Another exception is seen from trips made from Körve-village town and Nöo to Tartu city. Although Nöo is quite far from Tartu (23.42km) contrary to Körve-village town (7.09km), incoming and outgoing trips from the Nöo were higher (1.97%, 2.52%) than that of Körve-village town (1.01%, 0.59%) respectively. However, the case of Purtsi is a typical example of the distance decay phenomenon due to the lower percentage of both incoming and outgoing trips to Tartu city. At a distance of 76.83km, their incoming trips represented just 0.001% of total incoming trips while that of outgoing trips were 0.006%. It can be argued that certain factors such as the importance of a particular service, seasonality of events or activities and the absence of alternative locations which provide specialised functions can make people defy the distance decay theory and travel at great lengths to get their needs met

It was also observed from Figure 17 that there were settlement units which did not have any interactions and trips to Tartu City. These regions are sparsely populated and hence could explain the reason for the lack of public transport connections to the area. It could first of all be assumed that owing to the sparse population of these areas, it is not economically viable to have public transportation connections to these areas and hence the absence of connections in these settlement units. This economic viability can also be buttressed by one assumption of the Christaller's central place theory where he opines that places with higher population density leads to cheaper production. Hence, it is more economically viable to have public transport connections at settlements with high population densities than areas with low population densities, characterised by sparse populations. Another assumption can be made based on the Lorenz curve which represents income or wealth distribution within a population. As earlier stated, these settlement units without connections or commuter trips have populations below 600 and are also situated further away from Tartu city, which is the dominant node in the county, or its surrounding peripheries. It could therefore be assumed that very few people might get access to employment to earn them a source of income. In that regard, these areas could fall under the low-income population in the county and hence their inability to afford such public transportation, hence, its absence in these areas. In spite of this, seasonal occasions in the summer or the need to acquire services in other settlement units within the county, such as Tartu city could make people in these areas still look for options to be able to get their needs met. This means that an alternative means of transport should be provided for them so that they do not miss out on important activities in other settlement units.

5. Conclusion and Recommendation

5.1 Conclusion

In conclusion, this research sought to characterise the spatial structure and hierarchy of public bus transportation, identify variations in the seasonal rhythm of commuter trips as well as the influence of seasonal rhythm on distance trips. Tartu County can be classified into three settlement units which are cities, towns and villages. The structure of the County was classified based on strength, structure and symmetry. In terms of strength, Tartu County shows a strong interaction between Tartu City (its capital) and other settlement units in the county, due to its dominance. Tartu city's dominance comes as a result of the high inflows it receives from other settlements owing to the specialised functions it provides, which meets the needs of the County's dwellers. The County's structure is also non-hierarchical since it is not monocentric, that is, it has varied nodes which also provide varied functions although these functions may differ in the level of importance. In that regard, Tartu County is classified as not fully polycentric but has a form of polycentricity (centralised node) where there are interdependencies between both dominant and non-dominant nodes with strong interactions existing between dominant nodes and weaker interactions existing between non-dominant nodes. Also, with regards to symmetry, the direction of flow in the County is multi-directional and this makes it symmetrical.

Tartu County also experiences a variation in the seasonal rhythm of public transport, and this is seen during the summer and winter seasons where the former is more pronounced. The summer period from June to August therefore shows a high variation with an increase in both inflows and outflows, to and from Tartu as opposed to the other times of the year due to activities that go on in this season. This throws light on changes in commuter behaviour at different times of the year.

In addition distance decay can be defied by the importance of certain demands irrespective of their location or the absence of alternative locations which provide similar goods or services demanded by people.

Areas with sparse population would have the disadvantage of not benefiting from transportation because of their income level since they cannot afford public transport services as compared to people in highly populated areas which are seen as hubs for high demand of transport services and economically viable for investment.

The study emphasises how much human flows shape the urban system of a place as they seek to move to locations where their needs can be met. Knowledge of such impact is necessary for urban planning in the allocation of infrastructure or development of transport networks. It is also relevant for predictive purposes in forecasting surges in transport demand at certain times of the year to adapt to changes in commuter behaviour.

5.2 Recommendations

From the findings of this research, the following recommendations could be taken into account.

- There should be a revision of the county's urban plan to include major nodes in some areas with high population densities to enable people in settlement units, whose distances are far to get immediate access to specialised functions.
- It is also recommended that there should be an increase in public transport services during the two seasons (summer and winter) owing to the increase in the number of commuter trips during this season
- For the sparsely populated areas without public transport connectivity, efforts should be made to give them some form of transport connectivity that is economically viable and is also affordable enough. On demand transport can be a suitable solution to help solve the problem.

Summary

Using public bus data to assess urban seasonal rhythm of commuters in Tartu County

William Gyamfi Kumi

An urban system refers to a set of interdependent geographical spatial units that are connected to each other through flows, economic activities, social interactions, and information. In this research urban systems were characterised by human spatial flows as its constraints (for example physical distance) compared to other forms such as money and information is not instantaneous over time (Castells, 1996). This therefore serves as a more reliable way to assess the behaviour of commuters over a period within space.

The objectives of this master thesis were to characterise the spatial structure and hierarchy of public bus transportation, assess variations in the seasonal rhythm of commuters and how it influences travel distance in Tartu County. To achieve this bus data was sourced from MTÜ Tartumaa Ühistranspordikeskus. Tartu County settlement units were used as spatial units for the analysis which was also sourced from the Estonian Land Board Geoportal. The analysis utilised the S-three spatial interaction theory by Limtanakool et. al., (2007) to describe the spatial structure and hierarchy of public transportation in Tartu County. An analysis was made to identify the variations in the seasonal rhythm and how it influences travel distance in Tartu County using exploratory factor analysis.

Using the S-three spatial dimension, Tartu city could be classified as a dominant node. This phenomenon highlights the importance of such a location in meeting the needs of people and thus attracting a lot of inflows (Alderson & Beckfield, 2004). Tartu city also shows a non-hierarchical spatial structure which is type A4 (Limtanakool et al., 2007) which represents a moderately centralised node. The spatial structure of Tartu County can be characterised as a dominant central settlement unit (Tartu city) with a non-hierarchical structure.

The results from the exploratory factor analysis showed that, there were two main factors which represented seasonal variations in the dataset for the flow of commuters. The variations indicated a distinct summer trend from June to August, with an increase in commuters from the peripheral areas to Tartu city and a decrease in commuters from Tartu city to its peripheral areas.

For distance analysis, it was realised that settlement units with close proximity to Tartu city had the most inflows and outflows compared to settlement units further away (59km to 80km) from Tartu city with the exception of Elva city (37km) which although was relatively far away from Tartu, had the highest inflows and outflows.

In conclusion Tartu County has a non-hierarchical structure with Tartu city as a dominant centre while other settlement units act as peripheries. Seasonality also affected the flow of people to and from the centre, owing to its specialised functions and activities. Distance could also affect the movement of people, with high inflows or outflows seen in areas of close proximity to Tartu city with the exception of some settlement units.

A limitation to this study was the exclusion of trip purpose in the data which could have been useful for making objective conclusion on commuter's flows.

Busside avalike andmete kasutamine Tartumaa pendelrändajate hooajalise rütmi hindamiseks

William Gyamfi Kumi

Kokkuvõte

Linnasüsteem viitab vastastikku sõltuvuses olevate geograafiliste ruumiüksuste kogumile, mis on omavahel seotud voogude, majandustegevuse, sotsiaalsete suhete ja informatsiooni kaudu. Selles uuringus iseloomustasid linnasüsteeme inimeste ruumilised voolud, kuna selle piirangud (näiteks füüsiline kaugus) võrreldes muude vormidega, nagu raha ja informatsioon, ei ole ajaliselt ainult hetkelised (Castells, 2011). Seetõttu on see usaldusväärsem viis, et hinnata pendelrändajate käitumist ruumis teatud perioodi jooksul.

Käesoleva magistritöö eesmärgiks oli iseloomustada ühistranspordi bussiliikluse ruumilist struktuuri ja hierarhiat, hinnata pendelrändajate sesoonse rütmi varieeruvust ja selle mõju sõidukaugustele Tartumaal. Selle saavutamiseks saadi bussiandmed Tartumaa Ühistranspordikeskuselt. Ruumiliste üksustena kasutati Tartumaa asutusüksusi, mis saadi Maa-ameti geoportaalist. Analüüsis kasutati Limtanakool *et al.*(2007) S-kolme ruumilise interaktsiooni teooriat, et kirjeldada Tartumaa ühistranspordi ruumilist struktuuri ja hierarhiat. Selgitamaks välja sesoonse rütmi kõikumised ja kuidas see mõjutab läbisõidukaugust Tartumaal, tehti uuriv faktoranalüüs.

Kasutades S-kolme ruumimõõdet, võiks Tartu linna liigitada domineerivaks sõlmeks. See nähtus rõhutab sellise asukoha olulisust inimeste vajaduste rahuldamisel ja seeläbi suuremate sissevoolude ligitõmbamisel (Alderson & Beckfield, 2004). Tartu linn on ka A4 tüüpi mittehierarhiline ruumistruktuur (Limtanakool *et al*., 2007), mis kujutab endast mõõdukalt tsentraliseeritud sõlme. Tartumaa ruumistruktuuri võib iseloomustada kui domineeriva keskusega (Tartu linn) mittehierarhilist struktuuri.

Uuriva faktoranalüüsi tulemused näitasid, et pendelrändajate andmestikus esindasid hooajalist varieeruvust kaks peamist tegurit. Erinevused näitasid selget suvetrendi juunist augustini, kus äärealadelt Tartu linna sõitjate arv suurenes ja Tartu linnast äärealadele sõitjate arv vähenes.

Kaugusanalüüsiks selgus, et Tartu linna vahetus läheduses asuvatesse asustusüksustesse oli võrreldes Tartu linnast kaugemal (59km kuni 80km) asuvate asustusüksustega sisse- ja väljarändeid rohkem. Ainsa erandina käitus Elva linn (37km), mis oli küll Tartu linnast suhteliselt kaugel, kuid mille sisse-ja väljarändevood olid suurimad.

Kokkuvõtteks on Tartumaal mittehierarhiline struktuur, kus Tartu linn on domineeriv keskus ning teised asustusüksused käituvad perifeeriatena. Sesoonsus mõjutas inimeste rändevooge keskusesse ja sealt tagasi selle spetsiifiliste funktsioonide ja tegevuste tõttu. Kaugus võib mõjutada ka inimeste liikumist, kuna suurim sisse- ja väljarändevool on Tartu linna vahetus läheduses, mõne üksiku erandiga.

Selle uuringu piiranguks oli reisi eesmärgi väljajätmine andmetest, mis oleks võinud olla kasulik objektiivsete järelduste tegemiseks pendelrände voogude kohta.

Acknowledgement

My foremost gratitude goes to GOD for granting me grace and strength. I also want to acknowledge my supervisor, Anto Aasa, PhD, for his meticulousness and guidance in supervising this project work, as well as the golden opportunity to conduct this research.

Also, I am grateful to my friends and family for their enormous backing.

Thank you to all persons whose views and recommendation have made it possible towards the completion of this thesis.

References

Abrate, G., Piacenza, M. & Vannoni, D. (2009). The impact of Integrated Tariff Systems on public transport demand: Evidence from Italy. *Regional Science and Urban Economics* 39, pp. 120–127.

Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., & Liu, Y. (2015). *International Journal of Geographical Everyday space – time geographies : using mobile phone-based sensor data to monitor urban activity in Harbin , Paris , and Tallinn*. *July*. https://doi.org/10.1080/13658816.2015.1063151

Alberini, A. & Chiabai, A. (2007). Urban environmental health and sensitive populations:How much are the Italians willing to pay to reduce their risks?. *Regional Science and Urban Economics* 37, pp. 239–258.

Alderson, A. S., & Beckfield, J. (2004). Power and Position in the World City System. In *American Journal of Sociology* (Vol. 109, Issue 4, pp. 811–851). https://doi.org/10.1086/378930

Alvarez, P., Canito, J.L., Moral, F.J. & Rodriguez, F.L. (2007). Determination of the infrastructure needs for municipalities using an objective method. *Computers and Industrial Engineering* 52, pp. 344–354.

Anable, J. (2005). Complacent Car Addicts or Aspiring Environmentalists? Identifying travel behaviour segments using attitude theory. *Transp. Policy.* 12, 65–78.

Ares, J. & Serra, J. (2008). Selection of sustainable projects for floodplain restoration and urban wastewater management at the lower Chubut River valley (Argentina). *Landscape and Urban Planning 85*, pp. 215–227.

Avid, D., & Batten, F. (1995). Network Cities: Creative Urban Agglomerations for the 21st Century. In *Urban Studies* (Vol. 32, Issue 2).

Barry, J, J., Newhouser, R., Rahbee, A., & Sayeda, S. (2002). Origin and Destination Estimation in New York City with Automated Fare System Data. *Transportation Research Record*. https://doi.org/10.3141/1817-24

Barry, J. J., Freimer, R., & Slavin, H. (2008). Using Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City. *Transportation Research Record*, *August*.

Batten, D. F. (1995) Network cities: creative urban agglomerations for the 21st century, Urban Studies, 32(2), pp. 313–327.

Barabási, A.-L., & Albert, R. (1995). Emergence of Scaling in Random Networks. In Mat. Res. Soc. Symp. Proc (Vol. 74). www.sciencemag.orgSCIENCEVOL28615OCTOBER1999

Barthélemy, M. (2011). Spatial networks. In Physics Reports (Vol. 499, Issues 1–3, pp. 1–101). https://doi.org/10.1016/j.physrep.2010.11.002

Berry, B. J.L. (1964). Cities as systems within systems of cities. Papers and Proceedings of the Regional Science Association, XIII, p. 147-164.

Berry, B. J.L. (1973). *Growth Centers in the American Urban System*, 2 vol. Cambridge (Mass.). Ballinger Publications.

Bertaud, A. (2002). *Note on Transportation and Urban Spatial Structure*. *April*, 1–11.

Boyle, D. (2008). Passenger Counting Systems. *Transportation Research Board of the National Academies, Washington, D.C, Transporta*.

Bourne, L. S., & Simmons, J. W. (1978). Systems of Cities: readings on structure, growth, and policy. Oxford University Press.

Byrant, F., P. Yarnold, et al. (1999). Statistical Methodology: VIII. Using Confirmatory Factor Analysis (CFA) in Emergency Medicine Research. *Academic Emergency Medicine* 6(1): 54-66.

Castells, M. C. (2010). Preface. In The rise of the Network Society: Vol.I.https://doi.org/10.1007/978-1-60327-951-2

Chen, Y.-Z., Li, N., & He, D.-R. (2007). A study on some urban bus transport networks. Physica A: Statistical Mechanics and Its Applications, 376, 747–754. https://doi.org/https://doi.org/10.1016/j.physa.2006.10.071

Christaller, W. 1933. Die Zentralen Orte in Süddeutschland. Jena: Fischer. Translated by C.W.

Baskin. 1966. Central Places in Southern Ger-many. Englewood Cli¤s: Prentice-Hall

Chopde, N., N. Nichat, et al. (2013). Landmark Based Shortest Path Detection by Using A* and Haversine Formula," International Journal of Innovative Research in Computer and Communication Engineering, vol. 1, no. 2, pp. 298-302.

Clark, W.A.V., Hosking, P.L. & Rankin, D.G. (n.d.). Notes on Factor Analysis and the development of Urban Theory

Cornelis, J. & Hermy, M. (2004). Biodiversity relationships in urban and suburban parks in Flanders. *Landscape and Urban Planning* 69, pp. 385–401.

De Goei, B., Burger, M. J., van Oort, F. G., & Kitson, M. (2010). Functional polycentrism and urban network development in the greater South East, United Kingdom: Evidence from commuting patterns, 1981-2001. *Regional Studies*, *44*(9), 1149–1170. https://doi.org/10.1080/00343400903365102

Dai, L., Derudder, B., & Liu, X. (2018). Transport network backbone extraction: A comparison of techniques. Journal of Transport Geography, 69, 271–281. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2018.05.012

Dieleman, F. M., & Faludi, A. (1998). Polynucleated metropolitan regions in Northwest Europe: Theme of the special issue. European Planning Studies, 6(4), 365–377. https://doi.org/10.1080/09654319808720468

Duncan O.D., Scott W.R., Lieberson S., Duncan B. & Winsborough H. (1960). *Metropolis and Region*. Baltimore, Johns Hopkins Press.

Fabrigar, L. R., D. T. Wegener, et al. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 4: 272-299.

Furth, G. P., & Day, F. (1985). Transit Routing and Scheduling Strategies for Heavy Demand Corridors. *Transportation Research Record 1011*, 23--2.

Feng, S., Hu, B., Nie, C., & Shen, X. (2016). Empirical study on a directed and weighted bus transport network in China. *Physica A: Statistical Mechanics and Its Applications*, *441*, 85–92. https://doi.org/10.1016/j.physa.2015.08.030

Gim, T.-H.T. (2015). The relationship between land use and automobile travel utility: A Multiple Indicators Multiple Causes approach. *Transp. Res. Part Transp. Environ.* 41, 188–204.

González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. https://doi.org/10.1038/nature06958

Goonetilleke, A., Thomas, E., Ginn, S. & Gilbert, D. (2005). Understanding the role of land use in urban stormwater quality management. *Journal of Environmental Management* 74, pp. 31–42.

Gorsuch, R. (1983). Factor Analysis., Hillsdale, NJ: Erlbaum.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. Journal of Personality Assessment 68(3): 532-560.

Haseeb,J. (2017). https://www.aboutcivil.org/Chrystaller-Central-Place-Theory

Hair, J., R. Anderson, et al. (1995a). *Multivariate data analysis.* New Jersey, Prentice-Hall Inc.

Hall, P. et al. (1973). *The Containment of Urban England, Urban and Metropolitan Growth Processes.* Vol.1. London, Allen and Unwin.

Hamidi , S. , & Ewing, R. (2014). A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States. *Landscape and Urban Planning,* 128, 72-82.

Harman, H. H. (1976). *Modern Factor Analysis.* The University of Chicago Press, Chicago., 3rd Editio.

Hartmann, K., Krois, J., & Waske, B. (2018). E-Learning Project SOGA: Statistics and Geospatial Data Analysis. Department of Earth Sciences, Freie Universitaet Berlin.

Haseeb.J. (2017). *Chrystaller Central Place Theory - Definition and Principles*. https://www.aboutcivil.org/Chrystaller-Central-Place-Theory

Hayton, J. C., D. G. Allen, et al. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods* 7: 191-205.

Henk, van E., Werner, B., & Erhard, E. (2006). *Survey Design: The Past, the Present and the Future*.

Hora, J., Galvão, T., Camanho, A., & Sobral, T. (2017). Estimation of Origin-Destination matrices under Automatic Fare Collection : the case study of Porto transportation system. *Transportation Research Procedia*, *27*, 664–671. https://doi.org/10.1016/j.trpro.2017.12.103

Huang, S.L., Lai, H.Y. and. Lee, L.C (2001). Energy hierarchy and urban landscape system. *Landscape and Urban Planning* 53, pp. 145-161.

Hunecke, M., Haustein, S., Grischkat, S. & Böhler, S. (2007). Psychological, sociodemographic, and infrastructural factors as determinants of ecological impact caused by mobility behavior. *J. Environ. Psychol.* 27, 277–292

Hurley, A. E., T. A. Scandura, et al. Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior* 18: 667-683.

Inbakaran, C., & Kroen, A. (2018). *Travel Surveys – Review of international survey methods*. *January 2014*.

Järv, O., Ahas, R., & Witlox, F. (2014). Understanding monthly variability in human activity spaces : A twelve-month study using mobile phone call detail records. *TRANSPORTATION RESEARCH PART C*, *38*, 122–135. https://doi.org/10.1016/j.trc.2013.11.003

Jeong, R. H. (2004). *The prediction of bus arrival time using automatic vehicle location systems data the prediction of bus arrival time*. *December*.

Jiang, B., & Claramunt, C. (2004). Topological Analysis of Urban Street Networks. Environment and Planning B: Planning and Design, 31(1), 151–162. https://doi.org/10.1068/b306

Johansson, C.H., Svedung, I. & Andersson, R. (2006). Management of risks in societal planning– an analysis of scope and variety of health, safety and security issues in municipality plan documents. *Safety Science* 44, pp. 675–688.

Johnson, R. & Wichern, D.W. (2007). Applied multivariate statistical analysis, Sixth ed. Pearson International Edition.

Joly, I. (2004). *The link between travel Time Budget and Speed: a Key relationship for urban space-time dynamics*. AET Eur. Transp. Conf. 2004 – ETC 2004 4 - 6 Oct. 2004 Strasbg. 19 p.

Juniper Research. (2011). *Mobile Ticket to Ride, Whitepaper extract from: Mobile Ticketing Applications & Markets, Transport, Sport & Entertainment 2009-2014.*

Kandt, J., Rode, P., Hoffmann, C., Graff, A. & Smith, D. (2015). Gauging interventions for sustainable travel: A comparative study of travel attitudes in Berlin and London. *Transp. Res. Part Policy Pract.* 80, 35–48

King, L.J. (1985). Central Place Theory. Reprint. Edited by Grant Ian Thrall. WVU Research Repository, 2020.

Kloosterman, R. C., & Musterd, S. (2001). The polycentric urban region: Towards a research agenda. *Urban Studies*, *38*(4), 623–633. https://doi.org/10.1080/00420980120035259

Kotz, A. J., Kittelson, D. B., & Northrop, W. F. (2015). *Novel Vehicle Mass-Based Automated Passenger Counter for Transit Applications*. https://doi.org/10.3141/2536-05

Kuutti, J. (2012). *A Test Setup for Comparison of People Flow Sensors*. Aalto University.

Lankford, P. M. (1972). Regional Incomes in the United States, 1929-1967. Research Paper no. 145. Chicago, University of Chicago, Department of Geography.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters, 87*(19), 198701-1-198701–198704. https://doi.org/10.1103/PhysRevLett.87.198701

Limtanakool, N., Dijst, M., & Schwanen, T. (2007). A theoretical framework and methodology for characterising national urban systems on the basis of flows of people: Empirical evidence for France and Germany. *Urban Studies*, *44*(11), 2123–2145. https://doi.org/10.1080/00420980701518990

Lin, J., & Ban, Y. (2013). Complex Network Topology of Transportation Systems. *Transport Reviews*, *33*(6), 658–685. https://doi.org/10.1080/01441647.2013.848955

Liu, W., Tan, Q., & Liu, L. (2020). Destination Estimation for Bus Passengers Based on Data Fusion. *Mathematical Problems in Engineering, 2020*. https://doi.org/10.1155/2020/8305475

Marcus Berliant, (2005), Central Place Theory, Urban/Regional, University Library of Munich, German

Mathews, E., & Poigné, A. (2008). An echo state network based pedestrian counting system using wireless sensor networks. Proceedings of the 6th Workshop on Intelligent Solutions in Embedded Systems, WISES'08. https://doi.org/10.1109/WISES.2008.4623302

Meeteren, M. Van. (2018). *Christaller and " big data ": Recalibrating central place theory via the geoweb*. *January*. https://doi.org/10.1080/02723638.2017.1298017

Melchiorre, C., Matteucci, M., Azzoni, A. & Zanchi, A. (2008). Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology* 94, pp. 379–400.

Mikulski, J. (2014). *Telematics – Support for Transport*.

MMBW. (1951). *Survey of daily movement of people within Greater Melbourne, Melbourne: Opinion Research Centre*.

Netemeyer, R., Bearden, W., & Sharma, S. (2003). *Scaling Procedures.*

Niedomysl, T. (2007). Promoting rural municipalities to attract new residents: An evaluation of the effects. *Geoforum* 38, pp. 698–709.

Nosoohi, I. & Zeinal-Hamadani, A. (2011). Urban Planning with the Aid of Factor Analysis Approach: The Case of Isfahan Municipali

Nunes, A, A., Dias, T, G., & e Cunha, J, F. (2016). *Passenger Journey Destination Estimation From Automated Fare Collection System Data Using Spatial Validation*. *17*(1), 133–142.

Oloyede, Alaya S M, & Adewole, K. (2014). Development of an Online Bus Ticket Reservation System for a Transportation Service in Nigeria. *Issn*, *5*(12), 2222–1719. www.iiste.org

Patuelli, R., Reggiani, A., Gorman, S. P., Nijkamp, P., & Bade, F. J. (2007). Network analysis of commuting flows: A comparative static approach to German data. *Networks and Spatial Economics*, *7*(4), 315–331. https://doi.org/10.1007/s11067-007-9027-6

Pett, M. A., N. R. Lackey, et al. (2003). *Making Sense of Factor Analysis: The use of factor analysis for instrument development in health care research*. California, Sage Publications Inc

Pinna, I., & Chiara, B. D. (2016). *Automatic passenger counting and vehicle load monitoring. February 2010.*

Pitombo, C. S., E.Kawamoto, et al. (2011). An exploratory analysis of relationships between socio-economic, landuse, activity participation, variables and travel patterns. *Transport Policy* 18: 347-357.

Pitombo, C.S. & Gomes, M.M. (2014). Study of Work-Travel Related Behavior Using Principal Component Analysis. *Open J. Stat*. 4, 889–901.

Plevka, V., Segaert, P., Tampère, C.M.J. & Hubert, M. (n.d.). Analysis of Travel Activity Determinants using Robust Statistics.

Pred, A. (1981). *Social Reproduction and the Time-Geography of Everyday Life Author (s): Allan Pred Published by : Taylor & Francis , Ltd . on behalf of the Swedish Society for Anthropology and Geography Stable URL : https://www.jstor.org/stable/490994 TIME-GEOGRAPHY OF*. *63*(1), 5–22.

Pred, A. (1977). *City-systems in advanced economies*. Past growth, present processes and future development options. Hutchinson,london

Predić, B., Rančić, D., Stojanović, D., & Milosavljević, A. (2007). *Automatic Vehicle Location in Public Bus Transportation System*. *July 2014*.

Proost, S. & Dender, K.V. (2001). The welfare impacts of alternative policies to address atmospheric pollution in urban road transport. *Regional Science and Urban Economics* 31,pp. 383–411.

Puhe, M., Edelmann, M., & Reichenbach, M. (2014). *Integrated urban e-ticketing for public transport and touristic sites* (Issue January).

Reuter, G, L. (2003). *Passenger Counting and Service Monitoring A Worldwide Survey of*.

Rezessy, S., Dimitrov, K., Vorsatz, D.U. & Baruch, S. (2006). Municipalities and energy efficiency in countries in transition Review of factors that determine municipal involvement in the markets for energy services and energy efficient equipment, or how to augment the role of municipalities as market players. *Energy Policy* 34, pp. 223–237.

Rezania, A. A., & Darnis, F. (2020). Use of Haversine Formula in Finding Distance Between

Temporary Shelter and Waste End Processing Sites Use of Haversine Formula in Finding Distance Between Temporary Shelter and Waste End Processing Sites. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1500/1/012104

Roca Cladera, J., Marmolejo Duarte, C. R., & Moix, M. (2009). Urban structure and polycentrism: Towards a redefinition of the sub-centre concept. *Urban Studies*, *46*(13), 2841–2868. https://doi.org/10.1177/0042098009346329

Rodrigue, J.-P. (2020). *The Geography of Transport Systems* (5th ed.). Routledge.

Rodrique.J. (n.d.). *Dept. of Global Studies & Geography, Hofstra University*.
Smith, D. A., & Timberlake, M. F. (2001). *World City Networks and Hierarchies, 1977-1997*.

Rodriguez, M.C., Blanco, J.L. & Salinas, E.M. (2010). A geomorphologic GIS-multivariate analysis approach to delineate environmental units, a case study of La Malinche volcano (Central Mexico). *Applied Geography.* Article in press, pp. 1–10.

Roth, C. & Kang, S. M. A. N. D. B. M. A. N. D. B. M. (2011). Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLOS ONE*, *6*(1), 1–8. https://doi.org/10.1371/journal.pone.0015923

Rui, Y., & Ban, Y. (2014). Exploring the relationship between street centrality and land use in Stockholm. *International Journal of Geographical Information Science*, *28*(7), 1425–1438. https://doi.org/10.1080/13658816.2014.893347

Sanam, K., Bagasrawala, M., Shaikh, F., & , Anamta, S. (2018). Smart E-Ticketing System for Public Transport Bus. 2018 International Conference on Smart City and Emerging Technology (ICSCET), 1–7. https://doi.org/10.1109/ICSCET.2018.8537302.

Schwind, P. J. (1971). *Migration & Regional Development in the United States.* Research Paper no. 133, Chicago, University of Chicago, Department of Geography.

Siahaan, (2016). Adjustable Knapsack in Travelling Salesman Problem Using Genetic Process, *International Journal of Science & Techknowledge*, vol. 4, no. 9, pp. 46-55.

Sienkiewicz, J., & Hołyst, J. A. (2005a). Statistical analysis of 22 public transport networks in Poland. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *72*(4). https://doi.org/10.1103/PhysRevE.72.046127

Simmons, J. (1981). Urban systems the new regional geography. In: Espace géographique, tome 10, no. 2,. pp. 135-142

Sims, R.E. & Schuetz, J. (2009). Local regulation and land-use change: The effects of wetlands by laws in Massachusetts. *Regional Science and Urban Economics* 39, pp. 409–421.

Sinclair, R. (1983). *Changing Spatial Trends at the National, Regional and Metropolitan Levels of the Settlement System in the Developed Western Countries* (p. 108). Polish Scientif. Publ.

Smith, D. A., & Timberlake, M. (1995). Conceptualising and mapping the structure of the world system's city system. *Urban Studies*, *32*(2), 287–302. https://doi.org/10.1080/00420989550013086

Soh, H., Lim, S., & Zhand, T. (2010). Wighted complex netwrok analysis of travelling routes on the Singapore public transportation system. *Physica A, 389*(24), 5852–5863.

Sohn, K. & Yun, J. (2009). Separation of car-dependent commuters from normal-choice riders in mode-choice analysis. *Transportation.* 36, 423–436.

Statistics Estonia. (2017). *Population Census*. 2011.

Steg, L. (2005). Car use: lust and must. Instrumental, symbolic and affective motives for car use. Transp. Res. *Part Policy Pract.* 39, 147–162.

Stopher, P., Bullock, P., & Rose, J. (2003). *Simulating household travel survey data in Australia : Adelaide case study Simulating household travel survey data in Australia : Adelaide case study* (Issue September).

Sun, D., Zhao, Y., & LU, Q. C. (2015). Vulnerability analysis of urban rail transit networks: A case study of Shanghai, China. *Sustainability (Switzerland)*, *7*(6), 6919–6936. https://doi.org/10.3390/su7066919

Swain H. & MacKinnon R.D. (eds.) (1975). Proceedings of the Conference on National Settlement Systems and Strategies. IIASA CP-75-4. Laxenburg, Austria, International Institute for Applied Systems Analysis.

Tabachnick, B. G. & L. S. Fidell (2001). *Using multivariate statistics.* Needham Heights, MA, Allyn & Bacon.

Tao, S., Rohde, D., & Corcoran, J. (2014). *Examining the spatial – temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap*. *41*, 21–36.

Tavernia, B.G. & Reed, J.M. (2009). Spatial extent and habitat context influence the nature and strength of relationships between urbanization measures. *Landscape and Urban Planning* 92, pp. 47–52.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications.* Washington, DC, American Psychological Association.

Townes, M. S., & Hunter-zaworski, K. (1998). *Passenger Counting Technologies And Procedures*.

Turner, K., Lefler, L. & Freedman, B. (2005). Plant communities of selected urbanized areas of Halifax, Nova Scotia, Canada. *Landscape and Urban Planning* 71, pp. 191–206.

Ubbels, B. & Verhoef, E.F. (2008). Governmental competition in road charging and capacity choice. *Regional Science and Urban Economics* 38, pp. 174–190.

Van, H.T. & Fujii, S. (2007). Psychological Determinants of Behavioral Intention to Use Travel Modes in Ho Chi Minh City. Proc. East. *Asia Soc. Transp. Stud.* 2007, 98–98.

Virrantaus, K., & Patenková, O. 6. (2008). Identifying Critical Locations in a Spatial Network with Graph Theory. *Transactions in GIS*, *12*(1), 61–82.

Wang, Y.P., Wang, Y. & Wu, J. (2009). Urbanization and Informal Development in China: Urban Villages in Shenzhen. *International Journal of Urban and Regional Research* 33.4, pp. 957–73.

Wang, J., Mo, H., Wang, F., & Jin, F. (2011). Exploring the network structure and nodal centrality of China's air transport network: A complex network approach. *Journal of Transport Geography, 19*(4), 712–721. https://doi.org/10.1016/j.jtrangeo.2010.08.012

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442. https://doi.org/10.1038/30918

Wedel, M. & Kamakura, W.A. (2000). Market Segmentation. Springer US, Boston, MA.

Williams, B., T. Brown, et al. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine* 8(3).

Wittwer, R. (2014). Cluster-analytical-creation of a Typology of Young Adults 'Travel Behavior in Germany. *Procedia - Soc. Behav. Sci.* 160, 64–73.

Wu, F. (2009). Land Development, Inequality and Urban Villages in China. *International Journal of Urban and Regional Research* 33.4, pp. 885–9.

Xu, X., Hu, J., Liu, F., & Liu, L. (2007). Scaling and correlations in three bus-transport networks of China. *Physica A: Statistical Mechanics and Its Applications*, *374*(1), 441–448. https://doi.org/10.1016/j.physa.2006.06.021

Xu, J., Yeh, A. & Wu, F. (2009). Land commodification: new land development and politics in China since the Late 1990s. *International Journal of Urban and Regional Research* 33.4, pp. 890–913.

Xuehao Chu. (2010). *A Guidebook for Using Automatic Passenger Counter Data for National Transit Database (NTD) Reporting*. *December*.

Yang, X., Fang, Z., Yin, L., Li, J., Zhou, Y., & Lu, S. (2018). Understanding the spatial structure of urban commuting using mobile phone location data: A case study of Shenzhen, China. *Sustainability (Switzerland)*, *10*(5). https://doi.org/10.3390/su10051435

Yang, X. H., Chen, G., Chen, S. Y., Wang, W. L., & Wang, L. (2014). Study on some bus transport networks in China with considering spatial characteristics. *Transportation Research Part A: Policy and Practice*, *69*, 1–10. https://doi.org/10.1016/j.tra.2014.08.004

Yay, O.D., Alagha, O. & Tuncel, G. (2008). Multivariate statistics to investigate metal contamination in surface soil. *Journal of Environmental Management* 86, pp. 581–594.

Yeh, A.G.O. and Wu, F. (1996). The new land development process and urban development in Chinese cities. International Journal of Urban and Regional Research 20.2, pp. 330–53.

Zebardast, E., & Ghanooni, H. (2019). An Analysis of Urban Sprawl Using Factor Analysis Technique (Case: Qazvin City Districts). *Urban Economics and Management*, 7(2(26)), 65-84.

Zhang, Y., & Yao, E. (2015). *Splitting Travel Time Based on AFC Data : Estimating Walking ,* *Waiting , Transfer , and In-Vehicle Travel Times in Metro System*.

Zhang, H., Zhao, P., Gao, J., & Yao, X. M. (2013). The analysis of the properties of bus network topology in beijing basing on complex networks. *Mathematical Problems in Engineering, 2013*. https://doi.org/10.1155/2013/694956

Zhang, J., Wang, S., & Wang, X. (2018). Comparison analysis on vulnerability of metro networks based on complex network. *Physica A: Statistical Mechanics and Its Applications*, *496*, 72–78. https://doi.org/https://doi.org/10.1016/j.physa.2017.12.094

Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, *28*(11), 2178–2199. https://doi.org/10.1080/13658816.2014.914521

Zhu, G., Corcoran, J., Shyy, P., Pileggi, S. F., & Hunter, J. (2018). Analysing journey-to-work data using complex networks. *Journal of Transport Geography*, *66*, 65–79. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2017.11.006

Non-exclusive licence to reproduce thesis and make thesis public

I, William Gyamfi Kumi,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright "**Using public bus data to assess urban seasonal rhythm of commuters in Tartu County**" supervised by Anto Aasa, PhD.
2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

William Gyamfi Kumi

Tartu, 30.05.2022