JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 31 (2025), pp. 27–56 doi:10.5311/JOSIS.2025.31.365

RESEARCH ARTICLE

Implementing checkpoint-based movement data analysis using cordon networks: A framework validated with transportation case studies

Yaguang Tao, Qian (Chayn) Sun, and Matt Duckham

Department of Geospatial Science, RMIT University, Melbourne, Australia

Received: March 13, 2024; returned: June 12, 2024; revised: April 8, 2025; accepted: September 2, 2025.

Abstract: Checkpoint data are generated by movement past fixed "checkpoints," such as smart-card readers. The heterogeneity of checkpoint data sources, data structures, data models, and granularities means existing research on checkpoint-based movement analytics relies on bespoke or ad-hoc analytics frameworks. This work addresses this gap by developing a consistent framework for analyzing network-constrained checkpoint movement data based on the "cordon network." The cordon network is a simple, graph-based computational structure that captures the underlying spatial structure and heterogeneous granularity of movement through checkpoints. This paper explores the design, development, and testing of an analytics toolkit founded on the cordon network. The approach and its ability to handle heterogeneous checkpoint data added within transportation networks is validated using three diverse transportation case studies. While this study focuses on transportation-related checkpoint data, the discussion and conclusions outline key factors for extending the framework to other domains, providing guidelines for checkpoint movement analysis across different contexts.

Keywords: checkpoint-based movement dataset, cordon network, presence and transaction, network constrained movement, mobility

1 Introduction

This paper demonstrates a movement analytics framework for "checkpoint" movement data where analytics process can be composed with a collection of primitive operations. Based on the operations, it also explores the design and use of an open toolkit for efficient movement analytics using diverse sources of checkpoint movement data. Previous work has distinguished between two distinct types of data about moving objects: *trajectory* data, with position sampled at fixed time-points; and *checkpoint* data, with time sampled at fixed spatial locations called "checkpoints" [3, 6, 14, 24, 27, 44, 47]. This distinction is illustrated in Figure 1. Figure 1(a) illustrates the trajectory-based view, such as the sequence of coordinate positions that might be generated every few seconds using a GNSS (GPS) on-board a vehicle. By contrast, Figure 1(b) illustrates a checkpoint-based view, where the timestamp associated with entering, exiting, or passing fixed spatial locations is recorded. For instance, checkpoint data illustrated in Figure 1(b) might be generated through door access swipe cards, social media "check-ins," connections to WiFi access points, or indeed any combination of these. Many data sources supporting traffic management in particular fall in the category of checkpoint movement data (e.g., wireless sensor network and CCTV, cf. [2, 9]).

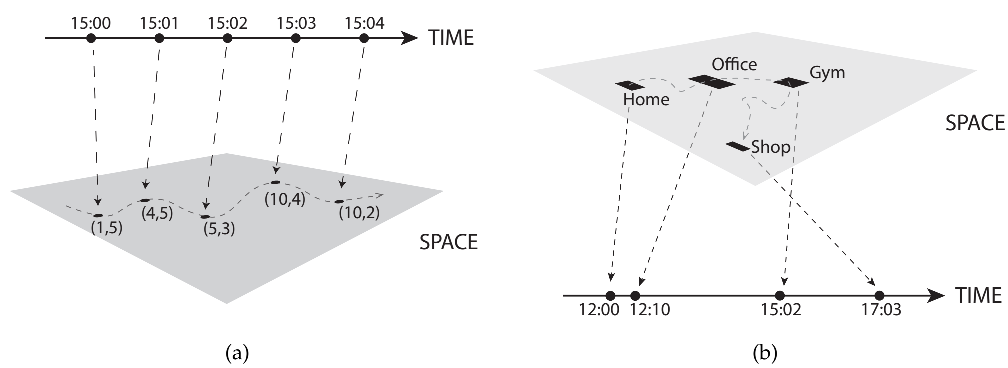


Figure 1: Contrasting trajectory-based and checkpoint-based data about movement: (a) Sample trajectory-based movement (e.g., GPS fixes); (b) Sample checkpoint-based movement (e.g., door access swipe card or social media "check-ins")

The majority of mobility research studies have relied on conventional GNSS (GPS) trajectory or movement data, with relatively fine spatial and temporal granularity. For instance, relative motion analysis [25, 26] uses movement segments, where sample points can be temporally aligned and motion geometry, such as motion azimuth, can be computed. Fine-grained trajectory data is frequently assumed for fundamental movement analytics operations such as segmentation [4, 7] or measuring movement similarity [36, 46]. Such analytics are designed for trajectory, not checkpoint data, based on embedded assumption about fine-enough temporal and spatial granularity to support geometry.

Checkpoint data have, of course, also been studied as the focus of movement analytics research. For instance, Bluetooth nodes have been used to analyze visitor movement at events [10], and sensor arrays have tracked acoustically tagged marine species [43]. Retailers utilize RFID readers to study customer shopping patterns [22, 33], while transit card transactions help model urban mobility [12, 28]. Additionally, cellular and WiFi networks serve as urban mobility checkpoints [15, 21, 32, 42], and traffic cameras monitor vehicle flows for congestion analysis [41]. The COVID-19 pandemic highlighted the value of multi-source mobility data, including QR code check-ins for contact tracing [18, 19, 49]. However, this diversity in data sources and applications has led to significant challenges in movement analytics. Each application typically develops its own bespoke framework, resulting in:

- incompatible data structures and formats across different checkpoint systems;
- difficulty in comparing or combining analyses across different checkpoint types;
- redundant development of similar analytics tools; and
- limited reusability of analytical methods across applications.

Thus, a key challenge addressed by this work and its predecessor [44] is to design a generic framework capable of representing and analyzing the diversity of checkpoint movement data that arises in these types of applications. While the previous work introduced a data model to encapsulate this diversity, the present study extends beyond structural representation by developing a comprehensive analytics framework. This framework enables rapid and scalable analysis of checkpoint movement data, integrating adaptable analytical functions tailored to different mobility datasets. By bridging the gap between data representation and analytical processing, this work facilitates a more unified and efficient approach to movement analytics, significantly advancing the practical applicability of checkpoint data analysis.

2 Literature review

Two key characteristics of checkpoint movement data present both challenges and opportunities for movement analytics: the heterogeneity of data representation and the salience of checkpoint locations.

2.1 Heterogeneity in checkpoint data

Checkpoint data, generated from diverse sources such as door access swipe cards, social media check-ins, WiFi connections, mobile phone records, credit card transactions, public transport smart cards, and electronic tolling, presents significant opportunities for movement analytics. Often generated as a by-product of essential services, this data offers wide-area coverage. However, its heterogeneity—stemming from differences in formats, models, and structuring—poses a key challenge, as analytics designed for one dataset often fail to generalize to others.

Past studies have adopted dataset-specific approaches to analyze checkpoint data. For instance, Delafontaine et al. [10] examined Bluetooth device movements through fixed sensors, modeling movements as sequences of bluetooth nodes within three-minute episodes and applying sequence matching to extract behavior patterns. Public transport movement has been analyzed as "trip chains" based on smart cards swipes [28, 31], while GPS trajectories have been represented as sequences of spatial regions of interest [16]. These methods rely on dataset-specific structuring, aggregation, and interpretation, limiting their applicability across different movement datasets.

2.2 Salience of checkpoint data

Checkpoint data is often captured at irregular spatial and temporal granularity, resulting in large gaps between check-ins. Unlike GPS trajectory data with fixed sampling frequency, which supports tasks like move-stop segmentation [13] and location inference via interpolation[17], checkpoint data lacks uniform sampling, making similar analyses nontrivial. The spatial checkpoint locations may also be at varying levels of precision, including regions (such as wireless hotspots), linear borders (such as crossing the boundary between two adjacent mobile phone cells), and more truly point-like "gates" (such as specific doors in a building).

Despite these challenges, checkpoint data frequently records *salient* locations—entry points to buildings, transport hubs, and intersections—implicitly encoding meaningful contextual information. For instance, traffic monitoring sites are strategically placed at congestion-prone areas, and public transport touch-on/off records inherently capture transit activity. Similarly, RFID systems are deployed as key locations for targeted tracking. Kurazono et al. [23] used RFID readers at seabird nests to analyze homing patterns, Schneider et al. [40] tracked honeybee foraging behavior with RFID-guided feeding structures, and Mandal et al. [29] monitored road congestion with RFID readers at intersections to measure travel times.

In comparison, trajectory data must be explicitly joined with other spatial data to enrich movement data with semantics. Studies such as [5, 35, 50] segment trajectories into stops and moves before associating them with contextual information, often using fixed spatial and temporal thresholds [37]. Such processing is unnecessary for checkpoint data, as its coarser sampling naturally partitions movement into meaningful segments, and checkpoints often have associated metadata (e.g., addresses) that facilitate direct integration with other datasets.

2.3 Network-constrained checkpoint-based movement

The focus of this work is restricted to network-constrained movement (i.e., movement performed in network-based infrastructure), as it is amongst the most common types of checkpoint data such as movement through transportation networks. Examples of network-constrained checkpoint-based movement include vehicles observed passing traffic intersections or electronic tolling, or movement through a security-controlled building.

Networks constraints are important not only for human movement, but also for animal movement. Movement ecologists, for example, understand the importance of connectivity of habitats in an ecosystem. Dendritic habitat structure, such as found in streams, hedgerows, and cave networks, imposes network constraints on population processes such as spread, growth, and survival [8, 34]. To study the persistence of riverine populations, for example, the river network topology has to be considered, such that the population of adult fish downstream is correlated with adult fish from upper stream and local young fish [30]. More widely, Jacoby and Freeman [20] has reviewed how networks have become important to analysis in movement ecology. For instance, graph metrics, such as connectivity, can indicate the importance of a specific habitat or land fragment, in turn helping in prioritizing habitat conservation.

Prior work has shown the potential of checkpoint data but remains fragmented due to dataset-specific solutions. The cordon network model[44] provides a promising foundation but lacks standardized operations and a reusable analytics framework. This work addresses these gaps by:

- demonstrating how the cordon network can provide a single consistent basis for movement analytics across a diversity of checkpoint movement data;
- designing a set of primitive movement analytics operations on movement data in cordon network;
- designing and developing an open-source Python software package to enable efficient and rapid analytics for checkpoint movement data; and
- exploring, through specific examples of analytics of checkpoint data, the key factors and guidelines important to successful checkpoint movement data analytics.

3 Checkpoint data analytics with cordon networks

Our analytics framework consists of data modeling component that integrates movement data, movement checkpoints, movement constraints, and movement analysis. An overview workflow for the framework is shown in Figure 2. Data about network structure and checkpoint locations is firstly used to construct a cordon network (described further below in Section 3.1). Checkpoint movement data, being either presence or transaction records, is then assigned to spatial checkpoint locations in the network. Within this consistent framework, analytics operations can also be applied, and ultimately results mapped or exported.

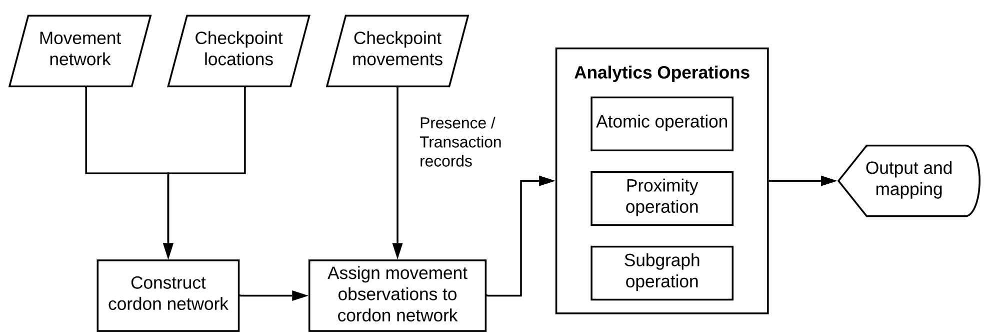


Figure 2: Overview workflow for the analytics framework.

The cordon network itself is the central abstraction that provides the ability to represent and analyze the diversity of checkpoint data sources in a single, consistent framework, independent of the specific geometry or granularity of the movement.

3.1 Cordon network construction

Data about movement past checkpoints is inextricably linked with the environment in which movement occurs. For instance, in a cell phone network, space is divided into services areas (Voronoi regions) of cell towers (Figure 3a). The irregular regions can be represented as a network, where each region is a node, and adjacency between regions forms the edges (Figure 3b). This adjacency graph is referred as a "cordon network."

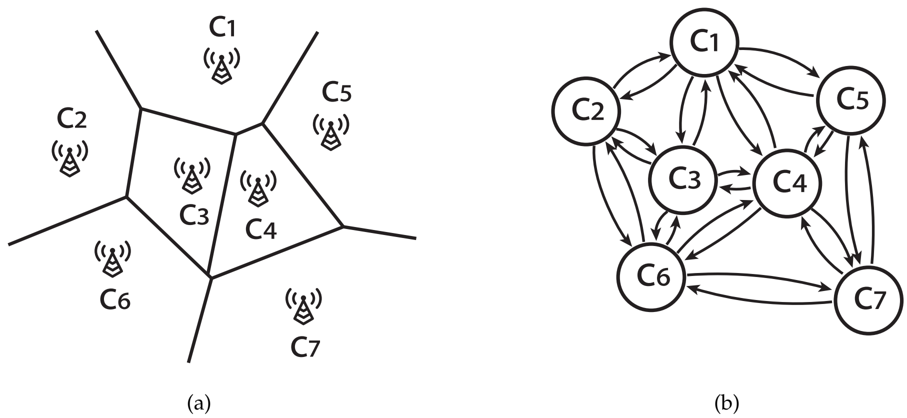


Figure 3: Coarse-grained and irregular regions and their corresponding adjacency graph (cordon network): (a) A cell phone network where the service area of different cell towers form a irregular partition of the space. (b) The cordon network based on this partition where each node represents a region and directed edges represent region adjacency.

A node in a cordon network corresponds to a spatial region (e.g., covered by a surveillance camera, cellphone, or wifi access point), and an edge represents the transition between two regions. This network structure can be applied to any movement scenario, whether in Euclidean space or a transportation network. For example, Figure 4 shows a road network partitioned into regions by camera checkpoints, which also form a cordon network.

3.2 Movement observations as presence or transaction records

The diversity of checkpoint data generated today derives fundamentally from two complementary types of movement "fixes," called *transaction records* and *presence records* and described in more detail in Tao et al. [44].

In brief, transaction records capture the *instantaneous* location and state of an object as it moves through space. Such records are generated when an object moves past a checkpoint, such as swipe card security system or an electronic tollbooth, transitioning from one partitioned region to an adjacent region.

In contrast, presence records capture the time-period *during* which an object is located within a region. Such records might be generated by video footage of a person traversing a monitored area of a city or the period in which a cellphone was within a particular mobile cell.

The choice of which types of record—presence and/or transaction—are used in a particular application or data set is somewhat arbitrary, because the two representations encode essentially the same information and are usually inter-convertible. As we shall see later,

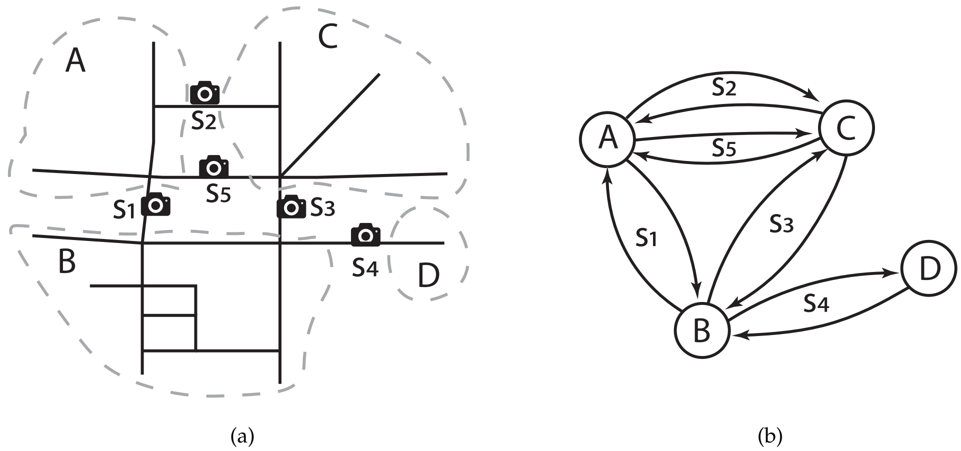


Figure 4: Space monitored by transaction sensors and its corresponding cordon network. (a) A street network environment fitted with surveillance cameras which record any vehicles passing by. The cameras partition the space into regions that within each a vehicle can move freely without being observed. (b) The cordon network with each region as a node and each camera as a pair of directed edges.

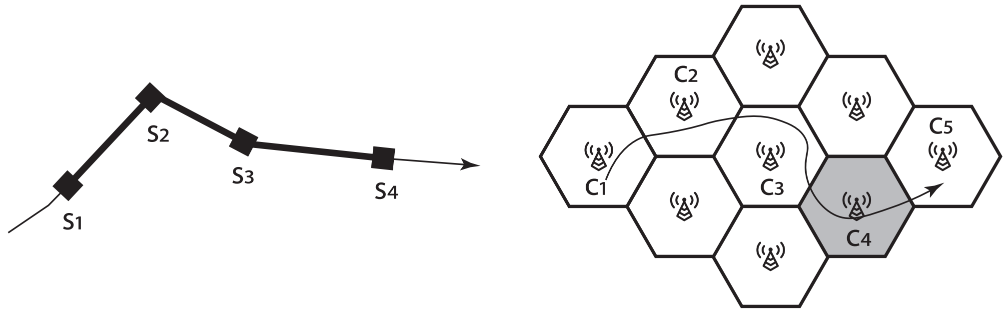
converting between these two representations are basic operations for a checkpoint data analytics system.

In practice, the characteristics of the sensors used to generate checkpoint data may lend themselves more "naturally" to one or the other of presence or transaction record. For example, Figure 5(a) depicts a section of toll road with four electronic toll sensors. Such a scenario lends itself naturally to a transaction record representation, recording the times at which vehicles pass the toll gates. By contrast, the cell phone network example in Figure 5(b) lends itself more naturally to the presence record representation, recording the time periods during which a vehicle traversing a region is contained within each cell. However, the alternate representation is always possible (e.g., a transaction-based representation of movement through a cellular network recording the times at which a mobile phone transitioned between one cell and the next, or a presence-based representation of movement in a network recording the period during which a vehicle was between two toll gates).

Together with the cordon network, presence and transaction records enable us to model the full range of checkpoint based movement. The cordon network represents the distinct areal or network regions in space (nodes) together with the checkpoints or boundaries that partition that space into distinct regions (edges). Presence records are therefore associated with nodes, and capture the time period during which each object was in a region. Transaction records capture the instant at which an object moved from one region to another.

3.3 Primitive checkpoint data analytics operations

Armed with the cordon network as the unifying conceptual model for checkpoint data, we turn next to the analytics operations afforded by such data. Reflecting the heterogeneous nature of checkpoint movement data, there exists a similar diversity of analytics operations



(a) (b)

Figure 5: Examples of "natural" transaction records (a) (e.g., electronic road tolling), and presence records (b) (e.g., cellphone trace)

that may be meaningful. However, the cordon network's graph structure makes neighborhood an obvious mechanism to classify different operation types.

In this paper, we propose an analytics framework based on neighborhood which is akin to the map algebra, familiar in raster data analytics [45]. The three types of operation in map algebra are: local operations, which compute the output value of a raster cell based on the input value of that cell alone; focal operations, which compute the output value of a target raster cell based on the input values of a coincident cell and its immediate neighbors; and zonal operations, which compute the output value of a target cell based on the input values of a coincident cell and all other cells in an arbitrarily specified region (zone) [45].

Map algebra uses the raster cell as the spatial unit, and adjacency between as the basis of neighborhood. Taking the map algebra as an analogy, nodes and edges in cordon network can be regarded as spatial units; and incidence between nodes and edges as defining neighborhood. The analogy provides a convenient basis for classifying checkpoint movement analytics operations into three categories: atomic operations, proximity operations and subgraph operations.

3.3.1 Atomic operations

Atomic operations concern analytics of movement data associated with a single cordon-network edge or node, much as map algebra local operations concern a single raster cell. For example, Figure 6 illustrates the atomic operation of computing the flow at each edge over time, based on the raw observations of movement at a cordon edge (transaction records). The observations of three individual objects moving from B to A (i.e., a transaction on edge S_1, Figure 6a) are summarized as numbers of objects passing over time periods (i.e., flows, Figure 6b). An example application of such an atomic operation would be to calculate a transport hotspot map, where individual vehicle movement records are summarized as a flow map.

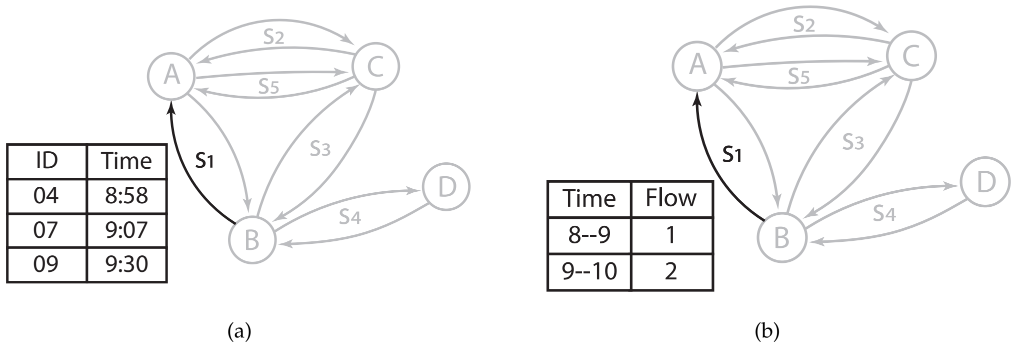


Figure 6: The atomic operation of calculating flows based on transaction records. (a) The raw transaction records at a particular edge S_1 capture the ID of an passing object and a timestamp. (b) Output of the operation, aggregating hourly traffic flow.

3.3.2 Proximity operations

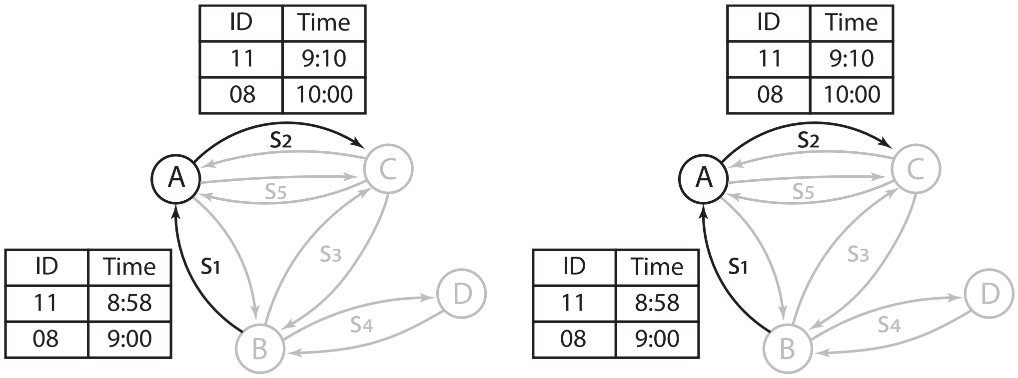
Most analyses depend not only on local movement information, but also on data about movement in the immediate vicinity. A proximity operation, the analog of the map algebra focal operation, is designed for this type of analysis. In defining a proximity operation, the neighborhood has to be specified first. For example, one can use the following extended definition of distance δ_{en} to compute the distance between network elements (nodes and edges) based on the standard definition of network distance $\delta(x, y)$ between nodes (i.e., the number of adjacent edges in the shortest path through the network between x and y).

$$\delta_{en}(x,y) = \begin{cases} 0 & \text{if } x = y \\ 2 \cdot \delta(x,y) & \text{if } x \text{ and } y \text{ are nodes} \\ 2 \cdot \min\{\delta(a_x, y), \delta(b_x, y)\} + 1 & \text{if } y \text{ is a node and } x \text{ is an edge } (a_x, b_x) \\ 2 \cdot \min\{\delta(x, a_y), d(x, b_y)\} + 1 & \text{if } x \text{ is a node and } y \text{ is an edge } (a_y, b_y) \\ 2 \cdot \min\{\delta(a_x, a_y), \delta(a_x, b_y), & \\ \quad \delta(b_x, a_y), \delta(b_x, b_y)\} + 2 & \text{if } x \text{ and } y \text{ are edges } (a_x, b_x) \text{ and } (a_y, b_y) \end{cases}$$

In Figure 3(b), for example, the network distance $\delta(C_2, C_5) = 2$. However, taking into account the edges in the extended network distance, with an edge denoted as a pair of nodes in parentheses, $\delta_{en}(C_2, C_5) = 4$ and $\delta_{en}((C_2, C_1), C_5) = 3$ and $\delta_{en}((C_2, C_1), (C_1, C_5)) = 2$. In short, our extended distance metric counts traversal of both edges and nodes as part of the total distance between cordon nodes/edges.

A proximity operation with a neighborhood of size m will involve analysis of movement information at all network components within a extended network distance of m or less, termed "m-adjacent." Thus, proximity operations are generalizations of atomic operations, with a 0-adjacent proximity operation equivalent to a atomic operation.

The actual size of a neighborhood required for a given proximity operation will be dependent on the specific application. For example, Figure 7 shows the proximity operation with $\delta_{en} \leq 1$ to calculate the duration of time moving objects stayed in each region, based



(a) (b)

Figure 7: Example 1-adjacent proximity operation, computing duration of presence from transaction records. (a) The original movement in transaction records (e.g., object 11 passed S_1 at 8:58 and S_2 at 9:10). (b) The proximity operation output calculated, with object 11 remaining in region A for 12 minutes.

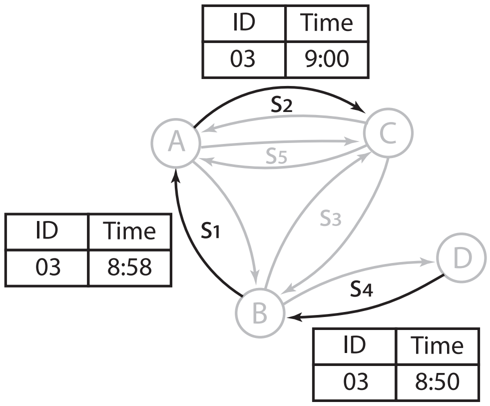
on the transaction records of neighboring transaction sensors. A corresponding operation can be defined for presence records, with transaction records calculated from 1-adjacent nodes.

An example usage of a basic proximity operation might be to identify sink and source nodes in the cordon network, where sink nodes have only inflow movements and source nodes only outflow movements.

3.3.3 Subgraph operation

Finally, subgraph operations are similar to proximity operation but based on arbitrary subgraphs, akin to "zonal" neighborhoods in map algebra. The semantics of the zones are highly dependent on specific application scenarios. Summarizing movement based on moving object identity is one instance of a subgraph operation. For example, Figure 8(a) shows all the transaction records related to the movement of object 03. A subgraph operation might compute the length of movement traces, using the entire recorded path of object 03 as the zone. More complicated subgraph operations might include computing which location has the longest total recorded journeys passing by, and hence might be a suitable site to place a new resting area.

In summary, this framework can model a wide variety of checkpoint data from diverse domains with associated basic analytics operations. For example, outlier detection might be implemented as an atomic operation (e.g., detecting a vehicle having unusually long or short duration in a node. In other cases, outliers might involve identifying anomalous movement between adjacent nodes (2-adjacent proximity) or arbitrary routes or regions (subgraph).



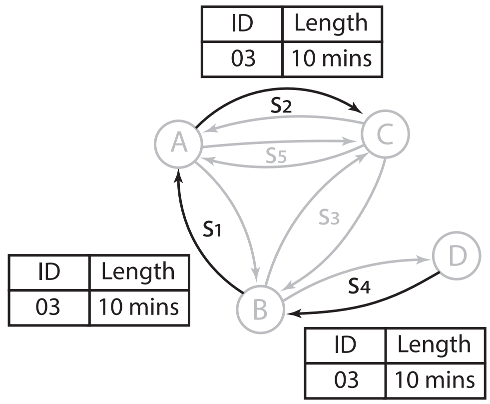


Figure 8: Computing the length of movement traces as a subgraph operation. (a) All edges passed by object 03 form a single zone. (b) The result of a subgraph operation on edge S_1 that the total duration of the journey passing S_1 of object 03 is 10 minutes.

4 Checkpoint analytics toolkit design

Our open-source analytics toolkit[1] is founded on the cordon network, and is structured around two complementary perspectives of movement (presence and transaction records) and three types of analytics operations (atomic, proximity, subgraph) described above (Figure 2). The toolkit facilitate five key stages in the analytics workflow (Figure 2): cordon network construction; checkpoint movement data input; graph database representation; analytics operations; and output and mapping.

4.1 Cordon network construction

A first step to checkpoint data analytics is to define the cordon network structure. A user may input the cordon network structure directly. More frequently, one may compute the cordon network based on spatial data about checkpoint locations (e.g., coordinates) and the movement environment (e.g., transportation network). For example, for movement in a transportation network and known checkpoint locations an algorithm for constructing the induced cordon network is given in Tao et al. [44]. Alternatively, computing the adjacency graph from a set of Voronoi regions is a standard process in any spatial database or GIS. This functionality of constructing cordon network based on checkpoint locations and movement environment is packed into the proposed toolkit.

There exist a range of well-known data structures that can be used for representing the connectivity of networks, such as adjacency lists and adjacency matrices [48]. We have successfully adopted both normalized tables, most commonly found in any spatial database. However, in this paper we describe an alternative implementation using a graph database (Section 4.5). Regardless, edge directions, lengths/weights, node weights, and in particular

[1]Source code: https://github.com/OzzyTao/cordon_network

geometries can also be associated to each network component. In most cases, users will opt to include network geometries, and this is indeed supported by our implementation.

4.2 Checkpoint movement data input

Once a cordon network is established, movement observation data acquired by various fixed sensors can then be referenced to the corresponding cordon network component. We assume users are able to format checkpoint data as relational tables such that each row represents one presence/transaction observation.

Fundamentally, a transaction record is a tuple $\langle s, t, d \rangle$ where s is a spatial reference to the location of the checkpoint (such as the identity of an edge in the cordon network or the coordinate location of a checkpoint); t is the timestamp of the observation; and d is some data about the movement, such as the identity of an object passing checkpoint s at time t.

In contrast, a basic presence record is a tuple $\langle s, t_s, t_e, d \rangle$, where t_s and t_e are the timestamps of the time period (start t_s and end t_e) during which d was observed.

As discussed previously, transaction and presence records encode essentially the same information, and so the choice of which to enter is arbitrary, usually based on the most "natural" representation provided by the checkpoint sensors themselves (cf. Figure 5). The analytics operations to convert between the two are available in our implementation. For example, a door swipe-card transaction sensor might record person P entering a room R at door D at 9am and exiting (D') at 10am (two transaction records, $(D, 9\text{am}, P)$ and $(D', 10\text{am}, P)$. Equivalently, a camera sensor might record person P remaining in room R between 9 and 10am (one presence record, $(R, 9\text{am}, 10\text{am}, P)$.

The implemented toolkit helps in storing the movement data and linking the records to corresponding cordon network components.

4.3 Analytics on the cordon network

Movement analytics operations can be performed based on combining checkpoint movement records and the cordon network. As described above in Section 3.3, analytics operations can be classified into three categories, based on the extent of neighborhoods required by the operation:

- Atomic operations, such as summary movement statistics for transaction/presence sensors.
- Proximity operations, such as computing presence duration of moving objects based on transaction records.
- Subgraph operations, such as trip statistics or summaries for moving objects.

The implemented toolkit supports the specification and execution of operations by providing python classes that users may inherent to customize proximity definition (for proximity operation and subgraph operation) and computation logic on each unit of computation.

4.4 Output and mapping

While the focus of this work is not mapping and visualization, our toolkit does provide basic mapping capabilities. The implementation offers just two basic visual parameters for the user to map quantities onto each cordon network component: color and weight for point and line, color and transparency for region. A wealth of previous work, however, already exists for visual communication of movement data (cf. [11]).

4.5 Graph database implementation

There already exist a range of relational database structures use to store graphs or multigraphs, such as the cordon network, in a spatial database [48]. Essentially, a relational database requires at minimum a table of edges, composed of start node and end-node identities, to encode the information in a graph. These can be used as the basis of an implementation of our framework, and indeed this option has been successfully developed and tested. However, graphs are not a natural structure to store as a relational table, and so retrieving information from a normalized relational table typically relies on multiple queries and join operations, for example to find whether two nodes are connected by a path.

In contrast, a graph database does offer a natural structure to store networks of semantic dependencies between entities with explicit relationships [38], such as the cordon network. Because of the network-oriented nature of graph databases, many graph database operations are more closely aligned with common queries on the cordon network, in particular concerning paths and connections between nodes and edges. Consequently, the cordon network analytics toolkit has also been implemented using a graph database, specifically, Neo4j[2]. Where SQL is the standard query language for relational databases, in a graph database analytics operations can be encoded as Cypher graph database queries. With reference to our three primitive analytics operation types, the graph database is defined in the following way:

- Cordon network nodes and edges are both modeled as nodes in the graph database, labeled "Cordon-Node" and "Cordon-Link" correspondingly;
- Subjects of movement are also modeled as nodes with a label of "Mover";
- Cordon network connectivity determines the relationships between "Cordon-Node" and "Cordon-Link" for the graph database (named "LINKS");
- Movement objects are related to "Cordon-Node" and "Cordon-Link" with relationships "STAYS" and "PASSES."

The graph database structure for the specific example introduced in Figure 4, is shown in Figure 9. In the database, the movement of object M_1 (e.g., a vehicle) is represented as relationships between the "Mover" and "Cordon-Node"/"Cordon-Link" nodes. Presence and transaction records are stored as properties of "STAYS"/"PASSES" relationships. For example, a "STAYS" relationship contains a start time and an end time. Similarly, contextual information related to movement can be stored as properties for nodes or relationships.

An advantage of using the graph database over a relational database for checkpoint analytics is the ease with which neighborhood can be defined and queried. The neighborhoods for the three primitive analytics operation types can be specified with Cypher query language as follows:

- *Atomic operations*: summarize information contained at individual "Cordon-Node" or "Cordon-Link". For example, to find all the moving objects that have stayed in a cordon network node "A":

[2] https://neo4j.com/

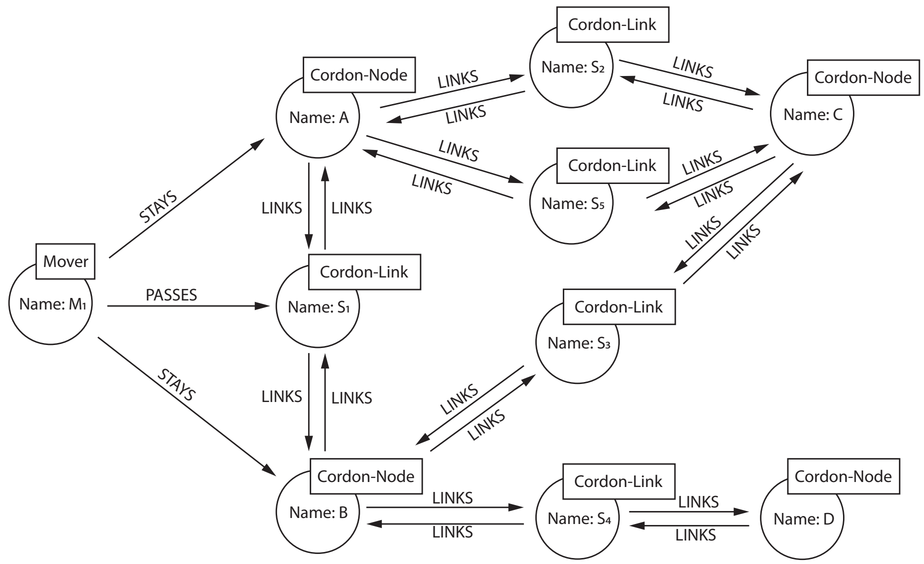


Figure 9: A labeled property graph data model for graph database: Moving object (e.g., Mover M_1) and movement environment are modeled in a single graph model.

```
MATCH (m:Mover)-[l:STAYS]->({name:"A"})
RETURN m, l
```

- *Proximity operations*: summarize based on a k-adjacent neighborhood of a size specified by the number of LINKS relationships contained in a path between pairs of nodes/edges. For example, to find all the objects that have moved withing a k-adjacent region of cordon network node "A":
```
MATCH ({name:"A"})-[:LINKS*1..k]-(n)<-[l:PASSES|:STAYS]-(m:Mover)
RETURN m, l
```
- *Subgraph operations*: summarize based on an arbitrary region specified by a subgraph. For example, to find all moving objects within a cordon network region specified by subgraph "G1":
```
MATCH ({subgraph:"G1"})<-[l:PASSES|:STAYS]-(m,Mover)
RETURN m, l
```

Thus, this analytics framework integrates the high-level conceptual model of the cordon network directly with a low-level graph database implementation. High-level queries, such as retrieving all objects moving through a region of the network, can be queried efficiently with a standard graph database query language. By including this graph database implementation in the toolkit, it offers a more flexible and efficient way for a user to customize different types of operations.

5 Case studies: Transportation movement checkpoint analytics

To validate our approach and show how it can be applied to a diversity of heterogeneous datasets, transportation network usage analysis using three different public transportation case studies with checkpoint movement data at different granularities were explored. The Zhengzhou bus dataset offers a typical example of aggregated commuting patterns, where numbers but not identities of passengers getting on and getting off at different stops are recorded. The Melbourne Myki dataset gives finer granularity insights into movement of individual passengers in a transportation network, where different journeys of anonymized individual users can be linked. The Sydney train occupancy data set has the coarsest granularity data, with train occupancy only for different journey segment and in ordinal values.

The three datasets are also heterogeneous in that the Myki and Zhengzhou datasets are natively in the form of transaction records (objects moving past a checkpoint); while the Sydney train occupancy data is natively in the form of presence records (occupancy of trains between two checkpoints). Table 1 summarizes the key characteristics of the datasets.

Table 1: Comparison of case study datasets

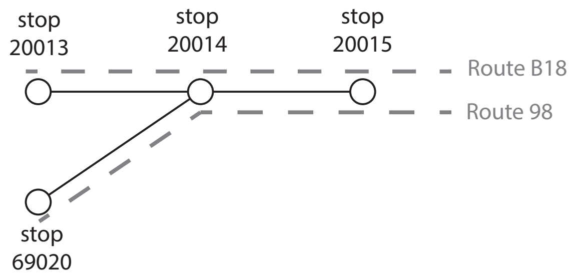
Dataset	Data source	Data records	Transaction or presence records	Application
Melbourne Myki	Passenger touch-on/off at stations	Individual passengers	Transaction	Station usage; route segment usage;
Zhengzhou bus	Passenger touch-on/off on vehicle	Group passenger counts	Transaction	Traffic speed estimation; bus route usage;
Sydney train	Observations on train	Ordinal occupancy level	Presence	Route segment usage; station net flow;

5.1 Zhengzhou aggregated bus passenger flow data

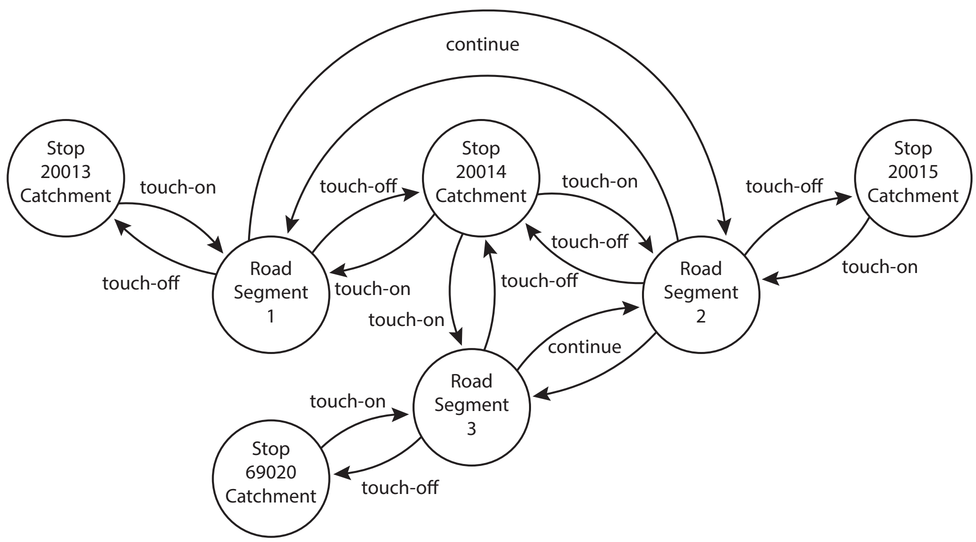
The Zhengzhou bus passenger dataset consists of passenger flow counts at each individual stop in Zhengzhou, both for boarding and alighting passengers. Each individual record in the dataset contains a bus stopping event: a timestamp; the identity of the bus; the identity of the bus stop; and the number of passengers touching-on or touching-off the bus at that stop.

To capture passengers' movement, the bus network is represented using our cordon network. Figure 10 shows part of Zhengzhou bus network and how it is modeled as a cordon network. The cordon network uses nodes to represent places where objects can remain present for a period of time (i.e., buses and stop catchments) and edges for instantaneous transactions between those places (i.e., touching on and off). Figure 10(a) shows two bus routes (i.e., Route B18 and Route 98) joining together at stop 20014 using a conventional planar map-based view. Figure 10(b) shows the cordon network, with individual touch-on/touch-off transaction records mapped to edges between nodes (stops and route segments).

While individual passenger movements are not available in the data, the movement flows still provide insights into bus movements in the road network. For example, the average speed for each road segment can be calculated by considering the time interval between



(a)



(b)

Figure 10: Part of bus network and its cordon network representation: (a) two bus routes (i.e., Route B18 and 98) travels pass stop 20013, 20014, 20015 and stop 69020, 20014, 20015 respectively; (b) the cordon network consists of nodes of both route segment and stop catchment and edges of touch-on/touch-off events and bus arrival/departure/continue events;

consecutive transactions. In terms of our primitive movement analytics operations, average speed can be constructed as a proximity operation comparing records between coincident edges. Figure 11 illustrates the results of just such a proximity average speed analysis, with the darker a road segments indicating faster average vehicle speed. As might be expected, this simple analysis shows slower traffic movement near intersections and in regions with dense road network. While individual buses can be probes for the traffic system, this speed profile can be further used to model general traffic speed.

Figure 12 shows the results of a different operation: the number of passengers using different bus routes (i.e. Route 30, 31 and 204) on a specific road segment over time. This

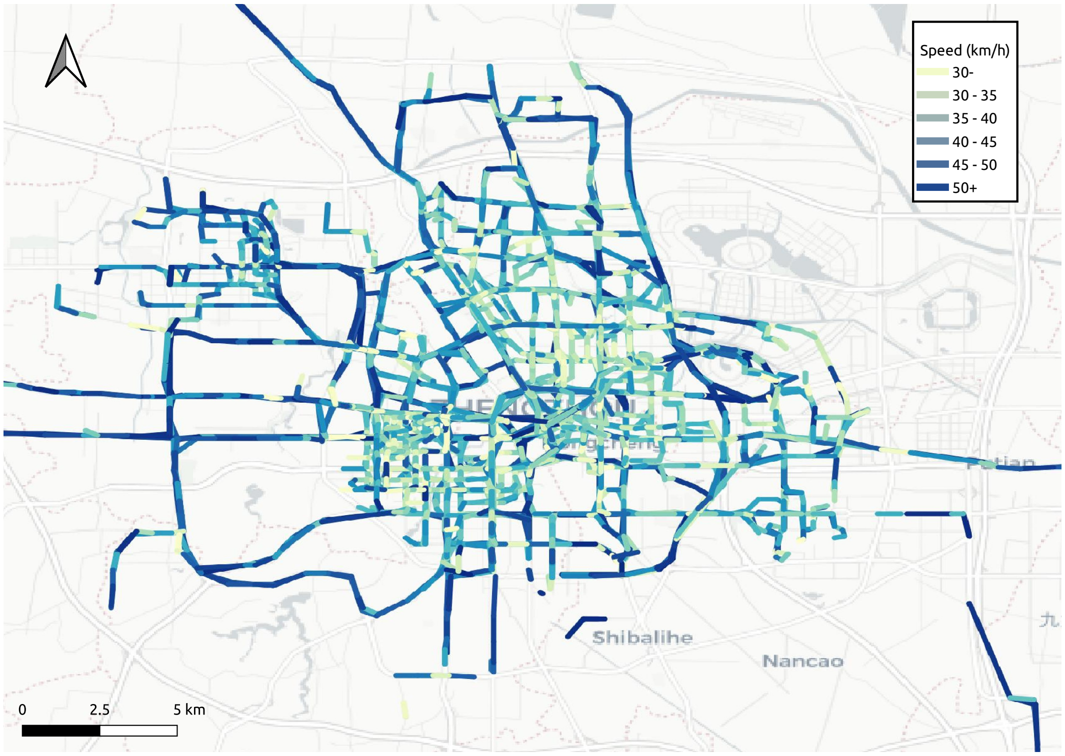


Figure 11: Zhengzhou bus speed map: the darker color a street segment is in, the higher speed a bus travels with on that segment.

analysis requires a subgraph operation, aggregating movements across multiple stops that make up the route segment (subgraph). From Figure 12 it is evident that the demand for Route 30 is constantly high over all periods the buses are operating. However, demand for Route 31 are mainly in the afternoon, while there's only significant demand for Route 204 at night.

5.2 Melbourne train transit card data

Myki is the smart card system for traveling with public transport within Victoria, Australia. For the train network, Myki requires users to touch on or touch off when entering or leaving a station. Like many sources of checkpoint data, data is collected primarily for fare collection purposes rather than movement analytics.

The physical location of Myki card readers for the Melbourne train network, at the entry to stations, means that our cordon network is subtly different than the one for the Zhengzhou bus environment (Figure 10). Having card readers at the entrance to train stations means that, in many cases, different train lines are not partitioned into different regions on the cordon network. In fact, once a passenger enters a train station, he or she can move freely throughout the train network until he or she exits.

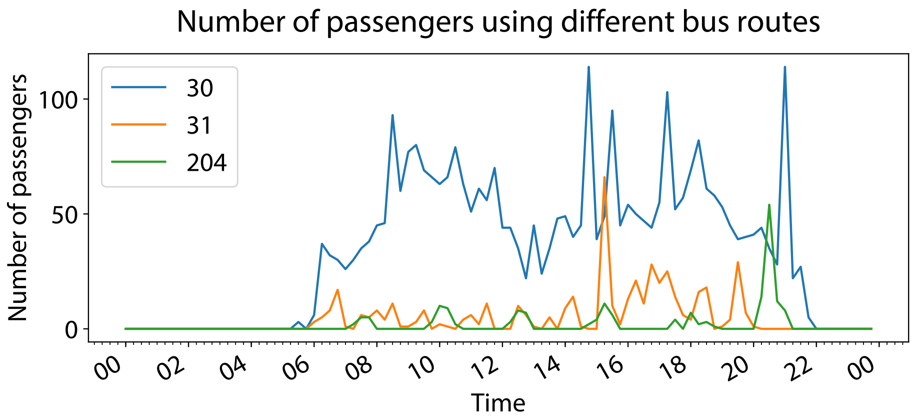


Figure 12: Passenger flow of different bus routes sharing a same road segment

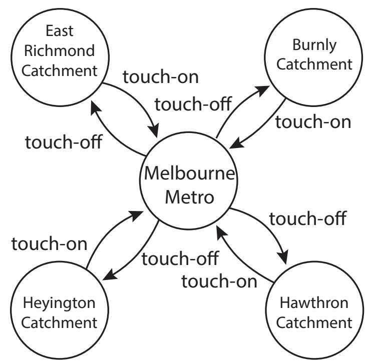
Figure 13(a) captures this as a cordon network, such that the entire train network is represented as a single a cordon node. Thus, although the transaction records are more detailed than Zhengzhou, capturing individual passengers, the spatial granularity is coarser with no ability to distinguish in general between different train lines.

However, some additional assumptions about passenger movement can be made to provide some insights into likely movements. For instance, let us assume a passenger always takes the shortest path to go from his or her origin to destination. Further, let's assume passengers tend to get onto the earliest train available on their route once they have entered a station.

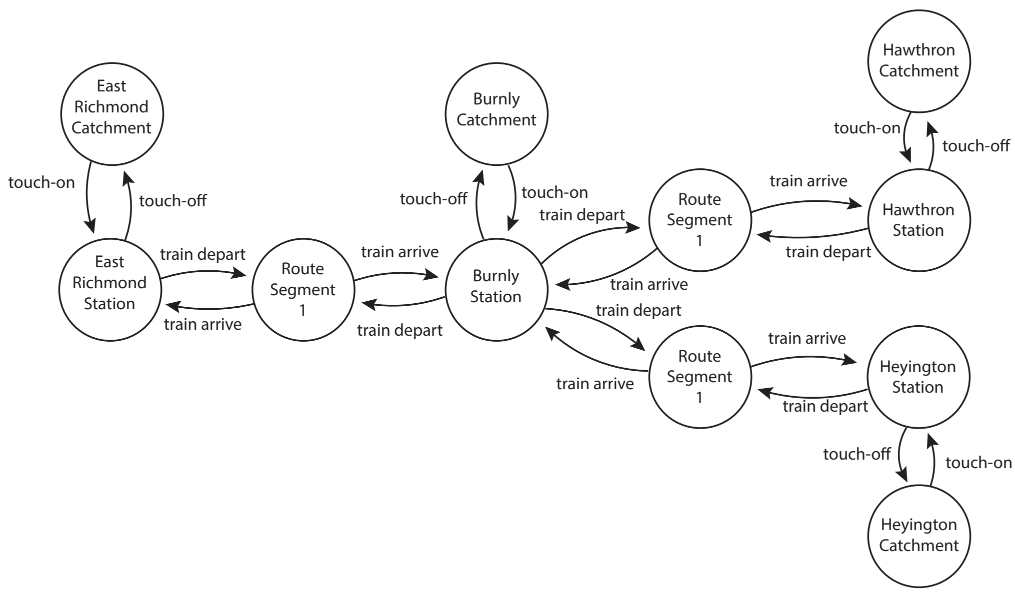
The cordon network makes it straightforward to relate and integrate such information into the analysis. Checkpoints are located at salient places in the network, i.e., entry or exit from train stations. Combining the train network with passenger and train service schedule data, it is possible to construct a cordon network of a finer granularity (showed in Figure 13(b)). In this case, each station is represented with a node for its platform area with another node for the catchment region of that station. Route segments are then also nodes, containing passenger movement. The edges between nodes then capture observed movements of passengers touching on and off at stations, with inferred movements between stations along routes, based on our knowledge of the underlying network connectivity and our assumptions about passengers taking the earliest, shortest route to their destination.

The most obvious analysis with our Myki data set is to extract stop usage information. This can be generated using a simple atomic operation, giving a summary of touch on/off over some time interval (e.g., per day, per hour, per minute, or so forth). Figure 14(a) illustrates the analysis results for morning passenger boarding flow (green circles in the figure) and morning passenger alighting flow (red circles in the figure). The size of the dots indicate the magnitude of the passenger flows, highlighting a clear trend of the morning commute to the major business districts.

While the raw data is transaction records, route segment usage statistics can be calculated with a 1st order proximity operation. The time when a passenger was traveling on a particular segment can be calculated based on the neighboring transaction records. Conse-

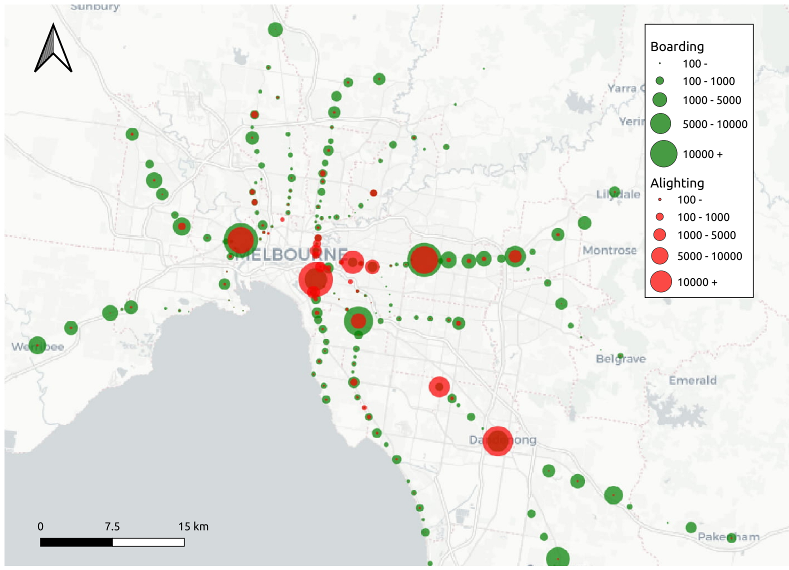


(a)

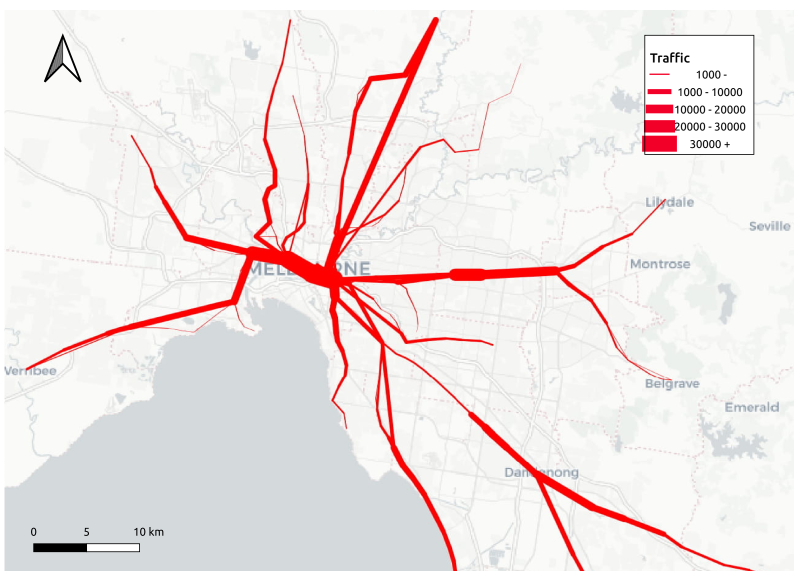


(b)

Figure 13: Cordon network of a train network monitored with Myki system: (a) models passengers' movement based on only touch-on/touch-off observations; (b) models passengers' movement near a station with both touch-on/touch-off observations and train schedule.

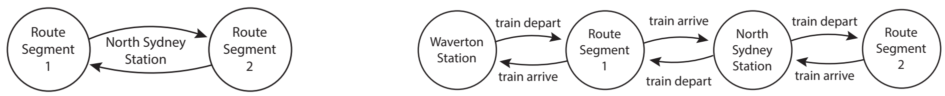


(a)



(b)

Figure 14: Melbourne Myki data movement pattern: (a) Morning origins and destinations extracted directly from local transaction records based on atomic operations: green dots are the places passengers start their morning journeys while red dots are the destinations with the size of the dots representing the volume of passengers; (b) Volume of passengers using each train route segment as the presence perspective: the thicker a line is, there are more passenger using the train route segment.



(a) (b)

Figure 15: Part of Cordon network for the Sydney train occupancy dataset: (a) model for train movement; (b) model for train passenger movement;

quently, the accumulated flow for each route segment for an entire day can be calculated, as shown in Figure 14(b). Different time frames could of course be specified to determine passenger flows over different time periods using the same method.

5.3 Sydney train occupancy data

Transport for NSW released a set of datasets of train occupancy level[3] that records occupancy level (i.e. "MANY_SEATS_AVAILABLE," "FEW_SEATS_AVAILABLE," "STANDING_ROOM_ONLY") on each train run between every two stops. Each record then contains the train identity, train occupancy level, the next station, and a timestamp for arriving at that station.

For this coarse granularity data, only train arrival time is provided. Hence, one option is to model each station as a single bidirected edge pair connecting route segments modeled as nodes, as in Figure 15(a). Such an approach has coarse spatial and temporal granularity, representing train stopping events as near-instantaneous transactions. However, at a finer spatiotemporal granularity, modeling each station as a node incident with two bidirected edge pairs enables explicit representation of the presence of the train and its passengers at the station (see Figure 15(b)). Such a representation is more faithful to the fine detail of the application, and allows the model to distinguish the arrival of the train at the station (as an in-edge) and the subsequent departure of the train after a time interval.

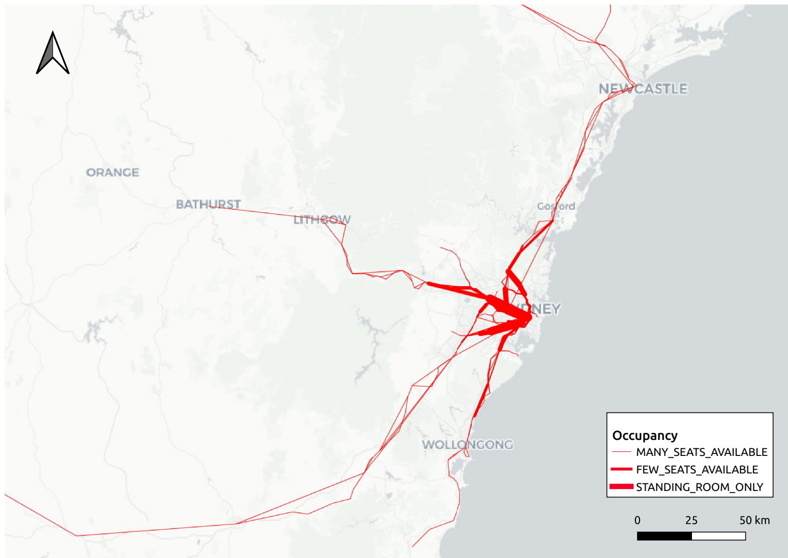
For example, the arrival time of a train at North Sydney Station marks a transaction between two regions (i.e., nodes): the route segment the train was traveling on and the station it has arrived at. Occupancy data is then captured as a presence record for each route segment containing train identity, time interval (time the train left the previous station and time it arrived at the next station), and recorded occupancy level.

In this case, given train occupancy level of each individual train journey, the average train occupancy level for each segment over time can be easily calculated with an atomic operation, as shown in the map Figure 16(a).

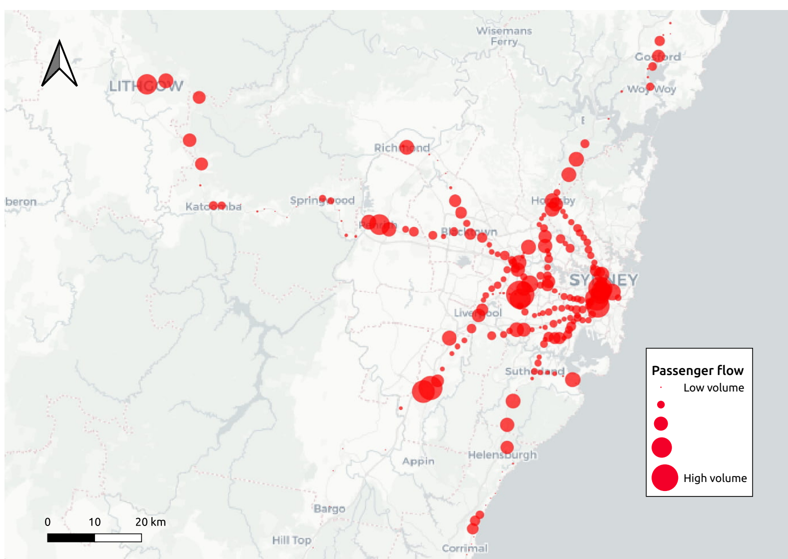
Based on the difference between two consecutive route segments, train station usage information can be deduced. Computing this analysis requires a second order proximity operation to consider all the neighboring route segments of a train station in cordon network. The result of such a proximity operation is visualized in Figure 16(b), showing net change in passenger flow at the station. It is noticeable from Figure 16(b) that some major train stops in Sydney CBD have small dots, perhaps indicating balanced passenger flow that does not change the overall occupancy level for trains passing these stops. However,

[3] https://opendata.transport.nsw.gov.au/dataset/train-occupancy-nov-2016-feb-2017

for other distant or less popular stops, a low occupancy change might suggest that there is only a small number of passengers getting on or off trains.



(a)



(b)

Figure 16: Sydney train passenger flow movement pattern: (a) train route occupancy level: the original presence observations attached to train route segments. The thicker a line is, more crowded a train segment is; (b) passenger flow at train stations: the size of a dot presents the volume of passenger flow.

5.4 Summary

We use the three case studies to demonstrate the heterogeneity of checkpoint data about movement, including transaction observations (i.e., Zhengzhou and Melbourne cases) and presence observations (i.e., the Sydney example). The three case studies also show diversity of granularity, in terms of both spatial, temporal, and attribute granularity. The Melbourne Myki dataset describes movement for individual passengers, but at relatively coarse spatial granularity. Conversely, the Zhengzhou dataset describes movement with relatively fine spatial granularity, but coarse passenger-flow attributes. The Sydney dataset only provides fine spatial granularity, but the coarsest attribute information about movement—train occupancy in qualitative measure.

Regardless of these differences, in each case the cordon network provides a single consistent model for organizing and analyzing checkpoint movement data. As shown in the Melbourne Myki case, the structure also facilitates the inclusion of information from multiple sources, such as train timetables, to refine the model.

6 Discussion

The previous section illustrates the diversity of checkpoint movement data, and demonstrates the flexibility of the cordon network to adapt to that diversity. Each individual case study is in itself not an especially challenging analytics task. However, the challenge this approach meets is the ability to analyze highly heterogeneous checkpoint movement data within a single, efficient toolkit underpinned by the simple, consistent conceptual model of the cordon network.

In all of the examples an important first step in using the cordon network is deciding whether to represent data as presence and/or transaction records, and the semantics of cordon network nodes and edges that follow from that decision. The following subsections review the three major factors that can contribute to making that decision:

6.1 Underlying sensor types

In many cases, the characteristics of the underlying sensor used to generate the data will lend itself more "naturally" to representation as an instantaneous (transaction) or durative (presence) observation. Location through access to Wifi hotspots (e.g., [39]), room occupancy measured by CO_2 sensors (e.g., [1]), or our train occupancy data provide observations in a region over time, and so naturally lend themselves to representation of presence records. In contrast, passing an electronic road toll gantry or using a smart card to pass through an access control gate, such as our Myki card example, are all events associated with a point in time. Such observations are near-enough instantaneous to be regarded as transaction records.

In general, where a sensor captures movement over a spatial region (such as an area of the plane or an edge in a transportation network) then presence records are a natural choice. Where a sensor captures movement past a point or a boundary then a transaction record is usually more natural.

6.2 Spatial and temporal granularity

The spatial and temporal granularity of checkpoint movement data, and the desired queries over that data, also affect the choice of how movement is modeled. For example, the underlying Sydney train data measures occupancy level at each station upon arrival. Hence, generalizing each station to a single edge connecting adjacent route segments, as in Figure 15(a), is an option for a coarser-grained representation of that application. Doing so has the advantage of conceptual simplicity, with one entity in the cordon network (a pair of bidirected edges) representing a station and another (nodes) representing route segments between stations.

However, this level of generalization also makes it harder to faithfully represent passenger movements on and off the train. In particular, it is not possible to represent connecting services at stations using the coarse-grained approach in Figure 15(a). To be able to represent the possible movement of passengers alighting from one train and then boarding another train at the same station requires finer spatial and temporal granularity than an edge and associated transaction records in the cordon network. Instead, the approach taken in Figure 15(b) explicitly represents the presence of the train and its passengers at the station as a node, distinguishing the arrival of the train at the station from its departure.

6.3 Movement entity

Finally, the semantics of the observed movement entity can affect the choice of whether to model movement as presence or transaction records. Where individual moving objects can be discerned apart, such as through an individual's smart phone or smart cards, representation through both presence or transaction records are usually possible. However, where the movement entity is an aggregate, such as groups or counts of individual moving objects, then presence records may be the more natural representation. Counts of objects are typically in a region or over time (i.e., flows), and so lend themselves more to durative or region-based presence records.

In addition, the identification of individual movement entity further affects the type of analytics operations can be applied. Without a persistent ID for each individual movement entity, multiple movement segments by the same entity can not be links. Therefore, it is more often to apply atomic or proximity operations on such movement data to categorize locations. For instance, the Sydney train occupancy data does not provide commuter's ID or any information related to movement of an individual. The opportunity resides in analyzing local group movement patterns, e.g., traffic of a train station. On the other hand, when persistent IDs are available, it becomes possible to characterize individual movement entity through long-term movement records. In such cases, subgraph operation becomes more useful as long-term movement is more likely to cover a larger spatial extent. This, however, does not suggest a fixed mapping relation between movement entity and analytics operations. Instead, the choice of analytics operations are determined by the logic of individual analytic tasks. For instance, subgraph operation was used to study bus routes with Zhengzhou bus passenger flow data, where individual passengers cannot be identified.

6.4 Privacy and ethical considerations

Movement data analysis, particularly in urban contexts, raises significant privacy and ethical concerns. It is crucial to ensure that sensitive data is handled responsibly, anonymized appropriately, and analyzed in compliance with relevant ethical guidelines and regulations. While our framework does not provide explicit mechanisms for data anonymization—this responsibility typically falls on the data owner—it does enhance the utility of aggregated data sources, such as Sydney train occupancy data, for analytical purposes. By enabling meaningful insights from aggregate datasets, our approach reduces the reliance on privacy-sensitive individual movement records. This, in turn, supports privacy-conscious analytics while still enabling valuable transportation and mobility studies.

7 Conclusion

This paper has explored the cordon network as a single, consistent model of checkpoint-based movement in transportation networks. Checkpoint-based movement data has distinctly different characteristics from more familiar trajectory data, such as GPS traces. In particular, the diversity of different types of checkpoint data and semantics has led to a similar diversity of analytics frameworks.

The paper demonstrates that this diversity of different movement data can be modeled and analyzed in a single consistent model, based on presence and transaction records associated with nodes (spatial regions) and edges (transitions between regions) in the cordon network.

The paper also presents a practical, open-source implementation of our cordon network framework used to perform a range of checkpoint-based movement analytics and basic mapping and visualization. Movement analytics in the approach structured into three types of operation, akin to those in the map algebra. Atomic operations analyze data at individual nodes and edges (akin to local operations in the map algebra); proximity operations combine data from neighboring nodes and edges (akin to focal operations in the map algebra); subgraph operations combine data over arbitrary regions, such as bus routes or suburbs (as for subgraph operations). The graph database implementation of cordon network allows network-based analytics operations using neighborhood to be computed efficiently.

Ultimately, increasing consistency of modeling and analysis increases the potential for sharing, comparison, and benchmarking of analytics operations and results. It also assists in more rapid exploration and analysis of checkpoint data, enabling analysts to spend less time building tools and analyses and more time understanding the patterns and drivers of movement in different applications.

This work, however, has not explored the full potential of cordon network. For simplicity, the application datasets discussed in this work are all constrained by networks. The network constraints together with checkpoint movement sensors partition movement environment into connected regions cordoned off by sensors. Essentially, any movement performed in an environment of connected regions cordoned off by movement sensors can be modeled with cordon network. An example of unconstrained movement data is CDR data of cellular network. Another issue has not been explored is the robustness of the model against imperfect sensors/movement records. The design of cordon network is based on the assumption that a fixed location movement sensor (i.e., checkpoint) is able to detect all movement within its range. However, this is sometimes not true in reality. For instance, a mobile user's movement may not be fully reflected by CDR records since CDR may only disclose locations where phone calls are made. Other movement data sources where movement tracking requires a user's active involvement (e.g., location-based social network check-ins) may also be hard be modeled with cordon network. Therefore, we have identified three possible future directions:

- Exploring the utility of cordon network for movement data other than network-constrained is one of the future directions of this work;
- Modeling sensor errors and imperfect records;
- Developing more GIS-user friendly query language based on Cypher, which offer shortcuts for frequent movement queries.

Disclosure statement

The authors report there are no competing interests to declare.

Data availability statement

The Sydney train occupancy dataset is openly available at https://opendata.transport.nsw.gov.au/dataset/train-occupancy-nov-2016-feb-2017. Other datasets that support the findings of this study are available from the corresponding author, Y.T., upon reasonable request.

References

[1] ANG, I. B. A., SALIM, F. D., AND HAMILTON, M. Human occupancy recognition with multivariate ambient sensors. In *Proc. IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (2016), pp. 1–6. doi:10.1109/PERCOMW.2016.7457116.

[2] ARTIKIS, A., WEIDLICH, M., SCHNITZLER, F., BOUTSIS, I., LIEBIG, T., PIATKOWSKI, N., BOCKERMANN, C., MORIK, K., KALOGERAKI, V., MARECEK, J., ET AL. Heterogeneous stream processing and crowdsourcing for urban traffic management. In *EDBT* (2014), vol. 14, pp. 712–723. doi:10.1007/978-3-662-44845-8_49.

[3] AVCI, B., TRAJCEVSKI, G., TAMASSIA, R., SCHEUERMANN, P., AND ZHOU, F. Efficient detection of motion-trend predicates in wireless sensor networks. *Computer Communications 101* (2017), 26–43. doi:10.1016/j.comcom.2016.08.012.

[4] BILJECKI, F., LEDOUX, H., AND VAN OOSTEROM, P. Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science 27*, 2 (2013), 385–407. doi:10.1080/13658816.2012.692791.

[5] BOGORNY, V., KUIJPERS, B., AND ALVARES, L. O. ST-DMQL: a semantic trajectory data mining query language. *International Journal of Geographical Information Science 23*, 10 (2009), 1245–1276. doi:10.1080/13658810802231449.

[6] BOTH, A., DUCKHAM, M., LAUBE, P., WARK, T., AND YEOMAN, J. Decentralized monitoring of moving objects in a transportation network augmented with checkpoints. *The Computer Journal 56*, 12 (2012), 1432–1449. doi:10.1093/comjnl/bxs117.

[7] BUCHIN, M., DRIEMEL, A., VAN KREVELD, M., AND SACRISTÁN, V. Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. *Journal of Spatial Information Science 2011*, 3 (2011), 33–63. doi:10.5311/JOSIS.2011.3.66.

[8] CAMPBELL GRANT, E. H., LOWE, W. H., AND FAGAN, W. F. Living in the branches: Population dynamics and ecological processes in dendritic networks. *Ecology Letters 10*, 2 (2007), 165–175. doi:10.1111/j.1461-0248.2006.01007.x.

[9] CHEN, Z., ELLIS, T., AND VELASTIN, S. A. Vehicle detection, tracking and classification in urban traffic. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on* (2012), IEEE, pp. 951–956. doi:10.1109/ITSC.2012.6338852.

[10] DELAFONTAINE, M., VERSICHELE, M., NEUTENS, T., AND VAN DE WEGHE, N. Analysing spatiotemporal sequences in bluetooth tracking data. *Applied Geography 34* (2012), 659–668. doi:10.1016/j.apgeog.2012.04.003.

[11] DEMSAR, U., BUCHIN, K., CAGNACCI, F., SAFI, K., SPECKMANN, B., DE WEGHE, N. V., WEISKOPF, D., AND WEIBEL, R. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology 3*, 5 (2015). doi:10.1186/s40462-015-0032-y.

[12] DEVILLAINE, F., MUNIZAGA, M., AND TRÉPANIER, M. Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2276 (2012), 48–55. doi:10.3141/2276-06.

[13] DODGE, S., WEIBEL, R., AND LAUTENSCHÜTZ, A.-K. Towards a taxonomy of movement patterns. *Information visualization 7*, 3-4 (2008), 240–252. doi:10.1057/PALGRAVE.IVS.9500182.

[14] DUCKHAM, M., VAN KREVELD, M., PURVES, R., SPECKMANN, B., TAO, Y., VERBEEK, K., AND WOOD, J. Modeling checkpoint-based movement with the earth mover's distance. In *Geographic Information Science* (2016), J. A. Miller, D. O'Sullivan, and N. Wiegand, Eds., Springer, pp. 225–239. doi:10.1007/978-3-319-45738-3_15.

[15] FINA, S., JOSHI, J., AND WITTOWSKY, D. Monitoring travel patterns in German city regions with the help of mobile phone network data. *International Journal of Digital Earth 14*, 3 (2021), 379–399. doi:10.1080/17538947.2020.1836048.

[16] GIANNOTTI, F., NANNI, M., PINELLI, F., AND PEDRESCHI, D. Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (2007), ACM, pp. 330–339. doi:10.1145/1281192.1281230.

[17] GÜTING, R. H., DE ALMEIDA, T., AND DING, Z. Modeling and querying moving objects in networks. *The VLDB Journal—The International Journal on Very Large Data Bases 15*, 2 (2006), 165–190. doi:10.1007/s00778-005-0152-x.

[18] HU, T., WANG, S., SHE, B., ZHANG, M., HUANG, X., CUI, Y., KHURI, J., HU, Y., FU, X., WANG, X., ET AL. Human mobility data in the COVID-19 pandemic: Characteristics, applications, and challenges. *International Journal of Digital Earth* (2021). doi:10.1080/17538947.2021.1952324.

[19] HUANG, X., LI, Z., JIANG, Y., YE, X., DENG, C., ZHANG, J., AND LI, X. The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the US during the COVID-19 pandemic. *International Journal of Digital Earth 14*, 4 (2021), 424–442. doi:10.1080/17538947.2021.1886358.

[20] JACOBY, D. M., AND FREEMAN, R. Emerging network-based tools in movement ecology. *Trends in Ecology and Evolution 31*, 4 (2016), 301–314. doi:10.1016/j.tree.2016.01.011.

[21] JIA, C., DU, Y., WANG, S., BAI, T., AND FEI, T. Measuring the vibrancy of urban neighborhoods using mobile phone data with an improved PageRank algorithm. *Transactions in GIS 23*, 2 (2019), 241–258. doi:10.1111/tgis.12515.

[22] JUNG, I.-C., AND KWON, Y. S. Grocery customer behavior analysis using RFID-based shopping paths data. *World Academy of Science, Engineering and Technology 59* (2011), 2011. doi:10.5281/zenodo.1081169.

[23] KURAZONO, H., YAMAMOTO, H., YAMAMOTO, M., NAKAMURA, K., AND YAMAZAKI, K. RFID and ZigBee sensor network for ecology observation of seabirds. In *Advanced Communication Technology (ICACT), 2013 15th International Conference on* (2013), IEEE, pp. 211–215.

[24] LAUBE, P. *Computational Movement Analysis*. Springer, 2014. doi:10.1007/978-3-319-10268-9.

[25] LAUBE, P., IMFELD, S., AND WEIBEL, R. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science 19*, 6 (2005), 639–668. doi:10.1080/13658810500105572.

[26] LAUBE, P., VAN KREVELD, M., AND IMFELD, S. Finding REMO–detecting relative motion patterns in geospatial lifelines. In *Developments in spatial data handling*. Springer, 2005, pp. 201–215. doi:10.1007/3-540-26772-7_16.

[27] LONG, J. A., WEIBEL, R., DODGE, S., AND LAUBE, P. Moving ahead with computational movement analysis. *International Journal of Geographical Information Science 32*, 7 (2018), 1275–1281. doi:10.1080/13658816.2018.1442974.

[28] MA, X., WU, Y.-J., WANG, Y., CHEN, F., AND LIU, J. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies 36* (2013), 1–12. doi:10.1016/j.trc.2013.07.010.

[29] MANDAL, K., SEN, A., CHAKRABORTY, A., ROY, S., BATABYAL, S., AND BANDYOPADHYAY, S. Road traffic congestion monitoring and measurement using active RFID and GSM technology. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on* (2011), IEEE, pp. 1375–1379. doi:10.1109/ITSC.2011.6082954.

[30] MARI, L., CASAGRANDI, R., BERTUZZO, E., RINALDO, A., AND GATTO, M. Metapopulation persistence and species spread in river networks. *Ecology Letters 17*, 4 (2014), 426–434. doi:10.1111/ele.12242.

[31] MCGUCKIN, N., AND NAKAMOTO, Y. Trips, chains and tours-using an operational definition. In *National Household Travel Survey Conference* (2004).

[32] MENESES, F., AND MOREIRA, A. Large scale movement analysis from WiFi based location data. In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2012), IEEE, pp. 1–9. doi:10.1109/IPIN.2012.6418885.

[33] NAKAHARA, T., AND YADA, K. Analyzing consumers' shopping behavior using RFID data and pattern mining. *Advances in Data Analysis and Classification 6*, 4 (2012), 355–365. doi:10.1007/s11634-012-0117-z.

[34] NATHAN, R. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences 105*, 49 (2008), 19050–19051. doi:10.1073/pnas.0808918105.

[35] PARENT, C., SPACCAPIETRA, S., RENSO, C., ANDRIENKO, G., ANDRIENKO, N., BOGORNY, V., DAMIANI, M. L., GKOULALAS-DIVANIS, A., MACEDO, J., PELEKIS, N., ET AL. Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR) 45*, 4 (2013), 42. doi:10.1145/2501654.2501656.

[36] RANACHER, P., AND TZAVELLA, K. How to compare movement? a review of physical movement similarity measures in geographic information science and beyond. *Cartography and geographic information science 41*, 3 (2014), 286–307. doi:10.1080/15230406.2014.890071.

[37] RENSO, C., BAGLIONI, M., DE MACEDO, J. A. F., TRASARTI, R., AND WACHOWICZ, M. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and information systems 37*, 2 (2013), 331–362. doi:10.1007/s10115-012-0511-z.

[38] ROBINSON, I., WEBBER, J., AND EIFREM, E. *Graph databases*. O'Reilly Media, Inc., 2013.

[39] SAPIEZYNSKI, P., STOPCZYNSKI, A., GATEJ, R., AND LEHMANN, S. Tracking human mobility using WiFi signals. *PLoS ONE 10*, 7 (2015), e0130824. doi:10.1371/journal.pone.0130824.

[40] SCHNEIDER, C. W., TAUTZ, J., GRÜNEWALD, B., AND FUCHS, S. RFID tracking of sublethal effects of two neonicotinoid insecticides on the foraging behavior of Apis mellifera. *PloS one 7*, 1 (2012), e30023. doi:10.1371/journal.pone.0030023.

[41] SHI, Y., DENG, M., GONG, J., LU, C.-T., YANG, X., AND LIU, H. Detection of clusters in traffic networks based on spatio-temporal flow modeling. *Transactions in GIS 23*, 2 (2019), 312–333. doi:10.1111/tgis.12521.

[42] SI, H., WANG, Y., YUAN, J., AND SHAN, X. Mobility prediction in cellular network using hidden Markov model. In *2010 7th IEEE Consumer Communications and Networking Conference* (2010), IEEE, pp. 1–5. doi:10.1109/CCNC.2010.5421684.

[43] STEHFEST, K. M., PATTERSON, T. A., DAGORN, L., HOLLAND, K. N., ITANO, D., AND SEMMENS, J. M. Network analysis of acoustic tracking data reveals the structure and stability of fish aggregations in the ocean. *Animal behaviour 85*, 4 (2013), 839–848. doi:10.1016/j.anbehav.2013.02.003.

[44] TAO, Y., BOTH, A., AND DUCKHAM, M. Analytics of movement through checkpoints. *International Journal of Geographical Information Science 32*, 7 (2018), 1282–1303. doi:10.1080/13658816.2017.1397675.

[45] TOMLIN, C. D. *Geographic information systems and cartographic modeling*. No. 526.0285 T659. Prentice Hall, 1990.

[46] TOOHEY, K., AND DUCKHAM, M. Trajectory similarity measures. *Sigspatial Special 7*, 1 (2015), 43–50. doi:10.1145/2782759.2782767.

[47] WANG, J., DUCKHAM, M., AND WORBOYS, M. A framework for models of movement in geographic space. *International Journal of Geographical Information Science 30*, 5 (2016), 970–992. doi:10.1080/13658816.2015.1078466.

[48] WORBOYS, M., AND DUCKHAM, M. *GIS: A Computing Perspective 2e*. CRC Press: Boca Raton, FL, 2004. doi:10.4324/9780203481554.

[49] WU, J., XIE, X., YANG, L., XU, X., CAI, Y., WANG, T., AND XIE, X. Mobile health technology combats COVID-19 in China. *Journal of Infection 82*, 1 (2021), 159–198. doi:10.1016/j.jinf.2020.07.024.

[50] YAN, Z., CHAKRABORTY, D., PARENT, C., SPACCAPIETRA, S., AND ABERER, K. SeMiTri: a framework for semantic annotation of heterogeneous trajectories. In *Proceedings of the 14th international conference on extending database technology* (2011), ACM, pp. 259–270. doi:10.1145/1951365.1951398.