Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 20 (2011) 514–523

Procedia
Social and Behavioral Sciences

14th EWGT & 26th MEC & 1st RH

Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns

Patrick Vogel[a,*], Torsten Greiser[a], Dirk Christian Mattfeld[a]

[a] University of Braunschweig, Decision Support Group, Mühlenpfordtstraße 23, D-38106 Braunschweig, Germany

Abstract

In this paper we analyze extensive operational data from bike-sharing systems in order to derive bike activity patterns. A common issue observed in bike-sharing systems is imbalances in the distribution of bikes. We use Data Mining to gain insight into the complex bike activity patterns at stations. Activity patterns reveal imbalances in the distribution of bikes and lead to a better understanding of the system structure. A structured Data Mining process supports planning and operating decisions for the design and management of bike-sharing systems.

Keywords: Bike-Sharing; Data Mining; Activity Patterns

1. Introduction

Bikes receive increasing attention in city transportation, mainly because they “provide the missing link between existing points of public transportation and desired destinations” (Midgley, 2009). Bike-sharing can be described as a short-term bicycle rental service for inner-city transportation providing bikes at unattended stations. In recent years, bike-sharing systems (BSS) have rapidly emerged in major cities all over the world.

Bike-sharing providers have to ensure high bike availability in order to satisfy customers. A challenging task, because movements of customers are highly dynamic and redistributing bikes is expensive. According to recent studies, the analysis of rides show spatio-temporal dependencies in bike usage (Borgnat et al., 2010, Froehlich et al., 2009, Kaltenbrunner et al., 2010). In addition, one-way use and short rental times lead to imbalances in the spatial distribution of bikes. There are different measures for overcoming imbalances, e. g. location planning of bike stations. With the help of Geographical Business Intelligence (Geo BI), ride data can be analyzed and processed in order to support location decisions (Feix, 2007).

* Corresponding author. Tel.: +49-531-391-3215; fax: +49-531-391-8144.
E-mail address: p.vogel@tu-bs.de.

doi:10.1016/j.sbspro.2011.08.058

This paper makes the following contributions: A taxonomy for measures alleviating bike imbalances in BSS is given in Section 2. A Geo BI process which includes data mining (DM) methods for location planning is presented in the following section. Furthermore, two years of operational ride data from Vienna's BSS "Citybike Wien" is analyzed according to the Geo BI process. We present results from an exploratory data analysis and from clustering stations according to their pickup and return activity to understand temporal and spatial causes of imbalances.

2. Design and Management of BSS

Over the last decades, traditional bike rental for leisure activities have evolved to modern BSS offering environment-friendly, individual urban mobility. According to Shaheen et al. (2010), primary goals of BSS are reduction of congestion and pollution, improvement of public health, increase in bike use and enhancement of public transport options. However, bikes are mainly suited for short trips and bike use strongly depends on weather and topography. Typically, stations are spread over the city providing free bikes or free boxes, respectively. Easy and quick city wide access of BSS is provided by information and communication technology that supports automated bike rentals and stores cyclists' rental activities for billing and control purposes.

Bike-sharing providers comprise transport agencies, local governments, profit and non-profit organizations as well as advertising companies (DeMaio, 2009). Advertising companies are the most common type of provider. They receive rights from municipalities to advertise on street furniture. In exchange a BSS is provided. In contrast to traditional rental services, BSS' pricing models are linear or progressive. The first minutes of a ride are often for free. After that, fees increase progressively in certain time steps, e. g. every 30 minutes. This is an incentive for customers to keep rental times short, making the bikes available for others to reach a high bike turnover.

The availability of free bike and boxes is crucial for the usage and acceptance of BSS. For instance, the provider of Arlington's BSS has to ensure that stations are not full of bicycles for more than 60 minutes during daytime and 180 minutes during nighttime (Zahory, 2009). Ensuring high availability is not an easy task. A common issue observed in BSS is imbalances in the spatial distribution of bikes at stations over time due to one-way use and short rental times. The number of bike boxes at stations is limited. It is impossible to return bikes at full stations and pick up bikes at empty stations. We identify three design and management measures alleviating these imbalances divided into different planning horizons:

- Strategic (long-term) network design comprising decisions about the location, number and size of stations.
- Tactical (mid-term) incentives for customer based distribution of bikes. For example, the Parisian BSS "Vélib'" grants 15 extra minutes for returning bikes at uphill stations (DeMaio, 2009).
- Operational (short-term) provider based repositioning of bikes. Here, staff relocates bikes from full to empty stations. E. g. costs for reposition of a bike in the "Vélib" amount to 3 $ (DeMaio, 2009).

Decisions on the strategic level have a great impact on the following levels. Once location and size of a station is implemented, it cannot be changed on a short-term basis. Unsophisticated design decisions have to be compensated by extensive effort on the tactical and operational level. Therefore fluctuations of pickups and returns on the operational level have to be anticipated and considered in the design phase. In this paper we present a Geo BI process for BSS design. On the one hand, the process consists of a DM part to gain insights into the complex bike operations at stations. On the other hand, an approach is presented how to incorporate the gained knowledge in BSS design decisions.

3. Using Data Mining to improve BSS design

Planning and operating of BSS receives attention in academia as well as in practice. However, scientific literature in this field is still rather scarce. A literature overview of DM and OR articles on BSS is given as follows:

- Borgnat et al. (2010) use DM to analyze the dynamics of bike movements in Lyon's BSS. Temporal patterns in the system-wide bike usage are examined. Weekdays show usage peaks in the morning, at noon and late afternoon, whereas usage is concentrated in the afternoon on weekend days. A statistical model for the prediction of the number of rentals on a daily and hourly basis is given. Furthermore spatial patterns are examined by clustering bike flows between stations. Spatial and temporal dependencies exist between stations of clusters interchanging many bicycles.

- Froehlich et al. (2009) have data about the number of available bikes and vacant bike boxes from Barcelona's BSS. Stations are clustered according to the number of available bikes and an activity score in the course of day. Visualization of the stations show spatial dependencies, e. g. uphill stations tend be empty and less active stations are located at the edge of the bike-sharing network.
- Kaltenbrunner et al. (2010) detect bike usage patterns in data from Barcelona's BSS. Results are similar to Froehlich at el (2009). Also a statistical model is presented that predicts the number of free bikes and boxes at stations some minutes ahead in time.
- From an OR perspective, Lin and Yang (2011) present a mathematical decision model to determine an adequate number and location of bike stations considering bike availability at stations. Also the network structure of bike paths between the stations and travel paths for users between their origin and destination are determined. Artificial bike demand data is used for testing the model.

Recent work either focuses mining of bike-sharing data or building decision models without incorporating real world BSS behavior. For instance, Lin and Yang (2011) only consider the availability of bikes and not the availability of boxes. In contrast to observations from real world BSS, hourly fluctuations in rentals are ignored in Lin and Yang's model. We present an integrated approach of DM and OR. DM provides information about the system structure which is used to determine attributes serving as input for a decision model or to refine the decision model structure (Meisel and Mattfeld, 2010). From a strategic point of view it is necessary to at least understand bike activities at stations on the operational level. Ignoring these operational activities can lead to suboptimal location decisions (Salhi and Rand, 1989). Location and size of station have to be determined in such way that fluctuations do not cause full or empty stations. Therefore, anticipating operational bike activities in long-term location planning is crucial for the actual bike availability at stations.

For anticipation of bike activities, a Geo BI approach for location planning in BSS is presented. Geo BI is based on knowledge discovery in databases (KDD) (Fayyad et al., 1996) extended by spatial analysis and visualization. Novel and understandable patterns are derived from data including spatial relations. Geographical information technology and DM methods are used to structure company internal and external data. Identified patterns are visualized with the aim of supporting business decisions (Feix, 2007). Potentially, DM can generate a multitude of patterns. Domain knowledge is necessary to separate interesting from uninteresting patterns. A pattern is interesting e. g. if it validates a hypothesis that sought to be confirmed (Han and Kamber, 2006). Concerning location planning in BSS, we are looking for spatial relations between bike usage at stations and location of stations. This leads to the following hypothesis: *Pickups and returns at bike stations as well as the type of customers using certain stations depend on the stations' surroundings.* If this assumption is valid, bike usage can be mapped to new stations according to their location.

With the help of the Geo BI process this hypothesis is evaluated. The bike activity at every station is derived from BSS ride data. This bike activity comprises stations' pickups and returns aggregated on a suitable timescale. Cluster analysis yields distinct groups of stations showing similar activity patterns. Also customer behavior is examined with cluster analysis leading to customer segments with different rental behavior. Furthermore location factors are derived from the stations' surroundings. Bike stations are clustered according to their location factors. These factors include e.g. access of stations to public transport, population, housing or commercial area. If station and customer clusters match the location clusters, the hypothesis can be confirmed.

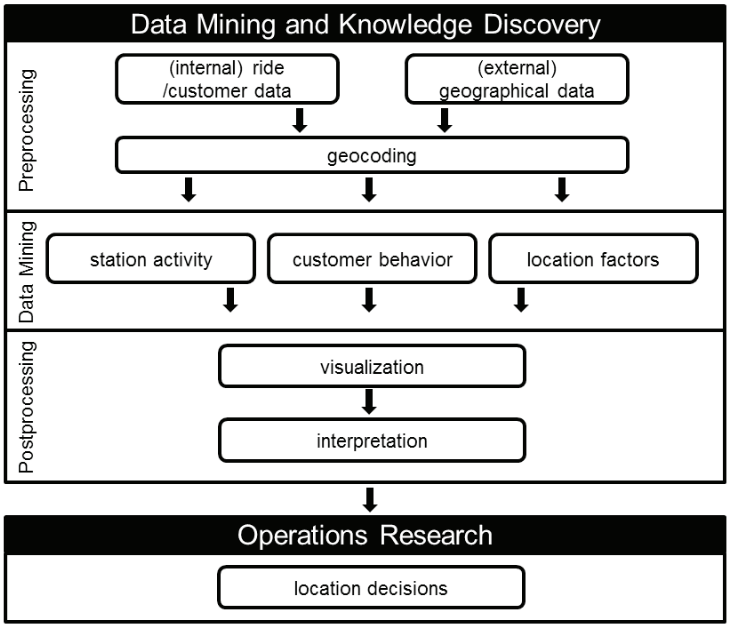

Figure 1. Geo BI process in style of Feix (2007)

When setting up or expanding a BSS, the pickups and deliveries at potential locations have to be anticipated. Location factors can be ascertained for every possible new location. Activity patterns and customer segments for already existing stations are then mapped to potential locations based on location factors. In a following location decision model an adequate number, location and size of stations have to be determined whereas the anticipated operating activities and bike availability have to be considered. The Geo BI depicted in Figure 1 process comprises DM and OR in the following steps:

- *Preprocessing*: In the preprocessing phase ride and customer data as well as location factors in terms of external geographical data are gathered. To make the data suitable for DM it has to be properly cleaned and selected. Also aggregation and normalization of data assures a solid basis for DM algorithms.
- *Data Mining*: With the help of cluster analysis station activity and location factor patterns as well as customer segments are identified. First, cluster validation is necessary in order to determine the clustering tendency of the data set by distinguishing whether there is a non-random structure in the data or not. Second, a cluster algorithm determines groups in which objects are similar. Last, cluster validation measures the adequacy of the cluster analysis.
- *Postprocessing*: The results from the cluster analysis have to be validated in a postprocessing step. Visualizing clusters with a geographical information system is needed for a spatial interpretation of clusters.
- *Operations Research*: Knowledge about the system structure and processed data from DM serves as input for the OR phase. The obtained knowledge and data lead to sophisticated OR models. The adequate number, size and location of stations are output of location decision models.

With the help of Geo BI, bike activity at stations and customer behavior can be anticipated at potential locations. This can improve the quality of strategic decisions. The following case study gives insight into temporal and spatial dependencies of bike pickups and returns at stations.

4. Case Study: Exploring Activity Patterns

In this section ride data is analyzed based on the Geo BI process. Since imbalances in the distribution of bikes have temporal and spatial reasons, the ride data is examined for spatio-temporal activity patterns. Therefore, bike stations are clustered according to pickups and returns in the course of day. Analysis of the stations surroundings validates determined clusters and gives insight into possible causes for imbalances in the distribution of bikes. The analyzed data covers two years of operational ride data from Vienna's BSS "Citybike Wien" (http://citybikewien.at). The system is operated by "Gewista Werbegesellschaft m. b. H." and consists of more than 60 bike stations. Rapidminer Community Edition from Rapid-I GmbH (http://rapid-i.com) is used for data analysis. Results are visualized with the Geographical Information System Google Earth (http://earth.google.com).

4.1. Preprocessing

The provided data sets consist of ride information in form of pickup station and timestamp as well as return station and timestamp (cf. Table 1). The data covers approximately 760'000 rides of the years 2008 and 2009. Trip durations are calculated by subtracting pickup from return timestamps. Also geographical coordinates are available for every station.

Table 1. Structure of ride data

Pickup station	Pickup timestamp	Return station	Return timestamp
1074	2008-01-01 17:13:27	1052	2008-01-01 17:25:54

According to the *preprocessing* phase, data objects that show the following characteristics are removed:

- Rides that start or end at test stations.
- Rides where bikes are reported as defect or stolen.
- Rides that show negative drip durations. This can happen if an error occurs at the bike box while returning a bike.
- Rides that last less than 60 seconds which start and end at the same station. This indicates that a bike is immediately returned after being picked up. An actual ride does not take place.
- Stations with only a few pickups or returns. These stations could show non-typical return and pickup patterns which distorts the clustering.

After removing the affected data sets, the number of data sets declines by 2 % to approximately 743'000. The number of stations drops to 59. Because of the extensive amount of data, an inspection of single rides does not lead to a general impression of bike activities. When looking for temporal patterns in bike activities, rides have to be aggregated at a suitable timescale. The tradeoff is the following: the smaller the time windows, the larger the fluctuations, whereas larger time windows might smooth relevant observations. Citybike Wien grants that the first 60 minutes of a ride a free of charge. This is the case for more than 90% of all trips. An average trip lasts 29 minutes, whereas the median trip duration is 15 minutes. A first glance at the data shows that bike activities are highly fluctuant with respect to time of day, day of week and type of day, e.g. holidays. That is why ride data is aggregated for 24 time windows per day of the week, leading to 168 values representing the two years of data. These time windows correspond to common approaches of traffic systems analysis (Pinkofsky 2006).

The total number of rented bikes in the course of week is depicted in Figure 2. Patterns for working days and weekends can be observed. Working days show three and weekend days two peaks during the course of day. On working days, the night peak occurs from 0-2 a.m. A reason for this might be that the subway stops service between 0-1 a.m. The least active hour is at 5 a.m. Due to morning commuters there is a peak in the number of rentals between 8-10 a.m. Another peak is situated in the late afternoon hours. Here, commuting and leisure activities overlap. On weekends, the night peak is more distinct, whereas the morning peak is absent. This indicates that the Citybike is predominantly used for leisure activities at weekends.

According to our hypothesis, bike pickups and returns depend on the stations' locations. Therefore it is necessary to break down the system wide usage patterns to station related pickups and returns activities. Since all working days and all weekend days show similar usage patterns, pickups and returns are aggregated for every station on an hourly basis. Working days and all weekend days are considered separately. Public holidays are excluded from the aggregation. Normalization of data is needed in order to compare the activity at stations. Thus, the activity at stations is defined by the proportion of pickups or returns in a certain hour of day divided by the total daily number of pickup or returns. This leads to 48 attributes describing the daily pickup and return activity for every station. An example for a station activity data set is stated in Table 2.

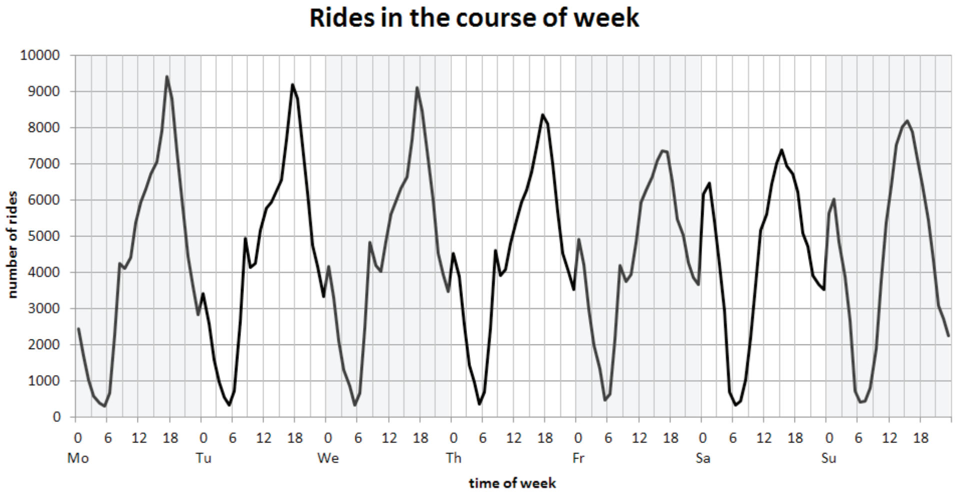

Figure 2. Number of rides in the course of week

Table 2: Normalized activity for station 1020 in the course of day

Station	Pickups 0-1	…	Pickups 23-0	Returns 0-1	…	Returns 23-0
1020	0.065148353	…	0,035205517	0,05418259	…	0,038980892

4.2. Data Mining

The normalized activity at stations serves as input for the DM phase. Here, cluster analysis is used in order to group stations according to their normalized bike pickup and return activity. The outcome in form of temporal patterns are spatially examined to discover location dependent reasons for activity patterns. First, the clustering tendency of the ride data is determined. Then, different cluster algorithms are applied and evaluated with validation measures. Visualization and interpretation of clusters complete this section.

Almost every cluster algorithms will find clusters, even if the data has no natural cluster structure (Tan et al., 2006). Thus, cluster validation is necessary in order to determine the clustering tendency of the data set by distinguishing whether there is a non-random structure in the data or not. With the help of the Hopkins-Statistic (Tan et al., 2006) the ratio of nearest neighbor distances between randomly generated and actual data points are calculated. Values near to 0.5 indicate randomly distributed data whereas values close to 0 or 1 indicate that data is highly clustered or regularly distributed. The Hopkins-Statistic applied to the normalized pickups and returns at stations yields a value of 0.743. This indicates that the data set is suitable for cluster analysis.

Cluster analysis is used to group stations according to their pickup and return activity. The goal is that data objects within a group are similar to each other and different from objects in other groups (Tan et al., 2006). Tan et al. (2006) suggest comparing the outcome of different cluster algorithms for a varying number of clusters. Therefore *k*-means (KM) (Tan et al., 2006), Expectation Maximization (EM) (Tan et al., 2006) and sequential Information-Bottleneck (sIB) (Slonim et al., 2002) are applied for cluster analysis. KM is one of the most well-known partitional cluster algorithms. Due to the high dimensionality, partitional cluster algorithms seem to be a promising approach (Berkhin, 2006). Here data objects are assigned to *k* clusters whereas the number of clusters has to be chosen beforehand. Based on an initial partitioning, objects are relocated by minimizing the distances of objects within clusters and maximizing the distance of objects in different clusters. The EM algorithm extends the KM paradigm. Each object is assigned to a cluster on basis of a weight that represents the probability of membership. sIB is an agglomerative clustering method originally designed for cluster analysis of documents. sIB is used because it is capable of dealing with high dimensional data. Cluster validation indices measure if a structure found with cluster analysis is adequate. No reference to external information is taken into account by the applied measures. The Davies-Bouldin-Index (Jain and Dubes, 1988), Dunn-Index (Abonyi and Feil, 2007) and Silhouette-Index (Tan et al., 2006) determine the cohesion and separation of clusters according to the distance between data sets among each other and their clusters. The higher the value for Dunn and Silhouette Index the better the clustering. The opposite holds for the Davies-Bouldin-Index. Cohesion and separation get better with ascending number of clusters. Therefore a first (local) optimum in the indices` values, called elbow, indicates a reasonable clustering (Tan et al., 2006). EM, KM and sIB are applied to the normalized activity data for *k* varying from two to ten. The best result of 100 iterations with different seeds for each algorithm is chosen using the Euclidian distance as similarity measure. A maximum of 500 steps is performed in every iteration. The additional parameter for EM is the minimum standard deviation for normal density computation set to 1E-10. For sIB the number of different solutions is two and the minimum number of changes in a single step is zero. The computed results are validated with three indices.

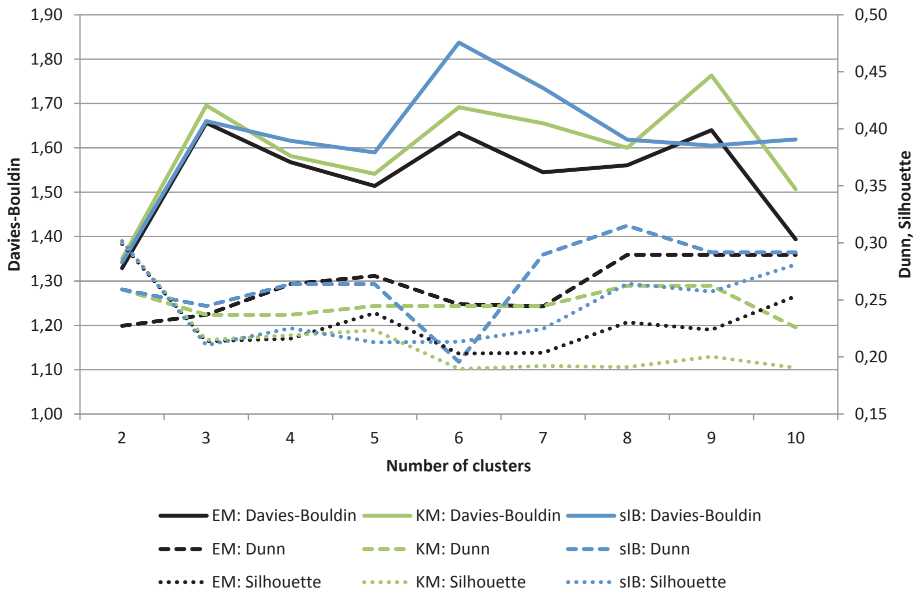

Figure 3. Index values for different cluster algorithms

Cluster validation values for working days are depicted in Figure 3. A clear elbow indicated by a drop in the Davies-Bouldin-Index and peaks for Dunn und Silhouette can be identified for the EM-Algorithm with 5 clusters. According to the elbow criterion, sIB and KM also yield the best clustering for $k = 5$. A comparison of cluster algorithms among each other shows that EM outperforms sIB and KM. Since cluster validation indices only determine the cohesion and separation of clusters, an interpretation of clusters themselves is needed. Results are discussed in the following section.

4.3. Temporal and spatial validation of clusters

According to our hypothesis, daily bike activity depends on the stations location. Thus, results of the EM algorithm for five clusters are temporally and spatially validated. The following validation is conducted in close consultation with the operator of the BSS. The centroid of every cluster is used for a temporal validation of clusters. A centroid represents the average point in space for a cluster. Regarding the activity at stations, the centroid is the average pickup and return proportion for every single hour of all stations belonging to the same cluster. The centroids are depicted in Figure 4. At blue cluster stations at average 6% of the total daily returns occur between 8 and 9 a.m., for example.

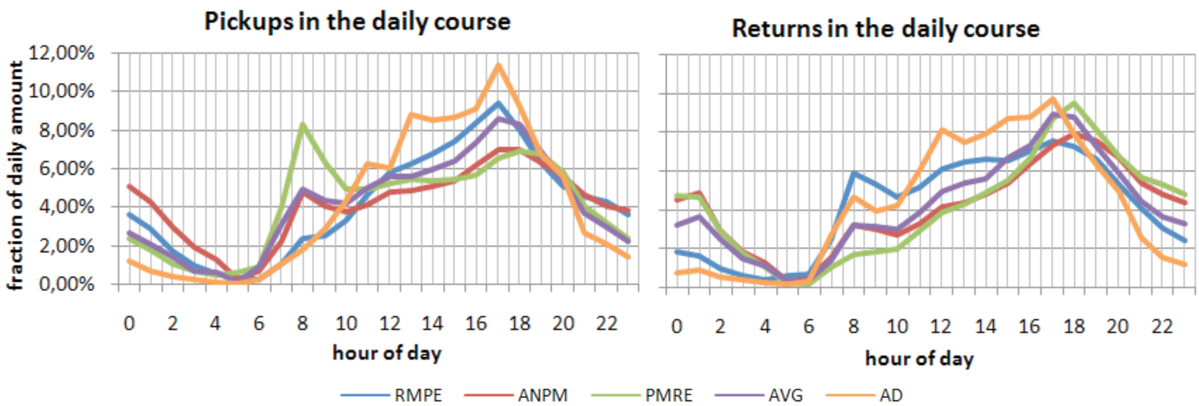

Figure 4. Activity patterns for pickups and returns in the daily course for working days

Five different patterns in the temporal pickup and return activity at stations are distinguished. For a better understanding clusters are labeled according to their daily pickup and return patterns:

- 16 stations are assigned to cluster *Returns Morning Pickups Evening (RMPE)*. These stations show a higher return activity in the morning compared to pickup activity (6 % vs. 2 %). In the evening, pickup activity is slightly higher than return activity.
- Cluster *Pickups Morning Returns Evening (PMRE)* consists of 13 stations which feature the opposite pattern. Morning return activities are four times higher than pickup activities whereas return activities dominate in the evening hours.
- The 12 stations of the cluster *Active Night Pickups Morning (ANPM)* shows the highest activity of all clusters at nighttime. Morning activities are slightly increased comparing to *PMRE*.
- The biggest cluster with 19 stations is labeled *AVG* because it reflects average pickup and return activities. When averaging the activity of all stations, the outcome is very similar to the pattern of this cluster.
- Cluster *Active Daytime (AD)* is the smallest cluster with only 3 stations. This cluster's pickup and return activities dominate over the other clusters during daytime.

With the help of a clustering algorithm stations are grouped according to their temporal activities. For location planning purposes it is necessary to examine if these activities depend on spatial factors. That is why the clusters' geographical distribution is visualized in Figure 5. Stations within the same cluster tend to be located in neighboring areas. Exploratory analysis of the clusters' surroundings leads to the following findings:

- The blue *RMPE* stations are all located in the inner city area which offers a high number of working places compared to other districts (Vienna Statistical Yearbook, 2009). This supports the assumption that workers use the Citybike for commuting. Also tourist attractions are found in the city center. Tourists and workers might be the cause for the high pickup and return activity in the evening hours.
- In the middle of cluster *RMPE* lays station 1021 of the orange *AD* cluster. Citybike tourist cards are handed out near this station. The other two *AD* stations are located near the famous tourist attractions castle Schoenbrunn in the southwest and the Prater carnival in the northeast. This confirms the high daily and low nightly activity of these stations which might be caused by tourists.
- Stations of the night active red *ANPM* cluster tend to be located at the edge of the network and close to the city center. For example, popular night clubs and bars lay in the immediate vicinity of station 1041 Falkostiege and 1042 Pilgramgasse causing high nightly activity.
- Cluster *PMRE* shows a similar activity pattern to *ANPM*, but is more active during the morning and evening. Green stations are even more located at the periphery with residential buildings.
- Although *AVG* stations are randomly spread over the network, they still occur in small groups. One reason for this might be that spatial factors overlap at these stations.

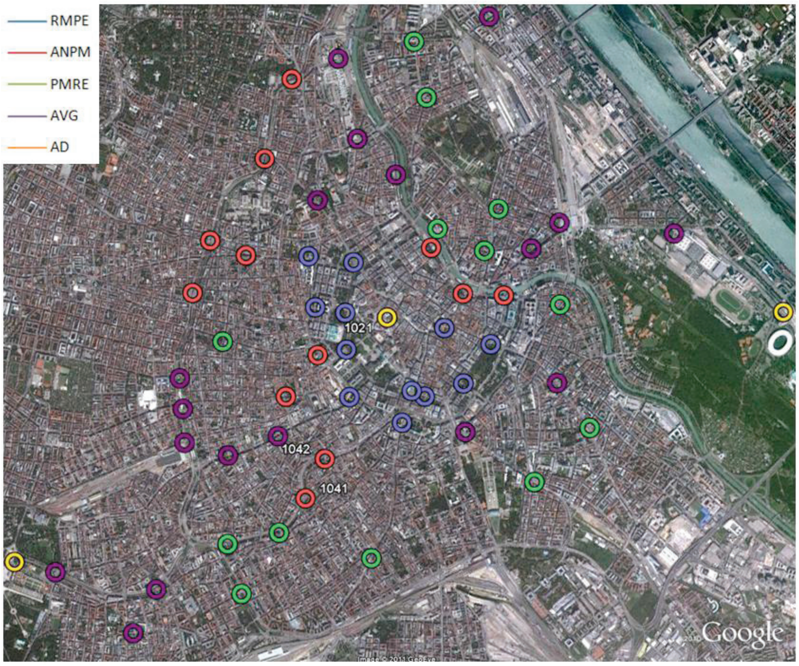

Figure 5. Geographical distribution of clusters (aerial view of the city of Vienna)

The presented findings from the spatio-temporal analysis show that reasons for certain pickup and return activities at bike stations are complex and diverse. The presumption that activity patterns and the station's location are correlated seems promising according to first results from the Geo BI process. A further investigation of the stations surroundings is needed to fully support the hypothesis.

5. Conclusion and Outlook

BSS enhance inner-city public transport options. Ensuring high bike availability is crucial for the acceptance of such systems. Due to one-way use and short rental times imbalances in the spatial distribution of bikes occur. One measure alleviating imbalances is a suitable location planning of bike stations with OR. A Geo BI process is presented to gain insights into the complex bike activity. This leads to better hypotheses about BSS structure and increases OR effectiveness.

We hypothesize that bike activity and demanding customers depend on the stations' locations. With the help of Geo BI, exploratory and cluster analysis of ride data reveal spatio-temporal dependencies of pickup and return activity patterns at bike stations. This supports the stated hypothesis. In future research, customer profiles and clustering of stations according to their location factors have to be determined with the aim of further supporting the hypothesis. Also ride data has to be adequately processed for the incorporation in location planning models.

Acknowledgements

We thank „Gewista Werbegesellschaft m.b.H.“ for providing data from their project “Citybike Wien” and discussing related issues and results.

References

Abonyi, J. & Feil, B. (2007).*Cluster Analysis for Data Mining and System Identification*. Birkhäuser Verlag AG.
Berkhin, P. (2006). *A Survey of Clustering Data Mining Techniques*. In: Grouping Multidimensional Data. 25 – 71.

Borgnat, P., Robardet, C., Rouquier, J.-B., Abry, P., Fleury, E. & Flandrin, P. (2010). *Shared Bicycles in a City: A Signal Processing and Data Analysis Perspective*.

DeMaio, P. (2009). Bike-sharing: History. Impacts. Models of Provision. and Future. In: *Journal of Public Transportation*. Vol. 12. No. 4. 2009.

Fayyad, U. M., Piattetsky-Shapiro, G., Smyth, P. (1996). *From Data Mining to Knowledge Discovery: An Overview*. In: Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Ed.): *Advances in Knowledge Discovery and Data Mining*. AAAI Press, Menlo Park, 1–34.

Feix, C. (2007). *Bedeutung von "Geo Business Intelligence" und Geomarketing zur Entscheidungsunterstützung unternehmerischer Planungsprozesse im Kontext wirtschaftlicher Liberalisierung.* Freie Universität Berlin.

Froehlich, J., Neumann, J. & Oliver, N. (2008). Sensing and Predicting the Pulse of the City through Shared Bicycling. (Proceedings of the International Workshop on Urban. Community. and Social Applications of Networked Sensing Systems).

Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann. San Francisco.

Jain, A. K. & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall.

Kaltenbrunner, A., Meza, R., Grivolla, J., Condina, J. & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*. 6(4). 455–466.

Lin, J.-R. &Yang, T.-H. (2011). Strategic design of public bicycle sharing systems with service level constraints. *In Transportation Research Part E: Logistics and Transportation Review* 47 (2), pp. 284–294.

Meisel, S. & Mattfeld, D. C. (2010). Synergies of Operations Research and Data Mining. *European Journal of Operational Research*. 206(1). 1–10.

Midgley, P. (2009). The Role of Smart Bike-sharing Systems. In: *Urban Mobility. Journeys*. May. 23 – 31.

Pinkofsky, L. (2006). *Typisierung von Ganglinien der Verkehrsstärke und ihre Eignung zur Modellierung der Verkehrsnachfrage*. Dissertation. Technische Universität Carolo-Wilhelmina zu Braunschweig. Institut für Verkehr und Stadtbauwesen.

Salhi, R. & Rand, G. K. (1989). The effect of ignoring routes when locating depots. *European Journal of Operations Research* 39 (1989). 150 – 156.

Shaheen, S., Guzman, S. & Zhang, H. (2010). Bikesharing in Europe. the Americas. and Asia: Past. Present and Future. *Transportation Research Record*.

Slonim, N., Friedman, N. & Tishby, N. (2002). Unsupervised document classification using sequential information maximization. In: Beaulieu. M. (Ed.): *SIGIR 2002*. Tampere. Finland. 129 – 136.

Tan, P.-N., Steinbach, M. & Kumar,V. (2006). *Introduction to Data Mining*. Pearson Addison-Wesley.

Vienna Statistical Yearbook (2009). http://www.wien.gv.at/statistik/pdf/bezirksportraets09.pdf

Zahory, M. N. (2009). *Tendering for the BSS in Arlington*. www.metrobike.net/index.php?s=file_download&id=18.