

Manuscript Number: JTRG-D-15-00385R2

Title: PERCEPTUALLY IMPORTANT POINTS OF MOBILITY PATTERNS TO CHARACTERISE
BIKE SHARING SYSTEMS: THE DUBLIN CASE

Article Type: Full Length Article

Keywords: Bike-sharing; Data mining; Mobility patterns; Turnover station
rate

Corresponding Author: Dr. Pilar Jiménez, Ph.D.

Corresponding Author's Institution: Technical University of Cartagena

First Author: Pilar Jiménez, Ph.D.

Order of Authors: Pilar Jiménez, Ph.D.; María Nogal, Ph.D.; Brian
Caulfield, Ph.D.; Francesco Pilla, Ph.D.

Abstract: Since the first Bike Sharing System (BSS) was introduced in Amsterdam (1965), studies about BSSs have constantly increased. BSSs studies are typically focused on user's socio-economic characteristics, bike sharing patterns and purpose of use in the city. This paper increases the knowledge of bike station classification due to users' mobility patterns based on data mining tools. For this purpose stations will be identified by a code based on joining three ratios: the load factor or number of available bikes ratio, the cumulative trips ratio, and the turnover station ratio. The latter is the new ratio proposed in this paper, which measures the effectiveness degree of each station. The higher the rate, the more effective the station is. Data mining tools to work with these three ratios are used in the proposed algorithm. Specifically, the perceptually important points (PIP) process to represent and index each time series of each station, and a rule set to classify the stations, are used. The results could support planning and operations decisions for re-design and management of BSSs in relation to the spatial implications of the stations and the users' mobility patterns, due to the classification reveals imbalances in the distribution of bikes and lead to a better understanding of the system structure. The proposed method is applied to the Dublin Bikes Scheme with good performance results.

PERCEPTUALLY IMPORTANT POINTS OF MOBILITY PATTERNS TO CHARACTERISE BIKE SHARING SYSTEMS: THE DUBLIN CASE.

Abstract

Since the first Bike Sharing System (BSS) was introduced in Amsterdam (1965), studies about BSSs have constantly increased. BSSs studies are typically focused on user's socio-economic characteristics, bike sharing patterns and purpose of use in the city. This paper increases the knowledge of bike station classification due to users' mobility patterns based on data mining tools. For this purpose stations will be identified by a code based on joining three ratios: the load factor or number of available bikes ratio, the cumulative trips ratio, and the turnover station ratio. The latter is the new ratio proposed in this paper, which measures the effectiveness degree of each station. The higher the rate, the more effective the station is. Data mining tools to work with these three ratios are used in the proposed algorithm. Specifically, the perceptually important points (PIP) process to represent and index each time series of each station, and a rule set to classify the stations, are used. The results could support planning and operations decisions for re-design and management of BSSs in relation to the spatial implications of the stations and the users' mobility patterns, due to the classification reveals imbalances in the distribution of bikes and lead to a better understanding of the system structure. The proposed method is applied to the Dublin Bikes Scheme with good performance results.

Key Words: Bike-sharing, data mining, mobility patterns, turnover station rate.

1 Introduction and Contributions

Since the first Bike Sharing System (BSS) was introduced in Amsterdam (1965), studies about BSSs have constantly increased. BSSs studies are typically focused on user's socio-economic characteristics, bike sharing patterns and purpose of use in the city.

A number of studies have begun to classify bike station on the basis of users' mobility patterns. For example, O'Neill and Caulfield (2012), after analysing the Dublin Bikes System, distinguish three different patterns of bike stations based on the pickup and return activity at stations in the daily course during the working week: Go-From stations, Go-To stations and Self-Sustainable stations. Under the same point of view, Vogel et al. (2011) add two more clusters: Active Night Pickups Morning stations and Active Daytime stations, according to data from Vienna's BSS (CityBike Wien). Lathia et al. (2012) assessed the impacts of the introduction of "casual" users to the shared bicycle scheme in London. For this purpose six clusters of docking stations were identified, although they are classified in three clusters: i) Day-Time Origins, ii) Day-Time destinations, and iii) Combined Origins/Destinations, at the same way that O'Neill and Caulfield (2012).

Among bikesharing studies at the urban level, García-Palomares et al. (2012) proposed a GIS-based method to calculate the spatial distribution of the potential demand for bikesharing trips in Madrid. Location determines the characteristics of each base, either as a trip generator or attractor, depending on whether its potential demand comes from residential areas or areas of economic activity. Four different types of bike station are distinguished based on data for potential demand: generators, mixed, attractors, and

high attractors. Recently, Chabchoub and Fricker (2014) have analysed the Parisian BSS, called Velib, in order to separate the Velib stations into three categories: underloaded, overloaded and balanced stations.

More generally, studies that seek to classify bike stations according to mobility patterns tend to identify three clusters:

- Generator stations of trips at early hours (Go-From, Pickups Morning Return Evening, Day - Time Origins, Underloaded). The activity in the morning is predominantly as a result of people taking bicycles out of the station, therefore the number of available spaces increases. In the evening, return activity is higher than pickup activity.
- Attractor stations of trips (Go-To, Returns Mornings Pickups Evenings, Day-Time Destination, Overloaded). The activity in the morning is predominantly as a result of people docking bicycles in the stations, therefore available spaces decrease. In the evening periods, the opposite happens as people leave the area, pickup activity is higher than return activity.
- Balanced stations when trips between arrivals and departures are balanced (Self-Sustainable, Average Station, Combined Origin/Destination, Balanced). The number of bicycles being docked at the station is similar to the number of bicycles being taken out from the station during the day.

The aim of this research is to increase the knowledge of bike station classification according to users' mobility patterns using data mining tools, that is, to check if the analysed bike sharing scheme works correctly based on a new bike station classification in order to identify the stations that could be improved.

For this purpose stations will be identified by a code based on joining three ratios:

- Number of Available Bicycles or load factor, previously used by Froehlich et al. (2009), Lathia et al. (2012) or O'Brien et al. (2014).
- Cumulative trips, applied by Chabchoub and Fricker (2014).
- Turnover Station ratio (TS), the new ratio proposed in this paper. It measures the effectiveness degree of each station, that is how many times the capacity of the station is used. The higher the rate, the more effective a station is.

Note that, Turnover Station ratio is a modification of the Turnover rate of Zhao et al. (2014) who analyse the turnover rate of public bikes, rate per bicycle per day, meanwhile the proposed turnover ratio is per station per day. The Zhao et al. (2014) paper evaluates factors affecting bikesharing daily use and turnover rate by analyzing data from 69 BSSs in China in order to know how many ridership BSS can attract, and what influences their effectiveness.

The combination of these ratios offers information about bike stations in relation to their spatial implications and the users' mobility patterns. The results could support planning and operations decisions for re-design and management of bike sharing systems in the city, due to the classification reveals imbalances in the distribution of bikes and lead to a better understanding of the system structure. In other words, results confirm which

stations work correctly, or which need a deep study to improve its resources (bikes and docks) and avoid the detected imbalanced or low use.

On the other hand, the other contribution of the paper is the difference in the mathematical tools used. Taking into consideration the various researches carried out about the temporal pattern discovery and clustering of BSSs, the supervised or concept learning algorithm has been selected instead of the unsupervised learning ones for the cluster analysis¹. That is, the proposed algorithm only has to classify the input data based on previous clusters that comes from previous works, training data and examples. Therefore, instead of using cluster analysis other tools of time series data mining have been used, specifically the perceptually important points (PIP) process to represent and index each time series of each station, because this process reduces the dimensionality of the original series but keeping the structure of the pattern. Next, a rule set to classify the data is applied.

This paper is structured as follows. In Section 2 the ratios and clusters used for the bike station classification are described. The data mining tools and algorithm used are presented in Section 3. In Section 4 the algorithm is applied to the Dublin Bikes Sharing Systems to show the performance of the methods. Finally, Section 5 provides some conclusions and possible future works.

2 Data analysis: ratios and clusters

2.1 Ratios

As mentioned before, the objective of the paper is to define bikesharing stations based on three ratios:

- Number of Available Bicycles, NAB (Froehlich et al. (2009)). It is calculated by dividing the current number of available bicycles in the station i at time t ($B_{i,t}$) by the station capacity (C_i) defined as its number of docks:

$$NAB_{i,t} = \frac{B_{i,t}}{C_i}. \quad (1)$$

- Cumulative Trips ratio, CumT (Chabchoub and Fricker (2014)). It is a ratio between arrivals and departures relative to each station during the whole day, that is:

$$CumT_{i,t} = \frac{a_{i,t} - d_{i,t}}{C_i}, \quad (2)$$

where $CumT_{i,t}$ is the ratio from cumulative trips to the station i at each time t , $a_{i,t}$ the cumulative number of arrivals at the station i between 0 hours and the time t , $d_{i,t}$ the cumulative number of departures from the station i at the same period of time of arrivals, and C_i the capacity of the station, defined as its number of docks. Both the NAB and CumT ratio show the variation in the resources of the station, that is, the availability of docks and bikes during all day.

¹Note that, an unsupervised learning does not need any priori knowledge about the processed data, only the number k of clusters is required as an input of the algorithm (see MacQueen (1967), Liao (2005) and Chabchoub and Fricker (2014) among other for more detail).

- Turnover Station ratio, TS. It is defined as the ratio between the total number of arrivals (a_i) at the end of the day at the station i and the station capacity C_i . The same ratio can be calculated for the total number of departures (d_i) at the end of the day at the station i :

$$TSa_i = \frac{a_i}{C_i}, \quad (3)$$

or

$$TSd_i = \frac{d_i}{C_i}. \quad (4)$$

In this paper, the ratios are applied to the Dublin BSS, called Dublin Bikes, whose scheme counts with 101 stations and 1500 bikes (see Section 4 for more detail). The data of the station located in Princess Street near to O'Connell Street for a week day (19th of March of 2015) are used to explain how the ratios work. This station is located in a commercial and tourist district of the city. For that reason, it belongs to an attractor station, with a high number of available bikes during the main hours. The number of bikes decreases when shops close however, suddenly, it increases again due to the activity night in the area.

NAB ratio may be interpreted as the percentage of the station's occupied docking spaces. It varies between 0 value, which means an empty station, and 1 value which means that the station is full, all bicycles are available. For example, Figure 1 shows the NAB pattern in a week day for the station in Princess Street, which indicates that this station has few available bicycles at very early morning, and near midday suddenly all docks are occupied ($NAB = 1$). Then, this percentage is basically maintained until last evening hours, when the number of available bikes decreases step by step until reaching similar availability of bicycles at the station than the start of the day. This is a typical scheme for an attractive station.



Figure 1: NAB ratio representation for a work day of the example station.

At the similar way, if cumulative trip ratio is represented for this station Figure 2 is obtained. Note that this ratio varies between -1 and 1 , therefore a value near to 0 indicates

that the station is balanced, the number of arrivals and departures are similar, that is the number of available bicycles is balanced during the day. Negative values indicate more departures than arrivals, meanwhile positive values indicate more arrivals than departures. In this case, the illustration shows that at the first hours of the day the number of available bikes and docks are similar. Then, the number of arrivals increases (positive values). This state is maintained roughly constant, to show a step by step departures towards the end of the day (negative slope).

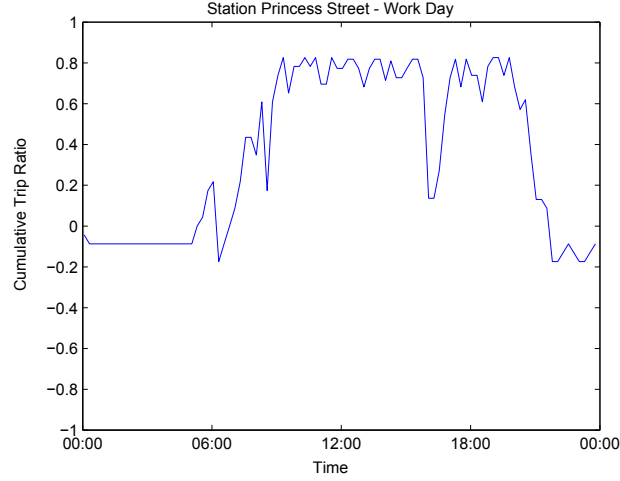


Figure 2: Cumulative trip ratio representation for a work day of the example station.

Note that the shape of both Figures 1 and 2 are the same and the information supplies are similar, that is the main pattern is clearly identified as an attractive station with morning return activity but evening pickup activity. The only difference is the values of each ratio, its scale.

However, the turnover ratio per station provides a new interpretation. The main difference is related to the period. The two previous ratios show a pattern during the day, whereas the turnover ratio is one value for one day: turnover arrivals rate or turnover departures rate, or both values at maximum features. This happens because the last value is required to know how many times the station capacity is used throughout a complete day. Figure 3 shows that the activity of bikes during weekends and bank holidays is lesser than during weekend days. Note that the values of both ratios are smaller than 1 on Saturday, that is the whole station capacity is not used. Some times, extra information can be identified. For example, in this week, Tuesday is the day with the lowest activity among the workdays because it was St's Patrick Day (17th of March), which is a public holiday in Ireland.

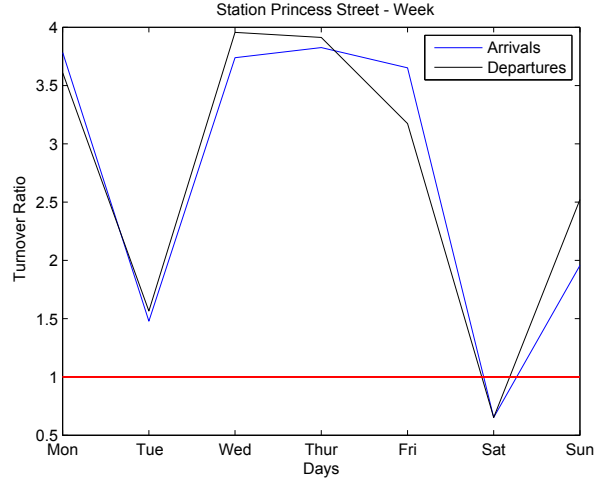


Figure 3: Turnover ratio representation (arrivals and departures) for a week of the example station.

2.2 Clusters per ratio

Due to the well-known general mobility pattern in bikesharing systems it is easy to distinguish a priori the cluster for each ratio, taking into account the interpretation of the scales of each one.

2.2.1 Number of Available Bicycles ratio clusters

Nine clusters have been identified for NAB ratio (see Table 1). The first classification depends on the gap between series extreme values throughout the day. Three groups can be distinguished:

- **Small Gap (SG).** The difference between the extreme values of data is equal or lower than 0.3. There are not high differences among bikes and slots during the day.
- **Medium Gap (MG).** The difference between the extreme values of data is in the range 0.3 and 0.7.
- **High Gap (HG).** The difference between the extreme values of data is equal or higher than 0.7, that is, there are high differences in the number of available bikes during the day within the station.

Note that, the considered reference values can be modified adding a small error in order to represent better the patterns of a particular bike sharing systems, thus each city has some particular issues to take into account.

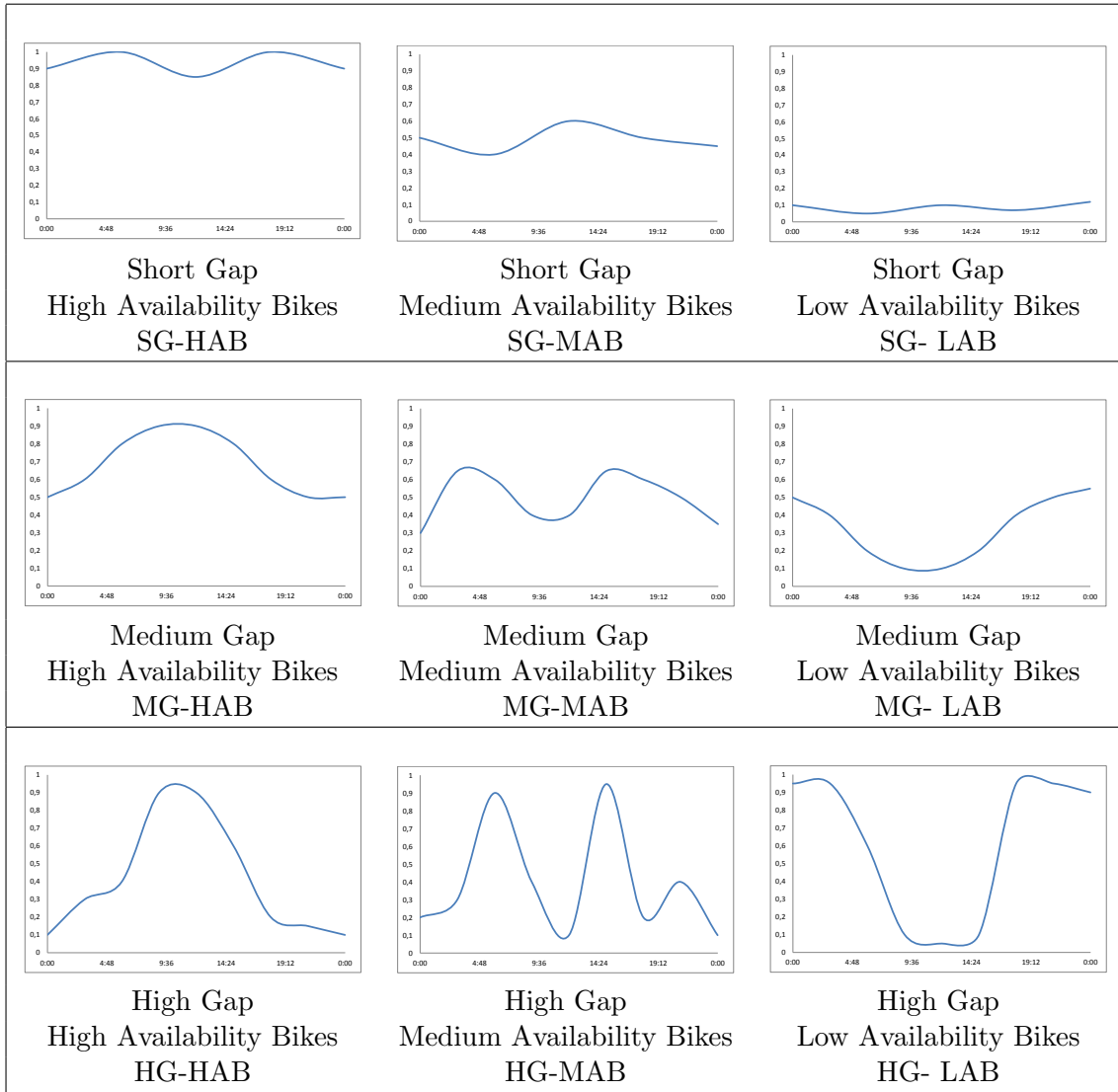
Furthermore, inside each previous group, other three clusters can be distinguished according to the amount of available bikes:

- **High Availability Bikes (HAB).** The most NAB values are near or equal to 1 during the day. A high number of bikes are available at the station.

- Medium Availability Bikes (MAB). The most NAB values vary around 0.5 during the day, that is a medium number of bikes are available at the station.
- Low Availability Bikes (LAB). The most NAB values are near or equal to 0 during the day. The common tendency of the station offers a low number of bikes.

Table 1 shows the identification of each cluster, its acronym, and the image of the pattern. The first three clusters (first row) represent the Small Gap group. They are HAB, MAB, and LAB based on the number of available bikes. In the same way the other clusters are represented. Note that the gaps are easily identified graphically.

Table 1: NAB ratio clusters.



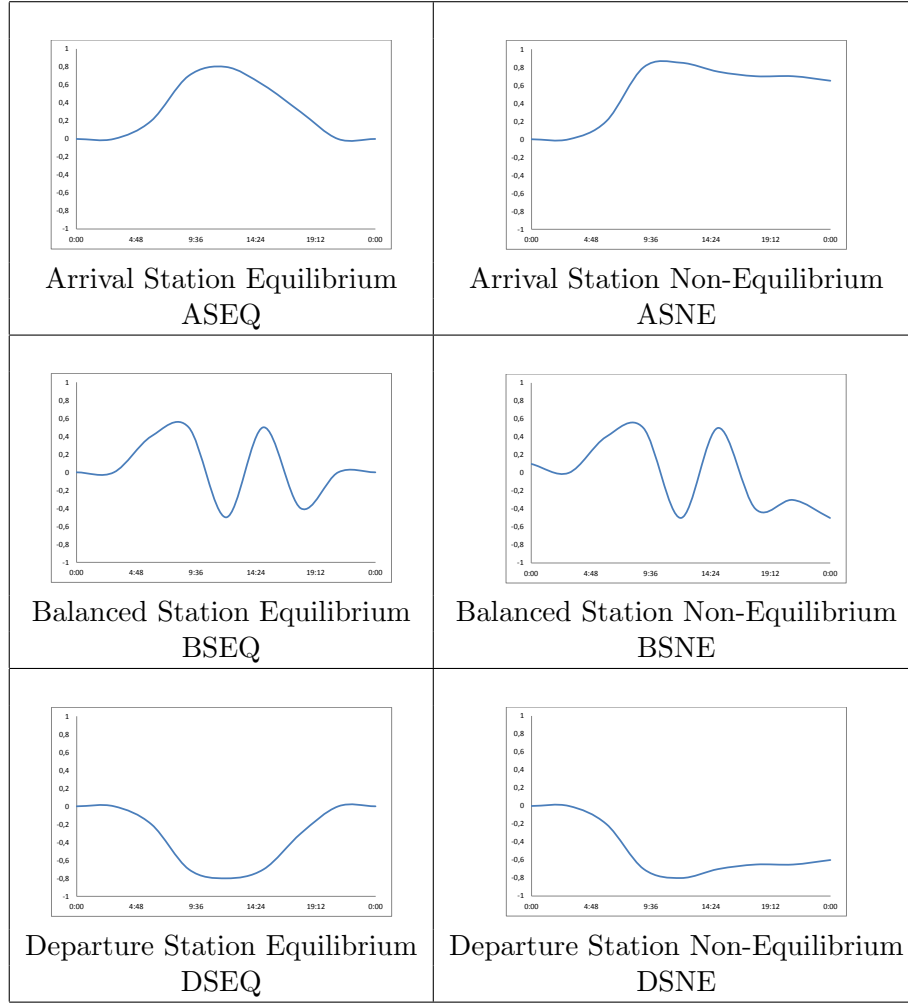
2.2.2 Cumulative Trip ratio clusters

Cumulative trip ratio has been split in six clusters (see Table 2). These clusters follow the well identified mobility patterns based on the attractive, generative or balanced characteristics of the station. Moreover, they can be distinguished between equilibrium or not equilibrium bike station, according to whether the end value of the day is equal or similar to the first value of the day. In other words, this classification adds a fast identification if the station is ready for the next day or, collecting or returning bicycles is necessary in order to meet station demand.

Table 2 shows the identification of each cluster: name, acronym and pattern layout for the cumulative trip ratio. The first column represents the equilibrium patterns and the non-balanced patterns are represented in the second column.

This classification could be a help to the reassignment bicycles problem associated with BSSs, thus if it could be reached that each station was balanced by itself, due to the equilibrium between arrivals and departures through a day. In such a case, the problem would become easier.

Table 2: Cumulative trip ratio clusters.



2.2.3 Turnover Station ratio clusters

In the case of turnover station ratio, five clusters are identified based on the number of times that the station capacity is exceed at the end of the day (see Table 3). If the turnover ratio is greater than 2.5 the station is considered too busy. This indicates that the station is well located, with a high demand. Maybe, in this case, an increase in the number of docks, or placing a new station near it, could improve the BSS. If turnover ratio varies between 1.5 and 2.5 the station is good. When the turnover ratio is between 0.5 and 1.5 the station is sufficient for the demand. However, if turnover ratio is under 0.5 the station could be miscalculated or, even the value is under 0.2, the station can be considered without use, therefore it can be thought about remove it.

Table 3: Turnover station ratio clusters.

High Turnover	$TS \geq 2.5$
Medium Turnover	$1.5 \leq TS < 2.5$
Normal Turnover	$0.5 \leq TS < 1.5$
Low Turnover	$0.2 < TS < 0.5$
Non Turnover	$TS \leq 0.2$

3 Description of the used methods and algorithm

Generally, bike sharing systems data provide station name, number of available bicycles and number of free slots every defined period of time (between 2 - 10 minutes) for each station. Thus, the algorithm is prepared to work with this type of base data.

Taking advantage of the advanced researches in the different fields of time series data mining (see Fu (2011) for a complete review), some of their tools to work with data will be used, in particular, the Perceptually Important Points (PIPs) process to represent and index each time series of each station. This process reduces the dimensionality of the original series but keeping the structure of the pattern. Next, a rule set to classify the data is applied.

The concept of Perceptually Important Points (PIPs) identification is based on the importance of data points. This importance is defined by the domination of a data point on the shape of the time series of whatever feature. A data point that has a greater domination on the overall shape of the series is considered to be more important. Let be a time series P with n data points: $P_1, P_2, \dots, P_n$. The PIP identification process consists on reordered data points by its importance in the time series. The first data point P_1 and the last data point P_n in the time series are the first and two PIPs respectively. The next PIP will be the point in P with maximum distance to the first two PIPs. The fourth found PIP will be the point in P with maximum vertical distance to the line joining its two adjacent PIPs, either in between the first and second PIPs or in between the second and the last PIPs. The PIP location process continues until the required number of PIPs is reached. Chung et al. (2001) were the first to introduced this identification process, and a detailed treatment can be found in Fu et al. (2008) for financial time series.

In the case of the research presented in this paper, only four PIPs are required to represent the ratios of the bike station data time series: the first and last point of the time series, and the maximum and minimum value on it (see Figure 4). Sometimes, the mean

value of the time series will be used to check that the cluster assignation is correct.

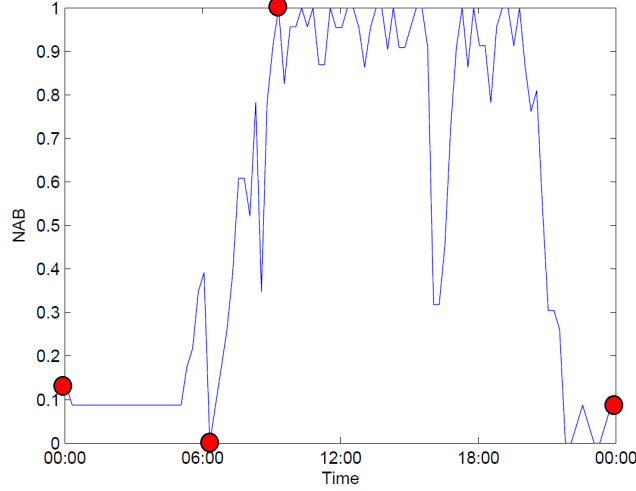


Figure 4: Time series compression by data point importance. In this case the pattern of bike station ratio time series is represented by four PIPs: First point, last point, maximum point and minimum point due to the well-known mobility pattern.

In relation to classification rules, note that they are a simplified version of the supervised decision tree algorithms. In this case, due to all prior information and the previous knowledge of the patterns cluster finding classification rules are easy. Therefore, a direct method with rules defined from training data of previous researches is applied. The formulation consists of a list of IF-THEN rules where IF part states condition over the data and THEN part includes a class label. Thus, the rules are of the form 'if A and B and C and ... then class X', where rules for each class are grouped together. A case is classified by finding the first rule whose conditions are satisfied by the case; if no rule is satisfied, the case is assigned to a default class.

Note that rules will be mutually exclusive, they will be independent of each other and every case will be covered at most one rule, that is, it will belong to one class. The base of each rule will be attribute-value comparisons.

3.1 Algorithm

This section proposed an algorithm to assign each station to the corresponding cluster of each parameter (NAB, CumT and TS) in order to identify new patterns from mixing previous clusters. Figure 5 shows a brief scheme of the algorithm.

Algorithm 1 (Bike station classification based on joining NAB, CumT and TS ratios)

INPUT: *Data set for the period of study of the bike sharing system analysed. Data set contains code station, number of bikes at the station, number of free slots and time stamp.*

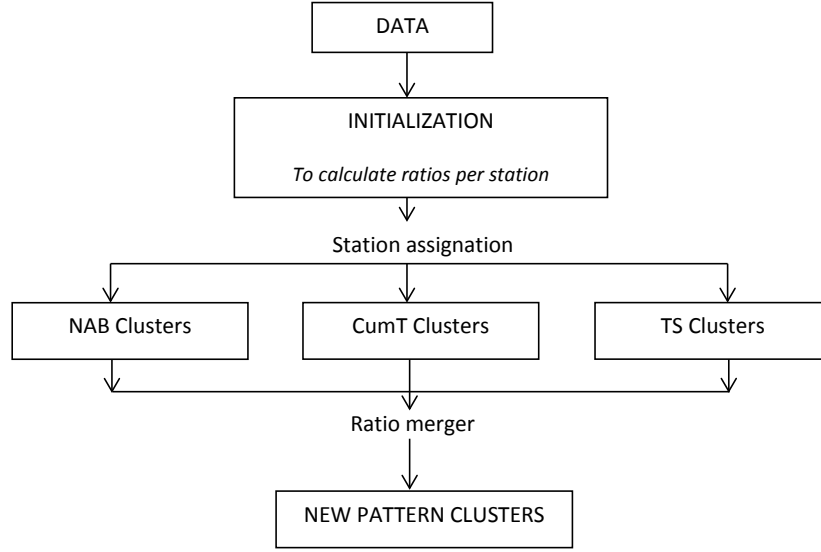


Figure 5: Algorithm scheme.

OUTPUT: *Four matrices:*

- **N** matrix with the new patterns identified to join the three ratios.
- **M** matrix where each row is a new cluster and the columns are filled with the stations that belong to each cluster.
- **M1** submatrix of the main result matrix (**M**) where the patterns with more stations are saved.
- **M2** submatrix of the main result matrix (**M**) where the station with some of the no desirable patterns are saved.

Note that, for CumT ratio all non-equilibrium clusters are considered no desirable because the number of bicycles or slots needs to be reassigned at the end of the day, and for turnover station ratio, their last clusters are not desirable in a BSS due to these station have small or null rotation.

Step 1: Initialization. *Select the weekday to represent. Split data to each station and calculate the three ratios for each one. From this step three matrices are obtained, that is, one for each ratio: NAB1, CT1, TS1. Each row of the matrices represents a station, and the columns indicate the ratio for each considered time stamp for the NAB and CumT ratio. In the case of turnover station ratio only two columns are needed to add the arrival and departure turnover ratio for each station.*

Step 2: Station assigation to NAB clusters.

Step 2.1: Perceptually important points identification. *The four PIPs are calculated:*

- **A:** *First point of the time series*

- *B: Maximum point of the time series*
- *C: Minimum point of the time series*
- *D: Last point of the time series*

Step 2.2: Application the rules to classify data. *Classification rules are applied to assign each station a cluster. The first rule assigns the times series based on the gap between maximum and minimum values of them. Note that, for this ratio the gap values for cluster can be lower than 0.3, between 0.3 and 0.7 and, greater than 0.7. Then, a set of if-then rules according to the tendency of the time series assigns the station to a final cluster.*

Step 2.3: Testing coverage. *Check if all stations are assigned a cluster, if not, save this station in a default cluster.*

Step 2.4: Middle results. *A matrix, P1, with the stations assigned per clusters is obtained.*

Step 3: Station assignation to CumT clusters.

Step 3.1: Perceptually important points identification. *Like Step 2.1 the four PIPs are calculated together with the mean value of the time series (E).*

Step 3.2: Application the rules to classify data. *Classification rules are applied to assign each station a cluster. The first rule assigns the times series based on the equilibrium of the time series, i.e., if the values of the first and last point are similar or not. Then, a set of if-then rules according to the tendency of the time series assigns the station a final cluster. Note that, the mean value of the time series will make sure if the time series is a generative station (negative values) or an attractive station (positive values).*

Step 3.3: Testing coverage. *Check if all stations are assigned a cluster, if not, save this station in a default cluster.*

Step 3.4: Middle results. *A matrix, P2, with the stations assigned per clusters is obtained.*

Step 4: Station assignation to Turnover Station clusters.

Step 4.1: Application the rules to classify data. *Classification rules are applied to assign each station a cluster. A set of if-then rules according to the usage limits is evaluated to assign the station a final cluster. Remember that for this ratio the limits are: 2.5, 1.5, 0.5 and 0.2.*

Step 4.2: Testing coverage. *Check if all stations are assigned a cluster, if not, save this station in a default cluster.*

Step 4.3: Middle results. *A matrix, P3, with the stations assigned per clusters is obtained.*

Step 5: New pattern are identified. *The three result matrix (P1, P2 and P3) of previous steps are joined to identify the new patterns which are composed of three elements each one for each ratio. The output matrices **N**, **M**, **M1** and **M2** are obtained.*

4 Examples of application

In this section the proposed method is applied to Bike Sharing System of Dublin. The algorithm performance is very good, providing very fast results and without numerical problems. All station were assigned a cluster.

4.1 Dublin Bikes

Dublin Bikes is a public bicycle rental scheme which has operated in the city of Dublin since 2009. At its launch, the scheme, which is sponsored by JCDecaux, used 450 bicycles with 40 stations. Dublin Bikes reached a landmark one million trips on the 14th of August 2010 and on the 12th of May 2011 it reached its two millionth trip. Such was the success of the scheme that 4 new stations were added to the network and an extra 100 bikes by 2011. In 2013 a major expansion scheme was announced and to date, the scheme counts with 101 stations and 1500 bikes, over 54,000 active long-term members and over 10 million journeys accumulated since its launch (www.irishcycle.com). All these figures makes Dublin Bikes system one of the most successful bike sharing schemes in the world.

4.2 Dataset description

The fundamental datum that J.C.Decaux makes available is the number of bikes docked at each station at a given point in time. The results are based on the data collected during the 3 weeks between 00.00 h, March 5th and 00.00 h, March 25th, 2015. Data from 101 stations during this time each 10 minutes, approximately, are collected, that is, about 200,000 registers. The capacity of the station varies between 40 and 16 slots (see Figure 6).

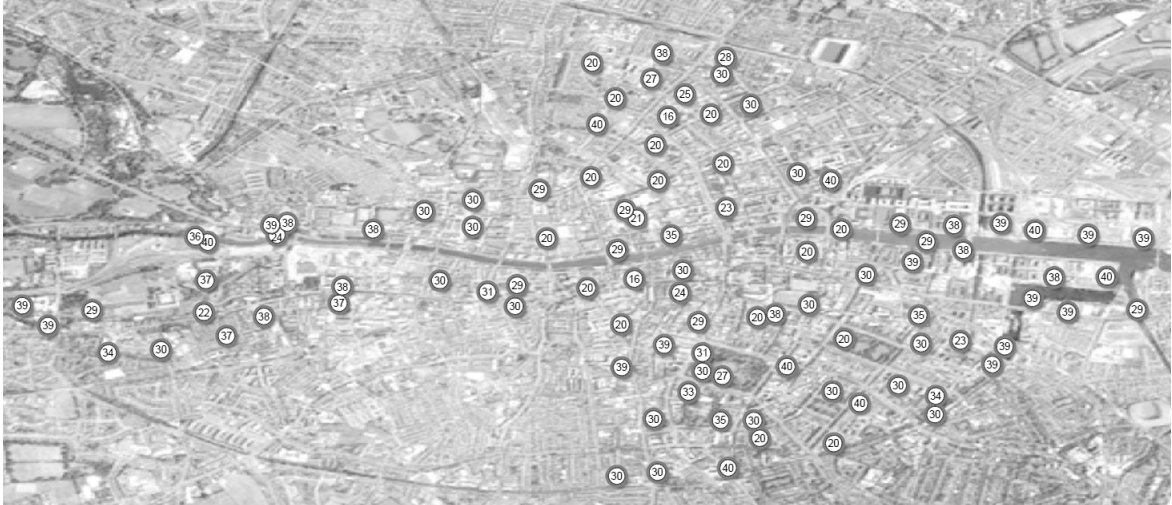


Figure 6: Dublin Bikes. Station distribution in the core of the city indicating its capacity (dock number).

The research is focused on a whole week, but to reduce the paper's length only the results of a workday (Thursday) are described, because, as the turnover station ratio and some other papers show, for example Borgnat et al. (2011), Nash (2011) or Kaltenbrunner et al. (2010) among others, weekdays show more use than weekend days, in which the use of public bikes are only concentrated in the afternoon.

4.3 Results analysis

Once the algorithm is applied to the data the results are shown graphically. Specifically, the intermediate matrix for each ratio: NAB, CumT and TS, and the final matrices **M1** and **M2** with the main identified patterns and the stations with no desirable patterns. Note that, all stations are assigned, there are not outliers or a default class, so the rules are exclusive and exhaustive.

As Figure 7 shows the first identified cluster of NAB ratio (Short Gap: high and medium availability bikes) does not exist in this scheme. The numerous clusters are High Availability and Low Availability Bikes for High Gap (HG-HAB, HG-LAB), that is the stations have a big difference of the number of available bikes throughout the day, and corresponding to the generative and attractive clusters. The clusters with pickups morning are situated in the scheme outline, whereas the return bikes at morning and pickups at afternoon clusters are in the centre of the scheme where banks, university and business are placed.

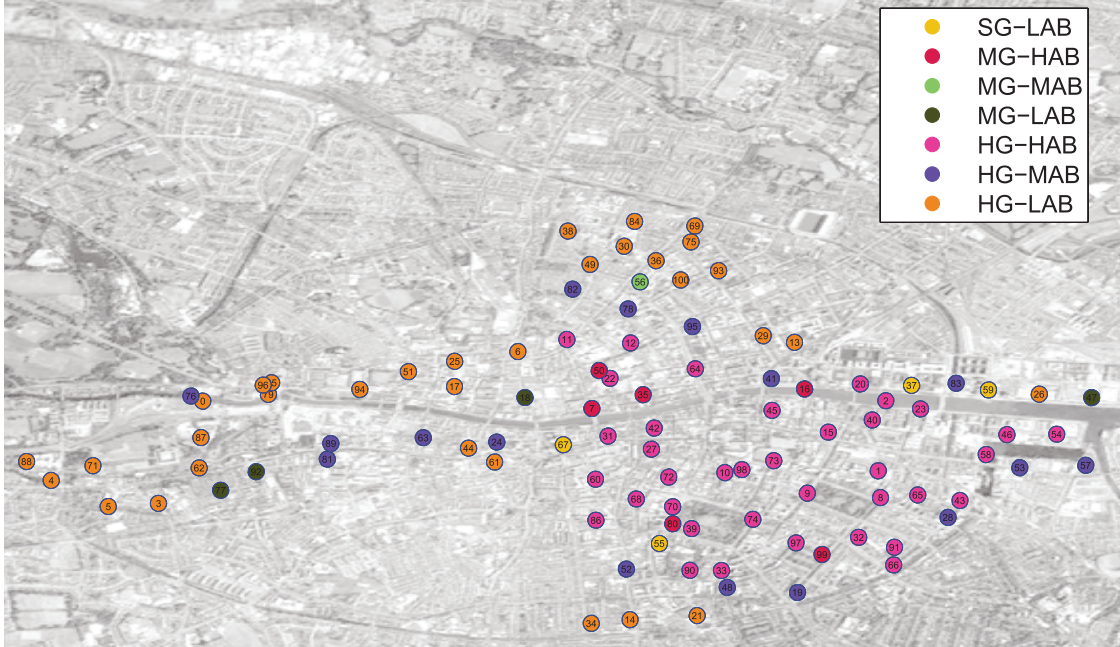


Figure 7: Dublin Bikes. Stations by number of available bicycles clusters.

Note that Dublin is a very extended city with very centralized services and that the BSS of Dublin is a core scheme; their bike stations are located in the city centre, as Figure 8 shows. In this area all buses lines go through and users have also access to train and tram.

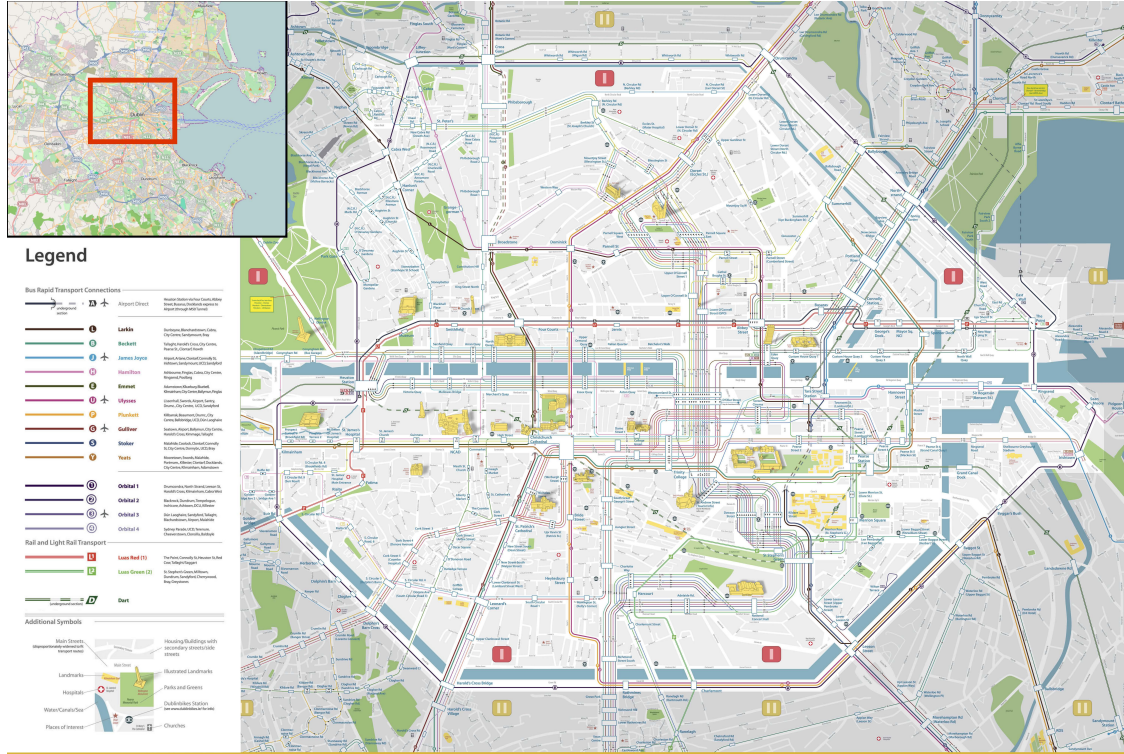


Figure 8: Urban structure and main line of public transport of Dublin city centre. Source: Aris Venetikidis.

Moreover, the city centre is the top tourist area (some of the most important building are shown in yellow in Figure 8), besides having commercial, cultural and financial service. Therefore, as the bike station classification confirms, intermodality works very good in the city centre. On one hand, many commuters arrive in the city core by bus, tram or train, and then they go to their work or to the university (Trinity College Dublin is there) on foot or by bicycle, and maybe they run some errands by bike. On the other hand, bicycles are used by tourists for moving through the city for sightseeing. Thus, the urban structure and coordination among public transport networks contribute to the system's good performance.

Figure 9 shows the results for the cumulative trip ratio. Similar results as NAB ratio are shown, that is the attractive balanced stations (blue) are located in the centre whereas the generative balanced stations (green) are in the edge of the scheme. It is easy to figure out the flow of bikes due the well balanced number of station of each complementary cluster.

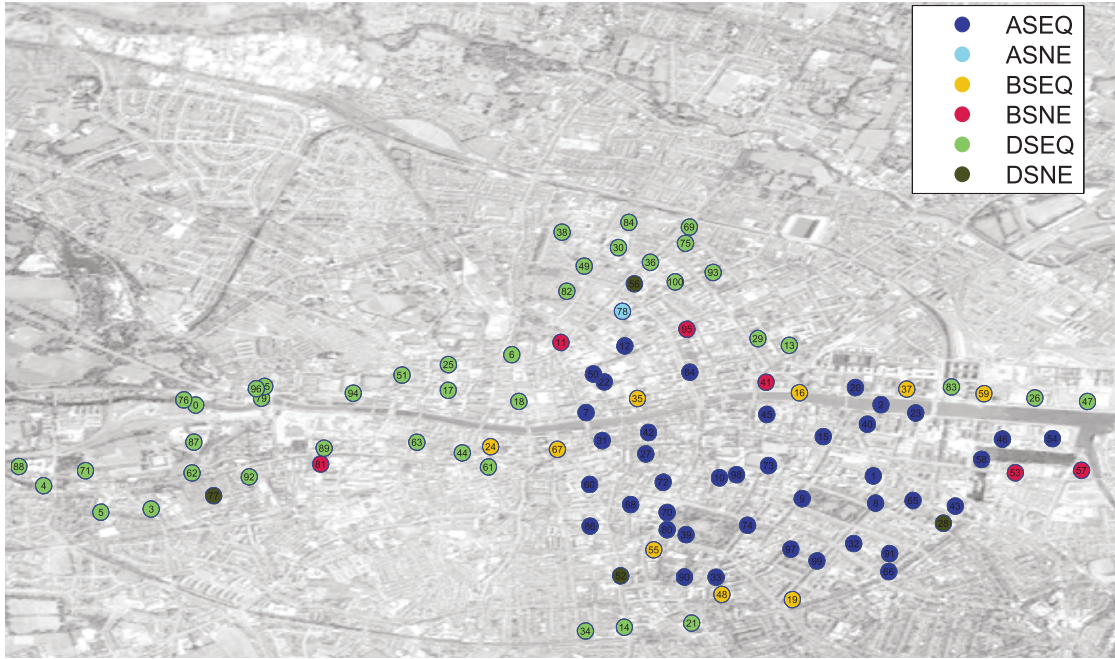


Figure 9: Dublin Bikes. Stations by cumulative trip clusters.

In relation to turnover station ratio, results are shown in Figure 10. Only a station has a low turnover rate for this weekday, but the 75% of stations are over a normal use, being the 22% with a high turnover. This points out the elevated use of the systems and its good performance.

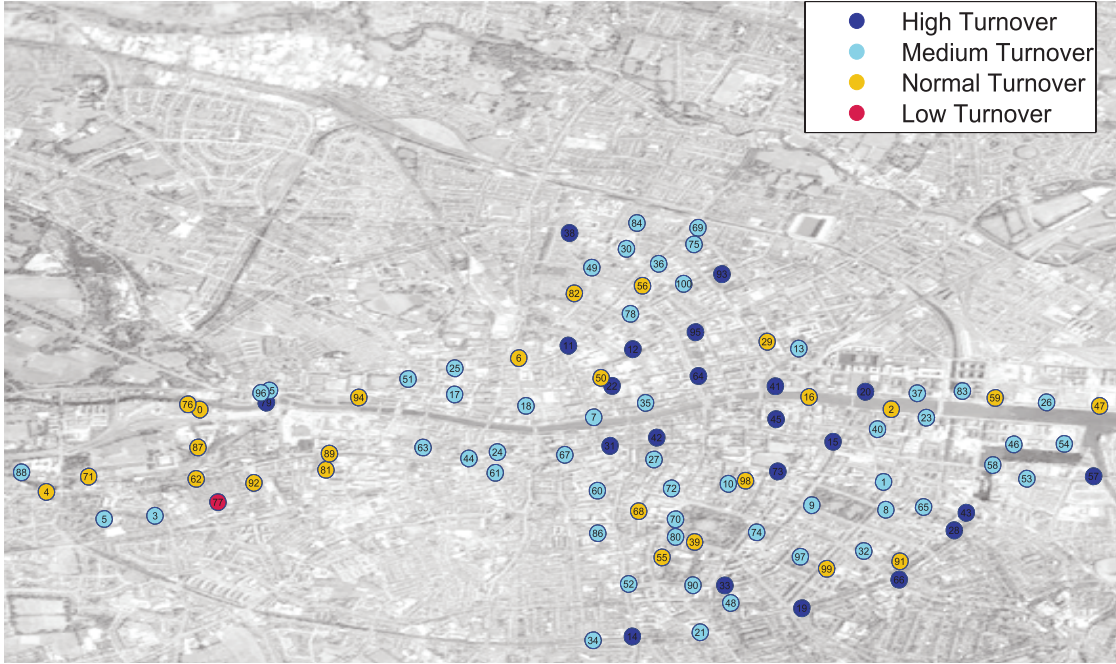


Figure 10: Dublin Bikes. Stations by turnover ratio clusters.

Once the three clusters are identified, the last algorithm step joins them to have a complete identification of the station. 27 combinations are feasible as Table 4 provides, where the non-desirable patterns are marked in bold. Note that, obviously the generative and attractive patterns of both ratios NAB and CumT are together and the key is the turnover station ratio, which allow the knowledge of the level of use of the station.

Table 4: New clusters after joining NAB, CumT and TS ratios. The number of stations in each new cluster is provided. The no desirable patterns are marked in bold.

Mixed Pattern	# of stations
SG-LAB + BSEQ + Medium Turnover	2
SG-LAB + BSEQ + Normal Turnover	2
MG-HAB + ASEQ + Medium Turnover	2
MG-HAB + ASEQ + Normal Turnover	2
MG-HAB + BSEQ + Medium Turnover	1
MG-HAB + BSEQ + Normal Turnover	1
MG-MAB + DSNE + Normal Turnover	1
MG-LAB + DSEQ + Medium Turnover	1
MG-LAB + DSEQ + Normal Turnover	2
MG-LAB + DSNE + Low Turnover	1
HG-HAB + ASEQ + High Turnover	12
HG-HAB + ASEQ + Medium Turnover	19
HG-HAB + ASEQ + Normal Turnover	5
HG-HAB + BSNE + High Turnover	1
HG-MAB + ASNE + Medium Turnover	1
HG-MAB + BSEQ + High Turnover	1
HG-MAB + BSEQ + Medium Turnover	2
HG-MAB + BSNE + High Turnover	3
HG-MAB + BSNE + Medium Turnover	1
HG-MAB + BSNE + Normal Turnover	1
HG-MAB + DSEQ + Medium Turnover	2
HG-MAB + DSEQ + Normal Turnover	3
HG-MAB + DSNE + High Turnover	1
HG-MAB + DSNE + Medium Turnover	1
HG-LAB + DSEQ + High Turnover	4
HG-LAB + DSEQ + Medium Turnover	21
HG-LAB + DSEQ + Normal Turnover	8

In Figure 11 the new main patterns are represented. There are three attractive patterns and three generative patterns, that is this system is well balanced and the mobility flows are easy to identify. On the other hand, Figure 12 shows the stations that a deep study to improve them will be recommended (their corresponding patterns are boldfaced in Table 4). The stations with a non balanced pattern are recommended to try to balance for themselves according to the bikes flows, and those stations with a low or null turnover ratio. In this case, the station located in St. James Hospital (Central) is the one with a low turnover station ratio and non balanced departure pattern. This may be possible due to the special characteristics of the health service.

To end this example, the stations of the main six new clusters are analysed in relation to its distance to the Luas² and train stations. However, there are not found significant results due to all station are an average distance of 500 metres and 700 metres to the

²Luas is the name of the tram network of Dublin.

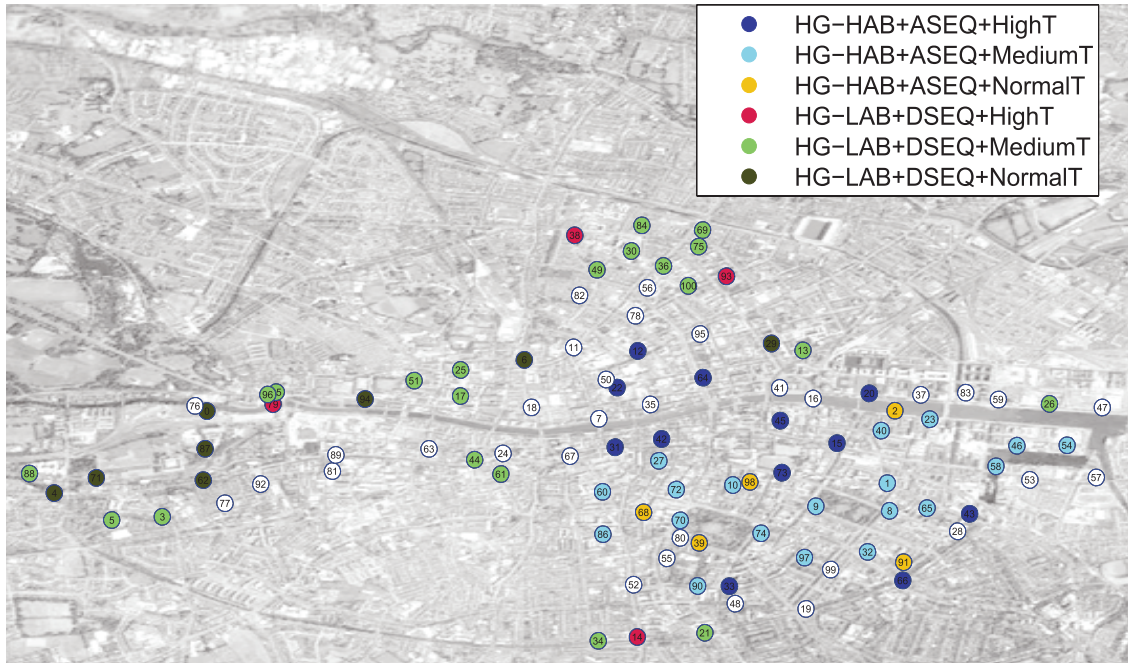


Figure 11: Dublin Bikes. Stations by main combinations of clusters.

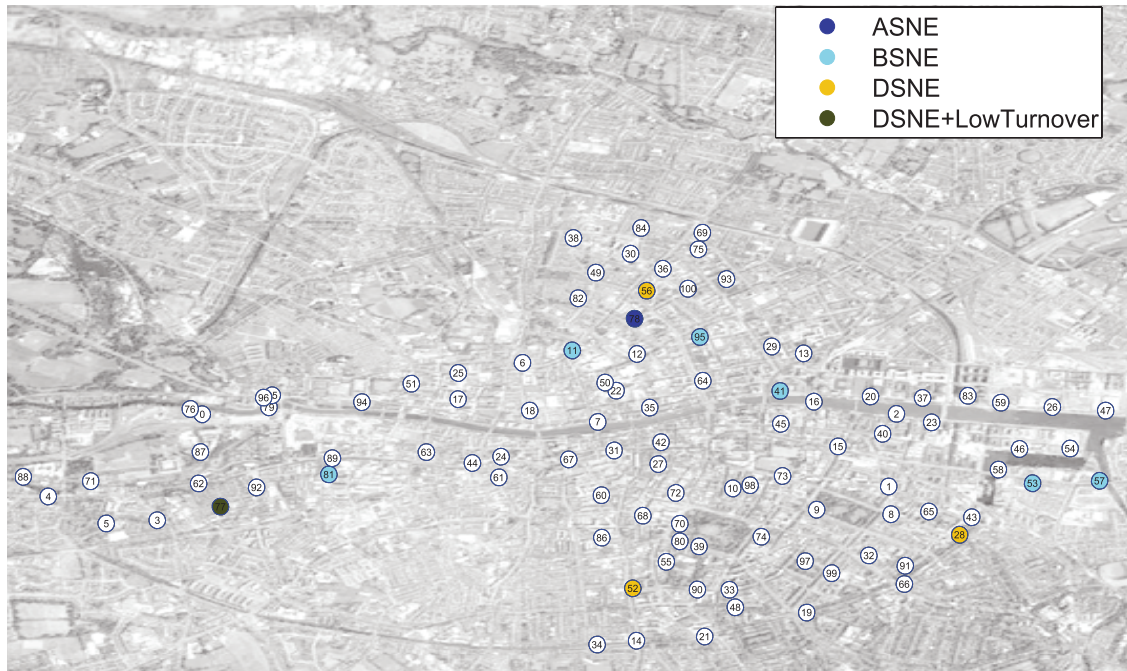


Figure 12: Dublin Bikes. Stations with some no desirable pattern.

Luas and train stations, respectively. This confirms the previous comment about the connection easiness among the different modes of transport that the city supplies. On the other hand, according to the Census 2011 of Ireland and Northern Ireland, and after located each station in the corresponding SAPS (Small Area Population Statistics) of Dublin, the relations among the main stations and their SAPS Data have been analysed, however as Dublin is a small core city only the 6% of the inhabitants of these areas use bicycles to move around the city. As Figure 13 shows on foot is the main mode to move around the city. This points out that ad-hoc surveys are needed to the users public bikes systems to know the detail of their movements as the work of O'Neill and Caulfield (2012) and Murphy and Usher (2015) have carried out.

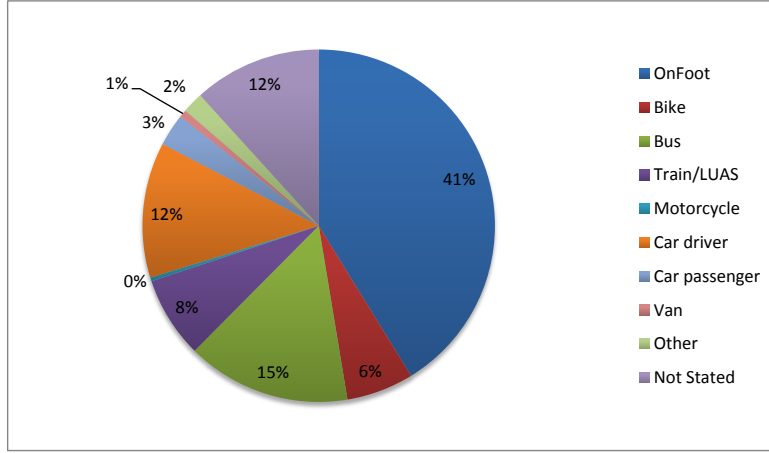


Figure 13: Modal split of the inhabitants in the core of Dublin. Source: Census 2011 of Ireland and Northern Ireland.

5 Conclusions

The main conclusions that can be drawn from this paper are the following:

1. In order to increase the knowledge about bike sharing station and to know if the service works correctly, a new ratio, the Turnover Station ratio, is presented. This ratio assesses how many times the station capacity is used in a complete day.
2. This new ratio together with the most useful ratios, Number of Available Bikes and Cumulative Trips, allow characterisation of the station in detail. Due to this union it is easy to identify the generative, balanced and attractive stations. Moreover, the Turnover Station ratio shows the effectiveness of the station.
3. The results could support planning and operations decisions for re-design and management of bike sharing systems, due to the classification reveals imbalances in the distribution of bikes and lead to a better understanding of the system structure in relation to services offered by the city. In other words, results confirm which stations work correctly, or which need a deep study to improve its resources (bikes and docks) and to avoid the detected imbalanced or low use.

4. Mathematical tools of time series data mining have been used. Taking advantages of the various researches carried out about the temporal pattern discovery and clustering of bike sharing schemes, a supervised learning algorithm has been applied. That is, clusters presented in previous works, training data and examples have been used. The algorithm only has to classify the input data. Therefore, instead of using cluster analysis, the perceptually important points (PIP) process to represent and index each time series of each station, and a rule set to classify the data, have been used.
5. The algorithm has been applied to the Dublin Bikes Scheme providing very fast results, without numerical problems and good performance due to all stations have been assigned the correct cluster. Moreover, the results show the good performance of Dublin Bikes system due to the 75% of the stations have a turnover ratio greater than 1.5 the station capacity, that is a high use, and the attractor and generator station are well balanced and distributed through the city.

In relation to future researches, it would be interesting to collect data from many cities around Europe to carry out a comparative study of bike sharing systems according to the results of the described algorithm. These results would allow the comparison of the new clusters and the performance of other bikesharing systems with a different spatial distribution. On the other hand, to improve these results together with users' surveys will be another field to explore in order to obtain more details about the users' decisions and try to understand the relationships among personal decisions about the choice of mode of transport and performance and management of the BSS. Finally, the seasonality study of the turnover station ratio in several BSSs would be another idea to develop in a future thus, the results can be generate interesting relationships among the use of the bicycle and weather, orography or accessibility to users among others.

6 Acknowledgements

The authors acknowledge Erasmus+ programme, specifically the Staff Mobility for Training - STT - 2014/2015 - for partial support of this work.

References

- Borgnat, P., Abry, P., Flandrin, P., Robardet, C., Rouquier, J., and Fleury, E. (2011). Shared bicycles in a city: a signal processing and data analysis perspective. *Advances in Complex Systems*, 14:3:415–438.
- Chabchoub, Y. and Fricker, C. (2014). Classification of the vélib stations usign kmeans, dynamic time wrapping and dba averaging method. *International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*., pages 1–5.
- Chung, F., Fu, T., Luk, R., and Ng, V. (2001). Flexible time series pattern matching based on perceptually important points. *International Joint Conference on Artificial Intelligence Workshop on Learning from Temporal and Spatial Data*, pages 1–7.

- Fröehlich, J., Neumann, J., and Oliver, N. (2009). Sensing and predicting the pulse of the city through shared bicycling. *Proceedings of the 21st International Joint Conference on Artificial Intelligence. Pasadena. California.*, pages 1420–1426.
- Fu, T. C. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24:164–181.
- Fu, T. C., Chung, F. L., Luk, R., and Ng, C. M. (2008). Representing financial time series based on data point importance. *Engineering Applications of Artificial Intelligence*, 21:2:277–300.
- García-Palomares, J. C., Gutiérrez, J., and Latorre, M. (2012). Optimizing the location of stations in bike-sharing programs: a gis approach. *Applied Geography*, 35:235–246.
- Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., and Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6:455–466.
- Lathia, N., Ahmed, S., and Capra, L. (2012). Measuring the impact of opening the london shared bicycles scheme to casual users. *Transportation Research C*, 22:88–102.
- Liao, T. W. (2005). Clustering of time series data-a survey. *Pattern Recognition*, 38:1857–1874.
- MacQueen, J. (1967). Some method for classification and analysis of multivariate observations. *5th Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297.
- Murphy, E. and Usher, J. (2015). The role of bicycle-sharing in the city: Analysis of the irish experience. *International Journal of Sustainable Transportation*, 9:2:116–125.
- Nash, O. (2011). Dublin bikes revisited. Accessed online:<http://olivernash.org/2011/02/02/dublin-bikes-revisited/>.
- O’Brien, O., Cheshire, J., and Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, 34:262–273.
- O’Neill, P. and Caulfield, B. (2012). Examining user behaviour on a shared bike scheme: the case of dublin bikes. *13th International Conference on Travel Behaviour Research*.
- Vogel, P., Greiser, T., and Mattfeld, D. C. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia Social and Behavioral Sciences*, 20:514–523.
- Zhao, J., Deng, W., and Song, Y. (2014). Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in china. *Transport Policy*, 35:253–264.

PERCEPTUALLY IMPORTANT POINTS OF MOBILITY PATTERNS
TO CHARACTERISE BIKE SHARING SYSTEMS:
THE DUBLIN CASE.

Pilar Jiménez

Department of Civil Engineering, Technical University of Cartagena, 30203 Cartagena, Spain

E-mail address: pilar.jimenez@upct.es

María Nogal

*Department of Civil, Structural and Environmental Engineering, Trinity College Dublin,
Dublin 2, Ireland*

E-mail address: nogalm@tcd.ie

Brian Caulfield

*Department of Civil, Structural and Environmental Engineering, Trinity College Dublin,
Dublin 2, Ireland*

E-mail address: brian.caulfield@tcd.ie

Francesco Pilla

*Department of Civil, Structural and Environmental Engineering, Trinity College Dublin,
Dublin 2, Ireland*

E-mail address: fpilla@tcd.ie

Corresponding author:

Pilar Jiménez

Associate Professor at Department Civil Engineering

Technical University of Cartagena

30203, Paseo Alfonso XIII, 52 Cartagena, Spain

E-mail:pilar.jimenez@upct.es

Tel.: +34 868 07 12 76

*Highlights

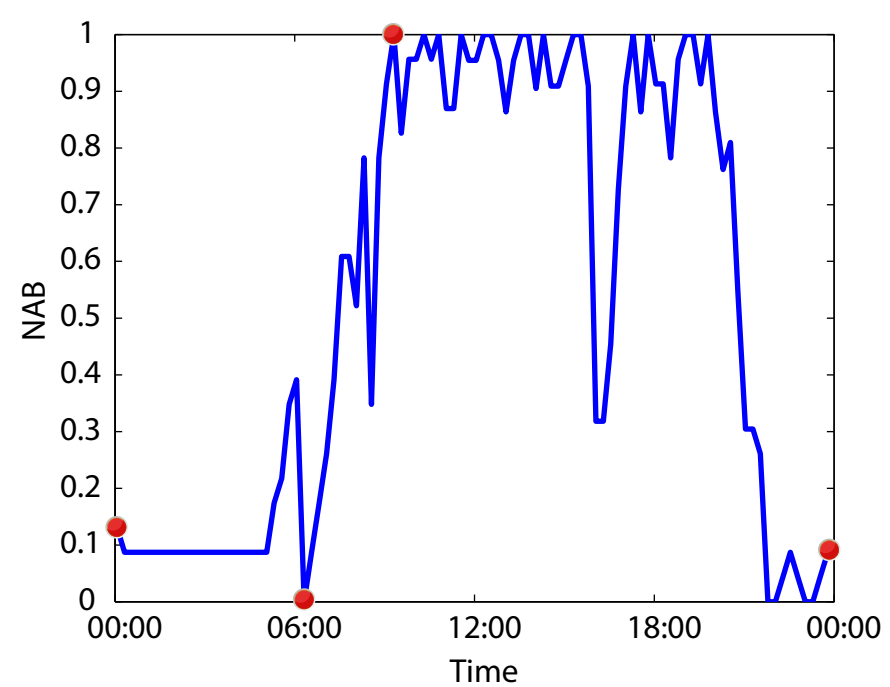
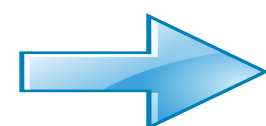
- A bike station classification based on users' mobility patterns is proposed
- 3 ratios are used: number of available bikes, cumulative trips and turnover station
- Turnover station ratio measures the effectiveness degree of each station
- Data mining tools are used: perceptually important points and rule sets
- Results support planning and operations decisions for management of BSSs

DATA
Mobility patterns for each station

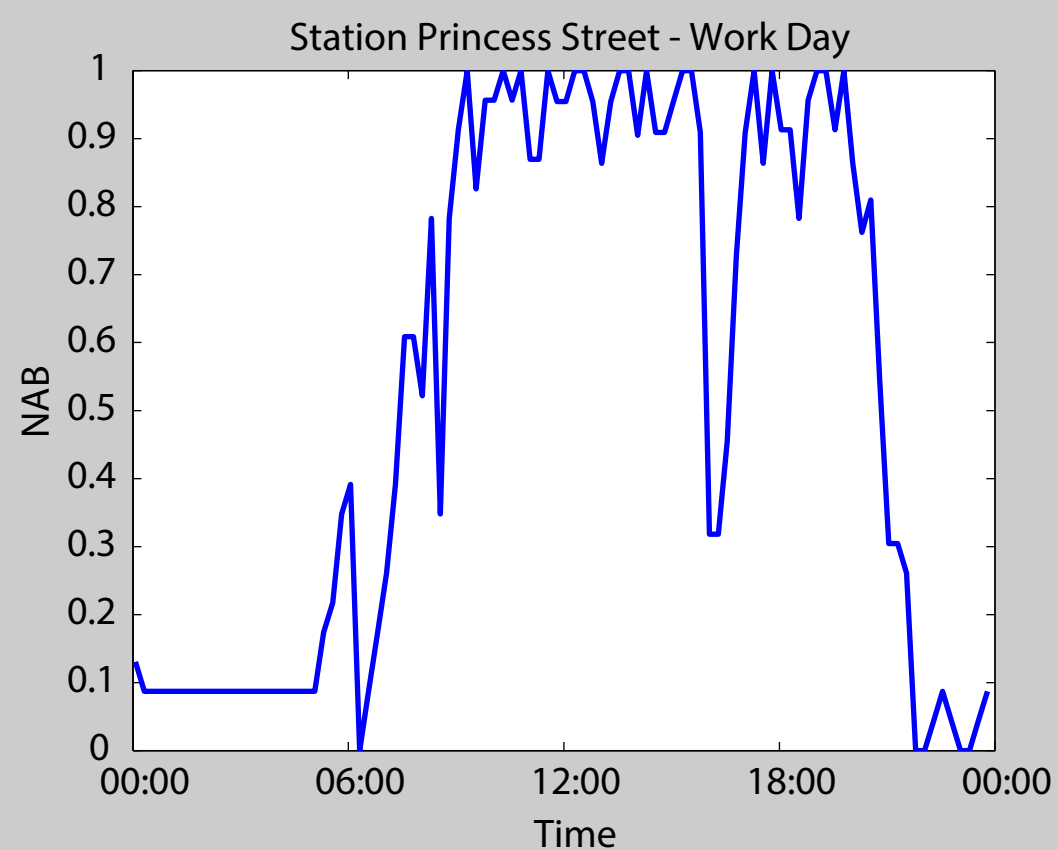


INITIALIZATION
PERCEPTUALLY IMPORTANT POINTS
To calculate ratios per station

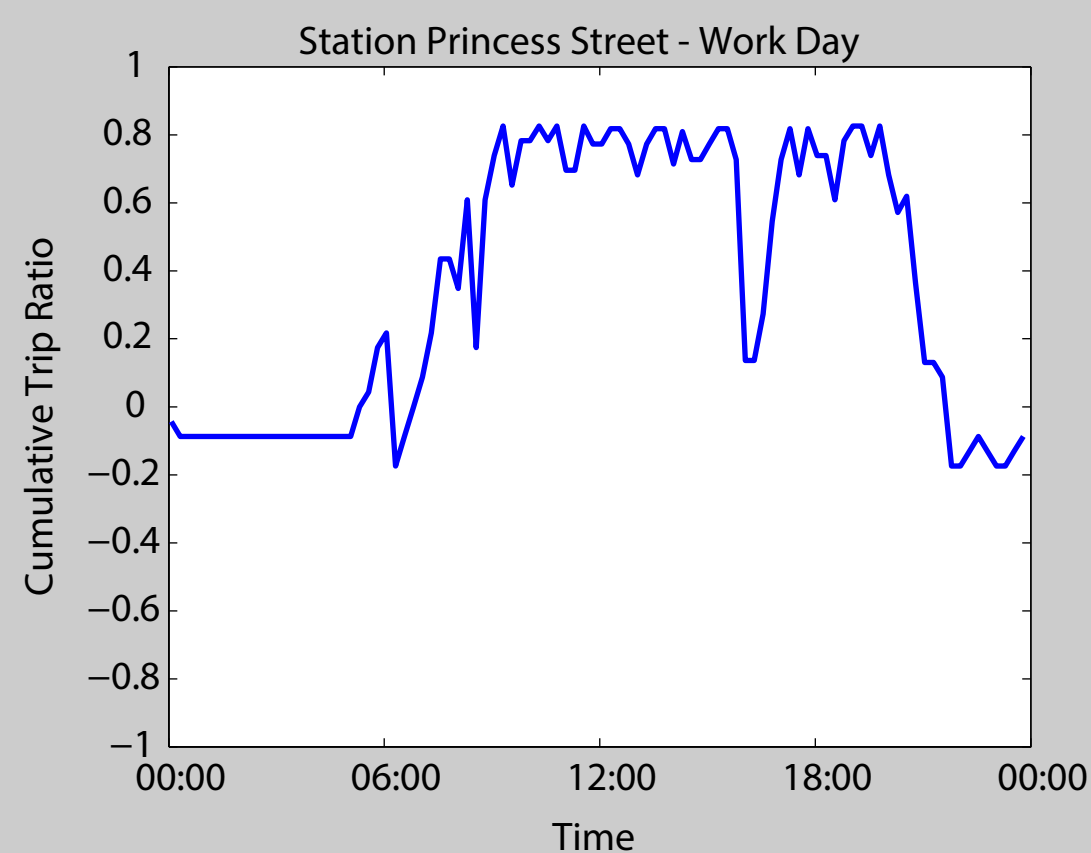
Assignment



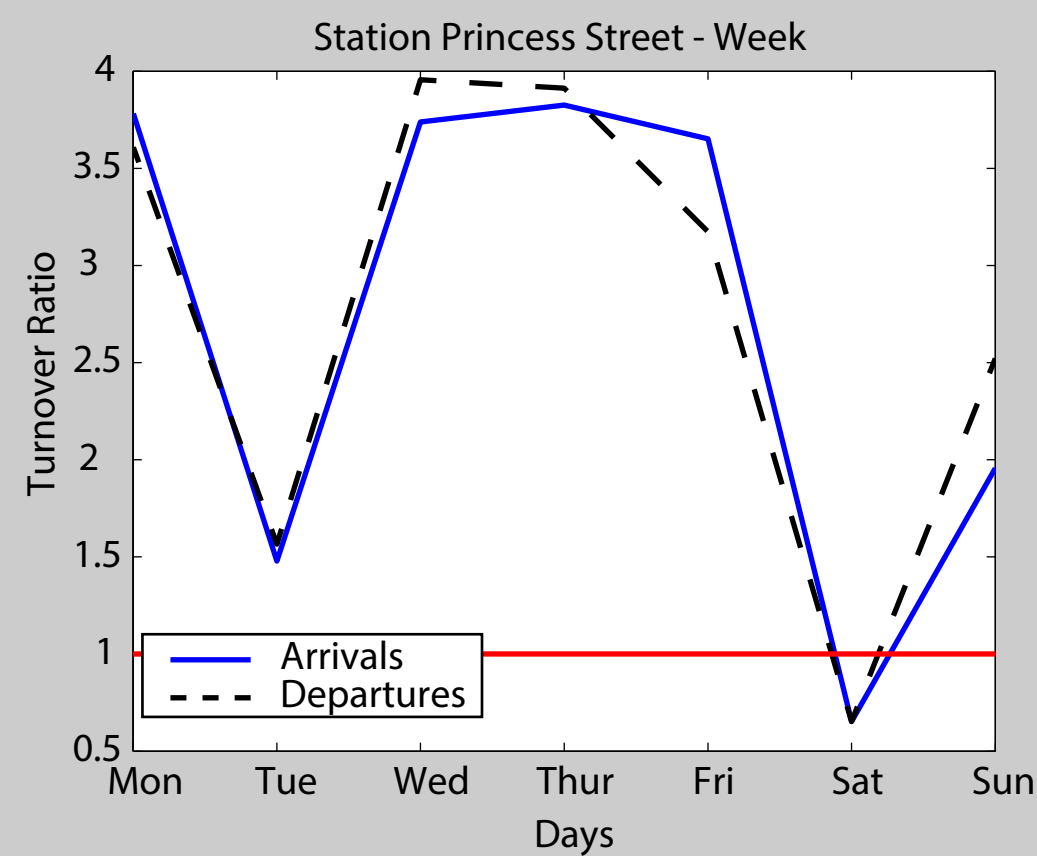
NAB CLUSTERS



CumT CLUSTERS



TS CLUSTERS



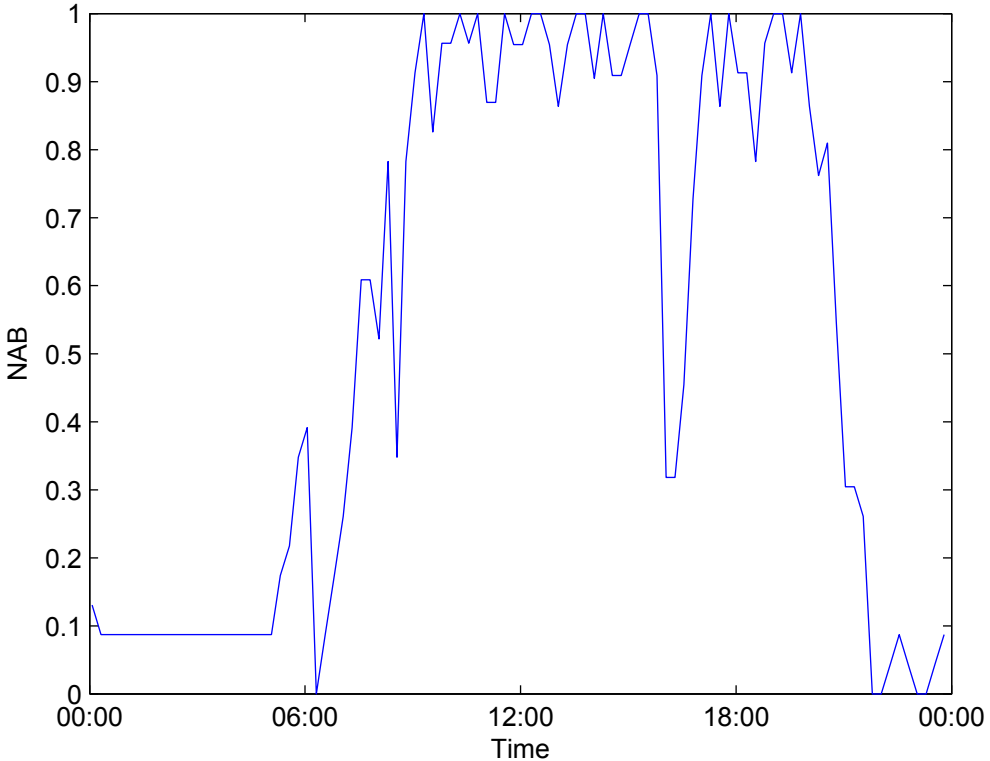
Ratio Merger



NEW PATTERN CLUSTERS
To classify bike station

Figure

Station Princess Street - Work Day

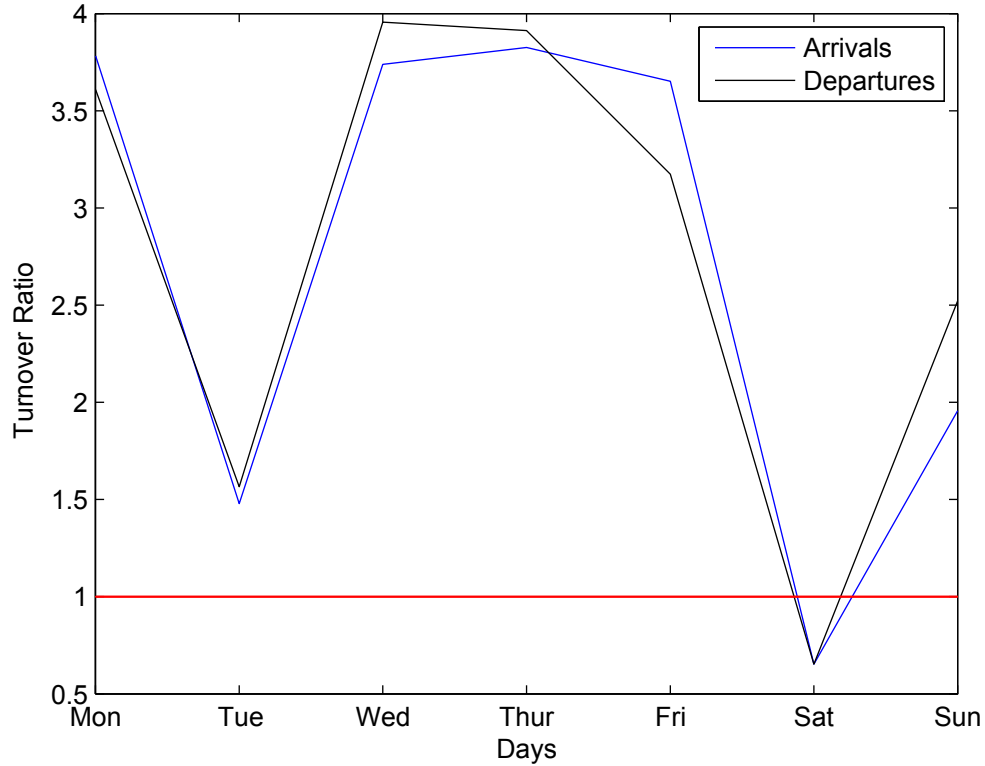


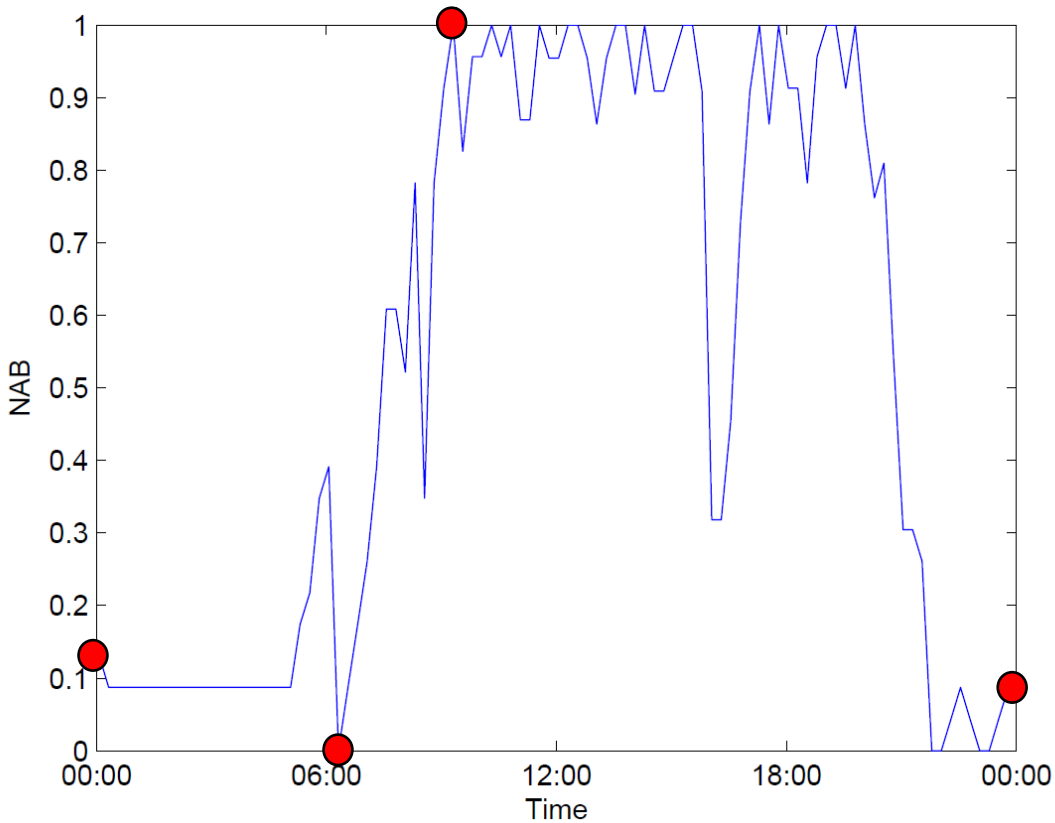
Figure



Figure

Station Princess Street - Week





DATA

INITIALIZATION

To calculate ratios per station

Station assignation

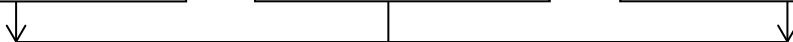
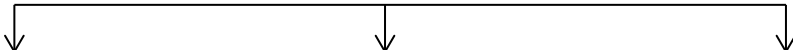
NAB Clusters

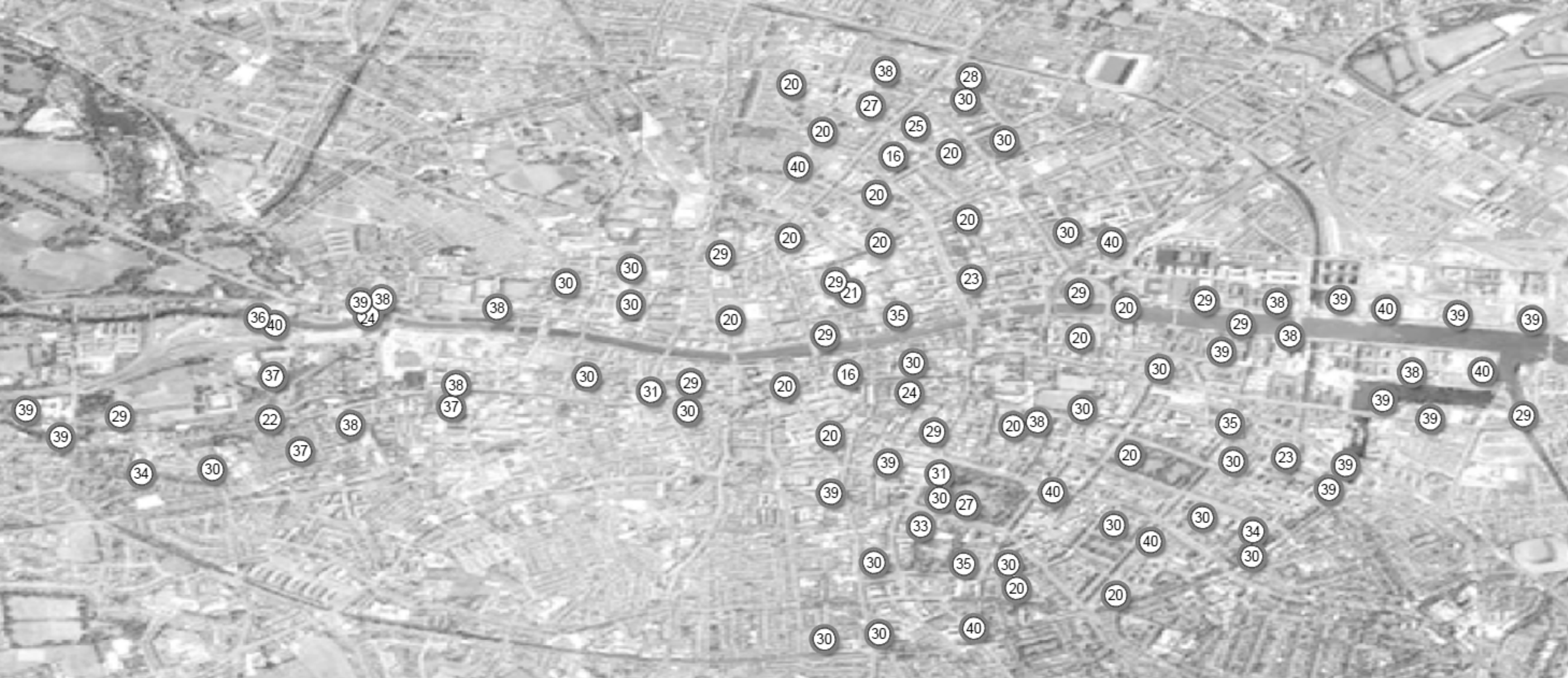
CumT Clusters

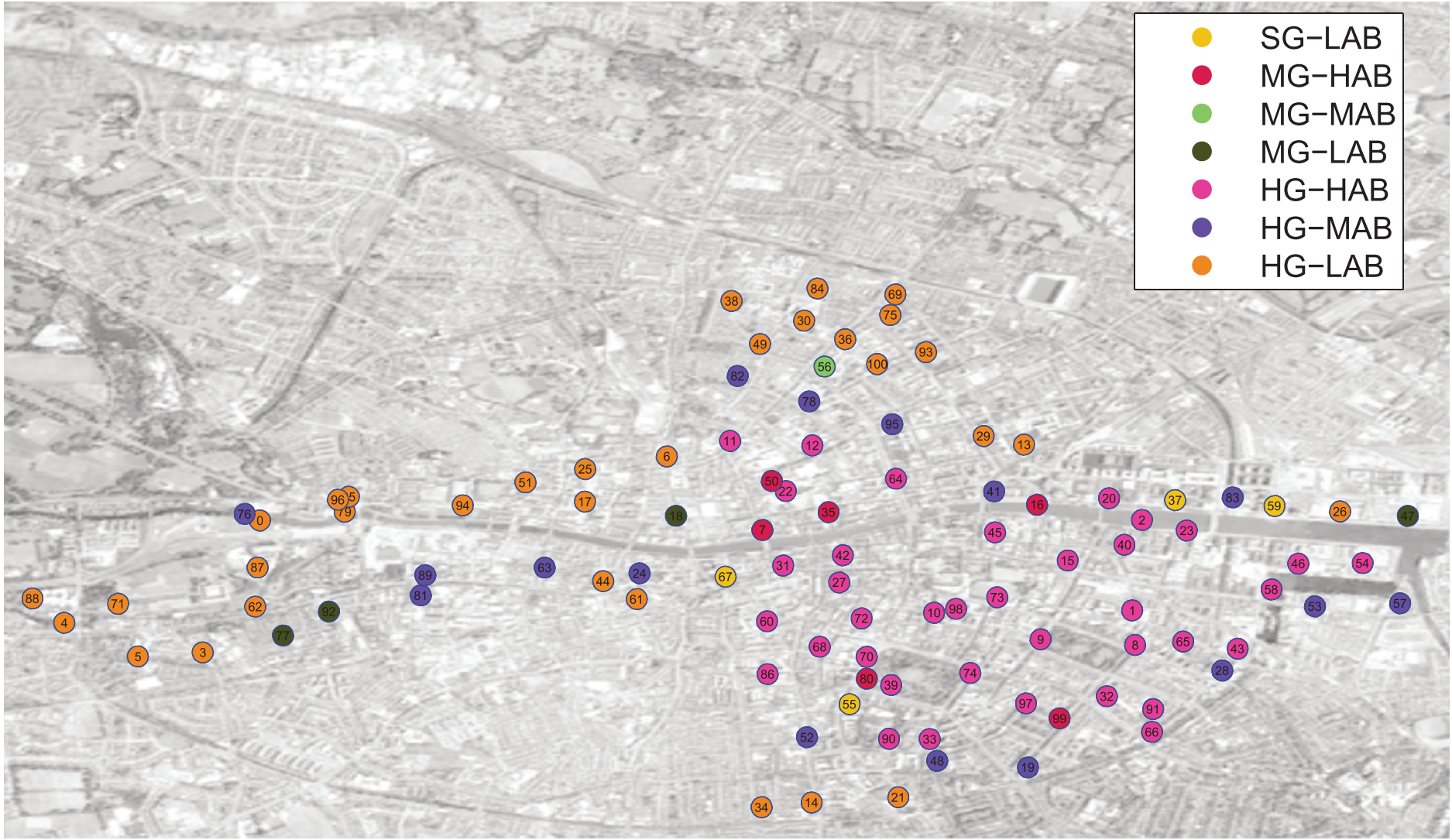
TS Clusters

Ratio merger

NEW PATTERN CLUSTERS

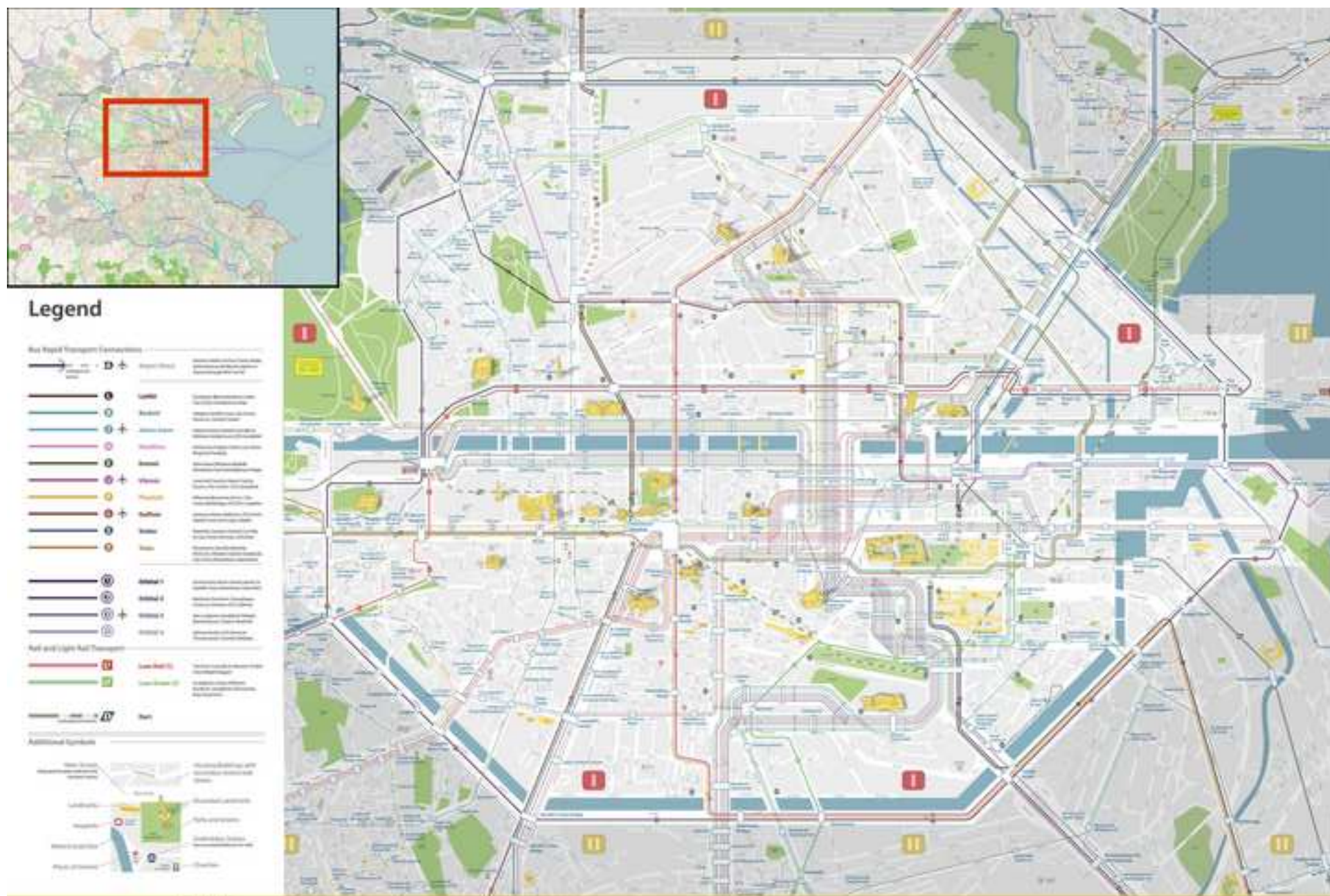


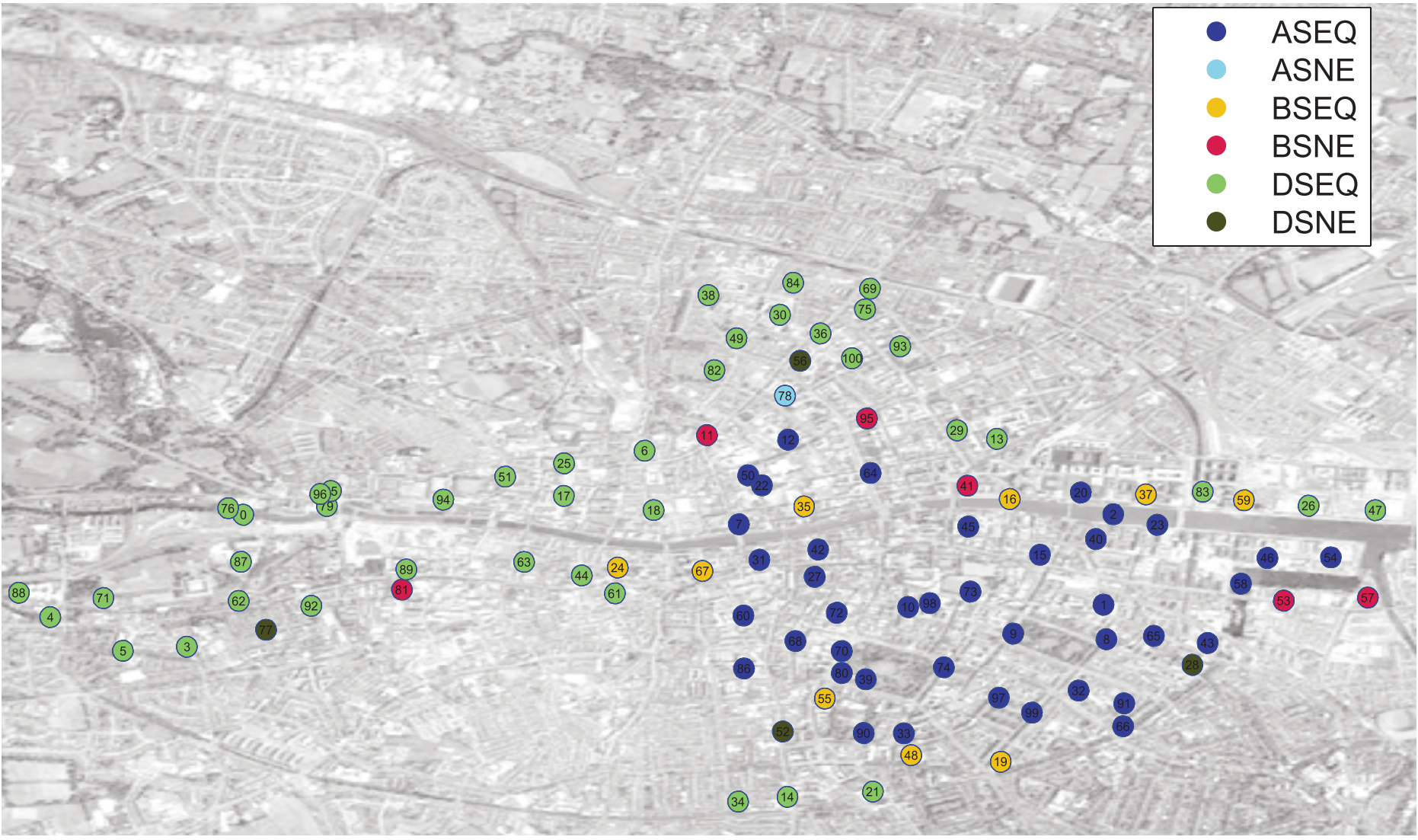




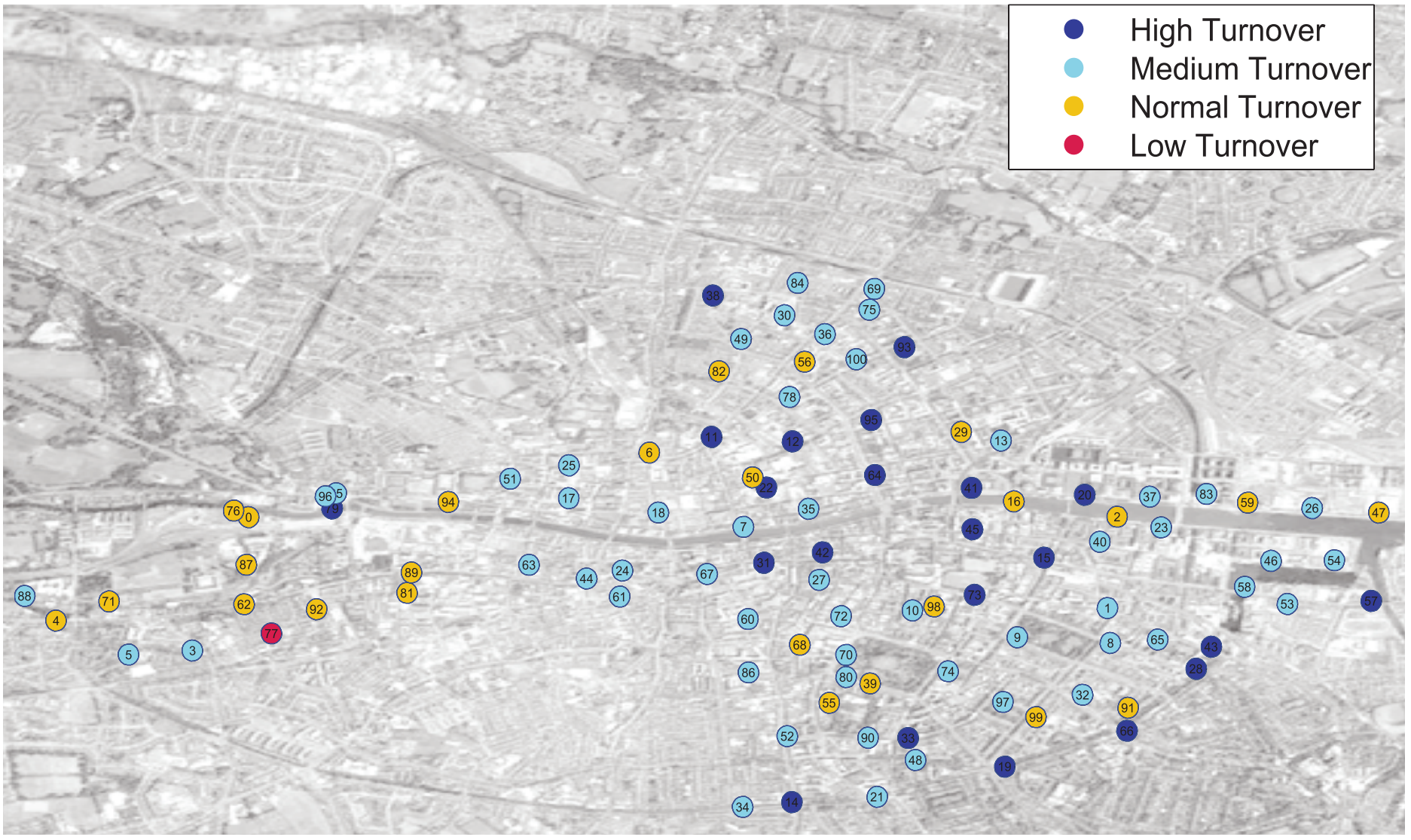
Figure

[Click here to download high resolution image](#)

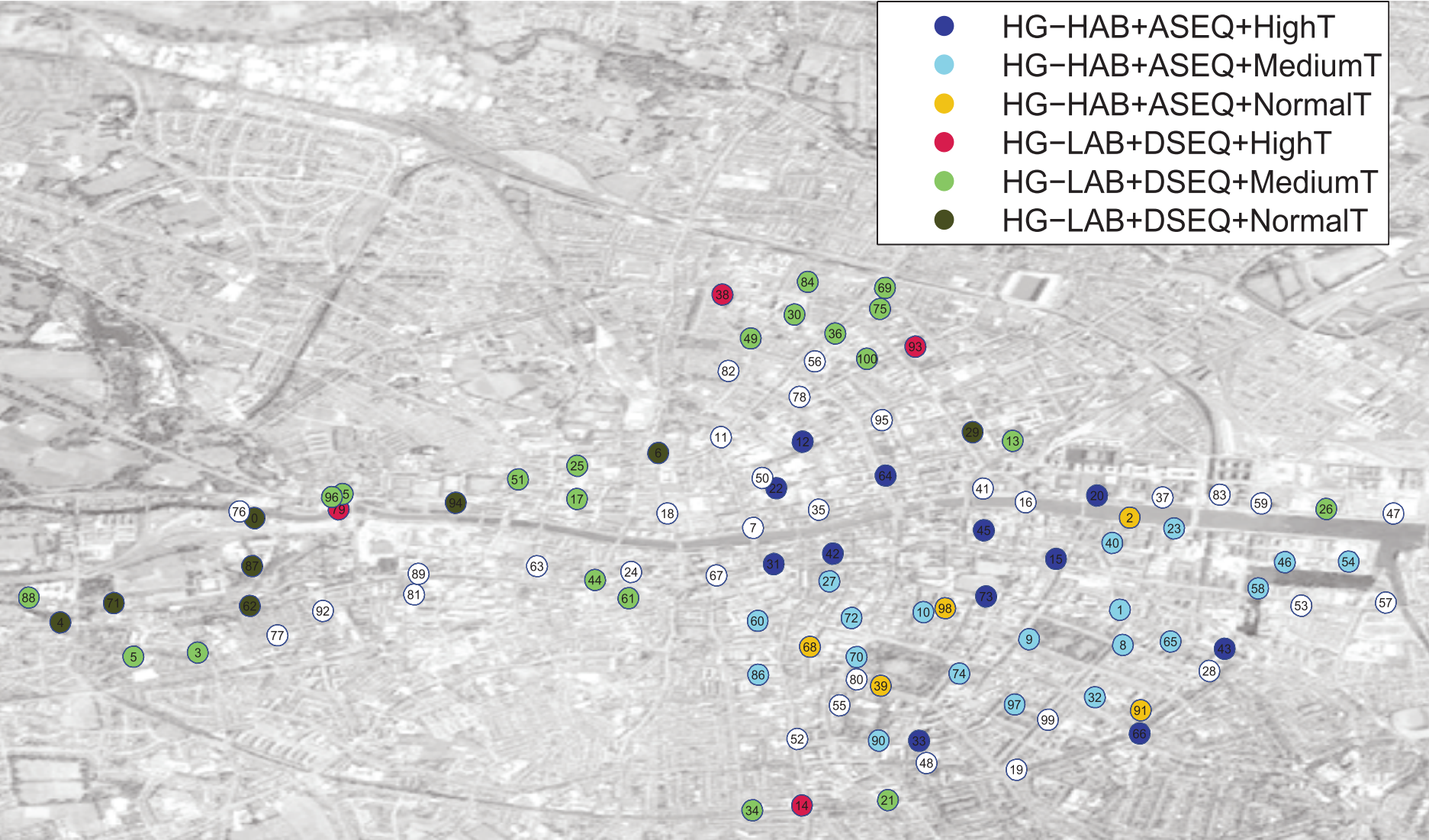




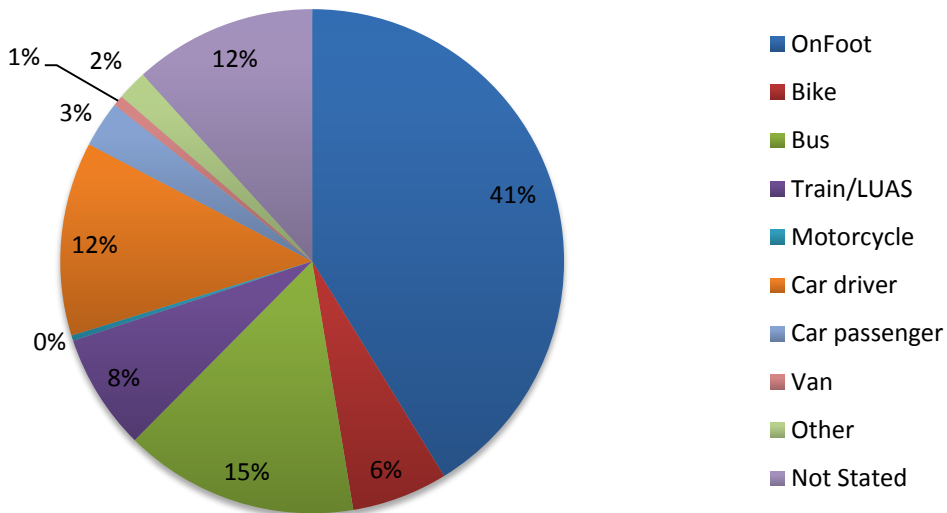
- High Turnover
- Medium Turnover
- Normal Turnover
- Low Turnover

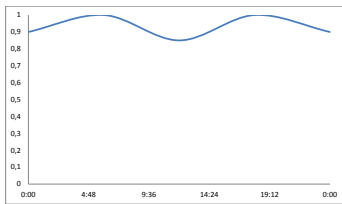


- HG-HAB+ASEQ+HighT
- HG-HAB+ASEQ+MediumT
- HG-HAB+ASEQ+NormalT
- HG-LAB+DSEQ+HighT
- HG-LAB+DSEQ+MediumT
- HG-LAB+DSEQ+NormalT

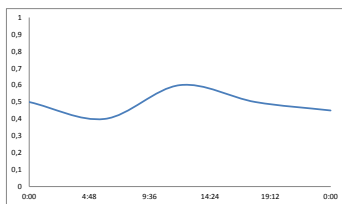




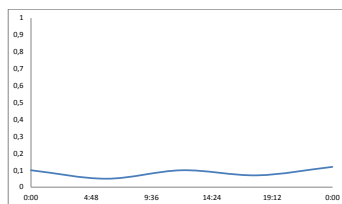




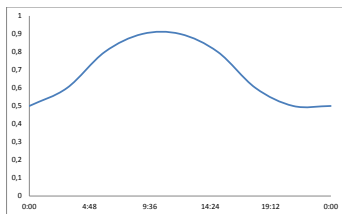
Short Gap
High Availability Bikes
SG-HAB



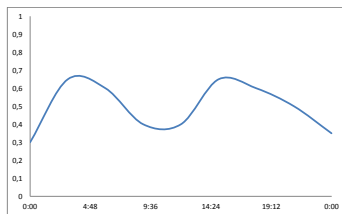
Short Gap
Medium Availability Bikes
SG-MAB



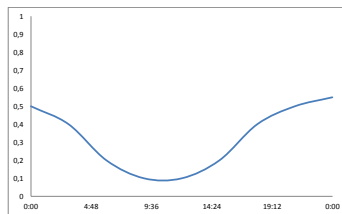
Short Gap
Low Availability Bikes
SG- LAB



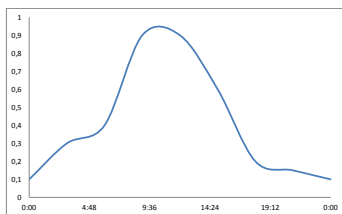
Medium Gap
High Availability Bikes
MG-HAB



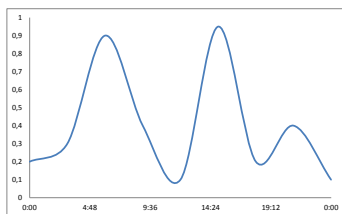
Medium Gap
Medium Availability Bikes
MG-MAB



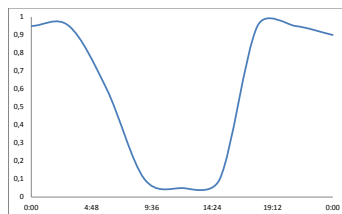
Medium Gap
Low Availability Bikes
MG- LAB



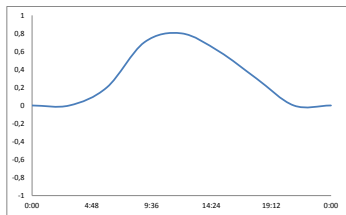
High Gap
High Availability Bikes
HG-HAB



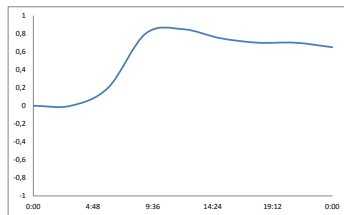
High Gap
Medium Availability Bikes
HG-MAB



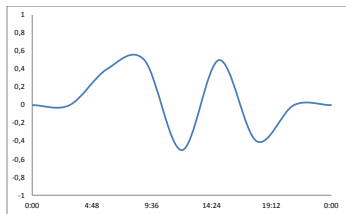
High Gap
Low Availability Bikes
HG- LAB



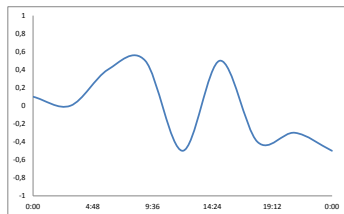
Arrival Station Equilibrium
ASEQ



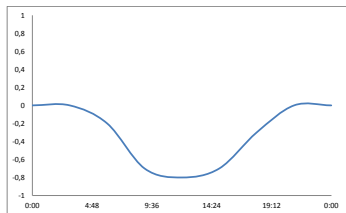
Arrival Station Non-Equilibrium
ASNE



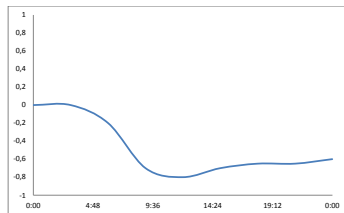
Balanced Station Equilibrium
BSEQ



Balanced Station Non-Equilibrium
BSNE



Departure Station Equilibrium
DSEQ



Departure Station Non-Equilibrium
DSNE

High Turnover	$TS \geq 2.5$
Medium Turnover	$1.5 \leq TS < 2.5$
Normal Turnover	$0.5 \leq TS < 1.5$
Low Turnover	$0.2 < TS < 0.5$
Non Turnover	$TS \leq 0.2$

Mixed Pattern	# of stations
SG-LAB + BSEQ + Medium Turnover	2
SG-LAB + BSEQ + Normal Turnover	2
MG-HAB + ASEQ + Medium Turnover	2
MG-HAB + ASEQ + Normal Turnover	2
MG-HAB + BSEQ + Medium Turnover	1
MG-HAB + BSEQ + Normal Turnover	1
MG-MAB + DSNE + Normal Turnover	1
MG-LAB + DSEQ + Medium Turnover	1
MG-LAB + DSEQ + Normal Turnover	2
MG-LAB + DSNE + Low Turnover	1
HG-HAB + ASEQ + High Turnover	12
HG-HAB + ASEQ + Medium Turnover	19
HG-HAB + ASEQ + Normal Turnover	5
HG-HAB + BSNE + High Turnover	1
HG-MAB + ASNE + Medium Turnover	1
HG-MAB + BSEQ + High Turnover	1
HG-MAB + BSEQ + Medium Turnover	2
HG-MAB + BSNE + High Turnover	3
HG-MAB + BSNE + Medium Turnover	1
HG-MAB + BSNE + Normal Turnover	1
HG-MAB + DSEQ + Medium Turnover	2
HG-MAB + DSEQ + Normal Turnover	3
HG-MAB + DSNE + High Turnover	1
HG-MAB + DSNE + Medium Turnover	1
HG-LAB + DSEQ + High Turnover	4
HG-LAB + DSEQ + Medium Turnover	21
HG-LAB + DSEQ + Normal Turnover	8

LaTeX Source Files

[Click here to download LaTeX Source Files: LatexJTRGD1500385.rar](#)