

---

# Community Structures, Interactions and Dynamics in London's Bicycle Sharing Network

**Fernando Munoz-Mendez\***

AECOM  
Bedford, UK  
*fernando.munozmendez@aecom.com*

**Ke Han**

Imperial College London  
London, UK  
*k.han@imperial.ac.uk*

\* These authors contributed equally to the work

**Konstantin Klemmer\***

University of Warwick &  
The Alan Turing Institute  
Coventry, UK  
*k.klemmer@warwick.ac.uk*

**Stephen Jarvis**

University of Warwick &  
The Alan Turing Institute  
Coventry, UK  
*s.a.jarvis@warwick.ac.uk*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

*UbiComp/ISWC'18 Adjunct*, October 8–12, 2018, Singapore, Singapore  
© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5966-5/18/10...\$15.00

<https://doi.org/10.1145/3267305.3274156>

**Abstract**

We apply a novel clustering technique to London's bikesharing network, deriving distinctive behavioral patterns and assessing community interactions and spatio-temporal dynamics. The analyses reveal self-contained, interconnected and hybrid clusters that mimic London's physical structure. Exploring changes over time, we find geographically isolated and specialized communities to be relatively consistent, while the remaining system exhibits volatility. We increase understanding of the collective behavior of the bikesharing users.

**Author Keywords**

Bikesharing; community detection; spatio-temporal analysis; clustering; urban mobility

**ACM Classification Keywords**

•**Information systems** → **Location based services**;  
•**Applied computing** → Transportation;

**Introduction**

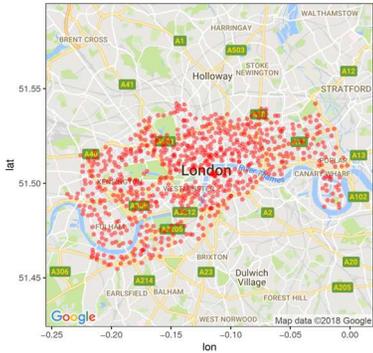
Bicycle Sharing Schemes (BSS) have become increasingly vital elements of urban mobility due to their complementary effect to conventional modes and last-mile connectivity to transit systems. By now, there are more than 600 BSS globally with the largest

systems located in China. The health benefits of bicycle use in cities even outweighs accident risk [9] and offer sustainable solutions to urban transportation by contributing to resolving crucial problems like congestion and pollution. In order to increase the expansion of BSS and attract new customers, it is vital to understand relevant travel patterns, and adjust design and management strategies (e.g. pricing, marketing, expansions) to encourage adoption. For example, if bikesharing is utilized for last mile travel, then a transfer fare could increase its usage and simultaneously promote public transit adoption [4]. A better understanding of trip patterns will also allow for advanced bicycle relocation strategies and more reliable service provision which, in turn, will make the system more attractive to users. The key challenge for any shared mobility system lies in the respective network complexity, noise and the resulting operational implications. To overcome this, we propose a novel, comprehensive and pervasive station-level characterization of the London network, based on spatio-temporal utilization features. Our framework extracts largely self-contained clusters which not only provide insight into mobility patterns, but also help with identifying bottlenecks and inefficiencies, and hence help decision makers to better understand supply and demand imbalances, plan operations and manage infrastructure. Comparing the explicitly non-spatial network model to the known geospatial structure of the system enables us to assess whether communities are a result of *space* (geography) or *place* (local features). We examine BSS rental data from London introducing a new approach to community detection in BSS networks and assess interactions between communities as well as the convergence of communities during different times-of-day.

The remainder of this paper is organized as follows. Section 2 reviews relevant studies of BSS and community analysis. A detailed description of the London BSS dataset is provided in Section 3. Section 4 describes the main methodologies and procedures involved in the data-driven analyses. Section 5 contains the discussion and concluding remarks.

### **Related work**

With the proliferation of smart data related to BSS there has been a significant amount of research dedicated to either improving our understanding of the BSS to support evidence-based policy making, or to perform logistic optimization methods for bicycle relocation. Based on Barcelona BSS data, [3] Froehlich et al. applied spatio-temporal analyses, clustering techniques and tested the performance of various machine learning algorithms. Kaltenbrunner [5] estimates station-level time-series using autoregressive predictive models on the same dataset. Other efforts include a characterization of the network based on the usage profiles of the stations. Vienna's BSS was analyzed to obtain distinct clusters using partitioning algorithms on usage time-series data in addition to a predictive method to forecast ridership volume [12]. London's BSS was examined to detect differences in the patterns of usage before and after the opening of the scheme to unregistered (casual) users and to identify commuters [6,15]. A big differentiating point between these studies is the type of data used: While some research only has access to availability data at station-level, others are able to use arrival and departure data, detailing every trip. This second set provides more detail as it captures information about periods of inactivity and activity, unlike availability data, which



**Figure 1** Locations of BSS stations in London

loses information when the net change in bikes at a station is small.

Another focus of analysis in BSS is the detection of groups of individual users or stations that exhibit a strong interdependence between one another. This allows for spatial aggregation and pattern extraction. Austwick et al. [15] analyze community structures in five urban BSS (London, UK; Boston, MA; Denver, CO; Minneapolis, MN; and Washington, DC), [1] Borgnat et al. apply Louvain community detection on data from Lyon. However, due to limited data availability, the researchers had to generate their own origin / destination (OD) matrices. Furthermore, their approach employs hierarchical community detection, which comes with shortcomings, especially when resolving the boundaries of different communities or relating nodes that do not share any connections [7]. Overall, we observe that existing literature on community structure detection in BSS lacks the use of non-modularity-based methods and more granular, empirical trip data. From a technical viewpoint, community detection using modularity maximization is known to be vulnerable to resolution restrictions, is limited to undirected information and assumes a process of endogenous network formation [2]. Furthermore, the interaction between extracted communities and the evolution of communities over time remain largely unexplored fields. With our paper, we seek to address these gaps in research.

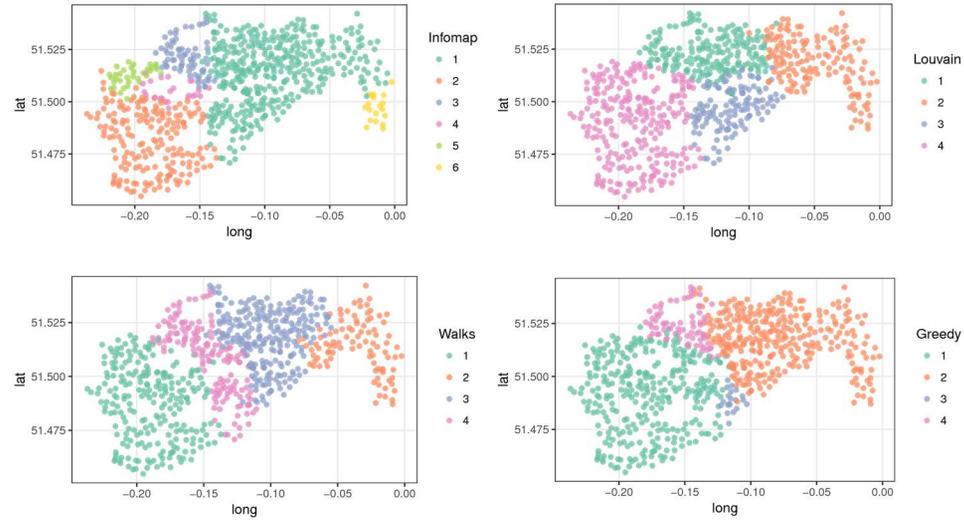
## Data

The data for our analysis comes from the Transport for London (TfL) Open Data API (accessible via: <https://api.tfl.gov.uk>) and contains information about the unique IDs for each bicycle, the names and IDs of

the origin and destination stations, a unique transaction ID (per trip) and the start/end times of each rental. The dataset covers every shared bicycle trip since 2012 and hence comprises millions of entries. Operating with such large amounts of data can be computationally expensive. We hence utilize an applicable subsample. First, however, we address some minor problems regarding the collection of trip information, that translates into missing retrieval for destination station IDs or trips without information about bicycle IDs. These issues are not temporally consistent across the dataset, with some months exhibiting higher error rates than others. To adapt the analysis accordingly, we select a small, particularly accurate interval where a crude cleaning of the data, i.e. a removal of incomplete entries, does not result in significant non-response bias. We clean our subset by removing the following entries:

- Trips that start or end at a repair station.
- Trips that do not report correct destinations and show a negative duration.
- Trips that do not report the bicycle ID.

Our final dataset is comprised of 1,469,945 unique shared bicycle trips in June and July 2014. While we also experimented with different periods, the selected interval reflects our overall observations best, while simultaneously containing a relatively small amount of noise and data errors. Weekends are disregarded, as their varying trip patterns and added noise could harm our analysis. Aggregated on 750 stations (see Figure 1), our data can be used to compile an Origin-Destination (OD) matrix. We can then formulate a graph  $G$  with each station describing a network node  $N_{\alpha}$ , linked to every other station in the network by a set of directed edges  $E_{\alpha}$  weighted by a flow  $w_{\alpha}$  equal to the number of trips observed.



**Figure 2** Comparison of community detection algorithms: BSS stations are colored according to their respective community assignment across the four techniques

**Algorithm 1:** Infomap algorithm

```

1: procedure Infomap( $G$ )
2:    $M(N_\alpha) \leftarrow M_\alpha$ 
3:   while  $\min L(M)$  do
4:     Order  $N_{1,2,\dots,n}$  randomly
5:     for Each  $N_\alpha$  do
6:       if  $\exists M(N_\alpha) \leftarrow M_\beta: L(M) \downarrow$ 
7:         then
8:            $M(N_\alpha) \leftarrow M_\beta$ 
9:         else
10:           $M(N_\alpha) \leftarrow M(N_\alpha)$ 
11:   return  $M(N_\alpha)$ 

```

**Methodology**

*Community detection*

Community detection techniques aim at reducing the complexity of a network to a degree that enables comprehensive insight into the underlying network structure. BSS are naturally suited for such applications. Simplifying BSS network descriptions and detecting clusters of stations that exchange many trips also have immediate operational implications by providing a valuable decision-support tool for network management and expansion. Yet, reliable results depend on an appropriate choice of the community detection algorithm. The most popular methods rely on modularity maximization; however, they do not seem to be applicable in our case, as outlined in the literature review. Most importantly, these methods assume an

underlying process of network formation which, in the case of station-based BSS, is not present. This problem also motivated the development of the *Infomap* algorithm, originally proposed by Rosvall & Bergstrom [10]. This method acknowledges that the system structure drives the flow in the system, leading to system-wide interdependencies. By partitioning the network, the length of the description of the movements can be longer or shorter (bigger and smaller cost of information). By choosing the partition that minimizes the description length, we find the division that provides the best representation of the community structures. Deploying the *Infomap* approach, we seek to partition the network nodes  $N_{1,2,\dots,n}$  into  $M_{1,2,\dots,m}$  modules by minimizing the information cost of describing the movements of a

random walker or, if available, the empirical flow (here the trips) through the network. This is implemented by the *map equation*:

$$L(M) = q_{\sim} H(Q) + \sum_{i=1}^m p_{\sim}^i H(\mathcal{P}^i),$$

where  $q_{\sim}$  gives the probability that the random walker leaves the current module,  $p_{\sim}^i$  gives the proportion the walker spends in the respective module  $\mathcal{P}^i$ ,  $H(Q)$  gives the index codebook entropy and  $H(\mathcal{P}^i)$  gives the module codebook entropy (Referring to Shannon’s source coding theorem [11]). Practically, this is carried out by assigning the modules  $M_{\alpha}$  to a neighboring module  $M_{\beta}$ , as long as this reduces  $L(M)$ . Dedicating a small fraction  $\tau$  of the probability flow randomly links every node in the network to every other node and hence prevents the random walker from becoming stuck. The *Infomap* algorithm can then be applied as outlined in *Algorithm 1*. The technical differences between the four approaches—discussed extensively in previous research [14]—manifest in the respective output communities. *Infomap* is the only method to detect known physical structures in London, such as Hyde Park and Canary Wharf. The first and largest community is (1) Central and East London, accounting for more than half of the trips in the network. It borders (2) South-West London, (3) Regent’s Park and (4) Hyde Park clusters in the West, which are the second, third and fourth largest clusters in terms of flow. These three clusters border the fifth largest (5) Notting Hill community. Finally, (6) Canary Wharf contains the least flow of any cluster and is remotely located in the South-East.

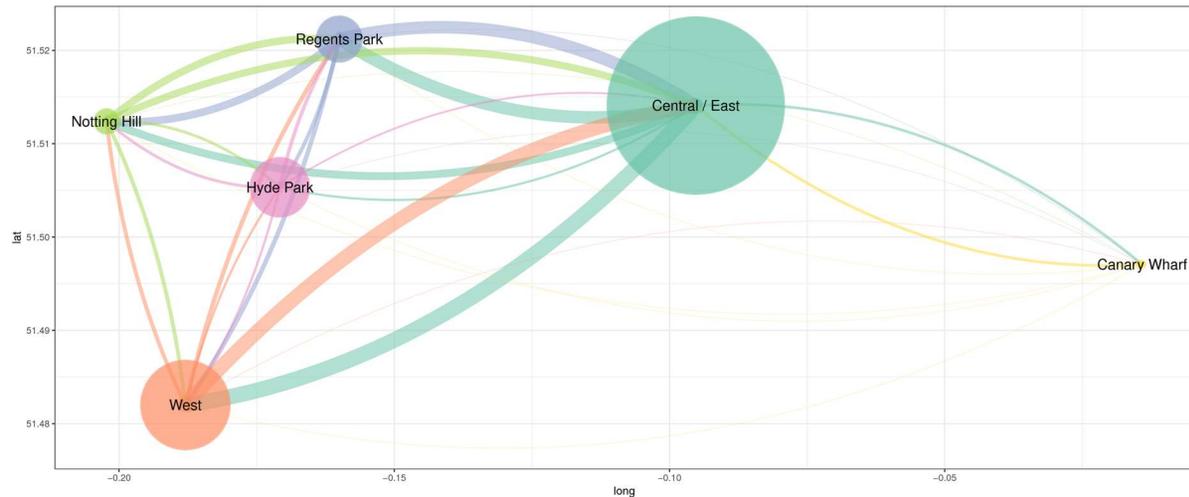
### *Community interactions*

Beyond the community detection, our results also enable insight into the interactions between the different communities. Around 75% of the observed trips start and end within the same cluster. Nevertheless, investigating the exchange of trips between clusters provides a deeper insight into the underlying mechanics of the observed system, particularly for the smaller, more interactive communities. Our simplified community network and the flow between communities are displayed in Figure 3. The size of the communities and the links respectively highlight the observed flow. The exact numbers of trips (within, outbound and inbound) are given in Table 1.

Cluster	Stations	Trips		
		<i>within</i>	<i>out</i>	<i>in</i>
Central / East (1)	408	760,404	112,263	118,146
West (2)	190	184,714	80,457	82,332
Regents Park (3)	71	48,259	75,618	71,990
Hyde Park (4)	26	80,354	63,617	60,289
Notting Hill (5)	35	17,481	30,443	29,084
Canary Wharf (6)	20	9,060	7,181	7,738

**Table 1** Summary of community cluster characteristics and interactions

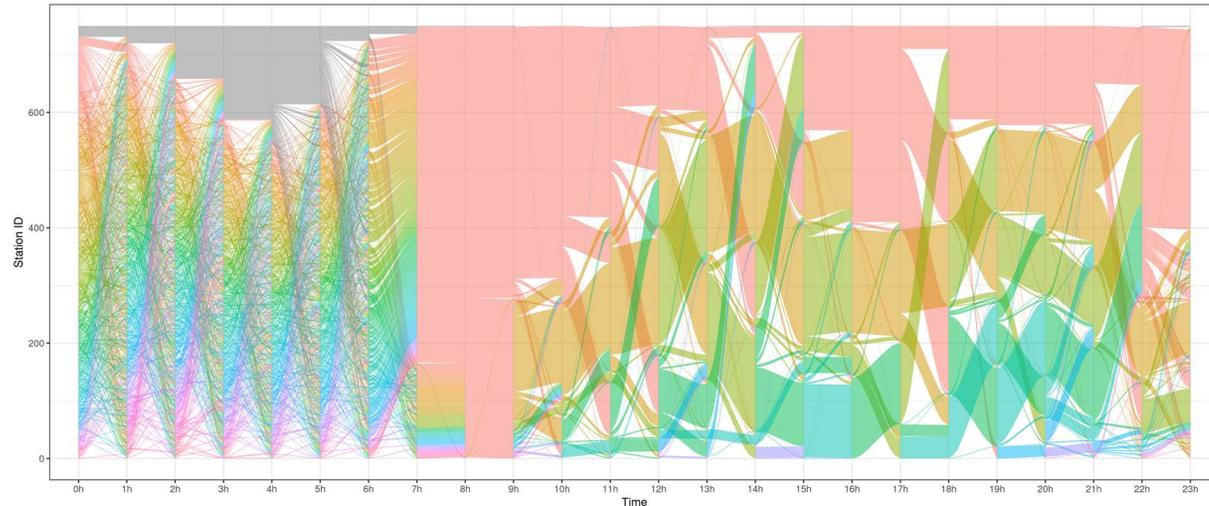
The Central London community accounts for around 50% of all trips in the network. Due to its large size, most trips occur within the community, while the connectivity with other clusters is particularly strong for the West London and Regents Park communities. This suggests that there might be additional hidden structures within the cluster that would allow for further



**Figure 3** Interactions and volume of London BSS communities: clusters are mapped at their geographic centroids. The size of point nodes and edges is scaled according to the observed flow within the community (nodes) and between the communities (edges)

simplification. The West London community behaves relatively similar, though having a higher share of trips interacting with other communities. The Regents Park and Notting Hill communities, both small in size, exhibit more interactive trips than within-cluster trips and hence suggest longer trip distance or special trip purposes. The last clusters—Hyde Park and Canary Wharf—are both small in size but nevertheless relatively isolated. For Hyde Park, a possible explanation is the specialized use of shared bicycles for leisure trips within the green-space. For Canary Wharf, this can be attributed to its remote geographic location on the Isle of Dogs, which reduces the attractiveness of bicycle trips to stations outside of the cluster. The community detection and interactions analysis enables us to gain novel insights into the spatial usage patterns of London's BSS. The noisy Central London cluster does

not seem to exhibit explicit community structures, which hints that the usage there is less community-driven and rather might be explained using temporal analyses (e.g. commuting peak times) or destination-based approaches. The smaller, more disconnected clusters, on the other hand, suggest a strong effect of community bounds on trip-making that might be explained looking into their respective location: while the Canary Wharf cluster is located in a business area where BSS trips might serve as last-mile connections to public transport, the emergence of the Hyde Park cluster within a large public green space suggests leisure activities. This seems especially likely as our observational period is during summer time, where mild weather conditions make parks particularly attractive. On the other hand, small but interactive clusters like Notting Hill and Regents park are based around mostly



**Figure 4** Community evolution over time: cluster assignment for each station and hour-of-day is given using color codes (grey for no assignment). Communities are ordered and colored top-to-bottom by size

residential areas and suggest the use of the BSS for commuting. As such, community detection and interaction together reflect London's physical environment and land use.

#### *Community dynamics*

While our previous network analyses presents novel insights into general system structures, aggregating the usage in such a large timescale results in the loss of temporal information about the way the network behaves at different times of the day. To provide a deeper understanding of the emergence and collapse of communities over time, we split our dataset into one-hour intervals for further examination. Apart from this, the methodology remains unchanged. We present the *Infomap* cluster assignments for each station and all 24 hours of the day in Figure 4. The results show the

evolution of the communities over the course of the day. It is apparent that during some hours---especially at nighttime---the noise in our data is considerably larger which prevents the algorithm from detecting community structures, leaving several hundreds of hardly relevant communities and even unassigned stations, due to the very low flow. This suggests a limitation of our analysis, but also relates to the general lack of observation during those unusual travel hours. From 7am, one dominant community abruptly emerges that subsumes almost all stations. This is due to the morning commuting peak, characterized by long trips connecting stations in residential areas or close to public transport facilities with the business districts. As the peak cluster disintegrates around 10am, three to four stable clusters emerge that resemble the general community structure outlined in the previous sections.

This period of stability contains the vast majority of observed trips and starts and ends with the respective morning and evening peak commuting times. From around 8pm, the daytime communities slowly disintegrate towards a full collapse at midnight. Looking at the spatial dimension, we can observe that during night-time, location does not seem to be of importance. Only starting from 10am, Central London emerges as a community, alongside clusters in Canary Wharf, Hyde Park and West London. These communities vaguely correspond to those extracted from our previous analyses (see Figure 2 and 3) and are mostly stable during daytime. Again, Canary Wharf stands out as the most isolated community, with almost all stations assigned to the same cluster from 9am to 9 pm.

### **Conclusion**

Each of our three methodological sections comes with novel findings and implications relevant to BSS users, providers and public authorities. (A) We present a new, information-theoretic method for BSS community detection that not only reflects known system features like directed links and exogenous network formation, but, unlike modularity-based approaches, is also able to detect known urban features like Hyde Park or Canary Wharf. We can infer valuable information on user behavior and the geographical boundaries of bikesharing trip-making. The findings may inform and motivate service adoption according to the detected communities. (B) To our knowledge, this is the first paper to explore bidirectional flow between BSS communities. The imbalances in trip flows are a crucial challenge for any shared mobility scheme as they often result in vehicles getting stuck in areas characterized by high destination attractiveness and low origin attractiveness. While the communities are largely self-

contained, trips between communities and particularly their imbalances can shed new light on this issue. (C) We present the first study to explore the emergence and collapse of BSS communities during the course of a day, thus evaluating noise and underlying mechanics in our system. We find that at nighttime—characterized by substantially less trips—the algorithm cannot detect clear community structures as the trips do not follow any distinct patterns. During daytime, communities emerge clearly structured during the morning commuting peak and stabilize afterwards. There is some spatio-temporal fluctuations between the clusters, which again seem to be mostly driven by peak commuting hours. Contrarily, we find communities of remote geographic location (Canary Wharf) and specific leisure usage (Hyde Park) to be the most consistent. Beginning with the afternoon commuting peak, communities contract and eventually collapse around midnight. By exposing structure-over-time, we examine how community presence is driven by spatio-temporal dynamics. Time-sensitive events like commuting hours mostly affect areas around public transport and business districts, while e.g. parks remain rather stable. These findings enable focused policies to address problems like supply shortages or congestion.

Altogether, our research reinforces the argument that BSS are inherently spatio-temporal, dynamic systems while also raises new questions regarding the driving factors of our observations. Further studies should extend the underlying questions from purely unsupervised learning problems to represent other layers of the urban system. For instance, research have recently shown that urban amenities or weather data can help with contextualizing patterns in shared mobility systems [8,13]. Our findings also suggest a

strong interconnection of said features with the bikesharing network, which implies that more holistic approaches are needed to draw meaningful operational and political conclusions.

### Acknowledgements

The authors gratefully acknowledge funding from the UK Engineering and Physical Sciences Research Council, the EPSRC Centre for Doctoral Training in Urban Science (EPSRC grant no. EP/L016400/1); The Alan Turing Institute (EPSRC grant no. EP/N510129/1).

### References

1. Pierre Borgnat, Patrice Abry, Patrick Flandrin, Celine Robardet, Jean-Baptiste Rouquier, and Eric Fleury. 2011. Shared bicycles in a city: a signal processing and data analysis perspective. *Advances in Complex Systems* 14, 3: 415–438.
2. Santo Fortunato and Marc Barthelemy. 2006. Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America*.
3. Jon Froehlich, Joachim Neumann, and Nuria Oliver. 2009. Sensing and predicting the pulse of the city through shared bicycling. *IJCAI International Joint Conference on Artificial Intelligence*, 1420–1426.
4. Zhan Guo and Nigel H.M. Wilson. 2011. Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. *Transportation Research Part A: Policy and Practice* 45, 2: 91–104.
5. Andreas Kaltenbrunner, Rodrigo Meza, Jens Grivolla, Joan Codina, and Rafael Banchs. 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.
6. Neal Lathia, Saniul Ahmed, and Licia Capra. 2012. Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies* 22: 88–102.
7. M. E J Newman. 2004. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter* 38, 2: 321–330.
8. Svenja Reiss and Klaus Bogenberger. 2015. GPS-Data Analysis of Munich’s Free-Floating Bike Sharing System and Application of an Operator-based Relocation Strategy. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 584–589.
9. David Rojas-Rueda, Audrey de Nazelle, Marko Tainio, and Mark J Nieuwenhuijsen. 2011. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *BMJ* 343:d4521.
10. M. Rosvall and C. T. Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105, 4: 1118–1123.
11. C. E. Shannon. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 3: 379–423.
12. Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. 2011. Understanding bike-sharing systems using Data Mining: Exploring activity patterns. *Procedia - Social and Behavioral Sciences*, 514–523.
13. Christoph Willing, Konstantin Klemmer, Tobias Brandt, and Dirk Neumann. 2017. Moving in time and space – Location intelligence for carsharing decision support. *Decision Support Systems* 99: 75–85.
14. Zhao Yang, René Algesheimer, and Claudio J. Tessone. 2016. A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports* 6.
15. Martin Zaltz Austwick, Oliver O’Brien, Emanuele Strano, and Matheus Viana. 2013. The Structure of Spatial Networks and Communities in Bicycle Sharing Systems. *PLoS ONE* 8, 9.