Understanding urban dynamics using graph-based analysis of public bike-sharing schemes

Yuanxuan Yang

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Geography

April, 2021

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The following chapters contain jointly authored manuscripts where Yuanxuan Yang is the lead author:

The work in Chapter 3 of the thesis has appeared in the publication as:

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2019. A spatiotemporal and graph-based analysis of dockless bike-sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems.* **77**, p101361.

Yuanxuan Yang was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Alexis Comber, Alison Heppenstall and Andy Turner are the supervisors and supervised this study. They provided conceptual direction, advised on study design, analyses and discussion, and made comments and edits on the manuscript from draft to proofs.

The work in Chapter 4 of the thesis has appeared in the paper as:

Yang, Y., Beecham, R., Heppenstall, A., Turner, A., & Comber, A. *Understanding the impacts of public transit disruptions on bike-sharing schemes and cycling behaviours using spatiotemporal and graph-based analysis: A case study of four London Tube strikes.* Submitted to a peer-reviewed journal.

Yuanxuan Yang was responsible for the study design, data analysis, preparation of figures, and writing the manuscript. Alexis Comber, Alison Heppenstall and Andy Turner are the supervisors and supervised this study. They provided conceptual direction, advised on study design, analyses and discussion, and made comments and edits on the manuscript. Roger Beecham provided data used in this study, advised on experiment design and made comments and edits on the manuscript from draft to proofs.

The work in Chapter 5 of the thesis has appeared in the publication as:

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. 2020. *Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems.* Computers, Environment and Urban Systems. 83, p101521.

Yuanxuan Yang was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Alexis Comber, Alison Heppenstall and Andy Turner are the supervisors and supervised this study. They provided conceptual direction, advised on study design, analyses and discussion, and made comments and edits on the manuscript from draft to proofs.

Acknowledgements

I would be nowhere in this bittersweet journey of PhD study without the kind care and support from the amazing people around me.

First and foremost, I would like to express my profound admiration and gratitude to my supervisors Lex Comber, Alison Heppenstall and Andy Turner. I greatly appreciate Lex for introducing me to the world of GIS and the comforting foods and beverages in the UK. Thanks Alison for her “red pens” - full of wisdom and encouragement, and thanks Andy for his tireless support and sharing the good stories of pioneers in GeoComputation. I have learnt a lot from my supervisors about how to be a brave and honest person, and to enjoy the frustration and gratification while pursuing study.

Thanks to my research support group panel members and transfer examiners, Andy Evans, Myles Gould, Richard Romano and Roger Beecham, who have reviewed parts of this research and shared insightful comments and suggestions.

Thanks to the School of Geography, especially the Centre for Spatial Analysis and Policy and Jacqui Manton for assisting me in this endeavour from application to completion.

Thanks to the China Scholarship Council (No. 201606420071) and The University of Leeds for funding this PhD project.

Huge thanks to my family and friends, for bearing with me and sharing with me the joy and frustration of completing this study.

Abstract

Urbanisation has contributed to more dense and diverse travel flows and interaction between people and urban spaces. In recent years, mobility systems have incorporated more Information technology and sensors, where the movement of people can be sensed and captured at high frequency. The increased availability of such data enables people and their interactions with the physical environment to be analysed and understood within the context of complex urban systems. It also has the potential to be used for evidence-based urban planning and management. However, problems such as congestion and air pollution continue to happen as results of high car-dependency in many cities. To overcome the challenges, sustainable modes of transportation such as cycling is (re)emerging through bike-sharing schemes as an increasingly common response. Analysing flow interactions is key to promoting and developing sustainable and integrated transportation systems in cities.

This study has aimed to better understand and model dynamics within sustainable urban transport systems through spatiotemporal and novel graph-based analysis. Large quantities of travel flows in different and newly emerged (e.g.dockless) bike-sharing schemes are investigated with high spatial and temporal granularity at the individual level. Weekly, daily and hourly graph structures are modelled to uncover and quantify how travel behaviours are impacted by the built environment, by changes in transportation systems, and as a result of events that disrupt the normal patterns of movement. The revealed mobility patterns and graph structural changes indicate underlying urban resilience, where individuals can rapidly adopt bike-sharing to fill transit gaps that emerged. Furthermore, this thesis, for the first time, quantifies the utility of temporal graph indices for enhancing machine learning models to making predictions on dynamics and demand in urban traffic systems. This research has great potential to help improve transport management and planning in urban areas.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Research context

With unprecedented global urbanisation, especially in developing countries (Chen et al., 2013; Glaeser, 2014; Jiang et al., 2017; Smith, 2019), 68% of the world population are projected to live in urban areas by 2050 (UNDESA, 2018). Cities provide people with housing, energy, employment, transportation, communication and other extensive systems for wide-ranging social and economic activities (Batty, 2018). Yet, intensive urbanisation also brings significant challenges for planning and management (Duhl, 1986). To solve problems like traffic congestion and public service equity, it is crucial to understand the interactions among different urban systems and those resident and moving in and through cities.

Cities can be viewed as complex systems constructed from the bottom up (Batty, 2013), the diverse behaviours of individuals collectively contribute towards the entire system ideally building liveable and sustainable urban areas (Makki et al., 2015; Batty, 2018; Freytag et al., 2018). These individual behaviours are the fabric of urban space and represent interactions among different people and entities. Intra-urban movement (e.g. commuting, trips for commercial and social activities) generate intra-urban flows, and they help to define how elements in cities are related to each other (Batty, 2018). These movements and flows “pulse” to periodic rhythms and reflect the activities, constraints and opportunities afforded to individuals and organisations (Froehlich et al., 2009; Graells-Garrido et al., 2017; Batty, 2018).

Considering the temporal nature of flows and interactions in cities is key to understand their dynamism, evolution and discontinuities (Batty, 2018; Anderson et al., 2020). Although there have been attempts to examine abrupt changes in urban form, activity and development caused by catastrophes (e.g. earthquake) and other singularities (Drakakis-Smith, 1995; Ghafory-Ashtiany et al., 2008), they are normally evaluated in the long term (years, even decades) rather than shorter periods (Johnson et al., 2017; Batty, 2018). In the last decade, more studies examining shorter term changes in urban flow interactions have been reported (Gao et al., 2013; Zhang et al., 2019). These are fuelled by larger and more detailed volumes of data that are made more available for research and technology innovations (e.g. sensors and increased computation power and software). In urban areas, huge volumes of diverse data about peoples' movements, activities and opinions are now readily gathered at a very fine spatiotemporal granularity, and at the individual level (Filipponi et al., 2010; Du et al., 2018). Such data are beginning to be used for urban planning and real-time management (Chen et al., 2018; Hossain et al., 2019). A number of the "smart city" monitoring and control centres have surged in big cities around the world (Du et al., 2018), where road and rail traffic, in particular is continually monitored (Batty, 2018), with responsive interventions ready to take place to ensure transport systems work smoothly and efficiently. In essence, the massive change in research focus and data availability from long-term to short-term, from aggregated to individual level, from coarse to fine spatial scale, has provided new opportunities to study and understand urban transport dynamics and the evolutions of cities in general.

Managing traffic is one of the most crucial issues to ensure efficient intra-urban flows and to maintain different urban systems (Batty, 2018). Despite huge investments, the current transportation systems in many big cities, such as Beijing and Los Angeles (Lee et al., 2014), still suffer problems of congestion and air pollution, related to over dependency on the use of polluting car vehicle-based transportation (Li et al., 2012; Chow et al., 2014; Fishman, 2016; Rodríguez et al., 2016).

To overcome the challenges presented by car dependence, sustainable modes of transportation such as cycling are (re)emerging as an increasingly common response in big cities (Pucher et al., 2012; Fishman, 2016). In the last 20 years or so, many different bike-sharing schemes were introduced in many cities throughout the world. These schemes allow people to use (borrow) bikes with low-cost and high convenience. In the age of the smart mobile phone, the bike-sharing mode of transportation has arguably become the fastest-growing transportation innovation in the world (Shaheen et al., 2016). In addition to replacing short-distance car journeys, bike-sharing has the potential to strengthen the existing public transit system, by connecting transit hubs (e.g. bus and metro station) to people's destinations such as their home and workspace, thus solving the so-called "first/last mile" problem (Shaheen & Chan, 2016). The forms of bike-sharing are also rapidly evolving and incorporating more technological innovations. Bike docking stations are able to provide a 24 hours service and allow people to borrow and return cycles at any time. Recent dockless systems equip bikes with GPS (Global Positioning

System) units and smart locking system, which allows them to be more widely distributed in urban spaces and not limited to docking stations. These new schemes generate much spatial and temporal flexibility for peoples' movements, especially in multi-mode transport journeys. As this innovation is new, there is a challenge and opportunity to study how it may impact the way people move around and how best to plan and (re)organise other transport systems accordingly. The high spatial and temporal resolution data about the availability or movement of such sharable bikes offers a new way to study the movements of users and how these patterns change at different times and during different events (Pase et al., 2020) in the transport system or in the cities generally.

Various studies have explored the dynamics of intra-urban flows, especially people's movements, utilising taxi, bus and metro travel records (Gao et al., 2013; Zhong et al., 2014; Sun et al., 2015). These studies (for the most part) have only helped to understand car-dependent flows characterised by relatively long journeys on the main roads in urban areas. Flows of shorter-distance journeys, many of which do not involve motorised traffic, require further studies to shed light on. This will also contribute to an improved understanding of urban sustainability (Curtis et al., 2020).

Successful interpretation of urban and traffic dynamics also requires suitable models to be utilised. Graph theory is a branch of mathematics, and it is well suited to representing the flows and structures encoded in complex urban systems (Batty, 2005; Anderson & Dragićević, 2020). Owing to its advantages in representing interactions between different entities, graph-based analysis has been utilised in a number of transportation and urban morphology studies (Austwick et al., 2013; Heckmann et al., 2015). As a bottom-up approach, modelling and analysing the interactions through graphs helps uncover the behaviours and patterns such as congestion and transport efficiency (Crucitti et al., 2006; Scellato et al., 2006). Many existing studies using graph theory focusing on characterising the static and general picture of the urban system. There is an ongoing challenge to adopt graph theory to understand and even predict the highly dynamic interaction flows. Measuring the temporal mobility graphs in urban contexts enables further insight to be obtained and helps answer vital questions regarding where, when and why people move. The knowledge obtained from mobility graphs will thus facilitate data-driven and analytics-powered smart and sustainable city management.

1.2 Research aims and objectives

The overall aim of this work was to investigate the characteristics of flow interactions and structures evolutions in sustainable urban traffic systems (bike-sharing) using novel temporal graph-based analysis. In addition, the study sought to build better models to forecast the transport dynamics and demand by incorporating graph theory. This was to improve urban planning and responsive transport management and operation practices.

This aim and desired outcomes can only be accomplished by breaking down the research into a number of objectives, and they were summarised in table 1.1.

Table 1.1 Research objectives and corresponding chapters.

	Research objectives	Corresponding chapter(s)
I.	(1) Design a method to convert travel flows from newly emerged sustainable transport mode (dockless bike-sharing) into graph structures. (2) Quantify the unique spatial and temporal characteristics of cycling behaviours in dockless bike-sharing. (3) Identify and understand the polycentric transformation in urban last mile through graph-based analysis.	Chapter 3
II.	(4) Quantify the interdependence between metro and bike-sharing in the context of different transit disruptions and interventions. (5) Demonstrate the utility of temporal graph-based analysis in capturing rapid changing interactions between people and places.	Chapter 3 and Chapter 4
III.	(6) Quantify the importance of temporal graph structures and indices in predicting short-term travel demand. (7) Design machine learning models to effectively process temporal graph features for better traffic modelling and demand forecasting.	Chapter 5

1.3 Thesis structure

This thesis is presented in the alternative format as described by the University of Leeds. The thesis consists of six chapters, chapter 3, 4 and 5 take the form of a self-contained manuscript, all of which have been published or submitted to peer-reviewed journals.

Chapter 2 provides a critical literature review of the methodology related to understanding urban flow interactions. The data and applications of constructing flow networks and graph structures in urban and transport systems (e.g. bike-sharing) are also introduced and critically examined. This chapter highlights the limitations and current research gaps: it is necessary to examine temporal urban flows interactions, with the context of local processes to understand and model the dynamics in mobility behaviours. The findings of this chapter would motivate the direction of research in the following chapters.

The work in Chapter 3 of the thesis has appeared in the publication as:

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems.* **77**, p101361.

Chapter 3 adopted geospatial statistics as well as graph-based approaches to reveal how urban flows in “capillary” are affected by “aorta”. The evolution in the flow network is examined using newly emerged dockless bike-sharing data. This new kind of travel mode can provide a large quantity of data in high spatiotemporal granularity and describe people’s movement in non-motorised traffic. Thus the study helps to shed more detailed lights on urban flows over the last mile.

The work in Chapter 4 of the thesis has appeared in the paper as:

Yang, Y., Beecham, R., Heppenstall, A., Turner, A., & Comber, A. *Understanding the impacts of public transit disruptions on bike-sharing schemes and cycling behaviours using spatiotemporal and graph-based analysis: A case study of four London Tube strikes.* Submitted to a peer-reviewed journal.

Chapter 4 sought to provide a more comprehensive picture of the interdependency between mass transit and dock-based bike-sharing scheme. Cycling flow and behaviours of users, and bike-sharing service provisions during transit disruptions are examined. A combination of spatiotemporal and graph-based analysis helped to demonstrate the flexibility nature of bike-sharing and resilience in the urban system.

The work in Chapter 5 of the thesis has appeared in the publication as follows:

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. 2020. *Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems.* Computers, Environment and Urban Systems. **83**, p101521.

Chapter 5 proposed a methodology of incorporating graph features for better modelling and forecasting transport dynamics and short-term travel demands. This chapter quantified and compared the importance of time-lagged graph indices to other commonly used model inputs. Machine learning models are examined in term of how they might be enhanced by knowledge obtained from temporal graph structures.

Chapter 6 is the conclusion of this thesis. The major findings of the research are consolidated, and this chapter summarised the contributions to the literature in the broader community of urban and transport studies. It is then followed by the critical discussion of limitations and outlooks.

References

Anderson, T. and Dragićević, S. 2020. Complex spatial networks: Theory and geospatial applications. *Geography Compass.***14**(9), pe12502.

Austwick, M.Z., O'Brien, O., Strano, E. and Viana, M. 2013. The structure of spatial networks and communities in bicycle sharing systems. *PloS One.* **8**(9), pe74685.

Batty, M. 2005. Network geography: Relations, interactions, scaling and spatial processes in GIS. In: *Re-presenting GIS.* Hoboken: Wiley press pp.149-170.

Batty, M. 2013. *The new science of cities.* Cambridge: MIT press.

Batty, M. 2018. *Inventing future cities.* Cambridge: MIT press.

Chen, M., Liu, W. and Tao, X. 2013. Evolution and assessment on China's urbanization 1960–2010: under-urbanization or over-urbanization? *Habitat International.* **38**, pp.25-33.

Chen, Y. and Han, D. 2018. Water quality monitoring in smart city: A pilot project. *Automation in Construction.* **89**, pp.307-316.

Chow, A.H., Santacreu, A., Tsapakis, I., Tanasaranond, G. and Cheng, T. 2014. Empirical assessment of urban traffic congestion. *Journal of advanced transportation.* **48**(8), pp.1000-1016.

Crucitti, P., Latora, V. and Porta, S. 2006. Centrality measures in spatial networks of urban streets. *Physical Review E.* **73**(3), p036125.

Curtis, S.K. and Mont, O. 2020. Sharing economy business models for sustainability. *Journal of Cleaner Production.* **266**, p121519.

UNDESA. 2018. 68% of the world population projected to live in urban areas by 2050, says UN. [Online]. [Accessed 1 December 2020]. Available from: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

Drakakis-Smith, D. 1995. Third World Cities: Sustainable Urban Development, 1. *Urban Studies.* **32**(4-5), pp.659-677.

Du, R., Santi, P., Xiao, M., Vasilakos, A.V. and Fischione, C. 2018. The sensable city: A survey on the deployment and management for smart city monitoring. *IEEE Communications Surveys & Tutorials.* **21**(2), pp.1533-1560.

Duhl, L.J. 1986. The healthy city: Its function and its future. *Health Promotion International.* **1**(1), pp.55-60.

Filipponi, L., Vitaletti, A., Landi, G., Memeo, V., Laura, G. and Pucci, P. 2010. Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors. In: *2010 Fourth International Conference on Sensor Technologies and Applications*. 18-25 July 2010, Venice. Hoboken: Wiley-IEEE press, pp.281-286.

Fishman, E. 2016. Bikeshare: A review of recent literature. *Transport Reviews.* **36**(1), pp.92-113.

Freytag, T. and Bauder, M. 2018. Bottom-up touristification and urban transformations in Paris. *Tourism Geographies.* **20**(3), pp.443-460.

Froehlich, J.E., Neumann, J. and Oliver, N. 2009. Sensing and predicting the pulse of the city through shared bicycling. In: *Twenty-First International Joint Conference on Artificial Intelligence*. 11-17 July 2009, Pasadena. California: IJCAI Organization, pp.1420-1426.

Gao, S., Wang, Y., Gao, Y. and Liu, Y. 2013. Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. *Environment and Planning B: Planning and Design.* **40**(1), pp.135-153.

Ghafory-Ashtiany, M. and Hosseini, M. 2008. Post-Bam earthquake: recovery and reconstruction. *Natural Hazards.* **44**(2), pp.229-241.

Glaeser, E.L. 2014. A world of cities: The causes and consequences of urbanization in poorer countries. *Journal of the European Economic Association.* **12**(5), pp.1154-1199.

Graells-Garrido, E., Ferres, L., Caro, D. and Bravo, L. 2017. The effect of Pokémon Go on the pulse of the city: a natural experiment. *EPJ Data Science.* **6**(1), pp.1-19.

Heckmann, T., Schwanghart, W. and Phillips, J.D. 2015. Graph theory—Recent developments of its application in geomorphology. *Geomorphology.* **243**, pp.130-146.

Hossain, M.S., Muhammad, G. and Alamri, A. 2019. Smart healthcare monitoring: A voice pathology detection paradigm for smart cities. *Multimedia Systems.* **25**(5), pp.565-575.

Jiang, L. and O'Neill, B.C. 2017. Global urbanization projections for the Shared Socioeconomic Pathways. *Global Environmental Change.* **42**, pp.193-199.

Johnson, J., Nowak, A., Ormerod, P., Rosewell, B., & Zhang, Y. C. 2017. *Non-equilibrium social science and policy: introduction and essays on new and changing paradigms in socio-economic thinking*. London: Springer Nature.

Lee, E.Y., Jerrett, M., Ross, Z., Coogan, P.F. and Seto, E.Y. 2014. Assessment of traffic-related noise in three cities in the United States. *Environmental research.* **132**, pp.182-189.

Li, Z. and Hensher, D.A. 2012. Congestion charging and car use: A review of stated preference and opinion studies and market monitoring evidence. *Transport Policy.* **20**, pp.47-61.

Makki, A.A., Stewart, R.A., Beal, C.D. and Panuwatwanich, K. 2015. Novel bottom-up urban water demand forecasting model: revealing the determinants, drivers and predictors of residential indoor end-use consumption. *Resources, Conservation and Recycling.* **95**, pp.15-37.

Pase, F., Chiariotti, F., Zanella, A. and Zorzi, M. 2020. Bike Sharing and Urban Mobility in a Post-Pandemic World. *IEEE Access.* **8**, pp.187291-187306.

Pucher, J.R. and Buehler, R. 2012. *City cycling.* Cambridge: MIT press.

Rodríguez, M.C., Dupont-Courtade, L. and Oueslati, W. 2016. Air pollution and urban structure linkages: Evidence from European cities. *Renewable and Sustainable Energy Reviews.* **53**, pp.1-9.

Scellato, S., Cardillo, A., Latora, V. and Porta, S. 2006. The backbone of a city. *The European Physical Journal B-Condensed Matter and Complex Systems.* **50**(1-2), pp.221-225.

Shaheen, S. and Chan, N. 2016. Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections. *Built Environment.* **42**(4), pp.573-588.

Smith, D.O. 2019. *Third world cities in global perspective: the political economy of uneven urbanization.* Abingdon: Routledge.

Sun, L., Jin, J.G., Axhausen, K.W., Lee, D.-H. and Cebrian, M. 2015. Quantifying long-term evolution of intra-urban spatial interactions. *Journal of The Royal Society Interface.* **12**(102), p20141089.

Zhang, Y., Cheng, T. and Ren, Y. 2019. A graph deep learning method for short‐term traffic forecasting on large road networks. *Computer‐Aided Civil and Infrastructure Engineering.* **34**(10), pp.877-896.

Zhong, C., Arisona, S.M., Huang, X., Batty, M. and Schmitt, G. 2014. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science.* **28**(11), pp.2178-2199.

Chapter 2
Literature review

Geography studies the integration of characteristics that define places, as well as connections between places and entities (National Research Council, 1997). It also aims to understand and model highly-connected spatial systems like cities and the sub-systems within (e.g. transportation) (Sun et al., 2015; Phillips et al., 2015). Researchers in GeoComputation have recognised the value of complexity science in understanding geography systems, with attempts to interpret and model the real-world from bottom-up (Batty, 2005; Crooks and Heppenstall, 2012), as well as represent the relationships between entities using graph structure and network science (Anderson and Dragićević, 2020).

The traditional top-down approach of using aggregate models to formulate policy and plans for the design and growth of cities have several drawbacks due to their aggregated and coarse treatment of individuals (Heppenstall et al., 2016). In contrast to the top-down way, bottom-up approaches seek to understand how different and heterogeneous entities (e.g. people, infrastructure) interact with each other and self-organise to form emergent patterns (Heppenstall et al., 2016; Anderson and Dragićević, 2020). These emerged behaviours (e.g. commute and travel), will lead to various socioeconomic outcomes for cities, such as congestions and the emergence of new urban centres (Zhong et al., 2014).

Examining the static structure of urban transportation systems is not sufficient to gain insights and understand the impact of changing dynamics; hence representing the dynamics in graphs and how they evolve becomes increasingly vital (Anderson and Dragićević, 2020). The inclusion of time is not simply gathering different snapshots (or periods) of the system, they need to be linked to the changed observations in the system, and then interpreted and understood on the basis of the local context and underlying processes (Batty, 2005; Anderson and Dragićević, 2020). The world is now more digitalised than ever before, and the behaviours and states of entities are recorded at an increasingly finer temporal granularity. There is an abundance of information encoded in various datasets to describe the interactions of people and places. New travel mode and mobility services (bike-sharing) have also generated valuable data to shed light on the movement of people. Hence new opportunities have emerged for uncovering more details in urban dynamics and flow interaction.

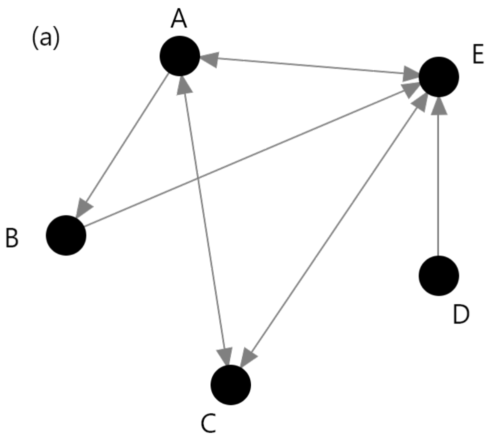
This chapter provides a comprehensive review of studies related to graph theory and their applications in urban and transport studies. The review suggests that existing research majorly focuses on static graph structures to understand the roles and characteristics within them. There is a need for exploring the temporal characteristics of real-world urban and transport systems. This will facilitate a better understanding and thus prediction, of urban and transport dynamics, for example, the behaviours in public transport and bike-sharing systems. **Section 2.1 introduces the basic definitions in graph theory and related measures that can be used to describe various states of graphs. Section 2.2 presents types and applications of graphs in geography, especially urban and transportation systems; Specifically, the flow and graph structures of bike-sharing systems are discussed in section 2.3.**

2.1 Graph and related indices

Graph Theory was born with the well-known story "Seven Bridges of Konigsberg" (Euler, 1741). In that story (model), Euler utilised graph theory to prove that it is impossible to walk through all the islands and cross each of those bridges once and only once. Therefore, graph theory was considered as a logic solving approach back to its origin (Euler, 1741), and traditionally often used for solving routing problems. It is concluded that finding (or not finding) the route is not a matter of intelligence; it is simply an inherent property of the graph (Anderson and Dragićević, 2020). Euler's findings formed the foundation of network science and gave way to the notion that graphs have different structural properties that can be discovered and catalogued using graph theory (Barabási, 2016). In the late 1990s (Vespignani, 2018; Heckmann et al., 2015; Newman, 2003b), networks (graphs) from very diverse fields of science were found to exhibit similar properties. This has resulted in considerable research into graph-based analysis over the last two decades (Vespignani, 2018).

Generally, a graph $G(N, E)$ is a structure consisting (figure 2.1 a) of a set of nodes (vertices) N which are connected by edges (links) E (Bollobás, 2013). The structure can be used to represent the relationships between different entities. For example, in a telephone network, people can be seen as graph nodes, while the action of ringing (a flux of information) can be cast as edges. Both nodes and edges can contain attributes, for example, edges may have weights to represent the strength of the relationship (frequency of calling), or the distance between two connected components. Edges may also have directions to indicates whether the relationship is unidirectional. A graph with N nodes can be mathematically represented as an adjacency matrix (figure 2.1 b) size of $N * N$, where $A_{ij} = 1$ if node N_i and N_j are connected by an edge , and $A_{ij} = 0$ if they are not.

(a)



(b)

	A	B	C	D	E
A	0	1	1	0	1
B	0	0	0	0	1
C	1	0	0	0	1
D	0	0	0	0	1
E	1	0	1	0	0

Figure 2.1 (a) An example of a graph structure; and (b) the adjacency matrix.

Once a graph is defined and constructed, a number of measures can thus be derived to describe the characteristics of different components within, or the overall status of the whole structure. In the context of urban and transport system and graphs, these different indicators may be used to measure system efficiency (Jiang, 2009), identify urban hubs and centres (Huang et al., 2015), understand socioeconomic patterns and activities (Strano et al., 2007), and evaluate the resiliency of the network structure to node failure (Wilkinson et al., 2012), etc. Several most commonly used measures are as follows.

2.1.1 Node degree

The node degree d is a centrality measure for describing an individual node N_i in a graph, and is defined as the sum of neighbours N_j that connected to N_i (Opsahl et al., 2010; Newman, M., 2008). In mathematical term, $d(i)$ of the node N_i is defined as:

$$d(i) = \sum_{j} m_{ij}$$

Where $m_{ij} = 1$ if there is an edge between node N_i and N_j, and $m_{ij} = 0$ if no such link exists (Otte and Rousseau, 2002). For example, In a graph representing co-authors, the degree centrality of an author is the number of other authors they have co-authored at least one article with (Otte and Rousseau, 2002). Degree can also have directions in directed graphs, calculated based on the number of edges that start from or end at N_i, so they are denoted as in-degree and out-degree, respectively (Newman, M., 2008). In the context of spatial networks, such as an airline graph, the in-degree of a city (airport) is the number of other cities (airports) that have direct flights to it.

2.1.2 Betweenness

In a connected graph, between every pair of nodes exists at least one shortest path. “Shortest” may be defined as the minimum number of edges that the path passes through, or as the sum of the weights of the edges are smallest. Hence, betweenness centrality is defined loosely as the frequency of a node N_i is needed/passed for a pair of other nodes to reach each other by the shortest path (Barthelemy, 2004).

Mathematically, the betweenness centrality of N_i, denoted as $b(i)$ is :

$$b(i) = \sum_{j,k} \frac{g_{jik}}{g_{jk}}$$

Where $g_{j,k}$ is the number of shortest paths from node N_j to N_k $(j, k \neq i)$, and g_{jik} is the number of shortest paths from node j to node k passing through node i (Otte and Rousseau, 2002; Anderson and Dragićević, 2020). In spatial graph structures such as road network, high betweenness implies a potentially high volume of traffic, and such area/road may have a greater risk of congestion.

2.1.3 Eigenvector centrality and PageRank

Eigenvector centrality is another indicator for measuring the level of influence/importance of a node within a graph. Similar to node degree, Eigenvector centrality is relative to the number of edges to other nodes. Unlike other measures, it weights connections to high-scoring centrality nodes contribute higher than low-scoring nodes (Bonacich, 2007). For example, a high degree node may be low in eigenvector centrality, if it is only linked with similarly low-scored nodes.

The PageRank is a variant of the Eigenvector centrality, and was first introduced by Google to evaluate the importance of a Web page (Brin et al., 1998). PageRank extends the idea in Eigenvector centrality that not counting links from all pages/nodes equally, and it majorly utilised/interested in the in-links. Therefore, PageRank is similar to voting, where a link from node N_i to N_j is like a N_i contributed to a vote of importance to N_j.

Assume node N_a has incoming edges from other nodes $N_1 \cdots N_n$, and the parameter d is a damping factor, $C(N_n)$ is defined as the out-degree of node N_n. The PageRank (PR) of node N_a is denoted as follows (Page et al., 1999; Brin and Page, 1998):

$$PR(N_a) = (1 - d) + d(\frac{PR(N_1)}{C(N_1)} + \cdots + \frac{PR(N_n)}{C(N_n)})$$

The $PR(N_a)$ can thus be calculated using an iterative algorithm that corresponds to the principal eigenvector of the normalised link matrix of the graph (Page et al., 1999; Brin and Page, 1998). Note that the PageRank forms a probability distribution over graph nodes, so the sum of all nodes' PageRank will be one. PageRank is an additional indicator of relative node importance and centrality in a graph. In a transportation network, this indicator can help to identify key nodes (places) in the system that have a high impact on transportation efficiency; urban hubs and centres also generally have high PageRank values in a travel flow graph (Huang et al., 2015).

2.1.4 Clustering coefficient

Clustering coefficient measures the degree to which nodes in a graph tend to cluster together (Schank and Wagner, 2005). In many real-world networks (graph structures), graph nodes are not randomly or evenly linked to each other. To some extent, they usually presented as tightly knit groups characterised by a relatively high density of links within the groups (Kim and Leskovec, 2012; Šíma and Schaeffer, 2006; Schaeffer, 2007). Hence the clustering coefficient is introduced to characterise and quantify this tendency.

The global clustering coefficient (i.e. transitivity) can be calculated based on triplets of nodes in a graph (Schank and Wagner, 2005). A triplet is three nodes that are connected by two or three edges, the former one is an open triplet, while the latter one is called a closed triplet. For example, in a triangle graph structure, which consists of three nodes and three edges, there are three closed triplets, because each of the three nodes can respectively be regarded as a triplet centre. The global clustering coefficient (Luce and Perry, 1949; Wasserman and Faust, 1994) is mathematically defined as:

$$C = \frac{N_{CT}}{N_{AT}}$$

Where N_{CT} is the number of closed triplets within the graph, and N_{AT} is the count of all triplets (open and close).

2.1.5 Communities and modularity

In a graph structure, community (i.e. modules or clusters) refers to the occurrence of groups of nodes that are more densely connected internally than with the rest (Fortunato, 2010). The heterogeneity of connections leads to some natural divisions within.

In spatial graph structures (e.g. transport system), these dense connections tend to occur between spatially proximate nodes (Saberi et al., 2018; Yang et al., 2019a). Therefore, finding the communities do have some practical applications In geography and planning, for example, design of efficient national, economic or administrative borders based on human mobility or economic interactions (Expert et al., 2011; Blondel et al., 2008; Wang et al., 2020; Liu, X. et al., 2015). The UK Travel to Work Area (TTWA) utilised a community detection approach to define a set of the TTWA for the whole of the UK, and it helps in labour market analysis and planning problems (Coombes and Bond, 2008).

Determining the right communities within a graph is relatively difficult because the number of communities and their (unequal) sizes are normally unknown. Several commonly used algorithms include Hierarchical clustering (Yin et al., 2015; Reichardt and Bornholdt, 2006), Girvan-Newman algorithm (Duch and Arenas, 2005; Despalatović et al., 2014), but the most popular ones are modularity maximisation methods (Fortunato and Hric, 2016).

Modularity is designed to measure the strength of division of a graph into communities, and Modularity Q (Newman, M.E. and Girvan, 2004) is defined as:

$$Q = \sum_{s=1}^{n_M} \frac{I_s}{E} - \left(\frac{d_s}{2E}\right)^2$$

Where n_M is the number of communities of the partition, I_s is the number of links inside community s, E refers to link counts in the graph, and d_s is the total degree of the nodes in s.

A graph with high modularity has dense connections between community members but sparse connections with nodes in different communities. Optimising this value theoretically results in the best possible segmentation of the nodes in a given graph structure. Different heuristics approaches that are used for solving this problem include greedy algorithms, simulated annealing, and spectral optimisation (Anderson and Dragićević, 2020). One of the most popular approaches is the Louvain method, and it first finds small communities by optimising modularity locally on all nodes, then each small community is grouped into one node, and the first step is repeated, in an iterative way. The division is determined/finalised when the global modularity can no longer be improved.

Although the Louvain method has been widely used in many spatial networks (Yang et al., 2019a; Saberi et al., 2018), the results sometimes are not successful in uncovering very small clusters (Anderson and Dragićević, 2020). In contrast to other approaches, such as greedy algorithms, tend to detect more clusters that are smaller in size.

2.1.6 Assortativity coefficient

The assortativity coefficient measures the preference of nodes in a graph to link to others that are similar (Noldus and Van Mieghem, 2015). Although the similarity can be determined in varying ways, normally, they are quantified using node degrees (Newman, M.E., 2003a; Newman, M.E., 2002; Anderson and Dragićević, 2020; Spiegel et al., 2017), and defined as the Pearson correlation coefficient of degree between all pairs of linked nodes in the graph. Mathematically, assortativity a is defined as:

$$a = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}$$

where M is the number of edges in the graph, j_i and k_i are the degree of the vertices at the two ends of the i-th edge, with $i = 1 \cdots M$ (Newman, M.E., 2002).

One example of high assortativity is the Rich-club effect (Xu et al., 2010), which exists in many real-world networks, such as scientific collaboration graph and transportation flow graphs (Wei, Y. et al., 2018). In these graph structures, well-connected nodes (rich-club members) are also found to connect to each other.

2.2 Graphs in urban and transport systems

Two different kinds of graphs are conceptualised to model urban and transport systems; these are planar and non-planar graphs (networks). Their structure and various indices have been extensively examined in various studies related to two main themes: (1) the topology of the urban/transport infrastructures (physical course of networks), and (2) the accessibility and centrality of different regions/locations (Anderson and Dragićević, 2020).

2.2.1 Planar graphs and physical course of transport networks

In graph theory, a planar graph is a graph that can be embedded in the plane, and no edge will cross each other (Barthelemy, 2018). Planar graphs were normally used to represent physical infrastructures, for example, roads (Cardillo et al., 2006; Strano et al., 2013; Qian et al., 2012), railways (Wei, S. et al., 2019), canals and rivers (Bogart, 2009) networks. In planar space, the graph links represent the physical course of road/rail/river, while the vertices are the intersections between them or the endpoints. Links intersect only at vertices (Barthelemy, 2011).

In urban and transport studies, measuring graph (network) properties of cities through examine street network topology is not new and can be traced back to the 1960s (Garrison, 1960), and this was popularised by Hillier and Hanson (1989) under the term of “space syntax”. For cities around the world, despite their varying geographical, climatic, historical, cultural and social-economical mechanisms that have shaped them in different ways, recent empirical studies have shown that, at least at a coarse-grained level, unexpected similarities exist in their graph measures by examining their road networks (Batty, 2018; Volchenkov and Blanchard, 2008; Abshirini and Koch, 2017). Generally, the structure of planar transportation networks is constrained by geographic space and proximity, thus limiting node degree values. As a result, such structures normally have a low average node degree. For example, in the street networks of many cities, an individual node’s degree rarely exceed five (Cardillo et al., 2006; Lämmer et al., 2006)

Lämmer et al. (2006) studied the German road networks and suggested that the distributions of betweenness centrality follow the power law. It indicates the strong heterogeneity within the graph in terms of potential traffic volume. The existence of several central nodes represents popular points (regions and road) and suggests some local congestion (Crucitti et al., 2006; Scellato et al., 2006). Other studies, for example, Strano et al. (2007) suggested that in transport systems, graph measures are related to socio-economical indices (Barthelemy, 2011; Jia and Jiang, 2012). In the context of urban areas, a clear correlation exists between the betweenness indicators and the presence of commercial activities (Strano et al., 2007).

In the past several decades, many works (Cardillo et al., 2006; Lämmer et al., 2006) have studied the transport and urban systems upon small and planar graphs, looking at the static topology of transport infrastructures (e.g. road network). This is partly due to the limited computational power and data availability (Ducruet and Lugo, 2013). The planar graphs are normally modelled using data extracted from city or country maps, while the rich and voluminous data detailing actual transport flows and movement of people remained hardly accessible in the past (Barthelemy, 2011).

2.2.2 Non-planar graph and flows

Batty (2018) suggested that there is increasing research interest in measuring physical traffic that uses physical infrastructures in more complete ways. Planar networks themselves only represent the fundamental infrastructures (e.g. physical roads), while how people use them and move around the space requires further understanding (Batty, 2018; Barthelemy, 2011). Individual flows reflect the location and movement of activities; thus, they are related to fundamental problems in geography and spatial economics. Moreover, the flows contain information on how diseases might transmit between population due to the close proximity and interactions while travelling (Sun et al., 2014). There are also applications in commercial domains, such as geomarketing or transport resource management. For example, the flows help to (1) identify where advertisement should be placed depending on how many people going through a location, (2) or how frequently should bus vehicle be dispatched, or sharing bikes fleet be rebalanced to meet the local travel demand. Clues on the statistics of human behaviours and flow interaction are thus important for understanding and managing the urban system (Barthelemy, 2011).

These physical travel flows, together with underlying capital and information flows, are the core of how different parts of the city may function and develop. Therefore, there is an increasing number of studies that examine the flows with non-planar graphs. In these graphs, a node normally represents a location or an entity (e.g. transport hub), while the link indicates the connection (e.g. trip origin-destination pair) between nodes, rather than physically representing a route. The links may have directions to indicate the flow direction and weights to represent the frequency or volume of the connection. Some examples of the non-planar graph, in the context of urban and transport systems, include bus and dock-based bike-sharing systems, where the stations are cast as graph nodes, and the travel flow between them are used for characterising graph edges. Other examples at a larger spatial scale (e.g. globally) include cargo and airline networks.

Non-planar transport graphs are also found to be constrained by geographic space, as well as the physical capacity of the node. For example, airport and cargo ships can only handle a limited number of connections because of space and travel time constraints (Amaral et al., 2000). However, non-planar graphs in the context of general urban commuting are found to be relatively less impacted by physical distance and space. The work of Chowell et al. (2003) analysed the simulated movement of 1.6 million individuals during a typical day in Portland (Oregon, USA). The results (Chowell et al., 2003) suggest that space is not constraining enough for general urban commuting. This is because the cost variations are too small, and the important features of spatial constraints (e.g. limited/few number of node degree) do not seem to appear evidently; hence graph measures (e.g. out-degree) are found to well-fitted by power laws (Montis et al., 2007; Chowell et al., 2003). However, Chowell et al. (2003) did not take the transport capacity into consideration. In other words, when certain urban transit systems are under disruption (Saberi et al., 2018), it might pose a severe challenge to the capacity and function of transport, thus constraining connections between graph nodes. While motorised traffic such as metro and bus are fast travel modes, which may be less impacted by long travel time/cost, other modes (e.g. cycling and walking) are less likely to have many long-distance links. Therefore, different travel modes may be differently impacted by physical space and distance.

While understanding the movement of people has evident benefits and wide applications (Sun et al., 2014; Barthelemy, 2011; Saberi et al., 2018), obtaining the movement (OD flows) of people is a very difficult and critical task. Currently, it largely relies on costly travel surveys. But the recent technological advances, for example, mobile phone and geosocial applications provide new opportunities and challenges for understanding individual movements, interactions and underlying non-planar graph properties.

2.2.3 Opportunities from new forms of data

Reliable flow data detailing OD pair or interaction between people are important inputs for planning urban and transportation initiatives. Various government sources, travel diary studies and roadside/household surveys have generated valuable and focused (bespoke) data to facilitate data-driven transport and urban studies. However, these data may also have disadvantages, for example, high expense, small sample size, lower update frequencies and potential sampling biases and reporting errors (Groves, 2006; Birkin, 2019).

But the world is increasingly digitalised nowadays, events and people's behaviours are recorded in a variety of new data unprecedentedly. Vehicles and bikes are equipped with GPS (Yang et al., 2019a), location-based service and mobile phones (Manley and Dennett, 2019) are popular among people, The widening availability of new dataset have the potential to represent the movement and interaction between people and places, also help to reveal underlying social and spatial characteristic and processes (Manley and Dennett, 2019; Barthelemy, 2011; Batty, 2018).

2.2.3.1 Mobile phone data

When a mobile phone is used for making calls, sending a text or using the Web, there will be footprints be left. Using methods like triangulating positioning algorithm, users location can be roughly calculated (Toole et al., 2012). The last decade has seen mobile phones, especially smartphones, increased their penetration rates (popularity) both in developed and developing countries (Manley and Dennett, 2019). Mobile phone data is able to characterise large-scale individual trajectories. Subsequently, the data have been used to identify “anchor points” where people stay and spend a lot of their time, and to reveal mobility patterns (Calabrese et al., 2011), route choice modelling, traffic model calibration, traffic flow estimation and obtaining OD (Origin and Destination) flows (Iqbal et al., 2014; Bachir et al., 2019).

Calabrese et al. (2011) used mobile phone data in Boston, USA, to quantify OD flows of people, the result is found to correlate well with US Census estimations at both the county and census tract level. Compared to traditional survey data, the flows generated from mobile phone data have some vital advantages, for example, much finer spatial and temporal granularity. It is, therefore, possible to look into flow and interactions in specific days or certain time periods (e.g. morning rush hour). Hence the data provides new opportunities to look deeper into individual mobility patterns.

The application of mobile phone data and related flows is not limited to transport and mobility studies. There is an increasing amount of research that has utilised mobile phone data for various purposes in different domains, for example, health (Vinceti et al., 2020), commerce (Birkin, 2019), land-use and urban planning (Manley and Dennett, 2019).

The work of Manley and Dennett (2019) combined mobile phone data and building use data to analyse activity and interaction in Dakar, Senegal, and elucidated their variation across land-use classification and time. Mobile phone data was also used to measure mobility restrictions and changes during Covid-19 health outbreaks. The reduced mobility, as identified from mobile phone data, are found to be inversely related to the daily number of newly diagnosed Covid-19 positive cases (Vinceti et al., 2020). An effective reduction in transmission occurs almost immediately after the tighter lockdown in several Italian regions, given the lag time of around ten days from asymptomatic infection to diagnosis (Vinceti et al., 2020). Therefore, measuring mobility behaviours through mobile phone data can be used for related emergence response and public health.

Mobile phone data not only helps to locate an individual's location and identify the movement, but the communications between them can also be used for constructing a spatial and social graph structure. For example, community detection was performed on Belgian mobile phone network to identify language communities (Blondel et al., 2008). Thus graph-based analysis played an interesting tool for defining administrative boundaries, and useful for policymaking and planning.

Despite the above applications, a shortcoming of mobile phone data is its relatively low positioning accuracy. For example, the accuracy for 2G, 3G, 4G networks are reported as 500 - 600 metres, 200m and 150-200m, respectively (Liu, Y. et al., 2017). Therefore, the data is much coarser than locations that are obtained by GPS (Global Positioning System), whose accuracy is normally within several meters (Owari et al., 2009). The nature of the coarse positioning accuracy in mobile phone data makes it more suitable for analysing the movement of long-distance such as commuting. If using mobile phone data to represent short-distance (e.g. within 400m) journey, it will bring greater uncertainty to the results. But the emerging 5G network may contribute to higher accuracy of cellular positioning, for example, within 1 meter (Liu, Y. et al., 2017). Therefore, the utility of mobile phone data will increase with the development of new communication technology in the future.

2.2.3.2 Social media data

With the increasingly popular social media and location-based services, more people are willing to share their location with the service company (e.g. Twitter, foursquare) and their friends in their post. While mobile phone data provide relatively coarse location accuracy, location-based social media provide location-specific data, as people check-in to different places, or at a finer spatial granularity of GPS coordinates. This means that the location and movement of users could be semantically enriched, by analysing the function and land-use related to the place.

A number of studies have utilised social network data to understand the travel behaviours, supplement semantic information (Sari Aslam et al., 2020) or trip purpose (Yang et al., 2019b) for travel OD flows. Beyond these utilities, social network data also contains valuable information detailing the connections and interactions between individuals. The underlying graph structure of social relations along with spatial information has been analysed in different studies

(Alizadeh et al., 2017; Backstrom et al., 2010; Scellato et al., 2010). For example, the work of Cho et al. (2011) suggests that people's short-ranged travel is less impacted by the social network structure, while if a person travels a long distance, then he/she is more likely to be spatially proximate to an existing friend in the graph. Therefore, the mobility and location of social media friends could be helpful to predict an individual's location (Cho et al., 2011; Backstrom et al., 2010).

2.2.3.3 Transaction data in mobility services

Transaction data that arise from interactions between customers and mobility service providers can also be used to obtain travel flows of people. Various public transport systems (e.g. train, bus) and increasingly popular shared-mobility services (e.g. Uber, Mobike) contain this valuable information, which helps to construct the OD matrix (Tang et al., 2020) and understand the travel behaviour (Manley et al., 2018) and flow interactions (Birkin, 2019; Batty, 2013).

Tap-in/out into stations, or boarding/alighting from vehicles are recorded by Automatic Fare Collection (AFC) systems in many public transport modes. By analysing these intra-urban flows and behaviours (Fuse et al., 2010), it is possible to shed light on local characteristics in different urban regions (Pelletier et al., 2011), such as local vitality (Sulis et al., 2018) and urban structure and hotspots (Roth et al., 2011; Wang et al., 2020). Taxi data is used in the work of Huang et al. (2015) to construct graphs, and then quantify the different centralities of urban regions, with high centrality (e.g. PageRank) areas potentially indicating hubs or centres in the city. The result obtained from the graph-based analysis is then validated by ground-truthing, a group of local drivers and experts have participated in confirming the findings. Overall the study (Huang et al., 2015) suggest that PageRank of taxi travel flow graph is found to be the most successful indicator of urban hubs, with a high agreement rate among all participates.

The recent years have witnessed the huge success and popularity of shared-economy travel mode (McKenzie, 2020). Companies that provide ride-hailing services (e.g. Uber) and micro-mobility (e.g. bike-sharing, scooter-sharing) have also generated large quantities of movement data at a fine spatiotemporal granularity. In particular, micro-mobility services are normally used for relatively short distance journeys and help solve the "first/last mile" problem. The low cost, as well as high convenience, made them very popular in recent years (McKenzie, 2020; Lovelace et al., 2020). This thesis puts a focus on different bike-sharing data and cycling flows. Therefore, a dedicated review on bike-sharing studies and cycling flows are presented in section 2.3.

Despite the new opportunities emerging with the increased availability of the above novel datasets, they all are the by-product of the systems such as communication services, social media and bike-sharing transactions. Therefore, these datasets are not designed specifically for research purposes. Disadvantages in sampling bias should be considered in conducting research and drawing conclusions for mobility patterns and urban phenomena.

Moreover, the effective exploitation of the above new opportunities depends on the collaboration between commercial to academic sectors (Birkin, 2019). All of the above datasets may contain user behaviours and identities at the individual level, hence related research ethics (e.g. data privacy) should also be taken care of.

2.2.4 Flows and graph structure evolution

Studying spatial graph structures have a long tradition; however, applications that incorporate temporal features are more recent, due to the increased data availability. Characterising real-world networks using graph information features is useful because it helps to identify, understand, and anticipate the processes that take place (Anderson and Dragićević, 2020). Therefore, increased popularity and interests have been observed in understanding the evolution of graphs and the underlying process. For example, the graph structures in airline networks vary over space and time due to the geographic, economic, political and historical factors, and thus are constantly changing as links between airports appear, disappear, expand or contract (Anderson and Dragićević, 2020). When the global airline graph (network) grows, it will have lower assortativity (Barrat et al., 2005), and nodes with a high degree tend to connect preferentially with nodes with a low degree.

Graph measures may look at the immediate environment of nodes by examining the changed adjacent neighbours (Ducruet and Lugo, 2013). Historically, the evolution of transport systems is examined from the perspective of physical infrastructures. In planar graphs, for example, a road network (Mohajeri and Gudmundsson, 2014), may evolve along with the development and expansion of the city. Some nodes in certain parts of the city may increase their degree and reach more places through new road segments of new directions. The node betweenness may also change, implying higher traffic volumes and potentially more congestions. But motivated by the recent availability of large-scale data on human activities and flows, the centrality and vitality of a location may be measured as to how people are attracted from other regions, and how diverse their activities are (Zhong et al., 2014; Zhong et al., 2017; Manley and Dennett, 2019). In non-planar transport flow graphs, questions might concern whether larger nodes are more central in structure due to urban development, or whether important nodes tend to connect to each other over time to form the so-called “rich-club” effect.

Centrality measures derived from graph analysis can be used for understanding the changed function and emergence of new urban centres. Zhong et al. (2017) used intra-urban travel flow data to present Singapore’s rapid development towards a polycentric urban form. They concluded that the downtown core has strongly gained importance (centrality), which is largely attributed to the extension of the public transit system and increased accessibility. Similar research is conducted in the work of Sun et al. (2015) and presented the temporal variation in graph communities, which becomes more connected and considered as the result of improved urban mobility. The above studies (Zhong et al., 2017; Sun et al., 2015) all look into the flow interactions and graph evolutions caused by the long-term urban growth and development. The comparisons, therefore, cover a relatively long period, for example, several years.

With the availability of trip data with higher spatial granularity, recent studies have begun to evaluate the evolution of graphs in a shorter temporal course; for example, the immediate changes caused by short-term indecent such as transport disruptions and adverse weather (Saberi et al., 2018).

There are also a number of studies that have tried to present the daily variance of graph structure during the course of the day (Huang et al., 2015). The work of Rahman et al. (2020) compared the bus travel flows and underlying graph structures at different times; significant differences were found during peak and off-peak hours, and weekdays against weekends. The variance indicates the travel pattern changes according to the temporal variable, and related information has been preserved in graphs. The temporal graph provides a different perspective for understanding travel behaviours, and have the potential to be used for better modelling and prediction.

To summarise, in recent years, many flow data are generated in real-time at a fine spatiotemporal resolution (Sun et al., 2015; Batty, 2018). Despite some privacy concerns, taking full advantages of such data in transport and urban planning would help researchers and stakeholders to better interpret and model urban dynamics, as well as contribute to smarter urban development.

2.3 Bike-sharing and cycling flows

Bike-sharing fills an important gap in the intra-urban transportation system between pedestrian and motorised vehicles. The low-cost and high convenience feature make it increasingly popular worldwide (Fishman, 2016; Lovelace et al., 2020). The schemes can bring various benefits to cities and their inhabitants – including better air quality, reduce congestion, creating livable streets, and health/well-being improvements for people (Shaheen et al., 2013). The recent technology development (Gu et al., 2019a) in bike-sharing contributed to a large quantity of flow data detailing people's movement at an increasingly fine spatial and temporal granularity. Hence, it provides new bedrocks for understanding the flow and graph structures in the urban "last mile" - the distance between home/workplace and bus/metro station (Shaheen et al., 2010).

2.3.1 Generations of bike-sharing

The history of bike-sharing can be traced back to 1965 (table 2.1), and the first generation provides bicycles to users without costs. Bikes are unlocked, normally placed haphazardly throughout an area for use. Despite its convenience, there is a fatal weakness of this scheme – the maintenance. Most first-generation bike-sharing schemes, such as Witte Fietsen (White Bicycle) (Amsterdam, 1965), La Rochelle (France, 1974), Green Bike Scheme (Cambridge, 1993), failed due to bicycle damages and thefts (Shaheen, S. et al., 2010).

The second-generation emerged in 1991 in the form of coin-deposit schemes (Shaheen et al., 2010). Bicycles are not freely available, and users have to use small deposits to unlock the bike from docking stations. Thus it is theft-resistant but with higher operating cost (especially human resources) than the previous generation.

The third-generation (table 2.1) of bike-sharing is similar to the second-generation, but incorporated more advanced technologies for bicycle management. Smart cards, mobile phones or codes may be used for bicycle check-in and check-out at docking stations. These schemes are relatively simpler to manage in terms of human resources, but requires a higher investment in IT (Information Technology). A great advantage of the IT-based scheme is the availability of 24-hour services. Besides, the system automatically records the transactions with information detailing borrowing and returning time, as well as related stations. This benefits research of cycling OD flows and travel behaviours (Lovelace et al., 2020; Beecham, 2015). Some example of the third-generation bike-sharing includes Santander Cycles in London and Citi Bike in New York.

Table 2.1 Generations of bike-sharing schemes.

Generation	Birth time	Scheme	Station/dock-based
First	1965	Free usage	No
Second	1991	Coin-deposit system	Yes
Third	1996	IT-based system	Yes
Fourth	2015	Dockless system	No

The fourth and latest generation, dockless schemes, emerged around 2015 and have spread to large numbers of cities, especially in China (Sun, 2018). Although services are provided by different companies (e.g. Mobike, OFO) and differ slightly, they have the same core feature: bikes can be borrowed and returned “haphazardly” (Chen et al., 2020), and no longer relies on docking stations. Cycles are equipped with Internet-enabled GPS units which can provide their real-time location to users and operators. In addition, every cycle has a built-in smart lock system. To start a trip, registered users can search for nearby vacant cycles through a smartphone app. Once located, the user then unlocks it by scanning the QR code (Quick Response Code) printed on the bike. When ending the ride, the cycle can be parked in any appropriate place, and it becomes available for the next user to locate and use.

In summary, dockless bike-sharing schemes are more flexible to use while also anti-theft (GPS traced). With competitive pricing, the scheme has become a huge success in Asia, especially in some Chinese cities. For example, the number of dockless sharing bikes in Beijing and Shenzhen reached 2.2 million and 0.89 million by 2018, respectively (Gu et al., 2019a). The vast amount, increased spatial flexibility, and real-time systems make travel flow data obtained from dockless bike-sharing an ideal proxy for measuring the activities and structures in urban last mile. However, despite its popularity in Asian countries, the expansion to European countries is not smooth at the first try. In 2017, Mobike, one of the industry leader of dockless bikes, chose Manchester as their 100th city, and the first outside Asia to host their dockless service. But the company ended the scheme in the following year (Sherriff et al., 2020). The failure is considered not due to the technological solution (dockless), but the lack of consideration in the context of existing infrastructure, local cycling environment and social exclusion (Sherriff et al., 2020).

2.3.2 Users' changed behaviours in bike-sharing schemes

Researchers have intensively examined various correlates with bike-sharing usage and travel behaviours in cities. Similar conclusions are drawn and suggested that bike-sharing trips can be influenced by user characteristics, scheme accessibility, weather and built environment (El-Assi et al., 2017; Gebhart and Noland, 2014; Gu et al., 2019 b, Lovelace et al., 2020; Younes et al., 2019). Most of these studies put a focus on confirming correlation, but causality and behaviour changes are less established (Wang & Lindsey, 2019).

One way to understand the changed behaviours in bike-sharing is to investigate travel pattern variations in response to different incidents. Gu et al. (2019b) examined the impact of a new metro line on an existing dock-based bike-sharing system, and concluded that the number of bike-sharing users and trips has increased. However, this work (Gu et al., 2019b) only examined patterns of trip numbers, while the underlying structural changes are not revealed. Gu et al. (2019b) also suggested that the potential impacts of new metro service on GPS-enabled dockless bike-sharing require further study. Bike-sharing is also reported to promote metro travels in return; Ashraf et al. (2021) indicated that in New York, a 10% increase in the number of bike-sharing trips could increase the average daily subway ridership by 2.3%. The presence of dedicated bike lanes and bike racks attracted more bike usage and increased subway ridership. Wang & Lindsey (2019) revealed in the circumstance of adding or changing the location of bike docking stations, the improved accessibility to bike-sharing service leads to a significantly higher frequency of usage among registered scheme members. Younes et al. (2019) analysed the impact of several metro stations' closure for maintenance on bike-sharing demand. They found that the temporary closing of stations can lead to between 24% and 45% more bike-sharing trips nearby the disrupted station. A transit disruption at a larger spatial scale, the Tube strike in London, is examined by Saberi et al. (2018). In addition to increased bike ridership, the graph-based analysis suggested that disruptions caused greater connectivity and more graph edges, node centrality measures still fit a power-law distribution but with greater scores (Saberi et al., 2018). It should be noted that the newly generated OD pairs are probably caused by the reduced availability of cycles and docks at each station. During mass transit disruption, the increased cycling demand requires extra maintenance to support the scheme's operation by providing sufficient resources (cycles and docks) at different stations. Otherwise, users may be forced to use nearby alternative stations to start and end their journey, thus bringing more diverse flows to the graph. However, more evidence is required to look into the patterns of resource apply and their effect on flow structures. Dockless bikes may relatively suffer less from the capacity problems in dock-based schemes because bikes may be located more haphazardly, rather than only at fixed docking stations.

The changed behaviours brought by various events may also vary depending on people's different socioeconomic characteristics and attitudes on cycling. For example, people can have diverse feelings on the scheme pricing, different level of knowledge (whether know how to ride bikes and use bike-sharing schemes), physical capability and self-efficacy (e.g. whether feel confident about cycling). These characteristics are likely to impact people's decisions and travel behaviours as well. Zhu et al. (2017) suggested that people with lower incomes are more likely to choose lower-cost alternative mobility services such as bike-sharing, rather than ride-hailing (e.g. Uber, Lyft) and taxi when large transit disruption happens.

Overall, travel behaviours in bike-sharing and underlying flow structure can be impacted by various events like adverse weather, transit disruptions and urban development, also influenced by user's socioeconomic characteristics.

2.3.3 Cycling flows and graphs

The trips in third-generation bike-sharing travel start and end at docking stations. Therefore, the cycling flow can readily be converted to a directed and weighted graph, using docking stations as graph nodes, and the trips represent the graph edge with the frequency variable.

The work of Austwick et al. (2013) analysed the flows and graph structures in 5 cities in North America and Europe, and suggest that high similarities exist in the graphs properties, such as link weight distribution. But this work focused and compared the system at a highly aggregated level, for example, generating graphs using travel record of several months. It may be desirable to further disaggregate the temporal variable to create time-dependent (e.g. hourly) graphs (Tang, J. et al., 2010; Casteigts et al., 2012). Measuring the changing graph metrics (Nicosia et al., 2013) at a fine temporal granularity helps to reflect varied use and behaviours in the scheme at different times, such as commuting in the morning rush hour and leisure-related trip during the weekend afternoon.

The recent dockless bike-sharing schemes and their nature of flexibility also bring challenges for graph-based analysis, in particular, the conceptualising of graph vertices. No docking stations exist, therefore the graph vertices should be cast based on other entities, for example, street segments, grid cells, transportation zones or regions that are divided by various algorithms (e.g. DBSCAN clustering). Grid cells are used in some studies (Yang et al., 2017) to aggregate trip counts of dockless sharing bikes. But it should be noted that the short distance of trips, which is a nature of cycling activity in the urban area, requires grid cells at a very fine spatial granularity to capture flow OD. If too coarse, it might lose information in the subtitle (short-distance) trips, because the origin and destination might fall in the same grid. However, if setting too small, there will be a lot of grid cells have very few or no trips were made, this results in a very sparse matrix, and unnecessarily increase the computation burden. How to conceptualise and construct graph structures for urban dockless bike-sharing schemes require further study.

Overall, it is increasingly interesting and challenging to examine the flows in bike-sharing schemes with finer spatiotemporal granularities, and to model and understand the process undertaken in urban space. The nature and characteristics of cycling and bike-sharing schemes, such as their functions in complementing the “last mile”, low-cost and replacing relatively short distance journey of other transport (e.g. bus, taxi), makes them unique while important in contributing to inclusive and sustainable urban development.

2.4 Chapter summary

This chapter provides a critical literature review of research fields relevant to graph theory and its application in urban and transport systems. The increased availability of novel movement data, enabled travel flows to be examined at individual level with richer details. Although many current studies put a focus on describing the static structure and function of urban spaces, transport hubs or other entities, the great benefits and utility of temporal graphs should not be ignored. Temporal graphs can reveal how different part of the system interact and evolved by interpreting local and global changes in various graph-based measures. More importantly, it helps to shed light on how these changes are associated with or caused by different kinds of interventions such as infrastructure changes and disruptions. How to take full advantage of temporal graphs to contribute to better interpretation and modelling of urban and transport systems remained to be explored.

The evolving bike-sharing schemes have incorporated more information and communication technology, which enable the graph-based analysis to look at the flow interaction at finer spatial and temporal resolution. Chapter 3 and 4 sought to answer the question of how flow data from different generations of bike-sharing can support a better understanding of the individual's behaviours and interactions within the urban environment. The knowledge and metrics obtained from temporal graphs are further explored in Chapter 5. Case studies are conducted (Chapter 5) to demonstrate their utilities for real-time management of the transport system and contributing to smart and sustainable cities.

References

Abshirini, E. and Koch, D. 2017. Resilience, space syntax and spatial interfaces: The case of river cities. *Journal of the Faculty of Architecture.* **14**(1), pp.25-41.

Alizadeh, M., Cioffi-Revilla, C. and Crooks, A. 2017. Generating and analyzing spatial social networks. *Computational and Mathematical Organization Theory.* **23**(3), pp.362-390.

Amaral, L.A.N., Scala, A., Barthelemy, M. and Stanley, H.E. 2000. Classes of small-world networks. *Proceedings of the National Academy of Sciences.* **97**(21), pp.11149-11152.

Anderson, T. and Dragićević, S. 2020. Complex spatial networks: Theory and geospatial applications. *Geography Compass.* **14**(9), pe12502.

Ashraf MT, Hossen MA, Dey K, El-Dabaja S, Aljeri M, Naik B. 2021. Impacts of Bike Sharing Program on Subway Ridership in New York City. *Transportation Research Record*. Advanced online publication. https://doi:10.1177/03611981211004980.

Austwick, M.Z., O'Brien, O., Strano, E. and Viana, M. 2013. The structure of spatial networks and communities in bicycle sharing systems. *PloS One.* **8**(9), pe74685.

Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M. and Puchinger, J. 2019. Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies.* **101**, pp.254-275.

Backstrom, L., Sun, E. and Marlow, C. 2010. Find me if you can: improving geographical prediction with social and spatial proximity. In: *Proceedings of the 19th International Conference on World Wide Web*, 26-30 April 2010, Raleigh. New York: Association for Computing Machinery, pp.61-70.

Barabási, A. L. 2016. *Network science.* Cambridge: Cambridge university press.

Barrat, A., Barthelemy, M. and Vespignani, A. 2005. The effects of spatial constraints on the evolution of weighted complex networks. *Journal of Statistical Mechanics: Theory and Experiment.* **2005**(05), pP05003.

Barthelemy, M. 2004. Betweenness centrality in large complex networks. *The European Physical Journal B.* **38**(2), pp.163-168.

Barthelemy, M. 2011. Spatial networks. *Physics Reports.* **499**(1-3), pp.1-101.

Barthelemy, M. 2018. Morphogenesis of spatial networks. Cham, Switzerland: Springer international publishing.

Batty, M. 2005. Network geography: Relations, interactions, scaling and spatial processes in GIS. In: *Re-presenting GIS.* Hoboken: Wiley press pp.149-170.

Batty, M. 2013. *The new science of cities.* Cambridge: MIT press.

Batty, M. 2018. *Inventing future cities.* Cambridge: MIT press.

Beecham, R. 2015. Using Bikeshare Datasets to Improve Urban Cycling Experience and Research Urban Cycling Behaviour. In: Gerike, R. and Parkin, J. eds. *Cycling Futures: From Research into Practice.* Farnham: Ashgate, pp. 267-283.

Birkin, M. 2019. Spatial data analytics of mobility with consumer data. *Journal of Transport Geography.* **76**, pp.245-253.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment.* **2008**(10), p10008.

Bogart, D. 2009. Inter-modal network externalities and transport development: evidence from roads, canals, and ports during the English industrial revolution. *Networks and Spatial Economics.* **9**(3), pp.309-338.

Bollobás, B. 2013. *Modern graph theory.* New York: Springer.

Bonacich, P. 2007. Some unique properties of eigenvector centrality. *Social Networks.* **29**(4), pp.555-564.

Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. In: *Proceedings of the 7th International World Wide Web Conference*, 14-18 April 1998, Brisbane. New York: Association for Computing Machinery, pp.107-117.

Calabrese, F., Di Lorenzo, G., Liu, L. and Ratti, C. 2011. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. *IEEE Pervasive Computing.* **10**(4) pp.36-44.

Cardillo, A., Scellato, S., Latora, V. and Porta, S. 2006. Structural properties of planar graphs of urban street patterns. *Physical Review E.* **73**(6), P066107.

Casteigts, A., Flocchini, P., Quattrociocchi, W. and Santoro, N. 2012. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems.* **27**(5), pp.387-408.

Chen, Z., van Lierop, D., & Ettema, D. 2020. Dockless bike-sharing systems: what are the implications? *Transport Reviews*, **40**(3), pp.333-353.

Cho, E., Myers, S.A. and Leskovec, J. 2011. Friendship and mobility: user movement in location-based social networks. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 21-24 August 2011, San Diego. New York: Association for Computing Machinery, pp.1082-1090.

Chowell, G., Hyman, J.M., Eubank, S. and Castillo-Chavez, C. 2003. Scaling laws for the movement of people between locations in a large city. *Physical Review E.* **68**(6), p066102.

Coombes, M. and Bond, S. 2008. *Travel-to-Work Areas: the 2007 review.* London: Office for National Statistics.

Crooks, A.T. and Heppenstall, A.J. 2012. Introduction to agent-based modelling. In: Heppenstall, A., Crooks, A., See, L.M.& Batty, M. eds. *Agent-based models of geographical systems.* New York: Springer, pp.85-105.

Crucitti, P., Latora, V. and Porta, S. 2006. Centrality measures in spatial networks of urban streets. *Physical Review E.* **73**(3), p036125.

Despalatović, L., Vojković, T. and Vukicevic, D. 2014. Community structure in networks: Girvan-Newman algorithm improvement. In: *2014 37th international convention on information and communication technology, electronics and microelectronics*, 26-30 May 2014, Opatija. New York: IEEE, pp.997-1002.

Duch, J. and Arenas, A. 2005. Community detection in complex networks using extremal optimization. *Physical Review E.* **72**(2), p027104.

Ducruet, C. and Lugo, I. 2013. Structure and dynamics of transportation networks: Models, methods, and applications. In: Rodrigue, J. P., Notteboom, T., & Shaw, J. eds. *The Sage handbook of transport studies.* Thousand Oaks: Sage publications, pp.20-48

El-Assi, W., Mahmoud, M.S. and Habib, K.N. 2017. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation.* **44**(3), pp.589-613.

Euler, L. 1741. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae.* **8**, pp.128–140.

Expert, P., Evans, T.S., Blondel, V.D. and Lambiotte, R. 2011. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences.* **108**(19), pp.7663-7668.

Fishman, E. 2016. Bikeshare: A review of recent literature. *Transport Reviews.* **36**(1), pp.92-113.

Fortunato, S. 2010. Community detection in graphs. *Physics reports.* **486**(3-5), pp.75-174.

Fortunato, S. and Hric, D. 2016. Community detection in networks: A user guide. *Physics Reports.* **659**, pp.1-44.

Fuse, T., Makimura, K. and Nakamura, T. 2010. Observation of travel behavior by IC card data and application to transportation planning. In: *Special Joint Symposium of ISPRS Commission IV and AutoCarto* 2010. 15–19 November 2010, Orlando. Hoboken: Wiley, Commission IV.

Garrison, W.L. 1960. Connectivity of the interstate highway system. *Papers in Regional Science.* **6**(1), pp.121-137.

Gebhart, K. and Noland, R.B. 2014. The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation.* **41**(6), pp.1205-1225.

Groves, R.M. 2006. Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly.* **70**(5), pp.646-675.

Gu, T., Kim, I. and Currie, G. 2019a. To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. *Transportation Research Part A: Policy and Practice.* **119**, pp.122-147.

Gu, T., Kim, I., & Currie, G. 2019b. Measuring immediate impacts of a new mass transit system on an existing bike-share system in China. *Transportation research part A: policy and practice*. **124**, pp.20-39.

Heckmann, T., Schwanghart, W. and Phillips, J.D. 2015. Graph theory—Recent developments of its application in geomorphology. *Geomorphology.* **243**, pp.130-146.

Heppenstall, A., Malleson, N. and Crooks, A. 2016. “Space, the final frontier”: How good are agent-based models at simulating individuals and space in cities? *Systems.* **4**(1), p9.

Hillier, B. and Hanson, J. 1989. *The social logic of space.* Cambridge: Cambridge university press.

Huang, X., Zhao, Y., Ma, C., Yang, J., Ye, X. and Zhang, C. 2015. TrajGraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. *IEEE Transactions on Visualization and Computer Graphics.* **22**(1), pp.160-169.

Iqbal, M.S., Choudhury, C.F., Wang, P. and González, M.C. 2014. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies.* **40**, pp.63-74.

Jia, T. and Jiang, B. 2012. Building and analyzing the US airport network based on en-route location information. *Physica A: Statistical Mechanics and its Applications.* **391**(15), pp.4031-4042.

Jiang, B. 2009. Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science.* **23**(8), pp.1033-1048.

Kim, M. and Leskovec, J. 2012. Multiplicative attribute graph model of real-world networks. *Internet mathematics.* **8**(1-2), pp.113-160.

Lämmer, S., Gehlsen, B. and Helbing, D. 2006. Scaling laws in the spatial structure of urban road networks. *Physica A: Statistical Mechanics and its Applications.* **363**(1), pp.89-95.

Liu, X., Gong, L., Gong, Y. and Liu, Y. 2015. Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography.* **43**, pp.78-90.

Liu, Y., Shi, X., He, S. and Shi, Z. 2017. Prospective positioning architecture and technologies in 5G networks. *IEEE Network.* **31**(6), pp.115-121.

Lovelace, R., Beecham, R., Heinen, E., Vidal Tortosa, E., Yang, Y., Slade, C. and Roberts, A. 2020. Is the London Cycle Hire Scheme becoming more inclusive? An evaluation of the shifting spatial distribution of uptake based on 70 million trips. *Transportation Research Part A: Policy and Practice.* **140**, pp.1-15.

Luce, R.D. and Perry, A.D. 1949. A method of matrix analysis of group structure. *Psychometrika.* **14**(2), pp.95-116.

Ma, X., Ji, Y., Yang, M., Jin, Y. and Tan, X. 2018. Understanding bikeshare mode as a feeder to metro by isolating metro-bikeshare transfers from smart card data. *Transport Policy.* **71**, pp.57-69.

Manley, E. and Dennett, A. 2019. New Forms of Data for Understanding Urban Activity in Developing Countries. *Applied Spatial Analysis and Policy.* **12**(1), pp.45-70.

Manley, E., Zhong, C. and Batty, M. 2018. Spatiotemporal variation in travel regularity through transit user profiling. *Transportation.* **45**(3), pp.703-732.

McKenzie, G. 2020. Urban mobility in the sharing economy: A spatiotemporal comparison of shared mobility services. *Computers, Environment and Urban Systems.* **79**, p101418.

Mohajeri, N. and Gudmundsson, A. 2014. The evolution and complexity of urban street networks. *Geographical Analysis.* **46**(4), pp.345-367.

Montis, A., Barthlemy, M., Chessa, A. and Vespignani, A. 2007. Structure of inter-cities traffic: a weighted network analysis. *Environment and Planning B: Urban Analytics and City Science.* **34**(5), pp.905-924.

National Research Council. 1997. *Rediscovering geography: New relevance for science and society.* Washington, D.C.: National Academies Press.

Newman, M. 2008. The physics of networks. *Physics Today.* **61**(11), pp.33-38.

Newman, M.E. 2002. Assortative mixing in networks. *Physical Review Letters.* **89**(20), p208701.

Newman, M.E. 2003a. Mixing patterns in networks. *Physical Review: E.* **67**(2), p026126.

Newman, M.E. 2003b. The structure and function of complex networks. *SIAM Review.* **45**(2), pp.167-256.

Newman, M.E. and Girvan, M. 2004. Finding and evaluating community structure in networks. *Physical Review: E.* **69**(2), p026113.

Nicosia, V., Tang, J., Mascolo, C., Musolesi, M., Russo, G. and Latora, V. 2013. Graph metrics for temporal networks. In: Holme, P. and Saramak, j. eds. *Temporal networks.* New York: Springer, pp.15-40.

Noldus, R. and Van Mieghem, P. 2015. Assortativity in complex networks. *Journal of Complex Networks.* **3**(4), pp.507-542.

Opsahl, T., Agneessens, F. and Skvoretz, J. 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks.* **32**(3), pp.245-251.

Otte, E. and Rousseau, R. 2002. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science.* **28**(6), pp.441-453.

Owari, T., Kasahara, H., Oikawa, N. and Fukuoka, S. 2009. Seasonal variation of global positioning system (GPS) accuracy within the Tokyo University Forest in Hokkaido. *The Bulletin of Tokyo Medical and Dental University.* **120**, pp.19-28.

Page, L., Brin, S., Motwani, R. and Winograd, T. 1999. *The PageRank citation ranking: Bringing order to the web.* Stanford: Stanford InfoLab.

Pelletier, M.-P., Trépanier, M. and Morency, C. 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies.* **19**(4), pp.557-568.

Phillips, J.D., Schwanghart, W. and Heckmann, T. 2015. Graph theory in the geosciences. *Earth-Science Reviews.* **143**, pp.147-160.

Qian, Y.S., Wang, M., Kang, H.X., Zeng, J.W. and Liu, Y.F. 2012. Study on the road network connectivity reliability of valley city based on complex network. *Mathematical Problems in Engineering.* **2012**, p430785.

Rahman, A.S., Magalingam, P., Kamaruddin, N.B., Samy, G.N., Maarop, N. and Perumal, S. 2020. Graph Analysis Study of a City Bus Transit Network. *Journal of Physics.* **1551**, p012004.

Reichardt, J. and Bornholdt, S. 2006. Statistical mechanics of community detection. *Physical Review E.* **74**(1), p016110.

Roth, C., Kang, S.M., Batty, M. and Barthelemy, M. 2011. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PloS One.* **6**(1), pe15923.

Saberi, M., Ghamami, M., Gu, Y., Shojaei, M.H.S. and Fishman, E. 2018. Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. *Journal of transport geography.* **66**, pp.154-166.

Sari Aslam, N., Zhu, D., Cheng, T., Ibrahim, M.R. and Zhang, Y. 2020. Semantic enrichment of secondary activities using smart card data and point of interests: a case study in London. *Annals of GIS.* pp.1-13.

Scellato, S., Cardillo, A., Latora, V. and Porta, S. 2006. The backbone of a city. *The European Physical Journal B-Condensed Matter and Complex Systems.* **50**(1-2), pp.221-225.

Scellato, S., Mascolo, C., Musolesi, M., & Latora, V. (2010). *Distance matters: geo-social metrics for online social networks*. In: Proceedings of the 3rd Wonference on Online social networks, 22 June 2010, Boston. Berkeley: USENIX Association, pp.8-17.

Schaeffer, S.E. 2007. Graph clustering. *Computer Science Review.* **1**(1), pp.27-64.

Schank, T. and Wagner, D. 2005. Approximating clustering coefficient and transitivity. *Journal of Graph Algorithms and Applications.* **9**(2), pp.265-275.

Schimohr, K., & Scheiner, J. 2021. Spatial and temporal analysis of bike-sharing use in Cologne taking into account a public transit disruption. *Journal of Transport Geography*. **92**, p103017.

Shaheen, S., Martin, E. and Cohen, A. 2013. Public bikesharing and modal shift behavior: a comparative study of early bikesharing systems in North America. *International Journal of Transport*. **1**(1), pp.35-54.

Shaheen, S.A., Guzman, S. and Zhang, H. 2010. Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record.* **2143**(1), pp.159-167.

Sherriff, G., Adams, M., Blazejewski, L., Davies, N. and Kamerāde, D. 2020. From Mobike to no bike in Greater Manchester: Using the capabilities approach to explore Europe's first wave of dockless bike share. *Journal of Transport Geography.* **86**, p102744.

Šíma, J. and Schaeffer, S.E. 2006. On the NP-completeness of some graph cluster measures. In: *International Conference on Current Trends in Theory and Practice of Computer Science*: Springer, pp.530-537.

Spiegel, O., Leu, S.T., Bull, C.M. and Sih, A. 2017. What's your move? Movement as a link between personality and spatial dynamics in animal populations. *Ecology Letters.* **20**(1), pp.3-18.

Strano, E., Cardillo, A., Iacoviello, V., Latora, V., Messora, R., Porta, S. and Scellato, S. 2007. Street centrality vs. commerce and service locations in cities: a kernel density correlation case study in Bologna, Italy. New York: *arXiv preprint, p0701111.*

Strano, E., Viana, M., da Fontoura Costa, L., Cardillo, A., Porta, S. and Latora, V. 2013. Urban street networks, a comparative analysis of ten European cities. *Environment and Planning B: Planning and Design.* **40**(6), pp.1071-1086.

Sulis, P., Manley, E., Zhong, C. and Batty, M. 2018. Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science.* **2018**(16), pp.137-162.

Sun, Y. 2018. Sharing and riding: How the dockless bike sharing scheme in China shapes the city. *Urban Science*, **2**(3), p68.

Sun, L., Axhausen, K.W., Lee, D.-H. and Cebrian, M. 2014. Efficient detection of contagious outbreaks in massive metropolitan encounter networks. *Scientific Reports.* **4**, p5099.

Sun, L., Jin, J.G., Axhausen, K.W., Lee, D.-H. and Cebrian, M. 2015. Quantifying long-term evolution of intra-urban spatial interactions. *Journal of The Royal Society Interface.* **12**(102), p20141089.

Tang, J., Scellato, S., Musolesi, M., Mascolo, C. and Latora, V. 2010. Small-world behavior in time-varying graphs. *Physical Review: E.* **81**(5), p055101.

Tang, T., Liu, R. and Choudhury, C. 2020. Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. *Sustainable Cities and Society.* **53**, p101927.

Vespignani, A. 2018. *Twenty years of network science*. London: Nature Publishing Group.

Vinceti, M., Filippini, T., Rothman, K.J., Ferrari, F., Goffi, A., Maffeis, G. and Orsini, N. 2020. Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking. *EClinicalMedicine.* **25**, p100457.

Volchenkov, D. and Blanchard, P. 2008. Scaling and universality in city space syntax: between Zipf and Matthew. *Physica A: Statistical Mechanics and its Applications.* **387**(10), pp.2353-2364.

Wang, J., & Lindsey, G. 2019. Do new bike share stations increase member use: A quasi-experimental study. *Transportation research part A: policy and practice*. **121**, pp.1-11.

Wang, T., Yue, W., Ye, X., Liu, Y. and Lu, D. 2020. Re-evaluating polycentric urban structure: A functional linkage perspective. *Cities.* **101**, p102672.

Wasserman, S. and Faust, K. 1994. *Social network analysis: Methods and applications.* Cambridge: Cambridge university press.

Wei, S., Teng, S.N., Li, H.J., Xu, J., Ma, H., Luan, X.l., Yang, X., Shen, D., Liu, M. and Huang, Z.Y. 2019. Hierarchical structure in the world's largest high-speed rail network. *PloS One.* **14**(2), pe0211052.

Wei, Y., Song, W., Xiu, C. and Zhao, Z. 2018. The rich-club phenomenon of China's population flow network during the country's spring festival. *Applied Geography.* **96**, pp.77-85.

Wilkinson, S.M., Dunn, S. and Ma, S. 2012. The vulnerability of the European air traffic network to spatial hazards. *Natural hazards.* **60**(3), pp.1027-1036.

Xu, X.-K., Zhang, J. and Small, M. 2010. Rich-club connectivity dominates assortativity and transitivity of complex networks. *Physical Review: E.* **82**(4), p046117.

Yang, Y., Comber, A., Heppenstall, A. and Turner, A. 2017. Exploring spatiotemporal features of station-free bike sharing trips: case study of Shenzhen. *GeoComputation 2017*, 7 September 2017, Leeds.

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2019a. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems.* **77**, p101361.

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2019b. Who, where, why and when? Using smart card and social media data to understand urban mobility. *ISPRS International Journal of Geo-Information.* **8**(6), p271.

Yin, C., Zhu, S., Chen, H., Zhang, B. and David, B. 2015. A method for community detection of complex networks based on hierarchical clustering. *International Journal of Distributed Sensor Networks.* **11**(6), p849140.

Younes, H., Nasri, A., Baiocchi, G. and Zhang, L. 2019. How transit service closures influence bikesharing demand; lessons learned from SafeTrack project in Washington, DC metropolitan area. *Journal of Transport Geography.* **76**, pp.83-92.

Zhong, C., Arisona, S.M., Huang, X., Batty, M. and Schmitt, G. 2014. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science.* **28**(11), pp.2178-2199.

Zhong, C., Schläpfer, M., Müller Arisona, S., Batty, M., Ratti, C. and Schmitt, G. 2017. Revealing centrality in the spatial structure of cities from human activity patterns. *Urban Studies.* **54**(2), pp.437-455.

Zhu, S., Masud, H., Xiong, C., Yang, Z., Pan, Y. and Zhang, L. 2017. Travel Behavior Reactions to Transit Service Disruptions: Study of Metro SafeTrack Projects in Washington, DC. *Transportation Research Record.* **2649**(1), pp.79-8

Chapter 3
A spatiotemporal and graph-based analysis of dockless bike-sharing patterns to understand urban flows over the last mile

Abstract

The recent emergence of dockless bike-sharing systems has resulted in new patterns of urban transport. Users can begin and end trips from their origin and destination locations rather than docking stations. Analysis of changes in the spatiotemporal availability of such bikes has the ability to provide insights into urban dynamics at a finer granularity than is possible through analysis of travel card or dock-based bike scheme data. This study analyses dockless bike-sharing scheme in Nanchang, China over a period when a new metro line came into operation. It uses spatial statistics and graph-based approaches to quantify changes in travel behaviours and generates previously unobtainable insights about urban flow structures. Geostatistical analyses support understanding of large-scale changes in spatiotemporal travel behaviours, and graph-based approaches allow changes in local travel flows between individual locations to be quantified and characterised. The results show how the new metro service boosted nearby bike demand, but with considerable spatial variation, and changed the spatiotemporal patterns of bike travel behaviour. The analysis also quantifies the evolution of travel flow structures, indicating the resilience of dockless bike schemes and their ability to adapt to changes in travel behaviours. More widely, this study demonstrates how an enhanced understanding of urban dynamics over the "last mile" is supported by analyses of dockless bike data. These allow changes in local spatiotemporal interdependencies between different transport systems to be evaluated, and support spatially detailed urban and transport planning. A number of areas of further work are identified to better understand interdependencies between different transit system components.

3.1 Introduction

Cities are complex systems, composed of people, places, flows, and activities (Batty, 2013). Quantifying their dynamics, system interdependencies and spatial structures can characterise urban morphology and metabolism. People as physical carriers, drive the flows of materials, money and information within urban spaces, and influence economic growth, social equity (Batty, 2013). Understanding the nature of these flows provides perspectives and insights into how socioeconomic and environmental problems such as urban development, transportation efficiency and air quality are being addressed (Borrego et al., 2006; Desouza and Flanery, 2013; Fishman et al., 2014).

Travel data can be used as proxies for urban flows because they describe people's movement. Traditionally, such data were obtained from household travel surveys, with high cost and time overheads. Recent research has used automated mass transit fare-collection data (e.g. travel cards of bus and metro travel), which is cheap and has high spatiotemporal granularity, to analyse urban flows travel behaviours and mobility patterns. However, very little research has considered urban morphology and metabolism. Zhong et al. (2014) and Gong et al. (2017) used smart card data (bus and metro) and graph-based approaches to quantify the dynamics of urban structures through the analysis of spatial networks. This characterises by medium-long distance travel but fails to reveal dynamics in local areas over short distances. Some research has used cell phone data to detect urban travel flows and some aspects of urban structure (e.g. home-to-work commuting structures) (Calabrese et al., 2011; Louail et al., 2014), but this lacks spatial detail due to cellular positioning, with median errors of hundreds of meters (Zandbergen, 2009). This results in large uncertainties when inferring people's movement over shorter distances. Thus, much previous work has examined broad scale urban flows, but with little consideration of finer scale "capillary" flows. These are characterised by non-motorised trips (walking, cycling), and they have the capacity to reveal the nature of urban flows over the "last mile".

This study examined dockless bike-sharing data from Nanchang, China over the period when a new metro line came into operation and compared "before" and "after" to reveal changes in travel behaviours, mobility patterns and flows over the last mile through spatial and analyses of dockless bike usage.

3.2 Background

Bike-sharing schemes have become increasingly popular in recent years, reflecting their environment friendly, low cost and convenient nature. They are understudied, with research focused on cycling behaviours associated with dock-based bike-sharing schemes. Vogel et al. (2011) examined geographical clusters of docking stations using spatiotemporal usage. Others have identified bike fleet rebalancing strategies for different types of stations and quantified links between bike demand and land use (Daddio and Mcdonald 2012; Jiménez et al., 2016; Kaltenbrunner et al., 2010; O'Brien et al., 2014).

However, bike-sharing schemes can play an important role in examining the "first/last mile" problem. This is the distance between home/workplace and public transport that is too far to walk (Fishman, 2016; Saberi et al., 2018; Shaheen et al., 2010), and bike schemes provide access to other forms of public transport and mass transit (train, metro, bus etc.): they act as the "capillaries" for the mass transit aorta. The advent of dockless bike schemes opens up the opportunity to examine the last mile in detail.

To understand the last mile using bike data, the provision of other transport systems needs to be considered as well. Many studies have examined how cycling and metro trips are combined (Martens, 2007; Lin et al., 2017), how this varies for different socio-economic groups (Zhao & Li 2017), are affected by pricing (Lin et al., 2017) and has sought to quantify the interdependencies between bike-sharing schemes and metro systems (Ma et al., 2015; El-Assi et al., 2017; Ding et al., 2019). Most of these studies have found a positive correlation between metro stations and bike-sharing trips, but some have questioned this (e.g. Tran et al., 2015) and suggested that underlying land use patterns drive bike trip (residential, industrial and commercial). Overall, these studies have focused on sharing trips preference, bike trip spatial clustering around other transportation hubs (metro, tramway and railway stations), and have ignored flows and structures in the last mile. Examining the relationships between bike and transportation flows and structures can lead to a deeper understanding of urban dynamics. Saberi et al. (2018) analysed spatiotemporal statistics and network (graph structure) properties of bike-sharing trips to examine the impact of metro strikes, identifying increases in bike use (numbers and trip distances). Chen et al. (2016) constructed a framework to predict the short-term over-demand periods for sharing bike station clusters considering of metro delay. Both studies used data from dock-based bike-sharing schemes. These have a number of important shortcomings, including service coverage (e.g. bikes may not be available in the suburbs), docks may have large distances from the actual origin/destination (OD), and there can be dock capacity/availability issues. Thus the inference derived from analyses of these data are limited.

The development of the IoT (Internet of Things) has the potential to revolutionise many aspects of our lives which are increasingly connected and sensed, generating large volumes of data with location and temporal attributes. In bike-sharing schemes, dockless bikes emerged around 2015 and rapidly became a success in a number of countries, including China (iiMedia Research, 2017; Xu et al., 2019). Unlike traditional bike-sharing schemes where bikes can only be borrowed and returned at docking stations, dockless schemes enable users to locate and borrow bicycles via a smartphone app, returning them to any suitable public location. Dockless schemes allow convenience and flexibility for users. The smart lock system and GPS unit on the bikes not only facilitate scheme operation and bike management, but also create a large quantity of spatiotemporal individual level data. An advantage over data from traditional dock-based bike-sharing schemes, is that the flows captured by the data are more detailed (i.e. with higher spatial granularity) and better capture people's actual activities, travel demands and behaviours.

Dockless bike-sharing studies are few and have focused on bike fleet management (Pal and Zhang, 2017), planning of related infrastructures (Bao et al., 2017), and bike distribution patterns (Liu et al., 2018). Liu et al. (2018) proposed combining a factor analysis and convolutional neural networks for inferring dockless sharing bike distribution in new cities. However, this work treated these spatial distributions as a temporally static problem and ignored any temporal changes over time. Other studies (Ai et al., 2018; Xu et al., 2019; Yang et al., 2018) have shown that bike distribution is significantly time-dependent, especially around metro stations, suggesting that dockless bike flows and activities are highly dynamic. Zhang et al. (2019) developed a framework for planning geo-fences to constrain dockless bike parking. Clustering analysis and location-allocation models were applied to assess the implications of spatial planning of geo-fences in different scenarios. While this framework incorporated spatial detail, it lacked explanatory social and economic information, for example, related to the cost of geo-fences and punishment/reward of parking bikes outside / inside geo-fences. Zhou et al. (2018) used questionnaires to examine attitudes over the effects of dockless bikes on metro commuting. Their results showed a significant positive influence, especially for non-motorised vehicle owners and metro stations outside city centres and highlighted the positive role of such schemes in mass transit systems.

Analyses of data from dockless bike-sharing schemes have the capacity to provide high resolution insights into people's non-motorised mobility patterns and behaviours and to reveal their relationship with other urban structures and processes, for example, flow networks in other mass transit system (e.g. metro), urban infrastructure development (e.g. new train stations, new bridge) and related urban updates. Critically such data allow such relationships to be examined dynamically over the last mile.

3.3 Study area and data

3.3.1 Study area

Nanchang in southeast China is the capital city of Jiangxi province with a population of 2.15 million. The city has a typical humid subtropical climate, characterised by hot and humid summers and mild winters. It has two metro lines as of September 2017. A new line of 17 stations, Metro Line 2, opened and started running on August 18, 2017. Figure 3.1 shows the transit map of Nanchang with the Gan river running through the city and the two metro lines. Other public transit systems include bus and dockless bike-sharing. This study analysed data of around 80,000 dockless bikes from the Nanchang urban area around the time of the opening of Metro Line 2, specifically to compare bike usage “before” and “after” the opening of the new metro line.

3.3.2 Data description

A program was set up to collect dockless bike availability data via the bike scheme API (application programming interface) for the month of August 2017. Queries to the API can return bike availability for any specified location (point), returning information on bike identifiers and their coordinates, with an in-built limit of the nearest 30 available bikes. The program iterated through the whole urban area collecting data on bike availability on a raster grid of 0.0015 degrees (length of sides equals to approximately 150 m). Most available bikes locations across Nanchang were captured approximately every four minutes due to the large urban area and the API query limits. There were some gaps in coverage due to bike GPS and communication unit signal receiving problems (e.g. GPS does not function well in tunnels). The data include bike ID (identifier), coordinates and timestamp information. Figure 3.2 (a) shows a snapshot of dockless bikes in Nanchang, with each point representing an available dockless bicycle.

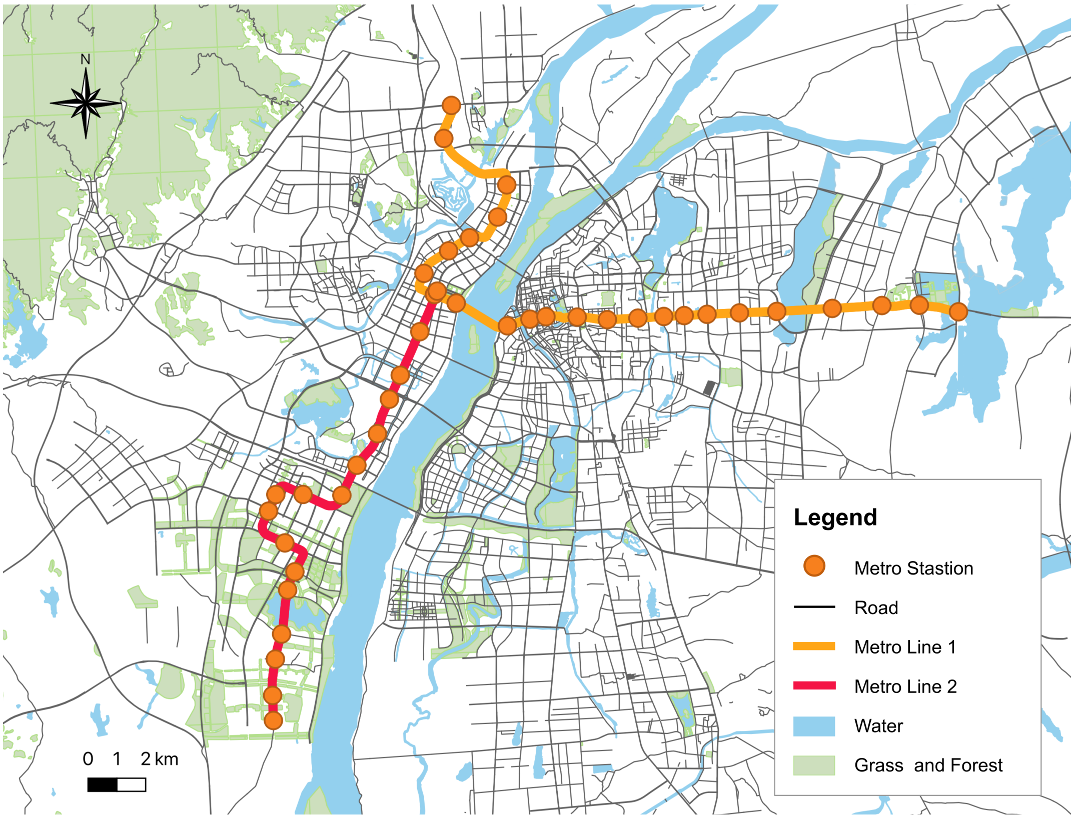
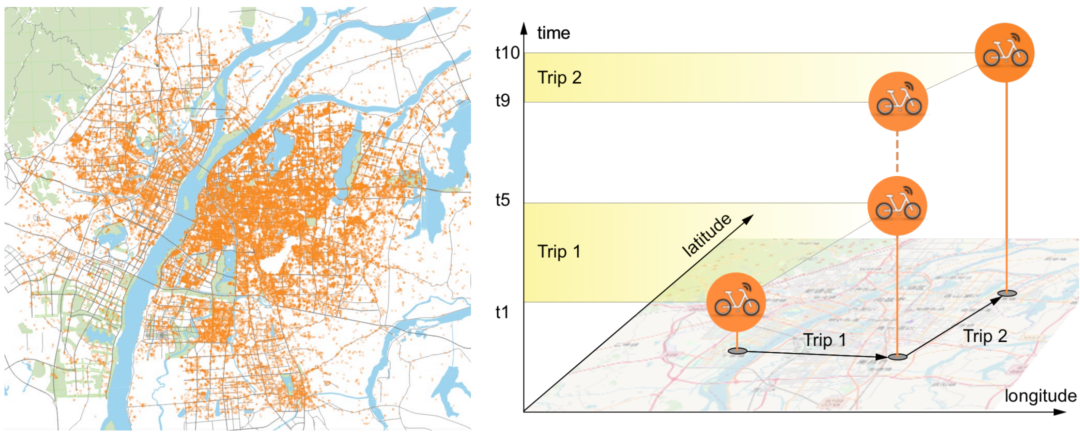


Figure 3.1 Study area of Nanchang, China. (source: OpenStreetMap)

Extreme weather conditions have been found to have significant negative effects on cycling activity (El-Assi et al., 2017; Zhou et al., 2017). To ensure consistency across both time periods, only data from rain-free days were analysed, with weather data from wunderground.com (one meteorological station, every 3 hours). Although other conditions (e.g. air condition, temperature) may have minor impacts on bike usage, these were relatively consistent over the study periods. Data for five weekdays before and five weekdays after the opening of the new metro line were selected for analysis, with all data collected in the same month and under the same general weather conditions on rain-free days.

Table 3.1 shows an example of the data collected, and table 3.2 shows the dates of the data used in the analysis. Open Street Map data were also used to provide basic background mapping and road network data.



(a) (b)

Figure 3.2 (a) Snapshot of dockless bike distribution; (b) detecting bike trips.

Table 3.1 Example of data records.

Time	Bike ID	Longitude	Latitude
2017-08-07 13:35:29	7910***002	115.9190	28.61415
2017-08-07 13:35:29	7910***748	115.8367	28.78435
2017-08-07 13:35:29	7910***911	115.8369	28.78446

Table 3.2 Date of collected data.

Period	Day of week	Date
Before	Monday, Tuesday	7, 8 August, 2017
	Wednesday -Friday	2, 3, 4 August, 2017
After	Monday- Friday	21-25 August, 2017

3.4. Methods

3.4.1 Detecting bike trips

The data provide snapshots of available dockless bike locations, bike identifier and timestamps. From the stacks of snapshots, it is possible to identify the changes in location of an individual bike and thus to derive dockless bike trip information. For this study, the threshold trip value was 100 metres: if the change in location of an available bike between two timestamps exceeded this threshold, then the two records were linked as a trip, with the earlier one providing the origin coordinates, and the latter one the destination. It should be noted that this method is unable to identify round trips where the origin is close to the destination. Figure 3.2 (b) shows an example trip. The bike changes its location from *t1* to *t5*, which are combined to form a bike trip. It stays in the same place from *t5* to *t9* and is considered as available in this period. Similarly, a second trip can be identified from *t9* to *t10*. The dockless bike-sharing scheme also allows bikes to be scheduled (booked) in advance of 15 minutes, but this still allows trip origin and destination to be identified, although with a small impact on trip duration.

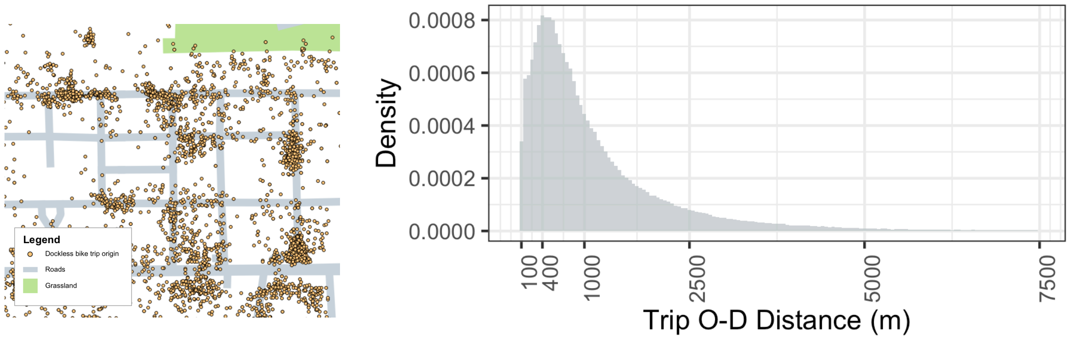
3.4.2 Constructing dockless bike mobility graph (network) structure

Converting the bike data into a graph structure to represent flows allows the changes arising from the introduction of a new metro line to be examined. Typically, any system composed of interconnected individuals can be viewed as a graph (i.e. network) with individual components are represented by nodes (i.e. vertices) and their interconnections by arcs or links (i.e. edges). Examples include online friendship networks, scientific collaboration networks and global airline routes. In urban and transportation studies, bike-sharing docking stations (Austwick et al., 2013), bus stops and metro stations are typically viewed as nodes. If there is at least one trip between a pair of nodes (stations), then a link is generated between them. By representing the relationship between transportation nodes (e.g. bike/bus/metro stations) as travel flow links, mobility graph structures can be constructed. For example, Saberi et al. (2018) studied the dynamics of the bike-sharing scheme in London, casting docking stations as graph nodes, and flows between stations as links. However, dockless bike-sharing schemes provide greater spatial detail about the patterns of bike movements and therefore their riders than docking station data. Because bikes and bike trips are not aggregated over docking stations, constructing graphs of dockless trips requires different considerations. Figure 3.3 (a) shows an example of bike trip origins in an urban area. Instead of being located at fixed docking stations, they are more loosely distributed along roads, and sometimes clustered at certain road segments. Figure 3.2 (a) also confirms that the majority of the bikes are located close to roads, with few of them distant to the road network in the north and southeast. These are locations outside the Nanchang urban area, and distant from the metro system.

This study uses road segments to aggregate bike locations and to cast them as vertices in a graph structure. The graph links represent trips originating from one road segment to another. Choosing the right spatial scale of aggregation is important because this can impose a source of bias in spatial analysis, i.e. the MAUP (modifiable areal unit problem) (Openshaw, 1979). Some research (Calabrese et al., 2011; Louail et al., 2014) using cell phone data to detect commuting flows in cities choose grid cells to aggregate flow origin and destinations with size varying from 500 m to 2 km. These scales may be useful for capturing work-home commuting flows and inferring coarse-scale urban structure, but here the focus was on identifying more spatially detailed structures in small areas requiring a finer spatial scale. Figure 3.3 (b) reveals that trip distances are commonly around 400 m and that 28% of trip O-D distance are less than 500 m, with 1km and 2km distances corresponding to 60% and 85% of trips respectively. Here, road segments of 200 m were used to aggregate dockless bike flows, since this distance captures most of the trips. More than 96% of bike travels' distance are larger than 200 m, so this threshold should be sufficient to capture all the flows between different segments, while larger distance (such as 400 or 600) may lead to a higher level of information loss.

The choice of using grid cells or road segments is relatively arbitrary and depends on the local characteristics of the road network, urban morphology and the people's mobility pattern. In the case study area of Nanchang, the service catchment of the new metro line does not have a dense road network, especially around the more rural end of it. Therefore, using street networks can help aggregate the movement more effectively, also avoid using a large number of grid cells to construct a large and sparse graph. The work of Gao et al. (2020) assessed the scale impact and MAUP of dockless bike trip aggregation using different aggregation units, size ranging from 100 m to 1000 m, with an interval of 100 m. The result (Gao et al., 2020) suggested that road length is the most stable factor under varying scales. In other words, the work of Gao et al. (2020) implied that road segments of equal or similar length could be a very suitable choice for aggregating dockless bike-sharing trip origins and destinations.

Figure 3.4 illustrates the process of casting bike trips to a graph using road networks as nodes. First, the Nanchang road network was split into segments based on road joins. Second, if a split segment was longer than 200 m, then it was divided into smaller segments of equal distance, all less than 200 m. Then, bike trip origin and destination were assigned to the nearest road segment. The result is that each trip has information about its origin and destination road segment, with a trip regarded as a flow from one node to another.



(a) (b)

Figure 3.3 (a) Examples of bike trip origin spatial distribution; (b) dockless bike trip distance probability distribution.

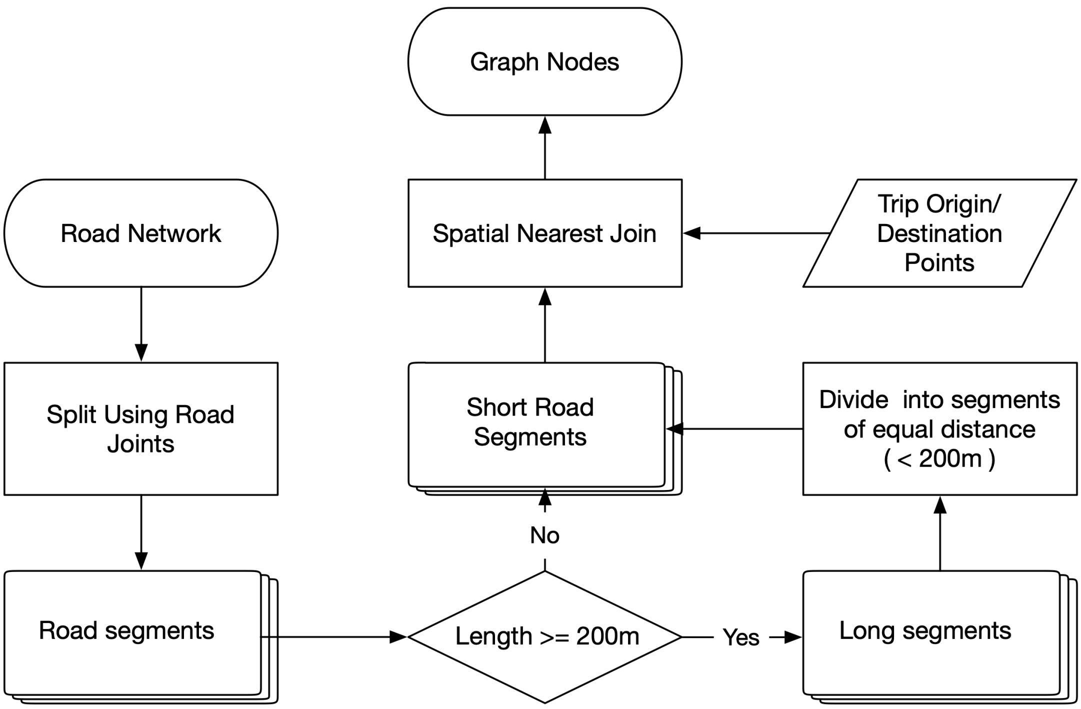


Figure 3.4 Diagram of creating graph nodes from the road network.

Dockless bike travel flows and the graph structures in the “before” period (5 days) are shown in figure 3.5, using a spatial layout and a Fruchterman Reingold layout. Each node represents a short road segment with the edge between any two nodes indicating travel flow between them, and the graph was trimmed using a threshold of at least ten trips between two nodes, to improve the visualisation. Figure 3.5 characterises nodes degree (number of connections) using both shading and node size. Node degree in an urban mobility network represents the connectivity or accessibility to destinations or activities across the network. In this case study, bike travel is split by the river, with few links crossing it (figure 3.5 a). Most of the trips are local, with most of the links connecting nearby areas (figure 3.5 a), and the graph structure has a clear multi-core (multi-cluster) spatial pattern. This is different to graph structures derived from other transportation modes such as metro and bus travel (e.g. Zhong et al. 2014) which tend to link different parts of the city over longer distances. The Fruchterman Reingold layout (figure 3.5 b) seeks to reduce the overlaps between nodes and maintain spatial topology. For example, two clusters of nodes presented in figure 3.5 (c) and (d) are shown in figure 3.5 (e) and (f), respectively. The group of nodes in figure 3.5 (c) and (e) are dominated by two nodes connected to surrounded nodes with low degree, while figure 3.5 (d) and (f) suggest a group of more evenly connected nodes. The structures imply many different patterns of how people move around and use urban space.

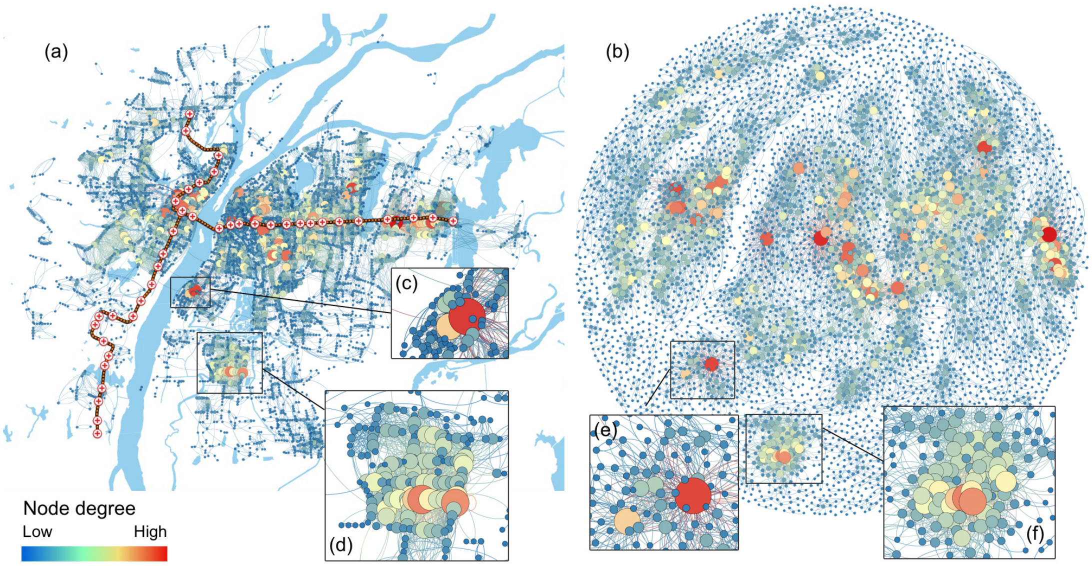


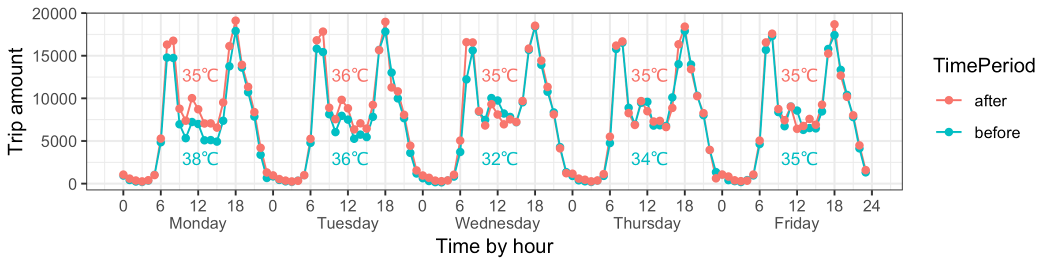
Figure 3.5 Dockless bike travel flows and graph structure of "before period". (a) Spatial layout; (b) Fruchterman Reingold layout, with detail in (c) to (f) as described in the text.

3.5 Result and discussion

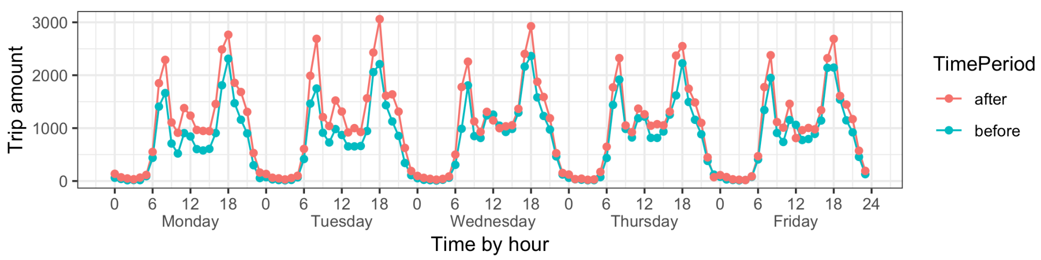
3.5.1 Spatiotemporal analysis

3.5.1.1 Temporal pattern

Temporal analysis of bike usage can reveal the dynamics in and characteristics of dockless bike-sharing schemes. Figure 3.6 (a) shows the temporal travel patterns of dockless bike trips in Nanchang (whole study area) with the highest daily temperature. Trip amount is the count of trips starting at different times in hourly intervals. Over the two periods, bike usage shows a similar overall temporal pattern with some difference in trip numbers. There are two significant rush hours each day, one from 07:00 lasting for two hours and another around 18:00, with the hourly trip number reaching approximately 19,000. There is also a lunchtime peak at 12:00. The trip amounts overlap with the exception of Monday at 12:00, when it is slightly lower, which is potentially due to the hotter weather (38°C) than any other days (figure 3.6 a). The total trip amounts for the two periods is 838,464 and 892,764, respectively, an increase of 6.5%.



(a)



(b)

Figure 3.6 Temporal pattern of dockless bike trips in (a) Nanchang city; (b) around new metro service catchments.

Figure 3.6 (b) defines service area catchments of 2 km around new metro stations. Trip amounts in these catchments increased by 28.0%, from 96,239 to 123,182. The highest rise was found in the early morning with 38.5%, 38.0%

and 31.3% in the hour intervals from 06:00-09:00, respectively. There are also substantial increases in the afternoon peak. The observations indicate the popularity of trips using the new metro with dockless bike as a preferred travel mode during morning and afternoon commuting.

3.5.1.2 Spatial pattern

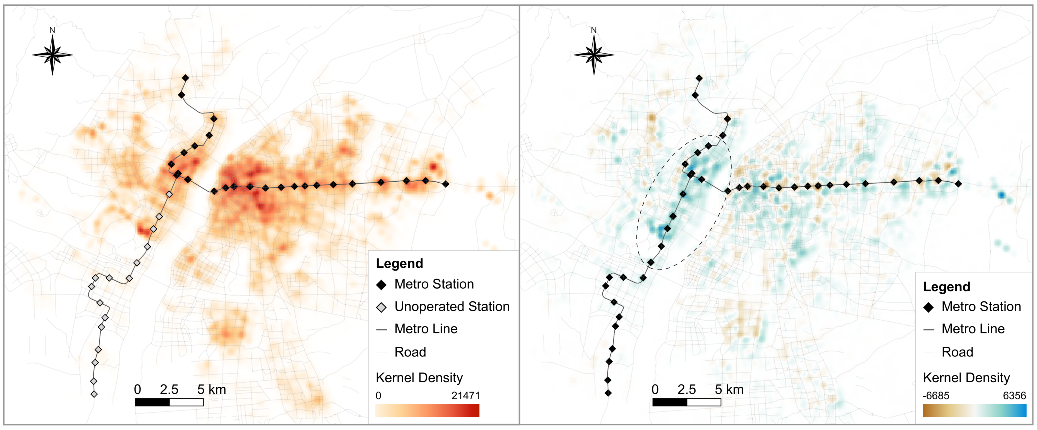
Dockless bike-sharing scheme offers flexibility, efficiency, and low cost making it an attractive way to fill the gaps in the public transit system (such as bus and metro). Bike-sharing travel has different spatial properties from other public transit modes, due to its non-motorised nature. For example, dockless bikes are not suitable for long distance travel, but they can be left immediately adjacent to the destination. Understanding the unique spatial pattern of dockless bike usage is important for scheme management (bike rebalancing) and the interdependence between metro and dockless bike system provides crucial information about local mobility. A spatial analysis was conducted to examine the characteristics and dynamics of dockless bike-sharing scheme in the study area. Figure 3.7 (a) shows the dockless bike trip kernel density across the city in the five weekdays of the before period based on trip origins. A KDE bandwidth of 118.2 m was determined as follows:

$$band\ width = 0.9 * \min\left(SD, \sqrt{\frac{1}{\ln(2)} * D_m}\right) * n^{-0.2} \quad (3\text{-}1)$$

where SD is the standard distance of the points spatial distribution, D_m is the median distance, and n is the number of points. The KDE has a 100 meters resolution.

The city centre from where most trips originate, straddles both sides of the river, and it has the highest kernel density. Several hubs can also be observed in the south, away from the city centre and the metro service.

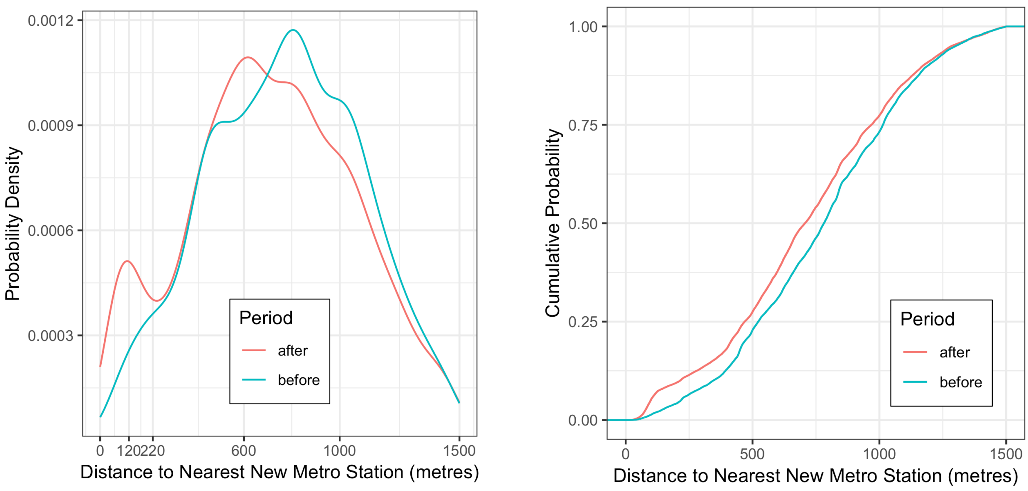
Figure 3.7 (b) shows the changes in kernel density estimation (KDE) between the two periods. The highlighted area in figure 3.7 (b) shows the new metro service catchments west of the river. It shows a clustering pattern of increased density. By contrast, in other parts of the city, areas of increased bike use are adjacent to or nearby to areas of decrease, suggesting an overall balanced and dynamic pattern. The observed spatial proximity between increase and decrease is due to the nature of dockless bike-sharing scheme: people are able to pick up and park bikes at places of their choice when necessary, subject to availability. The flexibility of travel is demonstrated by heterogenous use patterns, especially at fine spatial scales.



(a) (b)

Figure 3.7 (a) Kernel density of dockless bike trips in "before period"; (b) kernel density difference of dockless bike trips in the two periods.

To better understand the impact of the new metro service on dockless bike-sharing trips, a further statistical analysis was conducted to examine the probability density of dockless bike origin/destination spatial distribution. Figure 3.8 (a, b) illustrates the probability and cumulative probability of start and end distances from the newly opened metro stations.



(a) (b)

Figure 3.8 (a) Probability density; and (b) accumulative probability of distance between dockless bike trip origin/destination and nearest new metro station in the two periods.

In figures 3.8 (a), a new peak at around 120 metres in the “after” period is observed, which suggests that more bike trips originate from or end at areas very close (around 120 m) to new metro stations. This distance (120 m) can also be understood as how far metro travellers are typically walking to get or leave bikes around their metro usage. The first peak (figure 3.8 a) decreases and reaches a floor around 220 m, indicating weakened spatial clustering of bike trips origins and destinations. This distance can be interpreted as the affordable (walking distance) limit between the metro station and the bike, and provides information to support the planning of related bike parking facilities, suggesting in this case an upper limit for the provision of bike parking areas around the metro stations. The highest peak in the “before” period is located at around 800 m, but moves to 600 m in the “after” period (figure 3.8 a), suggesting the impact of the new metro stations on bike usage. The cumulative probability (figure 3.8 b) indicates that this impact is heterogeneous, with higher cumulative probability in areas at shorter distances to new metro stations in the “after” period. Figure 3.7 and 3.8 indicate the extent to which the areas closer to new metro stations experienced greater changes in daily dockless bike trips, in terms of distances people walked to get a bike and volume of trips.

Metro services not only change where people pick or park their bikes, but also influence local trip distance patterns. Building on the results in figure 3.8 (a), buffers of 250 m and 2000 m around new metro stations were used to select bike trips with origin or destination within them. Trips that start or end in the 250 m buffer can be understood as “last mile” trips, and trips in the 2000 m buffer capture flows within the metro service catchment area, indicating changes over larger areas. The travel distance patterns were examined their distributions compared (figure 3.9).

Figure 3.9 (a, c) indicates that bike trip counts increased significantly in both buffer zones and that the new metro service stimulated more trips, especially short bike travels. Figure 3.9 (b, d) describes the changes in probability distributions and indicates that trips with a travel distance of less than 1,000 m are more likely to be made with the introduction of the new metro service and long distance trips (greater than 1,000 m) less likely. *t*-tests were used to confirm the difference in dockless bike trip distances in the two periods (table 3.3) for the two trip types. The *p*-values in each case are less than 0.001, indicating the statistical significance of the observed reduction in trip distances with the introduction of the new metro line. Mean travel distance dropped from 1178 m to 1034 m for “last mile” trips, and trips in the larger service areas around new metros (radius of 2000 m) significantly decreased from 1317 m to 1150 m. Overall, the results suggest that new metro stations encouraged increases in cycling activity but with reduced trip distances, indicating improvements in local mobility and access to transportation services.

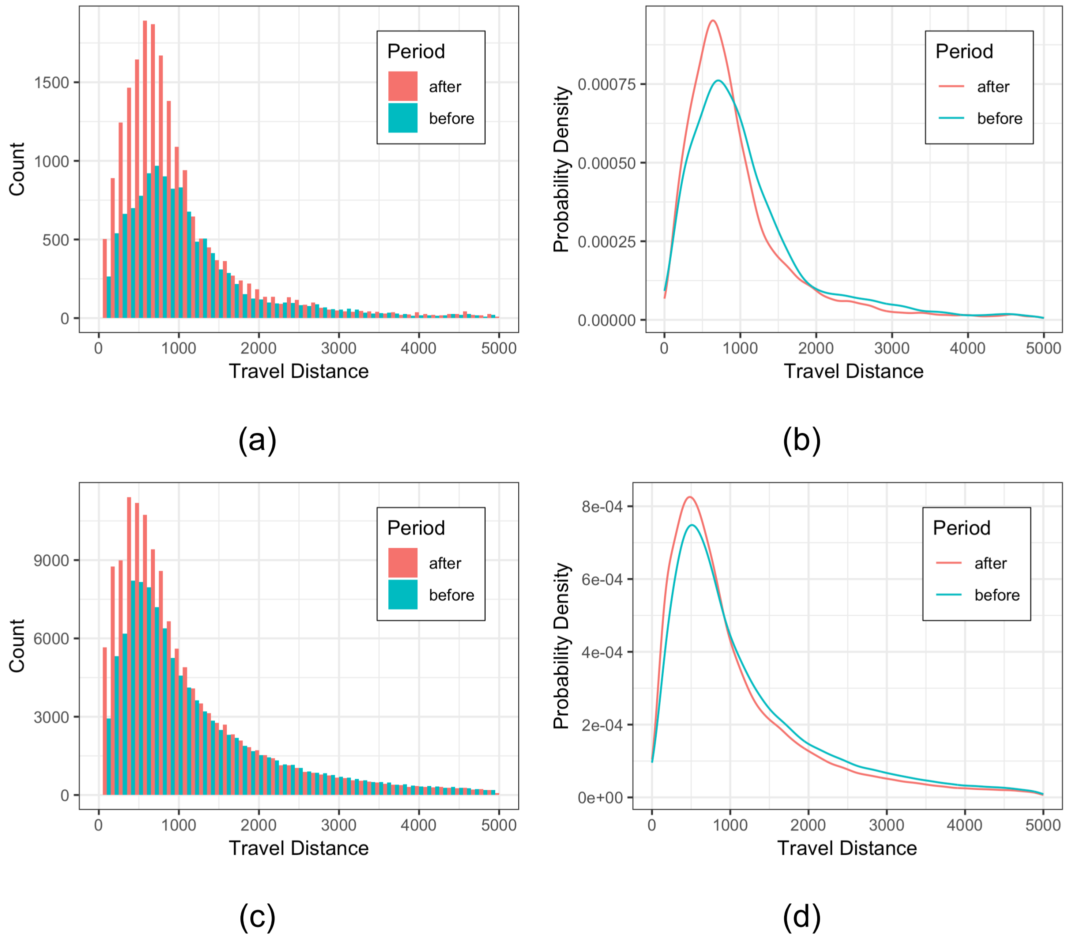


Figure 3.9 Distributions of dockless bike trip travel distance around new metro stations over the two periods: (a) Density distribution (buffer within 250 m); (b) probability density (buffer within 250 m); (c) density distribution (buffer within 2000 m); (d) probability density (buffer within 2000 m).

Table 3.3 *t*-test for travel distance (m).

buffer radius	p-value	mean value		95% confidence interval	
		before	after	before	after
250	< 0.001	1178	1034	1059-1297	865-1203
2000	< 0.001	1317	1150	1160-1473	973-1327

3.5.2 Graph-based analysis

Interpreting the changes in flow networks is important for understanding changes in mobility patterns and urban dynamics. Graph-based approaches were applied to quantify and compare the differences in the network structures. Firstly, two graph structures were constructed based on bike travel flow data for the two periods according to the method described above. Only trips with their origin and destination road segments in the 2000 m buffers around metro stations were selected, as this was considered the maximum distance people would travel using the combination of dockless bike and metro.

3.5.2.1 Network (Graph) statistical properties

Table 3.4 shows the network (graph) properties in the new metro service catchments. From the table, some changes between the two periods can be identified: the number of nodes increased from 984 to 1064, and the number of edges increased from 34948 to 41967. Network connectivity, δ, representing graph structure resilience, the higher value suggests its robustness in the graph structure to against disruptions of potential nodes failures.

Table 3.4 Graph properties in the two periods.

Properties	Before	After
Number of nodes (N)	984	1064
Number of edges (L)	34948	41967
$\delta = 2L/N^2$	0.072	0.074
Total Flux	85393	111920
Mean Node flux	173.6	210.4
Mean node degree	71.0	78.9
Variance of Node flux	343.2	385.6
Variance of Node degree	87.9	90.8
Mean clustering coefficient	0.6455	0.6560

Node degree helps evaluate the connectivity to and accessibility of destinations in a mobility graph (Zhong et al., 2014). Table 3.5 confirm the significance of the changes in node degree, with the mean value increasing from 71.0 to 78.9. Figure 3.10 (a) shows the CDF (Cumulative Density Function) of node degree in the two periods, both of them follows the power law, and the pattern implies a higher probability of observing a node with larger degree in the “after” period than “before”. The node flux of the graph, evaluate the total amount of trips that start from or end at a node, can be used to understand trip volumes. The rises in both total and mean flux (table 3.4 and 3.5) indicate that the demand for dockless bikes in local areas increased. The node flux CDF in figure 3.10 (b) also confirms this finding, indicating a higher probability of larger weights links. These changes describe the increased attractiveness of road segments, as well as their interaction strength in connection to other segments in the network. Node flux variability (table 3.4) increases from 343.2 to 385.6, the significant (p-value<0.001 as shown in table 3.6) rise indicates a more heterogeneous distribution of interaction strength across the network. Clustering coefficient (table 3.4) represents the extent to which nodes in a network tend to cluster (i.e. have links between them) (Saberi et al., 2018). Its average value shows an increase after the new metro opening, suggesting that the dockless bike trip network became more locally connected.

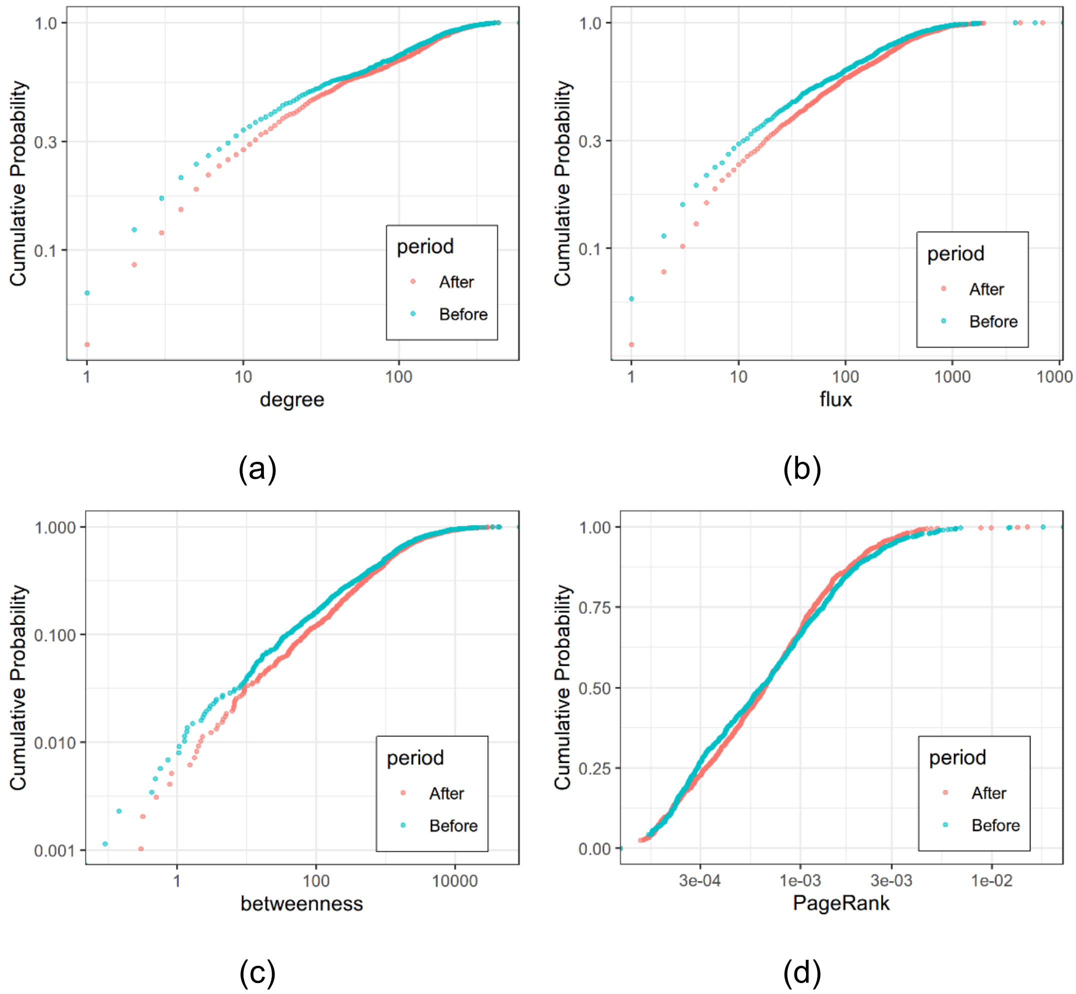


Figure 3.10 Cumulative Density Function of graph properties; (a) degree; (b) flux; (c) betweenness; (d) PageRank.

Table 3.5 *t*-test for degree and flux.

graph property	p-value	mean value		95% confidence interval	
		before	**after**	**before**	**After**
degree	0.046	71.0	78.9	65.5-76.5	73.4-84.3
flux	<0.001	173.6	210.4	152.1-195.1	187.2-233.6

Alternative hypothesis is true: both true differences in means of degree and flux are not equal to 0.

These results suggest that the dockless bike-sharing mobility network became denser and more heterogeneous after the opening of the new metro service. Similar patterns are evident in other graph properties. Firstly, changes in betweenness centrality were examined through "node betweenness" (Newman, 2005) rather than "edge betweenness" (Girvan & Newman, 2002). Node betweenness represents the extent to which nodes stand between each other in a graph, or serve as a bridge from one part to another (Newman, 2005). In the context of urban studies, this measure can identify hubs in flow networks (Zhong et al., 2014). Figure 3.10 (c) shows the cumulative probability of node betweenness centrality, and it indicates that the probability of low betweenness nodes dropped, while the probability of higher betweenness nodes increased. The pattern implies that nodes with higher levels of connectedness have a more intensive role in the "after" graph, which suggests the emergence of well-connected hubs.

Table 3.6 Variance test for degree and flux difference.

graph property	p-value	ratio of variances (before divide after)	ratio of variance (95% confidence interval)
degree	0.2989	0.9369	0.8289-1.0595
flux	<0.001	0.712	0.630-0.805

Alternative hypothesis (flux) is true: the true ratio of variances of flux is not equal to 1.

PageRank is another indicator of graph node importance. Generally, if the number of highly centred nodes (with very high PageRank) decreases while the number of secondary PageRank nodes increases, then this implies a polycentric change. These suggest that the influence of strongly centred nodes has gradually relaxed with their centrality increasingly shared with emerging subcentres (Zhong et al., 2014). Figure 3.10 (d) shows the CDF of PageRank, the “after” period exceeds the “before” period at around 0.75 of the cumulative probability (y-axis) and indicates that nodes are more likely to have secondary high PageRank, suggesting underlying polycentric transformation. The next two sections reinforce this finding by detecting community structures in graphs, as well as interpreting od flow changes.

Table 3.7 shows the result of Kolmogorov-Smirnov test to investigate the changes of the four graph indices as presented in figure 3.10. All p-values are smaller than 0.05 and confirmed the changes are statistically significant.

Table 3.7 Kolmogorov-Smirnov test results.

Graph Property	d-value	p-value
Degree	0.062099	0.03879
Flux	0.077167	0.004538
Betweenness	0.076525	0.00502
PageRank	0.066294	0.02236

3.5.2.2 Community detection and graph structure

Communities in a dockless bike-sharing mobility graph can be understood as a set of road segment clusters (graph subsets, or sub-graphs) that are more connected by trips internal to the cluster than external. This difference in internal and external connections in clusters is measured by modularity (Newman and Girvan, 2004). A modularity maximisation algorithm, the Louvain method (Blondel et al., 2008), was applied to detect the communities in the dockless bike network. Modularity characterises the density of edges inside potential communities relative to edges outside of the community. Networks with high modularity have dense connections between community members but sparse connections with nodes in different communities. Optimising this value theoretically results in the best possible segmentation of the nodes in a given graph structure. The Louvain method first finds small communities by optimising modularity locally on all nodes, then each small community is grouped into one node, and the first step is repeated in an iterative way.

Figure 3.11 shows the different communities detected in the two periods using colour shades. Although the analysis does not include node (road segment) location, the communities are spatially coherent, suggesting considerable spatial structure in this graph. The introduction of new metro stations resulted in the emergence of new communities in the graph structure (figure 3.11). One is located around the transfer station in the north, and another can be found near the centre of the network and demonstrates the impact of a new metro line on the structure of dockless cycling activities. Combined with the changes in other graph properties, it is possible to conclude that the new metro catchment area results in more bike trips and forms stronger local travel connections, which are more polycentric.

Another important indicator graph structure is assortativity. This describes the tendency of high degree nodes (nodes with many edges), to connect to other high degree nodes. For example, the structures in figure 3.5 (d) and (f) have higher assortativity than the structures of figure 3.5 (c) and (e). Structures with high assortativity are more robust to node removal or failure. Assortativity increased from 0.1832 to 0.2895 after the introduction of the new metro service, along with a slight increase in modularity, from 0.3236 to 0.3733. From a transportation perspective, the rise in the two indicators suggests that the new network alleviates congestion and enhances the efficiency of travel by non-motorised traffic (discussed in Sun et al., 2012). Here, the evolution in graph structure shows a quick self-adaptive process that meets the increasing and changing patterns of travel demand and traffic flows, and strongly implies underlying urban resilience. Whether the performance of the network will deteriorate with future increases in network size or a dramatic rise in trip numbers, remains to be seen. From the perspective of “urban metabolism”, the assortative structure suggests a robust spread and interaction of various urban flows of information, capital and materials through non-motorised traffic, but may also pose future challenges, for example, in the context of protecting against the spread of epidemics.

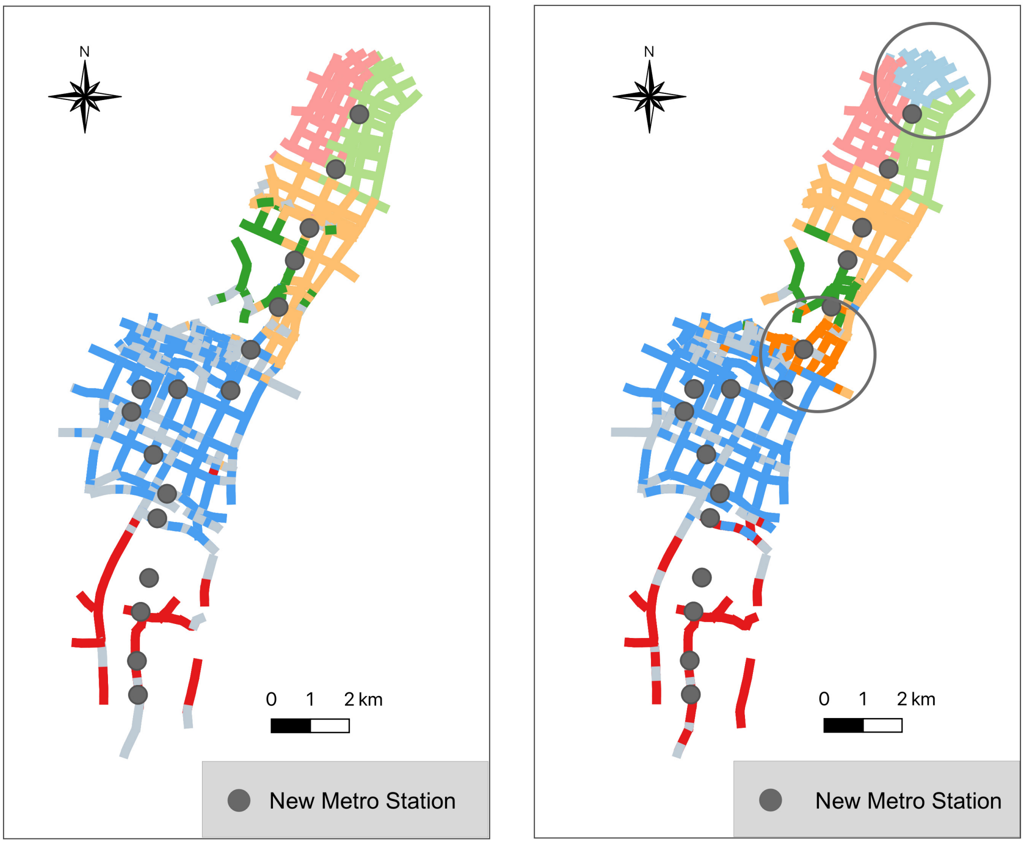
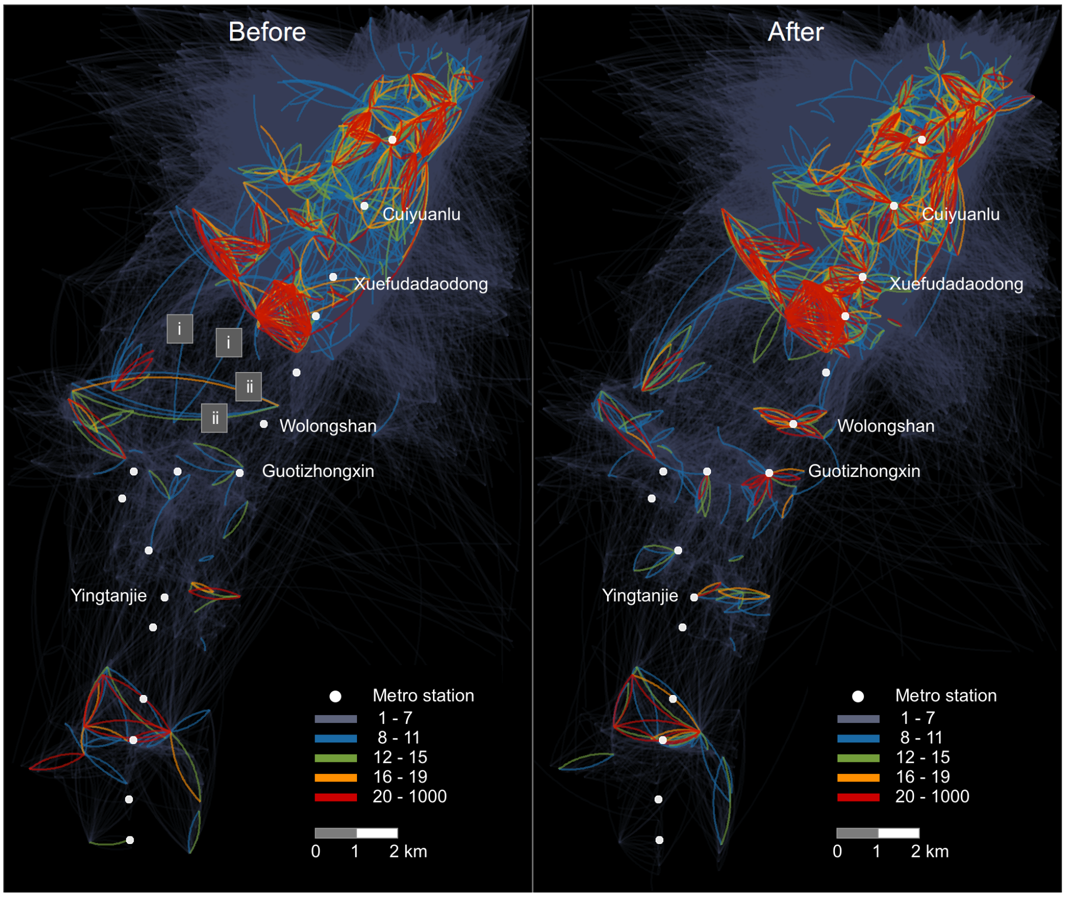


Figure 3.11 Communities of road segments for dockless bike-sharing trip in two periods.

3.5.2.3 Flowmap

A final exploration of the changes in bike flow through flow maps is shown in figure 3.12. This illustrates bike travel origin or destinations in the 2000 m buffers, and the colour shades are used to indicate trip frequency between different road segments, connected by the directed links. Several important differences can be observed, especially in areas close to the new metro stations. For example, Cuiyunlu, Xuefudadaodong, Wolongshan and Guotizhongxin stations all show large changes in flow and structures. These stations resulted in many new outward flows and regions of high trip destination density. The flows “i” and “ii” in figure 3.12 (a) suggest that bike users travel long distances from their origins to destinations in the “before” period, but these flows disappeared in the “after” period (figure 3.12 b). This is because travellers used the metro service instead of taking long-distance cycling trips, implying local mobility improvements. In the before period, there are many travel flows on the east of Yingtanjie Station (a large residential area). After the opening of the new metro service, more bike travel flows extended from and linked to Yingtanjie station, indicating that residents quickly changed their travel habits and started to combine dockless bikes trips with metro travel.



(a) (b)

Figure 3.12 Flow maps of dockless sharing bike trips in new metro service catchment, (a) “Before” period; (b) “After” period.

3.6 Conclusion

In order to capture the impact of metro service on the dockless bike-sharing system, this study applied a combination of geo-statistical and graph theory approaches. The analyses led to an in-depth understanding of the interaction and evolution of urban behaviours around non-motorised activity, by considering the changes in the spatial patterns of dockless bike-sharing as a result of a new metro line. The new metro service increased nearby dockless bike-sharing demand by 28%, and resulted in changes in other spatiotemporal patterns of travel behaviours, including bike travel distance and bike origin-destination spatial distributions. The observed changes in travel were not homogenous across the study area, with greater impacts closer to new metro stations. The dynamics and evolution in the graph structure capture the urban resilience of dockless bike schemes that are able to adapt to infrastructural changes such as new metro systems. Dockless bike trip structure has a tendency towards being polycentric (i.e. with more community structures), stronger local connectivity, higher assortativity reflecting increased travel demand and scheme robustness. This insight is not only useful for dockless bike-sharing schemes, but also provides a new perspective for the analysis on urban resilience and the interdependence between different urban complex systems. Observations from flow maps can be used to indicate improvements in local mobility, and the speed of adaption by people to combine dockless bikes for metro travels.

A number of policy and planning implications emerge from this study. First, it is important for bike-sharing operators to prepare more bike fleets in the new metro service areas as demand will increase. Second, bike parking facilities need to be planned around metro stations over distances up to 220 m. Third, analysis of the changes in the origin-destination and network structure can help to reveal which locations (roads) are more frequently used by bike users, thus supporting related planning. Last, bike fleet rebalancing strategies need to be redesigned to adapt to changes in flows.

There are several limitations to this work. First, it used short road segments (less than 200 m) to aggregate bike travel flows, which, although selected through analysis, may be specific to this study. Another analysis based on hexagon grid cells (size of 1 square km) has also been conducted and presented in the previous work of Yang et al. (2018); similar results and patterns in graph indices (e.g. degree and strength) has been found, which suggested that the findings are relatively robust under different travel aggregation methods and scale. Second, the study analysed data from a limited number of days. Climate (rainy days in summer) and the timing of the school year (the new term starts in September) were the major reasons for excluding data in order to minimise the variance of environmental factors. The changes in bike travel behaviours and associated flow structures could be confirmed if data for a longer period was available and through analysis of a similar case study area, with newly opened metro lines. Third, the data acquisition (section 3.3.2) procedure is only able to obtain information about the nearest 30 available bikes for each location, at each query. This potential data loss, if more bikes are available in that small area, introduces uncertainty into the analysis and results. Such shortcoming is also found in other dockless bike-sharing research (e.g. Shen et al., 2018; Liu et al., 2018; Ai et al., 2018).

However, since the data are consistent over the two periods, with similar sampling characteristics and uncertainties, the findings from “before” and “after” the new metro line are comparable. Finally, although dockless bike-sharing data provide a better representation of non-motorised travel behaviour than cell phone data, there is an inherent sampling bias because not all city inhabitants will be cyclists or will use a bike-sharing scheme. Future work will examine these issues further and will extend the analysis to integrate urban context, in order to develop a deeper understanding of the implications of changes in spatiotemporal patterns and graph structure evolutions. It may also examine the long-term spatiotemporal effects of new metro stations on dockless bike-sharing whose future patterns may change over time, as a result of deeper integration of the two travel systems.

References

Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W. and Ju, Y. 2019. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Computing and Applications.* **31**(5), pp.1665-1677.

Austwick, M.Z., O’Brien, O., Strano, E. and Viana, M. 2013. The structure of spatial networks and communities in bicycle sharing systems. *PloS One.* **8**(9), pe74685.

Bao, J., He, T., Ruan, S., Li, Y. and Zheng, Y. 2017. Planning bike lanes based on sharing-bikes' trajectories. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17 August, Halfiax. New York: Association for Computing Machinery, pp.1377-1386.

Batty, M. 2013. *The new science of cities.* Cambridge: MIT press.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment.* **2008**(10), p10008.

Borrego, C., Martins, H., Tchepel, O., Salmim, L., Monteiro, A. and Miranda, A.I. 2006. How urban structure can affect city sustainability from an air quality perspective. *Environmental Modelling & Software.* **21**(4), pp.461-467.

Calabrese, F., Di Lorenzo, G., Liu, L. and Ratti, C. 2011. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. *IEEE Pervasive Computing.* **10**(4), pp.36-44.

Chen, L., Zhang, D., Wang, L., Yang, D., Ma, X., Li, S., Wu, Z., Pan, G., Nguyen, T.M.T. and Jakubowicz, J. 2016. Dynamic cluster-based over-demand prediction in bike sharing systems. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing,*12-16 September 2016, Heidelberg. New York: Association for Computing Machinery, pp.841-852.

Daddio, D.W. 2012. Maximizing bicycle sharing: an empirical analysis of capital bikeshare usage. Msc. dissertation, University of North Carolina at Chapel Hill's.

Desouza, K.C. and Flanery, T.H. 2013. Designing, planning, and managing resilient cities: A conceptual framework. *Cities.* **35**, pp.89-99.

Ding, L., ZHANG, K.P. and XIE, B.I. 2019. Exploring the effects of building environments on the use of bike sharing: Case study of Shenzhen, China. In: *Proceedings of 2019 International Conference on Energy, Power, Environment and Computer Application,* 20-21 January 2019, Wuhan. Lancaster: DEStech Publications, pp.158-162.

El-Assi, W., Mahmoud, M.S. and Habib, K.N. 2017. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation.* **44**(3), pp.589-613.

Fishman, E. 2016. Bikeshare: A review of recent literature. *Transport Reviews.* **36**(1), pp.92-113.

Gao, F., Li, S., Tan, Z., Wu, Z., Zhang, X., Huang, G., & Huang, Z. 2021. Understanding the modifiable areal unit problem in dockless bike sharing usage and exploring the interactive effects of built environment factors. *International Journal of Geographical Information Science*. Advance online publication.

Girvan, M. and Newman, M.E.J. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America.* **99**(12), pp.7821-7826.

Gong, Y., Lin, Y. and Duan, Z. 2017. Exploring the spatiotemporal structure of dynamic urban space using metro smart card records. *Computers, Environment and Urban Systems.* **64**, pp.169-183.

iiMedia Research, 2017, *2017 Q1 China Renting Bicycle Market Research Report.* Beijing: Iresearch [Accessed 8 December 2020]. Available from https://www.iimedia.cn/149077640503942531.pdf

Jiménez, P., Nogal, M., Caulfield, B. and Pilla, F. 2016. Perceptually important points of mobility patterns to characterise bike sharing systems: The Dublin case. *Journal of Transport Geography.* **54**, pp.228-239.

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

Lin, J.J., Wang, N.L. and Feng, C.M. 2017. Public bike system pricing and usage in Taipei. *International Journal of sustainable transportation.* **11**(9), pp.633-641.

Liu, Z., Shen, Y. and Zhu, Y. 2018. Inferring dockless shared bike distribution in new cities. In: *Proceedings of the eleventh ACM international conference on web search and data mining*, 5-7 February 2018, Los Angeles. New York: Association for Computing Machinery, pp.378-386.

Louail, T., Lenormand, M., Cantu Ros, O.G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J.J. and Barthelemy, M. 2014. From mobile phone data to the spatial structure of cities. *Scientific Reports.* **4**(1), p5276.

Ma, T., Liu, C. and Erdoğan, S. 2015. Bicycle sharing and public transit: does Capital Bikeshare affect Metrorail ridership in Washington, DC? *Transportation Research Record.* **2534**(1), pp.1-9.

Newman, M.E. 2005. A measure of betweenness centrality based on random walks. *Social Networks.* **27**(1), pp.39-54.

Newman, M.E. and Girvan, M. 2004. Finding and evaluating community structure in networks. *Physical Review E.* **69**(2), p026113.

O'Brien, O., Cheshire, J. and Batty, M. 2014. Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography.* **34**, pp.262-273.

Openshaw, S. 1979. A million or so correlation coefficients, three experiments on the modifiable areal unit problem. In: Wrigley, N. eds. *Statistical applications in the spatial science.* London: Pion, pp.127-144.

Pal, A. and Zhang, Y. 2017. Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C: Emerging Technologies.* **80**, pp.92-116.

Saberi, M., Ghamami, M., Gu, Y., Shojaei, M.H.S. and Fishman, E. 2018. Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. *Journal of Transport Geography.* **66**, pp.154-166.

Shaheen, S.A., Guzman, S. and Zhang, H. 2010. Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record.* **2143**(1), pp.159-167.

Shen, Y., Zhang, X. and Zhao, J. 2018. Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation.* **12**(9), pp.686-700.

Sun, H. J., Zhang, H. and Wu, J. J. 2012. Correlated scale-free network with community: modeling and transportation dynamics. *Nonlinear Dynamics.* **69**(4), pp.2097-2104.

Tran, T.D., Ovtracht, N. and D'Arcier, B.F. 2015. Modeling Bike Sharing System using Built Environment Factors. *Procedia CIRP.* **30**, pp.293-298.

Vogel, P., Greiser, T. and Mattfeld, D.C. 2011. Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. *Procedia - Social and Behavioral Sciences.* **20**, pp.514-523.

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y. and Ratti, C. 2019. Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers, Environment and Urban Systems.* **75**, pp.184-203.

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2018. Examining the impact of new metro services on dockless bike sharing mobility patterns. In: *Proceedings of Association of Geographic Information Laboratories in Europe (AGILE) Conference 2018,* 12-15 June, 2018. Lund, Sweden. London: Springer, pp.1-4.

Zandbergen, P.A. 2009. Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning. *Transactions in GIS.* **13**, pp.5-25.

Zhang, Y., Lin, D. and Mi, Z. 2019. Electric fence planning for dockless bike-sharing services. *Journal of Cleaner Production.* **206**, pp.383-393.

Zhao, P. and Li, S. 2017. Bicycle-metro integration in a growing city: The determinants of cycling as a transfer mode in metro station areas in Beijing. *Transportation Research Part A: Policy and Practice.* **99**, pp.46-60.

Zhong, C., Arisona, S.M., Huang, X., Batty, M. and Schmitt, G. 2014. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science.* **28**(11), pp.2178-2199.

Zhou, M., Wang, D., Li, Q., Yue, Y., Tu, W. and Cao, R. 2017. Impacts of weather on public transport ridership: Results from mining data from different sources. *Transportation Research Part C: Emerging Technologies.* **75**, pp.17-29.

Zhou, S., Ni, Y. and Zhang, X. 2018. Effects of dockless bike on modal shift in metro commuting: A pilot study in Shanghai. In: *Proceedings of Transportation Research Board 97th Annual Meeting,* 7-11 January 2018, Washington DC. Washington DC: The National Academies of Sciences, Engineering, and Medicine, p18-04973.

Chapter 4
Understanding the impacts of public transit disruptions on bike-sharing schemes and cycling behaviours using spatiotemporal and graph-based analysis: A case study of four London Tube strikes

Abstract

Understanding the interactions between different travel modes is crucial for improving urban transport resilience, especially during times of disruptions and transit failures. As a flexible and sustainable travel mode, bike-sharing schemes are able to solve "first/last mile" problems in urban transit as well as provide an alternative to motorised traffic. This paper uses OD (origin and destination) trip data from the London Cycle Hire Scheme and temporal docking station bike availability data to explore the impact of four separate London Underground (Tube) strikes on bike-sharing usage and behaviours. The results suggest that bike-sharing usage generally rises in response to Tube disruptions, but the extent and nature of this rise in use varies according to the type of disruption. A novel measure of station pressure suggests that the scheme very quickly reaches saturated capacity and is unusable in certain parts of London during disruptions. A graph-based analysis reveals several changes in OD flow structures. This implies a modal shift from Tube to bike-sharing and a change of route behaviours amongst bike-sharing users. Weekday Tube strikes bring new behaviours and new OD pairs to the bike flow structures, whilst for weekend strikes, existing patterns are consolidated. The corollary is that more heterogenous OD trip patterns are introduced by higher volumes of commuting trips and intense competition of cycles/docks. Cyclists are forced into using alternative (second or third preference) docking stations with new behaviours, and possibly users, as journeys that would otherwise be made via the Tube are made via bike-sharing. Overall, this work comprehensively presents and compares the impacts of Tube strikes under varied circumstances and offers a detailed understanding of the changed cycling behaviours that could be used in transport planning and management.

4.1 Introduction

Public transit disruptions have become more frequent in recent years due to the increasing maintenance needs of ageing infrastructures, natural disasters as well as social and political events such as city-wide festivals and strikes (Zhu et al., 2017; Gonçalves et al., 2020; Rahimi et al., 2020). Such events and disruptions can significantly affect the resilience of transportation systems. Disruptions have many different consequences across the transport network, and characterising them, as well how travellers may respond to transit failures, can inform urban transport planning and decision-making. Among different travel modes in big cities, bike-sharing schemes are low-cost, highly flexible and convenient (Shaheen et al., 2013). In urban contexts, they fill an important gap between pedestrian and vehicular transport (Curran, 2008), and can provide a genuine alternative travel mode when other parts of the transportation system experience disruptions. Previous research (Green et al., 2012; Zhu et al., 2017; Younes et al., 2019) has shown that disruptions to metro and bus systems may result in a shift to bike-sharing schemes, especially for low-income groups, as bike-sharing is a low-cost alternative to, for example, private taxi and minicab services.

Understanding the changes in cycling behaviour during transit disruptions is crucial for minimising the impacts in the short-term, also benefits sustainable transport planning in the long-term (Dill et al., 2003; Zhu et al., 2010). Mass transit disruptions can prompt new behaviours and introduce new people to cycling. Research has shown that many new cyclists used bike-sharing in London during previous transit strikes (Green et al., 2012), and may have continued to use bike-sharing schemes subsequently (Zhu & Levinson, 2010). A more contemporary context is the reduced public transport capacity as a result of public health outbreaks, which require social distancing. This is likely to lead to an increase in cycling activities. Quinn (2020) predicts a multi-fold increase in cycling in London post-lockdown, and several plans have been made to overhaul the capital's streets and public space. These include: (1) the rapid construction of a strategic cycling network to help reduce crowding on bus and metro services; and (2) the transformation of local town centres to allow people to walk and cycle where possible (Quinn, 2020). The impacts of reduced public transport on shifts to cycling may be determined from the analysis of historical bike-sharing data, and the results can be used to inform related sustainable urban and transport planning.

During disruptions, a range of different travel behaviours emerge according to the spatiotemporal characteristics of the incidents. To understand the changes in travel behaviours, observational data need to be analysed within their spatial and temporal context. Such spatiotemporal analyses have been published (Vertesi, 2008; Zhu et al., 2017; Saberi et al., 2018), but not in a comprehensive and large-scale way, partly due to data availability. Large transit disruptions (e.g. Tube strikes) are relatively rare events, and difficult to compare whilst controlling across changing explanatory variables. Previous work also has focused on changes in demand, i.e. related to user behaviour (Vertesi, 2008; Zhu et al., 2017; Saberi et al., 2018) and has not considered the resource supply such as the provision of cycles and cycle docks. Information on dock availability could provide supporting context to explain user behaviours and may also support bike scheme operators in their fleet management strategies, which include manually redistributing bikes.

Until recently, reliable “impact” and “behaviour change” analysis has been problematic due to a relative lack of historical time series data. However, the London Cycle Hire Scheme (LCHS) has been in operation since 2010, and each timestamped user trip has been continuously recorded over this period. This greatly benefits long-term service analysis (Lovelace et al., 2020) and comparative studies, for example, characterising how usage and behaviours change in response to different events and interventions (Beecham, 2015).

The work presented in this paper tries to address the above problems by examining four London Tube strike events and their impacts on LCHS using freely available bike OD (origin-destination) usage data and station availability data. These strikes have varied temporal and spatial characteristics, which results in distinct patterns of change in LCHS. Temporal, spatial and structural patterns are examined to consider the impacts of Tube (i.e. metro, Underground) strikes on the bike-sharing scheme. The results have the potential to support LCHS service provision, to guide strategies for filling public transport gaps, and to strengthen transit resilience.

The structure of this paper is as follows: Section 4.2 reviews recent literature on bike-sharing studies related to disruptions and user behaviour, and section 4.3 introduces the methods as well as case study. The results are presented in section 4.4, providing insight on user behaviour changes and scheme dynamics in bike-sharing due to transit disruptions. Section 4.5 summarises and compares the findings to those of other research, before conclusions are drawn (section 4.6).

4.2 Background

Transit disruptions can adversely affect transport network reliability and bring substantial economic, social and safety impacts to cities and travellers (Wilson, 2007; Bauernschuster et al., 2017; Pregnolato et al., 2017; Yu et al., 2020). To minimise the impacts, people may change their travel behaviours according to the characteristics of the disruption (Cairns et al., 2002). For example, a short-term disruption (e.g. transport strike or a bridge closure) may lead to temporary changes in travel mode, choosing alternative destinations, reductions in journey frequency, etc. The behaviours in response to disruptions may also become permanent as new travel habits emerge (Zhu & Levinson, 2010).

A number of studies have analysed individual perceptions of and preferred reactions to transit disruption using questionnaires and survey data (Tsuchiya et al., 2008; Fukasawa et al., 2012; Teng et al., 2015). These suggest that patterns of temporary modal shifts during transit disruptions are related to income (Zhu et al., 2017). Different sharing-economy travel options offer new ways to minimise the impact of transit service failure, with wealthier people more likely to switch to taxis or car ridesharing (e.g. Uber, Lyft), and people with lower incomes choosing lower-cost mobility services such as bike-sharing (Zhu et al., 2017). Studies evaluating the impact of transit disruption on bike-sharing usage are critical for mitigating the impacts for disadvantaged groups.

Many large cities such as London have introduced bike-sharing schemes composed of a network of docking stations and bikes into their urban centres. They are used heavily by tourists and commuters, especially for short journeys that would otherwise be made by bus and metro (Shaheen et al., 2013). Bike-sharing works well when linked to public transport, solving the so-called `first/last” mile problem in urban transit - for example, by supporting short connecting trips from a major transport hub to a workplace or home (Yang et al., 2019). The majority of bike-sharing studies can be grouped into two classes (Beecham, 2015): exploratory studies analysing variations in scheme usage related to the built and social-spatial environment (Faghih-Imani et al., 2014; El-Assi et al., 2017); and more narrowly-focused studies developing algorithms for supporting fleet management and rebalancing (De Chardon et al., 2016). Nello-Deakin (2020) suggest that over the last twenty years, there has been an abundance of empirical studies with similar conclusions drawn, namely that urban environments with dedicated cycling infrastructure, traffic calming measures and moderate to high urban densities are associated with higher cycling rates (Nello-Deakin, 2020) and bike-sharing usage (El-Assi et al., 2017). Numerous algorithmic approaches have been proposed in the literature to solve the traffic prediction and rebalancing problems (De Chardon et al., 2016; Pan et al., 2019). In contrast, there are comparatively few studies examining the interdependence between bike-sharing and other transit modes, especially during disruption events or infrastructure changes.

Transit disruptions clearly have the most direct impact on public transport provision and influence bike-sharing use substantially (Chen et al., 2016). The work of Chen et al. (2016) combined bike-sharing usage data in New York with event data from multiple sources (Twitter, traffic data live feeds) to rank the impact of various social and transportation events on bike-sharing. They suggest that metro delays have a much larger impact on bike-sharing use than other disruption events such as surface road congestion and restrictions. Metro strikes (or closure for maintenance) are not included in Chen et al. (2016) 's events data set, and it is reasonable to speculate that they may have a higher or at least similar level of influence on bike ridership. This is due to the fact that strikes will make the metro service unavailable for a longer period, thus providing a more radical disruption than a delay to the schedule. Although highlighting the importance of understanding interdependence between metro and bike, the work of Chen et al. (2016) does not explore changes in spatiotemporal patterns of bike-sharing scheme use.

Further explorations into bike-sharing user's ridership changes may benefit scheme management activities, improve equity in mobility service provision, as well as long-term traffic planning. "Novice" cyclists who are not familiar with biking also make comparatively more trips on metro strike days (Green et al., 2012). These examples provide evidence that transit disruptions can introduce bike-sharing to new users, potentially promoting and increasing cycling rates in the long term.

Reviewing the literature has demonstrated that transport users may look to bike-sharing as an alternative to public transport during metro service failures. However, there is comparatively little research quantitatively examining changes in usage patterns in bike-sharing schemes while metro services are in disruption. Among the limited number of studies, Younes et al. (2019) analysed the impact of metro station closure for maintenance (a.k.a "surge") on bike-sharing demand. They found that the "surge" can lead to between 24% and 45% more trips in bike-sharing stations within 0.5 mile to the metro. However, "Surges" are very different to network- or line-level metro disruptions, because they operate over small spatial scales. Typically, only up to three metro stations are closed for maintenance (Younes et al., 2019);

therefore, travellers often find alternative routes within the metro station network. Network-level disruption is examined in the work of Saberi et al. (2018), which characterised the impact of a weekday Tube strike on all lines and stations in London, along with its effects on LCHS. The work suggests that the ridership increase shows a significant distance decay pattern: the closer a docking station is to metro lines, the higher ridership increase it will experience. However, the findings may only be applicable for weekdays, when a lot of journeys are made for commuting purposes to complete the "last mile" between the Tube or rail station and workplace. It is unclear whether this pattern still holds if the disruptions fall on holidays or weekends.

Over thirty strikes have occurred on London's Tube network since 2010 (Transport for London, 2017). These are caused by a mixture of factors, including disputes related to pay, safety, pensions and job security issues. The strikes may happen at the whole network-level, where at least the majority of the Tube service is unavailable, or at individual line-level, where only certain lines are disrupted. Impacts of these incidents on LCHS use will clearly be different depending on where and when disruptions occur as well as the duration of the disruption. According to Transport for London (2017), line-level strikes are more common in the London Tube than network-level strikes. Despite this higher frequency, the impact of line-level incidents on bike-sharing has rarely been examined in previous studies. An exception is Yang et al. (2019), which analyses how the introduction of a new metro line service stimulates bike-sharing ridership. Data from new forms of bike-sharing (dockless) are examined, which provide higher spatiotemporal granularity for understanding cycling activities and urban flows in the last mile. Yang et al. (2019) analyse the more flexible travel OD pairs in the dockless scheme, and the study indicates how a new metro line can rapidly boost local bike travel demand and result in emerging parking clusters within around 220 metres around new metro stations. The work also highlights the structural changes caused by new OD pairs (when cast into a graph), capturing system-level adaptations and responses to changes in demand. Although this work focused on the impact of new metro lines, the graph structure and metrics it used suggest research opportunities for quantifying large-scale behavioural responses to line-level disruption.

Whilst existing literature (Saberi et al., 2018; Yang et al., 2019; Younes et al., 2019) has evaluated the impact of events using travel records on patterns of use (demand), the dynamics of bike-sharing service supply have yet to be examined. Metro disruptions pose significant management problems for bike-sharing scheme operators (Younes et al., 2019), even if daily total trip frequencies do not substantially increase. Docking stations should always have cycles and empty docks available, but metro incidents may break this balance (Saberi et al., 2018). The dynamics in docking station capacity during disruptions have yet to be comprehensively analysed. The work presented in this paper starts to address the above gaps, by examining spatiotemporal and structural changes in bike-sharing schemes in relation to metro (Tube) strikes.

4.3 Method

4.3.1 Case study: London Cycle Hire Scheme (LCHS) and Tube strikes

The LCHS was launched by London's public transport authority (Transport for London, TfL) in June 2010, initially with 315 docking stations and 5000 bikes. The scheme expanded its service significantly with more bikes and larger coverage areas in March 2012 and December 2013. By 2014, it had become the world's second-largest bike-sharing scheme (Fishman, 2016), covering mostly central London. Due to the high prevalence of commuting activities, many LCHS trips are made to connect travellers' home, workplace and transit hubs. This presents a solution to the "first/last mile" problem by combining with train and Tube trips, but this combination can be interrupted during transit disruptions.

Whilst there have been more than 30 London Tube strikes over the last decade (Transport for London, 2017), several factors must be considered when studying the effect of such events on LCHS use. Both weather conditions and particular calendar events may also significantly change bike-sharing use, thus introducing uncertainties when comparing bike ridership patterns between strike and non-strike days. To control for these types of events, this work only focuses on rain-free day Tube strikes, and not on national holidays such as Christmas or bank holidays. A total of four Tube strikes have been selected (table 4.1), with two at network-level and two at line-level. Among these incidents, Strike 2 was at a weekend, while the others occurred on weekdays. A map of the London Tube and bike docking stations is shown in figure 4.1. The Piccadilly line is the fourth busiest line in the London Tube network, and serves many of London's key tourist attractions, including Harrods, Hyde Park and Buckingham Palace. It also connects with the major London King's Cross railway station. The Central line is the second busiest line, running through Oxford Street and the financial centre of the City of London. The Waterloo & City line is a shuttle line that runs between

Waterloo and Bank with no intermediate stops. Its primary traffic consists of commuters from southwest London, Surrey and Hampshire arriving at Waterloo station and connecting to the City, London's financial district.

Table 4.1 London Tube strikes.

Name	Category	Date	Tube Line	Day
Strike 1	Network-level	2015/07/09	All	Weekday
Strike 2	Network-level	2015/03/07	All	Weekend
Strike 3	Line-level	2018/09/27	Piccadilly	Weekday
Strike 4	Line-level	2018/10/05	Central, Waterloo & City line	Weekday

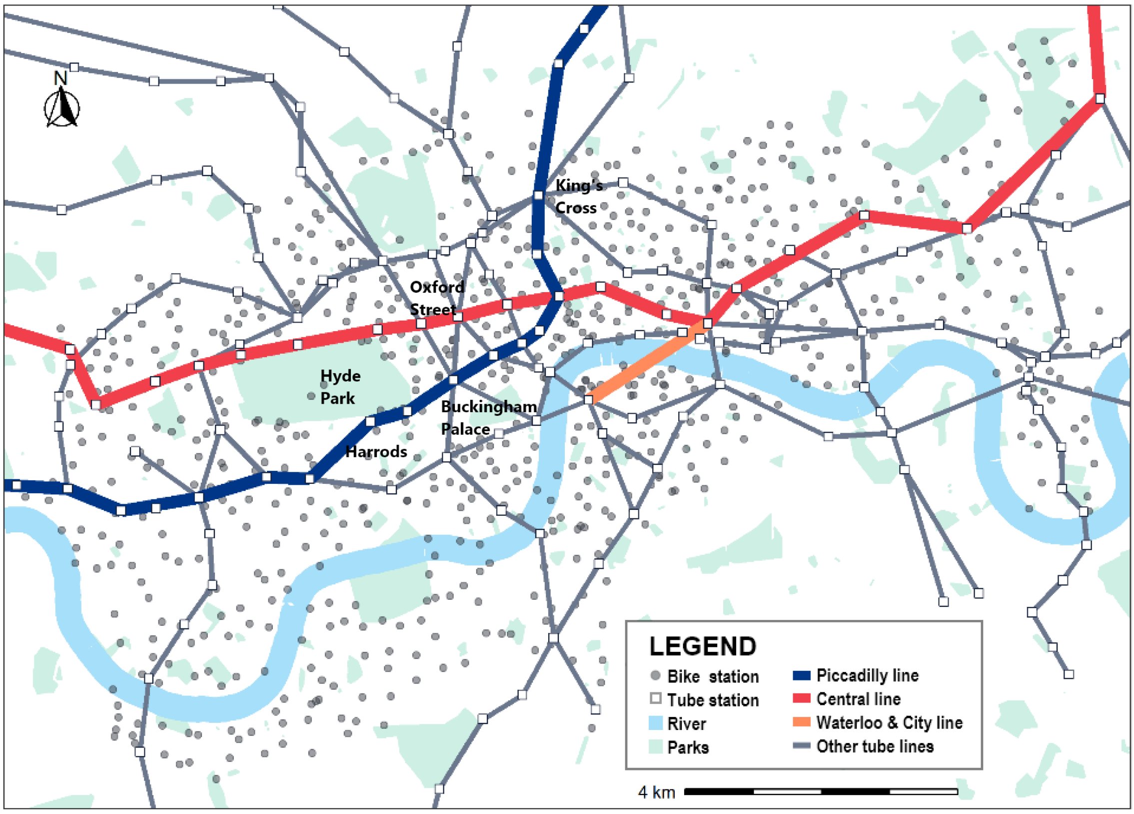


Figure 4.1 London Tube lines and bike-sharing docking stations.

4.3.2 Data

4.3.2.1 London Cycle Hire Scheme (LCHS) trip data

The LCHS data detailing trip origin-destination pairs (OD) is published by Transport for London (TfL), covering the period from 2010 to present (2020), and can be retrieved using the R package “bikeshare” (Padgham et al., 2017). Pre-processing work was carried out to remove redundant/duplicated and incomplete/faulty records from raw data. Each record in the cleaned dataset describes a single bike trip in the LCHS and contains complete information describing a trip’s start and end docking station, with associated timestamps. Therefore the trip OD flows can be examined in some spatial and temporal detail. Because the data covers over a decade, it can support longer-term analysis of the evolution of the scheme (Lovelace et al., 2020), or to understand and compare the behaviours and dynamics in LCHS during different events and periods, such as holidays, lock-down, and transit disruptions. In this research, bike travel records on four different strike days and their respective corresponding two non-strike days are examined and compared. To quantify the changes caused by Tube strikes, bike data on strike days (listed in table 4.1), their two nearest rain-free days of the same day of the week are used for comparison (Saberi et al., 2018). For example, if a strike happened on Friday, then data of the previous and subsequent Fridays (both rain-free) are used. Docking station location data, provided by the UK Consumer Data Research Centre[1], were applied to supplement spatial coordinates for trip origins and destinations. The data describes station-id, coordinates and several other variables such as the docking station’s opening date. When linked to LCHS trip OD records by matching on station-id, spatial details of travel flows can be obtained.

4.3.2.2 Docking station availability data

Data on the availability of bike docks are obtained from the LCHS live feed[2]. This records the number of available bikes and (empty) docks at each docking station every 10 minutes. The variables include station-id, timestamp, number of available bikes and number of empty docks. Stations can sometimes lack bikes or docks due to changes in bike-sharing demand throughout the day. For example, Beecham et al. (2014) revealed that many people ride bikes from their home for commuting in the morning. This renders some docking stations unavailable to users at specific periods. Examining the time-series of dock availability helps to evaluate variations in scheme “usability” during the course of the day, or in transit disruptions.

---

[1] https://www.cdrc.ac.uk

[2] https://tfl.gov.uk/tfl/syndication/feeds/cycle-hire/livecyclehireupdates.xml

4.3.3 Analysis

4.3.3.1 Spatiotemporal trip analysis

In order to shed light on when and how Tube disruptions may impact or increase bike travel, temporal analysis is conducted to compare the time-series of hourly bike travel counts over the four strike days and their comparison days. Within these periods, bike trip data were also aggregated over two different spatial units to characterise ridership change and its spatial patterns. Bike trips were first aggregated over a 500 m hexagonal grid (roughly 0.6 km2) covering the central London study area. The reason for using 500 m is that the average distance of nearby docking stations is reported as approximately 500 m (Duncan, 2015). Bike trip counts were allocated to the grid cells based on the origin station. The change in counts was determined by comparing counts from the control data (average value of 2 non-strike days for each disruption). The second approach was to aggregate bike trips over docking stations and to calculate changes (Saberi et al., 2018). Docking stations were categorised into different groups based on their shortest distance to disrupted Tube stations. A spatial interval of 250 metres was used as suggested in the work of Saberi et al. (2018).

4.3.3.2 Docking station availability analysis

When using bike-sharing schemes, travellers may encounter the problem of having no bike available at their desired trip origin, or more frustratingly, having no docks available at the desired destination. Increases in bike-sharing usage, and greater competition for bikes and docks, may lead to a higher rate of such “service outages” (De Chardon et al., 2016), decreasing the scheme’s reliability and attractiveness. Therefore, it is vital to ensure the availability of both bikes and docks during transit disruption. In LCHS, an individual station has 24 docks on average, and this work used a threshold of 15% to identify a station that is under low availability. That is to say, if a docking station has less than 15% bikes or 15% empty docks of its total capacity, then it is marked with the status of “low availability”. The proportion of 15% is an arbitrary choice, considering the typical dock capacity (24), it roughly equals to a mean value of three bikes or docks, and is a reasonable threshold for this analysis.

As described in section 4.3.2, availability data indicates the number of available bikes and docks at a frequency of ten minutes. Therefore, each station has six observations describing its status each hour. This paper defines the sum of “low availability” timestamps in each hour as a “service pressure” index, so it has the range of between zero to six. Higher values indicate that more timestamps have seen insufficient availability, whilst lower values suggest that the stations are often at a more balanced state. By analysing the fluctuation in service pressure in different groups of stations, the dynamics of the service provision can thus be evaluated, also reflecting their varying spatiotemporal patterns during strikes. The time periods of low bike usage, i.e. late night hours, were removed from the analysis due to the small number of trips made and the steady service provision at those times.

4.3.3.3 Graph analysis

Events in Tube networks may not only impact bike ridership volumes, but also change the structure of travel patterns and flows (Saberi et al., 2018; Yang et al., 2019). To quantify these structural changes, graph-based analysis was utilised in this study. First, the travel flows in LCHS were represented as directed and weighted graphs. A graph consists of a number of nodes (i.e. vertices), and they are connected by links (i.e. edges) to indicate their relationship and interactions, examples include social network graphs and global flight line graphs. To present LCHS as a directed and weighted graph, the bike docking stations were cast as nodes, while the cycling trips between them were defined as the links, with direction (from origin to destination) and weight (frequency) attributes. Once graphs are constructed, different indices can thus be derived to describe the state of the structure. By comparing non-strike and strike day graph indices, it is possible to characterise related structural dynamics and evolution, leading to a more comprehensive understanding of the changed relationships between OD pairs, and the different roles of docking stations. The indices can be categorised as measures for graph nodes, graph links and whole graph structure.

Node centrality measures are helpful for characterising graph nodes. They are indicators of the importance of individual vertices in a graph, and the definition of importance may vary over different indices. For example, the most common and basic node centrality is the degree. In this study, degree describes how many other bike stations in the graph are linked to a given bike station (with either in- or out-flows). There are also many other centrality measures such as flux, PageRank, node betweenness and eigenvector centrality. Eigenvector centrality (Bonacich, 2007) extends the idea of node degree by considering that nodes connected to other high centrality (degree) nodes should have a higher importance score than those connecting to low centrality nodes. It is also a relative measure, which ranges from zero to one, with higher values indicating larger importance and centrality.

Graph links can also be evaluated by various indices, with the most common being link weight. In the context of a bike-sharing network, link weight represents the number of bike trips (OD frequency) that travel from one docking station to another.

There are also graph (global) level indices or metrics that can be used to characterise the state of the whole structure, for example, graph transitivity and assortivity.

Transitivity (also called clustering coefficient), indicates the extent of graph nodes within a network cluster (i.e.community, subgroups, cliques). It captures the degree of local cluster (sub-graph) interactions compared to connections with nodes outside of the cluster (Saberi et al., 2018). It is calculated from the ratio between the observed number of closed triplets (triangles) and the maximum possible number of closed triplets in the graph structure.

Assortivity evaluates a preference for a graph's nodes to attach to others that are similar in centrality (Newman, 2002; Noldus et al., 2015). In the context of bike-sharing schemes, higher values imply that similar important docking stations, such as those close to transit hubs, are more likely to be connected by travel flows.

The various graph metrics were calculated for the periods during each Tube strike, and compared to those calculated for non-strike days, under the hypothesis that any changes in these may indicate the temporary structural responses to strike activity. Bike trips from the two comparison days were cast into two graphs, with their metrics calculated. Then the average value of the two was used for comparison. In addition to this, changes in the maps of flows were examined to confirm the findings and to provide context for interpreting the various graph metrics, thereby providing a deeper understanding of the observed trends. A particular focus was placed on OD pairs that experienced ridership increases, as these were hypothesised to represent higher transport needs. Thus, this work compares how different places are more strongly connected by LHCS users, which helps to interpret underlying travel behaviours.

4.4 Results

4.4.1 Bike-sharing usage characteristics

4.4.1.1 Temporal pattern

A consequence of Tube strikes is an increase in bike journeys, as shown in figure 4.2. A network-level weekday disruption (figure 4.2 a) was found to increase bike trip volumes from 37,070 to 69,734 (88%) throughout the day. At the weekend (figure 4.2 b) the numbers rise from 15,910 to 24,160 (52%), with a more significant peak time around 2-3 pm.

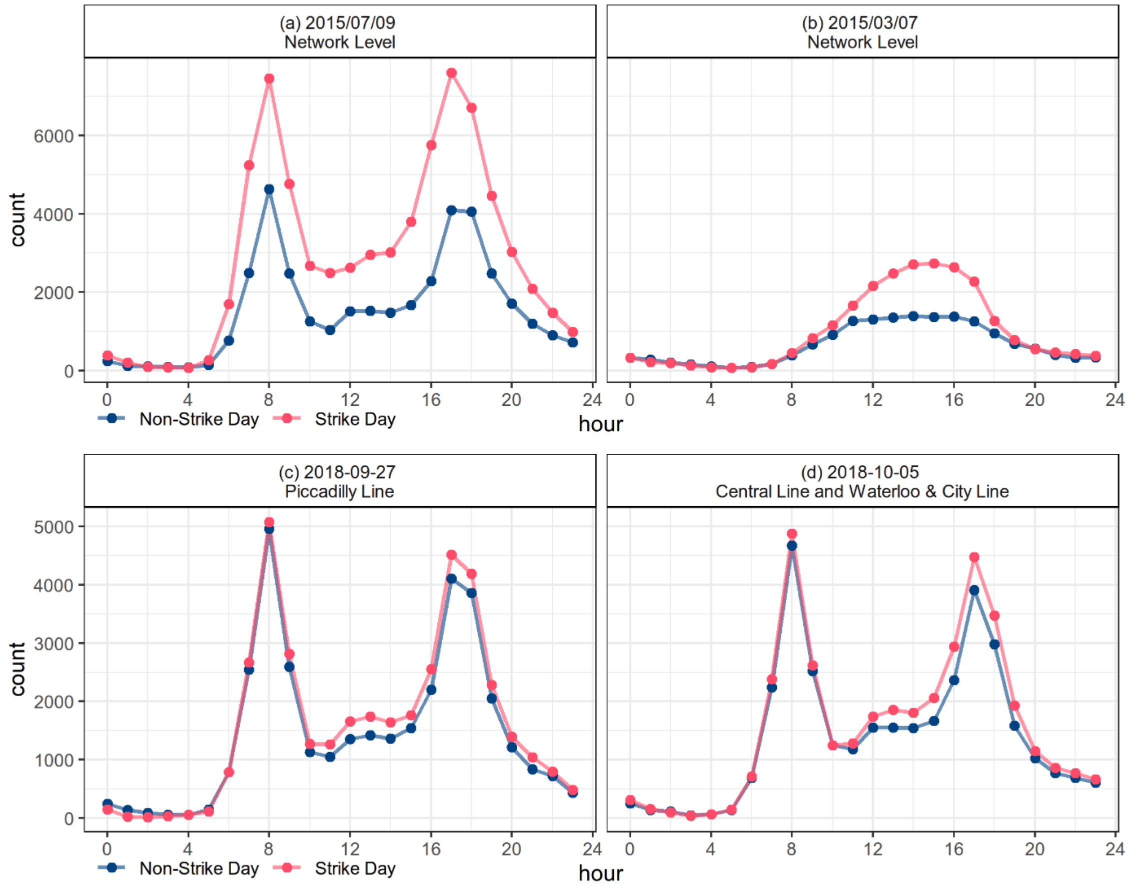


Figure 4.2 Hourly LCHS bike use on strike days and non-strike days; (a) Strike 1; (b) Strike 2; (c) Strike 3; (d) Strike 4.

The hourly use changes are also associated with the spatial scale of the incident. When line-level strikes occurred (figure 4.2 c, d), the increases are less pronounced. Total use increased by 9.9% during Strike 3, and for Strike 4 it is 12.2%. The smaller increases compared to network-level strikes is to be expected and suggests that additional increases in bike-sharing usage are likely to be constrained to parts of the city that fall within the disrupted line's service catchment. A more subtle but interesting pattern can also be observed: trip volumes (figure 4.2 c, d) increase from noon to 8 pm, while there is only a marginal impact on trip volumes in the morning (6-10 am). This may be due to the fact that there is generally less flexibility when travelling in the morning peak. The observational data shows that the morning peak is consistently narrower than the afternoon peak (figure 4.2 a, c, d). This has particular implications for the LCHS, where usage is heavily constrained by the number of bikes and docking spaces available. Intense competition for bikes and spaces at peak times means that the scheme can support only a limited number of additional trips - and this is exacerbated when disruption events such as Tube strikes contribute additional demand.

Overall, the different changes in patterns of hourly use indicate that network-level and line-level strikes have varied temporal impacts on people's modal shifting to bike-sharing. While network-level disruptions increase trip frequencies across most time slots, line-level strikes impact noon and afternoon periods more heavily. Weekend cycling has very different trends compared to weekdays, but increases are observed particularly in the afternoon of the strike days.

4.4.1.2 Spatial Patterns

The spatial pattern of LCHS ridership also varies during different Tube strikes, as illustrated in figure 4.3. Trip volume significantly increased in central London during Strike 1, with areas surrounding train stations experiencing the largest increases, for example, King's Cross, Paddington and Liverpool Street train station (marked as 1,2,3 in figure 4.3 a). Piccadilly Circus also shows significant increases (marked as 4). In contrast, during a weekend strike, areas with the highest increases are around Hyde park (marked as 5 in figure 4.3 b). These varied patterns suggest that the spatial distribution of increases in bike use is severely impacted by whether the strike occurred during a weekday or weekend.

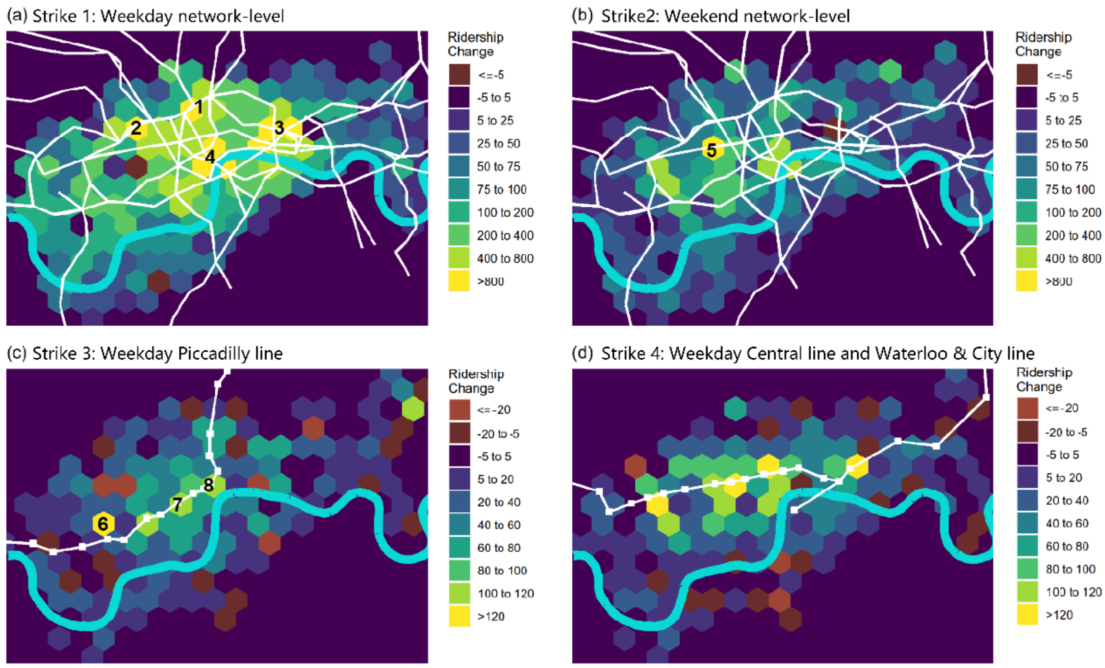


Figure 4.3 Ridership change (number of trips) in hexagonal grid, with white lines indicating the Tube routes that are disrupted, white lines are disrupted Tube lines; (a) Strike 1; (b)Strike 2; (c) Strike 3; (d) Strike 4.

For line-level strikes (figure 4.3 c, d), the impact is more localised, with regions closer to disrupted lines experiencing increases in use. In figure 4.3 (c), areas closest to south Kensington (marked as 6), where numerous museums are located, show the greatest increase in bike use. Green park (7) and Piccadilly Circus (8) also show marked increases. During Strike 4, most regions along Central line experienced higher LCHS trip counts.

The relationship between individual dock's distance to Tube and bike use changes are further examined by boxplots in figure 4.4. The x-axis indicates the distance to the nearest disrupted Tube stations, and the y-axis shows changes in trip frequencies at those docking stations (Saberi et al., 2018). A distance decay pattern can be observed during weekday disruptions (e.g. figure 4.4 a, d), with stations closer to Tube stations experiencing a greater increase in use, and these increases reducing with distance to affected Tube stations. This is consistent with the findings of Saberi et al. (2018). However, for weekend strikes (figure 4.4 b), the pattern is not replicated, rather increases are observed at greater distances with the peak at 1000-1250 m from the disrupted Tube stations. The differing patterns of distance decay may not only be driven by the density of docking stations, but also cycling purposes. On weekdays, bike use is more heavily associated with utility cycling (commuting), whilst weekend trips are more likely to be discretionary and made for leisure purposes. This inference can be further supported by comparing line-level strikes and associated trips. During Strike 4 (figure 4.4 d), the distance decay trend is more similar to Strike 1 (figure 4.4 a); whilst in Strike 3, no obvious distance decay exists. Tube lines affected by Strike 4 run through many employment-related locations and are associated with many cycling trips for commuting purposes. By contrast, The Piccadilly line, affected in Strike 3, covers comparatively more tourist attractions than does Central and Waterloo & City (Strike 4). So utilitarian commuting trips are less displaced in Strike 3, as shown by a lack of distance decay in figure 4.4 (c). A further interesting pattern is observed in line-level strikes (figure 4 c, d). There is an increase in trip frequencies for the most distant group (2000 m). This may be observed due to displaced Tube travellers switching to other travel options (e.g. overland local train services, buses, etc.) combined with bike travel for their journey, thereby contributing to more bike trips in typically less-heavily used parts of the LCHS network (Larcom et al., 2017). This phenomenon and related route changing behaviour is further analysed in later sections (section 4.4.3 and 4.4.4).

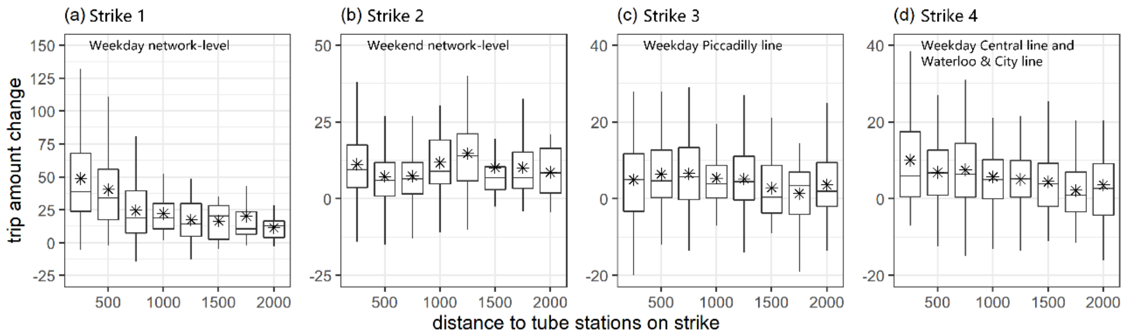


Figure 4.4 Box plots of bike use changes with distance to the nearest Tube stations affected by disruption; (a) Strike 1; (b) Strike 2; (c) Strike 3; (4) Strike 4.

4.4.2 Changes in docking station availability

The increased trip volumes during strikes pose challenges for service provision. Time-series of “service pressure” measures during network-level strikes are shown in figure 4.5. This shows the frequency with which stations suffer from “low availability” throughout the day.

For weekend strikes, the number of stations recorded as under pressure doubles during the weekday morning peak (8 to 10 am). Increased competition for bikes creates availability problems for the scheme as a whole. Service pressure at weekends also increases under strike events but in a different way to the weekday strikes, with a peak observed around 16:00 (figure 4.5 b) during the disruption.

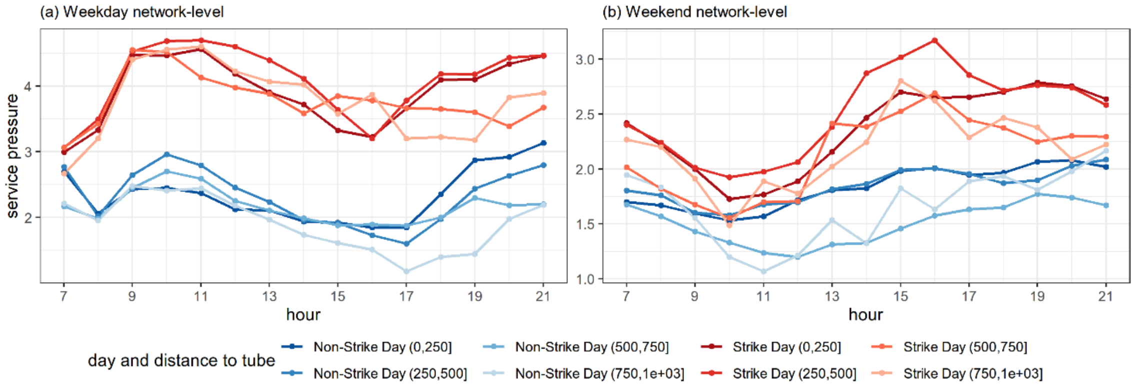


Figure 4.5 Time-series of service pressure during network-level Tube strikes, blue lines indicate non-strike day, red lines represent strike day; (a) Strike 1, weekday ; (b) Strike 2, weekend.

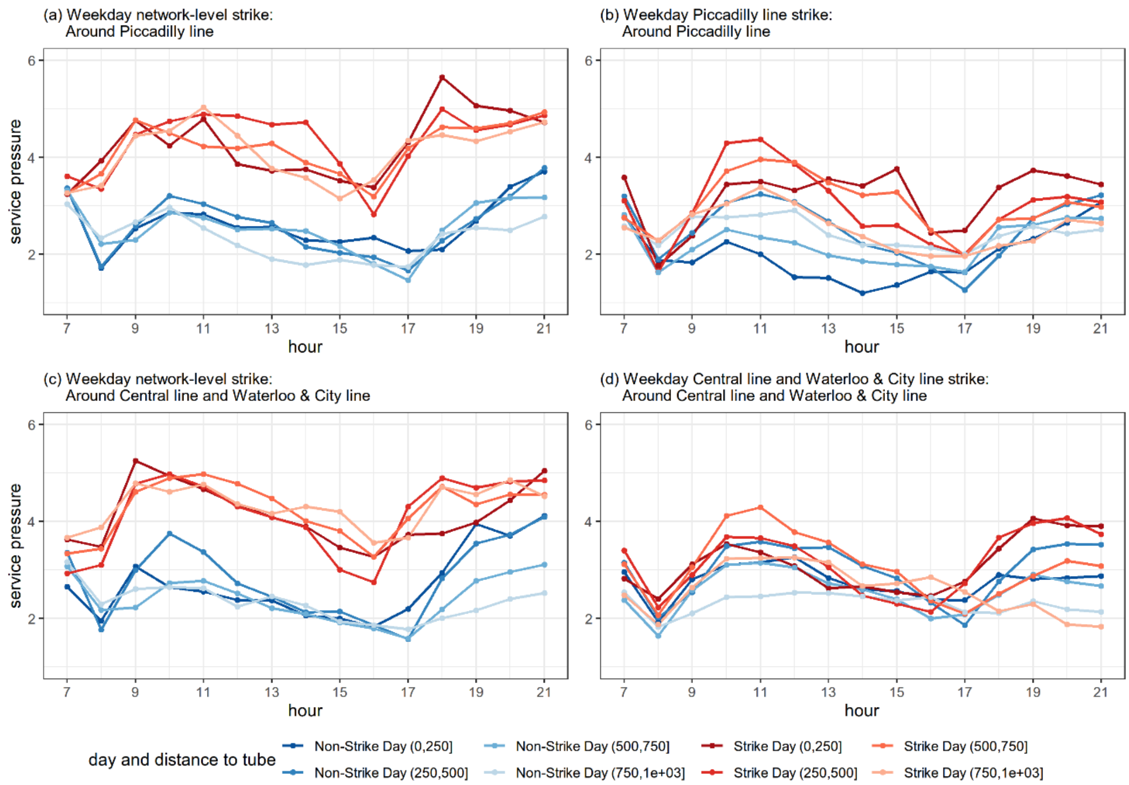


Figure 4.6 Time-series of service pressure in service catchment of specific Tube lines; (a) Piccadilly line - Strike 1 (b) Piccadilly line - Strike 3; (c) Central line and Waterloo & City line - Strike 1; (d) Central line and Waterloo & City line -Strike 4.

Line-level strikes lead to more localised trip increases, thereby bringing generally more service pressures to LCHS. Figure 4.6 shows service pressure around specific Tube lines in the event of network- and line-level Tube strikes on weekdays. When line-level strikes occur (figure 4.6 b, d), all distance groups generally show higher service pressure in the morning, but for evening periods (e.g. around 6 pm), bike stations further from the affected Tube lines (750-1000 m) do not exhibit such higher pressure.

The results from the service pressure analysis - characterising where, when and to what extent docking stations become unusable on strike days - could inform future targeting of rebalancing strategies. It should also be noted that disruption events may counter-intuitively lead to patterns of usage that are beneficial to fleet management. There is evidence that usage under strike conditions can become more heterogeneous. Analysing the dynamics in LCHS usage in greater spatial detail may therefore be instructive, and in the following section we present such an analysis, considering full OD flow data.

4.4.3 Graph statistics

Different graph statistics are derived from the full LCHS OD trip data, providing various insight into changes in spatial travel behaviours under disruption. Table 4.2 presents these statistics from network-level strikes along with what would be expected under normal conditions (a control group). Disruptions lead to higher graph connectivity (δ), which indicates more heterogeneous cycling behaviours - new OD pairs. The number of new OD pairs (L) increased by 80% for the weekday and 38% for weekend strike events. On one hand, this greater diversity in cycling behaviour may indicate that a wider and new set of users are attracted to the scheme during strike days. On the other hand, it is perhaps a function of parts of the scheme under service pressure becoming partially unavailable at peak times under strike conditions (established in section 4.4.2), and so existing users must find alternative routes; new OD pairs are introduced by bike-sharing cyclists forced to use second- or third- preference docking stations due to the additional competition for bikes/spaces during strikes.

Differences in centrality scores are also observed under strike events. Average node degree, $\bar{d}$, is larger under both events, implying that docking stations are linked to a larger set of other stations. But the coefficient of variation in node degree, *cv(d)*, shows a contrasting pattern. Whilst the indicator remained unchanged for the weekend strike, *cv(d)* increased from 0.62 to 0.68 for the weekday strike. The reasons are twofold: (1) during weekday disruptions, where commuting and utility cycling tends to be the dominant trip purpose, demand is concentrated at particular space-times. Because of increased service pressure, some bike-sharing cyclists are required to use alternative docking stations either to pick-up or drop-off bikes in intense demand regions. This will link many more nodes to one vertice; (2) and more importantly, due to the dominant commuting behaviours, parts of the bike-sharing network connecting key employment centres and transit hubs (e.g. King's Cross) experience disproportionally more trip frequencies and therefore serve a more diverse set of stations than other parts of the city that are less strategically important. During weekends, however, trip purposes are typically discretionary. There may be an overall increase in LCHS usage, but this is not so spatially concentrated, so the *cv(d)* remains unchanged.

Table 4.2 Graph statistics of network-level strikes.

N represent the number of nodes; L is the number of links; $\bar{d}$ is the average node degree, cv(d) is the coefficient of variation of node degree; $\bar{w}$ is the mean link weight; $\delta = 2L/N^2$ is defined as network connectivity; T is the sum of all link weights; a is the graph assortativity; t is the transitivity.

Graph Property	Strike 1 (weekday)	Strike 1 control group	Strike 2 (weekend)	Strike 2 control group
N	735	739	742	740
L	50,455	28,027	16,474	11,973
N/L	68.6	38	22.2	16.2
$\bar{d}$	137.3	75.9	44.4	32.3
cv(d)	0.68	0.62	0.63	0.63
$\bar{w}$	1.38	1.32	1.47	1.32
cv(w)	0.80	0.98	1.27	0.84
δ	0.187	0.103	0.06	0.044
T	69,734	37,070	24,160	15,910
a	0.001	0.098	0.071	0.069
t	0.31	0.23	0.17	0.14

Average link weight, $\bar{w}$ increases as a result of more trips being made, but the coefficient of variation, *cv(w)*, shows an interesting opposite pattern. The value of *cv(w)* decreased under the weekday strike while it increased under the weekend strike. Combining this pattern with the larger $\bar{w}$ in Strike 2 than Strike

1, a speculative explanation can be proposed: many of the new flows (OD pairs) occurring under weekday disruptions are of comparatively lesser weight than those occurring during the weekend strike event. It is likely that the new weekday pairs are between alternative or less-popular docking stations - made by commuting cyclists who are not able to complete their preferred trip (OD pair) due to the additional and highly concentrated demand. Further evidence to support this assumption is in the graph assortivity scores (*a*). For strike 1, assortivity shows an exponential reduction from 0.098 to 0.001. In contrast, under Strike 2, assortivity slightly increased (by 0.002). This difference indicates that under weekday disruptions it is more common to observe bike trips between an important, heavily used hub docking station and another, previously underused docking station; and by extension that weekday strikes induce new trip combinations (OD pairs) and possibly new users. In contrast, assortivity increased slightly under weekend strikes, suggesting that these more discretionary trips tend to be made typically between similarly important (popular) docking stations.

Overall, both Strike 1 and 2 have contributed to a denser cycling network, with a larger number of travel OD pairs and increased average link weights - a function of greater LCHS use during strike events. However, due to different travel purposes and docking station availability, the weekday strike (Strike 1) makes the graph structure more heterogeneous in terms of node centrality (degree), while the weekend strike (Strike 2) does not.

Further analysis on node centrality is shown in figure 4.7, and illustrates the CDF (Cumulative Distribution Function) of degree and eigenvector centrality. A high node degree indicates a node is connected to many other nodes, while a high eigenvector score means that a node is connected to many other nodes, which are also high in centrality (Oldham et al., 2019). All CDFs in figure 4.7 follow a power law. Figure 4.7 (a,b) suggests there is a higher probability of observing a node (a docking station) with a larger degree (connects to many other stations) during strikes on both weekdays and weekends; this accords with the patterns of $\bar{d}$ in table 4.2. Interestingly, figure 4.7 (c,d) shows a diverging pattern in eigenvector centrality: in figure 4.7 (c) there is a large difference between non-strike and strike days, while in figure 4.7 (d), no such difference exists. This further reinforces the patterns previously identified. When a low degree node links to a high degree node - a frequently used docking station connects to a less frequently used docking station - the eigenvector score of the low node (the less frequently used docking station) will increase. This kind of situation is more common in Strike 1, as indicated by the assortivity changes in table 4.2. During a weekend strike, however, docking stations are more likely to link to other docking stations with similar centrality. Therefore, many low-scoring nodes do not link and benefit from other high centrality nodes – during weekend strike events, there is a 'doubling-down' on existing travel behaviours with increased trips between already frequently used docking stations, presenting a clearer "rich-club" effect (Zhou et al., 2004).

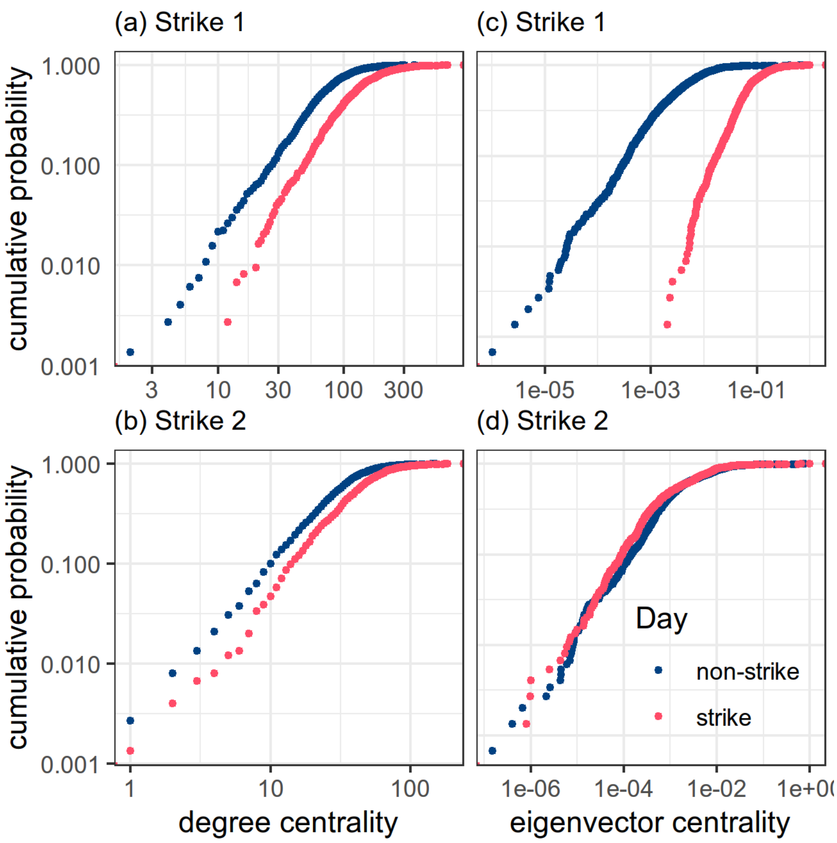


Figure 4.7 Node centrality change in network-level Tube strikes; (a) Strike 1: degree; (b) Strike 2: degree; (c) Strike 1: eigenvector centrality; (d) Strike 2: eigenvector centrality.

The above analysis on the cumulative probability of the graph properties are further examined by Kolmogorov–Smirnov test, and the results are shown in table 4.3. All p-values are smaller than 0.05, and confirmed that all the changes are statistically significant. The d-value of Eigenvector centrality (0.8045) is larger than the degree centrality (0.3694) during Strike 1; while contrasting results are shown in Strike 2, as eigenvector centrality (0.1014) is smaller than degree centrality (0.2272).

Table 4.3 Kolmogorov–Smirnov test results.

	graph property	d-value	p-value
Strike 1	Degree Centrality	0.3694	<0.0001
	Eigenvector Centrality	0.8045	<0.0001
Strike 2	Degree Centrality	0.2272	<0.0001
	Eigenvector Centrality	0.1014	<0.0001

Strike 3 and 4 are line-level events, which also happened on weekdays. Therefore, the changed graph statistics (table 4.4) share many similar trends with Strike 1.

Table 4.4 Graph statistics of line-level strikes on weekday.

Graph Property	Strike 3 (Weekday)	Strike 3 Control group	Strike 4 (Weekday)	Strike 4 Control group
N	784	785	785	784
L	29569	27207	28588	25932
N/L	37.7	34.7	36.4	33.1
$\bar{d}$	75.4	69.3	72.8	66.1
cv(d)	0.64	0.65	0.64	0.64
$\bar{w}$	1.28	1.27	1.3	1.28
cv(w)	0.74	0.64	0.68	0.65
δ	0.096	0.088	0.093	0.084
T	38321	34883	37648	33562
a	0.092	0.096	0.09	0.098
t	0.22	0.21	0.21	0.2

Higher $\bar{d}$ and $\bar{w}$ have been observed in Strike 1, 3 and 4. But a different pattern can be identified in *cv(w)*. While strike 1 (network-level) shows a lower value, decreasing from 0.98 to 0.8 (table 4.2), strike 3 and 4 both show increases in this index (table 4.4), and is due to the effect of changing Tube routes. For example, in the evening peak, when all Tube lines are unavailable (Strike 1), commuters may cycle a long distance directly from their workplace to home/train station. Due to the large variance in work-home/train station location pairs, the bike travel flows will present higher randomness to connect different docking stations. Hence, higher homogeneous link weights may be observed, and lead to lower *cv(w)*. When only certain Tube lines are affected (Strike 3,4), people may change their Tube route, and combine it with a shorter bike-sharing journey to their destination. In this case, the cycling origin/destination randomness is lower than Strike 1 because many bike trips are aggregated at (made from/to) certain docking stations close to unaffected Tube lines.

To obtain a more comprehensive understanding, and to confirm the changed travel behaviours as speculated here as well as in section 4.4.1.2, the next part provides a supplementary interpretation of flow maps variations.

4.4.4 Origin-Destination visualizations

Whilst the graph statistics provide useful aggregate-level summaries, visual analysis of the OD flow data allows us to characterise with greater richness the nature of changes in response to the strike events.

In figure 4.8, each cycled OD pair is represented as a Bézier curve, using asymmetric curves to indicate journey direction, with a straight end representing journey origin and a curved end indicating journey destination (Beecham et al., 2014). Flows are encoded according to the increases in flow frequencies recorded during network-level strikes (Strike 1 and 2) using colour value (lightness) and transparency. During Strike 1 (figure 4.8), increased trip frequencies are distributed across central London and the City of London. London’s large transit hubs can also be identified here. For example, increased trips from King’s Cross, Liverpool Street, Paddington and Waterloo. Notice that the increases recorded here are particularly in trips originating from these stations.

For Strike 2 (weekend), a different pattern is identified. There is a sense that bike-sharing users “double down” on their typical weekend travel OD - increases are observed for apparently leisure trips made within Hyde Park and West London. There is also a stronger bi-directional link between the southeast corner of Hyde Park and Piccadilly Circus, which are close to major shopping and entertainment areas in London’s West End. Higher trip counts between Hyde Park and the Regent’s park can also be identified. One of the most substantial increases in trip frequencies runs through the northeast and connects Victoria Park and the Broadway Market, and there are some comparatively long distance journeys from Liverpool street stations to Millwall Park. Overall, the increased OD pairs typically connect one leisure or shopping spot to another.

Figure 4.9 illustrates weekday line-level Tube strikes, and some differences to network-level disruptions (figure 4.8 a) can be found. Many major links in figure 4.8 (a) disappeared in figure 4.9. For example, in the case of train transfer, the stronger flow from Paddington to King's Cross station (figure 4.8 a) is not visible in figure 4.9 (a); instead, many more trip counts are observed connecting King's cross and central London (figure 4.9 a), but also trips starting and ending close to Tube line stations in central London and the City of London (e.g. Bakerloo line and Central line). This phenomena strongly implies the change of Tube routes among travellers. When all Tube services are unavailable, people may cycle from Paddington to King's Cross to transfer trains, while if only Piccadilly line is in disruption, transit users will firstly travel via other Tube lines to stations that are close to King's Cross, then combine a bike trip to reach their destination (King's Cross). This highlights the role of cycle hire as a flexible travel mode, and its advantages in providing a service under different conditions for travellers to complete their journey, and strengthen urban and mass transit resilience.

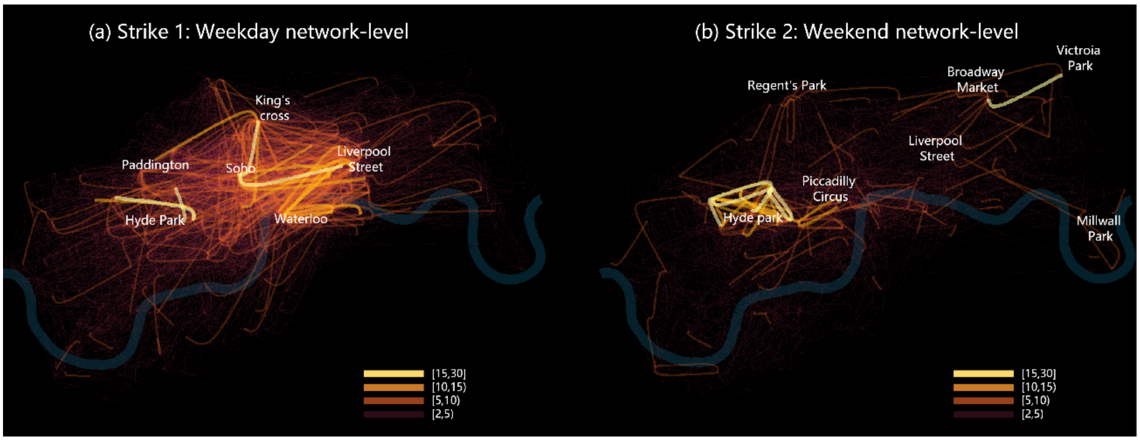


Figure 4.8 Increased OD flows during network-level Tube strike; (a) Strike 1; (b) Strike 2.

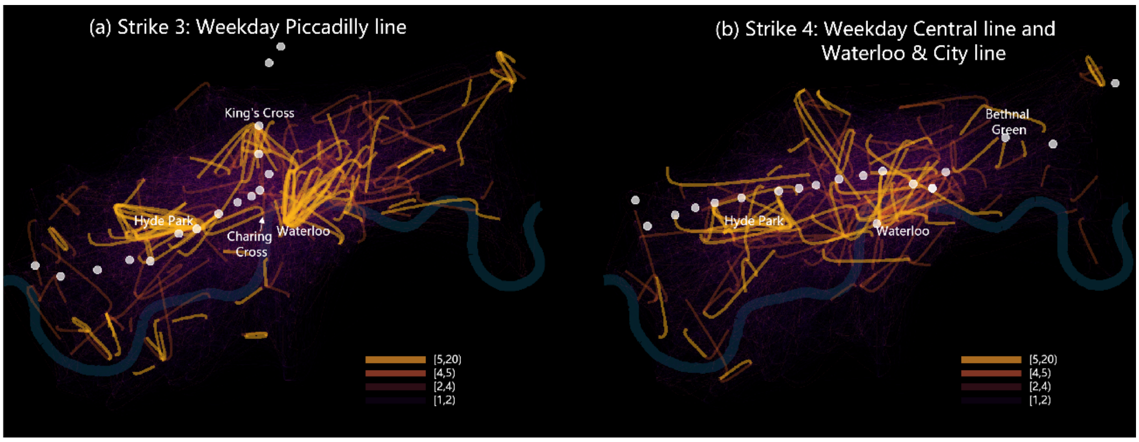


Figure 4.9 Increased OD flows during line-level Tube strike; (a) Strike 3; (b) Strike 4.

4.5 Discussion

Several findings can be abstracted from this analysis. Firstly, the effect of public transit disruptions such as Tube strikes on bike-sharing usage clearly varies depending on the nature of the public transit disruption. A network-level strike will increase overall bike trip frequencies throughout the day, whilst line-level disruption leads to higher usage mainly in the afternoons. There is also an expected distance decay effect to these increased trip frequencies, with docking stations closer to the disrupted Tube lines experiencing the largest increases in trip frequencies during the strike events. This is consistent with other observational studies (Saberi et al., 2018), but we add that this distance decay effect on individual docking stations is much stronger where the tube strike events occur in parts of the city that typically serve commuting journeys. The distance decay effect on individual station is much reduced, or even disappears entirely, for parts of the city or time periods are associated with leisure activities.

Secondly, the consequences of increased bike-sharing usage on the LCHS's usability is quantified by creating a novel service pressure index. Tube strikes are found to increase service pressure and the likelihood of docking stations either containing insufficient available bikes or docks. The consequences are most severe for weekday network-level strikes – the number of observed instances of docking stations under pressure generally doubles when compared to the non-strike control. For line-level strikes, the patterns of increased service pressure vary depending on the proximity of a docking station to the disrupted Tube stations.

Thirdly, graph-based analysis has identified several trends. The results indicate that Tube strikes can lead to a denser cycling network with higher numbers of trips and graph links. But weekday strikes tend to link nodes (bike docking stations) with varied centrality scores, and this is opposite to weekend disruption events when slightly more nodes with similar centrality are connected. These opposing patterns relate not only to differences in travel purpose (commute versus leisure), but also differences in observed service pressure during weekday and weekend disruption events. Moreover, visually representing changes in trip patterns via flow maps is instructive, with increased bike-sharing trips connecting central London with major rail terminals, suggesting that bike-sharing is being substituted for (commuting) journeys that would otherwise be taken by Tube. These all imply the importance and potential of bike-sharing as a flexible travel mode that might therefore strengthen urban transit resilience.

Whilst the LCHS is demonstrated to provide an alternative travel option when other parts of the transit network are under disruption, the scheme soon reaches capacity, especially at peak times. Efforts to maintain a functioning bike-sharing network under disruption events may have important social benefits - providing a cheap and accessible travel alternative, especially to lower-income groups whose employment may be more precarious, shift-based and less flexible (Green et al., 2012; Zhu et al., 2017; Younes et al., 2019). The maintenance and bike fleet rebalancing work should take the impacted behaviours of travellers into consideration. For example, when line-level disruptions happen, it is also important to provide more cycles around some unaffected Tube lines/stations, because they will experience higher demand due to travellers' changing Tube routes. Pop-up cycle lines may be set up between different hotspots locations as identified in this work. It helps to enlarge the space to meet the temporary increased cycling activities, provide a safer environment and eliminate potential congestion on cycle lanes.

Disruptions and reduced capacity in other public transport will lead to higher usage of bike-sharing, and these might be developed as new travel habits for people. The social distancing guidance may lead to much higher LCHS usage, combining with the increasing popularity of sustainable travel mode, more investment into LCHS, and improved service management and station capacity are required to meet the potential demand. This also helps to promote and facilitate greener and healthier travels of people.

A shortcoming of this work is the lack of user's socioeconomic information. Some studies (Green et al., 2012; Zhu et al., 2017) have suggested that low-income groups may benefit more from bike-sharing during mass transit failure, also more occasional users have seen using the sharing bikes. Therefore, future research will examine the characteristics of different groups of users and compare their modal shift behaviours. Surveys will also be used to confirm the trip purpose during varied strike events, to supplement understanding their attitude, preferences on using bikes as an alternative travel mode. Besides, the assumption of new users may be further confirmed by analysing the residual effect of the assortivity scores as its decreasing value during weekday strikes implies new OD trips are probably made by new scheme users. To validate this, further examination of whether assortivity scores stay at comparatively lower levels immediately post-strike (residual effect) will allow more insights to be obtained, and shed light on how transit disruptions attract new users and trips to the scheme.

4.6 Conclusion

This work combined spatiotemporal analysis and graph-based approaches to explore the changing behaviour of bike-sharing users in the LCHS as a result of four Tube strikes. Changes in user cycling activities and bike-sharing usability (the availability of bikes and docking points) were quantified. This study demonstrates that there is a distinct geography to affected travel behaviours and ridership that are consistent with previous studies (Saberi et al., 2018), with the docking station-level pattern is conditional on whether the disrupted parts of the city and the time periods relate to commuting activities. Systematic changes in travel behaviour were found by examining changes in OD flow graph structures. The observed variation in bike-sharing usage under disruption events demonstrates the flexibility of cycle hire schemes and their contribution to enhanced understandings of urban transit resilience. The findings of this work and the methods used provide useful information and tools for scheme operators to better manage system resources and to support cycle infrastructure policy-makers and planners in designing interventions aimed at incentivising cycling.

References

Bauernschuster, S., Hener, T. and Rainer, H. 2017. When labor disputes bring cities to a standstill: The impact of public transit strikes on traffic, accidents, air pollution, and health. *American Economic Journal: Economic Policy.* **9**(1), pp.1-37.

Beecham, R. 2015. Using Bikeshare Datasets to Improve Urban Cycling Experience and Research Urban Cycling Behaviour. In: Gerike, R. and Parkin, J. eds. *Cycling Futures: From Research into Practice.* Farnham: Ashgate, pp. 267-283.

Beecham, R., Wood, J. and Bowerman, A. 2014. Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems.* **47**, pp.5-15.

Bonacich, P. 2007. Some unique properties of eigenvector centrality. *Social Networks.* **29**(4), pp.555-564.

Cairns, S., Atkins, S., and Goodwin, P. 2002. Disappearing traffic? The story so far. *Proceedings of the Institution of Civil Engineers-Municipal Engineer.* **151**(1), pp.13-22.

Chen, L., Zhang, D., Wang, L., Yang, D., Ma, X., Li, S., Wu, Z., Pan, G., Nguyen, T.-M.-T. and Jakubowicz, J. 2016. Dynamic cluster-based over-demand prediction in bike sharing systems. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing,*12-16 September 2016, Heidelberg. New York: Association for Computing Machinery, pp.841-852.

Curran, A. 2008. *Translink public bike system feasibility study.* Vancouver: Quay Communications.

De Chardon, C.M., Caruso, G. and Thomas, I. 2016. Bike-share rebalancing strategies, patterns, and purpose. *Journal of transport geography.* **55**, pp.22-39.

Dill, J. and Carr, T. 2003. Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them. *Transportation Research Record.* **1828**(1), pp.116-123.

Duncan, R. 2015. *Developer guidance for Santander cycles*. [Online] *London: Transport for London.* [Accessed 1 December 2020]. Available from: http://content.tfl.gov.uk/developer-guidance-for-santander-cycles.pdf

El-Assi, W., Mahmoud, M.S. and Habib, K.N. 2017. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation.* **44**(3), pp.589-613.

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M. and Haq, U. 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography.* **41**, pp.306-314.

Fishman, E. 2016. Bikeshare: A review of recent literature. *Transport Reviews.* **36**(1), pp.92-113.

Fukasawa, N., Yamauchi, K., Murakoshi, A., Fujinami, K. and Tatsui, D. 2012. Provision of forecast train information and consequential impact on decision making for train-choice. *Quarterly Report of RTRT.* **53**(3), pp.141-147.

Gonçalves, L.A.P.J. and Ribeiro, P.J.G. 2020. Resilience of urban transportation systems. Concept, characteristics, and methods. *Journal of Transport Geography.* **85**, p102727.

Green, J., Steinbach, R. and Datta, J. 2012. The travelling citizen: emergent discourses of moral mobility in a study of cycling in London. *Sociology.* **46**(2), pp.272-289.

Larcom, S., Rauch, F. and Willems, T. 2017. The benefits of forced experimentation: striking evidence from the London underground network. *The Quarterly Journal of Economics.* **132**(4), pp.2019-2055.

Lovelace, R., Beecham, R., Heinen, E., Vidal Tortosa, E., Yang, Y., Slade, C. and Roberts, A. 2020. Is the London Cycle Hire Scheme becoming more inclusive? An evaluation of the shifting spatial distribution of uptake based on 70 million trips. *Transportation Research Part A: Policy and Practice.* **140**, pp.1-15.

Nello-Deakin, S. 2020. Environmental determinants of cycling: not seeing the forest for the trees? *Journal of Transport Geography.* **85**, p102704.

Newman, M.E. 2002. Assortative mixing in networks. *Physical Review Letters.* **89**(20), p208701.

Noldus, R. and Van Mieghem, P. 2015. Assortativity in complex networks. *Journal of Complex Networks.* **3**(4), pp.507-542.

Oldham, S., Fulcher, B., Parkes, L., Arnatkevičiūtė, A., Suo, C. and Fornito, A. 2019. Consistency and differences between centrality measures across distinct classes of networks. *PloS One.* **14**(7), pe0220061.

Padgham, M. and Ellison, R. 2017. bikedata. *Journal of Open Source Software.* **2**(20), p471.

Pan, L., Cai, Q., Fang, Z., Tang, P. and Huang, L. 2019. A deep reinforcement learning framework for rebalancing dockless bike sharing systems. In: *Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 27 January – 1 February 2019, Honolulu. Vancouver: PKP, pp.1393-1400.

Pregnolato, M., Ford, A., Wilkinson, S.M. and Dawson, R.J. 2017. The impact of flooding on road transport: A depth-disruption function. *Transportation Research Part D:Ttransport and Environment.* **55**, pp.67-81.

Quinn, B. 2020. London cycling could increase tenfold after lockdown, says TfL. *The Guardian.* [Online]. 6 May. [Accessed 1 December 2020]. Available from: https://www.theguardian.com/world/2020/may/06/cycling-could-increase-five-fold-in-london-after-lockdown

Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A. and Auld, J. 2020. Analysis of Transit Users' Response Behavior in Case of Unplanned Service Disruptions. *Transportation Research Record.* **2674**(3), pp.258-271.

Saberi, M., Ghamami, M., Gu, Y., Shojaei, M.H.S. and Fishman, E. 2018. Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. *Journal of transport geography.* **66**, pp.154-166.

Shaheen, S., Martin, E. and Cohen, A. 2013. *Public bikesharing and modal shift behavior: a comparative study of early bikesharing systems in North America.* Berkeley: University of California, Berkeley.

Teng, J. and Liu, W.-R. 2015. Development of a Behavior-Based Passenger Flow Assignment Model for Urban Rail Transit in Section Interruption Circumstance. *Urban Rail Transit.* **1**(1), pp.35-46.

Transport for London. 2017. *History of bus and tube strikes.* [Online]. [Accessed September 25]. Available from: https://tfl.gov.uk/corporate/transparency/freedom-of-information/foi-request-detail?referenceId=FOI-1931-1617

Tsuchiya, R., Sugiyama, Y. and Arisawa, R. 2008. A Route Choice Support System for Use During Disrupted Train Operation. In: *Proceedings of 15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting.* 16-20 November 2008. New York. Washington DC: The National Academies of Sciences, Engineering and Medicine, p01135975.

Vertesi, J. 2008. Mind the gap: The London underground map and users' representations of urban space. *Social Studies of Science.* **38**(1), pp.7-33.

Wilson, M.C. 2007. The impact of transportation disruptions on supply chain performance. *Transportation Research Part E: Logistics and Transportation Review.* **43**(4), pp.295-320.

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems.* **77**, p101361.

Younes, H., Nasri, A., Baiocchi, G. and Zhang, L. 2019. How transit service closures influence bikesharing demand; lessons learned from SafeTrack project in Washington, DC metropolitan area. *Journal of Transport Geography.* **76**, pp.83-92.

Yu, D., Yin, J., Wilby, R.L., Lane, S.N., Aerts, J.C.J.H., Lin, N., Liu, M., Yuan, H., Chen, J., Prudhomme, C., Guan, M., Baruch, A., Johnson, C.W.D., Tang, X., Yu, L. and Xu, S. 2020. Disruption of emergency response to vulnerable populations during floods. *Nature Sustainability.* **3**, pp.728-736.

Zhou, S. and Mondragón, R.J. 2004. The rich-club phenomenon in the Internet topology. *IEEE Communications Letters.* **8**(3), pp.180-182.

Zhu, S. and Levinson, D.M. 2010. A Review of Research on Planned and Unplanned Disruptions to Transportation Networks. In: *Proceedings of Transportation Research Board 89th Annual Meeting*, 10-14 January 2010, Washington DC. Washington DC: The National Academies of Sciences, Engineering, and Medicine, p10-2275.

Zhu, S., Masud, H., Xiong, C., Yang, Z., Pan, Y. and Zhang, L. 2017. Travel Behavior Reactions to Transit Service Disruptions: Study of Metro SafeTrack Projects in Washington, DC. *Transportation Research Record.* **2649**(1), pp.79-88.

Chapter 5
Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems.

Abstract

Short-term demand prediction is important for managing transportation infrastructure, particularly in times of disruption, or around new developments. Many bike-sharing schemes face the challenges of managing service provision and bike fleet rebalancing due to the “tidal flows” of travel and use. For them, it is crucial to have precise predictions of travel demand at fine spatiotemporal granularities. Despite recent advances in machine learning approaches (e.g. deep neural networks) and in short-term traffic demand predictions, relatively few studies have examined this issue using a feature engineering approach to inform model selection. This research extracts novel time-lagged variables describing graph structures and flow interactions from real-world bike usage datasets, including graph node Out-strength, In-strength, Out-degree, In-degree and PageRank. These are used as inputs to different machine learning algorithms to predict short-term bike demand. The results of the experiments indicate the graph-based attributes to be more important in demand prediction than more commonly used meteorological information. The results from the different machine learning approaches (XGBoost, MLP, LSTM) improve when time-lagged graph information is included. Deep neural networks were found to be better able to handle the sequences of the time-lagged graph variables than other approaches, resulting in more accurate forecasting. Thus incorporating graph-based features can improve understanding and modelling of demand patterns in urban areas, supporting bike-sharing schemes and promoting sustainable transport. The proposed approach can be extended into many existing models using spatial data and can be readily transferred to other applications for predicting dynamics in mass transit systems. A number of limitations and areas of further work are discussed.

5.1 Introduction

Research has shown that bike-sharing contributes to improved air quality and reduced congestion in cities as a part of a sustainable travel infrastructure (Shaheen et al., 2010; Lovelace et al., 2014). Its global popularity has increased in the last few years due to advantages in both cost and convenience over other forms of transport such as cars. A growing number of cities have operated such schemes to promote sustainable mobility, such as London Cycle Hire Scheme (London), Citi Bikes (New York) and more advanced dock-less systems (e.g. Mobike in Chinese cities). Bike schemes provide a key component of urban transportation infrastructures by providing an “extension service” for the “first/last mile” from other public transport hubs (Shaheen et al., 2010; Ma et al., 2015; Saberi et al., 2018).

While bike-sharing greatly enhances urban mobility as an affordable and sustainable traffic mode (Fishman, 2016), meeting the demand of users poses a challenge to scheme operators. This is due to the “tidal flows” of bike-sharing trips, with certain areas in the city facing the problem of insufficient bikes (Beecham et al., 2014). For example, during the morning rush hour, the number of commuting trips departing from residential areas will be high, potentially leading to a deficit of available bikes in those areas. This results in reduced service reliability and reduced user satisfaction (O’Brien et al., 2014; Fishman, 2016). Accurate and up-to-date estimations of travel demands across the city over the course of the day are crucial for successful bike scheme management and fleet rebalancing. This also has attracted a lot of research interest in recent years.

Researchers have used a combination of statistical models, machine learning and more recently, deep learning neural networks to forecast short-term travel demands (Karlaftis et al., 2011; Vlahogianni et al., 2014; Lin et al., 2018). While some studies have evaluated alternative predictive models for demand forecasting, fewer studies have focused on feature engineering, i.e. the identification and extraction of latent data features that can potentially improve the performance of predictive models (Borges et al., 2017). Recent thinking conceptualises cities as complex systems driven by the pattern of flows and networks of relations (Batty, 2013). One approach to understanding the temporality of urban dynamics and transportation flows is through the analysis of graph structures (Hoang et al., 2016; Yao et al., 2018b; Lin et al., 2018), which has been shown to support insights into different urban and transport problems. However, as yet little research has been undertaken that examines how information from temporal graphs can contribute to better traffic prediction in bike-sharing systems.

This paper evaluates the use of temporal information encoded in graph structures of bike traffic flow interactions for forecasting short-term bike-sharing demand. The experiments in this work retained initial model hyper-parameters to demonstrate the utility of the graph derived features. Section 5.2 introduces related work and reviews different models and inputs used for predicting bike travel demand. Section 5.3 presents the data and the concept of graph-based measures in the transportation network. It compares graph-based features to other commonly used variables such as meteorological data, in terms of importance, explanatory power and their potential contribution to demand prediction models. Section 5.4 presents and compares the results in detail and describes the relative benefits of including graph-based features for improved forecasting. The features, methods and their applicability to other application domains related to transport and flow predictions are discussed, and conclusions are drawn (Section 5.5).

5.2 Related works

There are two conventional approaches for dealing with dock-based bike-sharing travel demand forecasting problems: predicting at an individual station level or over aggregated groups/areas. The former approach models dynamics at each station (Lin et al., 2018), while the latter focuses on regional dynamics (Xu et al., 2013; Zhou et al., 2019b). Station level modelling can support bike-fleet management at finer spatial granularities, but can be less accurate due to higher levels of noise in the data. Many studies (Li et al., 2015; Zhou et al., 2019b) attempt to predict demand over small geographical areas for the following reasons. Firstly, bike docking stations are dynamic in urban areas over long periods. New stations may be added, with existing stations removed, or relocated. Analysing small clusters of stations allows local travel dynamics to be captured and supports a deeper understanding of these dynamics (Li et al., 2015; Zhou et al., 2019b). Secondly, the emergence and rise of dockless bike-sharing may change the nature of bike-sharing in the future. Dockless schemes allow individuals to borrow and return bikes at any location, rather than at fixed docking stations, this makes it both challenging as well as important to understand travel demand at the small area level (Cao et al., 2019; Yang et al., 2019). Finally, grouping stations into small area-based clusters supports bike fleet management regardless of the scheme type, with sufficient spatial grain to support rebalancing (Li et al., 2015).

A broad range of data-driven models have been proposed to forecast short-term travel demand in bike-sharing systems and other transportation systems such as the metro, buses and taxis (Vlahogianni et al., 2014). These can be categorised into parametric statistical models and nonparametric machine learning (ML) approaches (Zhang et al., 2019). Some examples of the former group include ARIMA (Autoregressive Integrated Moving Average model) and its variants (e.g. ARIMAX, seasonal ARIMA) and Bayesian Networks (Froehlich et al., 2009). Statistical models are easier to interpret but may have lower prediction accuracies when compared to ML models. Karlaftis and Vlahogianni (2011) observed a trend of research moving from statistical models to ML models as a result of both increased data accessibility and computing power.

Different ML models have been applied to forecast short-term traffic demand, such as support vector regression (Xu et al., 2013) and Regression Trees (Li et al., 2015). More recently, deep neural networks have attracted significant research interest due to their automatic feature extraction capacity and their success in handling temporal, spatial and semantic dependencies.

Temporal dependencies include snapshots of historical relationships, and have been widely used for traffic demand prediction problems (Froehlich et al., 2009; Giot et al., 2014; Li et al., 2018). For example, useful travel demand information is retained from the last few hours to suggest demand intensity trends. Deep neural networks such as Recurrent Neural Networks (RNNs) provide powerful tools for dealing with sequential information, and are suitable for analysing temporal dependencies. These recurrently connect hidden layers with different timestamps, identifying sequential characteristics and patterns that are then used to predict the next likely scenario. LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) Networks, both enhanced forms of RNNs, have been used to predict travel demand (Fu et al., 2016; Xu et al., 2018). They are able to overcome the “vanishing gradients” problem common in Neural Networks. This occurs when gradients of the loss function approach zero, making the neural network hard to train, which commonly happens when processing long-term temporal dependencies with standard RNNs.

The idea of spatial dependencies (Tobler, 1970) suggests that information from nearby locations can contribute to improved forecasting. Some studies (Ke et al., 2017) have applied Convolutional Neural Networks (CNNs) to capture spatial dependencies in traffic demand forecasting. CNNs were initially designed for the analysis of gridded data, such as images. They capture spatial dependencies between grid locations using localised filters or kernels. Previous research (Ke et al., 2017; Zhang et al., 2017) using this approach to analyse travel demand divided urban areas into two-dimensional grid cells and calculated the demand across each grid, with demand intensity represented as colour scales. However, the selection of grid size is critical and difficult to determine objectively: if the grid is too coarse, it will fail to capture sufficient spatial granularity to support bike fleet management. If it is too fine, then the computational burden increases significantly due to the large image-like matrices containing redundant information (grid cells with zero demand).

More recent studies have used semantic dependence. Semantically similar areas may not be contiguous or near each other. For example, metro stations located in two distant residential areas may have similar temporal patterns of bike travel demand. Characterising semantic dependencies from such similar areas may improve model performance. Some research has quantified the similarity of historical travel demand sequences over different sites and constructed semantic graphs to connect similar places (Hoang et al., 2016; Yao et al., 2018b). Lin et al. (2018) applied Multi-Graph Convolutional Neural Networks (MGCNN) to capture pairwise relations between bike stations, using spatial and semantic graphs to provide multi-graph embedding. However, the pre-processing requirements of capturing demand sequence similarities for MGCNNs are heavy, requiring at least one year's historical data to obtain good prediction accuracy in bike demand forecasting (Chai et al., 2018; Lin et al., 2018). This leads to significant limitations for analyses of sites and systems with insufficient historical travel records, for example, when new service stations or areas are introduced into a bike-sharing scheme.

Outside of the deep neural networks family, XGBoost (Chen et al., 2016b), an implementation of gradient boosted decision/regression trees, has been found to perform well in transport prediction problems and was the winner of the Kaggle bike-sharing prediction competition (Kaggle, 2015). Some research compared XGBoost to neural networks (Lin et al., 2018; Yao et al., 2018b; Ma et al., 2019; Yao et al., 2019; Zhou et al., 2019a). Most of these suggest that XGBoost is capable of obtaining better or similar performances in travel demand forecasting when compared to RNNs (LSTM, GRU), CNNs and to hybrid neural networks, for example, ConvLSTM (combine CNN with LSTM) and ST-ResNet (Deep Spatio-Temporal Residual Network). XGBoost is also found to have comparable performance to MGCNN in the work of Zhou et al. (2019a), in 50% of datasets, XGBoost produced better predictions than MGCNN. However, XGBoost may be inferior to some state-of-the-art fusion deep neural networks, such as Spatio-Temporal U-shape Networks (Zhou et al., 2019a) and Spatial-Temporal Dynamic Networks (Yao et al., 2018a).

Despite the intense competition among complex algorithms, whether one model outperforms the others is questionable. Li et al. (2019) compared various models for traffic demand forecasting, and concluded that the universally best model does not exist. When considering different specific areas and timestamps, several algorithms (e.g. LSTM and XGBoost) may offer better solutions depending on the nature of the spatial and temporal variables. Therefore, it could be beneficial to combine the prediction results of different models (KDD-Cup, 2017). There are also reproducibility issues in the literature; for example, ST-ResNet was found to outperform XGBoost in Chen et al. (2018). However, some studies (Ma et al., 2019; Yao et al., 2019) show contrasting results. Zhou et al. (2019a) found that MGCNN is worse than Xgboost on several datasets, while Lin et al. (2018) suggested a significantly better performance on New York bike-sharing. Differences in results may be due to the complexity of hyperparameter tuning in deep neural networks, varied model performance on different datasets, different preprocessing or unfair comparisons (Karlaftis & Vlahogianni, 2011). These make the “best” models even more of a challenge to identify.

Overall, short-term traffic forecasting is a highly dynamic and developing research arena with ever-growing literature that has mainly focused on testing and comparing the performance of alternative models (Vlahogianni et al., 2014). This focus on models has left other vital questions relatively unaddressed, for example, consideration of what kinds of variables should be included in models. The performance of a predictive model is not only associated with its generalisation ability but also its dependency on the input data and features (Hall et al., 1998). Deep learning neural networks require less effort to manually extract features from raw data (Goodfellow et al., 2016; Lin et al., 2018), but still may benefit from effective feature engineering, especially when the size of the training dataset is limited (Ketkar, 2017). Research has suggested that short-term traffic demand can be inferred from its spatiotemporal properties (e.g. historical travel demand) but may also benefit from other explanatory variables (Ke et al., 2017). However, there is only limited insight into the nature and direction of feature engineering, with studies generally using temporal features (e.g. time of day, day of the week) and meteorological features (e.g. temperature) to forecast travel demand (Giot & Cherrier, 2014; Salaken et al., 2015; Lin et al., 2018). For example, the work of Yang et al. (2016) suggests that average trips amount on weekdays are relatively smaller than during weekends (with the patterns being opposite for stations in residential areas). Both day of week and calendar events (Kim, 2018) are informative for modelling trip demand. Meteorological information, especially temperature, has a huge influence on user behaviours in bike-sharing systems. Good weather is strongly correlated with trip amount (Yang et al., 2016; Kim, 2018). In particular, temperature has been included in many studies and identified as a useful feature for predicting bike trip demand in various cities and regions (e.g. American, Asian, Europe) under different climates and cultural backgrounds (Rudloff et al., 2013; Li et al., 2015; Salaken et al., 2015; Yang et al., 2016). Some studies have used urban context such as land-use, Points of Interest (POI) (Tran et al., 2015; Xu et al., 2018) and event information (e.g. metro delays, concerts) (Chen et al., 2016a; Rodrigues et al., 2019) to improve forecasts. The work of Xu et al. (2018) suggests that land-use information derived from POI is not as helpful as meteorological features, but still can enhance prediction performance for neural network models. However, these are data enrichment approaches, requiring data from other sources (e.g. POI, textual data from Twitter), some of which are relatively difficult to obtain, process and merge into models. This leaves an important question: is it possible to derive additional useful information from the flow data itself, such as bike travel records, to improve the prediction performance further? In machine learning, feature engineering is the process of using domain knowledge to extract and transform raw data into explanatory features. The result is that machine learning algorithms are better able to detect patterns in input data, leading to better outcomes. As yet relatively little research has been undertaken using such approaches in this area to consider what features can be derived from raw travel data using domain knowledge, and whether they can improve different traffic prediction models. Here we examine the graph structures presented in bike travel records.

Research using graph theory has been successfully applied to analyse urban phenomena such as polycentric transformation, urban resilience, infrastructure updates and mobility change, through analysing and understanding urban flows such as travel (Batty, 2013; Zhong et al., 2014; Yang et al., 2019). Graph structures of travel flow spatial and temporal patterns may be used for interpreting urban dynamics as well as for traffic demand prediction (Zhang et al., 2017). Austwick et al. (2013) examined bike-sharing systems in different cities. They highlighted the use of graph analysis for understanding urban flow in spatial systems and whilst Zhang et al. (2017) argued that the historical regional inflows are related to outflows. Generally, studies examining short-term traffic demand forecasting have not fully exploited inflow interactions. The current state of the art in this area uses historical demand and common environmental variables (e.g. temperature) to predict future demand (Xu et al., 2013; Li et al., 2015; Feng et al., 2018; Li & Shuai, 2018; Lin et al., 2018; Yao et al., 2018b; Li & Axhausen, 2019). There are many kinds of graph information (e.g. degree, PageRank) that can be derived from bike travel data to describe flow interactions and to characterise the different urban places within the graph, for example, to infer the likelihoods of bike trips starting from specific regions. The utility of spatiotemporal graph properties to support short-term bike-sharing demand prediction has not been evaluated, and the research described in this paper starts to address this.

5.3 Methods

5.3.1 Study area and data

This study uses dock-based bike-sharing data from two cities to ensure the findings are not exclusive to a specific case. They are New York Citi bike and Chicago Divvy bike schemes, as shown in table 5.1.

Table 5.1 Bike-sharing schemes and data.

Scheme	**time**	**Number of stations**	**Variables**
New York Citi Bike	2016/11/01 – 2017/10-31	785	Departure time, End time, Departure station, End station
Chicago Divvy Bike	2016/10/01- 2017/09/30	569	

The datasets cover one year and contain variables describing bike trip departure and end time, departure station and end station. Corresponding hourly meteorological data were obtained from open weather map (https://openweathermap.org/), and the variables included temperature, humidity, wind speed, pressure and weather description (e.g. Cloudy, light rain).

5.3.2 Data pre-processing

5.3.2.1 Station groups

This study predicts regional (small area) demand and groups stations based on their spatial proximity. A hierarchical clustering method was applied to cluster stations into 120 and 80 groups in the New York and Chicago data, respectively (figure 5.1).

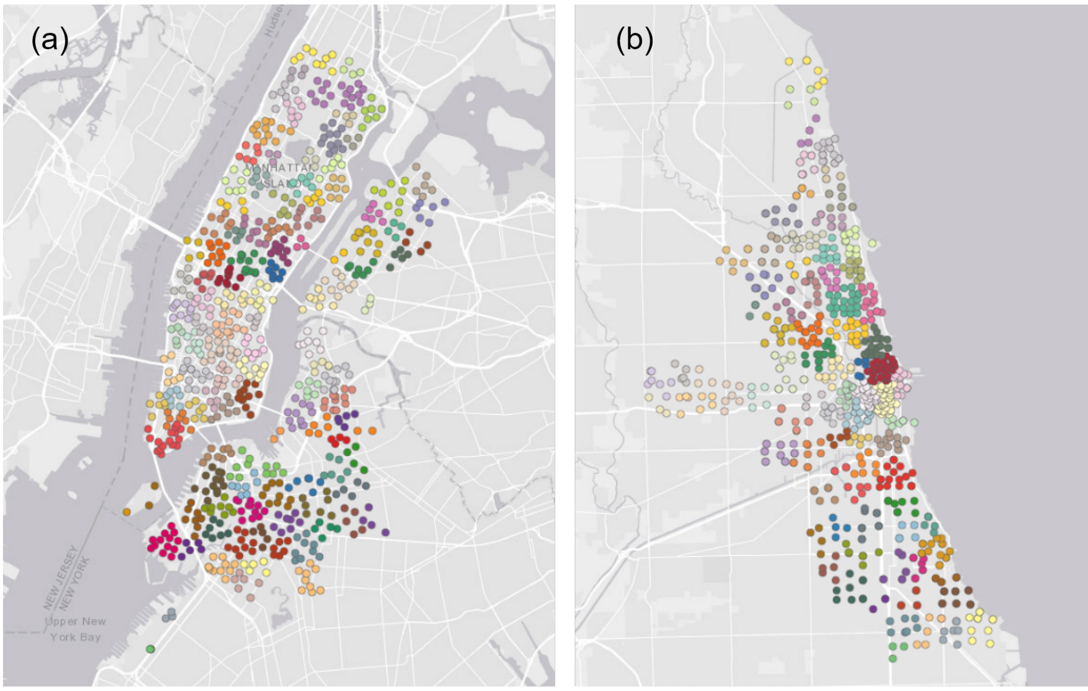


Figure 5.1 Groups of bike-sharing docking stations; (a) New York, (b) Chicago.

Li et al. (2015) analysed the choice of *k* when grouping (clustering) bike-sharing docking stations in New York, and concluded that the cluster number should be chosen by knowledge and real-world experience rather than purely based on optimizing clustering statistics (e.g. silhouette score or within cluster sum of squares), thus to facilitate real-world application of bike fleet rebalancing work. More clusters will generally lead to greater difficulties in prediction because travel amount may fluctuate tremendously at an individual station (Li et al., 2015) or in a small group. While few clusters, for example, when there is only one cluster, the demand is the entire traffic which can be predicted with high accuracy (Li et al., 2015), but may fail to reflect the spatial pattern in fine granularity, so the utility for guiding rebalancing work is reduced. Overall, this study followed the suggestion in the work of Li et al. (2015) and used a hierarchical clustering approach to obtain the groups of stations. The values of *k* were chosen to generate groups consisting of roughly 6 or 7 stations on average (table 5.1 shows total station number), which makes sense in real-world application for fleet management units (areas) as shown in figure 5.1(a, b), where the shading and plot characters indicate different clusters.

5.3.2.2 Travel flow graph structure construction

Graph theory is a mathematical approach for modelling pairwise relations between individuals. A graph structure typically consists of observations represented by nodes or vertices and their relationships represented by links or edges (although this can be reversed). A system formed of nodes and links that are interconnected is termed a graph. In urban and transport studies, public transportation systems have been viewed as complex networks (Zhong et al., 2014; Saberi et al., 2018; Yang et al., 2019) and represented as graphs in order to generate different scale-free graph-based measures pertaining to the network. Generally, transportation hubs and urban areas are regarded as graph nodes, and the travel flows between a pair of nodes generate links to connect them. Analysis of the network flows between nodes and their changes, for example over time, provides insights into spatiotemporal mobility characteristics in transportation systems. Saberi et al. (2018) used graph-based analysis to examine the impact of public transit disruptions on bike-sharing usage.

In this study, hourly graph structures were constructed from bike trip records. Each group of bike stations were cast as nodes, and the volume of hourly bike trips between any two nodes was used to generate edges to represent the origin-destination flows between them. This resulted in a series of temporally weighted and directed graph structures, from which a number of graph properties were calculated, describing the state of each node at different times. Following Zhong et al. (2014); Saberi et al. (2018); Yang et al. (2019), the graph properties were:

1. **Strength** – the total of the edge weights. In a directed and weighted graph structure, there are two strength measures, in-strength and out-strength. Here they represent the number of trips that start from and end at a node in the network. Out-strength can also be interpreted as the number of departures – i.e. travel demand.

2. **Degree** – the number of edges that are incidental to the node, indicating the number of neighbouring nodes. In-degree and out-degree account for the number of in-flow and out-flow links in a directed graph. A node is considered important if it is connected to many

neighbours, and for urban mobility networks, the degree can be used to describe the connectivity and accessibility to destinations or activities across the network (Zhong et al., 2014).

3. PageRank – a measure of node importance. This was first introduced by Google to evaluate the importance of a web page (Brin et al., 1998). The key idea behind PageRank in a graph context is that nodes with the same degree may not have the same importance in a graph. By not counting links from other nodes equally, PageRank treats an edge from a strongly connected node as more important than an edge linked to a node with few connections. Assume graph node *A* has incoming edges coming from a set of other nodes $T_1 \dots T_n$, and the parameter *d* is a damping factor (0.85 as the default value), $C(T_n)$ is defined as the out-degree of node T_n. The PageRank (PR) of a node *A* is denoted as follows:

$$PR(A) = (1 - d) + d(\frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)})$$

The PageRank of node *A* can thus be calculated using an iterative algorithm that corresponds to the principal eigenvector of the normalised link matrix of the graph (Brin & Page, 1998). Note that the PageRank forms a probability distribution over graph nodes, so that the sum of all nodes' PageRank will be one. PageRank is an additional indicator of relative node importance and centrality in a graph. In a transportation network, PageRank can help identify key nodes (places) in the system that have a high impact on transportation efficiency.

4. Betweenness – the number of links that pass through a node. The greater the betweenness, the more important it is (Newman, 2005). For each pair of nodes in a graph, there exists at least one shortest path between them. Node betweenness refers to the number of the shortest paths that pass through a node. Therefore, betweenness helps to represent the extent to which nodes are connected, and indicate transfers from one area to another in a transport system. Although bike-sharing trips generally do not rely on or are impacted by middle stations to reach the destination, they are still limited by station availabilities (available bike and empty docks) to start or complete journeys. The work of Saberi et al. (2018) suggests that in bike-sharing systems, the probability and spatial distribution of betweenness changed in response to urban public transit failure. Furthermore, betweenness is helpful to examine system changes during special events and adverse weather conditions.

Figure 5.2 (a) shows the map of the New York bike-sharing station groups, with each dot indicating the group's central position. Figure 5.2 (b-e) give examples of the graph information properties for each node, with the different properties normalised to [0,1] for visualisation purposes. The redder and larger the plot character, the higher value it has.

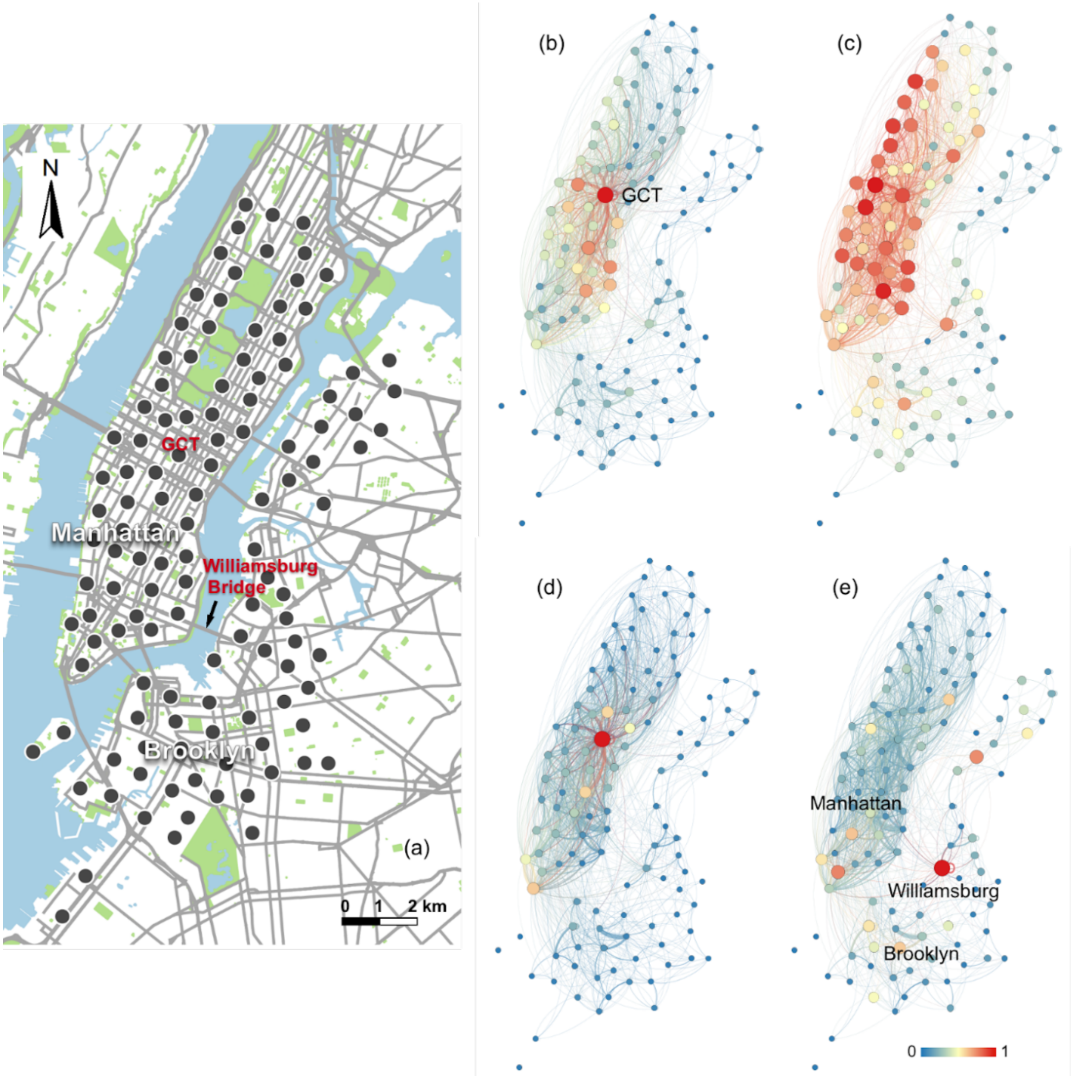


Figure 5.2 An example of the spatial distribution of graph properties using 1 hour of data; (a) station groups in New York, (b) out-strength, (c) out-degree , (d) PageRank, (e) betweenness.

This graph (figure 5.2) was constructed using 1 hour (8:00 to 9:00 am on October 25, 2017) of bike-sharing travel data, representing flow interactions in the morning rush hour. Figure 5.2 (b) shows the out-strength, directly representing area travel demand intensities. Areas close to Grand Central

Terminal (GCT) have the most bike trips with high numbers of trips in surrounding areas (Midtown Manhattan). Figure 5.2 (c) illustrates the distribution of out-degree, and suggests that different regions in Manhattan all have high levels of flow interactions indicated by the number of neighbours linked by travel flows. Interestingly, the GCT region does not have the highest out-degree. This is because the trip destinations are less diverse during the selected period. Figure 5.2 (d) shows the PageRank and has a similar pattern to figure 5.2 (b), emphasising the importance of GCT in the network. Figure 5.2 (e), shows that betweenness has a different spatial pattern to the other figures (figure 5.2 b, c, d). It highlights the region of Williamsburg, located at the east side of the Williamsburg bridge. The high betweenness value indicates its crucial role as a bridge in the graph connecting different parts of the city (e.g. Manhattan and Brooklyn). The different graph properties describe the flows and their interactions in the graph structure, allow the importance of each node to be characterised in different ways.

5.3.3 Analysis

5.3.3.1 Feature importance

Various models can be used to evaluate feature importance for making predictions, for example, Random Forest and Support Vector Machine. Among these approaches, XGBoost (extreme gradient boosting) is a gradient boosted regression tree algorithm and has been found to be one of the most powerful models in the literature (Lin et al., 2018; Li & Axhausen, 2019) and in competitions (Kaggle, 2015; KDD-Cup, 2017). It has been shown to have a comparable (or better) performance to several advanced deep neural networks such as MGCNN, ST-ResNet (Ma et al., 2019; Yao et al., 2019; Zhou et al., 2019a). Another advantage of XGBoost is that its results are easily explainable: once the boosted trees are constructed, importance (i.e. gain) scores of each feature can readily be retrieved. The importance metric provides an evaluation of how useful or valuable each feature is, based on the degree to which a feature is used to make key decisions in trees. Therefore, this study used XGBoost to evaluate feature importance.

There are potential multi-collinearity problems that may impact the feature importance identified from different models. Strong collinearity can affect model reliability and precision (Comber et al., 2018) and can result in unstable estimates of feature importance and therefore inferential and prediction biases (Dormann et al., 2013). As a result, model extrapolation may be erroneous, and there may be problems in separating variable effects (Meloun et al., 2002). For example, in a random forest model, the importance of a feature may be diluted by another highly correlated variable, because each tree is independent of others, and random choice will be made on features. XGBoost has been found to be relatively immune to the multi-collinearity problem (Chen & Guestrin, 2016b; Chen, 2018) because the algorithm does not re-focus on any specific link between feature and outcome after it has been made and learnt in the boosting process. Table 5.2 lists the input variables in the XGBoost model used to predict bike-sharing demand. Based on the literature reviewed, temporal and meteorological variables were included in the Basic Features group (Li et al., 2015; Zhou et al., 2019a).

Table 5.2 Variables in XGBoost.

Feature origin	Feature type	Variable
Basic Features	Temporal	Hour of day, Day of week, Holiday, Time-lagged travel demand (-1 hour).
	Meteorological	Temperature, Humidity, Wind speed, Weather description, Pressure.
Graph features	Time lagged (-1 hour) graph information	Out-strength, In-strength, Out-degree, In-degree, PageRank, Betweenness.

Bike travel flows were transformed into directed weighted graphs, allowing the strength and degree properties to represent the flow directions. Time-lagged travel demand is identical to time-lagged out-strength. All time-lagged properties (table 5.2) were obtained from the last hour to provide temporal dependence for the prediction (longer time-lags are examined in section 5.4.2 and 5.4.3). As XGBoost only accepts numeric values, categorical variables (e.g. hour of day) were processed using Multiple Correspondence Analysis (Meng et al., 2016) to generate lower-dimensional numeric representations.

5.3.3.2 Adding time and graph information

A good feature improves model performance (e.g. prediction) as it allows more parsimonious (less complex) models to be constructed, and non-optimised model hyperparameters to be included, whilst still generating good results. By continually adding different features into a machine learning model, changes in prediction results can be evaluated accordingly. A good feature will reduce forecasting errors, while bad features will result in higher errors (and more noise). In this study, Multi-Layer Perceptron (MLP) Neural Networks were constructed to confirm the usefulness of various input features. MLP is a class of feed-forward neural networks. It utilises backpropagation for training, and its multiple layers and non-linear nature contribute to its ability to distinguish data that is not linearly separable. As a neural network, it has a relatively simple structure, making it easier to construct and train than others. MLP also has been shown to have a strong performance in predicting short-term traffic demand (Lin et al., 2018; Li & Axhausen, 2019).

This study firstly constructed an MLP that is neither under- nor over-fitted, using the “basic features” listed in table 5.2 of meteorological and temporal features that included time-lagged travel demand of -1 hour. Different time-lagged travel demand variables and graph information properties were then sequentially added into the MLP, with the outputs evaluated accordingly. This identifies which lagged-time steps are more strongly associated with current travel demand and provides validation of the important features as identified by the XGBoost.

An investigation of the hyperparameters determined that an MLP with two layers of 32 and 8 units neither under- nor over-fitted models on both datasets. The mini-batch size was set to 1024 and training epochs to 150 (enough for convergence). The loss function used was RMSE (Root Mean Square Error), which is denoted as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2}$$

where A_i and F_i are the actual value and forecast value, respectively. There are alternative loss functions, such as Mean Absolute Error (MAE), which may be used for training machine learning models. However, the errors are squared before being averaged in RMSE, thereby giving relatively high weight to large errors. RMSE was used in this study due to the fact that in a bike-sharing system, large errors of demand estimation may pose significant difficulties to scheme operators for successful bike fleet rebalancing.

5.4 Results

5.4.1 Feature importance and variable selection

The data was split into training, validation and test datasets, with the data ordered by time. The first 80% of records were assigned as the training set, the following 10% as validation and the final 10% as the test set. The training datasets were inputted to the XGBoost models to rank the importance of different features with the results are shown in figure 5.3. Generally, temperature is considered as an important factor related to cycling activity (Miranda-Moreno et al., 2011; Thomas et al., 2013), and many bike travel demand prediction studies include temperature or a series of time-lagged temperatures as model inputs (Salaken et al., 2015; Zhou et al., 2019a). Interestingly here, in both case studies (figure 5.3 a, b), out-strength, in-strength, out-degree, in-degree and PageRank were all found have greater (or comparable) importance scores than temperature, indicating their potential utility in short-term demand prediction. This suggests that despite temperature being widely used for bike-sharing demand prediction studies, several graph features are potentially more important. Betweenness failed to outperform temperature, probably due to it being less associated with travel demand intensity. As observed in figure 5.2, strength, degree and PageRank are relatively similar in their spatial patterns, while betweenness is different as it describes the "bridge effect" of a node.

In summary, feature selection using an initial XGBoost model identified the following parameters for inclusion in subsequent models: out-strength (OS), in-strength (IS), out-degree (OD), in-degree (ID) and PageRank (PR). In the following section, the results of applying a different machine learning model (MLP) are described to confirm the utility of these graph information properties in solving bike-sharing demand prediction problems.

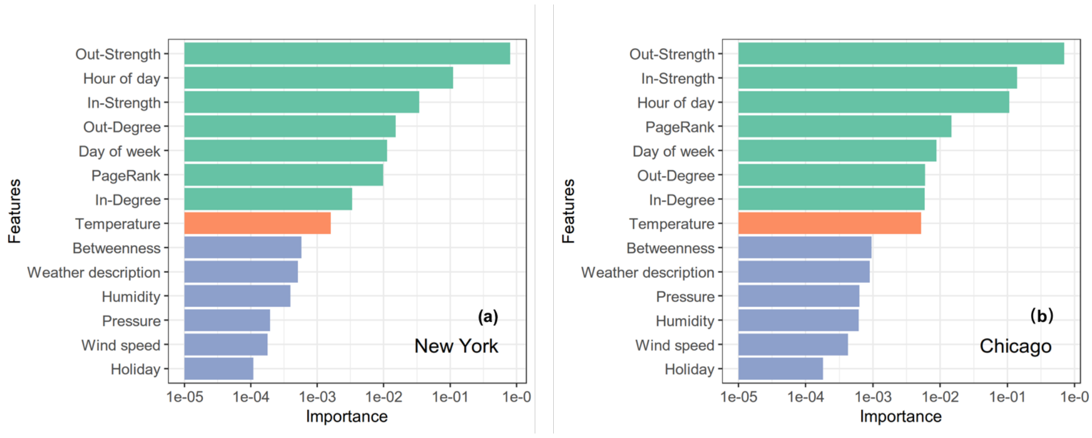


Figure 5.3 Feature importance ranked by XGBoost; (a) New York data, (b) Chicago data, with temperature as a typical benchmark shaded in orange.

5.4.2 Adding time and graph information comparison

Two types of MLP were constructed in the experiment, namely MLP-GI and MLP-DT. The former requires that graph information (GI) properties at a time lag of -1 hour are sequentially added into the model, with the order of out-strength (OS), in-strength in (IS), out-degree (OD), PageRank (PR) and in-degree (ID), as suggested in figure 5.3 (a). The MLP-DT model used time-lagged travel demand (DT) from only -1 hour to a group of -1 to -5 hours. This is a common approach, using multiple time-lagged demands from the last few hours provides a greater indication of temporal dependence in the models (Ke et al., 2017; Lin et al., 2018).

Figure 5.4 shows box plots of the distribution of the RMSE of 15 experiments, with the results evaluated on the validation set. Initially, the two models (MLP-GI and MLP-DT) are identical because they both used the travel demand (i.e. out-strength) number from the last hour. As more variables included, the MLP-GI models benefit from additional lagged graph information, with decreasing RMSE, in both average and median values. This is observed in both datasets (figure 5.4 a, b). Another finding is that adding OD (out-degree) and ID (in-degree) reduces prediction errors for the New York dataset (figure 5.4 a), but has less effect with the Chicago data (figure 5.4 b). The pattern accords with the previous finding in figure 5.3, where OD and ID much outperform the benchmark (temperature) in New York (figure 5.3 a), this again confirms the variable importance identified by XGBoost in section 5.4.1.

In the MLP-DT groups, there is a different pattern to MLP-GI. Although adding more time-lagged travel demand variables reduces errors initially, this improvement is reversed with longer sequences. In the case of New York (figure 5.4 a), the RMSE slightly increased after adding the travel demand of -5 hour. For the Chicago data (figure 5.4 b), the model shows underperformance after adding demand intensity of -4 hour, with a higher mean and median RMSE.

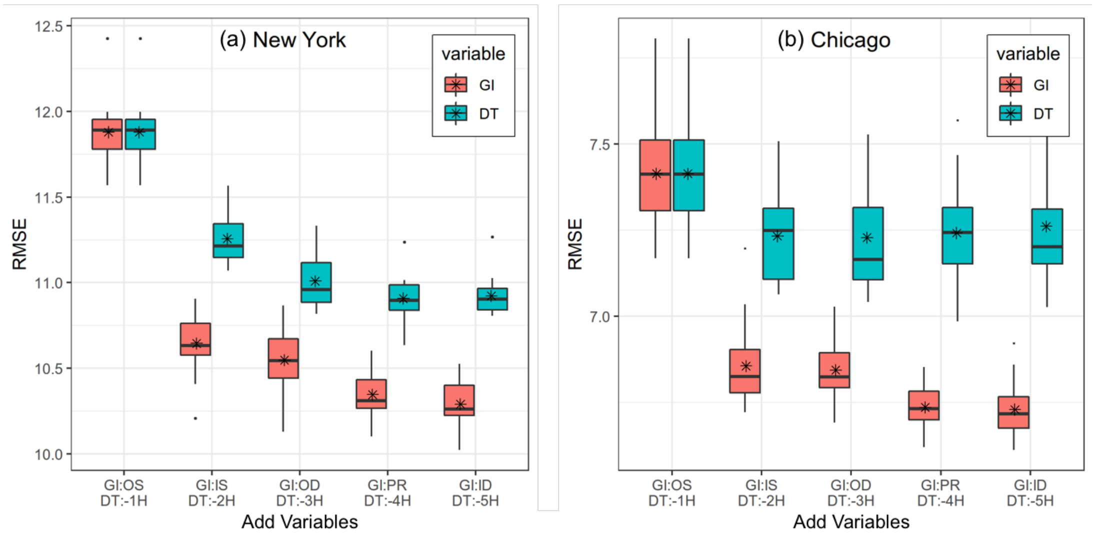


Figure 5.4 The impacts of adding different features into MLP models; (a) New York, (b) Chicago, with the mean indicated by a star and the median by a bar.

The possible underperformance with a longer temporal dependence is a general phenomenon, observed and discussed in many studies (Ke et al., 2017). The performance does not always improve when a long sequence of previous observations are fed into machine learning approaches modelling temporal dependency. The inclusion of information at less correlated timestamps can lead to poor forecasting. Therefore, the majority of previous studies only chose specific time steps to provide temporal dependence and to predict travel demand (Ke et al., 2017; Lin et al., 2018).

Comparing MLP-GI and MLP-DT with the same number of extra variables, MLP-GI always outperforms MLP-DT (see figure 5.4). The pattern indicates that using the groups of graph information properties is more effective than only using time-lagged observation of forecasting target (travel demand).

In summary, temporal dependence modelling is limited if only historical travel demand is utilised, because only a finite number of time lags will improve the prediction. However, better forecasting results may be obtained by introducing graph information properties.

5.4.3 Model comparison

The analysis and results from the previous sections indicate the potential usefulness of information derived from the bike flow interaction graph, but the properties were all derived from a single lagged timestep. This section examines how different ML models can comprehensively use varying lagged time-sequences of graph features and compares their performance with two other baseline approaches: HA (Historical Average) and ARIMA (Autoregressive Integrated Moving Average). The models are described as follows:

1. HA: uses the historical average demand for prediction. For example, the travel demand of Tuesday 12:00 is predicted as the average value of all past Tuesday's at 12:00 in the training dataset.

2. ARIMA: a statistical model, ARIMA is commonly used for analysing and forecasting time-series data. It has been widely applied in traffic prediction problems (Van Der Voort et al., 1996; Williams et al., 2003). In this work to predict demand at time T, the inputs to ARIMA were the demand observations from the first hour until T-1. It was undertaken using the automatic ARIMA model provided by the "forecast" package in R, a variation of the Hyndman-Khandakar algorithms (Hyndman et al., 2018). The model combines unit root tests, minimisation of the Akaike Information Criteria and Maximum Likelihood Estimation to construct the ARIMA. It should also be noted that the performance can be significantly influenced by model tuning, and there are also several variants, such as seasonal ARIMA, which may generate better results.

3. XGBoost: all features, including meteorological features, temporal features, and different groups of time-lagged graph information features are placed into a one-dimensional vector and used for prediction.

4. MLP: uses the same features as XGBoost, and like XGBoost, MLP does not differentiate between variables across time to model temporal dependencies.

5. LSTM: (Long-short Term Memory) is an improved version of RNN. Time-lagged variables are reshaped to a sequence and put into a bi-directional LSTM layer. Other temporal features (hour of day, day of week, holiday) and meteorological features are placed into a vector and processed using a densely-connected layer which is concatenated to the LSTM layer. The two branches are merged using another densely-connected layer. The LSTM unit is composed of three gates: input, forget and output gates. These gates determine whether to include new inputs, delete information and whether the hidden state of the current time step is carried over to the next time step (iteration). As a result, LSTMs suffer less from the vanishing gradients problem and can handle complex temporal dependencies.

XGBoost, MLP, and LSTM models have three variants, denoted as “-TD”,”-PGI”,”-FGI”, respectively. They all use the basic features including meteorological and temporal variables, but have different inputs in terms of graph information features.

1. **-TD:** uses time-lagged travel demand (out-strength) for prediction, as commonly observed in the literature (Lin et al., 2018).

2. **-PGI:** this uses part of the time-lagged graph information. Out-strength and in-strength are provided for temporal dependence modelling and demand forecasting.

3. **-FGI:** uses the full set of time-lagged graph information properties that were identified as more important than the baseline temperature variable; out-strength, in-strength, out-degree, in-degree in and PageRank.

Models with the same suffix (e.g. -PGI) used identical input features for travel demand prediction.

Incorporating the flexibility of feature engineering in machine learning models, allows them to achieve better results under the same or even reduced complexity. In this experiment, the hyperparameters of each -TD models were fine-tuned using grid search approaches, and the -PGI and -FGI models used the same hyperparameters. Therefore, -PGI and -FGI models do not significantly increase complexity in the algorithms (e.g. the learning rate in XGBoost, number of hidden layers in NN) compared to -TD models. For the neural networks, the Adam optimiser was applied as well as callbacks with a threshold of 10. This means that if the model performance (RMSE) does not improve for 10 epochs, the model will stop training to overcome potential overfitting problems.

In order to utilise hourly, daily and weekly periodicities in the model temporal dependencies (Zhang et al., 2019), time lags of today (-1 to -4 hour for the New York data; and -1 to -3 hour for the Chicago data), yesterday (-23 to -25 hours) and seven days ago (-167 to -169 hour) were selected, to provide three kinds of temporal dependence for forecasting. Graph information at these time lags was calculated and incorporated into the different models. Table 5.3 indicates the model forecasting results evaluation metrics. To eliminate randomness in NN outputs (MLP and LSTM), table 5.3 shows the average MAPE (Mean Absolute Percentage Error) and RMSE of multiple (9) experiments. MAPE is denoted as follows:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

This is a measure of relative error used to remove the scale effect of demand intensity levels, with lower MAPE generally indicating better prediction. Because bike trip number (A_t) may be 0 or near to 0 at certain places and times, leading to calculation and sensitivity problems, a threshold for A_t is usually set in MAPE evaluations (Ke et al., 2017). This study uses a threshold of 5.

Overall, the machine learning models (XGBoost, MLP and LSTM) outperform the two baseline approaches (HA and ARIMA). An important pattern is also evident: the more graph information that is included in an ML model, the lower the MAPE and RMSE values. However, different models have varying abilities in processing input features. XGBoost performed the best among the -TD models, similar to findings in other research. Lin et al. (2018) suggested that XGBoost outperforms LSTM and MLP in predicting New York bike-sharing demand, with historical travel demand included in the feature set.

When additional graph information is provided, the various -PGI models show significant improvements over the -TD models, and even lower errors with the remaining features (-FGI). Despite better forecasting results of using the full feature set, the performance of XGBoost -FGI becomes worse than LSTM -FGI. This is because time-lagged graph information properties are directly transformed into a vector for XGBoost and MLP. Although model improvements can be achieved, it is harder for them to differentiate information from different timestamps, and they fail to take full advantages of the long feature vector. However, LSTM, as a special RNN, leads to an improvement in forecasting (lower RMSE and MAPE) when using the complex full set (-FGI) of time-lagged graph information properties.

Overall, the results in table 5.3 confirm that the feature engineering in this study results in a better prediction and that different kinds of machine learning models can generally benefit from time-lagged graph information properties for bike travel demand prediction.

Table 5.3 Model result evaluation.

Model	New York		Chicago	
	MAPE(%)	RMSE	MAPE(%)	RMSE
XGBoost-TD	28.2	8.678	29.1	5.560
XGBoost-PGI	26.9	8.358	28.1	5.344
XGBoost-FGI	26.5	8.261	27.9	5.305
LSTM-TD	28.8	8.795	29.8	5.776
LSTM-PGI	27.0	8.299	28.4	5.360
LSTM-FGI	26.2	8.114	27.9	5.268
MLP-TD	29.2	8.833	29.8	5.845
MLP-PGI	27.8	8.301	28.4	5.394
MLP-FGI	27.1	8.178	28.3	5.303
ARIMA	47.1	18.273	48.8	12.49
HA	72.3	31.777	65.2	21.205

5.4.4 Spatial patterns of errors

Table 5.3 indicates that XGBoost is the strongest in the “-TD” group, and LSTM performs the best in “-FGI” family. These approaches are from two broad categories of machine learning models: regression tree and neural networks, respectively. Therefore, this section provides spatial interpretation as a supplementary analysis of the two models, and the results are shown in figures 5.5 and 5.6.

Figure 5.5 (a) indicates the total travel demand in each region (groups of stations) in the New York case study over the period of the test set. Figure 5.5 (b, c, d) shows how LSTM models benefit from additional graph information variables to forecast bike travel demand in New York. Areas close to Manhattan Midtown south (marked as “1” in figure 5.5 a) shows improvement in the LSTM-PGI model (figure 5.5 c), they are also areas with a high bike trip intensity. LSTM-FGI (figure 5.5 d) further improves the prediction by reducing MAPE in Upper East Side and Brooklyn (marked as“2” and “3” in figure 5.5 a), which presents medium-high demand. XGBoost also benefited from additional graph information properties (figure 5.5 e, f, g), but to a lesser degree than the changing patterns in LSTM, especially when the “-PGI” and “-FGI” models are compared. For example, less improvement was found in the Manhattan Midtown south area in figure 5.5 (e-f-g), compared to figure 5.5 (b-c-d). This pattern also accords with the findings in table 5.3, as LSTM experienced a significant decrease in MAPE from -PGI to -FGI. This again highlights LSTM’s ability to process complicated sequential information. It should also be noted that LSTM and XGBoost may outperform each other in different areas of the city, suggesting that no single ML algorithm will have the best performance at all areas, as discussed in the work of Li and Axhausen (2019).

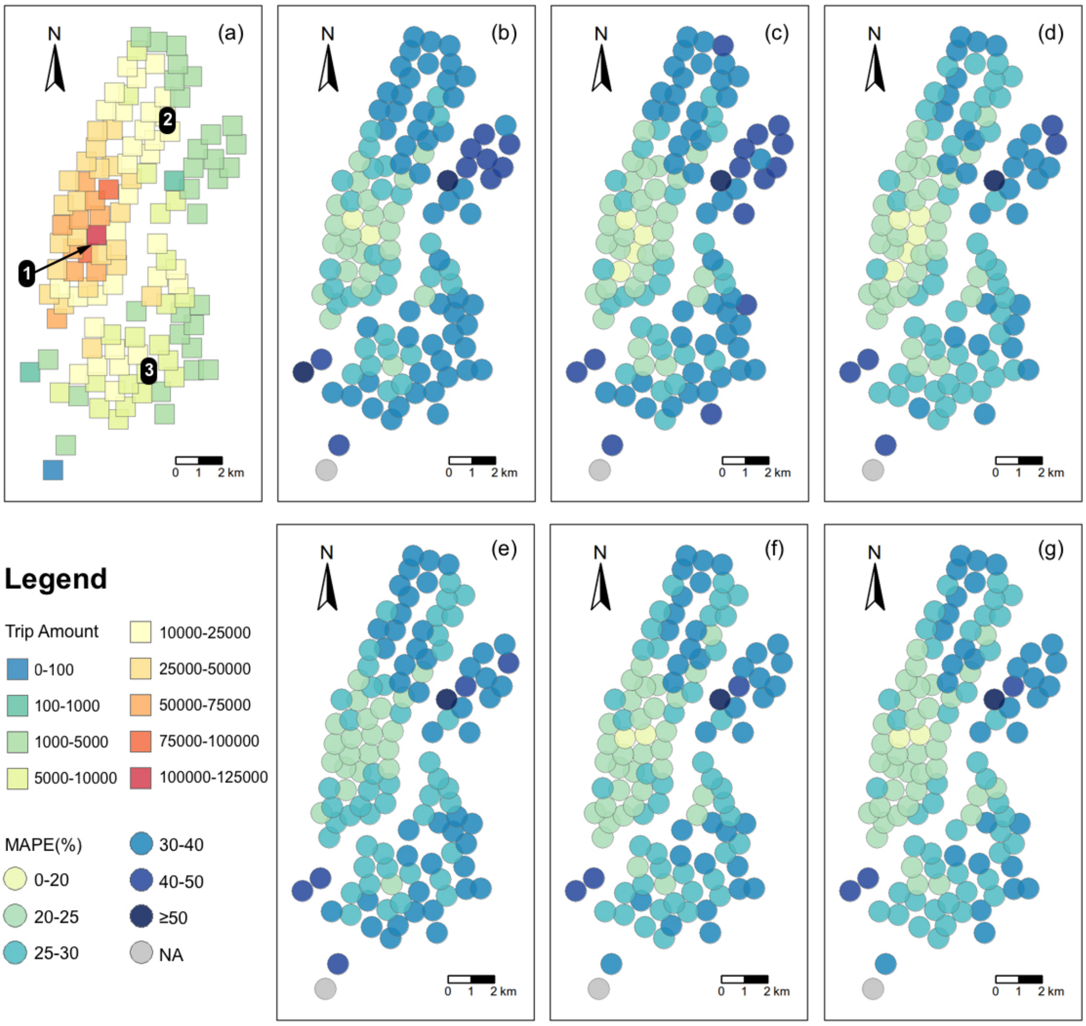


Figure 5.5 Travel demand and MAPE of different models on New York dataset; (a)Travel demand (b) LSTM-TD, (c) LSTM-PGI, (d) LSTM-FGI, (e) XGBoost-TD, (f) XGBoost-PGI, (g) XGBoost-FGI.

Similar patterns to figure 5.5 are observed in figure 5.6 for the Chicago case study. XGBoost outperformed LSTM in the “-TD” models (figure 5.6 b, e; table

5.3), but LSTM-FGI (figure 5.6 d) obtained better predictions than both LSTM-PGI (figure 5.6 c) and XGBoost-FGI (figure 5.6 g) in areas that have large numbers of travel demand around the city centre. This is helpful for bike fleet management because regions with higher demand may experience greater bike shortages and more precise forecasting benefits the rebalancing work of scheme operators.

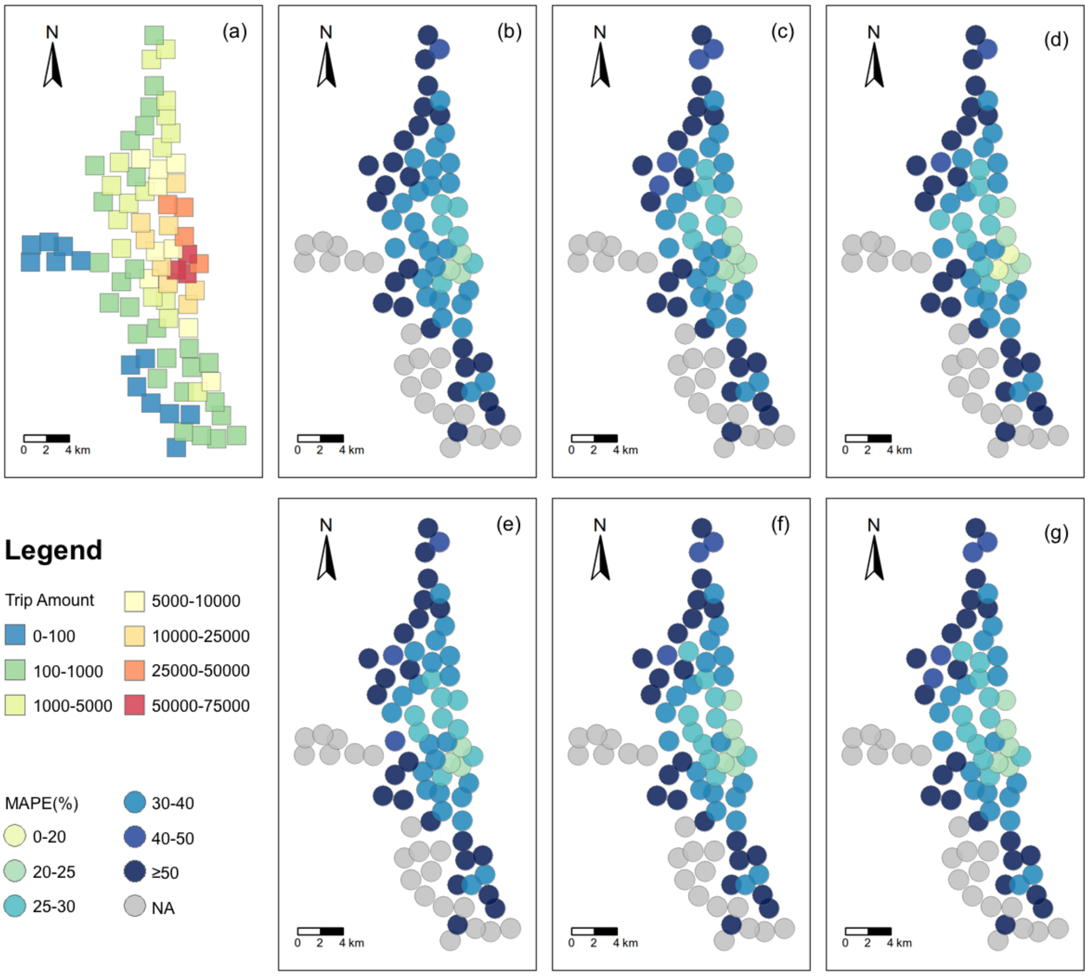


Figure 5.6 Travel demand; and MAPE of different model on Chicago dataset; (a) Travel demand (b) LSTM-TD, (c) LSTM-PGI, (d) LSTM-FGI, (e) XGBoost-TD, (f) XGBoost-PGI, (g) XGBoost-FGI.

5.5 Discussion and conclusions

By examining travel flow interactions in transport systems, it is possible to shed light on the underlying structural characteristics of regions. This work highlights the importance of domain knowledge and feature engineering in machine learning problems. Casting complex urban systems such as transport networks into graph structures allows graph derived measures such as node importance and centrality to be included in models to capture and represent travels flow and regional attractiveness patterns (Batty, 2013; Zhang et al., 2017; Yang et al., 2019). Related graph features improve and enhance modelling and prediction in both tree-based model and neural networks, demonstrating the utility of better feature engineering.

It should be noted that this research predicts demand at regional levels (groups of station), rather than at the individual station level in order to avoid the impacts of service change and to reduce noise in such a complex system (Li et al., 2015). There are other strategies to eliminate these, for example, Lin et al. (2018) sought to predict demand at individual station level, and removed more than half of the New York bike stations from the data in order to only focus on stations that persisted over time with relatively high travel demand. Despite finer spatial granularity (station level), their approach provided only a partial representation of actual demand patterns.

The station group/cluster size used in this work was a relatively arbitrary decision and may have affected the graph properties used in the models. Very large groups (areas) may result in many travel flows that start and end at the same region, making various centrality measures less representative of the actual dynamics and flows. Therefore, the choice of group number and clustering needs to find a balance between fine flow representations and system noises elimination.

There are several shortcomings in this study that will be improved and investigated in future work. First, more statistical time-series models could be used for comparisons, such as KNN and seasonal ARIMA. Other hybrid deep neural networks may also be applied to verifying the FGI improvement, examples include MGCNN and ST-ResNet. MGCNN can benefit from concatenating RNN (LSTM, GRU) layers to model time-lagged variables (Lin et al., 2018), and it may be enhanced by FGI further, just like LSTM has shown in this work. This study only compared the utility of time-lagged graph indices on three machine learning models, because they are recognised as powerful models (Li & Axhausen, 2019; Lin et al.,2018) in literature for travel demand forecasting, also belong to different model family (regression tree, feed-forward neural network and recurrent neural network). Some complex hybrid neural networks are relatively difficult to implement and may suffer from reproducibility issue to some extent. For example, MGCNN models do not have a “stable” performance for predicting short-term traffic demand; Zhou et al. (2019a) tested the performance of MGCNN on short-term demand forecasting in four different bike-sharing and taxi datasets. The result (Zhou et al., 2019a) suggested that MGCNN is worse than XGBoost in half datasets while outperforming the remaining ones. The work of Li & Axhausen (2020) also suggested that the MGCNN's outperformance is not stable in all periods and sub-regions. The result of several models are compared at each hourly interval during the course of the day (Li & Axhausen 2020); while obtaining lower RMSE, MGCNN is normally worse than random forest in terms of MAPE indicator; it only slightly outperforms random forest during the morning peak (3 hours from 6:00 to 9:00 am), while showing higher MAPE in the remaining 21 hourly intervals (Li & Axhausen 2020). The "unstable" performance is partly due to the difficulties in constructing the semantic graph. MGCNN's is capable of obtaining good prediction of dynamics based on planar networks (e.g. road networks) due to the clear concept of graph construction – roads are connected at intersections. However, for non-planar networks such as origin and destination graphs (e.g. bike-sharing travel flow graph. The nodes (docking stations) connection in semantic graphs is subject to several factors, for example, they depend on the relatively arbitrary and ambiguous choice of threshold (e.g., proximity, similarity, consistency), as well as specific preprocessing (e.g., removing less-used stations) (Chai et al., 2018; Lin et al., 2018). These lead to reproducibility issue in current MGCNN literature for predicting short-term travel demand. Therefore, this work did not include MGCNN for comparison to avoid bringing higher uncertainties into the results. Second, this study applied node-level graph information properties for better forecasting. Future work will examine the utility of including edge level (e.g. edge betweenness) and sub-graph level (e.g. modularity) information to improve transport demand forecasting. Third, this work only used data from two American cities, although similar patterns were identified, it is uncertain whether these findings are universally applicable to bike-sharing systems in other regions (e.g. Asia, Europe). Additionally, both datasets are from dock-based bike-sharing systems. Examining dockless bike-sharing systems as graphs (Yang et al., 2019) and deriving useful information for demand forecasting is an area for further study.

Overall, this study identified the importance and effectiveness of time-lagged graph information properties in bike-sharing travel demand forecasting. Analysis of real-world data from different cities suggests that several time-lagged graph properties are of greater relevance (importance) for predicting bike demand than more commonly used environmental measures. Graphs capture important structural information and system properties, and graph derived measures should be included in forecasting models. The follow-up experiments confirmed the improvements to several advanced machine learning approaches, noting that LSTM neural networks are able to effectively use a complex set of graph features, due to their ability to process sequential information.

A number of graph information variables were found to improve machine learning prediction of bike travel demand when included as lagged information in ML models: out-strength, in-strength, out-degree, in-degree and PageRank. Using in-strength can significantly decrease prediction errors, while the inclusion of the full set can lead to even lower average errors. The improvement also presents a spatial pattern and is more evident in areas with a medium and high volume of journeys, which is helpful in real-world applications. Unlike many data enrichment methods, this approach does not require data from other sources (e.g. land-use information from POI, Twitter) or extra processing, data cleaning and fusion. These features are easily derived from bike flow graphs and are relatively easy to include in existing models. Predictions using such data can inform bike scheme operators, help them to better understand and model demand patterns in different urban areas and to run more successful bike-sharing schemes, thereby promoting sustainable transport. The improved short-term demand predictions can also benefit "user-based rebalance" activities (Duan et al., 2019; Wu et al., 2019), which often have directed user incentives to help bike rebalancing work, and dynamically optimise service provision.

The key finding from this work is that time-lagged graph flow information derived from actual bike-sharing patterns were found to be stronger predictors of demand than several more commonly used temporal features and meteorological features. This is because graph structural information captures important spatial and behavioural properties. Our study also found LSTM neural networks to be the most effective at handling a complex set of graph features and at processing sequential information. Combining these, resulted in enhanced and more accurate demand forecasting in bike-sharing systems.

References

Austwick, M.Z., O'Brien, O., Strano, E. and Viana, M. 2013. The structure of spatial networks and communities in bicycle sharing systems. *PloS One.* **8**(9), pe74685.

Batty, M. 2013. *The new science of cities*. Cambridge: MIT press.

Beecham, R., Wood, J. and Bowerman, A. 2014. Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems.* **47**, pp.5-15.

Borges, J., Ziehr, D., Beigl, M., Cacho, N., Martins, A., Sudrich, S., Abt, S., Frey, P., Knapp, T. and Etter, M. 2017. Feature engineering for crime hotspot detection. In: *Proceedings of 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, 4-8 August 2017, San Francisco. Piscataway: IEEE, pp.1-8.

Brin, S. and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems.* **30**(1-7), pp.107-117.

Cao, Y. and Shen, D. 2019. Contribution of shared bikes to carbon dioxide emission reduction and the economy in Beijing. *Sustainable Cities and Society.* **51**, p101749.

Chai, D., Wang, L. and Yang, Q. 2018. Bike flow prediction with multi-graph convolutional networks. In: *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 6-9 November 2018, Seattle. New York: Association for Computing Machinery, pp.397-400.

Chen, C., Li, K., Teo, S.G., Chen, G., Zou, X., Yang, X., Vijay, R.C., Feng, J. and Zeng, Z. 2018. Exploiting spatio-temporal correlations with multiple 3d convolutional neural networks for citywide vehicle flow prediction. In: *2018 IEEE International Conference on Data Mining (ICDM)*, 17-20 November 2018. Piscataway: IEEE, pp.893-898.

Chen, L., Zhang, D., Wang, L., Yang, D., Ma, X., Li, S., Wu, Z., Pan, G., Nguyen, T.-M.-T. and Jakubowicz, J. 2016. Dynamic cluster-based over-demand prediction in bike sharing systems. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing,*12-16 September 2016, Heidelberg. New York: Association for Computing Machinery, pp.841-852.

Chen, T. and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 13-17 August 2016, San Francisco. New York: Association for Computing Machinery, pp.785-794.

Chen, T., Tong, H., Benesty, M., Yuan, T., 2018. *Understand your dataset with Xgboost.* [Online]. [1 December 2020]. Available from: https://cran.r-project.org/web/packages/xgboost/vignettes/discoverYourData.html

Comber, A., Chi, K., Huy, M.Q., Nguyen, Q., Lu, B., Phe, H.H. and Harris, P. 2018. Distance metric choice can both reduce and induce collinearity in geographically weighted regression. *Environment and Planning B: Urban Analytics and City Science.* p2399808318784017.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B. and Leitão, P.J. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography.* **36**(1), pp.27-46.

Duan, Y. and Wu, J. 2019. Optimizing rebalance scheme for dock-less bike sharing systems with adaptive user incentive. In: *Proceedings of 2019 20th IEEE International Conference on Mobile Data Management*, 10-13 June 2019, Hong Kong. New York: IEEE, pp.176-181.

Feng, S., Chen, H., Du, C., Li, J. and Jing, N. 2018. A hierarchical demand prediction method with station clustering for bike sharing system. In: Proceedings of 2018 IEEE Third International Conference on Data Science in Cyberspace, 18-21 June 2018, Guangzhou. New York: IEEE, pp.829-836.

Fishman, E. 2016. Bikeshare: A review of recent literature. *Transport Reviews.* **36**(1), pp.92-113.

Froehlich, J.E., Neumann, J. and Oliver, N. 2009. Sensing and predicting the pulse of the city through shared bicycling. In: *Twenty-First International Joint Conference on Artificial Intelligence*, 12-13 July 2009, Pasadena. Palo Alto: AAAI Press, pp.1-7.

Fu, R., Zhang, Z. and Li, L. 2016. Using LSTM and GRU neural network methods for traffic flow prediction. In: *Proceedings of 2016 31st Youth Academic Annual Conference of Chinese Association of Automation*, 11-13 Nov 2016, Wuhan. New York: IEEE, pp.324-328.

Giot, R. and Cherrier, R. 2014. Predicting bikeshare system usage up to one day ahead. In: *Proceedings of 2014 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems*, 9 – 12 December 2014, Orlando. New York: IEEE, pp.22-29.

Goodfellow, I., Bengio, Y. and Courville, A. 2016. *Deep learning.* Cambridge: MIT press.

Hall, M.A. and Smith, L.A. 1998. Practical feature subset selection for machine learning. *Journal of Computational Science.* **98**, pp.181-191.

Hoang, M.X., Zheng, Y. and Singh, A.K. 2016. Forecasting citywide crowd flows based on big data. In: *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* 31 October – 3 November 2016, San Francisco. New York: Association for Computing Machinery, pp.1-10.

Hyndman, R.J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. and Yasmeen, F. 2018. *forecast: Forecasting functions for time series and linear models.* [software]. [Accessed 1 December 2020].

Kaggle. 2015. *Bike Sharing Demand.* [Online]. [Accessed 22 September, 2019]. Available from: https://www.kaggle.com/c/bike-sharing-demand

Karlaftis, M.G. and Vlahogianni, E.I. 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies.* **19**(3), pp.387-399.

KDD-Cup. 2017. *Announcing KDD Cup 2017: Highway tollgates traffic flow prediction.*[Online]. [Accessed 22 September, 2019]. Available from: https://www.kdd.org/kdd2017/announcements/view/announcing-kdd-cup-2017-highway-tollgates-traffic-flow-prediction

Ke, J., Zheng, H., Yang, H. and Chen, X.M. 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies.* **85**, pp.591-608.

Ketkar, N. 2017. *Deep Learning with Python.* New York: Springer.

Kim, K. 2018. Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *Journal of Transport Geography.* **66**, pp.309-320.

Li, A. and Axhausen, K.W. 2019. *Comparison of short-term traffic demand prediction methods for transport services.Zurich: ETH Zurich*. [Accessed 1 December 2020]. Available from: https://www.research-collection. ethz.ch /handle/20.500.11850/356143

Li, A., & Axhausen, K. W. 2020. Short-term Traffic Demand Prediction using Graph Convolutional Neural Networks. In: *Proceedings of Association of Geographic Information Laboratories in Europe (AGILE) Conference 2020.* 18-20 June, 2020. Crete, Sweden. London: Springer, pp.1-14.

Li, Y. and Shuai, B. 2018. Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms. *Multimedia Tools and Applications.* **79**, pp.1-12.

Li, Y., Zheng, Y., Zhang, H. and Chen, L. 2015. Traffic prediction in a bike-sharing system. In: *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 3-6 November 2015, Seattle. New York: Association for Computing Machinery, p33.

Lin, L., He, Z. and Peeta, S. 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies.* **97**, pp.258-276.

Liu, J., Sun, L., Chen, W., & Xiong, H. (2016, August). Rebalancing bike sharing systems: A multi-source data smart optimization. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 13-17 August 2016, San Francisco. New York: Association for Computing Machinery, pp. 1005-1014.

Lovelace, R. and Philips, I. 2014. The ‘oil vulnerability’ of commuter patterns: A case study from Yorkshire and the Humber, UK. *Geoforum.* **51**, pp.169-182.

Ma, S., Guo, J., Guo, S. and Guo, M. 2019. *Position-Aware Convolutional Networks for Traffic Prediction.* New York：arXiv preprint,p.1904.06187.

Ma, T., Liu, C. and Erdoğan, S. 2015. Bicycle sharing and public transit: does Capital Bikeshare affect Metrorail ridership in Washington, DC? *Transportation Research Record.* **2534**(1), pp.1-9.

Meloun, M., Militký, J., Hill, M. and Brereton, R.G. 2002. Crucial problems in regression modelling and their solutions. *Analyst.* **127**(4), pp.433-450.

Meng, C., Zeleznik, O.A., Thallinger, G.G., Kuster, B., Gholami, A.M. and Culhane, A.C. 2016. Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics.* **17**(4), pp.628-641.

Miranda-Moreno, L.F. and Nosal, T. 2011. Weather or not to cycle: Temporal trends and impact of weather on cycling in an urban environment. *Transportation Research Record.* **2247**(1), pp.42-52.

Newman, M.E. 2005. A measure of betweenness centrality based on random walks. *Social Networks.* **27**(1), pp.39-54.

O'Brien, O., Cheshire, J. and Batty, M. 2014. Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography.* **34**, pp.262-273.

Rodrigues, F., Markou, I. and Pereira, F.C. 2019. Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach. *Information Fusion.* **49**, pp.120-129.

Rudloff, C. and Lackner, B. 2013. Modeling demand for bicycle sharing system–neighboring stations as a source for demand and a reason for structural breaks. *Transportation Research Record.* **2430**(1):1-11.

Saberi, M., Ghamami, M., Gu, Y., Shojaei, M.H.S. and Fishman, E. 2018. Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. *Journal of Transport Geography.* **66**, pp.154-166.

Salaken, S.M., Hosen, M.A., Khosravi, A. and Nahavandi, S. 2015. Forecasting bike sharing demand using fuzzy inference mechanism. In: *Proceedings of 2015 International Conference on Neural Information Processing*, 9-12 November 2015, Istanbul. New York: Springer, pp.567-574.

Shaheen, S.A., Guzman, S. and Zhang, H. 2010. Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record.* **2143**(1), pp.159-167.

Thomas, T., Jaarsma, R. and Tutert, B. 2013. Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling. *Transportation.* **40**(1), pp.1-22.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography.* **46**, pp.234-240.

Tran, T.D., Ovtracht, N. and d'Arcier, B.F. 2015. Modeling Bike Sharing System using Built Environment Factors. *Procedia CIRP.* **30**, pp.293-298.

Van Der Voort, M., Dougherty, M. and Watson, S. 1996. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies.* **4**(5), pp.307-318.

Vlahogianni, E.I., Karlaftis, M.G. and Golias, J.C. 2014. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies.* **43**, pp.3-19.

Williams, B.M. and Hoel, L.A. 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering.* **129**(6), pp.664-672.

Wu, R., Liu, S. and Shi, Z. 2019. Customer incentive rebalancing plan in free-float bike-sharing system with limited information. *Sustainability.* **11**(11), p3088.

Xu, C., Ji, J. and Liu, P. 2018. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies.* **95**, pp.47-60.

Xu, H., Ying, J., Wu, H. and Lin, F. 2013. Public bicycle traffic flow prediction based on a hybrid model. *Applied Mathematics & Information Sciences.* **7**(2), p667.

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems.* **77**, p101361.

Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J. and Moscibroda, T. 2016. Mobility modeling and prediction in bike-sharing systems. In: *Proceedings of the 14th annual International Conference on Mobile Systems, Applications, and Services*, 26-30 June, Singapore. New York: Association for Computing Machinery, pp.165-178.

Yao, H., Tang, X., Wei, H., Zheng, G. and Li, Z. 2019. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In: *Proceedings of 33rd AAAI Conference on Artificial Intelligence*, 27 January – 1 February 2019, Honolulu. Vancouver: PKP, pp.5668-5675

Yao, H., Tang, X., Wei, H., Zheng, G., Yu, Y. and Li, Z. 2018a. *Modeling spatial-temporal dynamics for traffic prediction*. New York: arXiv preprint p.1803.01254.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J. and Li, Z. 2018b. Deep multi-view spatial-temporal network for taxi demand prediction. In: *Proceedings of 32rd AAAI Conference on Artificial Intelligence*. 2–7 February 2018, New Orleans. Vancouver: PKP, pp.2588-2595.

Zhang, J., Zheng, Y. and Qi, D. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of *31st AAAI*

Conference on Artificial Intelligence. 4-9 February 2017, San Francisco. Vancouver: PKP, pp.1-7

Zhang, Y., Cheng, T. and Ren, Y. 2019. A graph deep learning method for short‐term traffic forecasting on large road networks. *Computer‐Aided Civil and Infrastructure Engineering.* **34**, p19.

Zhong, C., Arisona, S.M., Huang, X., Batty, M. and Schmitt, G. 2014. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science.* **28**(11), pp.2178-2199.

Zhou, Y., Chen, H., Li, J., Wu, Y., Wu, J. and Chen, L. 2019. Large-Scale Station-Level Crowd Flow Forecast with ST-Unet. *ISPRS International Journal of Geo-Information.* **8**(140), pp.1-16.

Zhou, Y., Li, Y., Zhu, Q., Chen, F., Shao, J., Luo, Y., Zhang, Y., Zhang, P. and Yang, W. 2019. A reliable traffic prediction approach for bike‐sharing system by exploiting rich information with temporal link prediction strategy. *Transactions in GIS.* **23**(5), pp.1125-1151.

Chapter 6
Discussion and conclusions

The thesis is concluded and the research undertaken is critically discussed in this chapter. A chapter-by-chapter summary is given in section 6.1, and the main findings of each chapter are outlined. Section 6.2 expounds the contribution of the thesis to the literature of urban and transport studies. Section 6.3 critically examines the limitations as well as future studies. A conclusion is provided in section 6.4, which also provides an outlook on understanding and modelling sustainable urban mobility.

6.1 Thesis summary

Chapter 1 demonstrated the research background and aims, the works in the thesis are devoted to deepening our understanding of dynamics and interactions in sustainable transport (bike-sharing) over the urban last mile, adopting spatiotemporal and graph-based approaches.

Chapter 2 provided a critical review of the relevant research on graph theory and its applications in urban and transportation systems. Firstly, the graph theory and various metrics were introduced, including their definition and potential indication in the geographical context. Secondly, the application of graphs in geography and urban systems are introduced, along with the data which has been used to generate graph structures to characterise urban mobility structures. Thirdly, chapter 2 reviewed the history of bike-sharing and the new research opportunities that emerged from the IT-based and dockless schemes to understand travel behaviours in urban last mile. Overall, the review concluded that urban travel flow interactions should be understood in the context of local processes, such as urban development and interventions. The breakdown to analysing temporal graphs will help interpret the variation and evolution in behaviours and the dynamics in urban space.

In chapter 3, the dockless bike-sharing scheme and the new opportunity and challenges it brings are presented. Through examining the flexible travel flows in the dockless bike-sharing scheme, the study revealed the quick adaption by people to combine a dockless bike with metro travels to complete their journeys. The integration with mass transit systems can rapidly boost local cycling demand, leading to new clusters of cycle trip origins and destinations emerging around new metro stations within walking distance. By aggregating flows origin and destinations by road segments in a novel way, related graph metrics are extracted and compared. This research found that with the development of the mass transit system, dockless bike-sharing is a flexible “first/last mile” solution, and the urban last mile has a tendency towards being polycentric and showing stronger local connectivity.

In chapter 4, dock-based bike-sharing (London Cycle Hire Scheme) were analysed to reveal how cycling travel behaviour changed during different transit disruptions in London. The results suggest that bike-sharing usage generally rises in response to transit disruptions, but the extent and nature of this rise varies according to the spatiotemporal characteristics of disruption. A novel measure of station pressure suggests that the scheme very quickly reaches saturated capacity and leads to insufficient service provision during disruptions. Graph-based analysis reveals several changes in OD flow structures, which imply a modal shift from Tube to bike-sharing and a change of route behaviours amongst bike-sharing users. Weekday Tube strikes may bring many new behaviours and new OD pairs to the bike flow structures, whilst for weekend strikes existing patterns are relatively consolidated.

Chapter 5 was inspired by one of the key findings that emerged from chapters 3 and 4: cycling flow structures rapidly evolve in response to different interventions in the urban system, and this varies according to different spatial (locations) and temporal variable (e.g. hour of the day). Flow structural dynamics represent the underlying changing mobility behaviours of people, therefore examining the temporal graph metrics is helpful for improving understanding of these behaviours and therefore modelling the system. This chapter extracted time-lagged graph metrics from real-world bike-sharing datasets, including graph node out-strength, in-strength, out-degree,in-degree and PageRank. They were added to different machine learning algorithms to predict short-term bike demand. Results indicated that the time-lagged graph metrics were more significant in predicting demand than commonly used meteorological information. Different machine learning approaches (XGBoost, MLP, LSTM) all benefit from the additional graph features. Deep neural networks were found to be better able to handle the sequences of the time-lagged graph variables than other approaches, resulting in more accurate forecasting. Thus, incorporating graph-based features can improve understanding and modelling of demand patterns in urban areas, supporting bike-sharing schemes and promoting sustainable transport.

6.2 Contribution to literature

In this section, the potential contribution of the research behind this thesis in a broader context of previous work are summarised. The contribution can be divided into four categories. (1) Contribution to the literature of dockless micro-mobility service and user behaviours; (2) Contribution to the literature of understanding urban dynamics over the last mile, especially how the structures change in response to different interventions; (3) Contribution to the literature of short-term travel demand prediction; and (4) Contribution to the literature of graph-based analysis on temporal travel flows.

6.2.1 Contribution to the literature of dockless shared-mobility services and user behaviours

In recent years, dockless bike-sharing schemes have emerged and grown rapidly in several countries. The fleets of these bikes in China surged in number to around 23 million in approximately two years, accounting for around 92% of all sharing bikes in China in 2018 (Gu et al., 2019). The proportion also grew rapidly (44%) in the USA (Shaheen et al., 2020). Given similar amounts of economic and social investment, dockless schemes are considered to operate more efficiently and provide more convenience for travellers than docked-based service (Mooney et al., 2019; Chen et al., 2020). Sharing vehicles in such a way radically improves mobility (McKenzie, 2020). Dockless electric powered scooters are another new trend that is radically changing the way that many people move around. In the future, shared self-driving vehicles (of whatever kind) might completely change the ways people live and move around and not just in urban areas. Problems such as congestion and car parking will be relieved by shared-mobility systems resulting in a more sustainable and more inclusive transport system. The changes are taking place first in more urban areas where there is a critical mass of potential users and where the problems of congestion and where to park vehicles is most acute. Therefore, understanding the dockless mode of shared-mobility is of vital importance in transport and urban planning. But because this is a relatively new phenomenon, there is little published academic literature, especially that based on quantitative, detailed and large scale studies.

This study addressed the challenges of understanding dockless micro-mobility service and user behaviours by:

- Proposing the method of detecting dockless bike trips from time-series of location snapshots, the trip origin and destinations can be determined by the location changes of individually available bikes.

- Identifying the role of dockless bike-sharing in rapidly serving for metro systems as the first/last mile solution, implying the quick adaption of new behaviour among people and the underlying urban resilience.
- Providing guidance of parking area planning for dockless bikes. For example, to promote the integration of metro and bike usage, parking areas should be located close to metro stations within an identified acceptable walking distance (220 m in the case study).

6.2.2 Contribution to the literature of understanding urban dynamics and structures over the last mile

The spatial structures of the city and the flow interactions within are becoming more complex with the growing population and new technology. People now engage in more diverse activities and have more sustainable travel options; these new forms and patterns of flow interactions remain to be understood. Despite the dynamics and structures of cities that have been explored in a number of studies (Zhong et al., 2014; Sun et al., 2015), the focus is normally on the broad-scale urban flows, for example, long-medium intra-urban travels in metro and taxi. Little knowledge of finer-scale "capillary" flows regarding individual non-motorised short-distance movement (e.g. cycling) is uncovered.

The research in this thesis contributes to a more comprehensive picture of the structure and dynamics over the urban last mile through analysing bike-sharing travel records at the individual level. These are summarised as:

- Quantifying the spatial distribution and temporal pattern (e.g. peak hours) of cycling behaviours in the urban last mile, within the dock-based and dockless bike-sharing scheme.
- Presenting the interdependence between the mass transit system and bike-sharing, and how the urban last mile activities may respond to different interventions. Disruptions in the metro system will lead to higher travel amounts in bike-sharing, and the station-level "distance decay" effect suggested in previous literature (Saberi et al., 2018) is further supplemented and explained by the work in this thesis. The effect is subject to the density of stations and commuting behaviours, thus can be weakened or eliminated by travels of other purposes (e.g. leisure).
- Introducing a novel "service pressure index" to quantify the challenges of bike-sharing schemes in providing sufficient resources. Spatial and temporal patterns of the index are presented within the context of mass transit disruption to demonstrate its utility to observe service provision shortages.

- Quantifying the structural changes in the urban last mile towards polycentric transformation with the development and extended services of mass transit system.

6.2.3 Contribution to the literature of travel demand prediction

Short-term traffic demand prediction is a key component for efficient transport and smart city management. In the last few decades, various prediction models have been extensively studied to forecast traffic dynamics. But it is very challenging to fully address this problem because of the complexity of transport and urban systems. Travel flows, demand and volumes are the outcomes of the interaction between people and the urban space. A number of factors, such as traveller's socioeconomic background, road network and infrastructures, traffic incidents, weather conditions, all can bring changes to the transport system. Although various models were studied, ranging from classical statistical models to relatively novel machine learning models, few studies have put a focus on the model inputs to answer the vital question: what hidden variables might be helpful to enhance different model's prediction performance. The study presented in this thesis addressed the question by:

- Using real-world datasets, this work quantified and proved that time-lagged graph node metrics (in-strength, out-strength, in-degree, out-degree and PageRank) are more helpful than meteorological features (e.g. temperature) for short-term travel demand prediction.
- Suggesting that different machine learning algorithms, such as gradient boosting trees and deep neural networks, can all benefit from extra input variable that derived from time-lagged temporal graph structures.
- Proposing to use LSTM neural networks layers to process the complex sequence of time-lagged graph variables, while concatenating with MLP layers to deal with the remaining variables (e.g. meteorological features). The model is able to obtain a better prediction for short-term travel demand in bike-sharing system, comparing to XGboost and MLP models.

6.2.4 Contribute to the literature of graph-based analysis on temporal travel flows

Using graph theory to mathematically characterise dynamics in travel flow is one of the key areas of novelty within this research. The generated graph structures covered a different range of temporal resolutions, from weekly (chapter 3) to daily (chapter 4), further broken down to hourly in chapter 5. The spatiotemporal and graph-based analysis provides the opportunities to better understand the flows and changing nature behind them. The proposed method and obtained findings are as follow:

- Proposing using short road segments to aggregate dockless bike trip origins and destinations for constructing an OD matrix, this approach has the potential to capture the spatial details in short-distance travel flows.
- Confirming that cycling flows in both dock-based and dockless bike-sharing present power-law distribution in graph metrics; disruptions (e.g. transit strike) or infrastructure changes (e.g. new metro line) will not change this nature.
- Investigating mobility graphs with high temporal resolution (hourly interval), This work showcased the utility of the highly dynamic temporal graph and the metrics in better modelling dynamics in transport systems(chapter 5).

6.3 Discussion and limitation of the study

A number of limitations, however, were encountered during the research. The limitations are highlighted in this section, and they should be considered while applying the research outcomes to urban and transport studies.

6.3.1 Limitation in understanding people's travel behaviours

The research presented in this thesis focused on analysing the structural changes in intra-urban cycling travel flows. An assumption is made that the variance in the graph structures represents the underlying changed behaviour of people, and associated with the potential transformation in the function of urban space over time or under interventions. However, it should be noted that when interpreting the driving force and mechanism of why people changed their travel options or routes, the consideration of individual socioeconomic information is vital but missing in this thesis. The changed behaviours should be further broken-down to link with the heterogeneity that lies in each individual. Lacking consideration in users' socioeconomic characteristics carries the risk of making an inaccurate inference. In chapter 3 and 4, although the ridership has increased, it is not clear whether the additional and changed journeys are made by the same group of users. It could be the case that large amounts of previous users gave up using bike-sharing, while the new users are from a completely different socioeconomic background. In this case, the changed travel flows and structures do not represent the adoption of new behaviours, but only reflect the mobility characteristics of different groups in the population. It is a paradox that despite the emerging transport data at finer spatiotemporal granularity, access to the individual's identity, socioeconomic or travel purpose is highly limited (Kitchin, 2013). Researchers are normally cut-off from the rich seam of potential studies unless the rare access is granted by private and public sectors of the data owner. Moreover, sometimes the socioeconomic variables are not collected because the travel records are generated only for management and engineering purposes rather than research, while only fundamental behaviours (e.g. action of departure) is available.

But there are possible solutions to supplement this understanding, for example, utilising social media, land use or POI data at the trips' origin and destination (Yang et al., 2019) can help infer travel purpose. The census data (Zhang et al., 2020) or other geodemographic indicators, such as IMD (index of multiple deprivations) (Noble et al., 2006), also have the potential to illustrate the characteristics of people as well trip origins and destinations.

6.3.2 Limitation in data bias and representativeness

The studies in chapter 3 and 4 sought to represent the urban dynamics over the last mile through analysing behaviours in bike-sharing schemes. However, the issue of data bias and representativeness should be noted. Since not everyone is a bike-sharing user, the findings and conclusions are drawn from a biased sampling of urban dwellers.

Much research has been devoted to answering the question of who uses bike-sharing. In a European (e.g. London) context, young males have higher incomes than average and are more likely to use bike-sharing (Fishman, 2016; Lovelace et al., 2020). Meanwhile, in some American cities (e.g. Washington), the users are more likely to be female with a lower household income (Buck et al., 2013). The user profiles may also change over time. The research of Lovelace et al. (2020) suggests that the London Cycle Hire Scheme has become more inclusive over the last decade, with an increasing amount of usage associated with lower-income areas in the city. In any case, dynamics within the bike-sharing schemes do not account for all the behaviours nor the whole population within the urban last mile. The results should not be over-generalised to infer patterns in private-owned bikes or pedestrian movements. Future work on comparing the patterns in different non-motorised traffic (e.g. pedestrian, scooter) will help to obtain a more nuanced picture of these different sustainable travel modes and the dynamics within. Moreover, a shortcoming of using data of observed activities in transit systems, is that they do not address fundamental preferences or barrier to usage; especially for disadvantaged groups who have limited access to mobility services. For example, in Shanghai, most users of dockless bike-sharing systems are aged between 20 to 40, with relatively high education level and middle-income level, although the rental fee of dockless bike-sharing is considered affordable for most people, with pricing of less than one Chinese yuan per hour, approximately 0.11 British Pound (Lyu et al., 2020). Pricing is not the only issue that prevents the disadvantaged group and older people to use innovative mobility services. Dockless bike-sharing, as well as many Internet ride-hailing services (e.g. Uber), tend to rely on smartphone applications. This inevitably leads to a selection bias. People suffering from digital poverty (e.g. not familiar with the Internet, do not have a smartphone) are being cut off from these mobility services. Disadvantaged groups' travel preference, behaviours, and barriers in using these travel modes are missing in the transaction-generated travel records. The development of and research on new and novel traffic mode should be aware of the underlying data bias issue. And it could be beneficial to combine travel records with survey data to obtain a more comprehensive understanding of different users and non-users, thus making sure the mobility service is inclusive.

6.3.3 Simplification of the spatial graphs

The graphs analysed in this thesis have been constructed using travel OD matrixes, and presented as weighted and directed graph structures. Although edges have attributes to indicate the flow direction as well as trip frequency, they do not include spatial attributes such as distances or travel time between different nodes.

OD graphs in the context of intra-urban travel (e.g. commuting) are found to be relatively less impacted by physical distance and space, as suggested by Chowell et al. (2003). However, cycling is an active travel mode that might be impacted by the distance to a greater extent. The changed spatial distribution of available bikes could also impact the flow structures, for example, not having enough dockless bikes may lead to fewer out-flows from certain places, thereby introducing uncertainties into the analysis and results.

In chapter 3, road segments are cast as graph nodes to aggregate the flexible dockless bike travels; there are other approaches, such as using grid cells, in the literature (Song et al., 2021). The choice is relatively arbitrary and depends on the local characteristics of the road network, urban morphology and the people's mobility pattern. In the case study area of Nanchang (chapter 3), the service catchment of the new metro line does not have a dense road network, especially around the more rural end of it. Therefore, using street networks can help aggregate the movement more effectively, and avoid using a large amount of area-based grid to construct a large and sparse graph.

The modifiable areal unit problem (Openshaw, 1981) should also be mentioned in relation to chapter 3 and 5. Point-based trip origins and destinations are aggregated into lines or regions (clusters), and the aggregation unit could bring uncertainties to the statistics in graph metrics, validation of using different aggregation length/size would help to consolidate the conclusions that are drawn in the two chapters.

Overall, future studies may investigate more aspects of travel flow spatial characteristics; these may include describing physical distance, available space and resources (e.g. parking place and vehicles) and suitable aggregation unit of graph nodes.

6.3.4 Limitation in incorporating more graph indices and models

Despite using different datasets from varied generations of bike-sharing schemes, the graph indices analysed in different chapters are similar. A focus was put on several node-level centrality measures such as degree and PageRank, with several structural-level indices (e.g. assortativity) used to characterise underlying characteristics in connections and the whole structure. The works may be further improved by incorporating additional graph indices. Other metrics (e.g. link betweenness, reciprocity) could be derived from graphs, and they should be explored and understood in the context of the urban and transport system. Moreover, the study in chapter 5 could benefit from evaluating more machine learning models, such as MGCNN, thus to consolidate the conclusion of the utility of graph indices in enhancing various machine learning models. Although LSTM is identified as suitable for dealing with the time-series of graph indicators, the work in chapter 5 does not provide comparisons to other types of RNN (e.g. GRU). There are also other mechanisms that could be incorporated into LSTMs. For example, adding the attention mechanism (Qin et al., 2017) could potentially enhance the model performance by equipping the neural network with the ability to further focus on an important subset of inputs in long sequences. Future works may investigate a wider range of graph indices and machine learning models to quantify their different utilities in understanding and predicting dynamics in transportation and urban systems.

6.4 Conclusion and outlook

In this thesis, spatiotemporal and graph-based approaches are proposed to better understand and model the dynamics over the urban last mile. Through examining flow data that are generated by different schemes and generations of bike-sharing, the research sheds more light on the relationship between people's changing travel behaviours and the associated built environment development and social interventions (e.g. strikes). The revealed mobility patterns and structural changes (e.g. polycentric transformation) imply urban resilience, where bike-sharing can rapidly be adopted by individuals to fill transit gaps that emerged with various spatial and temporal characteristics. A number of planning and management recommendations and strategies are drawn from the analysis presented in this thesis. More importantly, the work proposed that temporal graphs and the derived time-lagged node indices could be helpful to enhance machine learning models (e.g. neural networks) for making predictions on dynamics and demand in urban traffic systems, which benefits sustainable and smart urban development.

As the population in cities continue to grow, the flows and interaction of people and urban space are becoming more complex and diverse. However, the opportunities emerging from technological innovations provide new opportunities to capture detailed intra-urban travel flows at a fine spatial and temporal granularity. This enables the dynamics to be examined and understood in the context of local processes, policies and interventions.

The COVID-19 pandemic might change how people work and travel forever, while an increasing number of people are embracing sustainable options in daily life with greater environmental awareness. Growth and expansions in shared-mobility, especially dockless bike schemes, have been observed worldwide since 2016 (Fishman, 2016). The continued popularity among users and surge of private capital led operations makes the assessment of shared-mobility and "last mile" problems important (Lovelace et al., 2020). Understanding how, where, why, and by whom the sustainable transport is used is essential to ensure that cities and infrastructures are inclusive for all.

To conclude, the successful application in urban planning, transport management and policy-making relies on incorporating the knowledge of the complexity of urban travel flows and activities. The proposed spatiotemporal and graph-based approaches are likely to have considerable potential in practice, and can possibly improve the understanding of people's movement and promote sustainable urban development.

References

Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P. and Borecki, N. 2013. Are bikeshare users different from regular cyclists? A first look at short-term users, annual members, and area cyclists in the Washington, DC, region. *Transportation Research Record.* **2387**(1), pp.112-119.

Chen, Z., van Lierop, D. and Ettema, D. 2020. Dockless bike-sharing systems: what are the implications? *Transport Reviews.* **40**(3), pp.333-353.

Chowell, G., Hyman, J.M., Eubank, S. and Castillo-Chavez, C. 2003. Scaling laws for the movement of people between locations in a large city. *Physical Review E.* **68**(6), p066102.

Fishman, E. 2016. Bikeshare: A review of recent literature. *Transport Reviews.* **36**(1), pp.92-113.

Gu, T., Kim, I. and Currie, G. 2019. To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. *Transportation Research Part A: Policy and Practice.* **119**, pp.122-147.

Kitchin, R. 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography.* **3**(3), pp.262-267.

Lovelace, R., Beecham, R., Heinen, E., Vidal Tortosa, E., Yang, Y., Slade, C. and Roberts, A. 2020. Is the London Cycle Hire Scheme becoming more inclusive? An evaluation of the shifting spatial distribution of uptake based on 70 million trips. *Transportation Research Part A: Policy and Practice.* **140**, pp.1-15.

Lyu, Y., Cao, M., Zhang, Y., Yang, T., & Shi, C. 2020. Investigating users' perspectives on the development of bike-sharing in Shanghai. *Research in Transportation Business & Management.* Advanced online publication. https://doi.org/10.1016/j.rtbm.2020.100543

McKenzie, G. 2020. Urban mobility in the sharing economy: A spatiotemporal comparison of shared mobility services. *Computers, Environment and Urban Systems.* **79**, p101418.

Mooney, S.J., Hosford, K., Howe, B., Yan, A., Winters, M., Bassok, A. and Hirsch, J.A. 2019. Freedom from the station: Spatial equity in access to dockless bike share. *Journal of Transport Geography.* **74**, pp.91-96.

Noble, M., Wright, G., Smith, G. and Dibben, C. 2006. Measuring multiple deprivation at the small-area level. *Environment and Planning A: Economy and Space.* **38**(1), pp.169-185.

Openshaw, S. 1981. The modifiable areal unit problem. In: Wrigley, N., and Bennett, R. J., ed. *Quantitative geography: A British view.* Abingdon: Routledge & Kegan Paul, pp.60-69.

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G. and Cottrell, G. 2017. *A dual-stage attention-based recurrent neural network for time series prediction.* New York: arXiv preprint, p1704.02971.

Saberi, M., Ghamami, M., Gu, Y., Shojaei, M.H.S. and Fishman, E. 2018. Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. *Journal of Transport Geography.* **66**, pp.154-166.

Shaheen, S., Cohen, A., Chan, N. and Bansal, A. 2020. Sharing strategies: carsharing, shared micromobility (bikesharing and scooter sharing), transportation network companies, microtransit, and other innovative mobility modes. In: Deakin, E. ed. *Transportation, Land Use, and Environmental Planning.* Amsterdam: Elsevier, pp.237-262.

Song, J., Zhang, L., Qin, Z. and Ramli, M.A. 2021. A spatiotemporal dynamic analyses approach for dockless bike-share system. *Computers, Environment and Urban Systems.* **85**, p101566.

Sun, L., Jin, J.G., Axhausen, K.W., Lee, D.-H. and Cebrian, M. 2015. Quantifying long-term evolution of intra-urban spatial interactions. *Journal of The Royal Society Interface.* **12**(102), p20141089.

Yang, Y., Heppenstall, A., Turner, A. and Comber, A. 2019. Who, where, why and when? Using smart card and social media data to understand urban mobility. *ISPRS International Journal of Geo-Information.* **8**(6), p271.

Zhang, Y., Aslam, N.S., Lai, J. and Cheng, T. 2020. You are how you travel: A multi-task learning framework for Geodemographic inference using transit smart card data. *Computers, Environment and Urban Systems.* **83**, p101517.

Zhong, C., Arisona, S.M., Huang, X., Batty, M. and Schmitt, G. 2014. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science.* **28**(11), pp.2178-2199.

List of Abbreviations

a: Assortativity

b: Betweenness

AFC: Automatic Fare Collection

ARIMA: Autoregressive Integrated Moving Average model

c: Clustering coefficient

CDF: Cumulative Density Function

CNNs: Convolutional Neural Networks

cv(d): Coefficient of variation of node degree

D: Node degree

$\bar{d}$: Average node degree

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

DT: Model used time-lagged travel demand

E: Edge

FGI: Model uses the full set of time-lagged graph information properties that were identified as more important than the baseline temperature variable; out-strength, in-strength, out-degree, in-degree in and PageRank.

G: Graph

GCT: Grand Central Terminal

GI: Model used time-lagged graph information

GPS: Global Position System

GRU: Gated Recurrent Unit Networks

HA: Historical Average

ID: In-degree

IMD: Index of Multiple Deprivations

IS: In-strength,

IT: Information Technology

KDE: Kernel Density Estimation

L: Links

LCHS: London Cycle Hire Scheme

LSTM: Long Short-Term Memory

δ: Connectivity

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

MGCNN: Multi-Graph Convolutional Neural Networks

MLP: Multi-Layer Perceptron

ML: Machine Learning

N: Node

O-D: Origin and destination

OD: Out-degree,

OS: Out-strength,

PGI: Model uses part of the time-lagged graph information. Out-strength and in-strength are provided for temporal dependence modelling and demand forecasting.

POI: Points of Interest

PR: PageRank

QR code: Quick Response Code

RMSE: Root-Mean-Square Error

RNNs: Recurrent Neural Networks

ST-ResNet: Deep Spatio-Temporal Residual Network

t: Transitivity.

TD: Model uses time-lagged travel demand (out-strength)

TfL: Transport for London

TTWA: Travel to Work Area

$\bar{w}$: Mean link weight

XGBoost: Extreme gradient boosting