IEEE Open Journal of
Intelligent Transportation Systems

Received 22 August 2023; revised 9 November 2023 and 22 December 2023; accepted 31 December 2023. Date of publication 5 January 2024; date of current version 26 January 2024.

Digital Object Identifier 10.1109/OJITS.2023.3350213

Analyzing Shared Bike Usage Through Graph-Based Spatio-Temporal Modeling

DINH VIET CUONG[1], VUONG M. NGO[1,2], PAOLO CAPPELLARI[3], AND MARK ROANTREE[1]

[1] Insight Centre for Data Analytics, School of Computing, Dublin City University, Dublin 9, D09 V209 Ireland
[2] Information System Management Center, Ho Chi Minh City Open University, Ho Chi Minh City 700000, Vietnam
[3] City University of New York, College of Staten Island, Staten Island, NY 10314, USA

CORRESPONDING AUTHOR: V. M. NGO (e-mail: vuong.ngo@dcu.ie)

This work was supported by the Science Foundation Ireland under Grant SFI/12/RC/2289_P2.

ABSTRACT Bike sharing schemes can be used both to improve mobility around busy city routes but also to contribute to the fight against climate change. Optimization of the network in terms of station locations and routes is a focus for researchers, where usage can highlight the precise times at which bike availability is high in some areas and low in others. Locations for new stations are important for the expansion of the network, but spatio-temporal pattern analysis is required to accurately identify those locations. In other words, one cannot rely on spatial information nor temporal information in isolation, when making interpretations for the purpose of optimizing or expanding the network. In this research, a solution based on graph networks was developed to model activity in transport networks by exploiting properties and functions specific to graph databases. This generic approach adopts a broad series of analyses, comprising different levels of granularity and complexity, to enable better interpretation of network dynamics at a suitably granular level to help the optimization of transport networks. A large dataset provided by an electric bike company is used to address key research questions in both interpreting activity patterns and supporting network optimization.

INDEX TERMS Spatio-temporal graph analysis, smart city, transport networks.

I. INTRODUCTION

MOBY Move [1] is an electric bike sharing system operating in multiple countries using a mobile application platform to rent a fully electric bike. Customers are required to scan a QR code on the bike using their phone to unlock and start riding the bike. Moby Move is a *dockless* system, meaning people do not need to park the bikes at a physical station. In effect, this means that trips can begin and end at any location in cities where the bike scheme operates. Bike sharing schemes not only serve as efficient transport mechanisms, but by studying the dynamics of bike movements, knowledge can be generated to inform network expansion, new strategies and policies. For example, in [2], the authors examine the patterns in bike usage in medium sized Irish cities while also observing how it compares to 48 cities worldwide. This particular study focused on the influence of factors such as weather conditions, routes, and distance travelled on usage patterns. Other approaches, including where vehicles are airborne and unmanned [3] uses analyses of the network to efficiently plan routes through the network, offering in effect, an approach to exploiting under utilized routes in ground based transportation networks.

The study on shared bike usage has significant implications for improving sustainability in transportation, including: (1) reduced carbon emissions; (2) traffic congestion alleviation; (3) health benefits; (4) resource efficiency; (5) integration with public transport; and (6) promotion of sustainable urban planning. However, existing research on bike sharing systems lacks systematic methodologies and construction optimization for complex networks. Furthermore, there is a gap in exploring networks that incorporate both spatial and temporal patterns.

Historically, bike sharing systems were designed with docking stations in fixed locations across the city according

The review of this article was arranged by Associate Editor J. Liu.

to a predesigned topology from where customers collect and drop off bikes. More recently, providers started offering dockless bike sharing services, where bikes are picked up and dropped off at more casual locations, often referred to as *virtual stations*. Moby Move, as a dockless bike sharing provider, has designed an initial network of virtual stations. Considering that virtual stations are easily relocatable, it is in the best interest of the provider to monitor the bike usage and determine the optimal network configuration for such stations. In other words, the problem can be stated as "how close the currently deployed network of virtual stations resembles the network that optimizes bike usage?" Bike sharing systems often make their data available in a tabular format. However, tabular data is not well suited to modelling and analyzing this type of problem, and secondly, data associated with station locations and trips must be available at different levels of granularity to deliver both a high level perspective of network activity and more fine-grained analysis where required. Any solution must consider that decisions made at a global level may adversely affect individual stations, while adjustments to network topology to fix a localized problem may be to the detriment of the global network. More specific requirements by the organisation can be stated as follows:

- *Requirement 1.* A high level visualization is required to deliver a detailed picture of the busiest stations and routes, and also to identify little used stations and routes. Such a visualization is effectively a roll-up across dimensions of time and space.
- *Requirement 2.* For a deeper understanding of the popularity of certain stations and routes, a drill-down using the time dimension must be supported. In particular, there is a requirement to analyze any number of sub-networks over any time interval where these intervals may or may not, be disjoint.
- *Requirement 3.* The next requirement is to determine *similarities* across stations, using both time and location dimensions. Spatial similarities can be identified as part of Requirement 1 and temporal similarities can be observed as part of Requirement 2. However, the combination of these dimensions will identify those stations that exhibit similarities over specified time intervals.
- *Requirement 4.* A more complex global analysis requires an understanding of the relationship between each station and all other stations in terms of activity. Here, we refer to a *pattern* of a station as the route activity with *all other stations* in the network. This type of analysis identifies stations with similar travel patterns to all other stations. Conversely, it will also identify stations with highly unique travel patterns.

Contribution: Researchers at Dublin City University, in collaboration with City University of New York, teamed with Moby Move to develop a single framework approach to address each of the analytical requirements of the network provider. However, a parallel goal was to develop a generic solution to a range of spatio-temporal analyses with the Moby Move providing a challenging case study. Any solution must facilitate a better understanding of their transport network and its usage patterns, while acknowledging that spatio-temporal analysis is complex, making it easy to make incorrect inferences. Thus, our approach starts with a purely spatial analysis, before adding a temporal overlay, and finally moving to full spatio-temporal analysis.

Our contribution can be articulated as follows:

- the development of a framework that exploits graph technology to deliver a dimensional analysis of network traffic, in this case, a bike sharing scheme. By dimensional analysis, we facilitate the construction of high level graphs together with drill-downs across spatial and temporal dimensions;
- a temporal graph network which enables a drill-down over the time dimension which facilitates the specification and construction of graph subsets for the purpose of comparative analyses;
- an optimization function to adjust graph densities to remove low volume graph edges;
- the construction of a correlation matrix of time series data to create station profiles to model their interactions with all other stations, creating a mechanism to compare stations from a different perspective to comparisons using the temporal graphs.

The remainder of this paper is structured as follows: in Section II, we present a discussion on related work in this area; in Section III, the fundamental knowledge, concepts, and definitions necessary are provided for the subsequent development of this work; in Section IV, we introduce a 3-step methodology which constructs graphs with increasing levels of analytical capabilities; in Section V, we provide the key measures, dataset, and technologies employed for evaluating and implementing our study; in Section VI, we present our experiments together with a discussion of findings; and finally in Section VII, we summarize the paper and present some conclusions.

II. RELATED WORK

A. NETWORK CONSTRUCTION

Complex network analysis methods have been employed to investigate bike sharing systems worldwide. In these investigations, bike sharing records are modelled in network structures and through the analysis of these constructed networks, we can unveil the inherent characteristics of the bike sharing systems. Network modelling varies depending on the purpose of the analysis. Typically, traffic flows are investigated where spatial locations are viewed as nodes while trips form an edge between its starting location and destination [4], [5], [6], [7], [8], [9], [10].

The majority of the systems under examination are dock-based, meaning that customers are obliged to rent and return the bikes at specified stations. Therefore, it is logical to consider these stations as nodes. In contrast, with dockless systems users have the freedom to pick up and drop off their bikes anywhere, which requires researchers to define novel ways to describe and model the data. Zhang et al. [4] divided the area of interest into a grid of squares and treated them as nodes, while Yang et al. [5] used physical road segments as nodes. Certain studies adopt an alternative approach, wherein they group places together and treat each group as a single point within the network. For instance, Seo and Cho [6] categorized bike stations based on their surroundings, while Shi et al. [7] used clustering methods to group stations.

Other approaches to complex network construction have also been proposed for the analysis of bike-sharing systems. For instance, Batista et al. [8] built a network where each node represents a specific region. Within these regions, vehicles travel at the identical average speed, and nodes in the network are interconnected if the regions are adjacent to each other. The network helps to study relationships between the average distance travelled or travel time and the level of exhaust emissions along bike paths. Lin et al. [9] trained a graph convolutional neural network to learn the correlation among stations and predict station-level hourly demand. The correlation matrix that is obtained from this training is used to create an adjacency matrix for a network. By analyzing this network, more insights are gained about the spatial relationships between different stations. Furthermore, Ghandeharioun and Kouvelas [10] constructed road networks with pairwise edge correlation to estimate travel time of a route.

While complex network analysis has found widespread applications in the study of bike-sharing systems, the papers above have overlooked the need for a systematic framework for network construction and analysis, particularly in optimizing networks to enhance their analytical capabilities. Furthermore, the potential of correlation-based networks for uncovering spatial-temporal patterns remains underdeveloped in the context of bike-sharing systems. As a result, we aim to establish a formal framework for the application of complex network approaches and advocate for the utilization of correlation-based networks in bike-sharing data analysis.

B. DYNAMICS OF NETWORKS

In addition to static characteristics of networks, the evaluation and dynamics of bike sharing systems are also a concern. Therefore, constructed networks must be aggregated into different time intervals and periodicities. The choice of time interval depends on the analysis goal or is often determined by a domain expert. For example, [5] quantified changes in travel flows of a bike sharing system in Nangchang, China after a new metro line was introduced, by comparing network structures, five days before and after. Jia et al. [11] demonstrated the changes in bike-sharing systems in response to the outbreak and recovery of the pandemic. This was performed by analyzing networks that were projected onto different waves of COVID-19. In studies by Borgnat et al. [12] and Austwick et al. [13], patterns in bike usage during weekdays versus weekends and at different times of the day were successfully identified through network analysis.

In these studies, the choice of time intervals is often predetermined based on the specific research interests, which may not always be the ideal approach. Without domain knowledge, it can be challenging to identify similarities and anomalies in the time dimension, to select relevant periods for analysis. Analyzing every time-step would be time-consuming and inefficient. As a solution, we propose the utilization of a clustering-based method to aggregate similar time-steps and analyze the resulting representations instead.

C. NETWORK ANALYSIS

In terms of network analysis, networks for bike sharing systems have been analyzed through the same methods, i.e., network metrics, community detection with the support of visualization and domain knowledge. Global metrics illustrate the overall structure of the networks, while local metrics show the roles or properties of the nodes in a network. Most commonly used metrics include the number of nodes, the number of edges, degree and strength, which indicate the degree of activity and connectivity at a location [4], [14].

Austwick et al. [13] identified similarities in the strength and edge weight distribution among networks created from various bike sharing systems. Yang et al. [5] and Jia et al. [11] recommended incorporating a diverse set of network properties to encompass different facets of a network, including connectivity metrics (degree and node flux), spatial distribution (clustering coefficient), as well as measures of interaction (accessibility), network stability (network connectivity), efficiency (network efficiency), and equity (Gini coefficient). Other centrality metrics, such as betweenness and PageRank, as utilized by [5], have also been employed in the context.

Community analysis, which is a recurring theme in network research, plays a crucial role in understanding a network's structure. Community detection algorithms group nodes within the network into distinct communities, where nodes exhibit stronger internal connections than external ones. The Louvain algorithm in [15] is most commonly used. However, Shi et al. [7] noted that different algorithms produced varying community characteristics based on the measurement criteria. Although network analysis is now well developed, the methods are mostly applied to networks where edges are trips. We would like to apply similar approaches, such as visualization, complex metrics like strength, closeness, betweenness, local clustering coefficients and community detection to more complex, correlation-based networks.

D. SUMMARY

In summary, despite the numerous network-based analyses of bike sharing systems, several limitations persist in contemporary studies: (1) existing studies lack systematic methodologies for applying complex networks in the analysis of bike sharing systems; (2) the construction of the networks is often neglected and not optimized, hindering subsequent network analysis; (3) when studying the dynamics of these systems, the choice of the time periods for analysis highly depends on domain-specific knowledge, such as grouping hourly data into similar patterns or identifying evolving station periods; (4) correlation networks have the potential to reveal not only spatial but also temporal patterns, but their application in network analysis is under explored.

Motivated by these observations, we propose a systematic comprehensive method that focuses on integrating complex network theory into analyzing bike sharing systems. The method involves building and analyzing 3 types of networks: spatial graph networks, temporal graph networks and spatio-temporal graph networks. Spatial graph networks provide an overview of spatial interactions, while temporal graph networks capture the dynamics of the systems. Spatio-temporal graph networks are correlation-based networks that reveal temporal patterns in the context of spatial relations.

The construction of our three networks is optimized using a proposed algorithm for selecting an appropriate trimming threshold. This optimization enhances the analysis of the networks, particularly with respect to shortest distance-based centrality metrics and community detection. Finally, we also suggest an algorithm to assist in the analysis of the dynamic networks by automatically clustering periods with similar network structures.

III. PRELIMINARIES

In this section we present the pre-requisite knowledge, concepts and definitions, that will be used in the development of this work.

A. SPATIAL GRAPH NETWORK

The Spatial Graph Network (SGN) is the simplest form of network including notes and edges. Nodes are connected via undirected edges where an edge between two nodes is created if there exists at least one trip between the node representing the origin node (source) and the node representing the destination node (target). The SGN provides a complete aggregated view of the network topology as well as the volume of activities through the network. Each edge is also associated with a weight describing the number of trips that occurred on the route between the two connected nodes over the entire time period of interest. More formally, the SGN is defined in Def. 1.

Definition 1 (Spatial Graph Network): The *SGN* is a pair $SGN = \langle S, J \rangle$, where S is a set of vertices (nodes) and J a set of edges, connecting vertices. A vertex (or node) $s \in S$ is a pair $s = \langle lat, lon \rangle$, where *lat* is the latitude and *lon* is the longitude at which the vertex is spatially located. An edge $j \in J$ is a triple $j = \langle o, d, a \rangle$, where $o, d \in S$ are the origin and destination vertices of the edge, and a is the activity between such vertices.

B. TEMPORAL GRAPH NETWORK

The aggregated nature of the SGN can hide nuances and continual evolution of networks. In order to study network evolution over time, there is the need to explicitly represent time as a dimension in the graph. Segmenting data, especially graph or unstructured data, helps in the analysis of data at a more granular level and is more efficient in terms of query response time [16].

For the above reason, we introduce the temporal graph network (TGN). A TGN is an ordered sequence of graphs defined over nodes and edges where each graph in the sequence describes the state of the network at a specific point in time. More precisely, each graph in the TGN describes the status of the network as a snapshot of the SGN taken during a time window, over the course of the period of interest. The TGN enables both a more fine-grained analysis of the network and also the ability to compare across graph projections, thus facilitating analysis of the network's evolution. Formally, a TGN is defined in Def. 2.

Definition 2 (Temporal Graph Network): A *TGN* is a set $TGN = \{\mathcal{SGN}_1, \mathcal{SGN}_2, \ldots, \mathcal{SGN}_T\}$, where SGN_i is a time-bounded *SGN* and represents the status of the network during a specific interval of time. Each SGN_i is a pair $SGN_i = \langle S, K \rangle$, where S is a set of vertices (nodes) and K a set of edges, connecting vertices. A vertex (or node) $s \in S$ is a pair $s = \langle lat, lon \rangle$, where *lat* is the latitude and *lon* is the longitude at which the station is spatially located. An edge $k \in K$ is a tuple $k = \langle o, d, t, \delta, a \rangle$ representing the volume of activity and journeys between two vertices during an interval of time, where: $o, d \in S$ are the origin and destination vertices of a journey; t is the time at which the observation period starts; δ is the duration of the interval of time with origin t and during which the observation occurs; and a is the activity, defined as number of trips between the vertices o, d during the observation period starting at time t through the interval δ of time.

C. SPATIO-TEMPORAL GRAPH NETWORK

The Spatio-Temporal Graph Network (STGN) has an entirely different structure to the previous graph representations. It has a network structure where edges model the *similarity* between two vertices (nodes) in terms of temporal *patterns* in their activities. This enables the treatment of data as timeseries patterns [17], with the potential to identify trends, seasonal or cyclical components, irregular components and potentially, the diversity [18] within the data. The STGN is defined formally in Def. 3.

Definition 3 (Spatio-Temporal Graph Network): A STGN is a tuple $STGN = \langle S, H, T_s, T_e, \gamma, \upsilon \rangle$, where S is a set of vertices (nodes) and H a set of edges, connecting vertices, T_s is the time at which network observation commenced, T_e is the time at which the network observation finished, γ is the time increment between start times of network activity observations, and υ is the duration of each observation. A vertex (node) $s \in S$ is a pair $s = \langle lat, lon \rangle$, where *lat* is the latitude and *lon* is the longitude at which the vertex is spatially located. An edge $h \in H$ is a triple $h = \langle s_1, s_2, r \rangle$ describing the similarity between nodes $s_1, s_2 \in S$ in terms of a temporal pattern, where r is the similarity value between the two nodes calculated over the series of activity observations over the time periods.

IV. METHODOLOGY

Networks such as those implemented by bike sharing systems are time evolving systems that can be characterized via the following core features: space, time, and volume of activity. Space and time can be considered dimensions, whereas activity is a measure whose value depends on how one looks at the dimensions. Here, we model the bike sharing system as a graph network that analyzes the measure (trip) according to one or more dimensions. Specifically, we build three network types. The first network considers the spatial dimension only. The second network, although it incorporates spatial properties, considers only the time dimension for the purpose of analysis. The third network combines the spatial and temporal dimensions. Each network increases in complexity and power compared to the preceding network. However, simpler networks are easier to construct, and as the cost of building more complex networks increases, it is important to understand the capabilities of each network type. We will now discuss how networks are built, the analyses possible using each network and finally, discuss the application of each network.

Algorithm 1 find_threshold(edges): Binary Search to Find the Largest *threshold* for Strong Connectivity From Edge

Input: *edges* ← A list of all edges in the network, each with a weight
Output: *threshold*: Largest threshold value for strong connectivity, *NULL* if there is none.

1: *edgeValues* ← unique weights extracted from edges sorted in ascending order
2: *low* ← 0
3: *high* ← length(*edgeValues*) − 1
4: *threshold* ← *NULL*
5: **while** *low* ≤ *high* **do**
6:     $mid \leftarrow \left\lfloor \frac{high+low}{2} \right\rfloor$
7:     *network* ← build_network(*edges*, *edgeValues*[*mid*])
8:       ▷ Removes edges with weight < *edgeValues*[*mid*]
9:     **if** is_strongly_connected(*network*) **then**
10:         *threshold* ← *edgeValues*[*mid*]
11:         *low* ← *mid* + 1
12:     **else**
13:         *high* ← *mid* − 1
14:     **end if**
15: **end while**
16: **return** *threshold*

A. SPATIAL BIKE GRAPH NETWORK

We successfully utilized a graph network to model the structure of textile patterns [19] and a historical knowledge base [20]. Consequently, we constructed a Spatial Bike Graph Network (SBiGN) based on the SGN to describe the spatial features of the bike sharing usage, including station connections and trip volumes between stations. This representation is invaluable for analyzing the *traffic flow* of the transportation system. In a straightforward sense, SBiGN's nodes represent bike stations, while its edges describe journeys (trips) between stations. Each edge is also associated with a weight that quantifies the number of trips that occurred on the route between the two connected stations over the entire time period of interest. More formally, the SBiGN can be found in Def. 4.

Definition 4 (Spatial Bike Graph Network): The *SBiGN* is a pair $SGN = \langle S, J \rangle$, defined as per Def. 1, where a vertex and an edge in *SGN* correspond to a bike station and a journey (trip) in *SBiGN*, respectively. Additionally, the activity a between two stations within a trip can be viewed as the cardinality of that trip.

We refine the SBiGN by removing statistically insignificant data which add noise rather than value to the analysis. First, we identify and remove so-called *weak* edges, which are edges with small weight values relative to the aggregated values in the network. This is a common step in network optimization [21] where potentially high numbers of insignificant data points can slow network analysis while providing little new information. The threshold is set to guarantee the network is strongly connected, ensuring network algorithms such as closeness and betweenness are well-defined while minimizing the network size. Thereby, this optimization enhances the quality as well as the speed of the analysis. A binary search version of the threshold algorithm is described in Algorithm 1, where the algorithm (line 5) continually checks the graph's strongly connected property, to ensure it never drops below a specified threshold. Essentially, the `is_strongly_connected` function validates that all nodes remain reachable from any starting node [22].

Second, we remove *looping* edges, which are defined as edges where the source and destination are the same node (i.e., the bike trip originates and ends at the same station), as they add little information and may cause confusion for some graph algorithms. Loops may however, be an important part of the nature of the transportation system so they should be analyzed in a separate network.

Other than for purposes of analysis, a significant benefit of the spatial graph is that it can be easily overlayed on the geographical map where the bike sharing network is deployed. As part of **Requirement 1**, the SBiGN in Figure 1 uses data over a 15 month period from June 2020 to August 2021 to easily identify the location of the busiest stations and the spatial distribution of a set of busy stations. In the visualization, the size of the node is proportional to the total number of trips beginning or ending at the station. The top 10 biggest nodes (stations) are depicted as red nodes with other nodes in blue. Only the top 10% of edges are illustrated with the most frequent routes in red and the remainder in blue.

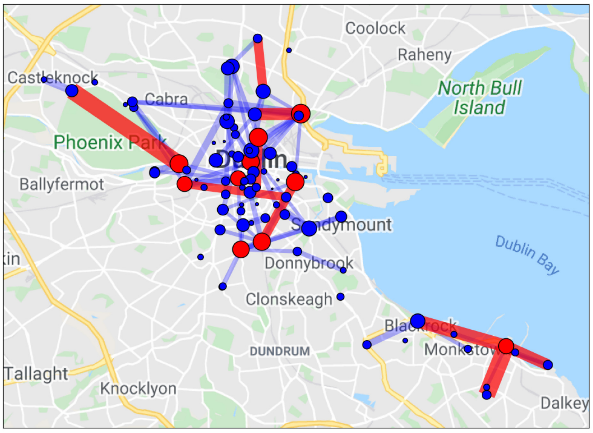


FIGURE 1. Geographic Map overlaid with SBiGN.

B. TEMPORAL BIKE GRAPH NETWORK

From a practical point of view, we create a Temporal Bike Graph Network (TBiGN) by constructing a sequence of time-bounded SBiGNs, as defined in Def. 5. The TBiGN follows a specified temporal order and a predetermined time scale, where each SBiGN_i captures network activity pertinent to a particular time interval.

These intervals can have any duration and may or may not overlap. This is necessary to model the fact that temporally adjacent networks describe how the system evolves over time. Defining the length of these intervals determines the similarity between adjacent networks. Intuitively, a shorter interval yields to potentially less similar networks, but it enables capturing the system evolution at a finer level of granularity, as per **Requirement 2** in the introduction. Conversely, a longer interval yields to potentially more similar networks, because aggregates over longer periods may converge to similar values, at the cost of missing out on some of the system evolution behaviours. Yet they are necessary to capture properties such as seasonality and stationarity [18].

Definition 5 (Temporal Bike Graph Network): A *TBiGN* is a set $TBiGN = \{SBiGN_1, SBiGN_2, \ldots, SBiGN_T\}$, where $SBiGN_i$ is a time-bounded *SBiGN* and is a pair $SGN_i = \langle S, K \rangle$. This corresponds to Def. 2, where a vertex and an edge in SGN_i corresponds to a bike station and a trip in $SBiGN_i$, respectively.

Typically, for networks implemented by bike sharing systems, intervals are defined in the range of hours, days, weeks etc., in order to analyze periodical and seasonal factors. As part of the solution to address **Requirement 3**, networks are then clustered into groups where their centroids represent the entire cluster and are used for analysis. This simplifies the analysis of a wide range of nodes as the analysis can then focus on a small group of centroid nodes.

C. SPATIO-TEMPORAL BIKE GRAPH NETWORK

The Spatio-Temporal Bike Graph Network (STBiGN) was designed to use STGNs as a fundamental building block and is defined in Def. 6. This type of network is used to address **Requirement 4** where a different perspective on station by station correlation can be explored.

The STBiGN construction comprises three steps:

1) In the first step, a time series is constructed for every station for a specified timescale. The timescale which is controlled by the values chosen for γ and υ, has the following semantics:
   - $\gamma = \upsilon$: creates contiguous non-overlapping intervals, over the observation period T_s, T_e.
   - $\gamma < \upsilon$: creates overlapping intervals, where the same activities may contribute to the activity volume count in multiple intervals.
   - $\gamma > \upsilon$: generates intervals that may not account for all activities that occurred during the observation time.

   Generally, analyses are conducted on contiguous, non-overlapping intervals, such as hourly, daily, weekly, monthly, yearly, or a timescale to explore the periodic nature of the transportation systems, e.g., morning or evening rush hours. Once the timescale is known, a station is represented by a time series that is the series of numbers of trips at every time-step window. This way the activity of a station can be studied for the specified timescale.
2) In the second step, a similarity score is computed for *every* station pair based on the popular Pearson correlation coefficient, shown in equation (1). Here, (x_i) and (y_i) are two time series of length n, where $n = (T_e - T_s)/w$ is the number of time intervals in which the activity is measured, and with $w = \gamma = \upsilon$; $\bar{x} = \sum_j^n x_j$ is the sum of x scores; and $\bar{y} = \sum_j^n y_j$ the sum of y scores. The function measures the linear correlation between two variables on the range $-1 \leq x \leq 1$. As most network algorithms work with non-negative edge weights, we normalize the coefficient to have a value between 0 and 1 by $S(x, y) = (S(x, y) + 1)/2$.

$$S(x, y) = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})}\sqrt{\sum_i^n (y_i - \bar{y})}}. \quad (1)$$

3) In the final step, the STBiGN is constructed based on a correlation matrix of the Pearson coefficients. The similarity matrix formed by all normalized coefficients is the adjacency matrix for the network. In the STGN, nodes represent stations and there *always* exists an edge between every pair of stations with the corresponding similarity score as the weight. As network metrics like closeness, betweenness or local clustering coefficient only work on unweighted networks and badly on fully connected networks, weak edges that are statistically insignificant are removed. The strongly connected property is ensured when optimizing the graph (trimming edges).

Definition 6 (Spatio-Temporal Bike Graph Network): A *STBiGN* is a tuple $STGN = \langle S, H, T_s, T_e, \gamma, \upsilon \rangle$, corresponding to Def. 3, where a vertex and an edge in $STGN_i$ refer to a bike station and a trip in $STBiGN_i$, respectively.

V. MEASURES AND DATASET

In this section, we first introduce the main metrics to lead to a quantitative analysis of the model and of the experiments; then, we describe the dataset used in the analysis; finally, we briefly mention the technology used to develop the implementation part of the study.

A. GRAPH METRICS

1) GRAPH METRICS FOR SBIGN AND TBIGN

The key techniques for analyzing the structure of the SGN/TGN, as well as SBiGN/TBiGN, revolve around the concept of centrality metrics. These metrics quantitatively measure the importance of a node in a network [23], [24]. When analyzing centrality metrics, it's important to pay attention to nodes with exceptionally high scores, as they could either be outliers or nodes that play pivotal roles within the network.

- *Strength*. Node strength is defined as the sum of the weights of the edges connecting to immediate neighbors [25]. In our SBiGN/TBiGN, the weight represents the volume of activity. Therefore, station strength represents the total number of trips occurring at a specific station in either directions to or from all adjacent stations.
- *Degree.* The degree of a node is defined as the number of edges that connect it to its adjacent neighbor nodes, which are the nodes to which it is directly connected [26]. In the SBiGN/TBiGN graph, the degree of a station represents the number of stations with which it shares at least one trip.
- *Closeness.* The closeness of a node is one measure of its centrality in the network, that is how close, on average, the node is to all other nodes [27], [28]. In this context, we determine the distance between nodes as the length of the shortest path between source and destination, i.e., the minimum number of hops (trips) required. Therefore, in the SBiGN/TBiGN, the closeness score quantifies how proximate a station is to all other stations based on all trips, reflecting the overall movement within the network.
- *Betweenness.* The betweenness score of a node indicates how many shortest paths between any two nodes pass through the specific node under consideration [28], [29]. Indeed, the higher the betweenness score, the more significant the station is considered within the SBiGN/TBiGN's connectivity structure. In practice, stations with high betweenness scores typically act as "bridge" points, serving as common waypoints connecting different sub-networks.
- *Local Clustering Coefficient.* This metric reflects the likelihood that the neighboring nodes of a node are interlinked [30]. In the context of SBiGN/TBiGN, a high coefficient suggests that individuals departing from a specific station tend to move to or from only a few other stations
- *Communities.* Communities present a metric where nodes exhibit strong connections within a specific community of nodes [15], [31]. Essentially, in SBiGN/TBiGN, a community represents a group of stations where there are more trips occurring within the community itself than to or from stations outside of the community.

2) GRAPH METRICS FOR STBIGN

The same graph functions are used again as metrics. But due to the difference in nature of STBiGN from previous networks, graph metrics in such correlation-based networks [32], [33], [34] have different interpretations derived from its definitions and network construction:

- *Strength.* The node strength, which is the total scaled correlations connected to the node, demonstrates the *degree* of similarity between the node with other stations.
- *Degree.* In a STGN, the degree is the number of stations that have similar behavior to the station under observation.
- *Closeness.* The distance between two stations in a STGN is interpreted as the smallest number of significant similarities which form a path from one station to another. The more distant the two stations, the less similar are the temporal patterns in their activities.
- *Betweenness.* Here, a community is a group of stations that have similar temporal activity patterns. Therefore, high scoring stations are usually similar to two or more groups of stations.
- *Local clustering coefficient.* High coefficient stations in a STGN mean their neighbours are likely to also be similar to each other. Stations with a high local clustering coefficient form a group that share more or less the same activity patterns.
- *Communities.* Hence the connections between nodes are built upon similarity of their activity timeseries. Members of a community share common temporal patterns.

B. DATASET

Our original dataset included data from 86 stations and covered a total of 52,936 bike rentals spanning the period between June 2020 and September 2021 as summarized in Table 1. Additionally, this dataset also contained detailed information regarding both pick-up and drop-off times and locations. However, the original dataset has many *virtual* pick-up locations, drop-off locations and trips, requiring cleaning before importing to our graph networks.

Firstly, if the pick-up and drop-off locations do not correspond to existing stations, we assess their proximity to nearby bike stations. If these locations are within a 1km

TABLE 1. Dataset overview.

Measures	Original Dataset	Cleaned Dataset
Duration of data	June 2020 - Aug 2021	
Station count	86 stations	
Rental count	52,936 trips	36,181 trips
Max trips (per station)	2,775 trips	1,936 trips
Min trips (per station)	97 trips	82 trips

radius of at least one station, we reassign them to the nearest station. It's important to note that a single trip may involve both pick-up and drop-off locations that belong to the same station, resulting in what we refer to as a "loop-trip". However, if these locations are more than 1km away from any station, we exclude them from the dataset. In this first step, there are 654 reassigned trips (approximately 1.2%) and 6,650 removed trips (approximately 12.5%).

Secondly, certain trips within the original dataset can be created by system errors or errors made by renters. These errors can encompass issues like app compatibility problems, mistakes in app usage, and failures to properly start or end a trip. Specifically, very short trips are those lasting less than 10 minutes or covering less than 100 meters, while very long trips were defined as those lasting more than 1 day. As a result, during the second step, 10,105 trips (approximately 19%) were removed. It's important to note that we have conducted a pre-experiment analysis on the original dataset to identify suitable values for the parameters mentioned above.

The cleaned dataset now comprises 36,181 trips and 86 stations in Dublin city, Ireland with summaries shown in Table 1. On average, there are 2,412 trips monthly across all stations, with an average of 28 trips per station. The busiest station recorded 1,936 trips over 15 months which is equivalent to 129 trips per month, while the least frequented station had only 82 trips over 15 months, equivalent to 5 trips per month.

C. MATERIALS AND TOOLS

The validation environment included Python with libraries scipy,[1] matplotlib,[2] and NetworkX[3] all used to run experiments. Pandas[4] are used for data processing, data analysis, and graph construction, with Matplotlib utilized for data visualization. Graph visualization used Google Maps API[5] and NetworkX. The Neo4j graph database[6] was used to manage and store all graphs. Additionally Neo4j's Graph Data Science Library[7] with built-in centrality metrics and community detection were used for all network metrics.

1. https://scipy.org/
2. https://matplotlib.org/
3. https://networkx.org/
4. https://pandas.pydata.org/
5. https://developers.google.com/maps
6. https://neo4j.com/
7. https://neo4j.com/docs/graph-data-science/current/

TABLE 2. Node strength in SBiGN.

Station	Strength
Fairview Avenue Lower	1,478
Mountjoy Square South northside opposite No. 40-45	1,385
Outside Criminal Courts of Justice	1,350
Grand Canal Docks, outside Fresh	1,322
Beside Penneys O'Connell Street	1,272
...	...
Benson Street	43
Start of St. Stephen's Green terrace.	43
Pearse Street, outside Subway	26
Phibsborough Rd, outside Broadstone Hall Studen...	23
Merrion Square North opposite Oscar Wilde House	11

VI. ANALYSIS AND DISCUSSION

In this section, we validate our approach against the set of requirements outlined in Section I.

A. OUR SPATIAL BIKE GRAPH NETWORKS

1) SBIGN CONSTRUCTION

As in Def. 4, stations are modelled as nodes in SBiGN, and they are linked by undirected weighted edges if there is a single trip between them in the entire 15 months period. The edge weight is the number of trips, counting both directions. For improving the clarity, readability, and interpretability of the graph visualization, we eliminated loop and weak links in SBiGN.

To pinpoint weak connections, we systematically decrease the edge weight threshold until the network achieves strong connectivity. In these experiments, the final threshold is set at 11 trips, as outlined in Algorithm 1. The ultimate network density, calculated as the ratio of actual edges to potential edges, is approximately 19%. Following the removal of 9,544 loop trips and 7,920 trips associated with weak links, our SBiGN comprises 18,717 trips connected by 686 edges, with the node count remaining unchanged at 86. The diagram shown in Figure 1 illustrates the SBiGN with 69 edges, which account for approximately 10% of the total edges, and carry the most significant weights.

2) SBIGNS - ANALYSIS AND DISCUSSION.

Basic graph analytics are employed to analyze bike movement across the entire network. In Table 2, stations are listed with their respective **strength** values in descending order. Notably, the station at *Fairview Avenue Lower* exhibits the highest strength at 1,478, closely followed by stations located near the city centre: *Mountjoy Square South* (1,385), *Criminal Courts of Justice* (1,350), *Grand Canal Docks* (1,322) and *O'Connell Street* (1,272). The decline in station strength follows a slightly negative-exponential and near linear pattern. Smaller stations tend to have few trips over the observation period, often less than several dozen. Interestingly, many of these smaller stations are located in close proximity to larger stations, which is a noteworthy finding even based on these basic graph analytics.

TABLE 3. Edge weight in SBiGN.

Source	Target	Weight
Dun Laoghaire Dart	Honeypark Neptune Way	247
Criminal Courts of Justice	Phoenix Park Gate	243
Blackrock Main St.	Dun Laoghaire Dart	190
Drumcondra Road Upper	Fairview Avenue Lower	179
Dun Laoghaire Dart	Sandycove Beach	165
...	...	...
End of Adelaide Road	Talbot Hotel Stillorgan	11
Richmond Row	Dun Laoghaire Dart	11
Sandymout Village	Dun Laoghaire Dart	11
Grand Canal Docks	Seapoint	11
Ballsbridge	Monkstown	11

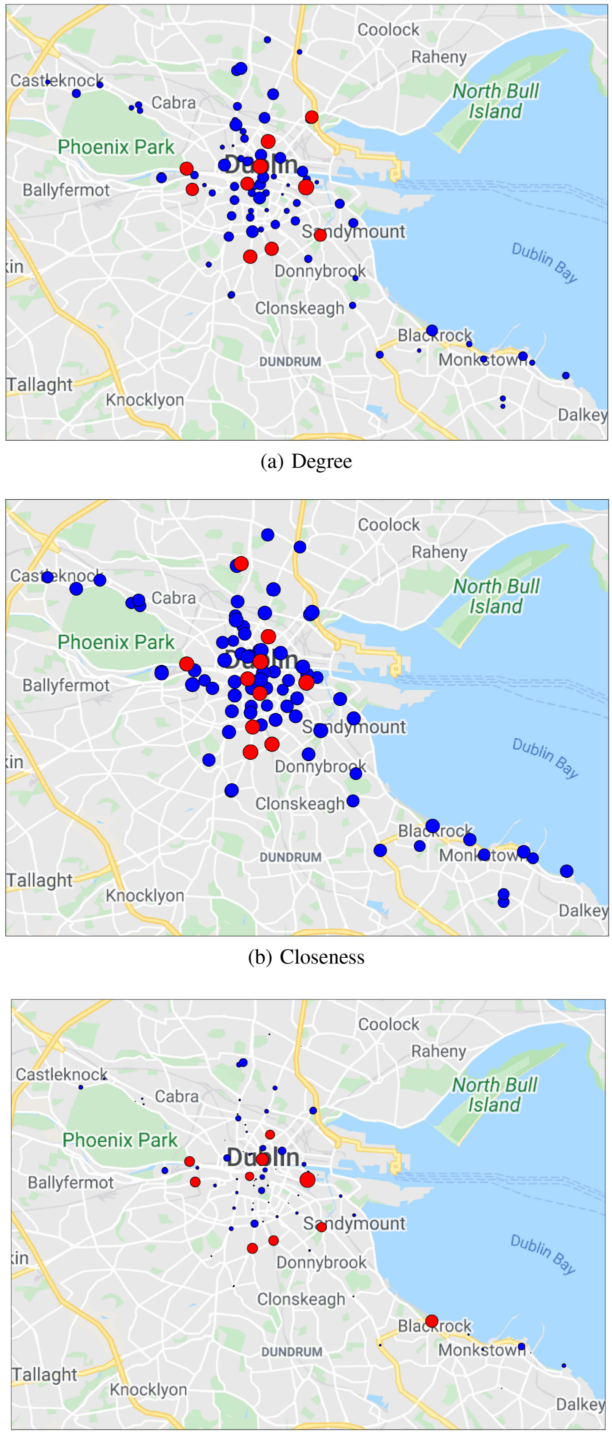
Table 3 presents the edge weights. The most frequently traveled routes are typically associated with commute trips or follow the Blackrock-Monkstown coastal line. The edge weights follow an exponential decrease, with the most popular routes having nearly double the weight of the third most popular route. Unsurprisingly, the least traveled routes have negligible weights. Although certain stations appear at the top of both node and edge metrics, there are distinct variations in popularity patterns between routes and stations, with stations showing a more gradual decline in strength.

Figure 2 highlights important geographical areas using a station's node size to signify its individual importance. In Figure 2(a), **degree** is strongly correlated with **strength**, with the exception of the *Monkstown* node, which is geographically isolated from other stations. High degree nodes, shown in red, are typically located in the center of residential areas, serving as transportation hubs for people traveling to and from these areas.

The **closeness** metric provides additional insight into the significance of stations within the network. Stations with high closeness values (depicted in red) mean that stations are well-connected to a large number of other stations, while stations with lower scores tend to be isolated. In Figure 2(b), the top 10 closeness stations are in the center of the city, indicating that these stations have extensive connections with numerous other stations throughout Dublin. The network's average closeness score is 0.59, implying that, on average, a station is approximately 1.7 trips away from other stations.

As **closeness** values do not vary much across stations, this means that there is no clear standout station that serves as a central hub for reaching many other stations via existing trips. This is advantageous, as it means that no single station closure or removal would significantly disrupt movement throughout the network.

In Figure 2(c), nodes with high **betweenness** scores are represented in red. These high scoring nodes are primarily situated along the river in the city center or in the Blackrock area. When a high-betweenness node like Blackrock is isolated, it suggests that the closure of this station could potentially divide the network into two separate segments, with renters and bikes confined within their respective networks.



(a) Degree

(b) Closeness

(c) Betweenness

FIGURE 2. Graph Analytics for Spatial Bike Graph Networks.

The final analysis for SBiGNs identifies communities using spatial data only where Figure 3 shows nodes with the same color as belonging to the same group. Four communities are clearly identified: the north-east (purple), the north-west (green), the south-city group (yellow) and the south-suburban (Blackrock and Monkstown) area (red). It is an interesting observation that three communities exist

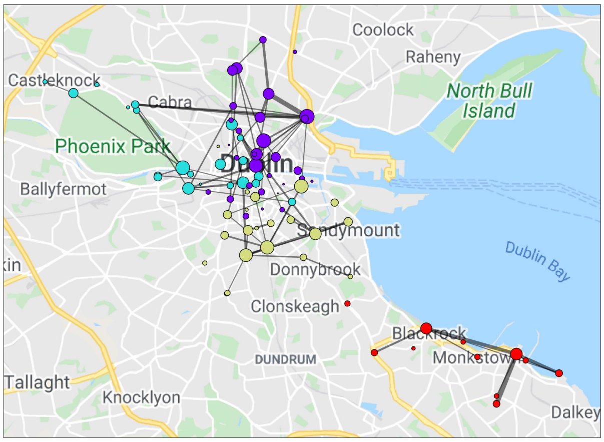


FIGURE 3. SBiGN Community Detection.

close to the centre of the city with two on the north side and a single cluster on the south side. This would indicate more widespread usage on the north side, perhaps indicating that more customers use the bikes for work or might be an indication of socioeconomic properties. The final community (Blackrock-Monkstown) is quite isolated from the other three as it is located far from Dublin city center. Trips here are usually along the coastline or from nearby residential areas to the coast but rarely cross over to other communities. This is an interesting finding given there exists communities with no overlap whereby customers and bikes remain "trapped" inside that network.

3) SBIGNS: SUMMARY

In SBiGNs, we addressed **Requirement 1** by providing a high-level visualization that crosses both time and space. Information contained in Tables 2 and 3, as well as Figures 2 and 3, offer further insights or context into the visualization provided by the network. These include the **strength**, **degree**, **closeness**, **betweenness**, different station **communities**, along with the **weight** of routes over the span of 15 months.

B. OUR TEMPORAL BIKE GRAPH NETWORKS

TBiGNs are employed to analyze different movement patterns for specific time intervals. In this experiment, daily and monthly graph projections were constructed to facilitate both detailed and broader analyses, respectively.

1) DAILY TBIGNS - NETWORK CONSTRUCTION

In this experiment, 7 TBiGNs are generated for each day of the week, covering a span of 66 weeks within a 15-month period from June 2020 to August 2021. These 7 TBiGNs have been categorized into two groups: "weekday", representing 5 TBiGNs of 5 weekdays, and "weekend" representing 2 TBiGNs of the 2 days of the weekend. From 36,181 trips of the cleaned dataset, similar to SBiGN, we have also removed loop and weak links in TBiGNs. However, in TBiGNs, weak links are identified as edges with fewer than 5 trips. On each day and on average, we have eliminated approximately 1,364 loop trips, accounting for 26.4% of the data and 551 trips connected to weak edges, representing 10.6%.

TABLE 4. Node strength: Weekday vs weekend.

Station	Weekday		Weekend	
	Strength	**Rank**	**Strength**	**Rank**
Fairview Avenue Lower	155.0	**1**	145.5	**3**
Mountjoy Square South	116.8	**2**	110.5	8
Warehouse	113.2	**3**	15.5	50
Criminal Courts of Justice	102.2	**4**	191.0	**1**
Ranelagh Village	102.0	**5**	123.0	**5**
Dun Laoghaire Dart	97.8	7	169.5	**2**
Blackrock Main St.	48.8	20	134.5	**4**

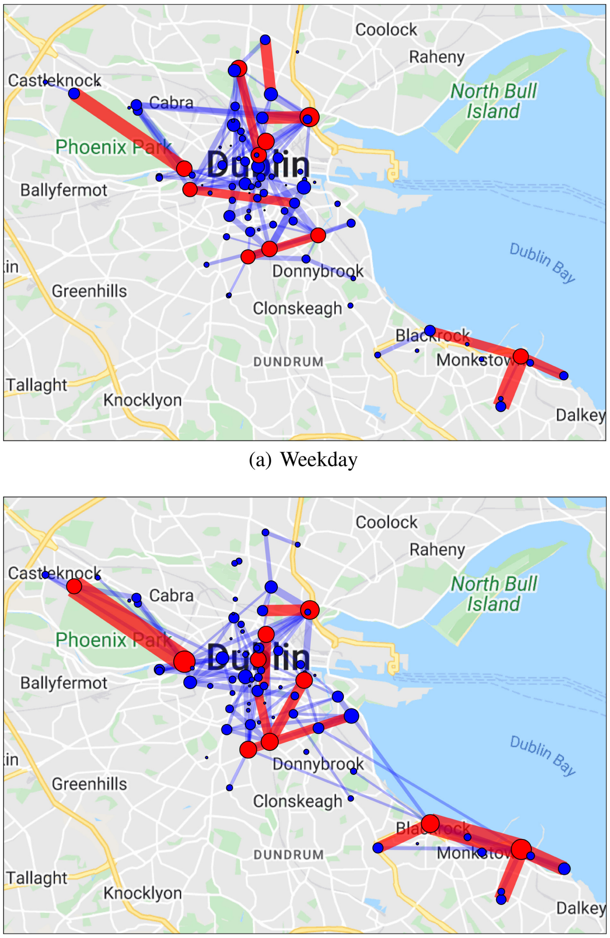
TABLE 5. Edge weights: Weekday vs weekend.

Source	Target	Weekday		Weekend	
		Wgt.	Rank	Wgt.	Rank
Dun Laoghaire Dart	Honeypark Neptune Way	36.0	**1**	31.5	**3**
Criminal Courts of Justice	Phoenix Park Gate	29.6	**2**	47.5	**1**
Drumcondra Road Upper	Fairview Avenue Lower	26.2	**3**	24.0	**5**
DCU Glasnevin	Drumcondra Rd	23.4	**4**	9.0	38
Warehouse	Old Finglas Road	23.0	**5**	5.5	73
Dun Laoghaire Dart	At The Forty Foot	20.8	**6**	30.5	**4**
Blackrock Main St.	Dun Laoghaire Dart	19.8	**7**	44.5	**2**

2) DAILY TBIGNS - ANALYSIS AND DISCUSSION

Table 4 presents the 5 most **strength** stations, categorized by weekdays and weekend, sorted by weekday strength. We use station strength to determine popularity, with each station's strength normalized to a daily value for comparison between the 5-day weekday and the 2-day weekend. The table shows that the 5 busiest weekend stations are also among the top 20 on weekdays. While, the 5 busiest weekday stations are still within the top 50 on weekends. This indicates a significant contrast in station usage patterns. However, it's worth noting that 3 out of the top 5 weekday stations also maintain their top 5 positions on weekends. Exceptions to the general patterns are observed in the cases of *Warehouse* and *Mountjoy Square*, where most trips are concentrated on weekdays for commuting. Additionally, *Dun Laoghaire Dart (station)* and *Blackrock Main St.* serve as popular weekend destinations.

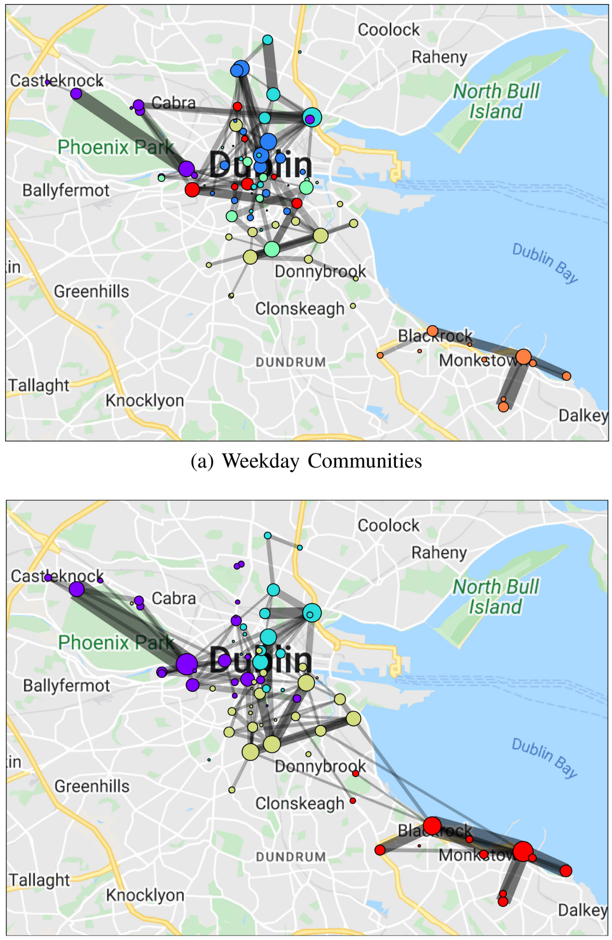
When examining the frequently traveled routes, as indicated by their **weights** in Table 5, the same variance can be observed. However, it's worth noting that the top 5 weekend routes are encompassed within the top 7 weekday routes. Furthermore, the strength in both weekday and weekend networks present a similar exponential decline as seen in



(a) Weekday

(b) Weekend

FIGURE 4. Daily Activity Networks. (Representation Networks of Weekend and Weekday Clusters).



(a) Weekday Communities

(b) Weekend Communities

FIGURE 5. Daily Activity Communities.

the SBiGN network, although the decline rates are steeper. Similarly, edge weights in the weekday and weekend TBiGN networks also demonstrate an exponential decline that is comparable to the SBiGN network. As is the case with Table 4, edge weight values are normalized to facilitate a comparison between a 5-day weekday and a 2-day weekend.

Weekday and weekend networks are illustrated in Figures 4(a) and 4(b), respectively. These figures show **strength** of stations and **weight** of edges in respective networks. The main difference is that in weekday networks, connections between residential areas and office areas are clearly identified by large red edges, especially on the north side of Dublin. On weekends, there are more trips between the centre and the Blackrock-Monkstown area, most likely leisure trips.

Figure 5 shows the community structure on weekday and weekend basic networks from the entire dataset. Subfigure 5(a) contains all weekday data, while Subfigure 5(b) contains all weekend data. The weekend network displays four distinct communities, whereas weekday networks exhibit spatially mixed communities, particularly around the centre of Dublin. Larger node sizes and wider edges denote higher trip volumes. In both weekday and weekend networks, Blackrock-Monkstown constitutes an independent community, and stations with strong connections in the suburbs belong to the same community throughout the entire week. The TBiGN provides an intriguing level of detail, revealing that weekday networks are divided into a larger number of communities compared to weekends. This observation suggests that on weekdays, people tend to follow established routes supported by a few highly connected edges, which in turn create many distinct communities. Conversely, on weekends, users tend to travel over a broader spatial range in a less predictable manner, resulting in larger and more interconnected communities.

3) MONTHLY TBIGNS - NETWORK CONSTRUCTION

Monthly TBiGNs are employed to examine variations in network activity on a monthly timeframe.

Our approach involved defining a rolling window size, with *size* = *4 weeks*, starting at week 1. This created 63 networks for the 66 week period in question, from the beginning of June 2020 to the end of August 2021. Nodes represent stations, and edges are weighted based on the number of trips occurring within each 4-week window. In addition to eliminating loop trips, we optimized the network

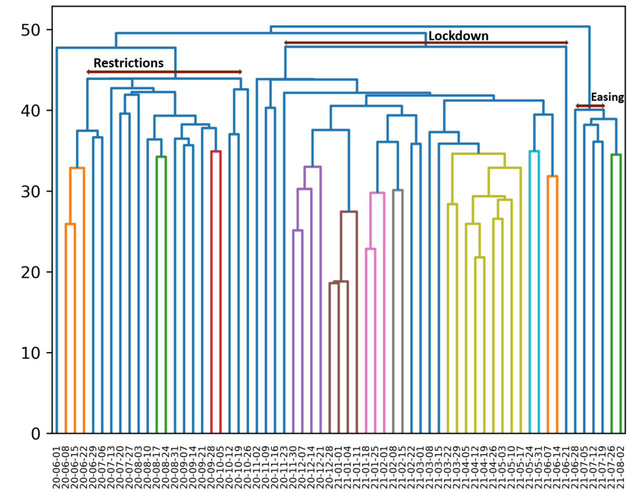


FIGURE 6. **Rolling Window Monthly Clustering.**

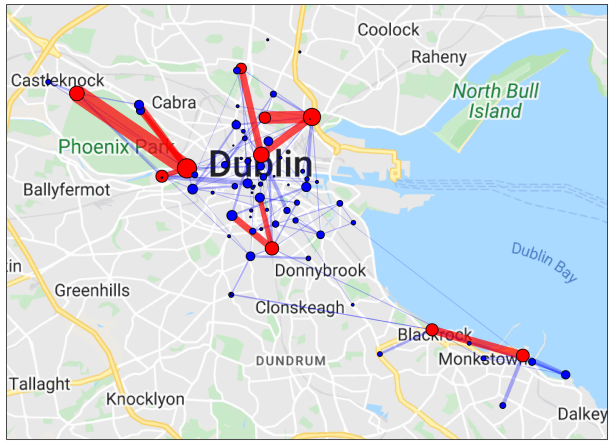
by removing edges with fewer than 3 trips, using a weekly threshold of 0.75. In each TBiGN, around 384 trips were removed, accounting for a total of 1,628 trips (approximately 23.6%).

4) MONTHLY TBIGNS - ANALYSIS AND DISCUSSION

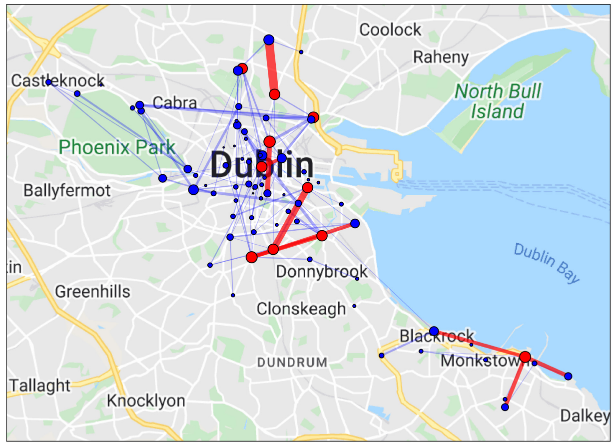
We used agglomerative hierarchical clustering with Euclidean distance, as explained in [35], to create a dendrogram showing network changes over time. Figure 6 displays the results of hierarchical clustering on a series of monthly basic networks, with the x-axis indicating the start date of each 4-week period. Notably, three clusters emerged. The first spans from June 2, 2020, to November 15, 2020, the second from November 2, 2020, to July 4, 2021, and the third covers the remainder, from June 28 to the end of August 2021. These clusters align with two waves of COVID-19 lockdowns (restrictions, lockdown) and reopening (easing), revealing how a major event can impact network activity in monthly temporal graphs.

Figures 7 illustrates the differences among the three TBiGN clusters over time.

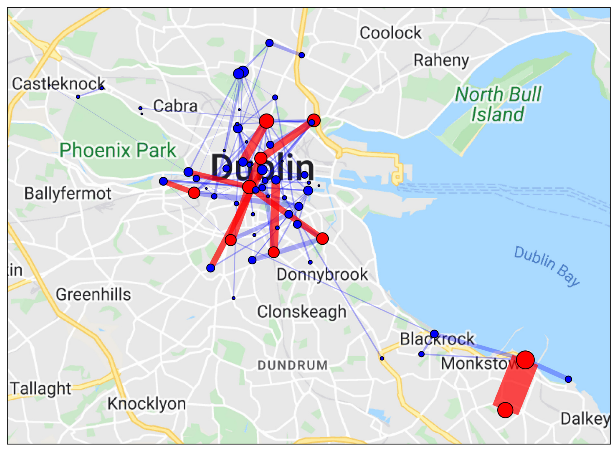
- *Restrictions Cluster.* In the cluster representing the first period from June to mid-November 2020 (Figure 7(a)), most of the activity occurs at only a few stations and routes, most notably *Phoenix Park* and *Fairview Park* to the city center. However, we find that the biggest stations and routes are not concentrated on a single area but are more widespread across central areas, with activity mainly near the city.
- *Lockdown Cluster.* In the second period from November 2020 to early July 2021 (Figure 7(b)), the volume of activity has reduced considerably due to stricter lockdown. Also, the top stations and routes become smaller and have changed with the top 10 most active stations no longer to the west of Dublin.
- *Easing Cluster.* In the final period in July and August 2021 (Figure 7(c)), activity levels have almost recovered. Trips occur more at the city centre resulting in top stations placed around that location. The top routes connect residential areas to the center.



(a) Restriction Cluster



(b) Lockdown Cluster



(c) Easing Cluster

FIGURE 7. **Geographical Plot of By Month Clusters. (Representation Networks of the Clusters).**

5) TBIGNS: SUMMARY

In TBiGNs, we addressed **Requirement 2** by presenting daily TBiGNs categorized and normalized into a weekday TBiGN and a weekend TBiGN. The data found in

Tables 4 and 5, as well as Figures 4 and 5 provides insights into various aspects, including **strength** and **communities** of stations, **weight** of edges, and activity networks (illustrating **strength** and **weight**).

Additionally, this type of graph addressed **Requirement 3** by clustering monthly TBiGNs over time. Figure 7 shows information about **strength** of stations and **weight** of edges in clusters relating the COVID-19 restrictions, lockdown and easing periods.

C. OUR SPATIO-TEMPORAL BIKE GRAPH NETWORKS

Spatio-Temporal Graphs enabled the analysis of temporal patterns that occur between spatial locations. Three levels of granularity, where the similarity between stations can be analysed across three temporal dimensions are supported in the system. These are hourly, where each of the 24 hourly intervals are averaged across the entire dataset, daily, where each day of the week is averaged, and monthly, where a single monthly value is calculated for each station over the 15 months of the study.

1) STBIGN CONSTRUCTION

A typical investigation for this type of network involves the growth of stations in proportion to the similarity between them. Nodes are stations and an edge connects two stations depending on the degree of similarity in terms of the number of trips over the period of time under analysis. The edge weight is the Pearson correlation coefficient between two timeseries. The similarity is only considered significant when the edge formed between 2 nodes has a coefficient value that exceeds the threshold, $T = 0.533$. This threshold ensures the strong connectivity of the network while maintaining the least possible number of edges, where $T = 0.533$ is statistically significant. The density of the network is 0.098, where density reflects how well connected the graph is, calculated as $\#edges \div \#possibleEdges$. For STBiGNs, the higher the value the greater the similarity between 2 stations. However, lower density values can indicate stations that have unique travel patterns.

2) MONTHLY STBIGNS - ANALYSIS AND DISCUSSION

It is important to note that large nodes in STBiGNs do not mean that stations are popular with many trips. Instead, the meaning is there are more stations that share the same properties and thus, 2 low activity stations may have large values. Furthermore, the edges in correlation networks are less affected by geographical locations and more by the properties of the station area. In essence, this graph identifies stations exhibiting analogous temporal patterns, characterized by similar trip volumes observed at corresponding time intervals. Figure 8 displays the network on a Dublin map, where five communities have been detected.

Overall, the transportation system sees growth until August 2020, then experiences a decline in the first few months of 2021. For the rest of the year, the total number of

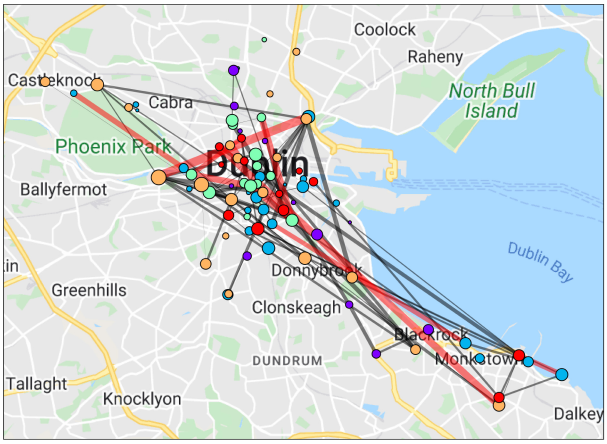


FIGURE 8. STBiGN Network: Monthly Timescale.

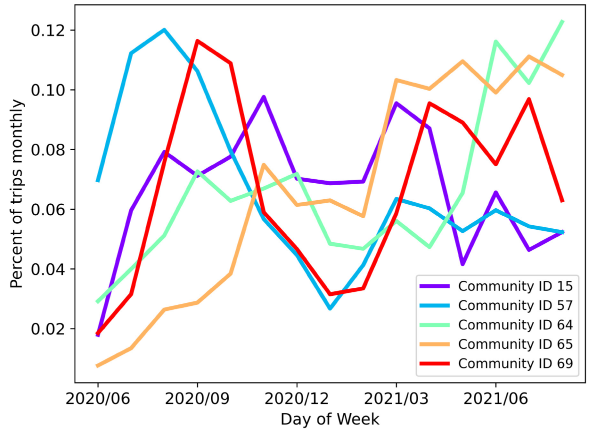


FIGURE 9. Timeseries Communities in Monthly STBiGNs.

trips almost recovers to its previous peak, a trend observed in TBiGN graphs but with more fine-grained detail here. Figure 9 plots the averaged monthly timeseries (every month in the dataset) for the communities identified in Figure 8 and provides insights into why stations (nodes) formed communities in distinct parts of the city. Community IDs are simply labels with nothing inferred from label numbers. In terms of size: community 15 (Purple) has 13 stations; 57 (Light Blue) has 26 stations; 64 (Light Green) has 15 stations; 65 (Orange) has 20 stations; and 69 (Red) has 12 stations. The light blue stations, primarily located in the south city center and the red stations, encompassing the surroundings of the city center, show quick growth in the first few months before experiencing a significant drop in activity in January 2021. While the red stations almost fully recover their level of activity, the light blue stations barely see any increase in the number of trips. The light green stations, found mainly in the north of the city center, and the orange stations, which are scattered mostly in the suburbs, display a steady growth, with a small decrease noticeable for the light green ones. Lastly, the purple stations, largely situated in the south where Dublin and Blackrock connect, initially show a growth in activity but maintain a stable level

TABLE 6. Node (station) strength in monthly STBiGNs.

Station	Strength	Avg. Cor.	#Pos	#Neu	#Neg
DCU Alpha	52.9	0.622	40	43	2
Rathmines	52.5	0.618	36	48	1
Dun Laoghaire Dart	52.5	0.618	33	51	1
...					
Pearse Street	43.8	0.515	15	59	11
Rathgar	43.7	0.514	23	38	24
Cathal Brugha Street	42.1	0.496	18	46	21
Sandymout Village	42.1	0.495	13	62	10
...					
Irishtown Rd	38.4	0.451	14	43	28
Parnell Street	38.3	0.451	6	56	23
Phoenix Park Gate	34.6	0.407	10	37	38

for most of the period, with a slight drop in the final few months.

To obtain more detail, it is necessary to analyse some of the raw data. Table 6 shows the node strength as well as the average weights, and the number of stations that are positively, neutrally and negatively correlated with the node in the monthly STBiGN. Columns represent the following values: *Strength* is the sum of weights to all other stations; *Avg. Cor.* is the average weight for that station; *#Pos*, *#Neu*, *#Neg* are the number of stations that are positive, neutral or negative to the station. Thresholds were computed as 0.65 and 0.35. A station has a *positive* relationship to another station if $w > 0.65$ is the weight between them, *negative* where $w \leq 0.35$, and otherwise, the relationship is *neutral*.

A complete Table 6 would comprise 86 rows but for the purpose of this discussion, we examine high, low and neutral strength stations. Highest strength stations such as *DCU Alpha* and *Rathmines*, are regarded as *central* nodes, exhibiting similarity to 47% and 42% of stations respectively. This indicates that they are good stations to study (perhaps using more detailed networks) to understand activity patterns of most stations. Interestingly, both of these stations are clustered into the purple community which has mixed patterns with all other communities. Neutral stations with average coefficients close to 0.5, such as *Pearse Street* and *Sandymount Village*, have minimal correlations with other stations. This may be an indication that they are not useful for analysis as they are almost equally positively and negatively correlated with the activity patterns of other stations. Low strength stations, including *Phoenix Park Gate*, *Parnell Street* and *Irishtown Road*, demonstrate monthly temporal similarities with only a few stations (up to 14), and have no correlations or are negatively correlated with the majority of the stations. However, this may make these stations interesting for analysis as they demonstrate unique activity patterns.

3) DAILY STBIGNS - ANALYSIS AND DISCUSSION

Dimensional modelling and analysis [36] often involves a hierarchy of analyses, especially when analyzing over time. Analyzing data at higher levels of abstraction can often

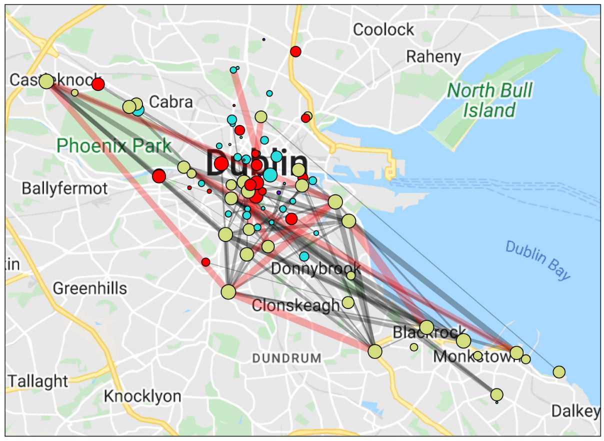


FIGURE 10. Daily Correlation Network.

identify periods that require more detailed scrutiny. A daily correlation network provides an analysis on a day by day basis for any period of time where activities are aggregated per day. Thus, two stations are connected when activities are similar in terms of daily comparisons. Timeseries in this network are vectors of length 7 whose entries corresponds to each day of the week. A timeseries data point is the total number of trips occurring at a specific day of the week at each individual station. An edge is retained only when the Pearson correlation coefficient between the two timeseries is above a threshold T, where the goal is to maintain a low threshold while ensuring a strongly connected network. For these experiments, we have $T = 0.609$ and a network density $D = 0.238$. Figure 10 shows the daily correlation network for the entire dataset. Only the top 5 percent of the edges are illustrated in the figure. The node colors represent detected clusters of the nodes. The top 10 edges are colored red while others are blue.

Three major communities and one 3-member one are detected in the day-of-week correlation network. It can be seen that the communities are slightly affected by their spatial locations. Cardinalities are: Community purple with 3 stations; light blue with 31 stations; light brown with 34 stations and red with 18 stations. The largest community (light brown) is mainly in the south of Dublin, including the entire Blackrock-Monkstown area but also includes some stations in the Phoenix Park. The light blue stations are mainly around the river Liffey. The red nodes are distributed around the centre and the north of Dublin. The light brown community is well connected with stations scoring high in strength and also including the highest station (Castleknock). This identifies all stations that have very common activity patterns. On the other hand, the light blue community contains stations with significantly lower strengths and without strong connections. We can also see that red stations can be strongly connected, even though they have a lower number of members.

Figure 11 displays daily timeseries for each community, where each day is averaged across the entire dataset.

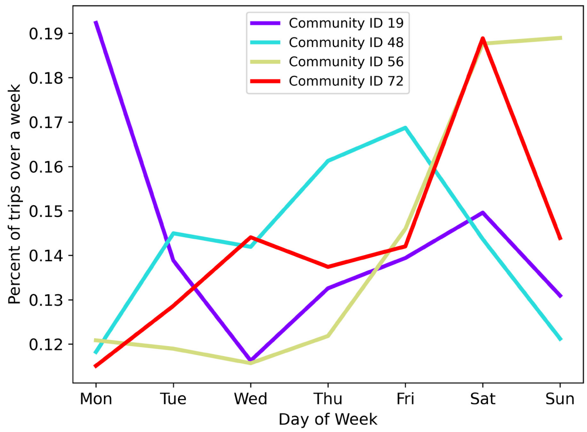


FIGURE 11. Timeseries Communities in Daily STBiGNs.

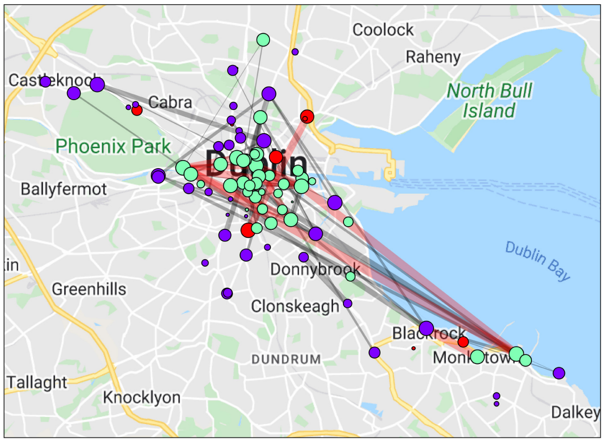


FIGURE 12. Hourly Correlation Network.

TABLE 7. Node (station) strength in daily STBiGNs.

Station	Strength	Avg. Cor.	#Pos	#Neu	#Neg
Castleknock	63.1	0.742	62	18	5
Luke Street	62.8	0.739	63	17	5
Irishtown Rd	62.5	0.735	62	16	7
...					
Westmoreland street	34.4	0.405	6	43	36
Blessington Street	30.4	0.358	7	29	49
Warehouse	26.5	0.312	12	23	50

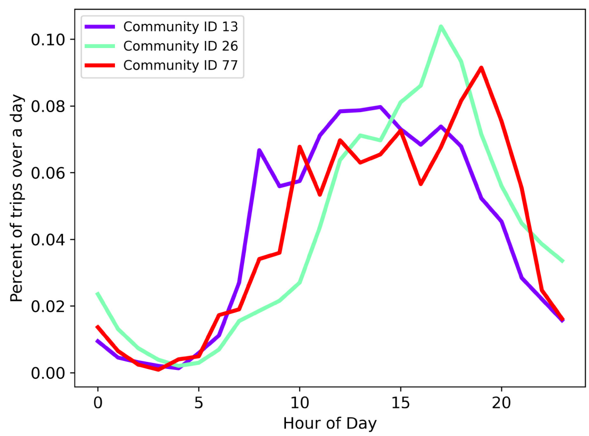


FIGURE 13. Timeseries Communities in hourly STBiGNs.

Generally the Moby Move service tends to be more active on the weekend, especially on Saturday. From analyses using the simpler networks, it was possible to determine that the light brown stations in these graphs are *weekend* stations. These stations are significantly inactive on weekdays, but starting from Friday the number of trips increases dramatically and peaks on Sundays. Similarly, red stations have low activity on Mondays and Tuesdays, but the rest of the week is steady except for Saturday where it peaks. As opposed to the other two communities, light blue stations are weekday stations where people use the service more on weekdays.

Table 7 uses the same structure as Table 6 to highlight the strongest stations in terms of daily correlations. The strengths and average coefficients have greater variance, ranging from 0.31 to 0.74. This indicates that a clearer pattern should emerge, forming a bigger community in the network. It also means that the characteristics of these communities are more distinguishable.

4) HOURLY STBIGNS - ANALYSIS AND DISCUSSION

Hourly correlation networks facilitate analyses of activity patterns on an hourly basis. Timeseries have a length of 24 points, corresponding to 24 hours. An entry of the timeseries is the total number of trips that happen at a specific hour at the station. The same criteria is again used to form edges between nodes, with $T = 0.737$ and a network density $D = 0.370$. Figure 12 shows the daily correlation network with just 3 communities and where Community 13 (purple) has 38 stations; 26 (light green) has 40 stations and 77 (red) has 8 stations. Node and edge sizes are proportional to the sum of coefficients of nodes and edges, respectively. Only the top 5% of edges are shown with node colors representing their clusters while the top 10 edges are red with all others being charcoal.

There are two large communities detected along with a small one. The light green community concentrates at the center of Dublin, the purple one surrounds the first community (mainly located in the suburbs) and the very small red scattered over a wide location. The central (light green) stations have the highest strength scores indicating a well connected community, sharing strong similarities. However, most of the strongest connections (edges) are in the purple group, even where station strength is small. This indicates that the activity pattern among them is very clearly defined. Similar to daily networks, we observe that stations can be strongly connected even though they are geographically remote. This indicates that remote stations can share some underlying characteristics, such as commuting behavior or activities in their social lives.

Figure 13 plots the averaged timeseries of the 3 communities with the same colors as used in Figure 12. The light green community is more active in the afternoons and

TABLE 8. Node (station) strength in hourly STBiGNs.

Station	Strength	Avg. Cor.	Pos	Neu	Neg
Grand Canal Docks	75.1	0.883	84	1	0
Irishtown Rd	75.0	0.882	85	0	0
Criminal Courts of Justice	75.0	0.882	83	2	0
...					
Dean Street	62.6	0.737	81	4	0
Clanbrassil St Lower	56.8	0.669	43	42	0
Warehouse	55.1	0.648	45	40	0

evenings, especially between 3pm to 7pm, while there are much less trips during the mornings. In contrast, purple stations are busier in the mornings and at noon but less during afternoons. The small red community provides little information except around 7pm where it peaks.

Similar to monthly and daily analyses, Table 8 displays the node strength for hourly networks, where here the threshold $T = 0.65$. Given that all stations exhibit a similar pattern with limited activity during the night and increased activity during day, it is reasonable that even the smallest strength nodes reach values as high as 55.1 or an equivalent average edge weight of 0.648, meaning that the station is significantly correlated with half of the stations. On the other hand, the highest strength stations such as *Grand Canal Docks* and *Criminal Courts of Justice*, exhibit significant correlations with almost every other station. These stations are typically large hubs with high traffic volumes, so it is understandable that they have mixed behaviors and shared patterns with other stations, as observed with fine grained hourly STBiGNs.

Having now generated all 3 types of STBiGNs using the entire dataset in each case, the heterogeneity across the 3 graphs is surprising and unexpected. There is no station that appears in the top 5 of each. Only *Irishtown Road* appears in the top 5 of two STBiGNs, occupying the 3rd place in Daily and 2nd place in Hourly. If we extend our analysis to the top 10 stations, once again, no station appears in all 3 graphs. Only the *Rathmines* station, which occupies 2nd place in Monthly and 6th in Daily and *Dun Laoghaire Dart*, which occupies 3rd place in Monthly and 7th in Daily appear in 2 top 10 rankings. Thus, stations are quite different when comparing these similarity patterns at different levels of granularity.

5) STBIGNS: SUMMARY

In STBiGNs, we addressed **Requirement 4** by supplying correlation networks (illustrating **strength**, **weight** and **communities**), and timeseries **communities** and **strength** of stations in Monthly (Figures 8 and 9, and Table 6), Daily (Figures 10 and 11, and Table 7), Hourly (Figures 12 and 13, and Table 8). These outputs provide insights regarding: the relationship between each station and all other stations; in addition to relationships between each community and all other communities based on timeseries comparisons.

VII. CONCLUSION

Launched in 2019, Moby Move is a Dublin bike sharing scheme with an extensive network across Dublin city. A research collaboration between Dublin City University and the City University of New York teamed with Moby Move to analyze network usage patterns with the goal of detecting areas of improvement in terms of optimizing the overall network. The organisation behind Moby Move needed to extract analytics from their trip database to: have a high level visualization of the levels of activity across the network; have the ability to examine spatial regions over specified time intervals; and to identify stations that are similarly connected in terms of spatio-temporal metrics. As spatio-temporal analyses can be difficult to interpret, we adopted a graph-based model as it enables a smoother representation of spatial data with temporal aspects captured in relationship between stations. A gradual approach to more deeper analyses is enabled by 3 different types of graph networks. Our proposed method is transferable and not limited to Moby Move. It enables a more precise interpretation of network dynamics at a granular level, contributing to the optimization of transportation networks and enhancing sustainability in shared transportation systems.

Spatial Graph Networks are easiest to construct and have the additional benefit of fast identification of strong (many trips) stations but also the closeness metric quickly identifies popular station-to-station trips. The Betweenness metric easily identifies a *bridge* station through which many paths must cross. This typically means that the removal of this station could segment the network and suggests the need for a new station nearby. To drill-down the information supplied by SBiGNs, the Temporal Graph Networks add the time dimension across different levels of granularity. Using monthly networks, trends (academic year) or events (extreme weather or in the case of this dataset, COVID-19) can be observed, as discussed in the findings on the activity levels during COVID-19 restrictions and subsequent easing.

Spatio-Temporal Graph Networks are used to compare stations in terms of their activity patterns over a specified time period. This can be useful in observing how 2 stations that may be in quite different locations, can have the same levels of activity; or identify a station with quite a unique activity level; or detect potential anomalies such as stations which should have very similar activity patterns but belong to different communities. What was unexpected from this analysis was the discovery that almost none of the stations had the same activity patterns across different time granularities. The uncovered hidden patterns offers strong evidence for the usage of graph-based systems when modelling transportation networks.

REFERENCES

[1] Moby. “Bikes as a service.” 2019. [Online]. Available: https://mobybikes.com/about-moby

[2] B. Caulfield, M. O’Mahony, W. Brazil, and P. Weldon, “Examining usage patterns of a bike-sharing scheme in a medium sized city,” *Transp. Res. Part A, Policy Pract.*, vol. 100, pp. 152–161, Jun. 2017.

[3] R. Papa, I. Cardei, and M. Cardei, “Generalized path planning for UTM systems with a space-time graph,” *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 351–368, 2022.
[4] 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,” *J. Clean. Prod.*, vol. 313, Sep. 2021, Art. no. 127930.
[5] 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,” *Comput., Environ. Urban Syst.*, vol. 77, Sep. 2019, Art. no. 101361.
[6] 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 *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, 2022, pp. 5–8.
[7] X. Y. Shi, Y. Wang, F. Lv, W. Liu, D. Seng, and F. Lin, “Finding communities in bicycle sharing system,” *J. Visual.*, vol. 22, no. 6, pp. 1177–1192, Dec 2019.
[8] S. F. A. Batista, M. Ameli, and M. Menéndez, “On the characterization of eco-friendly paths for regional networks,” *IEEE Open J. Intell. Transport. Syst.*, vol. 4, pp. 204–215, 2023.
[9] 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,” *Transp. Res. Part C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018.
[10] Z. Ghandeharioun and A. Kouvelas, “Link travel time estimation for arterial networks based on sparse GPS data and considering progressive correlations,” *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 679–694, 2022.
[11] 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, Nov. 2023, Art. no. 104554.
[12] P. Borgnat, C. Robardet, J.-B. Rouquier, P. Abry, P. Flandrin, and E. Fleury, “Shared bicycles in a city: A signal processing and data analysis perspective,” *Adv. Complex Syst.*, vol. 14, no. 3, pp. 415–438, 2011.
[13] M. Z. Austwick, O. O’Brien, E. Strano, and M. Viana, “The structure of spatial networks and communities in bicycle sharing systems,” *PLoS ONE*, vol. 8, no. 9, 2013, Art. no. e74685.
[14] 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, p. 5425, 2019.
[15] H. Lu, M. Halappanavar, and A. Kalyanaraman, “Parallel heuristics for scalable community detection,” *Parallel Comput.*, vol. 47, pp. 19–37, Aug. 2015.
[16] M. Roantree and J. Liu, “A heuristic approach to selecting views for materialization,” *Softw. Prac. Exp.*, vol. 44, no. 10, pp. 1157–1179, 2014.
[17] J. Hamilton, *Time Series Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1994.
[18] F. Bahrpeyma, M. Roantree, P. Cappellari, M. Scriney, and A. McCarren, “A methodology for validating diversity in synthetic time series generation,” *MethodsX*, vol. 8, Jan. 2021, Art. no. 101459.
[19] V. M. Ngo, S. Helmer, N.-A. Le-Khac, and M.-T. Kechadi, “Structural textile pattern recognition and processing based on hypergraphs,” *Inf. Retr. J.*, vol. 24, pp. 137–173, Apr. 2021.
[20] V. M. Ngo, G. Munnelly, F. Orlandi, P. Crooks, D. O’Sullivan, and O. Conlan, “A semantic search engine for historical handwritten document images,” in *Proc. 25th Link. Int. Conf. Theory Prac. Digit. Libr. (TPDL)*, 2021, pp. 60–65.
[21] D. Xin, J. Han, X. Li, Z. Shao, and B. W. Wah, “Computing iceberg cubes by top-down and bottom-up integration: The starcubing approach,” *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 1, pp. 111–126, Jan. 2007.
[22] K. Mehlhorn, S. Näher, and P. Sanders, “Engineering DFS-based graph algorithms,” 2017, *arXiv:1703.10023*.
[23] H. Almeida, D. Guedes, W. Meira, and M. J. Zaki, “Is there a best quality metric for graph clusters?” in *Proc. Mach. Learn. Knowl. Discov. Databases*, 2011, pp. 44–59.
[24] J. M. Hernández and P. V. Mieghem, “Classification of graph metrics,” Fac. Electr. Eng., Math., Comput. Sci., Delft Univ. Technol., Delft, The Netherlands, Rep. 20111111, Nov. 2011.
[25] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks,” *Proc. Nat. Acad. Sci.*, vol. 101, no. 11, pp. 3747–3752, 2004.
[26] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1979.
[27] G. Sabidussi, “The centrality index of a graph,” *Psychometrika*, vol. 31, no. 4, pp. 581–603, Dec. 1966.
[28] U. Brandes and C. Pich, “Centrality estimation in large networks,” *Int. J. Bifurcat. Chaos*, vol. 17, no. 7, pp. 2303–2318, 2007.
[29] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.
[30] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.
[31] S. Fortunato, “Community detection in graphs,” *Phys. Rep.*, vol. 486, nos. 3–5, pp. 75–174, 2010.
[32] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: Uses and interpretations,” *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, Oct. 2010.
[33] K. Steinhaeuser, N. V. Chawla, and A. R. Ganguly, “Complex networks as a unified framework for descriptive analysis and predictive modeling in climate science,” *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 4, no. 5, pp. 497–511, 2011.
[34] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. “Complex networks in climate dynamics,” *Eur. Phys. J. Spec. Topics*, vol. 174, no. 1, pp. 157–179, Jul. 2009.
[35] P. Baby and K. Sasirekha, “Agglomerative hierarchical clustering algorithm- a review,” *Int. J. Sci. Res. Publ.*, vol. 83, p. 83, Mar. 2013.
[36] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Hoboken, NJ, USA: Wiley, 2011.