

Graph-Based Optimisation of Network Expansion in a Dockless Bike Sharing System

Mark Roantree[✉]

Insight Centre for Data Analytics
School of Computing, Dublin City University
Dublin, Ireland
Email: mark.roantree@dcu.ie

Dinh Viet Cuong

Insight Centre for Data Analytics
School of Computing, Dublin City University
Dublin, Ireland
Email: dinh.cuong2@mail.dcu.ie

Niamh Murphy

School of Computing
Dublin City University
Dublin, Ireland
Email: niamh.murphy88@mail.dcu.ie

Vuong M. Ngo[✉]

School of Computing
Dublin City University, Ireland
Ho Chi Minh City Open University, Vietnam
Email: vuong.ngo@dcu.ie and vuong.nm@ou.edu.vn

Abstract—Bike-sharing systems (BSSs) are deployed in over a thousand cities worldwide and play an important role in many urban transportation systems. BSSs alleviate congestion, reduce pollution and promote physical exercise. It is essential to explore the spatiotemporal patterns of bike-sharing demand, as well as the factors that influence these patterns, in order to optimise system operational efficiency. In this study, an optimised geo-temporal graph is constructed using trip data from Moby Bikes, a dockless BSS operator. The process of optimising the graph unveiled prime locations for erecting new stations during future expansions of the BSS. The Louvain algorithm, a community detection technique, is employed to uncover usage patterns at different levels of temporal granularity. The community detection results reveal largely self-contained sub-networks that exhibit similar usage patterns at their respective levels of temporal granularity. Overall, this study reinforces that BSSs are intrinsically spatiotemporal systems, with community presence driven by spatiotemporal dynamics. These findings may aid operators in improving redistribution efficiency.

Index Terms—Bike Sharing, Urban Movement, Graph Theory, Clustering, Community Detection, spatiotemporal Analysis

I. INTRODUCTION

Commercially viable technology capable of tracking the movement of bikes has enabled the rapid expansion of Bike Sharing Systems (BSS) in cities across many countries [1]. By August 2022, the number of bikes in these schemes had grown to almost 9 million across 1,914 systems, spanning 92 countries and 1,590 cities [2]. This substantial growth has transformed the bike sharing market into a high performing industry, which according to Statista [3] has a valuation of 9.46 billion US dollars. However, as bike networks grow, improving their efficiency to meet the demands of users is crucial for the survival of BSS operators [4]. One of the primary objectives of this endeavor is to identify the optimal number of stations in suitable locations, a task that necessitates spatial analysis employing appropriate tools and methods. Graph networks have been widely utilized in various applications, including

textile structure [5], [6], ontology [7], [8], [9], knowledge-based systems [10], [11] and electronic health records [12]. In the context of bike-sharing systems, graph-based approaches have been employed to model locations, trips, and associated time intervals, facilitating the investigation of system dynamics. In these projects, bike journeys were modelled as network structures and through the analysis of these networks, the inherent characteristics of journey dynamics were studied [13], [14], [15], [16]. We examine these dynamics by representing spatial locations as nodes, with trips forming the edges between starting and ending locations.

A. Problem Statement

An in-depth understanding of usage patterns can greatly improve the impact of management strategies, making the service more attractive to users. However, two key obstacles exist: the availability of data with the necessary features and the ability to model the complexities of the network dynamics [17]. The first problem requires either some form of online data acquisition such as [18] where live data is processed into temporally aware data marts, or, collaboration with the bike rental company. The second problem requires a comprehensive spatiotemporal characterisation of the travel patterns in order to identify high usage locations with no station on the network.

The data issue was resolved by collaborating with Moby Bikes¹ who, after anonymisation, made all data available for the purpose of this research. Trips made using Moby's fleet generated significant quantities of time and location specific data to enable the study of travel behaviour and mobility at trip level granularity. The multi-dimensional, interconnected nature of the spatiotemporal data motivated the use of a graph database to resolve the second issue. The Neo4j [19] database was used to construct a number of different graph networks and to facilitate graph based (community detection) clustering as a

¹<https://mobybikes.com/>

means of understanding network traffic and to ensure that newly created stations observed the same activity patterns as existing stations. Given the data and infrastructural solution available, the research questions are: can trip data be modelled so as to identify optimal candidate locations for network expansion (new stations)? can spatial analysis be facilitated at different levels of temporal granularity? and finally, is there a means of validating new stations so that they are not outlier stations with activity patterns unrepresentative of other nodes in the network (outliers)?

B. Contribution and Paper Structure

The Moby network comprises 92 fixed stations and in prior work [20], a distance function (discussed later as eq. 2) was employed to reassign non-station source and destination nodes to their closest fixed station. This enabled the construction of spatiotemporal graphs to facilitate time-based comparison of different stations and the construction of station profiles to model their interactions with all other stations. However, it could not identify precise locations for the placement of new fixed stations to reduce traffic bottlenecks. Analysis showed that many virtual stations could be created to accommodate the lack of fixed stations in key locations but this represents uncontrolled expansion of the network. The contribution of this work can be articulated as follows:

- Optimizing the expansion of the network through a novel algorithm for identifying new station locations;
- A methodology for facilitating spatiotemporal analysis using graphs with different temporal granularities;
- A validation mechanism using community detection to ensure that new stations had similar behaviour or characteristics as existing stations.

Paper Structure. The remainder of this paper is organised as follows. Section 2 briefly reviews relevant studies of BSSs with a spatial, temporal, or graph theory focus. A detailed description of the dataset and the preprocessing steps performed are presented in Section 3. Section 4 describes the main methodologies. A summary of the results and a validation of the results are contained in Section 5. Finally, Section 6 concludes the paper with a discussion and recommendations for future work.

II. RELATED WORK

In recent years there has been a surge in research interest in the spatiotemporal characteristics of BSSs. Sun et al. [21] studied a number of factors that affected BSS usage and found that BSSs' usage increased during peak hours on weekdays, indicating that BSSs are predominantly used for commuting purposes. Zhou [22] uncovered a similar finding by constructing a bike flow similarity graph of a Chicago BSS and used a fast greedy algorithm to detect spatial communities with unique travel patterns on weekdays and weekends. Munoz-Mendez et al. [17] developed a novel modified Infomap clustering approach to capture the spatiotemporal patterns observed in a London-based BSS. Through clustering, self-contained and

interconnected community structures were discovered, with approximately 75% of the observed trips starting and ending within the same community [17]. In separate work, the authors used global metrics to capture the overall structure of the network while local metrics were used to identify prominent nodes across the network. Commonly used metrics encompass the count of nodes, edges, degree, and strength, which signify the level of activity and connectivity within a given location [23], [24].

The previously discussed research focuses on conventional dock-based BSSs. The usefulness of these studies may be limited as users of dockless BSSs are not confined to fixed stations, meaning the spatial distribution is drastically different. There are fewer relevant studies that have analysed the spatiotemporal characteristics of dockless BSSs on real-world trip data. In [25], Lin et al. constructed a spatially embedded network using dockless BSS data from Beijing. Origin and destination locations were represented by nodes and trip flows by edges. Spatial community detection was used to reduce the complexity of the network. On average, 77% of trips began and ended inside the same community, indicating that these communities were largely self-contained.

The study conducted by Austwick et al. [26] revealed commonalities in the distribution of strength and edge weight within networks generated from diverse bike-sharing systems. In similar work, researchers [16], [27] recommended incorporating a more holistic set of network properties to capture different network features including connectivity metrics such as degree and node flux, spatial distribution using the clustering coefficient algorithm, network stability or connectivity (centrality algorithms), network efficiency, and equity (using the Gini coefficient). More traditional centrality metrics such as *betweenness* and *PageRank*, have also been employed to describe a network's features [16].

Limited research has also been conducted using data from Dublin based BSSs. Research has been performed using publicly available data from DublinBikes [28], [29]. These studies classify bike stations according to users' mobility patterns and analyse usage patterns of stations, respectively.

Previous studies have illustrated the suitability of representing BSS data as a graph. BSSs in cities around the world have been shown to exhibit temporal patterns and largely self-contained sub-networks. However, the majority of the existing literature focuses on dock-based BSS. The few relevant studies that involve dockless BSSs do not group rental locations according to both usage patterns and temporal features. This paper seeks to address this gap in research by analysing a Dublin-based BSS to assess whether observed communities are a result of spatial or temporal features.

Community analysis, a recurrent theme in network research, plays a pivotal role in comprehending a network's structure. Community detection algorithms categorize nodes within the network into distinct communities, where nodes exhibit stronger internal connections than external ones. The Louvain algorithm, as discussed in [30], stands as the most

commonly employed method in this realm. However, Shi et al. [31] observed that different algorithms yield diverse community characteristics depending on the measurement criteria. Despite the advancement in network analysis, these methods are predominantly applied to networks with edge connections representing direct interactions. We aim to extend similar methodologies, including visualization and the utilization of complex metrics such as strength, closeness, betweenness, local clustering coefficients, and community detection, to more intricate correlation-based networks.

III. DATA

The data used in this research has been collected and provided by Moby Bikes [32]. Moby’s bike fleet consists of 95 fully electric bikes, and, with a *dockless* mode of operation, users can collect or drop bikes at any suitable public location. However, to reduce operational overheads, Moby introduced fixed charging stations where customers are financially incentivised to return rented bikes to these fixed locations which we refer to as *stations*.

The GPS units on the bikes have created a large amount of individual level data which if properly managed, can be modelled as a spatiotemporal graph network. The data collected is stored in two SQL tables, Rental and Location. The Rental table contains information about each logged rental during the period from 3rd January 2020 to 19th September 2021. There are 62,324 records in the original Rental table. The Location table provides finer-grained information on each of the locations a bike was either rented from or returned to, during the same 21 month period, with a total of 14,239 records. Interestingly, COVID-19 was declared a pandemic by the World Health Organisation in March 2020, meaning the majority of the data used was created during the pandemic.

There were some minor issues with the dataset: references to non-existent data and inadmissible or spurious locations. The dataset was cleaned by removing the following entries:

- Locations outside Dublin (as rentals should only involve journeys within city locations) and rentals that started or ended at these locations.
- Locations that are not on land and associated rentals.
- Locations that are missing latitude or longitude coordinates and associated rentals.
- Rentals that do not report a Rental Location ID or a Return Location ID.
- Rentals with a Rental Location ID or a Return Location ID that is not in the Location table.
- Location IDs in the Locations table that are not referenced in the Rental table.

TABLE I
DATASET OVERVIEW

Measures	Original Dataset	Cleaned Dataset
Duration of data	Jan 2020 - Sept 2021 (~21 months)	
#stations	95	92
#rental	62,324	61,872
#location	14,239	14,156

After removing these entries, the Rental table comprised 61,872 entries and the Location table comprised 14,156 entries across 92 stations, as Table I.

IV. METHODOLOGY

A methodology was devised as a three-step process: (1) graph construction, involving a preprocessing step to determine the spread of trips and begin/ending locations for generating candidate stations; (2) a process to select suitable stations from the candidate stations, merging them with pre-existing stations. This process is guided by a selection and ranking algorithm; and (3) community detection to understand the impact of the new stations on the network.

A. Graph Construction

The majority of bike-sharing systems operate on a dock or station-based model, where customers are required to rent and return bikes at designated stations. Consequently, it is natural to regard these stations as nodes within the network. In contrast, dockless systems afford users the flexibility to pick up and drop off bikes anywhere, without the constraint of fixed stations. If we consider these sources and destinations to be a virtual station, this creates a graph with an unlimited number of stations, creating a complex network where nodes may potentially have very low degree scores as very few trips begin or end at precise locations. This requires a method to optimise the graph by reducing the number of nodes (virtual stations) in the network. While our previous work [20] reassigned non-station source and destination nodes to their closest fixed station, the goal here is the optimised expansion of the network before this reassignment task, in order to retain as much spatial information regarding trips, as possible.

A novel strategy was developed to use hierarchical agglomerative clustering (HAC) to group geographically close (starting or ending trip) locations with logical thresholds defined to identify *candidate* locations for new stations that consider *current* fixed stations. HAC is a bottom-up clustering algorithm that begins by considering each data point as an individual cluster and merges the clusters based on a distance measurement until a single cluster containing all of the data points emerges [33]. The distance metric used for geographic locations was the Haversine distance, shown in equation 1. The function calculates the shortest distance between two points on a sphere and was chosen as it remains accurate for computations at small distances unlike calculations based on the spherical law of cosine [34].

$$d = 2R \arcsin \sqrt{\sin^2\left(\frac{\varphi_1 - \varphi_2}{2}\right) + \cos \varphi_1 \cos \varphi_2 \sin^2\left(\frac{\lambda_1 - \lambda_2}{2}\right)} \quad (1)$$

Equation 1 shows the distance function where R is the radius of Earth; φ_1 and λ_1 represent latitude and longitude for Location A; φ_2 and λ_2 represent latitude and longitude for Location B; and d is the distance between two locations. The Complete Linkage criterion was used to determine the distance between two clusters based on the largest distance over all possible pairs [35].

Preprocessing. Pre-existing fixed stations were set as immovable locations and set as their own group's centroid. To adhere to the criterion of groups' centroids being at least 50 metres apart, any location that was within a 50-metre radius of a fixed station was assigned to that station's group and was excluded from clustering.

B. Station Ranking and Selection

After conducting several analyses of the data, particularly focusing on instances where a high number of distinct locations were less than three meters apart, we established a set of parameter settings before initiating the clustering process. These settings formed a set of rules that governed the output of the clustering algorithm.

Rule 1: Cluster-Boundary.

The distance between 2 locations L_1 and L_2 inside a given cluster C may not exceed 100m.

Rule 2: Cluster-Proximity.

The distance between any pair of cluster centroids C_{ci} and C_{cj} cannot be less than 50m.

Rule 3: Degree-Threshold.

The degree D for candidate station S cannot be less than S_{min} , the minimum degree from the set of fixed stations.

Rule 4: Secondary-Distance.

A second threshold was set to ensure that nodes which represent a cluster are not within 250m of a station.

The procedure for selecting stations from candidate stations is outlined below:

- 1) Non-fixed station locations underwent clustering. The trips starting and ending at each location in each cluster.
- 2) As rule conflicts will occur if all candidates become fixed stations, clusters are ranked in descending order for degree metrics. Before each candidate station is converted to a fixed station, it is checked against each of the rules. Unless all 4 rules are true, the candidate station is rejected.
- 3) When the process completes, unconverted candidate locations are reassigned to the nearest station.

C. Community Detection

In order to understand the behaviour of new stations, a community detection algorithm was used to study the spatiotemporal patterns at different levels of temporal granularity. Community detection was chosen as it can identify sub-regions, in which the nodes are more strongly connected to one another than to nodes in other sub-regions. Thus, community detection was used to reduce the complexity of the network and to facilitate the study of the network's underlying structure as new stations (nodes) are added. Furthermore, it can serve as a validation of the overall network through the verification of good partitions [36]. To ensure a robust validation, community detection is performed at three different levels of temporal granularity across three different network structures.

Temporal Granularity.

- T_{Null} : includes no temporal features.
- T_{Day} : uses the day of the week that trips took place.

- T_{Hour} : uses the time of day that trips began.

The objective of experimenting with different levels of temporal granularity using the inherently spatial data was to detect clusters of stations that exchange many trips and that have similar temporal and/or spatial characteristics. Thus, providing a better understanding of usage patterns and a valuable decision-support tool for fleet management.

The Louvain algorithm is a popular greedy community detection algorithm. This algorithm was chosen for community detection due to its rapid convergence properties, high modularity, hierarchical partitioning and its ability to incorporate weighted edges [37]. The Luovain algorithm presented in Algorithm 1 was run on three slightly different bidirectional graphs.

Algorithm 1 Station Selection Algorithm

Input: List of new station candidates; List of pre-existing stations

Output: Selected new stations

```

1: Threshold  $\leftarrow$  minimum degree of pre-existing stations
2: for each station candidate do
3:   score of the station  $\leftarrow$  degree
4:   if degree < Threshold then
5:     score of the station  $\leftarrow$  0
6:   else if distance to the nearest pre-existing station  $\leq$  0.25
7:     then
8:       score of the station  $\leftarrow$  0
9:   end if
10: end for
11: repeat
12:   for each pair of station candidates with score > 0 do
13:     if distance between the two stations  $\leq$  0.25 then
14:       score of the lower-degree station  $\leftarrow$  0
15:     end if
16:   end for
17: until no station candidates with score > 0 are near each other
18: Sort station candidates based on the score
19: return station candidates with score > 0 as new stations

```

Network Structures.

Based on 3 temporal granularities T_{Null} , T_{Day} and T_{Hour} , we have 3 structured graphs as follows:

- G_{Basic} was constructed with stations represented by nodes, and trips represented by edges, weighted by the number of trips.
- G_{Day} has a similar node structure but each trip is now represented by a unique edge with a property denoting the day of the week that the trip took place.
- G_{Hour} is similar to G_{Day} but each edge contains a property indicating the time of the day that the trip started.

Modularity is the objective function used by the Louvain algorithm. It was also used to evaluate the resulting community structures. Modularity quantifies the quality of an assignment of nodes to a community by comparing the number of edges

that are within communities to the number of edges expected if the edges were placed at random [38]. Modularity scores are scaled between -1 and +1, with a positive score indicating the presence of a community structure.

$$Q = \sum_{c_i \in C} \left[\frac{\sum_{in}^{c_i}}{2m} - \frac{(\sum_{tot}^{c_i})^2}{4m^2} \right] \quad (2)$$

The calculation for modularity (denoted by Q) is shown in equation 2, where C is the set of communities; m is the ratio of edges within the community to the total number of edges in the graph; $\sum_{in}^{c_i}$ is the sum of edges belonging to community c_i over all of the nodes in the community; and $\sum_{tot}^{c_i}$ is the sum of the degree of all nodes in community c_i .

V. RESULTS AND DISCUSSION

A. Candidate Stations

Upon completing the graph construction outlined in Section IV-A, 1,172 clusters are generated, yielding a total of 1,080 potential locations for new stations, in addition to the existing 92 stations. The candidate graph, incorporating candidate stations, was aggregated and represented by weighted edges. The resulting graph is illustrated in Figure 1, and the corresponding data is summarized in Table II.

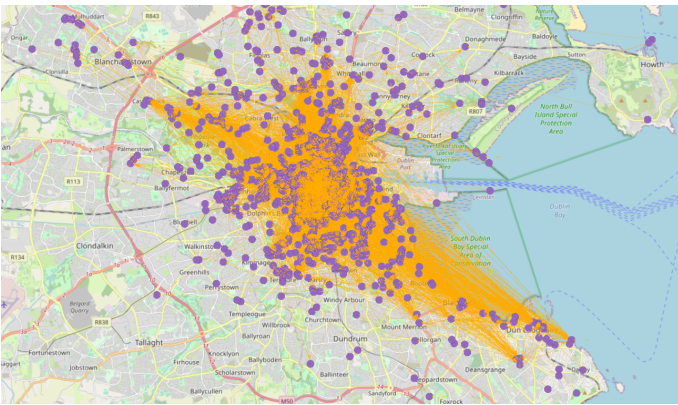


Fig. 1. The candidate graph generated by HAC, including the pre-existing stations. Nodes are shown in purple and edges in yellow.

TABLE II
THE DETAILS OF THE CANDIDATE GRAPH GENERATED BY HAC

Measures	Numbers
#nodes	1,172
#undirected edges	8,240
#undirected edges (no loops)	7,820
#directed edges	16,042
#directed edges (no loops)	15,604
#trips	61,872

B. Selected New Stations

Following the execution of our station ranking and selection algorithm in Section IV-B, a selected graph is generated, introducing 146 new stations and raising the overall count to 238 stations. The augmented graph is visualized in Figure 2 and detailed in Table III. All trips from non-selected stations

were redirected to the nearest existing station, ensuring that the total number of trips remains unchanged. Notably, the newly added stations are predominantly concentrated around Dublin City Centre, extending into the adjacent suburbs beyond the positions of the existing stations.

TABLE III
DETAILS OF THE SELECTED GRAPH

	Stations	Trips		Edges	
		From	To	From	To
Pre-existing	92	54,670	54,727	6,437	6,310
Selected	146	7,202	7,145	2,072	2,199
Total	238	61,872		8,509	

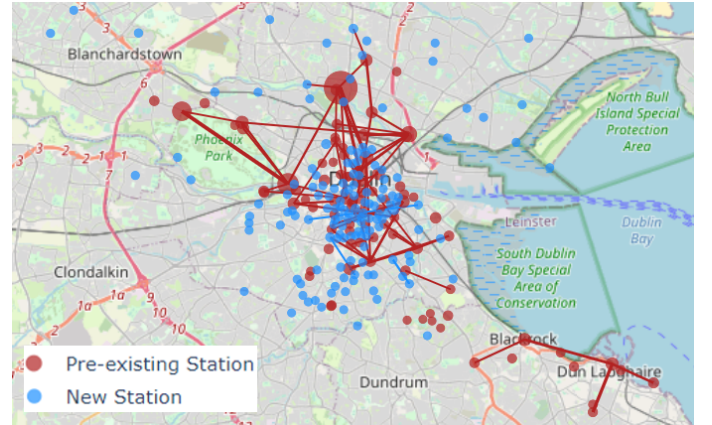


Fig. 2. The selected graph includes pre-existing stations and selected new stations. Node size is scaled according to the number of self-contained trips (self-edges). Edge width is scaled according to the number of directed trips between nodes (out-edges). Only edges with a weight in the top 1% percentile of weights are shown.

C. Community Detection

We need to be able to distinguish between new and existing stations. Are they spread across all communities? Can we update the table to reflect the location of these stations?

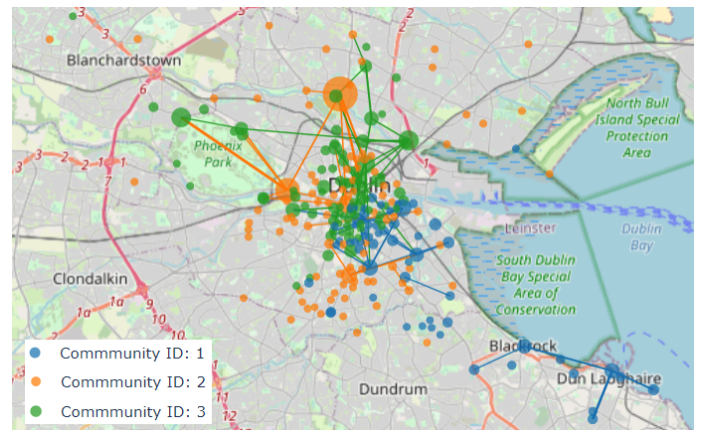


Fig. 3. Community detection for G_{Basic} . Stations are coloured according to their respective community assignment.

1) *Trips*: The first round of community detection used the G_{Basic} graph with no temporal information. The Louvain

TABLE IV
DETAILED INFORMATION ABOUT COMMUNITIES IN G_{Basic}

Community		Stations			Trips			
ID	Color	Old	New	Total	Within	Out	In	Total
1	Blue	40	18	58	12,012	5,238	5,255	22,505
2	Orange	4	94	98	9,158	4,078	3,995	17,231
3	Green	48	34	82	24,494	6,892	6,958	38,344

algorithm yielded three communities with a modularity score of 0.25. Given the low number of communities detected, this modularity score indicates that the sub-networks are non-trivial. The resulting communities are displayed in Figure 3. These communities have unique spatial properties. Stations belonging to the blue community are exclusively on the southside of Dublin; stations belonging to the orange community are typically less central than the other communities' stations and are spread around the suburbs of Dublin; while the green community's stations are generally in the city centre or are on the northside of Dublin.

The results of this community detection provided insight into the interactions between different communities, with a summary of the number of trips for each community provided in Table IV. For each community, this table details the exact number of trips that started and ended in that community (within), the number of trips that started in the community but ended in another (out), and the number of trips that started in another community but ended in the community (in). Approximately 74% of the trips start and end in the same community. This finding is consistent with previous studies into the self-contained nature of BSSs in London and Beijing where 75% and 77% of the trips started and ended in the same communities [17], [25].

Around 50% of all trips in the network start in the green community which is unsurprising as it is the most centrally located community. This community also has the highest proportion of within-community trips whereas the other 2 communities have a similar lower proportion of within-community trips. This could suggest trips involving stations in these 2 communities cover longer distances or are used for special purposes.

2) *Trips and Day of Week*: The second round of community detection used the G_{Day} network, incorporating the day of the week that a trip took place. On this occasion, 7 community structures were identified, with a modularity score of 0.32. The communities are displayed in Figure 4 and Table V, while the proportion of each community's trips that took place on each day of the week is shown in Figure 5.

Distinct temporal patterns can be seen among the larger communities. For instance, usage was lowest during the weekends in Community 2, Community 4, and Community 6. Bikes in these communities are likely largely used for weekday commuting purposes. These uncovered patterns could be used to assist with fleet re-balancing strategies. Conversely, usage peaked on Saturday in Community 1, Community 3, and Community 7. This could be explained by these communities' geographical locations and their stations' proximity to popular weekend spots. Community 1 is mainly composed of stations that are

within close proximity to the Phoenix Park and many stations in Community 7 are within a short distance of Blackrock/Dún Laoghaire. These uncovered patterns could be used to assist with fleet re-balancing strategies. For example, bikes could be moved from Communities 2, 4, and 6 to Communities 1, 3, and 7 each Friday night to prepare for the shift in demand over the weekend.

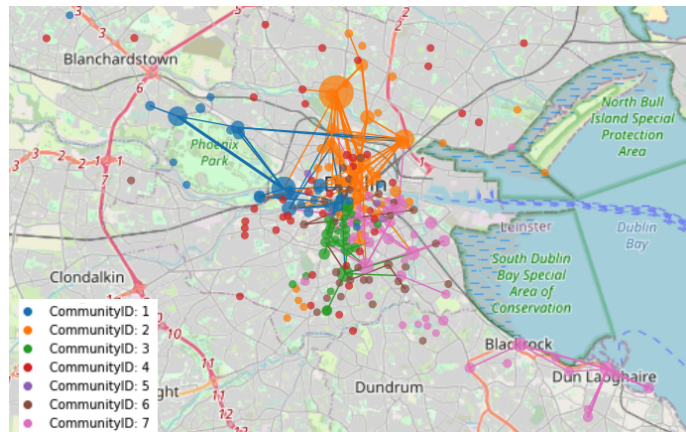


Fig. 4. Community detection for G_{Day} .

TABLE V
DETAILED INFORMATION ABOUT COMMUNITIES IN G_{Day}

Community		Station			Trips			
ID	Color	Old	New	Total	Within	Out	In	Total
1	Blue	15	16	31	8,517	3,516	3,522	15,555
2	Orange	0	22	22	551	227	238	1,016
3	Green	14	16	30	3,983	3,995	4,049	12,027
4	Red	0	27	27	551	179	170	900
5	Purple	36	16	52	11,555	4,949	4,933	21,437
6	Brown	0	32	32	1,411	450	414	2,275
7	Pink	27	17	44	16,328	5,660	5,650	27,638

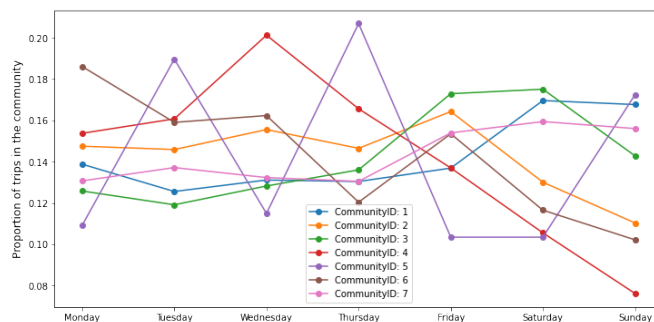


Fig. 5. Daily travel patterns per community in G_{Day}

3) *Trips and Time of Day*: The final round of community detection used the G_{Hour} network, which utilised the time of the day that trips began, detecting 10 communities with a strong modularity score of 0.54. These communities provide further support that the network is heavily influenced by spatiotemporal patterns. The communities are displayed in Figure 6 and Table VI. While, a breakdown of usage per hour of the day is shown in Figure 7.

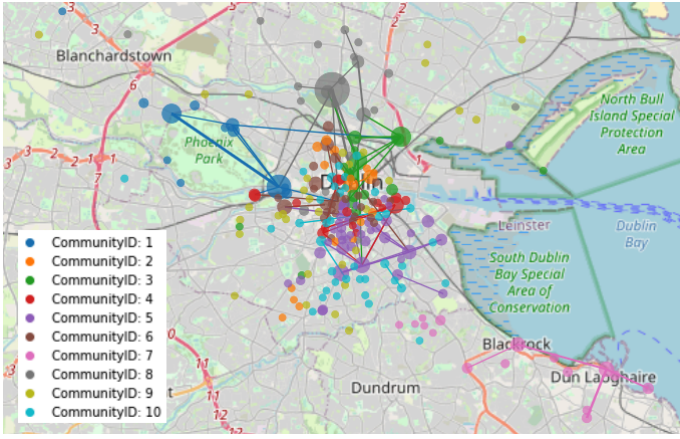


Fig. 6. Community detection for G_{Hour}

TABLE VI
DETAILED INFORMATION ABOUT COMMUNITIES IN G_{Hour}

Community		Station			Trips			
ID	Color	Old	New	Total	Within	Out	In	Total
1	Blue	9	4	13	5,422	1,706	1,704	8,832
2	Orange	13	11	24	1,774	1,930	1,944	5,648
3	Green	11	9	20	4,762	4,062	4,083	12,907
4	Red	10	9	19	2,379	2,833	2,825	8,037
5	Purple	14	0	14	8,313	4,974	4,991	18,278
6	Brown	15	14	29	3,234	3,613	3,656	10,503
7	Pink	6	18	24	4,186	1,161	1,175	6,522
8	Gray	9	17	26	5,450	2,310	2,256	10,016
9	Olive	1	30	31	767	221	207	1,195
10	Cyan	4	34	38	1,912	863	832	3,607

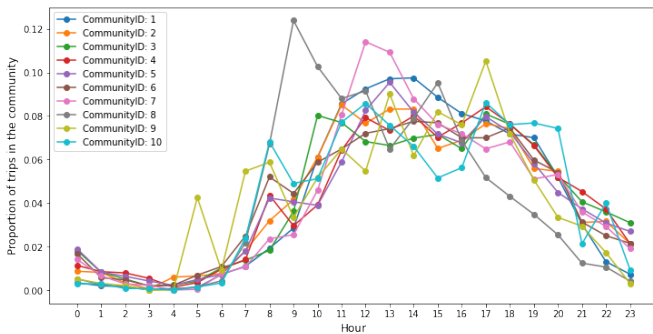


Fig. 7. Hourly Travel patterns per community in G_{Hour} .

Communities that experience a spike in demand between 7 am and 9 am and again at around 5 pm, such as Community 9 and Community 10, are likely mainly used by commuters. These communities have a large proportion of stations in the suburbs and/or in the city centre. On the other hand, usage in Communities 1 and 7 spikes at around midday, and are comprised of stations around the Phoenix Park and Dún Laoghaire, respectively.

VI. CONCLUSIONS

This study utilised Moby Bike trip data to reveal temporal and spatial patterns within the complex BSS network. The geospatial nature of the data motivated the use of graph theory to accomplish this objective. This study was initially impeded by the sheer number of locations and software limitations. While research with similar objectives exists, there are a limited number of research papers available that examine the usage patterns of dockless BSSs. Therefore, a novel approach to intelligently condense the size of the network using hierarchical clustering and logical thresholds was developed. The results from this network optimisation process were validated visually and by comparing the number of new stations to the number of pre-existing stations. The output of the optimisation process could potentially be used by system operators to identify locations for new fixed stations.

Community detection at three different levels of temporal granularity was performed on the optimised graph. The temporal dependence of the network is exposed by examining the community structures. Spatiotemporal fluctuations are observed in the detected communities using the day of the week and the time of day. Peak usage appears to be driven by commuting hours and specific leisure usage. Furthermore, the largely self-contained nature of the resulting communities, revealed by their strong modularity scores, suggests that bikes are not freely flowing between communities, which emphasises the importance of redistributing the bikes. These findings enable focused policies to optimise fleet rebalancing strategies and to meet user demand.

The outcomes of this research enables evidence-based policies to address network expansion and supply shortages. However, this study has several limitations. The criterion for clustering, that no two locations in a single cluster should exceed 100 metres, and the threshold that a new station must be at least 250 meters away from all other stations, were not motivated by empirical evidence. Instead, these distances were set based on pragmatic reasoning. Another limitation of this study is that it does not explore external factors such as weather or urban amenities that influence BSS usage. This study suggests an interconnection between leisure spots and bike-sharing usage patterns, which implies that a more holistic approach would be useful in uncovering additional network dependencies. Finally, this work did not experiment with different community detection algorithms.

Future studies should compare the results of a range of community detection algorithms, such as the Infomap algorithm and the Label Propagation algorithm. Further research should also investigate the effect of different graph optimisation strategies on community detection, particularly if more computational resources are available to allow for larger graphs.

Funding Acknowledgement. This work was supported by Science Foundation Ireland through the Insight Centre for Data Analytics (SFI/12/RC/2289_P2) and the Vistamilk SFI Research Centre (SFI/16/RC/3835).

REFERENCES

- [1] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258–276, 2018.
- [2] Bike-Sharing World Map Team, *The Meddin Bike-sharing World Map Report (December 2022 edition)*. PBSC Urban Solutions, 2022.
- [3] Statista, "Bike sharing worldwide," <https://www.statista.com/outlook/mo/shared-mobility/shared-rides/bike-sharing/worldwide>, 2024, retrieved 5th January 2024.
- [4] P. Lin, J. Weng, S. Hu, D. Alivanistos, X. Li, and B. Yin, "Revealing spatio-temporal patterns and influencing factors of dockless bike sharing demand," *IEEE Access*, vol. 8, pp. 66 139–66 149, April 2020.
- [5] S. Helmer and V. M. Ngo, "A similarity measure for weaving patterns in textiles," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, 2015, p. 163–172.
- [6] V. M. Ngo, S. Helmer, N.-A. Le-Khac, and M.-T. Kechadi, "Structural textile pattern recognition and processing based on hypergraphs," *Information Retrieval Journal*, vol. 24, no. 2, pp. 137–173, 2021.
- [7] T. Cao, K. Le, and V. Ngo, "Exploring combinations of ontological features and keywords for text retrieval," in *Proceedings of The 10th Pacific Rim International Conference on Artificial Intelligence (PRICAI-2008)*, ser. Lecture Notes in Artificial Intelligence (LNAI), vol. 5351. Springer, 2008, pp. 603–613.
- [8] V. Ngo, T. Cao, and T. Le, "Combining named entities with wordnet and using query-oriented spreading activation for semantic text search," in *Proceedings of the 2010 IEEE International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF-2010)*. IEEE, 2010, pp. 1–6.
- [9] V. Ngo, T. Cao, and V. Le, "Wordnet-based information retrieval using common hypernyms and combined features," in *Proceedings of The 5th International Conference on Intelligent Computing and Information Systems (ICICIS-2011)*. ACM, 2011, pp. 1–6.
- [10] L. H. Nguyen, N. T. H. Pham, and V. M. Ngo, "Opinion spam recognition method for online reviews using ontological features," *Journal of Science, Special Issue: Natural Science and Technology*, vol. 61, no. 95, pp. 44–59, 2014.
- [11] V. M. Ngo, G. Munnely, F. Orlandi, P. Crooks, D. O'Sullivan, and O. Conlan, "A semantic search engine for historical handwritten document images," in *Linking Theory and Practice of Digital Libraries*, G. Berget, M. M. Hall, D. Brenn, and S. Kumpulainen, Eds. Springer International Publishing, 2021.
- [12] V. M. Ngo, P. Kearney, F. Donohue, C. Buckley, G. Sood, and M. Roantree, "Using hl7-fhir as an integration platform for chronic disease services management and planning in the irish healthcare sector," in *Joint Int. Conference of ISEH, ICEPH ISEG on Environment and Health 2024*, 2024.
- [13] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258–276, 2018.
- [14] Z. Ghandeharioun and A. Kouvelas, "Link travel time estimation for arterial networks based on sparse gps data and considering progressive correlations," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 679–694, 2022.
- [15] I.-J. Seo and J. Cho, "Structural features of public bicycle transportation networks over times of the day: The case of seoul public bicycle," in *2022 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2022, pp. 5–8.
- [16] Y. Yang, A. Heppenstall, A. Turner, and A. Comber, "A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile," *Computers, Environment and Urban Systems*, vol. 77, p. 101361, 2019.
- [17] F. Munoz-Mendez, K. Han, K. Klemmer, and S. Jarvis, "Community structures, interactions and dynamics in london's bicycle sharing network," in *Proceedings of ACM UbiComp&ISWC 2018*. ACM, 2018, pp. 1015–1023.
- [18] M. Scriney, S. McCarthy, A. McCarren, P. Cappellari, and M. Roantree, "Automating data mart construction from semi-structured data sources," *Comput. J.*, vol. 62, no. 3, pp. 394–413, 2019.
- [19] *Neo4j Graph Data Science Library Manual v2.5*, Neo4j, 2023. [Online]. Available: <https://neo4j.com/docs/graph-data-science/current/>
- [20] D. Cuong, V. Ngo, P. Cappellari, and M. Roantree, "Analyzing shared bike usage through graph-based spatio-temporal modelling," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 5, pp. 1–18, 01 2024.
- [21] F. Sun, P. Chen, and J. Jiao, "Promoting public bike-sharing: A lesson from the unsuccessful pronto system," *Transportation Research Part D: Transport and Environment*, vol. 63, pp. 533–547, 08 2018.
- [22] X. Zhou, "Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in chicago," *PLOS ONE*, vol. 10, no. 10, pp. 1–20, 10 2015.
- [23] H. Zhang, C. Zhuge, J. Jia, B. Shi, and W. Wang, "Green travel mobility of dockless bike-sharing based on trip data in big cities: A spatial network analysis," *Journal of Cleaner Production*, vol. 313, p. 127930, 2021.
- [24] Y. Yao, Y. Zhang, L. Tian, N. Zhou, Z. Li, and M. Wang, "Analysis of network structure of urban bike-sharing system: A case study based on real-time data of a public bicycle system," *Sustainability*, vol. 11, no. 19, 2019.
- [25] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, p. 258–276, Dec 2018.
- [26] M. Zaltz Austwick, O. O'Brien, E. Strano, and M. Viana, "The structure of spatial networks and communities in bicycle sharing systems," *PLOS ONE*, vol. 8, pp. 1–17, 09 2013.
- [27] J. Jia, C. Liu, X. Wang, H. Zhang, and Y. Xiao, "Understanding bike-sharing mobility patterns in response to the covid-19 pandemic," *Cities*, vol. 142, p. 104554, 2023.
- [28] P. Jiménez, M. Nogal, B. Caulfield, and F. Pilla, "Perceptually important points of mobility patterns to characterise bike sharing systems: The dublin case," *Journal of Transport Geography*, vol. 54, pp. 228–239, 2016.
- [29] T. Thi, J. Timoney, S. Ravichandran, P. Mooney, and A. C. Winstanley, "Bike renting data analysis: The case of dublin city," in *25th GIS Research UK Conference (GISRUK)*, 2017.
- [30] H. Lu, M. Halappanavar, and A. Kalyanaraman, "Parallel heuristics for scalable community detection," *Parallel Computing*, vol. 47, pp. 19–37, 2015.
- [31] X. Shi, Y. Wang, F. Lv, W. Liu, D. Seng, and F. Lin, "Finding communities in bicycle sharing system," *Journal of Visualization*, vol. 22, no. 6, pp. 1177–1192, Dec 2019.
- [32] Moby, "Bikes as a service," <https://mobybikes.com/about-moby>, 2019.
- [33] Y. Feng, R. C. Affonso, and M. Zolghadri, "Analysis of bike sharing system by clustering: the vélib' case," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 12 422–12 427, 2017, special Issue for the 20th IFAC World Congress.
- [34] R. A. Azdy and F. Darnis, "Use of haversine formula in finding distance between temporary shelter and waste end processing sites," *Journal of Physics: Conference Series*, vol. 1500, no. 1, p. 012104, apr 2020.
- [35] L. Ramos Emmendorfer and A. M. de Paula Canuto, "A generalized average linkage criterion for hierarchical agglomerative clustering," *Applied Soft Computing*, vol. 100, p. 106990, 2021.
- [36] M. Signorelli and L. Cutillo, "On community structure validation in real networks," *Computational Statistics*, vol. 37, no. 3, pp. 1165–1183, Jul 2022.
- [37] X. Que, F. Checconi, F. Petrini, and J. A. Gunnels, "Scalable community detection with the louvain algorithm," in *2015 IEEE International Parallel and Distributed Processing Symposium*, 2015, pp. 28–37.
- [38] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.