University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

Master Thesis in Geoinformatics for Urbanized Society (30 ECTS)

Measuring urban connectivity using bike-share data: network analysis approach

Muhammad Hamza Zubair

Supervisors: Veronika Mooses, MSc
Siiri Silm, PhD

Tartu 2021

Measuring urban connectivity using bike-share data: network analysis approach

Abstract

Connectivity refers to the ability or condition to create and maintain connections between spatial points Having high connectivity between different neighbourhoods in the city is essential for sustainable and liveable cities. While being one of the core characteristics of networks, connectivity has been so far mainly applied for investigating the transportation infrastructure (e.g. street networks). However, a focus on infrastructure network only does not tell how well different parts of the city are actually connected in terms of human mobility. For this reason, the current study aims to measure urban connectivity in Tartu.

To achieve this aim, bike-share GPS data has been used by applying social network analysis. In this study, different approaches were used in constructing the mobility flows between districts. In the **aggregated approach,** only the origin and destination districts of a bike-share trip were considered, and in the **detailed approach**, all the districts that constituted the trip were included for constructing the mobility flows. Social network analysis metrics were then applied to the mobility flows for measuring urban connectivity.

The results revealed that the city's central districts were more connected to each other in terms of trips and users movements than with the other districts, reflecting Tartu city's monocentricity. The outskirt districts showed poor connectivity in comparison to central districts. Also, as part of the study, both approaches were compared, and it was found that overall, the detailed approach was better as the results were not confined only to the districts with docking stations. It means that this approach enables to include these districts into the analysis that do not have docking stations and thus, give a more accurate picture of urban connectivity.

Keywords: urban connectivity, social network analysis, origin-destination matrices, spatial mobility, public transportation system, bike-share system

CERCS code: S230 – Social geography

Linnalise ühenduvuse mõõtmine rattaringluse andmetel: võrgustiku analüüsi lähenemisviis

Lühikokkuvõte

Ühenduvus viitab võimele või tingimustele luua ja säilitada seoseid ruumis paiknevate punktide vahel. Jätkusuutlike ja elamiskõlblike linnade jaoks on oluline erinevate linnaosade omavaheline kõrge ühenduvus. Ühenduvust kui võrgustike ühte põhiomadust on aga siiani peamiselt rakendatud transpordi infrastruktuuri (näiteks tänavavõrgud) uurimiseks. Keskendumine ainult infrastruktuuri sidususele ei anna aga teadmist, kui hästi on erinevad linnaosad inimeste liikuvuse alusel ühendatud. Sellest lähtuvalt on antud uuringu eesmärgiks mõõta linnalist ühenduvust Tartus inimeste tegeliku liikuvuse põhjal.

Eesmärgi saavutamiseks on sotsiaalse võrgustiku analüüsi teel uuritud rattaringluse GPS-andmeid. Uuringus kasutatakse erinevaid lähenemisviise liikumisvoogude loomiseks asumite vahel. **Agregeeritud lähenemisviisi** kasutamisel arvestati üksnes rattaringluse reiside lähte- ja sihtkoha piirkondi ning **detailse lähenemisviisi** kasutamisel kaasati liikumisvoogude tekitamiseks kõik reise moodustanud asumid. Sellele järgnevalt rakendati linnalise ühenduvuse mõõtmiseks sotsiaalse võrgustiku analüüsi mõõdikuid.

Tulemustest selgus, et linna keskosad olid reiside ja kasutajate liikumise osas omavahel rohkem ühendatud võrreldes teiste asumitega, mis peegeldab Tartu linna ühekeskuselist struktuuri. Äärelinna asumite ühenduvus võrreldes linna keskel paiknevate asumitega oli madal. Uuringus võrreldi ka mõlemat lähenemisviisi ning leiti, et detailne lähenemine on parem, sest see annab võimaluse analüüsi kaasata ka neid piirkondi, kus ei ole rattaringluse parklaid. See lähenemisviis annab seega täpsema pildi linnalisest ühenduvusest.

Märksõnad: linnaline ühenduvus, sotsiaalsete võrgustike analüüs, lähte-sihtkoha maatriksid, ruumiline liikuvus, ühistranspordisüsteem, rattaringluse süsteem

CERCS klassifikaator: S230 - Sotsiaalne geograafia

Table of contents

Introduction

Connectivity refers to the ability or condition to create and maintain connections between spatial points (Reggiani et al., 2015). The concept of connectivity originates from ecology and landscape ecology that refer to the structure and strength with which species and resources disperse, migrate and interact across patches (Biggs et al., 2015). The connectivity concept, however, also gains importance in studying urbanised areas. For example, as urban growth is linked closely to the fragmentation of natural landscapes, it is suggested to study urban sprawl in terms of urban connectivity (Behnisch et al., 2019). In the context of urban ecology, connectivity depicts incidental contact with or direct exposure to the natural or built environment in cities (Kareem, 2017). The quality or condition of being connected or connective, is closely related to urban morphology, design and regeneration (Nel et al., 2018). Having high connectivity between different neighbourhoods in the city is essential for sustainable and liveable cities (Australia et al., 2018).

To understand the place, it is necessary to understand the flows and to understand the flows one should understand networks (Press, 2021). An intrinsic characteristic of connectivity is that it is related to networks, which are a set of interconnected nodes. A node can be a person, firm, city, country or other spatial entity (Kunaka, 2019). Connectivity can be explored at different scales, from the local to the regional and global scales. Connectivity is, therefore, an attribute of a network and is a measure of how well connected any one node is to all other nodes in the network (Kunaka, 2019).

Tracing human mobility flows is one crucial indicator of urban connectivity, which is a topic of deep interest for urban planners, policymakers, and geographers. Human mobility refers to the movement of an individual or a group of people from their origin geolocation to their destination (Ebrahimpour et al., 2020), and it has a tremendous impact on the economy, society and environment (Li et al., 2020). Understanding how and where people move in cities has implications for urban policy and planning (Galpern et al., 2018), namely estimating commuting flows, traffic forecasting and urban planning (Barbosa et al., 2018). Analysing mobility flows is necessary to understand urban connectivity which helps in better spatial and urban planning, namely in the fields of transport and infrastructure. Transport plays a crucial role in the functioning of cities because it provides access for people to different activities, including education, markets, employment, recreation, health care and other key services (Makarova et al., 2017). Sustainable cities cannot exist without sustainable mobility which envisions more integrated and multimodal public transport systems and with land use distribution matching the needs of the population, business and institutions, shifting mobility to active transportation modes such as walking and cycling and public transport (Gil & Read, 2013).

In terms of connectivity, there is already a substantial research on studying urban connectivity on the level of transportation and street infrastructure (Dill, 2004; Makarova et al., 2017; Yang et al., 2019) which is one of the core characteristics of urban networks, but there are less studies which utilise human mobility data (Barbosa et al., 2018). The focus on infrastructure networks does not tell how well different parts of the city are actually connected in terms of human mobility, i.e., how much people are using the streets and light traffic lanes for example. With the increasing use of GPS, mobile positioning, social media and other means of ICT, there is a massive amount of geospatial data that enable researchers to study individual and collective mobility flows (Willberg, 2019, Lu et al., 2019; Galpern et al., 2018; Silm, Jauhiainen et al., 2020; Barbosa et al., 2018; Lin et al., 2019). The public bike-share system is an active mode of transport that is gaining popularity in different parts of the world due to its social, economic, environmental and health-related benefits. Most bike-share systems are equipped with GPS devices that collect spatio-temporal data on bike-share trips (Mooney et al., 2019), providing a good opportunity to identify well connected and poorly connected parts of the city that need special attention.

This research aims to **measure the connectivity between the districts of Tartu** using bike-share data provided by Tartu Smart Bike System (hereinafter: TSB) which operates as part of the city’s multimodal transport system. The study examines mobility flows in terms of the number of trips and users and applies social network analysis (hereinafter: SNA) metrics to describe the connectivity between the districts of Tartu. In this study the city is considered as a directed network with the districts of the city are considered as nodes of this network. The reason for applying social network analysis is that network analysis helps to find the most important elements and their interactions on a network level (Haznagy et al., 2015). This study captures the urban connectivity from two angles. In the **aggregated approach,** the connectivity is estimated based on bike-share trips between origin and destination districts. In the **detailed approach**, the connectivity is measured by including all districts that constitute a bike-share trip. A comparison of both methods for describing connectivity in the study area is provided.

Research questions are as follows:

- Which districts are more connected and least connected with each other?
- What is the extent of connectivity between all districts? Are there bridges or hubs in the city network?
- What is the connectivity of the whole network?

This research contributes to future mobility studies that utilise bike-share datasets and it can also help in infrastructure planning. It is important for planners to identify districts that are highly connected with other parts of the city and act as hubs and also those areas that are poorly connected and need special attention. Bike-share based connectivity can show the dynamics of urban flows and it has the potential to reflect the existing socioeconomic disparities among different neighbourhoods of the city (Wang & Su., 2019; Yang et al., 2019). This study can be used as a reference point for choosing the best suitable approach for utilising bike-share data for measuring urban connectivity.

In the first part of the thesis, a concise overview of urban connectivity, mobility and the use of social network analysis in urban connectivity and mobility studies is outlined. In the second part, an overview of the data and used methods is provided. The third part outlines the results according to the research questions and lastly comes the discussion and conclusion part.

1. Theoretical background

1.1 Human mobility and urban connectivity

While the term mobility has multiple connotations, in the context of this study, it refers to the movement of human beings (individuals as well as groups) in space and time (Barbosa et al., 2018; Wang et al., 2019). Different transport modes are used for human mobility. The transport modes can be either human-powered (walking) or using cars, public transit systems, or air or water transport. Studies conducted in Europe and the United States found that the average household spending on transportation is between 15 and 25 per cent of the total expenditures, making transportation the second-largest expenditure category after housing (Barbosa et al., 2018). Transportation is also the second source of greenhouse gas emissions to the atmosphere. From these few examples, it should be clear that mobility has an enormous impact on human societies. An accurate quantitative description of human mobility is of fundamental importance to understand the processes related to human movement and their impact on the community and the natural and living environments (Barbosa et al., 2018).

The emerging concept of “New Urbanism” strongly advocates the reintegration of spatial form and built-up environments to generate a perfect neighbourhood to support active forms of human mobility (Lu et al., 2019). According to the Physical Activity Through Sustainable Transport Approaches report (Mueller et al., 2015), active mobility reflects the sense of safety, resulting from favourable urban infrastructure policies and it translates into a shift away from the use of motorised transport. The knock-on effect of this is a reduction in air pollution and traffic noise levels in cities and an overall improvement in the quality of life for all residents. Active mobility is a favourable public policy towards sustainable urban development and human health. Most inner-urban car trips have <5 km in European cities, which active modes of transport can easily do. Cycling is one of the popular modes of active transport. Cycling has been one of the best examples in the scope of spatial mobility and transportation research where improved data availability and versatility has met the great societal need for knowledge (Willberg, 2019.). The primary reason for the increase in the cycling trend is the public bike system in which a network of bicycles is made available for shared use to the public on a short-term basis.

The urban form has diverse impacts on travel distance and mode. The sustainable urban form concept is a useful starting point in determining sustainability criteria and factors for a neighbourhood (Moroke et al., 2019). Compact development patterns, featured by relatively high population density, mixed land use, and easily accessible facilities, are beneficial for walking, cycling and public transportation (Song et al., 2017). The relationship between neighbourhood design, land use, and transportation has been studied extensively and across many disciplines. The connection between built form and physical activity is found throughout public health and transportation planning research. As a result, there is now a general consensus that an undeniable connection exists between transportation decisions and the built form (Rybarczyk & Wu, 2014). The spatial distribution and temporal dynamics of urban activity density are fundamental determinants of urban mobility (Psaltoglou & Calle, 2018). Measuring human mobility can help study how much different urban areas are connected or disconnected, reflecting the underlying urban form. In the most general sense, connectivity is about relatedness (Peponis et al., 2008). It can help in measuring the degree of urban connectivity between different areas of the city.

1.2 Data and methods used in measuring mobility and connectivity

The traditional method of assessing urban connectivity relies on the street network infrastructure. Advocates of New Urbanist and neo-traditional planning concepts include street connectivity as a critical component for good neighbourhood design. More grid-like street networks are preferred over networks that include many cul-de-sacs and long blocks, thus increasing distances between destinations (Dill, 2004). Peponis et al., (2008) focused in their research on street connectivity on the basis of intersections and directional distances. Extensive road coverage and the network robustness in maintaining the connectivity of urban systems. In their study, Mohamad & Said (2014) used centrality measures on street infrastructure for measuring connectivity. Usually, the neighbourhood connectivity is also measured by the notion of how cohesive is the street infrastructure (Watts et al., 2015). However, the connectivity can also be measured using mobility flows based on various datasets. With the advent of mobile communication and wireless network technology and the availability of open geolocation data, a wide range of human digital footprints are now easily accessible. There is a plethora of research that uses different datasets for mobility studies. Wang et al., (2019) complied in a comprehensive study using various datasets for mobility studies, including census, banknotes, passive mobile positioning data (for example call detail records CDR), GPS data, social media check-in data, and public transport transaction data.

Barbosa et al. (2018) wrote about using **census data** where periodically collected census data's questions related to workplace location and current and previous residences are analysed for mobility studies. Wang et al. (2019) explained the use of **banknotes** in studying human travelling statistics. **Call details record (CDR)** is perhaps the most important dataset of the last decade for inferring human mobility with a high granularity than, for example, census data (Barbosa et al., 2018). CDR data is also sometimes referred to as passive mobile positioning data. Ahas et al., (2010) described how the passive mobile positioning data could be used for mobility studies. Passive mobile positioning data is normally collected with the precision of network cells. Mobile operators can aggregate anonymous geographical data from log files, such as location points or movement vectors, and researchers can use for scientific purposes. Another study (Silm, Jauhiainen, et al., 2020) used mobile phone data for cross border mobility analysis. CDR data can also be useful for identifying interregional mobility patterns. One of the studies by Ahas et al., (2010) developed a methodology for monitoring the population's short-term mobility using mobile positioning data. Mobile positioning dataset can also be used to detect spatial and temporal differences in everyday activities in cities (Ahas et al., 2015). However, Some drawbacks of passive mobile positioning data are that it usually has no or little background information on the phone users, and access to data is limited, while mobile network operators are hesitant to provide their data (Silm, Järv, et al., 2020).

In smart cities, smartphones and vehicles equipped with a GPS receiver can record the trajectories of user's movement with a high degree of accuracy and continuous spatiotemporal resolution (Wang et al., 2019). These GPS logs data can be used for connectivity studies. One of the studies used vehicle **GPS data** for inter, and intra-city connectivity, a subset of the GPS logs of commercial vehicles, such as trucks and taxis, collected by Toyota Tsusho Nexty Electronics Thailand was used to analyse the temporal variation of mobility (Miyazaki, 2019).

Social media check-in data is also a valuable source collected by social network providers. The geotagged data is made up of coordinates, time, photos, and comments. Therefore, the movement trajectories of users can be obtained from the sequence of published locations. Through analysing the information, some important metrics, i.e., the radius of gyration, jump length and visit frequency can be calculated (Wang et al., 2019). In their research, Lin et al., (2019) used Sina Weibo, a Chinese microblogging website like Twitter for measuring inter-city connectivity.

The public transport system is becoming more and more developed, and it makes people's lives more convenient in urban cities. People usually carry a smart card to travel for social activities, which generate massive trip information, e.g., card I.D., trip origin, boarding time, trip destination, alighting time, and trip expense. This public transport transaction data captures people’s travel behaviour precisely and provides a new source for exploring human mobility patterns (Wang et al., 2019).

Bike-share systems generally collect data in the form of GPS records (Song et al., 2021), which have been widely applied in spatial mobility studies (e.g. Willberg., 2019; Chen et al., 2015; Li et al., 2020). Some studies that focus specifically on urban form and connectivity with bike-share data have also emerged; however, there are few examples. In their study, Hong et al., (2015) used bike trip data in Manhattan bike share data considering bike stations as nodes and connections between them as edges to form a spatio-temporal graph. They calculated “black holes” – a sub graph of the spatio-temporal graph with the overall inflow more significant than the overall outflow – and “volcanos” - a subgraph with the overall outflow greater than the overall inflow by a threshold. Jurdak (2013) focused on characterising the direct impact of cost on urban mobility and compared its relative importance with network topology using docked bike-share system data for Boston and Washington cities. In another similar study by Borgnat et al., (2011), they used the V´elo’v system, a docked bike-share system in Lyon, France. In this study, the bike-share system was interpreted as a dynamical network for quantitative analysis of movements using bicycles in the city. Wang & Su (2019) analysed neighbourhood connectivity using human mobility patterns of bike share data in Washington DC. They measured how well bike-sharing helps riders travel between different parts of the city. They counted how many different neighbourhood destinations each starting neighbourhood generated and the number of different starting neighbourhoods reflected in each receiving neighbourhood. They called this measure of exposure to different neighbourhoods as “destination diversity” and “receiving diversity”.

The movements of people, vehicles, or goods from one geographical location to another can be aggregated and visualised with origin-destination (OD) matrices (Caceres et al., 2007). OD matrices can be essential for mobility flow spatiotemporal analyses (Song et al., 2021). When OD matrices are useful for simplifying and visualising the huge amount of data, further network analysis helps to find the most important elements and their interactions on a network level (Haznagy et al., 2015).

1.3 Network theory & social network analysis

1.3.1 Network theory

Network theory is the study of graphs or network. The terms network and graph are used interchangeably (Barabási, 2014). A network consists of actors or nodes and a set of ties of a specified type that link them (Sine, 2011). Graph theory is the study of graphs, which are mathematical structures used to model pairwise relations between objects and a network can be defined as a graph with node/vertex and edges, where nodes are the object and edges are the relationships or links connecting those nodes. Representing a problem as a graph can make a problem much more straightforward. More accurately, graph theory can provide the appropriate tools for solving the problem. Network theory is part of graph theory, and very often, the terms network and graphs are used interchangeably (Table 1).

Table 1: Synonymous terms in network science and graph theory (Barabási, 2014)

Network Science	**Graph Theory**
Network	Graph
Node	Vertex
Link	Edge

1.3.2 Social network analysis

Network theory is the backbone of the social network analysis (SNA). Social network analysis techniques help in understanding the dependencies between social entities in the network, characterising their behaviours and their effect on the whole network and over time (Tabassum et al., 2018). In simple words, SNA is the process of investigating social structures using network theory.

Two primary parameters of a network are nodes and links. The number of nodes represented by N represents the objects of the network. N is called the size of the network. To distinguish the nodes, they are labelled with i = 1, 2, ..., N. Number of links are often denoted by L, representing the number of total interactions between the nodes. Links help in quantifying the interaction between the nodes. Figure 1 shows a simple network with the points showing the nodes, and the lines connecting the points are links of the network. The network shown in Figure 1 has N=4 and L=6.

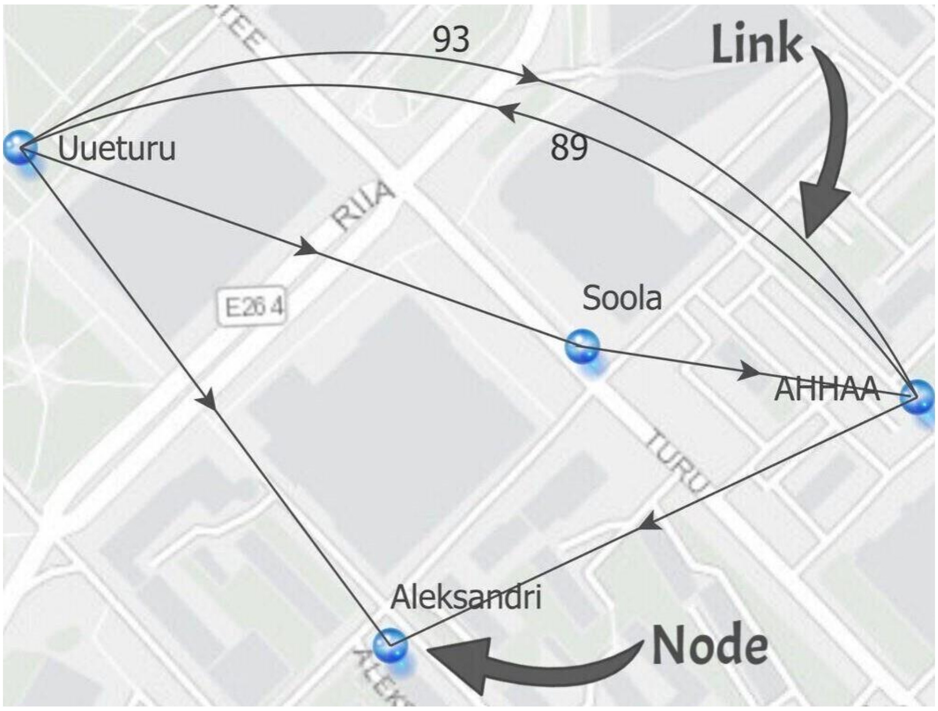


Figure 1: Example of a network with nodes and links

The links of a network can be directed or undirected. Some systems have directed links, like the WWW, whose uniform resource locators (URL) point from one web document to the other, or phone calls, where one person calls the other. Other systems have undirected links, like romantic ties: if I date Janet, Janet also dates me, or like transmission lines on the power grid, on which the electric current can flow in both directions (Barabási, 2014). A network is called directed (or digraph) if its links are directed and undirected if all its links are undirected. If the weights are assigned to all the node's edges, it becomes a weighted network. Figure 1 shows the network with the inflow and outflow information between the nodes. It is showing a directed network of four nodes and six directed links. Here nodes are docking station names in the city of Tartu. This network shows that 89 interactions (in this case, number of people) were made between AHHAA to Uueturu and 93 interactions were made between Uueturu to AHHAA which are called weights of the links in SNA terminology. The directed networks help quantify and analyse the patterns and are important in calculating flow movements.

Social Network Analysis (SNA) has become a widely applied method in research and business for inquiring into the web of relationships on the individual, organisational and societal level (Ghali et al., 2012). Wu & Kim (2020) emphasised using complex network theory to examine the network structure of the bike-sharing system and quantify the important nodes (bike stations) in the network. Shanmukhappa et al., (2018) proposed a novel methodology called supernode graph structuring for modelling the bus transport network. Supported by the social network analysis method, Zhang et al., (2020) used the Tencent Location Big Data for analysing the spatial patterns of interregional population mobility among 328 Chinese cities and aimed to identify the influential factors associated with population flows.

The network properties describe the overall structure of a network quantitatively. Centrality indicators, such as degree, are among the widely adopted measures. Degree centrality offers a microscopic view of a network: for example, the locations of influential nodes that exert attractive power on their neighbouring nodes and the presence of these nodes are critical to the stability of the whole network (Song et al., 2021). Zhang et al., (2020) used degree centrality, a metric of SNA, to analyse the hierarchy of population flow. It helped in the characterisation of population mobility networks. Centrality is an index to study the status and function of each region in a network. The more central a region is in a network, the more “influence” it has in the network.

1.4 Transportation system as a network

In urban and transport studies, public transportation systems have been often viewed as complex networks where components are represented by nodes and their interconnections by arcs or links (i.e. edges) (Yang et al., 2019). Interpreting the changes in flow networks is important for understanding changes in urban dynamics (Yang et al., 2019). Complex networks highlight the topological characteristics of the system structure and the geometric properties of complex systems play a primary role in the network’s dynamic behaviour. Public transportation can be depicted by traffic flows on complex networks (Jia et al., 2019).

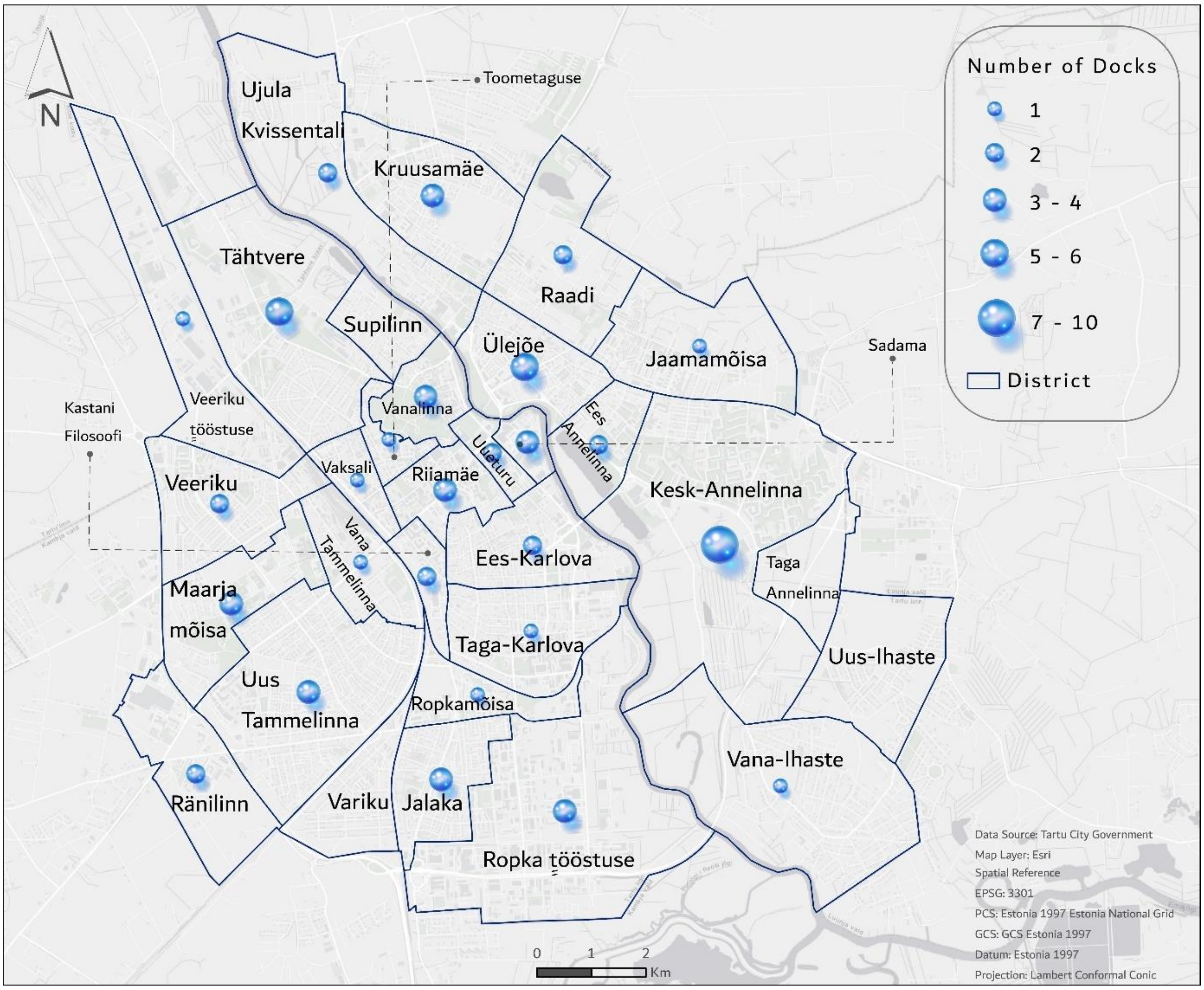
The public bike system, part of multimodal public transport system is gaining popularity as an alternative transportation mode in many countries. It has health benefits (Woodcock et al., 2014) as well as it brings several social, environmental, and economic benefits such as saving transportation time and expenses, alleviating traffic congestion, reducing greenhouse gases (GHG) emissions and air pollutants, improving multimodal transport connections, and increasing the productivity of local economic activities (Bullock et al., 2017). Currently, two main service models of bike-share exist, namely, docked and dockless. Docked bike-share schemes first gained popularity in the late 1990s and require users to start and end trips at a bike-share docking station. Dockless systems were developed as a response to the difficulty associated with accessing docking stations (Mooney et al., 2019).

The bike-sharing network is different from other public transit networks (e.g., bus and metro). Transport planners plan the routes of public transit systems, but the movement path of bikes in the bike-sharing networks is determined by travellers. Thus, the bike-sharing network is self-organised (Wu & Kim, 2020). Yang et al., (2019) converted the bike data into a network to capture the impact of metro service on the dockless bike-sharing system. This study applied a combination of geo-statistical and network theory approaches. Another study (Song et al., 2021) analysed the spatial-temporal dynamics of the bike-share system from the perspective of trip spatial patterns and linkage between traffic zones. This approach helps understand the trip demands of bike-share and model the spatiotemporal dynamics of cycling flows by representing the dockless bike-share system as a directed network to model the linkage among cycling flows in different regions. In one of the study, (Yao et al., 2019) analysed bike-share system network by applying social network analysis measures of comparing degree, strength, radiation distance and community structure for the network.

2. Data and methods

2.1 Study area

The study area in this research is Tartu, the second-largest city of Estonia. Tartu is officially divided into 17 bigger administrative units (“linnaosad” in Estonian) and 31 smaller administrative units (“asumid” in Estonian). In this research 31 smaller administrative units (hereinafter referred to as "districts") were chosen (Figure 2). The reason for choosing districts with smaller spatial coverage can, firstly, give a better overview of which parts of the city are well connected and which are not. Poor connectivity of a bigger spatial unit can occur due to poor connections of smaller spatial units inside that district. Secondly, using smaller districts requires less aggregation of docking stations. The information of docking stations, aggregated on the district level, is shown in Figure 2.



Data Source: Tartu City Government
Map Layer: Esri
Spatial Reference
EPSG: 3301
PCS: Estonia 1997 Estonia National Grid
GCS: GCS Estonia 1997
Datum: Estonia 1997
Projection: Lambert Conformal Conic

Figure 2: Tartu districts and bike-share docking stations.

2.2 Data

The primary data used in this research was provided by TSB a public, self-service bike share system, a part of multimodal transport system established in Tartu city for short trips with a dedicated docking stations network around the city to offer an environmentally friendly commute. There were 69 docking stations and 750 bikes in the used dataset, where two-thirds of the fleet was equipped with electric-assist motors that provide riders with an extra boost when pedalling. Every bicycle has a GPS device that captures location after every 5 seconds.

The **GPS file** used for this analysis consisted of GPS tracks of the bike-share trips. It contained information on 692,347 trips. This dataset contained route codes, cycle number, latitude, longitude, coordinates date, coordinates, time and user ids (Table 2): 1) **route code** contained a unique code assigned to each trip; 2) **user id** contained unique anonymous id's assigned to each registered user of TSB; 3) **cycle number** was also evident but this variable was not used in the study; 4) spatial information in latitude and longitude form in the WGS-84 coordinate system, which was converted into the Estonian coordinate system of 1997 (EPSG:3301); 5) **coordinate date** field contained the date of the trip; 6) **coordinate time** field contained time for each captured point of each trip's trajectory. Based on this field all the captured points were combined to form the trajectories for each trip. Table 2 shows the sample data of the GPS file.

Table 2: Sample data from GPS log file

Route code	**Cycle number**	**Latitude**	**Longitude**	**Coordinate date**	**Coordinate time**	**User id**
1563397276221	2519	58.37809	26.72629	7/18/2019	00:01:53+00	1
1563397276221	2519	58.37809	26.72629	7/18/2019	00:01:58+00	1
1563397276221	2519	58.37809	26.72629	7/18/2019	00:02:03+00	1
1563397276221	2519	58.37809	26.72629	7/18/2019	00:02:08+00	1
1563397285255	2477	58.37424	26.71481	7/18/2019	00:01:58+00	1
1563397285255	2477	58.37424	26.71481	7/18/2019	00:02:02+00	2
1563397285255	2477	58.37435	26.71511	7/18/2019	00:02:12+00	2
1563397285255	2477	58.37436	26.71512	7/18/2019	00:02:17+00	2

The secondary dataset used in this research was district boundaries provided by Tartu City Government (last updated on 06-12-2020). It contained the name of each district (“nimi” in Estonian) and district code (“kood” in Estonian). These district codes were used for joining and analysis purpose. Details on district names, district codes and the number of docking stations for each district is listed in Annexe 1.

2.3 Methods

To achieve the aim of this study, the methodology was divided into three parts. Figure 3 shows the workflow of the study. First, the **GPS file** was spatially joined with the **Districts Boundary** file to add district names and district codes to trips trajectories data.

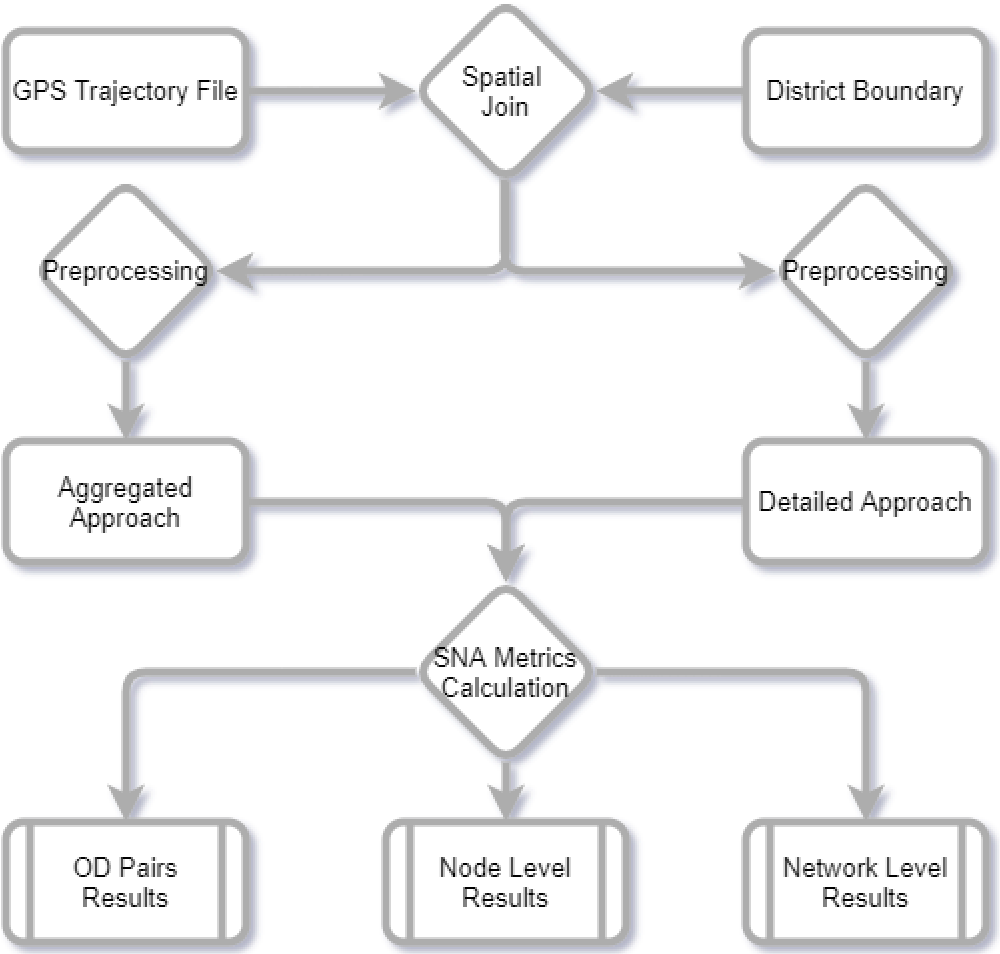


Figure 3: Workflow of Analysis

Second, the result was then processed using two separate Python scripts made for data cleaning and for calculating mobility flows that is used as an input for the third step. Two different approaches were taken to calculate the mobility flows – namely aggregated and detailed approach – for comparative purpose. In the **aggregated** approach, only the origin and destination districts of every trip were considered in calculating the mobility flows. Some data discrepancies were removed. For example, there were few points where the **coordinate time** field was null and hence removed. Also, some trips had null geometry, i.e. no trajectory but rather only a single point, so those points were also excluded. Finally, the points in which the origin and destination districts were the same were removed. 522,795 trips were left after data cleaning from the original dataset of 692,347 trips.

In the **detailed approach** all the districts which constituted the trip were included for constructing the mobility flows. The trips were splitted over all the districts so that the destination district of the first splitted part of the trip was the origin of the second and so forth. In this approach also some similar data discrepancies were removed like null field records for **coordinate time** and null geometry records. 568,920 trips were left after data cleaning from the original dataset of 692,347 trips. This dataset was splitted into 9,447,386 trip segments. For this approach those part of trips were excluded if and only if their origin and destination districts were the same.

After data cleaning origin-destination codes (OD codes) were assigned to each trip. It consists of the first letter **O** denoting "origin" followed by the origin district code, and **D** refers to the "destination" followed by destination district code. For example, **O12D15**, here 12 is the origin district, and 15 is the destination district. This convention was followed throughout the analysis, where assigning the OD codes to the trip records was necessary. Table 3 shows an example of a single trip from file prepared for aggregated approach, the trip started from district **15** and ended in district **08**.

Table 3: Sample record for aggregated approach file

Route code	**User id**	**Origin id**	**Origin**	**Destination id**	**Destination**	**od code**
1563397276221	1	15	Uueturu	08	Kruusamäe	O15D08

In the detailed approach the same single trip was divided into four parts (Table 4) as the trip started from district **15,** and then the user passed through districts **30, 09, 07** respectively, before ending the trip in district **08**.

Table 4: Sample record for detailed approach file

Route code	**User id**	**District list of trip**	**od code**
1563397276221	1	[15, 30, 09, 07, 08]	O15D30
1563397276221	1	[15, 30, 09, 07, 08]	O30D09
1563397276221	1	[15, 30, 09, 07, 08]	O09D07
1563397276221	1	[15, 30, 09, 07, 08]	O07D08

Mobility flows are represented by origin-destination pairs. OD matrix is an n × m matrix where n is the number of different “origin” districts, m is the number of “destination” districts. There are 31 districts in Tartu, so the OD pairs were calculated using the permutation formula below:

$$P(n,r) = \frac{n!}{(n-r)}$$

The number of all possible relations between districts were 31*31=961, but as those OD pairs where origin and destination districts were the same were eliminated resulting in 930 possible OD pairs left. Each OD pair was depicted with the **OD code** using the convention mentioned earlier. For each OD pair the number of trips and number of users were calculated for For both aggregated and detailed approaches.

Third, SNA metrics were calculated. The network in this study was a weighted directed network with number of trips and number of users as weights. Different measures were selected to characterise the connectivity of the network on local (district) and global (city) levels. Centrality is a characteristic that has been extensively discussed for complex networks. There are various centrality measures. From a local perspective, degree centrality, weighted degree or strength, and betweenness centrality measure different node’s features. From a global perspective, average degree, density and modularity, and others depict underlying characteristics of the whole network (Lin., 2017). Based on existing studies, the local and global metrics shown in Table 5 were selected and calculated for this study.

Table 5: Selected SNA metrics

Local or Node Level Metrics	**Global or Network Level Metrics**
Degree	Average Degree
Weighted Degree/Strength	Density
Betweenness	Community Detection

Degree centrality

Degree centrality is the count of the total number of connections to any node. In the case of directed networks, the degree is the sum of in-degree and out-degree. Despite its simplicity, the degree is an effective measure to identify the most important and influential nodes in the network. The degree is used to describe the exposure to the network. Denoted as C_D, the formula for the degree is:

$$C_D(i) = k_i^{in} + k_i^{out}$$

in-degree is a count of the number of ties directed to the node (head endpoints),

$$k_i^{in} = \sum_{j=1}^{n} a_{ij}$$

and out-degree is the number of ties that the node directs to others (tail endpoints)

$$k_i^{out} = \sum_{j=1}^{n} a_{ji}$$

where a_{ij} is the entry of the ith row and jth column of the adjacency matrix and vice versa (Tabassum et al., 2018).

Weighted Degree / Strength

The weighted degree or strength of a node considers both the connectivity and the weights of the links (Antoniou & Tsompa., 2008). It is the same as the degree but with the additional weight assigned, which is the inflow and outflow of the people's movements and trip counts. It quantifies the hub nodes (nodes with most connections) in the network.

$$S_i = S_i^{in} + S_i^{out}$$

in-strength is a count of the number of ties directed to the node (head endpoints), including the weights from ith to jth node,

$$S_i^{in} = \sum_{j=1}^{n} a_{ij}\, w_{ij}$$

and out-strength is the number of ties that the node directs to others (tail endpoints), including the weights from jth to ith node.

Betweenness Centrality

Betweenness centrality is a way of detecting the amount of influence a node has over the flow of information in a network. Betweenness centrality quantifies the number of times a node acts as a **bridge** along the shortest path between the other connected nodes. Nodes with high betweenness occupy critical roles in the network structure since they usually have a network position that allows them to work as an interface between tightly-knit groups, being "vital" elements in the connection between different regions of the network (Tabassum et al., 2018). Nodes with high values of betweenness centrality participate in many shortest paths. For directed network betweenness centrality often denoted by $C_{.B.}$ can be calculated by the following formula:

$$C_B = \Sigma \frac{xx}{(n-1)(n-2)}$$

where xx is:

$$xx = \frac{g_{kt\,(n)}}{g_{kt}}$$

here g_{kt} (n) is the total number of shortest paths between nodes k and t that pass-through node n and gkt is the total number of shortest paths between nodes k and t.

$$S_i^{out} = \sum_{j=1}^{n} a_{ji}\, w_{ji}$$

Average Degree

The average degree is the number of links per nodes. The average degree is simply the mean of the degrees of all nodes in a network. The average degree can be used to measure the connectivity of the whole network (Tabassum et al., 2018). It can be calculated as

$$k_{avg} = \frac{1}{n}\sum_{i=1}^{n} k_i$$

where k_i is the degree of i^{th} node

Density

Density is an important network-level measure, which can explain the general level of connectedness in a network (Tabassum et al., 2018). Density is defined as the actual number of ties in a network, expressed as a proportion of the maximum possible number of ties. It is a number that varies between 0 and 1.0. When density is close to 1.0, the network is said to be dense otherwise, it is sparse (Ghali et al., 2012). It is the ratio of actual connections by potential connections.

$$\rho = \frac{Actual\ Connections}{Potential\ Connections}$$

Actual connections are also called the number of links between the network nodes, and potential connections in the directed network case can be calculated as n(n-1), where n is the number of network nodes.

Community Detection

Community structure is one of the key properties of complex networks and detecting communities is a problem of considerable interest. Community structure in the context of networks refers to the occurrence of groups of nodes in a network that is more densely connected internally than with the rest of the network (Ghali et al., 2012). Community is a subset of nodes in a network with more connections than the rest of the network. Community detection works on the modularity optimisation algorithm to detect communities in the network based on their modularity. Modularity is a measure of the network's structure, measuring the density of connections within a module or community (Newman & Girvan., 2003). Modularity ranges from -1 to 1. It is positive if the number of links inside the group or density is more than the expected number, i.e. expected density of the assumed random network. Variation from zero indicates difference with the random case.

For data preparation, cleaning and spatial joining purposes Python’s pandas and geopandas libraries were used. For SNA calculation python’s networkx library and Gephi software was used. For the visualisation of the results different maps were done using ArcGIS Pro software. Bivariate mapping technique was used and flow maps were constructed for showing the mobility flows between different districts of Tartu. Choropleth maps were used for showing the degree distribution and betweenness centrality. Bar charts on maps were displayed to show the weighted degree or strength of the nodes/districts.

3. Results

3.1 Inter-district connectivity

This section discusses the links between the districts based on trip counts and user movement for both approaches. It is noteworthy that the results explained in this section are shown as the percentage of the total for a reason. It necessary for comparison between aggregated approach and the detailed approach as the number of records was impacted hugely because of splitting each trip to all districts in the detailed approach. It should also be noted that for visualisation purposes in Figure 4, 5, 6 and 7, those OD pairs are not shown where the trip count or users count were less than 500 to avoid cluttering.

For aggregated approach, the result in Figure 4 shows that Tähtvere, Maarjamõisa, Ülejõe, Vanalinn, Riiamäe, Ees-Annelinn, Kesk-Annelinn and Uueturu were the districts with most inflow and outflow of trips. Taga-Annelinn, Uus-Ihaste, Variku, Supilinn were the districts with the least or zero connectivity.

As far as the OD pairs were concerned, the highest-ranking OD pair in terms of trip count is Ees-Annelinn and Kesk-Annelinn, which covers 1.35 % of total trips. This is followed by Kesk-Annelinn and Sadama with 1.34% of total trips. Sadama and Kesk-Annelinn has 1.31 % of total trips. 1.26% of total trips were in Kesk-Annelinn and Ees-Annelinn. Uueturu and Kesk-Annelinn hold 1.12% of total trips. As for the least connected OD pairs, Taga-Annelinn, Uus-Ihaste, Variku, Supilinn does not have any docking station, so these districts were making some 118 OD pairs out of 930 total pairs which have zero trip counts.

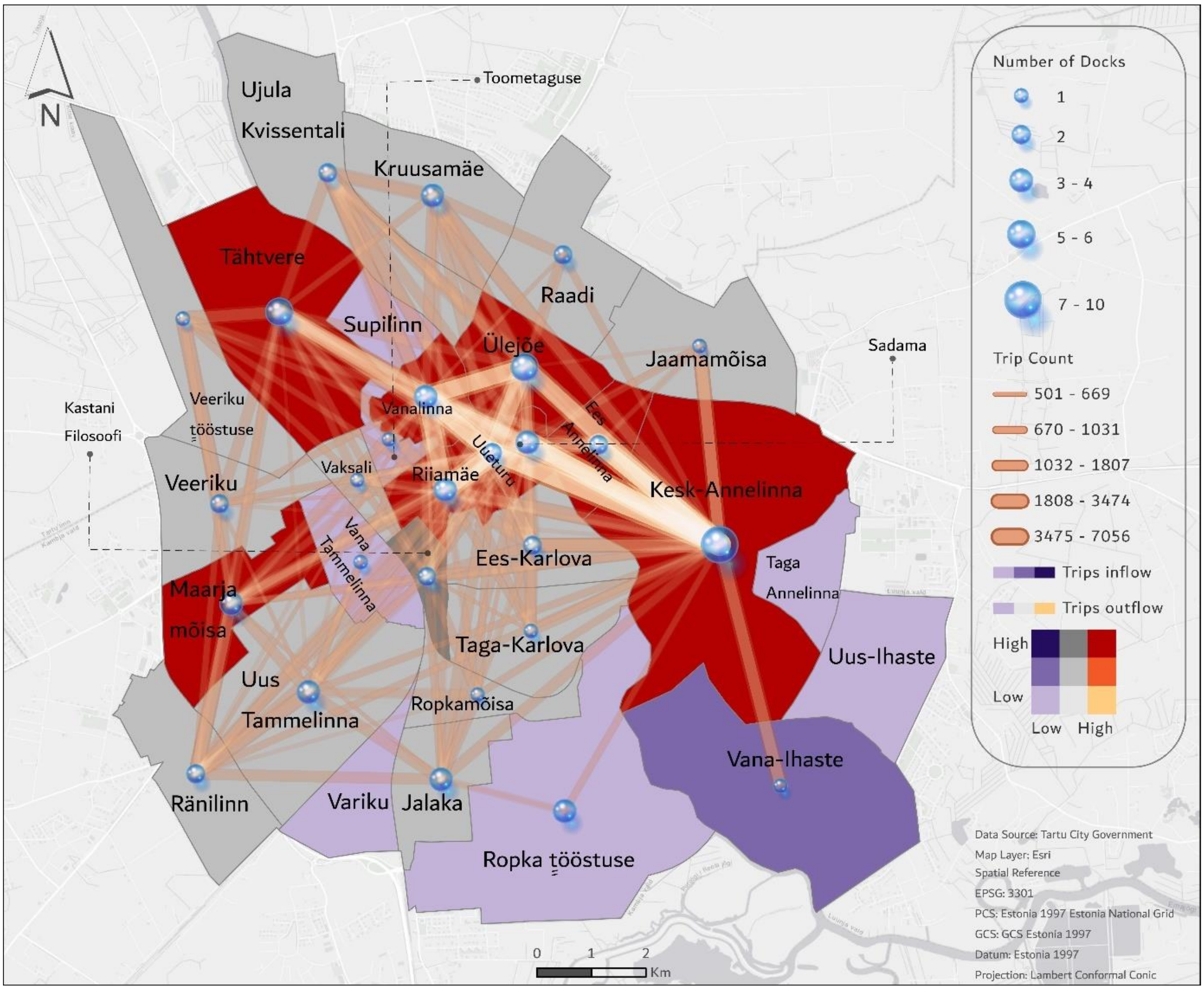


Figure 4: Map of trip count for aggregated approach.

For aggregated approach based on users count, the results slightly differ from trip count results. High number of bike-share users occurred in Tähtvere, Ülejõe, Vannalinn, Riiamäe, Sadama, Kesk-Annelinn, and Uueturu districts as shown in Figure 5. Even though Maarjamõisa district had a high number of trip inflow and outflow (Figure 4), this did not occur for the number of users (Figure 5). This means that high number of trips were made by comparatively small number of people. The highest OD pair in terms of number of users was Sadama and Kesk-Annelinn, with the highest of 2992 people in this OD pair.

The top 5 OD pairs in terms of number of users were Sadama and Kesk-Annelinn with 1.06% of all users, followed by Ees-Annelinn and Kesk-Annelinn with 1.05%, Vanalinn and Uueturu with 1.04%, Kesk-Annelinn and Sadama with 1.02%, and Uueturu and Kesk-Annelinn with 0.96%.

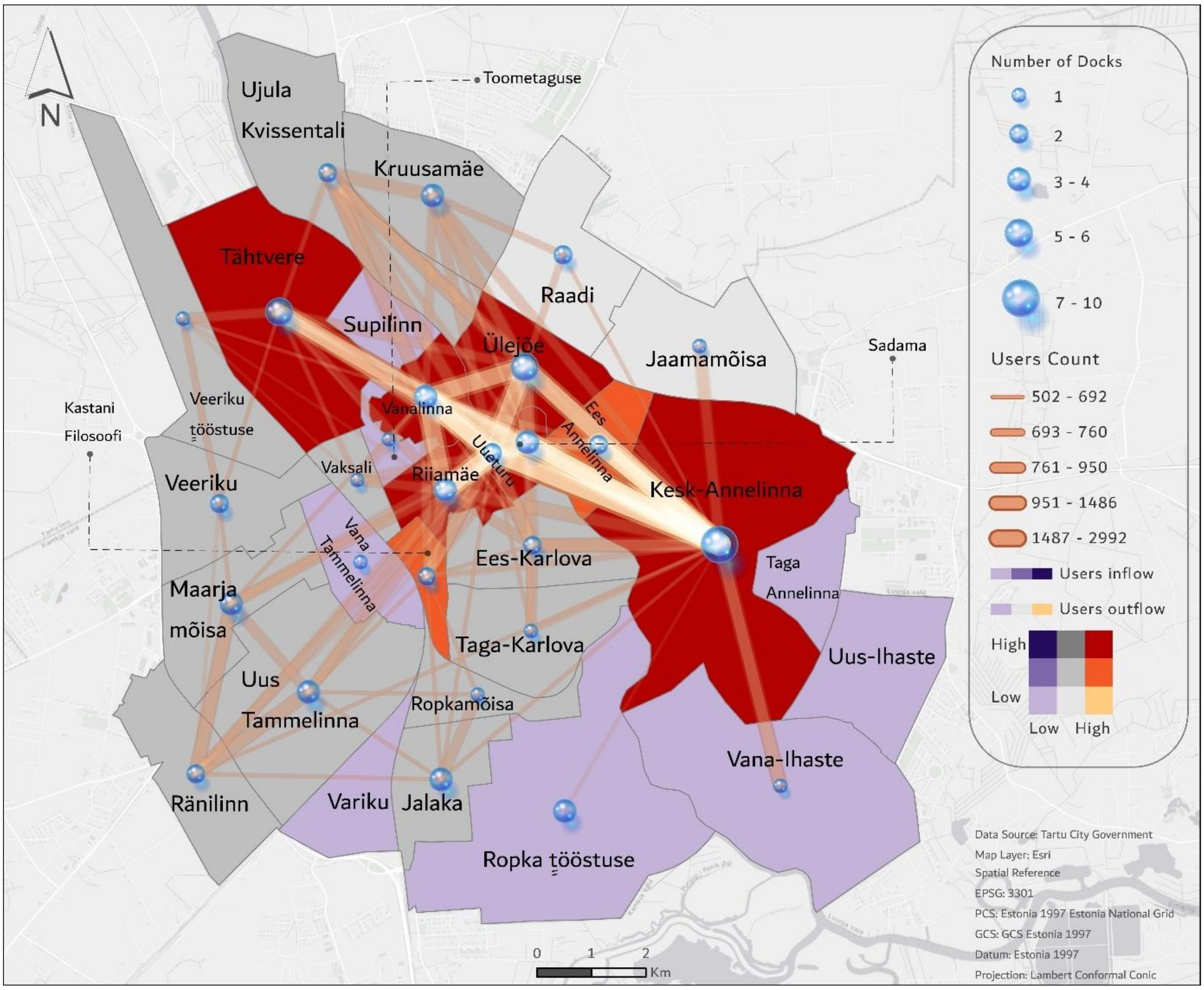


Figure 5: Map of users count for aggregated approach.

In detailed approach, it is interesting that despite having the largest number of docking stations in Kesk-Annelinn, this district did not have the highest number of trips. Uus-Tammelinn, Vaksali, Vanalinn, Uueturu, Riiamäe, Ees-Annelinn, Ees-Karlova and Ülejõe were among top districts with most incoming and outgoing trip flow.

The results of connectivity patterns using trip counts among the districts of Tartu from detailed approach are shown in Figure 6. Detailed approach showed the connectivity among those districts also where there were no docking stations. The top 5 OD pairs were Riiamäe and Uueturu making 3.29% of total trips. Uueturu and Sadama made 3.23% of total trips followed by Uueturu and Vanalinn with 2.92%. Uueturu and Riiamäe has 2.89% and 2.76% trips out of total number of trips were made between Vanalinn and Uueturu districts. Interestingly, Uueturu has only 2 docking stations but was dominant district in the top 5 OD pairs for trip count in detailed approach. Out of 930 OD pairs, 155 pairs were having zero trip counts. Interestingly, these OD pairs were mostly not among those districts which do not have any docking stations but in those districts where docking stations are installed. For example, Ränilinn and Kesk-Annelinn, Kesk-Annelinn and Maarjamõisa, Ülejõe and Ränilinn, Veeriku tööstuse and Kesk-Annelinn.

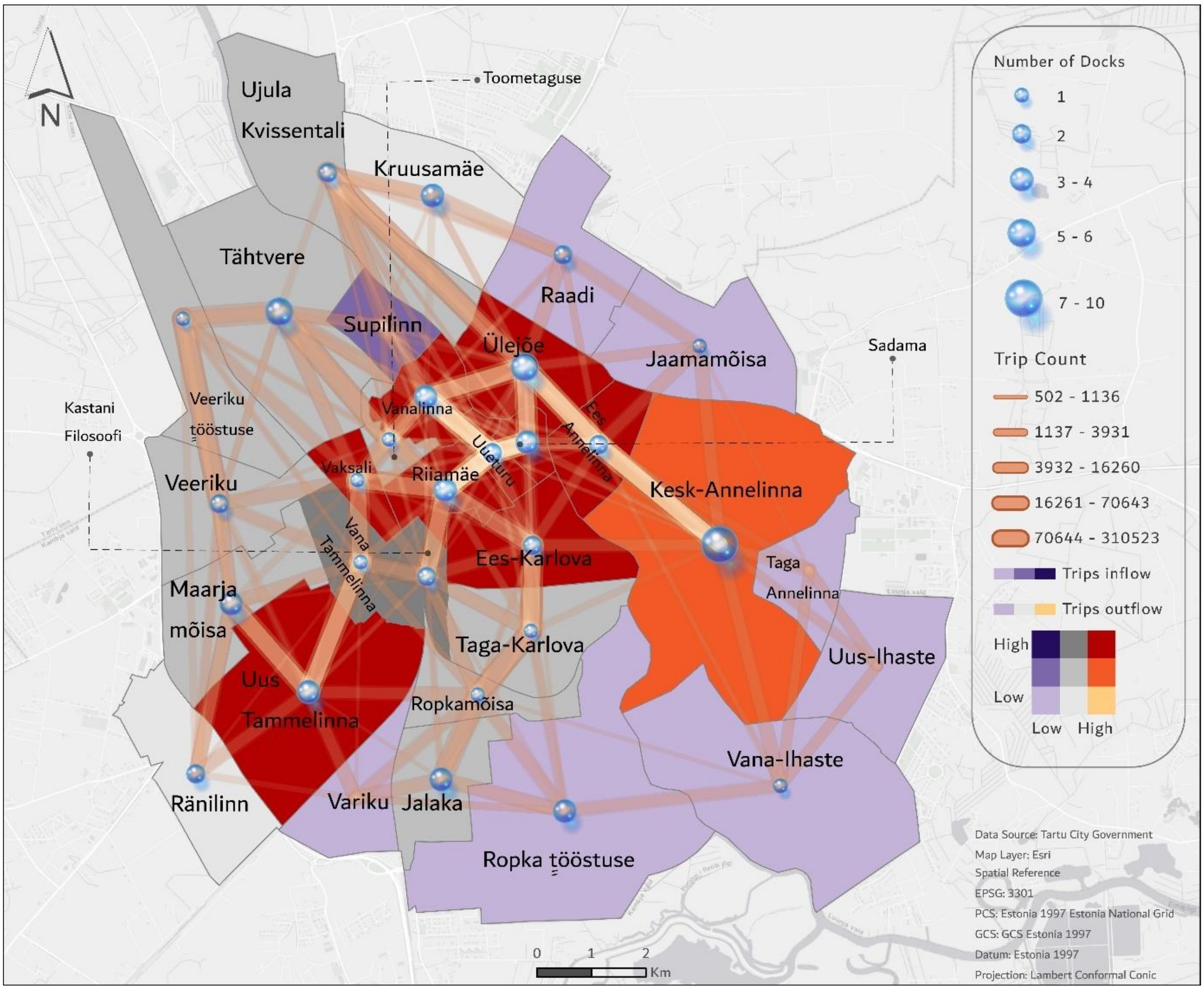


Figure 6: Map of trip count for detailed approach.

Number of users between different districts using the detailed approach is shown in Figure 7. In top OD pairs here Uueturu was again the dominant district. Uueturu and Vanalinn was top OD pair with 2.21% of total users moving between these two districts. Uueturu and Sadama made up 2.13%, followed by Uueturu and Sadama with 2.07%, and Vanalinn and Uueturu with 2.07% of total users. Uueturu and Riiamäe had 1.96% of all users moving in this OD pair. The southwest districts like Uus-Tamelinn, Maarjamõisa and Vana-Tammelinn also showed a high number of users, which was otherwise not evident in aggregated approach. Maarjamõisa and Uus-Tammelinn consists of 1.18% of all users and Uus-Tammelinn and Vana-Tammelinn has the percentage of 1.28.

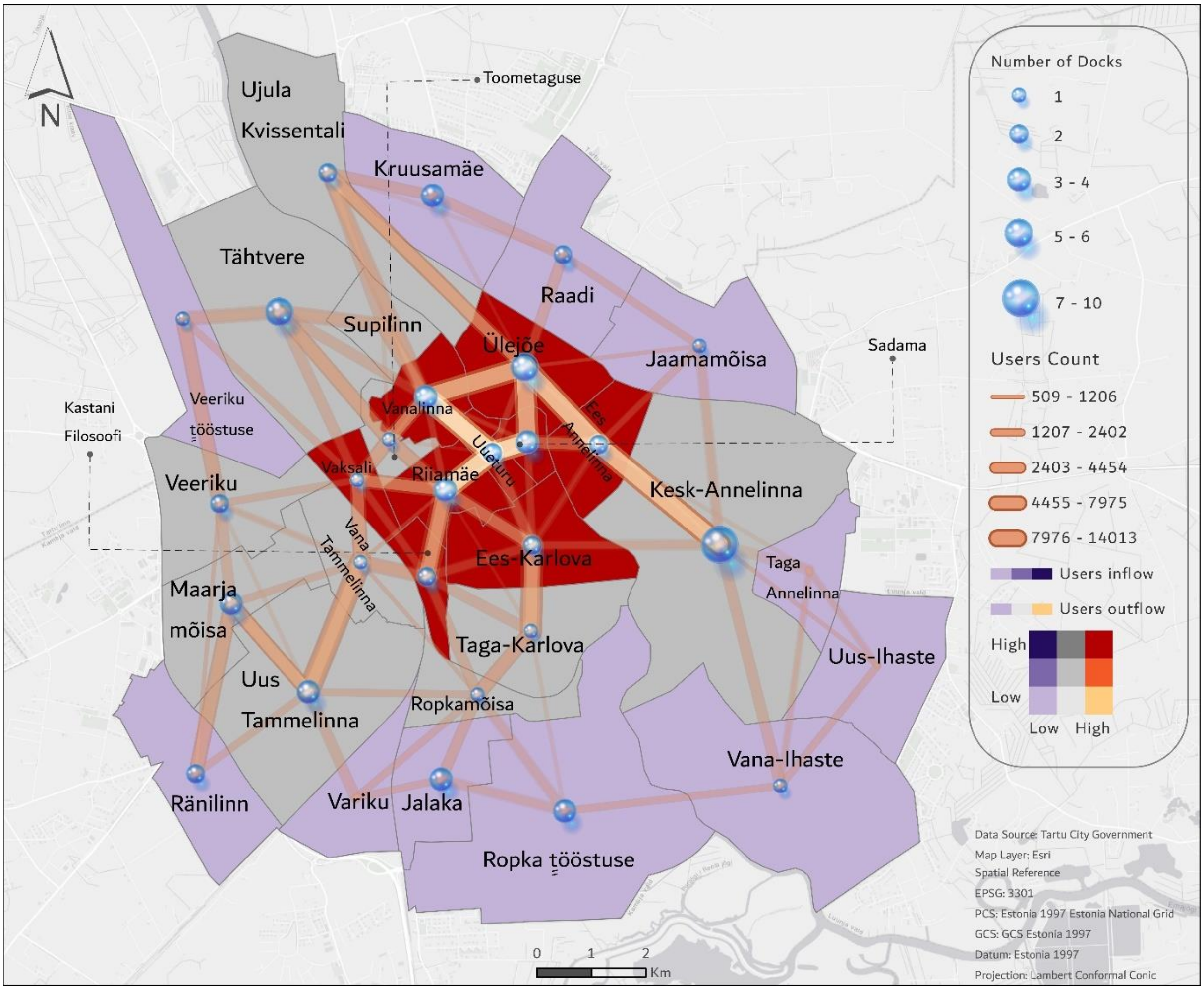


Figure 7: Map of users count for detailed approach.

The comparison between both approaches showed that high inter-district connectivity occurred in seven districts among the top ten OD pairs (Table 6). These districts are Ees-Annelinn, Kesk-Annelin, Riiamäe, Sadama, Ülejõe, Uueturu, and Vanalinn. These districts had the most connectivity with each other in terms of trips counts.

Table 6: Comparison of top ten connected OD pairs for trip count. Common OD pairs are highlighted in grey. Their ranking, however, varied.

Sr #	Aggregated approach		Detailed approach	
1	Ees-Annelinna	Kesk-Annelinna	Riiamäe	Uueturu
2	Kesk-Annelinna	Sadama	Uueturu	Sadama
3	Sadama	Kesk-Annelinna	Uueturu	Vanalinna
4	Kesk-Annelinna	Ees-Annelinna	Uueturu	Riiamäe
5	Uueturu	Kesk-Annelinna	Vanalinna	Uueturu

6	Vanalinna	Uueturu	Sadama	Uueturu
7	Kesk-Annelinna	Uueturu	Kesk-Annelinna	Ees-Annelinna
8	Uueturu	Vanalinna	Ees-Annelinna	Kesk-Annelinna
9	Uueturu	Riiamäe	Ülejõe	Ees-Annelinna
10	Ülejõe	Vanalinna	Ees-Annelinna	Ülejõe

When comparing the least connected OD pairs, in aggregated approach 118 ODs had zero number of trip counts, and the same figure for the detailed approach was 155 ODs. There were 58 common OD pairs with no connectivity in both approaches (Table 7). As such, the results from the detailed approach showed more promising results for urban connectivity analysis as its results were not that much influenced by those districts that had more docking stations or those districts where there were no docking stations.

Table 7: Common OD pairs in both approaches with no connectivity in terms of number of trips.

Sr #	Origin	Destination	Sr #	Origin	Destination
1	Supilinn	Taga-Annelinna	30	Taga-Annelinna	Supilinn
2	Tähtvere	Taga-Annelinna	31	Vana-Ihaste	Supilinn
3	Veeriku tööstuse	Taga-Annelinna	32	Uus-Ihaste	Supilinn
4	Variku	Taga-Annelinna	33	Variku	Supilinn
5	Ränilinn	Taga-Annelinna	34	Jalaka	Supilinn
6	Maarjamõisa	Taga-Annelinna	35	Taga-Annelinna	Tähtvere
7	Ropkamõisa	Taga-Annelinna	36	Taga-Annelinna	Veeriku
8	Vana-Tammelinna	Taga-Annelinna	37	Uus-Ihaste	Veeriku
9	Vaksali	Taga-Annelinna	38	Taga-Annelinna	Veeriku tööstuse
10	Kastani-Filosoofi	Taga-Annelinna	39	Uus-Ihaste	Veeriku tööstuse
11	Toometaguse	Taga-Annelinna	40	Variku	Veeriku tööstuse
12	Ees-Annelinna	Uus-Ihaste	41	Uus-Ihaste	Uueturu
13	Kruusamäe	Uus-Ihaste	42	Jaamamõisa	Variku
14	Tähtvere	Uus-Ihaste	43	Raadi	Variku
15	Veeriku	Uus-Ihaste	44	Supilinn	Variku

16	Veeriku tööstuse	Uus-Ihaste	45	Uueturu	Variku
17	Taga-Karlova	Uus-Ihaste	46	Toometaguse	Variku
18	Variku	Uus-Ihaste	47	Taga-Annelinna	Ränilinn
19	Ränilinn	Uus-Ihaste	48	Taga-Annelinna	Maarjamõisa
20	Maarjamõisa	Uus-Ihaste	49	Uus-Ihaste	Maarjamõisa
21	Ropkamõisa	Uus-Ihaste	50	Taga-Annelinna	Uus-Tammelinna
22	Uus-Tammelinna	Uus-Ihaste	51	Uus-Ihaste	Uus-Tammelinna
23	Vana-Tammelinna	Uus-Ihaste	52	Taga-Annelinna	Vana-Tammelinna
24	Vaksali	Uus-Ihaste	53	Uus-Ihaste	Vaksali
25	Toometaguse	Uus-Ihaste	54	Taga-Annelinna	Kastani-Filosoofi
26	Variku	Jaamamõisa	55	Uus-Ihaste	Vanalinna
27	Variku	Raadi	56	Variku	Vanalinna
28	Taga-Annelinna	Kruusamäe	57	Taga-Annelinna	Toometaguse
29	Variku	Kruusamäe	58	Variku	Toometaguse

3.2 Degree of connectivity of the districts

The results of the node-level degree calculation are discussed in this subchapter and shown in Figure 8. Degree measure showed that in detailed approach all nodes (districts) of the network (Tartu) have a connection with each other making it a complete network. A term “complete network” is used for such networks in which all nodes have connections with each other. In detailed approach, all nodes were related to each other, and since it was a directed network, the highest possible degree was 62, which means 31 for in-degree and 31 for out-degree.

In aggregated approach, **Kesk-Annelinn** and **Sadama** districts had the highest degree of 59 connections. It means Kesk-Annelinn and Sadama were the hub components of this network. A hub is component of a network with the highest degree node/s. Though this simple measure gives some glimpse of the network structure, it does have some drawbacks as well as it shows only the node connections but does not consider how many connections each node has, and that is further characterised by strength or weighted degree measure.

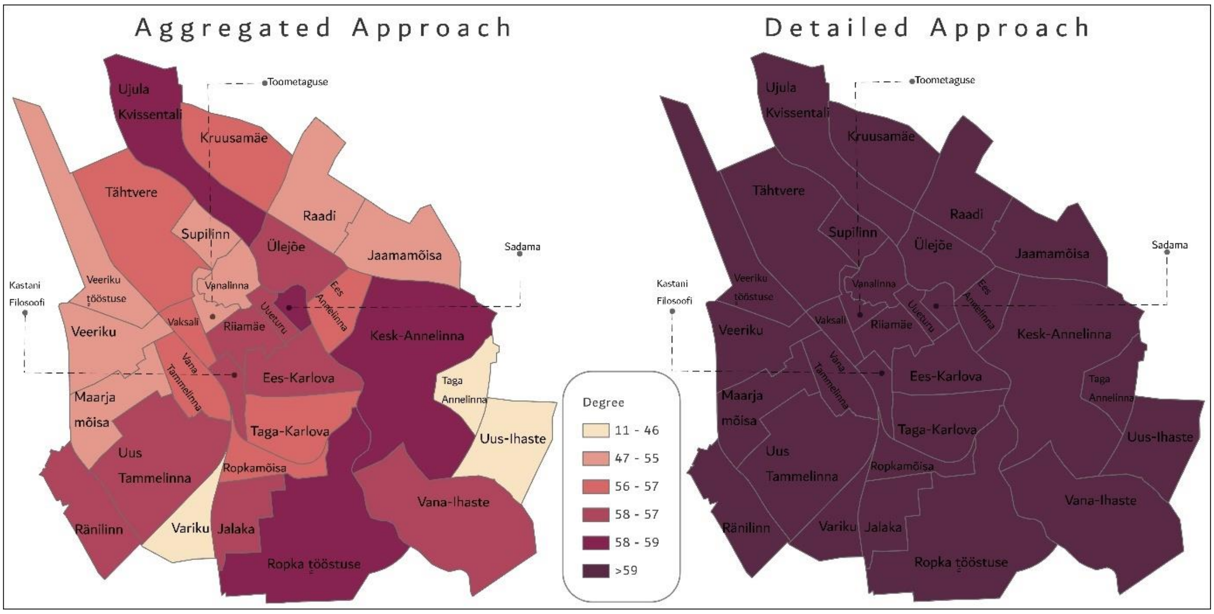


Figure 8: Degree measure of nodes for aggregated and detailed approach.

A bridge node is an important centrality measure to highlight crucial nodes in the network. For calculating bridge node, detailed approach results could not be used since for the detailed approach the network is complete – betweenness centrality is zero for all the nodes (Figure 9). Calculations based on the aggregated approach, however, showed that **Sadama** district was the bridge node with the highest value of 16.53, followed by the second-highest value of 16.45 for the Kesk-Annelinn district. It means that Sadama is a bridging/junction node to keep the whole network connected. It unifies the whole network and is the most influential node. The detailed table showing degree, betweenness and strength values for all nodes for aggregated and detailed approach are listed in Annexe 2 and Annexe 3 respectively.

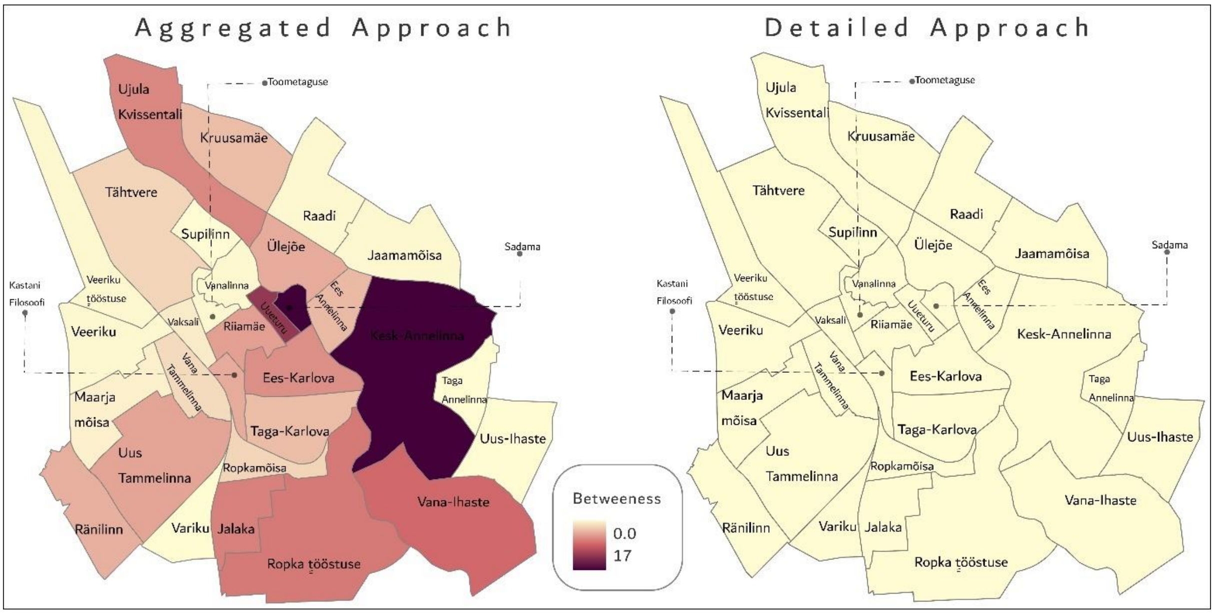


Figure 9: Betweenness measure of nodes for aggregated and detailed approach.

The strength of the network is calculated by assigning weights to the nodes. The weights are the number of links each node has, i.e., the number of trips. As for both approaches, the trip counts vary a lot that is why strength for each node was showed and compared in terms of percentage of its total in figure 10. For aggregated approach, out of total trips, 10 percent were accounted for Kesk-Annelinn alone, followed by Uueturu with 9%, Vanalinn with 8% and Sadama with 6% of whole trips. Comparatively, for the detailed approach, Kesk-Annelinn accounts only 4% of total trips. The results show that the mobility flows were not only confined to the city centre and to the area with more docking stations in case of detailed approach. This metric also supports the detailed approach as it was not biased by more docking stations in the district and did not leave out the districts with no docking stations like Taga-Annelinn and others.

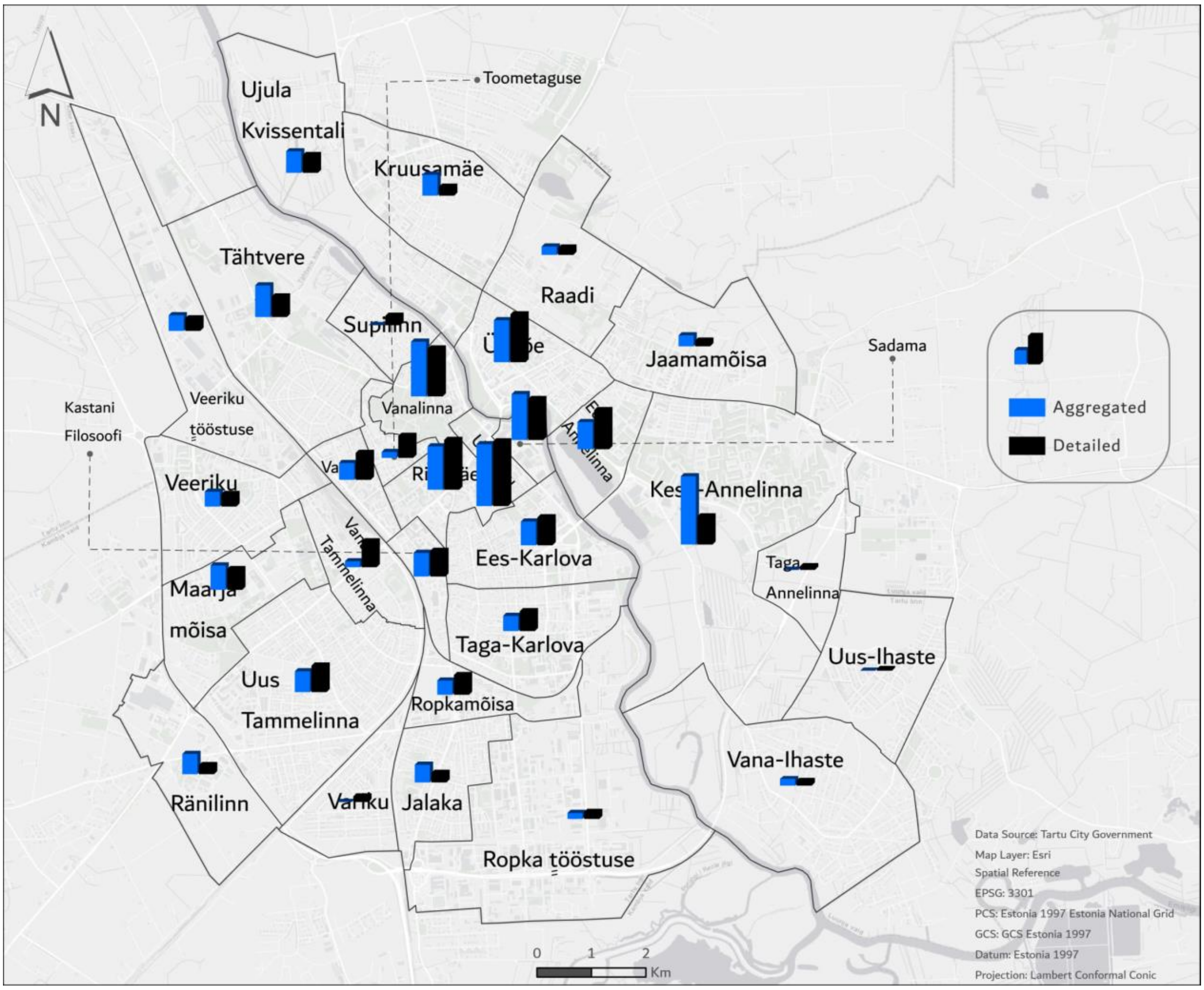


Figure 10: Strength measure of nodes for aggregated and detailed approach.

To conclude, this section found that Kesk-Annelinn and Uueturu are the hub nodes, and Sadama district is the bridge node of the network.

3.3 City level connectivity

When in previous subchapters connectivity was explored on the level of districts, this section explores the connectivity on a global (city) level that reflects the network structure. Table 8 shows all the properties of calculated global SNA metrics.

In aggregated approach, the **average degree** of the network was ~26, which meant that on average, ~26 districts are connected with each other. However, some districts did not have any docking stations, so this method was heavily affected by that. On the other hand, the detailed approach shows an average degree of 31, which means that all nodes are connected and it is a complete network.

The network in aggregated approach had a density of 0.8, which means it is more sparse than the network calculated by detailed approach, where the density is 1. Network with density of 1 is a fully dense network.

Table 8: Network-level statistics for both approaches.

Metric	**Aggregated approach**	**Detailed approach**
Average Degree	26.1	31
Density	0.8	1.0
Modularity	0.1	0.2

Modularity helped to determine whether there was any division of the nodes in the network, which could form communities. Modularity looks for the group of nodes that are more densely connected to each other to detect the hidden patterns of the network as the modularity value lies from -1 to 1. Being positive, i.e. 0.1 and 0.2 for both approaches, means that the algorithm showed the number of links inside communities was more than the expected number. Modularity, thus, measures the density of connections within a module or community. Community detection helped to explore the connections among the communities or clusters of districts that tend to have significantly more connections (in terms of trips) than the rest of the districts. The results from the modularity algorithm showed three communities for each method. However, the number of districts in each community varies for both methods. Community detection based on both aggregated and detailed approaches found that the smallest community is the second one, with 7 and 9 districts, respectively (Table 9), and the largest is the third community with 14 and 11 districts, respectively. The results from this section could be more explored in-depth in terms of underlying reasons for clustering or segregation of those communities.

Table 9: Division of communities.

	Aggregated approach	Detailed approach
Community 1	10 districts	11 districts
Community 2	7 districts	9 districts
Community 3	14 districts	11 districts

Some nodes were overlapping in both approaches, but the communities formed based on a detailed approach (Figure 11) were more logical, as results from previous sections also suggested. For example, in the case of the detailed approach, Kesk-Annelinn, Ees-Annelinn, Sadama, Uueturu were the participant districts of community 1 (Figure 11) among other districts with more connections than the others.

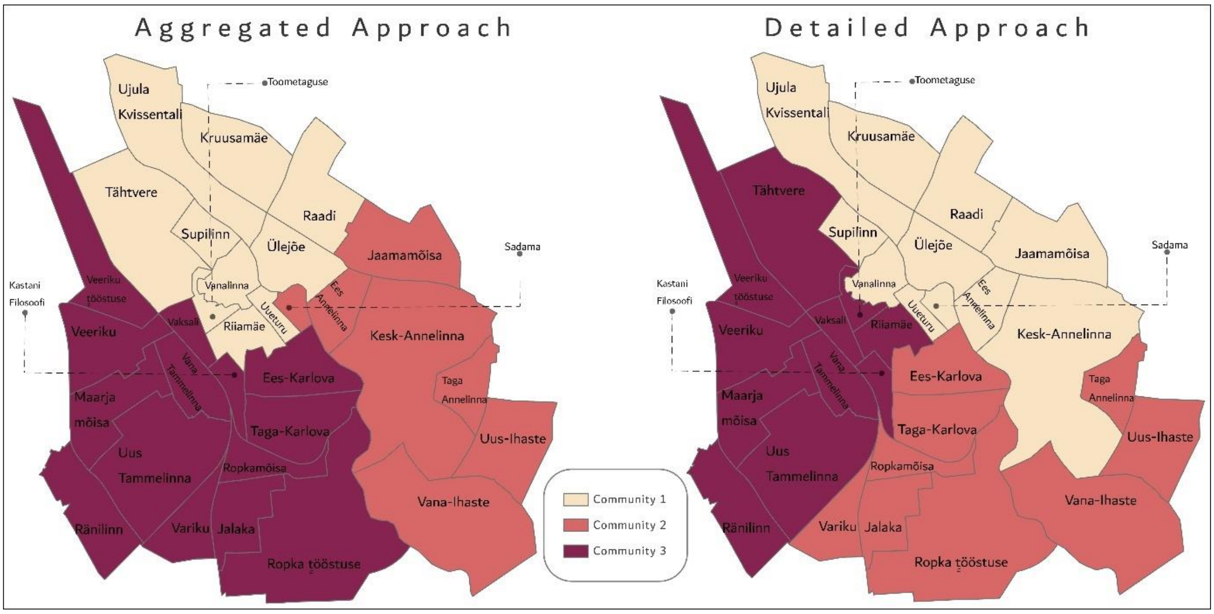


Figure 11: Comparison of communities for aggregated and detailed approach.

4. Discussion

Connectivity is a topic that has been mostly measured from infrastructure perspective – how well streets and roads are connected (Dill, 2004; Mohamad & Said, 2014; Peponis et al., 2008; Watts et al., 2015). Less attention has been paid to the connections between urban areas based on human mobility, which brings in an important dimension that reflects how people are using urban space. This study used bike-share data to investigate urban connectivity in Tartu.

The link level results based on aggregated approach showed that Tähtvere, Maarjamõisa, Ülejõe, Vanalinn, Riiamäe, Ees-Annelinn, Kesk-Annelinn and Uueturu were the districts with most inflow and outflow of trips. Taga-Annelinn, Uus-Ihaste, Variku, Supilinn were the districts with the least or zero connectivity. In terms of number of users, more bike-share commuters mobility was among Tähtvere, Ülejõe, Vanalinn, Riiamäe, Sadama, Kesk-Annelinn, and Uueturu districts. Maarjamõisa district, despite having high number of trips was not high ranking district in terms of users meaning that high number of trips were made by comparatively small number of people.

According to detailed approach, Uus-Tammelinn, Vaksali, Vanalinn, Uueturu, Riiamäe, Ees-Annelinn, Ees-Karlova and Ülejõe were among the top districts with the most incoming and outgoing trip flow. Kesk-Annelinn, despite having the highest number of docking stations, was not among those districts which the highest number of trips. Out of 930 OD pairs, 155 pairs had zero trip counts. Interestingly, those OD pairs were mainly not among those districts that do not have any docking stations but those with docking stations. For example, Ränilinn and Kesk-Annelinn, Kesk-Annelinn and Maarjamõisa, Ülejõe and Ränilinn, Veeriku tööstuse and Kesk-Annelinna. For the number of people, Uueturu and Vanalinn, Sadama, Riiamäe were among high-ranking districts in terms of users movements. Uueturu was the dominant district in most high ranking OD pairs. The southwest districts like Uus-Tammelinn, Maarjamõisa and Vana-Tammelinn also showed the high movement of users, which was otherwise not shown in aggregated approach.

The comparison between both methods showed the highest inter-district connectivity among the top ten OD pairs was in the following seven districts: Ees-Annelinn, Kesk-Annelinn, Riiamäe, Sadama, Ülejõe, Uueturu, and Vanalinn. Built environment and urban form are important determinant factors for mobility (Koohsari et al., 2014) which could be seen in case of Tartu as well. The central districts were more connected with each other reflecting Tartu city's monocentricity.

The node level results showed higher degree not only for central districts or districts with high number of docking stations but in outer districts also showing more diverse mobility flow coming from outer districts. For example, Ujula-Kvissentali district only had 2 docking stations but the degree for this district was 58 which was among the highest ones. The results also showed that Kesk-Annelinn and Sadama were found as hub nodes of the city network. It means that these nodes had the most connections with the other nodes (i.e. districts). Such nodes are generally of higher importance for the connectivity of a network (Sun & Guan, 2016). Sadama district was found as a bridge node, which means this node acted as a bridge for connecting any two other nodes for the highest number of times compared to other nodes. Disruption in this node could disrupt network connectivity (Sun & Guan, 2016). This information could help in future infrastructure planning.

Network level measures showed that overall the network is densely connected. For the aggregated approach, the density was 0.8. For the detailed approach, it was 1, making it a fully dense network. Based on the aggregated approach, the network had an average degree of ~26, but in the case of detailed approach the average degree was 31, i.e. equal to the nodes, which indicates a complete network. Community detection helped to examine which parts of Tartu city are so connected that they can be considered separate communities. It refers to the grouping of nodes in the network which are more densely connected internally than with the rest of the network and helps to show internal relations of the network (Yao et al., 2019). Both approaches resulted in three communities, but the composition of districts was different. It shows that these three communities have dense connections within themselves than the rest of the network. Districts with strong and weak strength showed a clear geographical distribution in communities.

It was important to analyse connectivity on link level (mobility flow), node level (district) and network level (city) with both approaches to get more rich insight into the urban network. On the most general level the second approach showed more precise results as there was connectivity among all the districts, even there were no docking stations were installed. However, it was not possible to find hubs and bridging districts with detailed approach, because the results indicated that Tartu is a complete network. Similary finding bridge node in the network was not possible using the detailed and aggregated approach have shown that Vannalinn and Uueturu were commonly the most critical nodes of the network.

Additionally, the second approach shows that Ülejõe and Riiamäe have a higher percentage of inflow and outflow. All the old town area and city centre shows the most of the people movements and trips. For the link-level, there were some similarities of the network in both approaches. However, for such analysis, the detailed approach was more favourable than the aggregated approach. The cluster of the trip counts for both approaches showed similar patterns, but as far as people’s movements are concerned, the detailed approach showed more precise results that people are using the docking stations in the city centre to move away to the adjacent neighbourhoods from the city core. Summarising it all, only considering the origin and destination neighbourhoods, i.e. the aggregated approach, would not have helped get the broader and more concise view of the connectivity of neighbourhoods as it is compared with the detailed approach.

When interpreting the results of the analysis, the limitations of the study must be considered. One of the limitations of this analysis was that not all the population of Tartu uses the bike-share system as the primary mode of communication. According to the provided dataset, there were only 29956 registered users of TSB. According to the Statistics Estonia website (as of 10-05-2021), the population of Tartu was 96,123. The registered users of TSB are not even half of the population, so it heavily affected the mobility study, which influences the results of urban connectivity. Another limitation of using bike-share data for measuring urban connectivity was that seasonal variation in the usage of the bike-share system affected the results. There was also a decline in bike-share use in winter months as compared to summer months.

Conclusion

Good urban connectivity creates sustainable, liveable, and accessible cities economically, socially, and environmentally. Tracing urban movements and human mobility is a crucial indicator of urban connectivity, which is increasingly becoming a topic of deep interest for urban planners, policymakers, and geographers. The study of human mobility is also important for urban planning (Barbosa et al., 2018). Understanding social activities in urban activity centres can benefit both urban authorities and citizens. Traditionally, monitoring large social activities usually incurs significant costs of human labour and time. Fortunately, with the recent booming of urban open data, a wide variety of human digital footprints have become openly accessible, providing us with new opportunities to understand the social dynamics in the cities (Longbiao et al., 2015). Recent years have witnessed an explosion of extensive geolocated datasets related to human movement, enabling scientists to study individual and collective mobility patterns quantitatively and to generate models that can capture and reproduce the spatiotemporal structures and regularities in human trajectories.

This research focused on identifying the urban connectivity between all the districts of the city of Tartu and identifying mobility patterns of people using social network analysis techniques.

This research employed TSB data for mapping the mobility patterns and understanding the inter-district connectivity. The temporal resolution of data was from June 2019 to December 2019. TSB dataset was aggregated on the district level, and then it was divided into two different approaches. In aggregated approach, only the trip's start and ending districts were used. In detailed approach the trajectory of the trip was splitted into all the intersecting districts.

The analysis mainly involved GIS operations, data analytics, social network analysis methods. In social network analysis, degree, strength and betweenness centralities were calculated. At the network level, average degree, density and modularity were calculated for both approaches. The results suggested that Ees-Annelinn, Kesk-Annelinn, Riiamäe, Sadama, Ülejõe, Uueturu, and Vanalinn were the most connected districts. Also, Uueturu, Kesk-Annelinn were the hub districts as they were the districts with the most connections of the most trip counts. Sadama was the bridge district, a binding node of the network. In conclusion, there was the concentration in the city centre and adjacent neighbourhoods for both approaches in terms of the trip count. The detailed approach showed more accurate situation of the people mobility as it was not concentrated in the city centre only. For future studies, this dataset can be used with other mobility datasets to show more concrete results and help validate results better.

Summary

Measuring urban connectivity using bike-share data: network analysis approach

Muhammad Hamza Zubair

Urban planners and policymakers are interested in tracing human mobility patterns to find the urban connectivity for making better data-driven decisions to improve the quality of life of the city dwellers (Galpern et al., 2018). Analysing mobility patterns helps in understanding urban connectivity that facilitates transport and infrastructure planning which is crucial for urban development (Makarova et al., 2017). There is already substantial research on studying urban connectivity on the level of transportation and street infrastructure (Dill, 2004; Makarova et al., 2017; Yang et al., 2019), but less which utilise human mobility data (Barbosa et al., 2018). The public bike-share system is an active mode of transport that is gaining popularity in different parts of the world due to its social, economic, environmental and health-related benefits. Most bike-share systems are equipped with GPS devices that collect spatio-temporal data on bike-share trips (Mooney et al., 2019). This kind of data provides an excellent opportunity to study the urban connectivity in terms of actual human mobility, which help urban planners to identify districts that are highly connected with other parts of the city and act as hubs, and also those areas that are poorly connected and need special attention.

In this master thesis, the social network analysis metrics were applied to the bike-share data for analysing urban connectivity in Tartu on the district level. The bike-share dataset comprising of GPS logs contained the trajectories of trips made by Tartu smart bike system users from June 2019 to December 2019. The bike-share dataset was spatially joined with the district’s boundaries layer. Mobility flows based on bike-share dataset were constructed using two different approaches. In the aggregated approach, the mobility flows were calculated by only including origin and destination districts of bike-share trips. In the detailed approach, the mobility flows were constructed by including all districts that constitute the trip. The trips were splitted in an order that the destination of the first splitted part would be the origin of the second and so forth. The mobility flows were summed for OD pairs based on the attributes of trip counts and user counts.

In this research, the city was considered as a directed network with districts as nodes of the network and bike-share trajectories as links among those nodes. Social network analysis (SNA) metrics were applied to the mobility flows to characterise urban connectivity in Tartu. SNA

measures were applied on local (node i.e. district) and global (network i.e. city) level. From a local perspective, degree centrality, weighted degree, and strength and betweenness centrality measures were calculated. From a global perspective, average degree, density, and modularity measures were calculated. These measures depict underlying characteristics of the whole network (Lin, 2017).

The results revealed that the city's central districts were more connected with each other in terms of trips and users movements than the other districts, which reflects Tartu city’s monocentricity. The outskirt districts showed poorer connectivity with themselves and rest of the network in comparison to the central districts. Taga-Annelinn, Uus-Ihaste, Variku and Supilinn do not have any docking station, so these districts had no connectivity in the aggregated approach, which was tackled using the detailed approach.

Based on the aggregated approach, the trip count showed that Tähtvere, Maarjamõisa, Ülejõe, Vanalinn, Riiamäe, Ees-Annelinn, Kesk-Annelinn, and Uueturu districts had the most incoming and outgoing trips among top ten OD pairs. Based on the aggregated approach and users movements, there were more commuters between Tähtvere, Ülejõe, Vanalinn, Riiamäe, Sadama, Kesk-Annelinn, and Uueturu districts as these were the contributing districts of top ten OD pairs. Based on the detailed approach, trip count indicated that Uus-Tammelinn, Vaksali, Vanalinn, Uueturu, Riiamäe, Ees-Annelinn, Ees-Karlova, and Ülejõe were the high-ranking districts with the most incoming and outgoing trips in between them. Uueturu, Vanalinn, Sadama, and Riiamäe had the highest inflow and outflow of commuters.

Local (district level) analysis showed that Kesk-Annelinn and Sadama were found as hub nodes of the city network. It means that these nodes had the most connections with the other nodes (i.e. districts). Such nodes are generally of higher importance for the connectivity of a network (Sun & Guan, 2016). Sadama district was found as a bridge node, which means this node acted as a bridge for connecting any two other nodes for the highest number of times compared to other nodes. Disruption in this node could disrupt network connectivity. This information could help in future infrastructure planning.

Global level social network analysis showed that the city as a whole is densely connected. The community detection showed that the network is forming three communities. It shows that these three communities have dense connections within themselves than the rest of the network.

There were some limitations in the analysis. First, not the whole population of Tartu is using the bike-share system. Second, other factors like the certain dominant age groups and seasonal variation affect bike-share system usage, which may influence the analysis.

Kokkuvõte

Linnalise ühenduvuse mõõtmine rattaringluse andmetel: võrgustiku analüüsi lähenemisviis

Muhammad Hamza Zubair

Linnaplaneerijad ja poliitikakujundajad on huvitatud inimeste liikumisvoogude jälgimisest, et teha paremaid andmepõhised otsuseid linnaelanike elukvaliteedi parandamiseks (Galpern et al., 2018). Liikumismustrite analüüs aitab mõista linnalist ühenduvust, mis hõlbustab linnaarengu jaoks üliolulise transpordi ja infrastruktuuri kavandamist (Makarova et al., 2017). Linnalist ühenduvust on eeskätt uuritud transpordi ja tänavate infrastruktuuri tasandil (Dill, 2004; Makarova et al., 2017; Yang et al., 2019), kuid palju vähem inimeste tegelike liikumisvoogude tasandil (Barbosa et al., 2018). Rattaringlus on transpordiliik, mis on oma sotsiaalsete, majanduslike, keskkonna ja tervisega seotud eeliste tõttu kogumas populaarsust erinevates maailma paikades. Enamik rattaringluse süsteeme on varustatud GPS-seadmetega, mis koguvad ajalis-ruumilisi andmeid rattaringluse reiside kohta (Mooney et al., 2019). Sellised andmed annavad suurepärase võimaluse uurida linnalist ühenduvust inimeste tegeliku liikuvuse alusel, mis aitab linnaplaneerijatel tuvastada linnaosasid, mis on tihedalt seotud teiste linnaosadega ja toimivad sõlmpunktidega, ning ka piirkondi, mis on halvasti ühendatud ja vajavad erilist tähelepanu.

Magistritöös uuriti Tartu asumite ühenduvust, kasutades rattaringluse andmeid ja sotsiaalse võrgustiku analüüsi. GPS-logidest koosnev rattaringluse andmestik sisaldas Tartu rattaringluse kasutajate reiside trajektoore alates juunist 2019 kuni detsembrini 2019. Rattaringluse andmed ühendati ruumiliselt asumite kihiga. Rattaringluse andmetel tuginevad liikumisvood koostati kasutades kahte erinevat lähenemist. Agregeeritud lähenemisviisi puhul arvutati liikumisvood kaasates üksnes rattaringluse reiside lähte- ja sihtkoha asumeid. Detailse lähenemisviisi puhul koostati liikumisvood kaasates kõiki reise moodustavaid asumeid. Liikumisvood konstrueeriti lähtudes reiside ja kasutajate arvust.

Antud uuringus vaadeldi linna kui suunatud võrgustikku, kus asumid on võrgustiku sõlmed ja rattaringluse trajektoorid seosed nende sõlmede vahel. Tartu linnalise ühenduvuse iseloomustamiseks rakendati liikumisvoogudele sotsiaalse võrgustiku analüüsi (*social network analysis* – SNA) mõõdikuid. SNA mõõdikuid rakendati kohalikul (sõlme ehk asumi) ja globaalsel (võrgustiku ehk linna) tasandil. Kohaliku tasandi vaatenurgast arvutati kesksuse aste

(degree centrality), kaalutud aste *(weighted degree)*, tugevuse ja vahendatud kesksuse *(strength and betweenness centrality)* näitajad. Globaalsest vaatenurgast arvutati keskmine aste *(average degree)*, tiheduse *(density)* ja modulaarsuse *(modularity)* näitajad. Need näitajad kujutavad endast kogu võrgustiku põhiomadusi (Lin, 2017). Tulemused tõid esile, et linna keskosad on reiside ja kasutajate liikumise osas omavahel rohkem seotud kui teised asumid, mis peegeldab Tartu linna ühekeskuselist struktuuri. Äärelinna asumid on kesklinna piirkondadega võrreldes kehvemini üksteisega ja ülejäänud võrgustikuga ühendatud. Taga-Annelinnas, Uus-Ihastes, Varikus ja Supilinnas ei olnud ühtegi rattaringluse parklat, mistõttu need piirkonnad olid agregeeritud lähenemisviisi puhul isoleeritud, kuid detailse lähenemisviisi puhul siiski teiste asumitega ühendatud.

Agregeeritud lähenemisviisi tulemused reiside arvu põhjal näitasid, et kõige rohkem sissetulevaid ja väljuvaid reise on seotud Tähtvere, Maarjamõisa, Ülejõe, Vanalinna, Riiamäe, Ees-Annelinna, Kesk-Annelinna ja Uueturu asumitega. Agregeeritud lähenemisviisi tulemused kasutajate arvu põhjal näitasid, et Tähtvere, Ülejõe, Vanalinna, Riiamäe, Sadama, Kesk-Annelinna ja Uueturu asumite vahel on kõige rohkem rattaringluse kasutajaid. Detailse lähenemisviisi tulemused reiside arvu põhjal näitasid, et Uus-Tammelinn, Vaksali, Vanalinn, Uueturu, Riiamäe, Ees-Annelinn, Ees-Karlova ja Ülejõe vahel oli kõige rohkem sisenevaid ja väljuvaid reise. Uueturu, Vanalinna, Sadama ja Riiamäe asumites oli kõige rohkem sisse ja välja pendeldajaid.

Kohaliku tasandi (asumid) analüüs näitas, et Kesk-Annelinna ja Sadama asumid on linnalise võrgustiku sõlmpunktid. See tähendab, et nendel sõlmpunktidel on kõige rohkem ühendusi teiste sõlmedega (asumitega). Sellised sõlmed on võrgustiku ühenduvuse seisukohalt üldiselt suurema tähtsusega (Sun & Guan., 2016). Sadama asum toimib sillasõlmena *(bridge node)*. Selle sõlme katkestamine võib häirida võrgustiku ühenduvust. See teave võib aidata kaasa infrastruktuuri kavandamisel tulevikus.

Globaalse taseme sotsiaalse võrgustiku analüüs näitas, et linn tervikuna on tihedalt ühendatud. Kogukondade tuvastamine *(community detection)* näitas, et võrgustikus on rattaringluse andmete põhjal kolm kogukonda. See näitab, et neil kolmel kogukonnal on tihedad ühendused nii omavahel kui ka ülejäänud võrgustikuga. Analüüsis olid ka mõned puudujäägid. Esiteks ei kasuta rattaringluse süsteemi kogu Tartu elanikkond. Teiseks mõjutavad rattaringluse süsteemi kasutamist muud tegurid, näiteks teatud vanuserühmade domineerimine ja aastaajast tulenev varieeruvus, mis võivad analüüsi tulemusi mõjutada.

Acknowledgements

I sincerely acknowledge the cooperation and help of my Supervisors **Veronika Mooses and Dr Siiri Silm.** They had been a constant source of guidance throughout this research, and without their counsel and support, I would not have been able to conduct this research. My profound sense of gratitude is due to my all teachers who made me capable of what I am today.

I am highly indebted to the Archimedes Foundation for supporting my studies. It will always be a great honour for me that the University of Tartu is my Alma Mater. My heartfelt thanks to Estonia for showing me a great deal of love. It will always be my second home no matter where I go and will always be dear to my heart. Finally, thanks to my Mother, Family and Friends who have supported me on this journey.

References

Ahas, R., Aasa, A., Yuan, Y., Martin, R., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., & Zook, M. (2015). Everyday space–time geographies: Using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science*, *29*. https://doi.org/10.1080/13658816.2015.1063151

Ahas, R., Silm, S., Järv, O., Saluveer, E., & Tiru, M. (2010). Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology*, *17*(1), 3–27. https://doi.org/10.1080/10630731003597306

Antoniou, I. E., & Tsompa, E. T. (2008). Statistical Analysis of Weighted Networks. *Discrete Dynamics in Nature and Society*, *2008*, 1–16. https://doi.org/10.1155/2008/375452

Australia, Parliament, House of Representatives, Standing Committee on Infrastructure, T. and C., & Alexander, J. (2018). *Building up & moving out: Inquiry into the Australian Government's role in the development of cities*.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, *734*, 1–74. https://doi.org/10.1016/j.physrep.2018.01.001

Behnisch, M., Schorcht, M., Kriewald, S., & Rybski, D. (2019). Settlement percolation: A study of building connectivity and poles of inaccessibility. *Landscape and Urban Planning*, *191*, 103631. https://doi.org/10.1016/j.landurbplan.2019.103631

Biggs, R., Schlüter, M., & Schoon, M. L. (Eds.). (2015). *Principles for Building Resilience: Sustaining Ecosystem Services in Social-Ecological Systems*. Cambridge University Press. https://doi.org/10.1017/CBO9781316014240

Borgnat, P., Abry, P., Flandrin, P., Robardet, C., Rouquier, J.-B., & Fleury, E. (2011). Shared bicycles in a city: a signal processing and data analysis perspective. *Advances in Complex Systems*, *14*(03), 415–438. https://doi.org/10.1142/S0219525911002950

Bullock, C., Brereton, F., & Bailey, S. (2017). The economic contribution of public bike-share to the sustainability and efficient functioning of cities. *Sustainable Cities and Society*, *28*, 76–87. https://doi.org/10.1016/j.scs.2016.08.024

Chen, L., Yang, D., Jakubowicz, J., Pan, G., Zhang, D., & Li, S. (2015). Sensing the Pulse of Urban Activity Centers Leveraging Bike Sharing Open Data. *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, 135–142. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.43

Dill, J. (2004). Measuring Network Connectivity for Bicycling and Walking. *83rd Annual Meeting of the Transportation Research Board: Washington, D.C.*

Ebrahimpour, Z., Wan, W., Velázquez García, J. L., Cervantes, O., & Hou, L. (2020). Analyzing Social-Geographic Human Mobility Patterns Using Large-Scale Social Media Data. *ISPRS International Journal of Geo-Information*, *9*(2), 125. https://doi.org/10.3390/ijgi9020125

Galpern, P., Ladle, A., Alaniz Uribe, F., Sandalack, B., & Doyle-Baker, P. (2018). Assessing urban connectivity using volunteered mobile phone GPS locations. *Applied Geography*, *93*, 37–46. https://doi.org/10.1016/j.apgeog.2018.02.009

Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., & Snasel, V. (2012). Social Networks Analysis: Tools, Measures and Visualization. In A. Abraham (Ed.), *Computational Social Networks* (pp. 3–23). Springer London. https://doi.org/10.1007/978-1-4471-4054-2_1

Gil, J., & Read, S. (2013). *Patterns of sustainable mobility and the structure of modality in the Randstad city-region*. *11*, 027:1-027:22.

Haznagy, A., Fi, I., London, A., & Nemeth, T. (2015). Complex network analysis of public transportation networks: A comprehensive study. *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 371–378. https://doi.org/10.1109/MTITS.2015.7223282

Hong, L., Zheng, Y., Yung, D., Shang, J., & Zou, L. (2015). Detecting urban black holes based on human mobility data. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 1–10. https://doi.org/10.1145/2820783.2820811

Jia, G.-L., Ma, R.-G., & Hu, Z.-H. (2019). Urban Transit Network Properties Evaluation and Optimization Based on Complex Network Theory. *Sustainability*, *11*(7), 2007. https://doi.org/10.3390/su11072007

Jurdak, R. (2013). The Impact of Cost and Network Topology on Urban Mobility: A Study of Public Bicycle Usage in 2 U.S. Cities. *PLoS ONE*, *8*(11), e79396. https://doi.org/10.1371/journal.pone.0079396

Kareem, B. (2017, December 9). Urban Connectivity is a Catalyst for Leaving No One Behind. *The Nature of Cities*. https://www.thenatureofcities.com/2017/12/09/urban-connectivity-catalyst-leaving-no-one-behind/

Koohsari, M. J., Sugiyama, T., Lamb, K. E., Villanueva, K., & Owen, N. (2014). Street connectivity and walking for transport: Role of neighborhood destinations. *Preventive Medicine*, *66*, 118–122. https://doi.org/10.1016/j.ypmed.2014.06.019

Kunaka, C. (n.d.). *Why Connectivity Matters*. 13.

Li, W., Wang, S., Zhang, X., Jia, Q., & Tian, Y. (2020). Understanding intra-urban human mobility through an exploratory spatiotemporal analysis of bike-sharing trajectories.

International Journal of Geographical Information Science, *34*(12), 2451–2474. https://doi.org/10.1080/13658816.2020.1712401

Lin, Jingyi. (n.d.). *Spatial analysis and modeling of urban transportation networks*. 90.

Lin, Jinyao, Wu, Z., & Li, X. (2019). Measuring inter-city connectivity in an urban agglomeration based on multi-source data. *International Journal of Geographical Information Science*, *33*(5), 1062–1081. https://doi.org/10.1080/13658816.2018.1563302

Lu, S., Huang, Y., Shi, C., & Yang, X. (2019). Exploring the Associations Between Urban Form and Neighborhood Vibrancy: A Case Study of Chengdu, China. *ISPRS International Journal of Geo-Information*, *8*(4), 165. https://doi.org/10.3390/ijgi8040165

Makarova, I., Pashkevich, A., Shubenkova, K., & Mukhametdinov, E. (2017). Ways to Increase Population Mobility through the Transition to Sustainable Transport. *Procedia Engineering*, *187*, 756–762. https://doi.org/10.1016/j.proeng.2017.04.434

Miyazaki, H. (2019). *Measurement of Inter- and Intra-city Connectivity Using Vehicle Probe Data*. 36.

Mohamad, S., & Said, I. (2014). A review of variables of urban street connectivity for spatial connection. *IOP Conference Series: Earth and Environmental Science*, *18*. https://doi.org/10.1088/1755-1315/18/1/012173

Mooney, S. J., Hosford, K., Howe, B., Yan, A., Winters, M., Bassok, A., & Hirsch, J. A. (2019). Freedom from the station: Spatial equity in access to dockless bike share. *Journal of Transport Geography*, *74*, 91–96. https://doi.org/10.1016/j.jtrangeo.2018.11.009

Moroke, T., Schoeman, C., & Schoeman, I. (2019). Developing a neighbourhood sustainability assessment model: An approach to sustainable urban development. *Sustainable Cities and Society*, *48*, 101433. https://doi.org/10.1016/j.scs.2019.101433

Mueller, N., Rojas-Rueda, D., Cole-Hunter, T., Nazelle, A., Dons, E., Gerike, R., Götschi, T., Int Panis, L., Kahlmeier, S., & Nieuwenhuijsen, M. (2015). Health impact assessment of active transportation: A systematic review. *Preventive Medicine*, *76*, 103–114. https://doi.org/10.1016/j.ypmed.2015.04.010

Nel, D. H., Higgins, C. D., & Bruyns, G. (2018). Urban Design, Connectivity and its Role in Building Spatial Resilience. *XXV International Seminar on Urban Form 2018 Urban Form and Social Context: From Traditions to Newest Demands*. XXV International Seminar on Urban Form 2018 : Urban Form and Social Context: from traditions to newest demands. https://research.polyu.edu.hk/en/publications/urban-design-connectivity-and-its-role-in-building-spatial-resili

Network Science by Albert-László Barabási. (n.d.). Retrieved February 18, 2021, from http://networksciencebook.com/

Newman, M. E. J., & Girvan, M. (2003). Finding and evaluating community structure in networks. *ArXiv:Cond-Mat/0308217*. https://doi.org/10.1103/PhysRevE.69.026113

Peponis, J., Bafna, S., & Zhang, Z. (2008). The connectivity of streets: Reach and directional distance. *Environment and Planning B: Planning and Design*, *35*(5), 881–901. https://doi.org/10.1068/b33088

Press, T. M. (n.d.). *The New Science of Cities | The MIT Press*. The MIT Press. Retrieved January 6, 2021, from https://mitpress.mit.edu/books/new-science-cities

Psaltoglou, A., & Calle, E. (2018). Enhanced connectivity index – A new measure for identifying critical points in urban public transportation networks. *International Journal of Critical Infrastructure Protection*, *21*, 22–32. https://doi.org/10.1016/j.ijcip.2018.02.003

Reggiani, A., Nijkamp, P., & Lanzi, D. (2015). Transport resilience and vulnerability: The role of connectivity. *Transportation Research Part A: Policy and Practice*, *81*, 4–15. https://doi.org/10.1016/j.tra.2014.12.012

Rybarczyk, G., & Wu, C. (2014). Examining the Impact of Urban Morphology on Bicycle Mode Choice. *Environment and Planning B: Planning and Design*, *41*(2), 272–288. https://doi.org/10.1068/b37133

Shanmukhappa, T., Ho, I. W.-H., & Tse, C. K. (2018). Spatial analysis of bus transport networks using network theory. *Physica A: Statistical Mechanics and Its Applications*, *502*, 295–314. https://doi.org/10.1016/j.physa.2018.02.111

Silm, S., Järv, O., & Masso, A. (2020). Tracing human mobilities through mobile phones. In M. Büscher, M. Freudendal-Pedersen, S. Kesselring, & N. Grauslund Kristensen, *Handbook of Research Methods and Applications for Mobilities* (pp. 182–192). Edward Elgar Publishing. https://doi.org/10.4337/9781788115469.00025

Silm, S., Jauhiainen, J. S., Raun, J., & Tiru, M. (2020). Temporary population mobilities between Estonia and Finland based on mobile phone data and the emergence of a cross-border region. *European Planning Studies*, 1–21. https://doi.org/10.1080/09654313.2020.1774514

Sine, W. D. (2011). Organization Science. In R. Mitchell & R. Dino, *In Search of Research Excellence* (p. 14204). Edward Elgar Publishing. https://doi.org/10.4337/9781849807630.00024

Song, J., Zhang, L., Qin, Z., & Ramli, M. A. (2021). A spatiotemporal dynamic analyses approach for dockless bike-share system. *Computers, Environment and Urban Systems*, *85*, 101566. https://doi.org/10.1016/j.compenvurbsys.2020.101566

Song, Y., Shao, G., Song, X., Liu, Y., Pan, L., & Ye, H. (2017). The Relationships between Urban Form and Urban Commuting: An Empirical Study in China. *Sustainability*, *9*(7), 1150. https://doi.org/10.3390/su9071150

Sun, D. (Jian), & Guan, S. (2016). Measuring vulnerability of urban metro network from line operation perspective. *Transportation Research Part A: Policy and Practice*, *94*, 348–359. https://doi.org/10.1016/j.tra.2016.09.024

Tabassum, S., Pereira, F. S. F., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *WIREs Data Mining and Knowledge Discovery*, *8*(5). https://doi.org/10.1002/widm.1256

Wang, J., Kong, X., Xia, F., & Sun, L. (2019). Urban Human Mobility: Data-Driven Modeling and Prediction. *ACM SIGKDD Explorations Newsletter*, *21*(1), 1–19. https://doi.org/10.1145/3331651.3331653

Wang, R., & Su, Y. (2019, March 4). *What does bikeshare trip data reveal about connections across DC neighborhoods?* Urban Institute. https://www.urban.org/urban-wire/what-does-bikeshare-trip-data-reveal-about-connections-across-dc-neighborhoods

Watts, A., Ferdous, F., Diaz Moore, K., & Burns, J. M. (2015). Neighborhood Integration and Connectivity Predict Cognitive Performance and Decline. *Gerontology and Geriatric Medicine*, *1*. https://doi.org/10.1177/2333721415599141

Willberg, E. (n.d.). *Master's thesis Geography Geoinformatics*. 114.

Woodcock, J., Tainio, M., Cheshire, J., O'Brien, O., & Goodman, A. (2014). Health effects of the London bicycle sharing system: Health impact modelling study. *BMJ*, *348*, g425. https://doi.org/10.1136/bmj.g425

Wu, C., & Kim, I. (2020). Analyzing the structural properties of bike-sharing networks: Evidence from the United States, Canada, and China. *Transportation Research Part A: Policy and Practice*, *140*, 52–71. https://doi.org/10.1016/j.tra.2020.07.018

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. (2019). A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems*, *77*, 101361. https://doi.org/10.1016/j.compenvurbsys.2019.101361

Yao, Y., Zhang, Y., Tian, L., Zhou, N., Li, Z., & Wang, M. (2019). Analysis of Network Structure of Urban Bike-Sharing System: A Case Study Based on Real-Time Data of a

Public Bicycle System. *Sustainability*, *11*(19), 5425. https://doi.org/10.3390/su11195425

Zhang, W., Chong, Z., Li, X., & Nie, G. (2020). Spatial patterns and determinant factors of population flow networks in China: Analysis on Tencent Location Big Data. *Cities*, *99*, 102640. https://doi.org/10.1016/j.cities.2020.102640

Annexes

Annexe 1: District codes with number of docking station

District code	District name	Number of docking stations
01	Ees-Annelinna	2
02	Taga-Annelinna	0
03	Kesk-Annelinna	10
04	Vana-Ihaste	1
05	Uus-Ihaste	0
06	Jaamamõisa	1
07	Raadi	2
08	Kruusamäe	3
09	Ülejõe	6
10	Ujula-Kvissentali	2
11	Supilinn	0
12	Tähtvere	5
13	Veeriku	2
14	Veeriku tööstuse	1
15	Uueturu	2
16	Ropka tööstuse	3
17	Ees-Karlova	2
18	Taga-Karlova	1
19	Variku	0
20	Ränilinn	2
21	Maarjamõisa	4
22	Ropkamõisa	1
23	Jalaka	3
24	Uus-Tammelinna	4
25	Vana-Tammelinna	1
26	Vaksali	1
27	Kastani-Filosoofi	2
28	Vanalinna	4
29	Toometaguse	1
30	Riiamäe	4
31	Sadama	3

Annexe 2: Node level SNA metrics aggregated approach

District Name	In-degree	Out-degree	Degree	Hub	Betweenness	Weighted in-degree	Weighted out-degree	Weighted degree	Modularity
Ees-Annelinna	28	28	56	0.19	3.48	21589	20746	42335	1
Ees-Karlova	28	29	57	0.19	5.71	18018	18957	36975	2
Jaamamõisa	27	27	54	0.19	0.17	8802	8218	17020	1
Jalaka	29	28	57	0.18	6.48	14452	13983	28435	2
Kastani-Filosoofi	28	29	57	0.19	4.22	18611	18280	36891	2
Kesk-Annelinna	30	29	59	0.19	16.45	53333	53083	106416	1
Kruusamäe	28	28	56	0.19	3.30	16640	16036	32676	0
Maarjamõisa	28	26	54	0.18	0.59	18575	19462	38037	2
Raadi	27	27	54	0.19	0.17	6560	6490	13050	0
Ränilinn	29	28	57	0.19	3.96	15883	15736	31619	2
Riiamäe	29	28	57	0.19	4.89	34127	33526	67653	0
Ropka tööstuse	29	29	58	0.19	6.85	4468	4798	9266	2
Ropkamõisa	28	28	56	0.19	1.99	11065	11067	22132	2
Sadama	29	30	59	0.19	16.53	35699	35850	71549	1
Supilinn	23	27	50	0.19	0.00	299	940	1239	0
Taga-Annelinna	8	3	11	0.02	0.00	18	10	28	1
Taga-Karlova	28	28	56	0.19	3.30	11751	11752	23503	2
Tähtvere	28	27	55	0.19	2.00	25404	23955	49359	0
Toometaguse	27	27	54	0.19	0.17	4165	4995	9160	0
Ujula-Kvissentali	29	29	58	0.19	6.19	17502	16055	33557	0
Ülejõe	28	29	57	0.19	4.22	32056	33475	65531	0
Uueturu	29	28	57	0.19	11.74	48144	48242	96386	0
Uus-Ihaste	9	10	19	0.07	0.00	84	63	147	1
Uus-Tammelinna	28	29	57	0.19	4.63	16417	15970	32387	2
Vaksali	28	27	55	0.19	0.83	12472	13533	26005	2
Vana-Ihaste	28	29	57	0.19	7.86	5635	5188	10823	1
Vanalinna	27	27	54	0.19	0.17	43194	43060	86254	0
Vana-Tammelinna	28	27	55	0.18	1.74	4191	5127	9318	2
Variku	13	17	30	0.12	0.00	74	165	239	2
Veeriku	27	27	54	0.19	0.17	11405	11696	23101	2
Veeriku tööstuse	27	27	54	0.19	0.17	12162	12337	24499	2

Annexe 3: Node level SNA metrics for detailed approach

District Name	In-degree	Out-degree	Degree	Hub	Betweenness	Weighted in-degree	Weighted out-degree	Weighted degree	Modularity
Ees-Annelinna	31	31	62	0.18	0	580869	520702	1101571	0
Ees-Karlova	31	31	62	0.18	0	407339	379912	787251	1
Jaamamõisa	31	31	62	0.18	0	75113	83150	158263	0
Jalaka	31	31	62	0.18	0	122402	163423	285825	1
Kastani-Filosoofi	31	31	62	0.18	0	362334	360575	722909	2
Kesk-Annelinna	31	31	62	0.18	0	329701	463571	793272	0
Kruusamäe	31	31	62	0.18	0	103411	146993	250404	0
Maarjamõisa	31	31	62	0.18	0	274469	308571	583040	2
Raadi	31	31	62	0.18	0	105737	102510	208247	0
Ränilinn	31	31	62	0.18	0	98019	154395	252414	2
Riiamäe	31	31	62	0.18	0	710933	692829	1403762	2
Ropka tööstuse	31	31	62	0.18	0	97279	98494	195773	1
Ropkamõisa	31	31	62	0.18	0	281355	271175	552530	1
Sadama	31	31	62	0.18	0	569859	581990	1151849	0
Supilinn	31	31	62	0.18	0	143637	121039	264676	0
Taga-Annelinna	31	31	62	0.18	0	57032	44510	101542	1
Taga-Karlova	31	31	62	0.18	0	286371	270151	556522	1
Tähtvere	31	31	62	0.18	0	272753	306962	579715	2
Toometaguse	31	31	62	0.18	0	332282	285256	617538	2
Ujula-Kvissentali	31	31	62	0.18	0	249091	251337	500428	0
Ülejõe	31	31	62	0.18	0	695175	653973	1349148	0
Uueturu	31	31	62	0.18	0	925738	899018	1824756	0
Uus-Ihaste	31	31	62	0.18	0	26276	20413	46689	1
Uus-Tammelinna	31	31	62	0.18	0	375545	371448	746993	2
Vaksali	31	31	62	0.18	0	387672	367080	754752	2
Vana-Ihaste	31	31	62	0.18	0	68630	73291	141921	1
Vanalinna	31	31	62	0.18	0	688612	677899	1366511	0
Vana-Tammelinna	31	31	62	0.18	0	385953	342330	728283	2
Variku	31	31	62	0.18	0	85800	70181	155981	1
Veeriku	31	31	62	0.18	0	171963	181975	353938	2
Veeriku tööstuse	31	31	62	0.18	0	173328	179525	352853	2

Non-exclusive licence to reproduce thesis and make thesis public

I, Muhammad Hamza Zubair,

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,

"Measuring urban connectivity using bike-share data: network analysis approach" supervised by Veronika Mooses and Dr.Siiri Silm.

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.

Muhammad Hamza Zubair

Tartu, 24.05.2021