Computers & Graphics (2024)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Topic modelling for spatial insights: uncovering space use from movement data

Gennady Andrienko[a,b,*], Natalia Andrienko[a,b], Dirk Hecker[b]

[a] City, University of London, London EC1V 0HB, United Kingdom
[b] Fraunhofer Institute IAIS, Sankt Augustin, 53757 Germany

ARTICLE INFO

Article history:
Received April 16, 2024

Keywords: Visual Analytics, Movement

ABSTRACT

We present a novel approach to understanding space use by moving entities based on repeated patterns of place visits and transitions. Our approach represents trajectories as text documents consisting of sequences of place visits or transitions and applies topic modelling to the corpus of these documents. The resulting topics represent combinations of places or transitions, respectively, that repeatedly co-occur in trips. Visualisation of the results in the spatial context reveals the regions of place connectivity through movements and the major channels used to traverse the space. This enables understanding of the use of space as a medium for movement. We compare the possibilities provided by topic modelling to alternative approaches exploiting a numeric measure of pairwise connectedness. We have extensively explored the potential of utilising topic modelling by applying our approach to multiple real-world movement data sets with different data collection procedures and varying spatial and temporal properties: GPS road traffic of cars, unconstrained movement on a football pitch, and episodic movement data reflecting social media posting events. The approach successfully demonstrated the ability to uncover meaningful patterns and interesting insights. We thoroughly discuss different aspects of the approach and share the knowledge and experience we have gained with people who might be potentially interested in analysing movement data by means of topic modelling methods.

1. Introduction

Understanding how people, vehicles, or animals utilise space as they move through different environments is important for various applications, such as urban planning, transportation management, wildlife conservation, and many others. Analysis of movement data plays a crucial role in uncovering patterns and trends in space use.

Generally, analysis of movement data can concentrate on three core dimensions: the moving entities, the spatial realms they traverse, and the temporal dynamics of their motion [1]. The selection of analytical methods is contingent upon the dimension of interest. In this paper, we focus on the spatial dimension, with a primary objective to reveal interconnections between areas in the underlying space through the trajectories of moving objects traversing them. To accomplish this, we explore the applicability and analytical potential of topic modelling techniques.

Traditionally, places are characterised according to points of interest they include, events that happened in places, or categories of moving objects visiting them, taking into account the timing of visits [2]. In this work we characterise places according to their connectivity to (or accessibility from) other places, as inferred from trajectories. Analysis of connectivity between places and accessibility of places is an important task in urban planning, transportation systems, animal ecology and ge-

*Corresponding author:
e-mail: gennady.andrienko.1@city.ac.uk (Gennady Andrienko)

ographical analysis (e.g., [3, 4]). Traditional approaches have primarily focused on physical infrastructure, landscape, habitat fragmentation, and distance-based measures to understand connections between locations. However, in recent years, there has been a growing recognition that connectivity encompasses more than just physical proximity.

A novel approach gaining prominence is to examine linkages between places based on the trajectories that traverse these places (e.g., [5, 6]). This new approach offers several advantages over traditional methods. By considering trajectories instead of solely relying on distances or pathway topology, it provides a more comprehensive understanding of how places are interconnected. Trajectories capture the movement patterns and flows of people, animals, goods, and information, thereby uncovering hidden connections and interdependencies between places. By analysing common trajectories, researchers can gain insights into the functional relationships and shared characteristics of different locations. This approach aligns with the increasing availability of large-scale mobility data and advancements in data analysis techniques.

There are two aspects of space use by moving objects: which places are visited and how the objects proceed from place to place. Respectively, our problem statement includes two major questions addressing these aspects.

Problem statement. Given a set of discrete places and a set of trajectories of moving entities that visited these places, answer the following questions:

1. What groups of places are highly interconnected as evidenced by frequent co-appearance in trajectories of entities visiting them?
2. What are the primary transitions taken by the entities as they move from one place to another and so connect the places?

To answer these questions, we represent trajectory data in two complementary ways: as a list of visited places (to address question 1) and as a list of moves (transitions) between the places (to address question 2). Using these representations, we investigate and compare two distinct approaches to analysing interconnections among places or among transitions. The *first approach* relies on a numerical measure that assesses the extent of linkage between places or transitions based on their co-occurrence in the same trajectories. Subsequently, this measure is employed in embedding and clustering techniques. In contrast, the *second approach* treats the lists of places or transitions as textual documents. These document-like lists can be subjected to analysis using topic modelling methods. An advantage of the second approach is its capacity to uncover complex structures beyond pairwise relationships, revealing interconnected patterns composed of multiple places or transition links.

An early version of this work was initially presented at the EuroVA 2023 workshop [7]. The workshop contribution is reiterated in Section 4.3 of the current paper. Subsequently, we undertook a more comprehensive exploration of the abilities of topic modelling in analysing space use and connectivity. In parallel, we conducted a comparative evaluation of this approach with the use of a quantitative measures of place interconnectivity within embedding and clustering methods. We also extended the scope of our investigation beyond network-constrained trajectories to movement data with different properties. Here we report our case studies on applying topic modelling to episodic trajectories derived from georeferenced social media posts and the unconstrained movements of a ball during a football game. We also provide an extensive discussion of the practical strengths and limitations of employing topic modelling as a tool for spatial analysis using movement data.

Contribution. The primary contribution of this paper lies in the presentation of a novel approach to analysing space use by moving entities. The presentation includes a demonstration of the capabilities of the approach in application to three distinct examples, an empirical comparison with widely used analysis strategies employing embedding or clustering methods, and an extensive discussion covering various aspects of our approach.

In the following Section 2, we introduce the key concepts that are necessary for presenting our approach. After considering the related work in Section 3, we introduce in Section 4 our approach on a previously studied mobility data set with known properties and compare it with applications of embedding and clustering to the same data (Section 4.2). Next, we present two use cases where we apply the approach to two real-world data sets of human mobility expressed through social media activity (Section 5) and analysis of football tactics based on ball movement trajectories (Section 6). Section 7 with a detailed discussion of our approach and lessons learnt concludes the paper.

2. Background

Regardless of the methods used for collecting movement data, these data sets typically consist of records that include four elements: the entity in motion, spatial position, time, and (optionally) associated attributes [1, 8]. The three fundamental dimensions in movement data analysis are the population of moving entities, the spatial dimension, and time. Consequently, the analysis of movement data can center on one of these essential dimensions while considering the others. In this research, our primary focus is on the spatial dimension as it relates to the movement of objects.

2.1. Discretised representation of space

Many analytical approaches for movement data treat space as a collection of discrete relevant places. When these places are not predefined by the application, they can be identified from the available movement data. For instance, clusters of locations corresponding to specific movement events like stops, turns, or speed reductions can define these relevant places. Alternatively, spaces can be divided into compartments, ensuring the division is fine enough to capture significant variations in space utilisation at the desired level of detail while still being manageable. Various methods can be employed for spatial division. Some rely on pre-existing geographical or administrative boundaries, while others divide space based on specific geographical objects such as street segments or intersections. Divisions into equally-sized rectangles or hexagonal grids are also quite common.

In many of our research activities we utilise data-driven tessellation dividing space into places using Voronoi polygons. These polygons are defined based on the proximity of locations to specific seeds, such as centres of dense spatial clusters of distinctive points from the movement trajectories [9]. This division can be adjusted to attain the desired level of detail while retaining essential spatial information. Notably, our empirical findings demonstrate that altering the level of spatial abstraction in representing vehicle traffic retains crucial relationships between traffic volume and movement speed [10].

2.2. Properties of movement data

It is important to note that, depending on factors like the movement environment, the characteristics of moving objects, the physics of movement, and the methods of data collection, movement trajectories need to be treated in one of two distinct ways: as quasi-continuous or as episodic [11]. The primary distinction between these categories lies in the feasibility of reconstructing the movements that occurred between recorded positions. Quasi-continuous data allow for the interpolation of intermediate positions between sequential data points, effectively filling in gaps between them. In the case of episodic data, sequentially recorded positions may be temporally and/or spatially distant from each other, making it impossible to make valid inferences about the intermediate movements.

This differentiation has important implications for place-based data analysis, particularly when trajectories are transformed into sequences of visited places and transitions between them. In quasi-continuous data, consecutive places in these sequences are typically spatial neighbours. In situations where an occasional gap occurs in the trajectory, interpolation can be used to estimate the movement between non-neighbouring places by identifying an optimal path (e.g., the shortest path) between them. Conversely, for episodic data, one must refrain from attempting interpolation or reconstruction of the complete place sequence, as well as from interpreting the data as sequences of direct transitions between the places.

Another relevant distinction is between trajectories reflecting network-constrained and free movements. In network-constrained trajectories, like those of vehicles on roads, one can expect a large number of repeated subsequences of visited places, since moving entities tend to follow similar routes. Unconstrained trajectories, on the other hand, are much more varied. If the division of the underlying space is overly fine, the resulting place sequences from these trajectories may lack the necessary similarity to reveal common movement patterns throughout the space. Furthermore, data-driven space tessellation can yield results that may appear arbitrary and challenging to interpret. In such cases, opting for a regular grid or a division based on domain-specific spatial semantics might offer a more straightforward and meaningful representation.

2.3. Topic modelling

Topic modelling is a method for discovering abstract themes or topics in a collection of documents [12]. It is widely used in text mining as a tool for uncovering hidden structures in text data. The two most commonly used methods are Latent Dirichlet Allocation (LDA) [13] and Non-negative Matrix Factorisation (NMF) [14]. Beyond text analysis, topic modelling methods can be applied to abstract documents where "terms" are labels or identifiers representing objects of any nature, for instance, nucleotides in DNA [15].

Topic modelling treats each text document as a combination of terms. It finds groups of documents containing similar terms and groups of terms occurring in similar documents. These groups of terms are called *topics*. The output of an algorithm consists of two matrices. The first matrix provides a definition of each topic in the form of a multidimensional vector of term weights expressing the importance of each term for the topic. The group of terms having high weights is supposed to express the meaning of the topic, which can be understood by a human, while formally a topic is merely a distribution of weights over the set of all terms. The second matrix represents the content of each document as a combination of weights of the topics. In our work, we mostly use the topic-term matrix.

Topic modelling considers each document as a bag of terms, disregarding the word order. However, it is possible to define terms as ordered pairs of words that appear in documents one after another [16]. This approach increases the size of the vocabulary, but it can provide more useful results when the word order is important.

3. Related work

In this section, we review the related work on the topics of space use exploration, utilisation of space embedding and clustering techniques in visual analytics, utilisation of topic modelling, particularly, for non-textual data, and visualisation techniques included in our analysis workflows.

3.1. Visualisation and analysis of space use

Trajectory data analysis is an important topic of research in data management [17], data mining [18, 19] and visual analytics [1]. Among various applications, trajectories of moving objects are studied for identifying meaningful places in the space and evaluating their interconnections. Analysis of place connectivity finds application in diverse domains, including transportation [20, 21] and animal ecology [8], which tend to develop their specific methods. In transportation, there are studies focusing on movements between selected locations, the assessment of connectivity for specific places like transportation hubs or residential districts, and the analysis of accessibility to particular classes of locations. Examples include computing isochrones of travel times to the nearest hospital [22] and generating a visual overview of a train line's daily schedule [23]. Several approaches are proposed for understanding origin-destination flows, frequently involving exploration of connections from a chosen origin or to a given destination through multiple maps [24] or map-like representations like OD maps [25].

Research on connectivity of places in animal ecology is focused on understanding how animals move across landscapes and interact with their environment. Landscape connectivity is defined as the movement of organisms and materials among

patches on landscapes, and functional connectivity is defined as the connectivity of a landscape from the perspective of a focal organism [26]. Review [27] provides an overview of landscape connectivity analysis techniques, genetic connectivity, and the implications of habitat fragmentation for conservation and management.

Among generic techniques for movement data analysis, clustering of trajectories by proximity of origins and/or destinations and by similarity of the routes [28, 1] can provide some insights into the extent of connections between different places through movement. However, it's important to note that this approach may lead to significant underestimation of connectedness, as it does not capture the diversity of trajectories that traverse the same groups of places but do not fit within the same cluster.

Human movement can be represented as a network (graph) with nodes corresponding to the places and links to aggregated transitions between the places. This enables application of graphs analysis methods, in particular, node clustering and community detection, to analyse place interconnections. Thus, Rinzivillo et al. [29] extract relevant clusters of nodes and map then back onto the territory, finding a good match with the existing administrative borders. A downside of this method is that it divides the places into disjoint groups and does not reveal connections between the groups. Brilhante et al. [30] use community detection [31] to find groups of points of interest that are highly connected by the mobility of the individuals. Due to the existence of community detection algorithms capable to find overlapping communities of graph nodes [32], overlapping groups of interconnected places can be discovered. However, the algorithms have been criticised for missing important structural properties of the communities, particularly, the memberships of the nodes. Besides, the problem of effective visualisation of the results in the spatial context has not been properly addressed.

As an alternative to the graph-based community detection, we investigated the potential of topic modelling techniques, which do not group places or links directly but express their relatedness through vectors of topic weights.

3.2. Embedding and clustering in visual analytics

Dimensionality reduction [33] and cluster analysis [34] are commonly utilised techniques in visual analytics [35, 36] for handling various types of complex data. Particularly, dimensionality reduction is often employed to embed data instances in a two-dimensional abstract space, facilitating visualisation through scatterplots . In this role, dimensionality reduction is called *space embedding* [37], as well as *multidimensional projection* or, simply, *projection* [38, 39]. Both embedding and clustering are used to explore similarities between data items and find groups of similar items by employing numeric measures of pairwise similarity [40], also called distance functions. A large variety of similarity measures have been proposed for different data types, including numeric and binary [41], categorical [42], sequential [43], geo-spatial [44], and temporal data [45]. Despite the availability of numerous similarity measures, it may be necessary to develop a specialised distance function to capture specific notions of similarity. In our work, we have implemented a custom function to measure the similarity between places based on the number of trajectories connecting them.

3.3. Topic modelling in visual analytics

Topic modelling methods are very sensitive to their parameters such as the desired number of topics, frequency of words to be considered or ignored by the method, initialisation procedure, just to name a few issues. There exist sophisticated visual analytics tools for user-steerable topic model optimisation, for example [46, 47].

Beyond analysing texts in natural languages, topic modelling methods have been successfully applied to different types of data, for example, DNA codes [15], software repositories [48], user activities in interaction with software systems [47], and multivariate time series [49]. These examples demonstrate the versatility of the topic modelling methods.

Works on applying topic modelling to movement data, namely, taxi trajectories, have been reported in papers by Chu et al [50] and Liu et al [51], respectively. Both papers construct a vocabulary from street segments and their ordered pairs, with the goal of finding patterns in the trajectories and focusing on the moving objects. Our work expands on these ideas with a different goal: finding patterns in space, understanding space structure in terms of place interconnections, and revealing how the space is used by movers.

3.4. Visualisation techniques involved in our approach

To visualise topic modelling results for places, we create maps with pie charts positioned at the locations of the places. Although pie charts tend to have bad reputation since the experimental study conducted by Cleveland and McGill [52], there have been other studies showing that pie charts can be superior to other representations in tasks on estimating proportions [53]. It is important to note that all experiments evaluated the accuracy of assessing elementary values, whereas this task has low relevance in maps intended to provide an overview and enable detection of spatial patterns. Unfortunately, we are not aware of any studies on perception of maps with multiple pie charts, although such representations are quite popular in cartography since the pioneering works by Charles Minard [54].

Another technique that we use is similarity-based assignment of colours to data items obtained by means of a 2D embedding with a continuous colour scale spread over the embedding space. The items receive the colours corresponding to their positions in the embedding. Properties of different two-dimensional colour scales that can be used for this purpose have been discussed and compared by Bernard et al. [55]. We routinely use this technique over many years to assign colours to clusters (e.g., [56]), but it can be applied to any kind of data items given a measure of similarity between them. For instance, this technique was applied to value combinations from multivariate time series [57, 58]. Notably, the use of embedding for assigning colours to data is not discussed in a recently published survey on the use of embeddings in visual analytics [37].

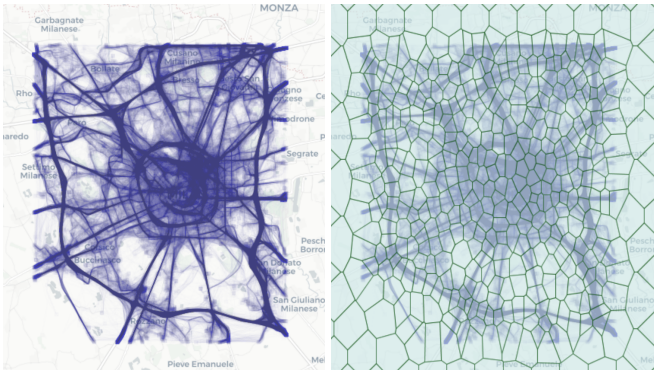

Fig. 1. Trajectories of cars (left) and a result of data-driven tessellation of the territory (right).

4. Approaches

As previously mentioned, we investigated and compared two distinct approaches to analysing space use by moving entities. In one approach, we employed space embedding and clustering techniques utilising a numeric measure of pairwise similarity between places derived from the trajectories connecting them. The other approach leveraged topic modelling. Our rationale for juxtaposing topic modelling with embedding and clustering stemmed from the widespread use of the latter two methodologies in visual analytics for exploring and analysing various data types, given their universal applicability to any data for which a numeric measure of similarity can be chosen or defined. Conversely, the potential of topic modelling for non-textual data remains relatively unexplored.

In this section, we outline both methodologies utilising a dataset capturing the movement of 17,000 cars in Milan over a one-week period, totaling approximately 2,000,000 positional records (Fig. 1, left). This data set has been extensively analysed in multiple previous studies (e.g., in [1, 9]), so the major patterns in the data are already known. This provides us with a valuable opportunity to validate our new findings against the earlier extracted knowledge.

4.1. Representation of trajectories

As mentioned in Section 2, our analysis relies on representing space as a discrete set of places. When the places are not predefined, we divide continuous space into compartments by means of data-driven tessellation [9]. The method allows adjusting the sizes of the compartments so as to obtain a set of places that is sufficiently big for uncovering essential differences in space use at a desired level of abstraction but not so large that it becomes unwieldy. For our running example of Milan traffic, we target at compartments of about 1km radius based on the size of the city and structure of its road network. After cleaning the trajectory data and dividing them into 51,498 trips, we applied the tessellation algorithm and obtained 451 polygons (Fig. 1, right), 385 of those were crossed by the trajectories. We also obtained 2156 directed links between the polygons that emerged due to the transitions made by the moving entities. In the following, we shall use the term *transition link* or, in short, *link*, to refer to links between places that result from transitions.

The polygons form our set of places $p_1, p_2, ..., p_{385}$. Each trajectory t_i that starts in place $p_{t_i^0}$ and ends in place $p_{t_i^{N_i}}$ receives two complementary representations suitable for finding answers to the questions 1 and 2, respectively, formulated in the problem statement (Section 1):

1. List of visited places: $p_{t_i^1}, p_{t_i^2}, ..., p_{t_i^{N_i}}$
2. List of transition links between the places: $p_{t_i^1} \rightarrow p_{t_i^2}, p_{t_i^2} \rightarrow p_{t_i^3}, ..., p_{t_i^{N_i-1}} \rightarrow p_{t_i^{N_i}}$

In our investigation, we employ these two representations in two distinct approaches: (1) utilising a suitable measure of place similarity for embedding and clustering, referred to as approach S (or S-approach), and (2) employing topic modelling, referred to as approach T (or T-approach).

4.2. Approach S: Utilising a co-visiting similarity measure

We define a measure of similarity between two places A, B based on the number of trajectories that visited both places. The measure is expressed as a distance function (i.e., zero corresponds to the highest possible similarity) computed as follows (Equation 1):

$$D_{A,B} = 1 - \frac{|T^A \cap T^B|}{|T^A \cup T^B|} \quad (1)$$

T^A and T^B denote the subsets of trajectories that visited the places A and B, respectively. As can be seen, the distance equals 0 when T^A and T^B coincide and 1 when they are disjoint. The same formula is used to compute co-visiting-based distances between transition links.

Once a matrix of pairwise distances between places and/or links has been computed, analytical techniques involving embedding or clustering can be applied to the matrix. Let's begin with representing the data as lists of visited places. We experimented with several embedding methods, including MDS [59] and t-SNE [60], which yielded consistent outcomes. A comprehensive comparison of different techniques (see) is beyond the scope of our paper. Interested readers can be referred to the survey by Ayesha et al. [61]. We present the results of the t-SNE embedding in Fig. 2: the embedding space with points representing the places (top left), colour-coded representation of the space using a radial colour map (top right), and the use of the embedding-based colours in the geographic map of places (bottom). This technique effectively assigns similar colours to places that have small distances between them, as determined by the distance matrix. Examining the map reveals distinct patterns that align with specific geographical features, including the belt road, highways, major radial roads, the city center, and its outskirts.

Another way of utilising a distance matrix in analysis is to apply clustering methods that permit this type of input. Not all methods are suitable; thus, the popular k-means requires data representation in the form of multidimensional vectors. In our example, we employ density-based clustering using OPTICS [62]. The results displayed in Fig. 3 are consistent with

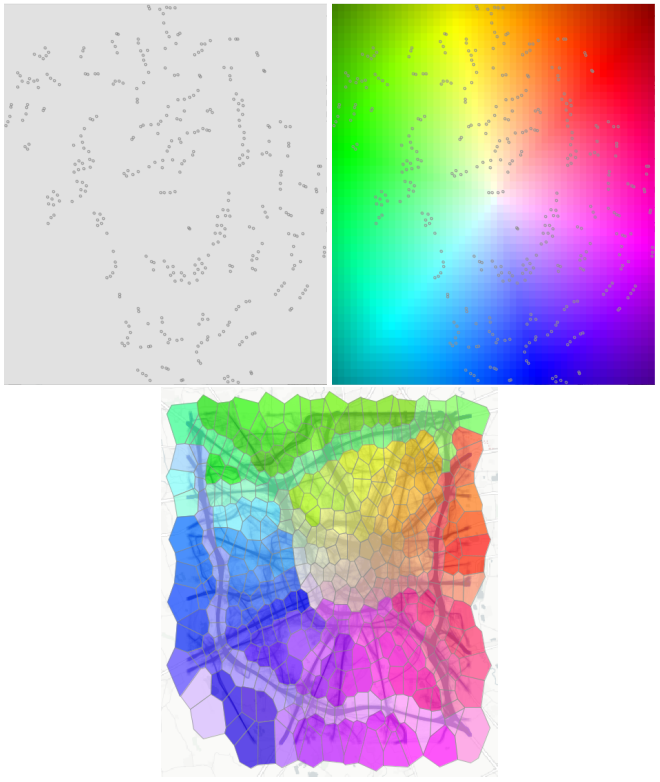

Fig. 2. t-SNE embedding of places using co-visiting-based distance matrix (top left). A 2D colour map is applied to the embedding (top right), and the colours are used for painting places on he map (bottom).

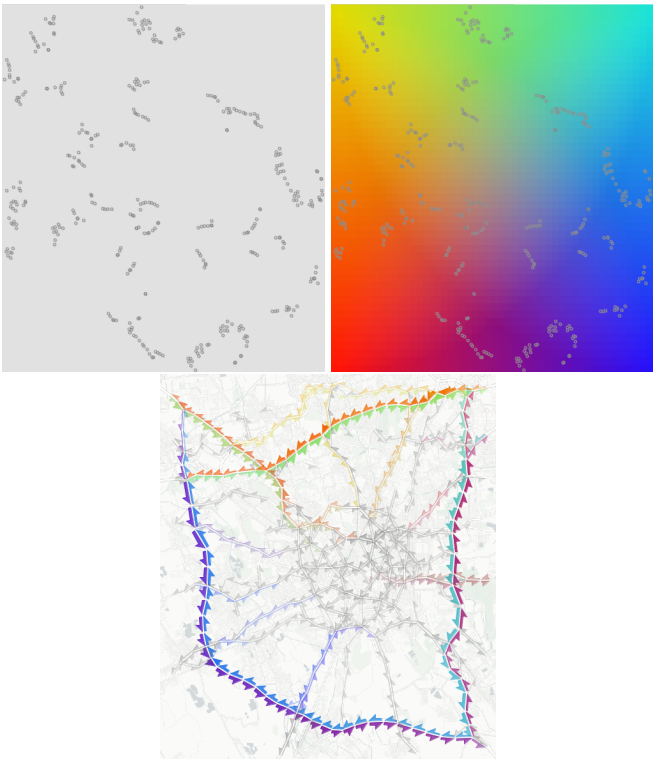

Fig. 4. t-SNE embedding of 2156 transition links using pairwise similarity distance matrix (top left). A 2D colour map is applied to the embedding (top right), and the colours are used for painting the half-arrow symbols representing the links on the map (bottom).

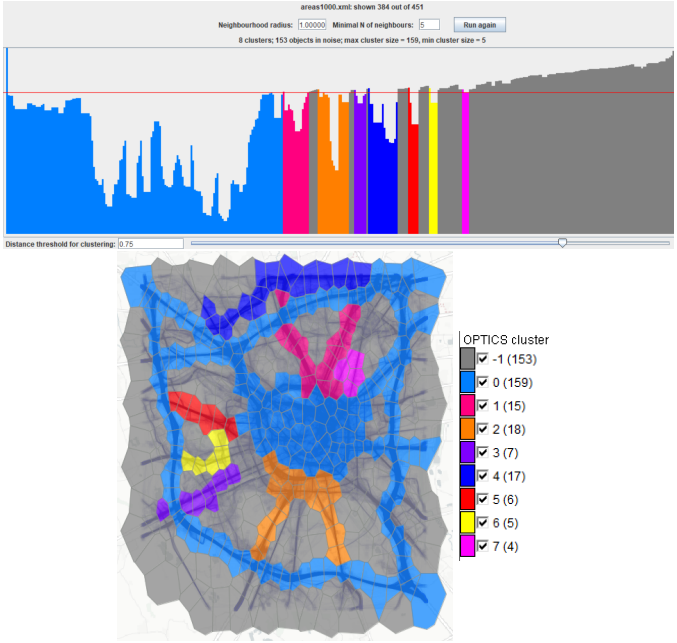

Fig. 3. Clustering of places by Optics: reachability plot (top) indicating clusters, map (bottom left) and legend (bottom right). The noise (labelled -1 in the legend) is shown in grey while the other colours correspond to clusters.

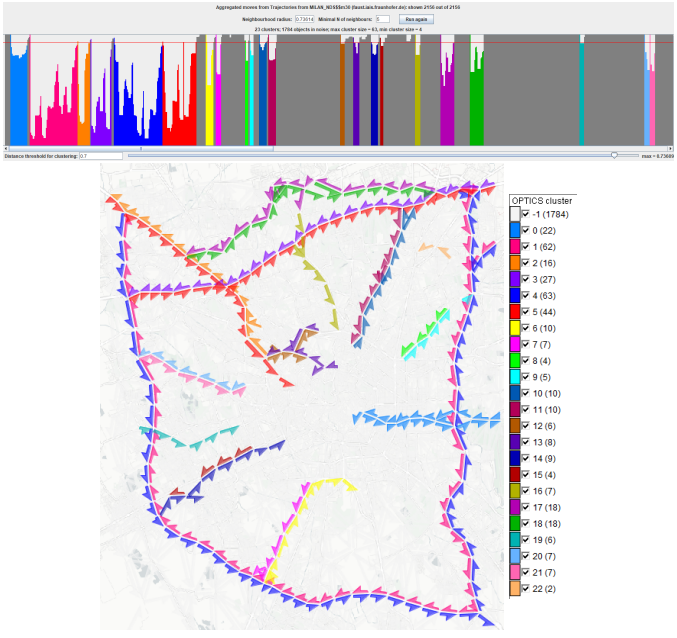

Fig. 5. Clustering of links by Optics: reachability plot (top) indicating clusters, map (bottom left) and legend (bottom right).

the findings from the t-SNE embedding. Notably, we can identify clusters that correspond to major components of the road network, effectively reflecting the overall urban structure. The shape of the reachability plot (Fig. 3, top) indicates the presence of hierarchical clusters, as was explained by the authors of the method [62]. Nonetheless, even minor adjustments to the clustering parameters lead to significant alterations in the resulting clusters. A slight reduction of the neighbourhood radius

can cause the division of certain clusters, while others vanish because their members no longer meet the necessary neighbour count. Additionally, a considerable part of the city falls within the *noise* category (represented by grey colouring), signifying that these places do not fit in dense neighbourhoods of other places defined by the parameters of the method. This, however, does not necessarily mean their high dissimilarity to others.

Figures 4 and 5 demonstrate the results of applying embedding and density-based clustering, respectively, to a distance matrix comprising pairwise distances between 2156 transition links. The transition links are represented on maps by flow symbols in the shape of half-arrow, as suggested by Tobler [63]. In both the embedding and clustering processes, links associated with significant segments of the road network were grouped together, while the majority of minor links connected to a limited number of trajectories remained unclustered.

Our experiments have shown that the S-approach can result in creating visualisations that exhibit meaningful spatial patterns. In our running example, the visible patterns align with our knowledge derived from earlier analyses. However, there are several disadvantages. First of all, the results of applying both classes of techniques are highly sensitive to parameter settings. Thus, altering the perplexity in t-SNE or choosing another embedding method changes the distribution of the places in the embedding space, which, in turn, changes the colouring and, hence, the visual patterns on the cartographic map. Similarly, adjusting the thresholds determining the neighbourhoods in OPTICS results in a different division of the set of places into clusters and noise, which may be difficult to compare with the previously obtained outcomes.

Besides, the results of embedding techniques are characterised by very high stress indicating substantial distortions of the distances in the produced 2D output. Another limitation of embedding is that patterns are constructed only in the mind of an observer lacking an explicit representation that could be used in the further analysis process. On the other hand, clustering methods produce tangible results, namely, cluster memberships, but they divide data into disjoint groups and miss any relationships between them. The noted limitations raise a need in finding other approaches to studying place connectivity. To address this need, we explore the possible uses of topic modelling techniques.

4.3. Approach T: Applying topic modelling

Here we introduce our novel approach T, in which topic modelling is applied to trajectories treated as documents.

4.3.1. General ideas

As noted in Sections 2 and 3.3, topic modelling methods are applicable to abstract "documents" where "terms" can be labels of objects of any kind. In our work, terms are identifiers of either places in space or directed links (i.e., ordered pairs of places) representing possible transitions between places. A trajectory transformed to a sequence of visited places or transitions (Section 4.1) is treated as a document consisting of words $p_{t_i^1}, p_{t_i^2}, ..., p_{t_i^{N_i}}$ or of words $p_{t_i^1} \rightarrow p_{t_i^2}, p_{t_i^2} \rightarrow p_{t_i^3}, ..., p_{t_i^{N_i-1}} \rightarrow p_{t_i^{N_i}}$. The set of trajectories in our running example is considered as a *corpus* consisting of 51,498 *documents*.

When topic modelling is applied to trajectories represented as combinations of visited places, the resulting topics are distributions of weights over the set of all places. From the perspective of the topic modelling algorithm, places are merely distinct terms. However, we can take into account the spatial positions of the places and thus treat each topic as a *spatial distribution* of weights. The places with high weights constitute the core of the topic. The topic tells us that these places tend to be used conjointly in multiple trajectories, which is a particular pattern of space use.

Similarly, when topic modelling is applied to trajectories represented as combinations of transitions between places, each resulting topic is a distribution of weights over the set of all links. It can be treated as a directed weighted graph with vertices corresponding to the places and weighted edges to the transition links. In the spatial context, the topic core is a particular spatial configuration of places and transition links which shows how the places are interconnected.

4.3.2. Topic modelling in application to places

Let's start with the representation in the form of the lists of places. The *vocabulary* of the document corpus consist in this case of 385 distinct *terms*, each corresponding to one place. In text mining, a corpus with such characteristics is considered as a suitable subject for applying topic modelling methods. Taking into account that the documents are rather short, and the vocabulary is not very extensive, it appears preferable to apply NMF [14] instead of more popular LDA [13]. NMF was found to perform better than other methods in comparative studies involving analysis of short texts, such as posts in social media [12].

It is known [64, 65] that the results of topic modelling can vary significantly based on the number of topics desired. To determine a suitable number of topics, we use an ensemble approach as described by Chen et al. [47]. This involves running NMF multiple times within a specified range of target parameters, combining all obtained topics into one table, and reducing the dimensionality of the topics using t-SNE [60]. The number of well separated groups of points observed in the embedding space indicates the number of stable topics existing in the data. Figure 6, top, shows an embedding of the NMF outputs from 11 iterations for the target number of topics ranging from 15 to 25. Strong clustering of topics is observed, indicating consistent results with only slight variations. Based on the number of well-separated groups of points, we conclude that the distribution can be appropriately represented by 21 topics (Fig. 6). We are aware that t-SNE and other dimensionality reduction methods are quite sensitive to hyperparameter settings [65]. To validate the number of topics, we performed sensitivity analysis by varying the perplexity value, and found that the results remain stable.

A single run of a topic modelling method generates two output matrices: topic-term and document-topic. The first matrix represents 21 topics by assigning non-negative weights to each of the 385 terms (places). The second matrix assigns non-

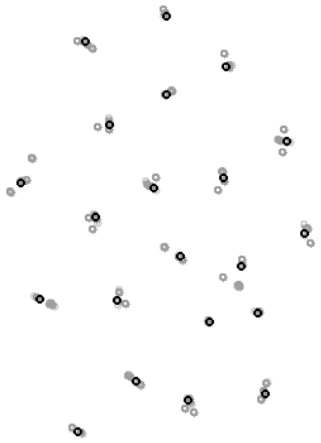

Fig. 6. t-SNE embedding of NMF outputs of 11 runs (in grey); the results of the run that creates 21 topics is marked in black.

negative weights of the 21 topics to each of the 51,498 documents, i.e., trajectories. These weights reflect the strength of the association of each term or document with the topic.

To interpret the acquired topics, it is necessary to study their distribution over the set of terms. As the terms represent places in our case, each topic can be expressed as a geographic distribution of the weights associated with the places. The spatial distributions of all topics are shown in the small multiple choropleth maps in Fig. 7. We see that each topic has a spatially compact core consisting of places with high topic weights represented by darker shading. The core is usually associated with components of the transportation infrastructure. One topic has the core in the city centre. The core is surrounded by adjacent places with lower topic weights representing the "catchment regions" of the respective infrastructure features. There are also spatially scattered distant places that are weakly associated with the topic. Since their topic weights are low, they can rather be considered as noise than a part of the spatial pattern.

One place may be associated with multiple topics, while the strengths of the associations usually differ. In Fig. 8, each place is coloured according to the dominant topic assigned to it, i.e., the topic having the highest weight. We see how the dominant topics divide the territory into regions, such that connectedness of the places within the regions is stronger than between the regions. It is important to note that the colours assigned to the topics indicate the similarity of the topics based on the closeness of their vectors of the term weights. To assign these colours, the set of 21 topics described by the vectors of 385 term weights is projected onto a 2D space using one of existing dimensionality reduction methods, namely, MDS [59]. The positions in this space are colour-coded using the Cube Diagonal Cut B-C-Y-R colour map [55, 66]. We stress the need of using different embedding methods for different purposes. The neighbourhood-preserving t-SNE is used for selecting the optimal number of topics, while the better preserving long distances MDS is used for assigning colours to topics.

Places having sufficiently high weights of two or more topics can be considered as connectors between the spatial regions represented by the topics. It is hard to identify such connectors by examining the small multiples display, where each topic is shown separately from the others. Hence, we need to visualise the topics on a single map in such a way that the combinations of the topic weights in the places can be seen. The map with pie charts demonstrated in Fig. 9 serves this purpose. We see that some individual places, as well as larger contiguous groups of places, are primarily linked to a single topic, and we also observe areas comprised of places that exhibit a combination of two or more topics. These areas show us where and how strongly the connectivity regions revealed by the dominant topics are interlinked with other (typically adjacent) regions. This kind of result, i.e., division into overlapping regions, cannot be obtained by means of clustering. The capability of producing such results is an indubitable strength of topic modelling.

4.3.3. Topic modelling in application to links

We have also applied NMF to the representation of the data set as a *corpus* of *documents* composed from the *vocabulary* consisting of 2,156 distinct *terms* representing the directed transition links between the places $p_i \rightarrow p_j$. The iterative execution of NMF with setting the number of topics to values in the range from 20 to 35 followed by visual exploration of the embedding space suggested acquiring 30 topics. We obtain two matrices, one with non-negative term weights for the topics and another with non-negative topics weights for the documents. Based on the first matrix, each topic consists of a combination of directed links with non-zero weights. It can be considered as a directed weighted graph with spatially anchored nodes (= places). Such a graph can be represented visually by a node-link diagram whose layout is determined by the spatial positions of the nodes.

As an example, the graph representing one selected topic is shown in Fig. 10. The links are represented by white curved lines with the curvature increasing in the link direction and the width proportional to the topic weight. The representation of links by curved lines follows the established practices, e.g., [67, 68]. For choosing between the representations by straight arrows (as in Figs. 4 and 5) and by curved lines, we compare two variants of the flow maps regarding the display clarity, amount of visual clutter, and aesthetics of the appearance.

In the background of the map in Fig. 10, the black lines represent the trajectories that contributed to defining the topic, i.e., whose having non-zero weights of the topic. Since the trajectories are quasi-continuous, the links connect neighbouring places. Like in the topics composed of places, there is a topic "core" consisting of links with high weights. These high weighted links form a connected component of the graph. There are also spatially scattered links with low weights, which may rather be treated as noise than as significant components of the topic.

From the perspective of movement, the core of a link topic can be considered as a major pathway for travelling through the space. The system of links connected to the core show the ways that connect this pathway with the places in the surrounding territory. Different pathways existing in Milan with their satellite link systems can be seen in the small multiples display in Fig. 11, which shows the spatial graphs for all 30 link topics.

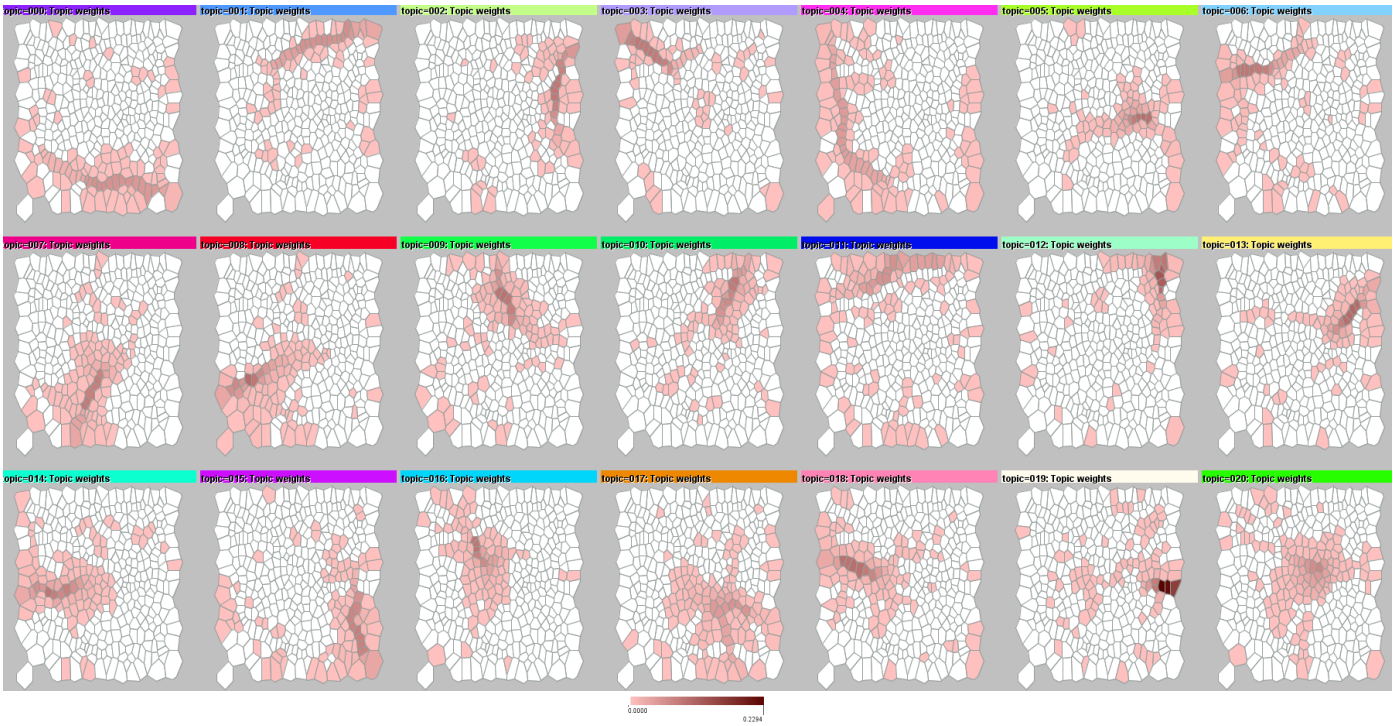

Fig. 7. Spatial distributions of 21 place visiting topics are shown by small multiples. The darkness of the shading is proportional to the topic weight for a given place; white corresponds to zero weights.

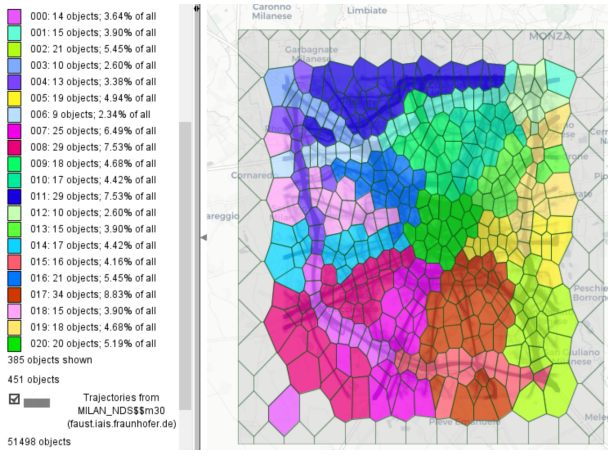

Fig. 8. Colouring of the places represents the topics having the highest weights for the places. The colours are assigned to the topics according to the topics' positions in an MDS embedding using the 2D Cube Diagonal Cut B-C-Y-R colour map.

The links with low topic weights are drawn with low opacity, which makes the topic cores more prominent.

To see all major pathways in a single map, we paint the links in unique colours assigned to the topics, as shown in Fig. 12. Assuming that a topic core consists of links for which this topic has the highest weight, the colouring according to the dominant topic fully represents the topic cores, i.e., the major pathways, whereas the satellite link systems get truncated and intertwined with other link systems. Hence, such a map is mostly suitable for observing the major pathways and their spatial relationships. Thus, we see that many of them have counterparts going in the opposite direction. By comparing the map in Fig. 12 with the maps in Fig. 8 and Fig. 9, we note that there is a correspondence between the major pathways and the connectivity regions.

Hence, the place-based and link-based topic modelling produce consistent results. Both place-based and link-based topics reveal similar connectivity regions and catchment areas of major road network elements. Additionally, the link-based topics reveal the major pathways with satellite systems of connecting links, which are used for travelling through and between the connectivity regions.

4.3.4. General notes

We used the Milan data example to introduce the novel approach T, the essence of which is application of topic modelling to specially represented trajectories followed by visual exploration and interpretation of the topics. The presented analysis demonstrates that the T-approach can work well when applied to quasi-continuous trajectories reflecting movements constrained by a transportation network. As explained in Section 2, such data entail strong connectedness of neighbouring places and frequent occurrence of same or similar routes. Therefore, the result of applying the T-approach includes contiguous regions of connectivity and unbroken pathways surrounded by systems of connected weaker links. However, it would be wrong to believe that similar types of findings can be obtained regardless of data and movement properties. To investigate the capabilities and possible limitations of the T-approach more comprehensively, we apply it to data with different characteristics and to free movements.

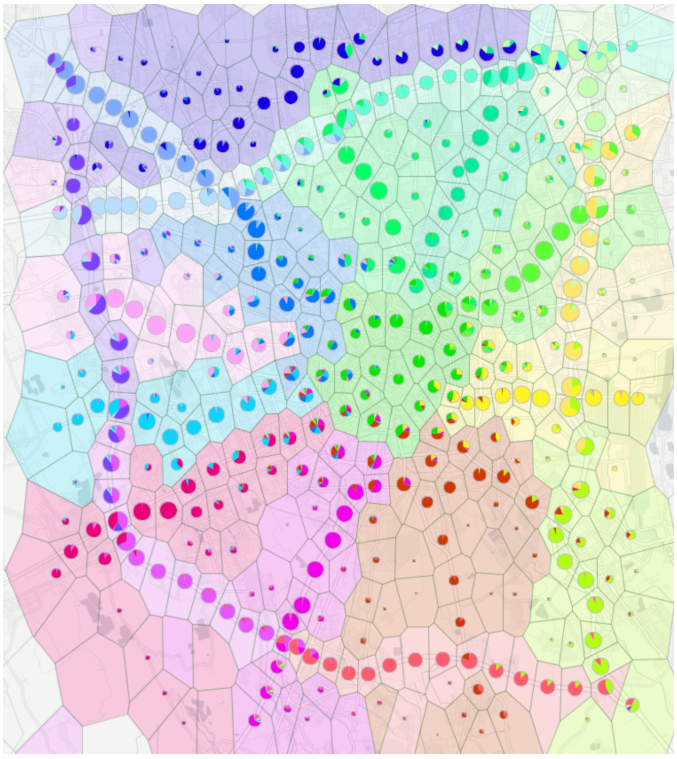

Fig. 9. The combinations of the NMF weights of the topics for the places are represented by pie charts. The pie sectors corresponding to the topics are painted in the same colours assigned to the topics as in Fig. 8.

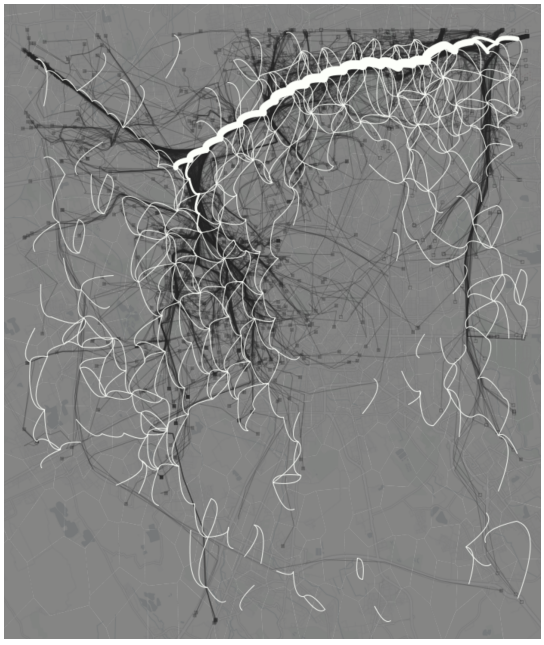

Fig. 10. One example topic is visualised as a node-link diagram with the nodes anchored in the space. The visibility of the nodes is reduced. The links are represented by curved white-coloured lines with the widths proportional to the weights of the topic for the links. In the background, 1,332 trajectories having non-zero weights of the topic are represented by black-coloured lines.

5. Case study 1: Application to episodic movement data

The data set consists of trajectories of Twitter users constructed from their geo-located Twitter messages posted on the territory of Greater London in the time frame from 05.11.2012 till 25.09.2013. For the analysis, we selected 40,246 trajectories (consisting in total of 15,246,565 points) with the minimal duration of 30 days, which are likely to be created by residents or frequent visitors of Greater London. This subset was analysed in an earlier paper [69]. The trajectories are shown on a map in Fig. 13. Using the data-driven tessellation method, where we set the desired place radius to be about 3 km, we obtained 450 places and 93,542 transition links, 29.2% of which were used only once, and 98.25% (91,905) are links between non-neighbouring places. Such links appear due to the absence of recorded visits of the Twitter users in intermediate places.

Our previous analysis [69] revealed high correspondence between the spatial patterns of the Twitter users' mobility and the topology of the transportation network of Greater London, with strong radial flows from the peripheral areas to the city centre and back. We expect to detect similar patterns by means of topic modelling, and we also want to see whether it will give us any additional information.

To select a suitable number of topics for the places, we iteratively run NMF with setting the parameter value in the range from 15 to 35. In a common projection of all resulting topics, we observe that the result with 27 topics is the best in terms of representing the topic distribution in the projection space. In Fig. 14, top, the dominant topics of the places are represented by place colouring and the compositions of the topic weights by pie charts. We see that the pie chart sizes in the city centre are much larger than on the remaining territory, and the sizes decrease in the directions from the centre to the periphery. This aligns with the spatial distribution of the place visits counts.

Consistent with the previous analysis, the application of the topic modelling reveals the central-radial structure of the Twitter users' mobility. There are areas of dominance of different topics extending radially from the centre to the periphery, and there is little intersection between neighbouring areas. A prominent exception is the area of the Heathrow airport on the west of the city, which is covered by three places containing relatively big pies composed of many differently coloured sectors, which indicate connectedness of these places with many different areas. Opposite to this, there are many places in the centre where the charts appear as large unicoloured circles. This kind of appearance means that a place is either very weakly connected to places from other areas or all connections have equal strengths and thus do not contribute to the definition of the topic associated with this place.

Additional understanding can be gained from the small multiples display in the upper part of Fig. 15 presenting the spatial distributions of the individual topics. The topic weights for the places are represented by proportional areas of circle symbols. We see two kinds of distributions: (1) large groups of neighbouring places with "cores" formed by places with higher weights, plus multiple disjoint places with very low topic weights scattered over the territory, and (2) one place with very high weight in the centre and many extremely scattered places with very low weights. The distributions of the first kind cor-

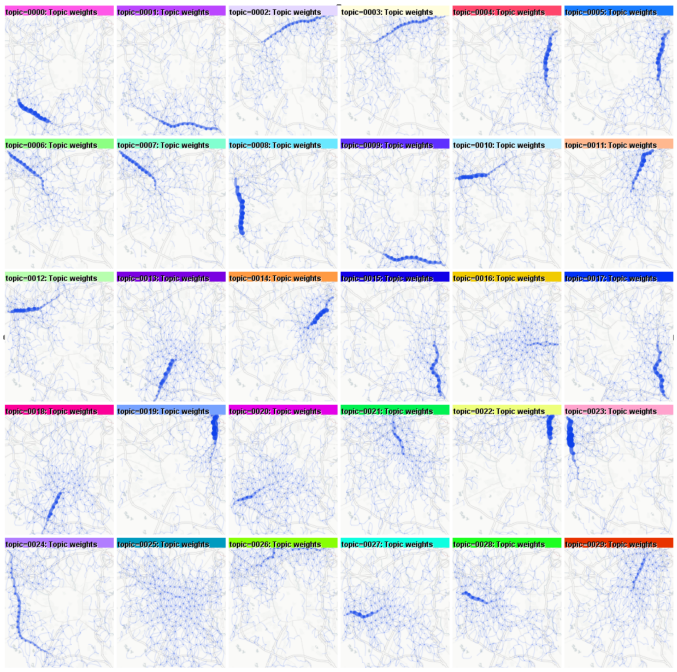

Fig. 11. Small multiples with 30 move topics represented as spatial node-link diagrams.

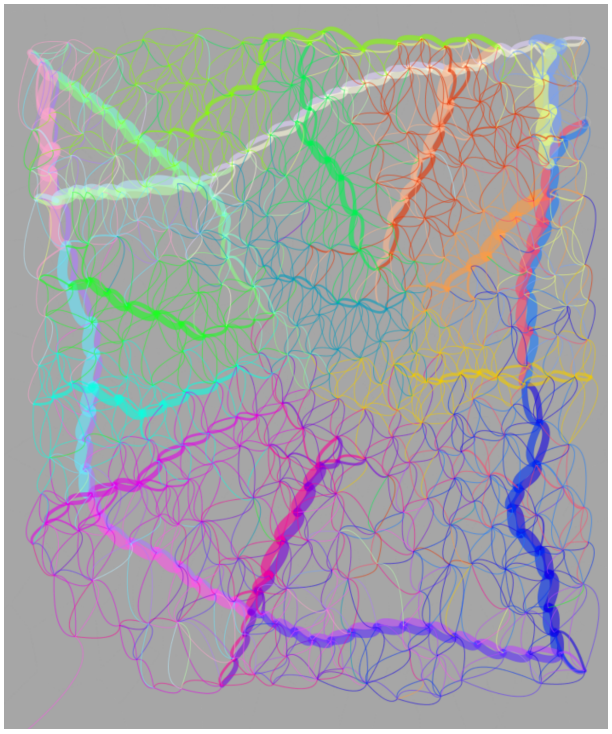

Fig. 12. Transition links coloured according to the dominant topics.

respond to the areas stretching radially from the centre to the periphery that we have seen in Fig. 14, top. The distributions of the second kind show that places in the centre have relationships with many places over the entire territory, but most of these relationships are weak, with a few exceptions. In particular, the area of Heathrow contains slightly larger circles than in the majority of places in many of the small maps. It can also

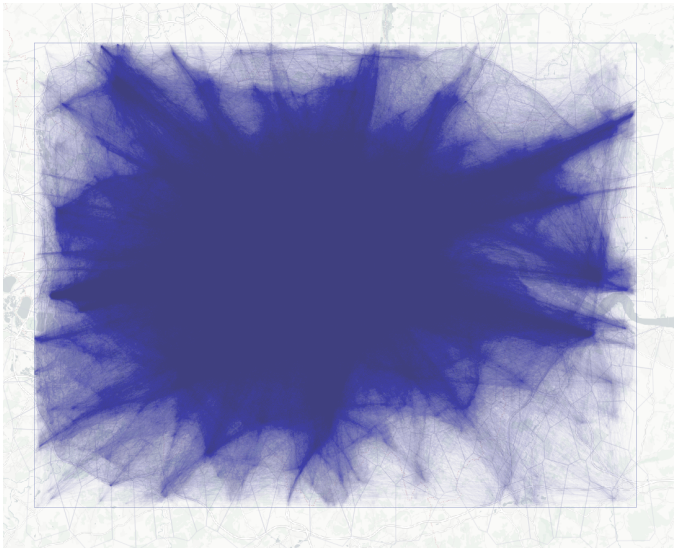

Fig. 13. Episodic trajectories of resident Twitter users in London. The trajectories are rendered with about 99% transparency.

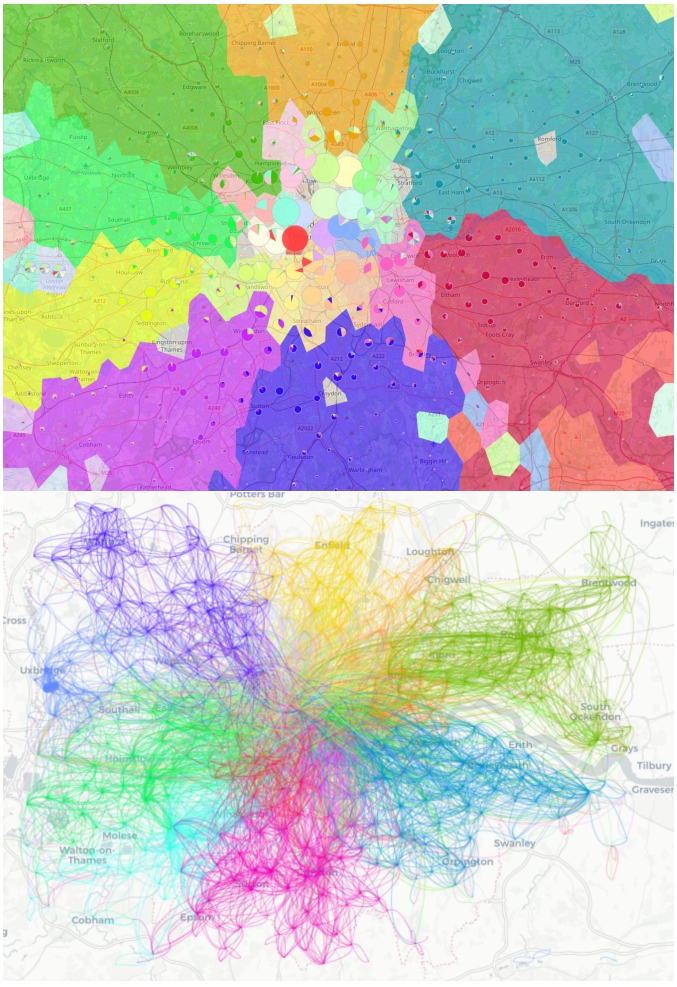

Fig. 14. Top: Spatial distribution of 27 place-based topics. The places are coloured according to the dominant topics. Pie charts show the composition of the topic weights. Bottom: 30 topics obtained for the transition links are represented by colouring the links according to their dominant topics. The links with the dominant topic weights below 0.001 are omitted.

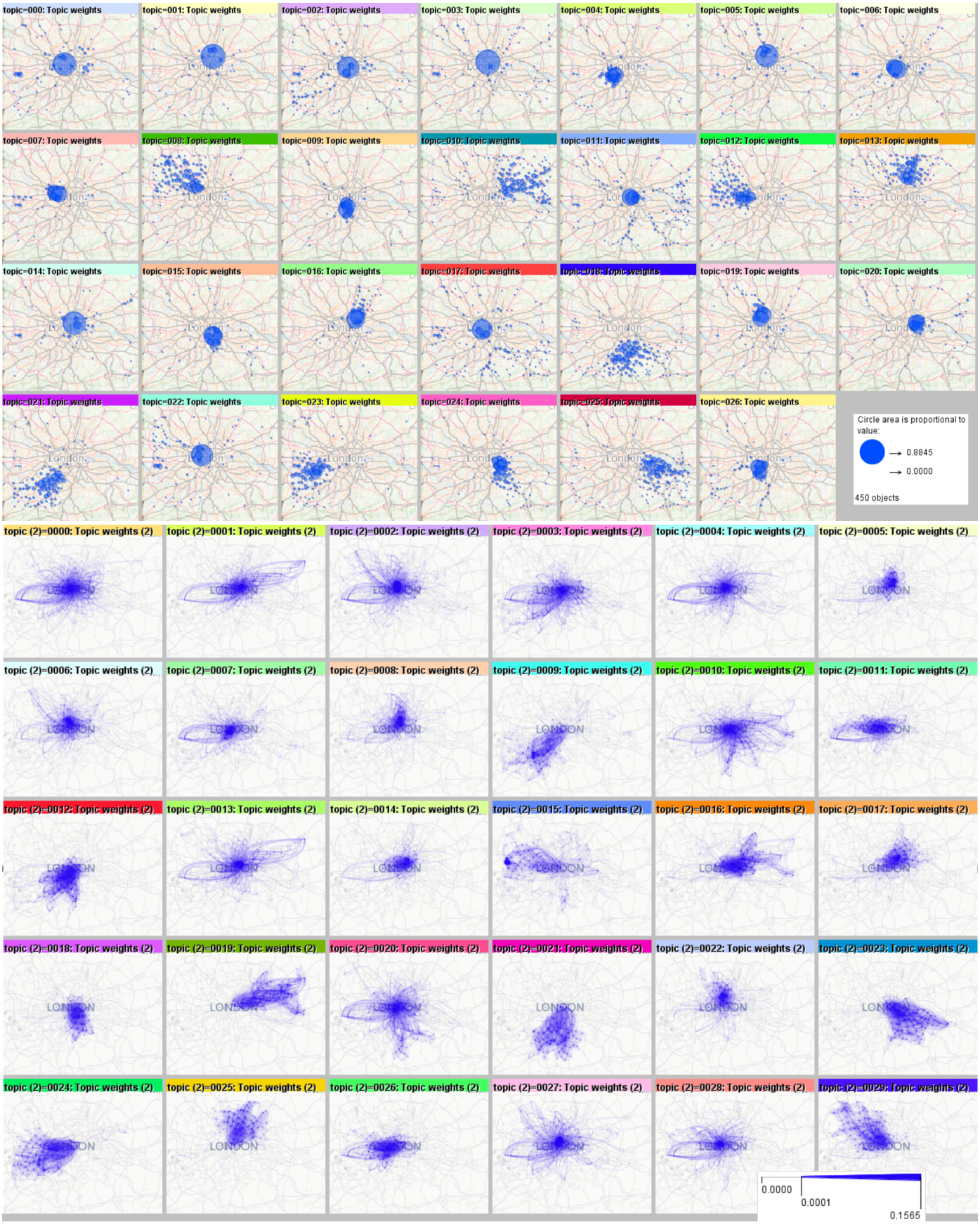

Fig. 15. Small multiples show the spatial distributions of the individual place-based (top) and link-based (bottom) topics.

be noticed that in some maps the places with low weights (i.e., small circles) are aligned along radial streets.

It should be admitted that any individual topic with such highly scattered spatial distribution where most of the places have very low topic weights is not very interesting by itself. We need to consider all these topics together to make the general observation that places in the city centre have relationships with the whole surrounding territory. If it is planned to involve the results of topic modelling in subsequent analysis, it may be appropriate to aggregate all these topics in a single topic, similarly to what we did in another work [49].

To verify and complement the observations made using the place-based topics, we also apply NMF to the lists of transition links. We set the number of topics to 30, which is slightly more than the number of the place-based topics. We avoid creating too many topics, which will require much effort to consider and interpret. Besides, the result obtained for the places showed us that some topics may be excessively fine and provide low level of abstraction.

The results of the link-based topic modelling are presented in the lower parts of Fig. 14 and Fig. 15, so that they can be compared with the place-based topics. In Fig. 14, the curved lines representing the links are painted in the colours of the dominant topics. Like for the places, see spatial clusters of uniformly coloured links stretching in different directions from the centre to the periphery. This spatial pattern is also consistent with what was uncovered in the previous study [69]. There are no prominent differences between the line widths representing the topic weights, which shows that the weights are approximately equal and low.

Complementing the patterns perceived from the map of the dominant topics, the small multiples in Fig. 15, bottom, tell us that, apart from the radial configurations of related links, there are configurations composed of many links clustered in the centre and weak connections between the centre and more distant areas. These patterns, like the patterns of type 2 for the place-based topics, tell us about high interconnectedness of the places in the centre and multitude of links between the centre and different places throughout the city. Again, each of these topics by itself is not highly useful; they need to be considered in combination.

In this case study, the link-based topic modelling does not distinguish pathways going in opposite directions. This can be attributed to the fact that each city resident's mobility over an extended period is considered as a single trajectory, which is expected to encompass routine patterns involving repetitive movements in opposite directions (e.g., home to work and work to home). The detection of unidirectional patterns would require dividing such trajectories into individual trips. However, due to the nature of social media data, most trips would be represented by only a few positions, making pattern detection challenging.

This case study demonstrates that, despite the challenging properties of episodic movement data, it is possible to uncover meaningful patterns of movement through the space, while the types of patterns significantly differ from what we have seen in the case of quasi-continuous trajectories constrained by the transportation network. In the next case study, we apply topic modelling to quasi-continuous unconstrained trajectories.

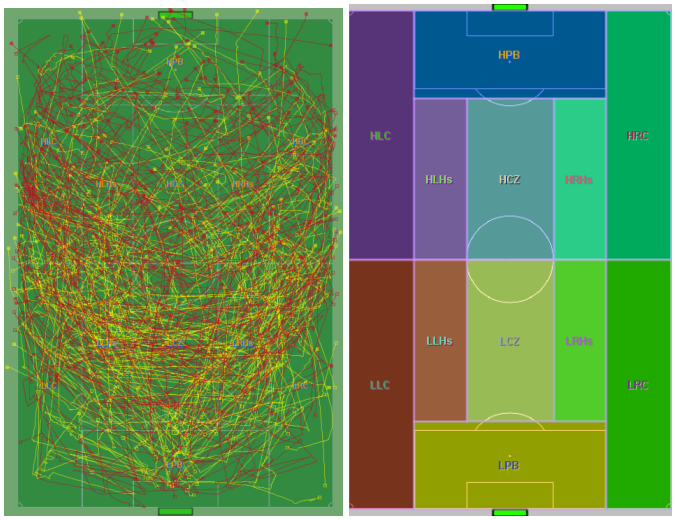

Fig. 16. Left: Fragments of the ball trajectory under the possession of two teams mapped onto a coordinate system where the pitch is oriented vertically and the upward direction is the direction of the attacks. Right: Division of the football pitch into zones.

6. Case study 2: Application to unconstrained movement (football)

In this case study, we apply topic modelling to data representing movements of the ball in a football game. The goal of the analysis is to uncover different patterns of progressing the ball towards the opponents' goal. From the entire trajectory of the ball, we exclude the time intervals when the ball was out of play and divide the remaining parts of the trajectory into 454 (=230+224) fragments corresponding to the possession by the two teams. Next, we transform the data to a uniform coordinate system with the vertical axis oriented in the attack direction, as shown in Fig. 16, i.e., the attacked goal is on the top regardless of which team possesses the ball. We apply one of existing schemes of dividing the pitch into tactical zones [70], namely, in two halves (called low and high), five lanes (left and right channels, left and right half-spaces, and central lane), and two penalty boxes, as shown on the right of Fig. 16. We want to find out how the teams use these zones to progress the ball towards the goal of their opponents. For this purpose, we consider sequences of ball transitions between the zones using only those trajectory fragments that make at least two transitions. There are 229 (=102+127) such fragments making in total 1,715 transitions. The transitions are represented by terms having the format $Z_i\_Z_j$, where Z_i and Z_j are identifiers of two pitch zones. The overall "vocabulary" consists of 60 distinct terms.

We apply NMF to the transition sequences. To decide on the number of topics to derive, we run NMF iteratively setting the desired number of topics to values from 5 to 25 and obtain 315 topics in total. We then apply dimensionality reduction to the term weights vectors of the topics to put them in a common 2D embedding space. In the embedding plot, we see 14 relatively dense clusters of points suggesting the existence of 14 topics that remain stable across many runs.

To visualise the results of topic modelling, we use small multiple displays containing maps of two types: flow maps (Fig. 17, left) and trajectory maps (Fig. 17, right). Each small map represents one topic. Flow maps portray aggregated transitions between the zones. The transitions are represented by flow symbols in the shape of half-arrow, as suggested by Tobler [63]. The widths of the symbols are proportional to the weights of the transitions in the topic represented on the map. In trajectory maps, fragments of the ball trajectory are represented by lines with widths proportional to the topic weights. When the weight equals zero, the trajectory is not visible. The lines are coloured in yellow and red according to the ball possession by two teams.

As mentioned earlier (Section 5), the number of derived topics influences the degree of abstraction of the patterns that can be observed. This can be clearly seen in Fig. 17, which presents, from top to bottom, results of deriving 8, 10, and 14 topics. When we consider the flow maps on the left of Fig. 17 focusing our attention on the prominent (i.e., sufficiently thick and bright) flow symbols, we see that the more topics we derive, the smaller are the groups of links having high weights in the same topic. These smaller groups of links manifest finer patterns of ball movements, whereas the larger groups visible in the result with 8 topics reveal patterns of a larger spatial scale.

While the flow maps show the patterns in an abstract and schematic manner, the trajectory maps on the right of Fig. 17 give us complementary information. First, we see the geometric shapes of the trajectories contributing to the topics. Second, we can estimate and compare the amounts of trajectories contributing to different topics. In particular, we see that some of the 14 topics presented in the lower part of the figure come from quite a small number of trajectories. Third, we can compare the tactics of two teams in terms of choosing the ways to move the ball. Thus, we see that the "red" team often moved the ball through the central channel to the upper left half-space, whereas the "yellow" team almost never used this pattern. The "yellow" team often moved the ball starting from their goal, while this was rarely done by the "red" team. The "yellow" team made more movements across the pitch in the lower half and the "red" team had more such movements in the upper half of the pitch.

In this case study, topic modelling was helpful, first, for obtaining an abstracted view of the movement as a combination of several general patterns and, second, for revealing groups of similar trajectories in terms of progressing the ball through the pitch zones.

7. Discussion

The strategic goal of our work was to thoroughly investigate the potential of applying topic modelling to movement data for studying particular aspects of space use, the ways to do this and to visualise the results, the meanings of the results, and the problems that may arise in order to present the knowledge and experience we gained to other researchers and practitioners, so that they can be aware of the existing possibilities and know how to make use of them. In the following, we discuss different aspects of our approach, share our experiences, heuristics, and lessons learned, and, where possible, give recommendations for those who would like to use the approach T or parts of it.

Analysis goals and interpretation of topics. Topic modelling can be applied to two representations of the same trajectories: as lists of visited places and as lists of transitions between the places. These two ways of application are oriented to different analysis goals. Lists of places are used when the goal is to see what connectivity regions exist in the space (question 1 in the problem statement). A place-based topic represents a region consisting of interconnected places, such that places within the region have predominantly stronger relationships than those with places from other regions. However, two or more regions may have some places in common indicating relationships between the regions. So, place-based topic modelling reveals regions of high internal connectivity as well as connections between them. Generally, the regions may be spatially discontinuous, as in the case study with episodic trajectories.

Topic modelling is applied to lists of transitions when the goal is to reveal major pathways by which moving entities traverse the space (question 2 in the problem statement). However, this goal can be achieved only when the trajectories are quasi-continuous, so that each trajectory consists of transitions between neighbouring places. In a case of episodic data, topic modelling is used to find groups of transitions that tend to co-occur in multiple trajectories. A topic in this case is a spatial configuration of weighted links. The places connected by the links can also be considered as parts of this configuration, which can be thus treated as a spatial graph, or network. Hence, link-based topic modelling can reveal spatial connectivity networks, which show not only groups of interconnected places but also how they are interconnected. However, connectivity networks do not explicitly represent the relative importance of the places they include. At the end of this section we shall outline a possible research direction towards overcoming this limitation.

In one analysis process, either one of the two goals or both may be pursued. In the latter case, it is reasonable to begin with obtaining and analysing place-based topics, i.e., connectivity regions. As we discuss later on, this usually gives a hint of how many link-based topics may be needed to reveal the movement patterns leading to the formation of these regions.

Preparation of data. Our approach relies on a representation of space as a discrete set of places, so that trajectories can be transformed to lists of visited places and/or list of transitions between the places, depending on the question that needs to be answered (see the problem statement in Section 1). The set of suitable places can be determined by the application domain and analysis goals. For applications where there are no predefined places or domain-specific ways to define places, we suggest to identify and delineate relevant places by applying our data-driven tessellation algorithm [9] or its extension [71] to the available movement data. A useful feature of the algorithm is the possibility to adjust the spatial extents of the resulting places to the desired level of spatial abstraction [10].

Before applying topic modelling to lists of places or transitions, it makes sense to filter out the lists consisting of a single element, i.e., only one place or only one transition. Such data

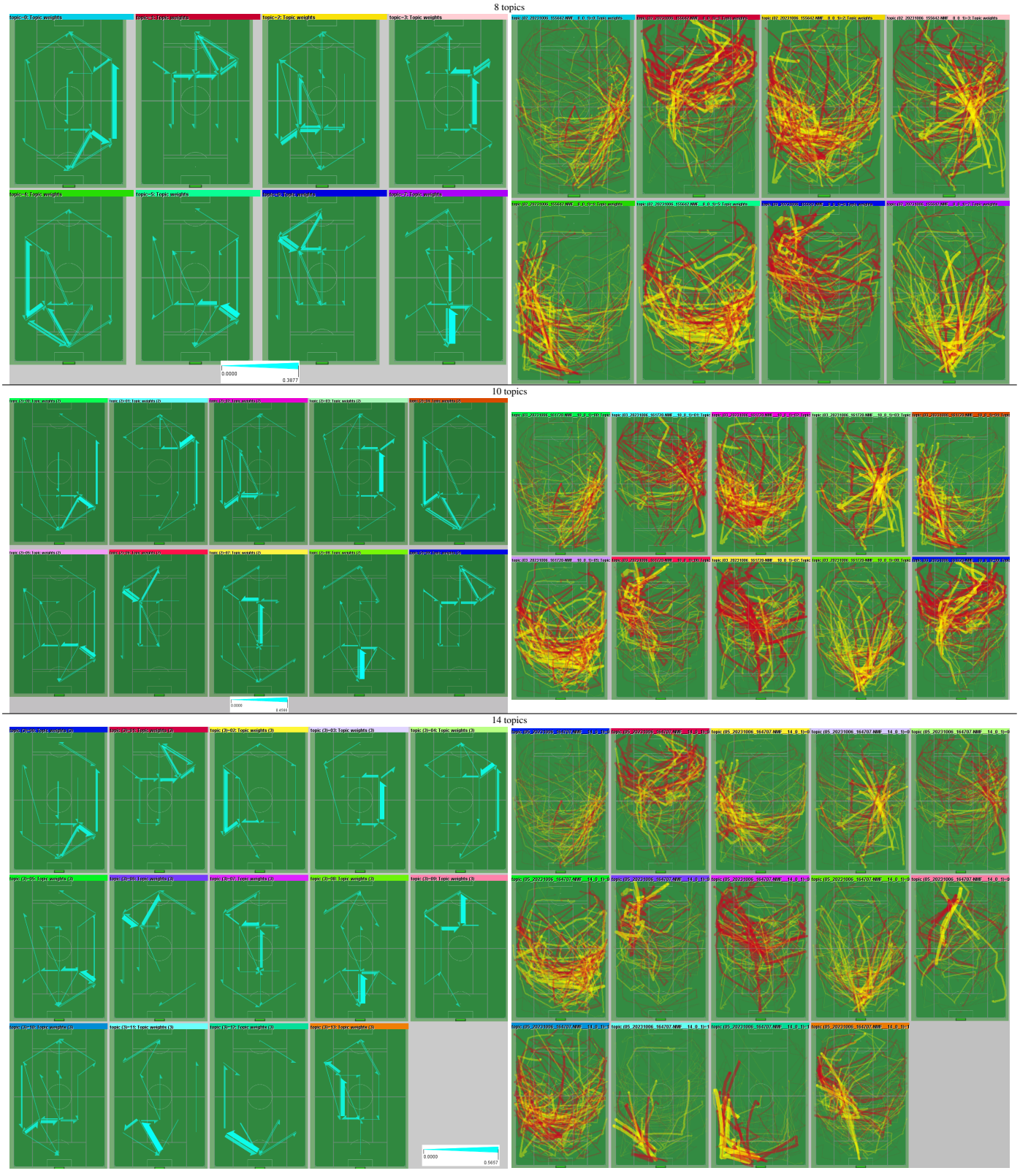

Fig. 17. Small multiples showing results of topic modelling obtained by means of NMF with setting the number of topics to 8 (top), 10 (center), and 14 (bottom). Each of the small maps corresponds to one topic. On the left, the topics are presented on flow maps where half-arrow symbols (as suggested by Tobler [63]) represent aggregated movements of the ball between the pitch zones. The widths of the symbols are proportional to the topic weights for the movements. On the right, the lines represent fragments of the ball trajectory corresponding to ball possession by two teams, which is signified by yellow or red colour. Line widths are proportional to the topic weights for the trajectory fragments.

are useless for the analysis, as they can not contribute to quantifying relationships between places or between links, respectively.

Choosing the topic modelling method. Numerous existing topic modelling methods have been discussed and compared in surveys, such as [12]. However, in all comparative studies, different algorithms have been applied to texts in natural languages. Therefore, the results of these studies may not readily translate to other types of data. Non-Negative Matrix Factorisation (NMF) [14] has demonstrated superior effectiveness over other methods when applied to short documents [72, 73]. Our proposed approach involves transforming trajectories into a form akin to short 'documents,' suggesting that NMF may be considered as a potentially suitable choice.

Advantages of NMF stem from its enforcement of non-negativity and sparsity constraints, facilitating the capture of local structures within the data. This characteristic makes the method apt for deriving interpretable topics from short documents with limited context. Since these constraints are not exclusive to text data, we anticipate that NMF will exhibit similar capabilities when applied to abstract "documents", aligning well with our application where local structures are of interest.

Additionally, another compelling reason for preferring NMF is its the deterministic nature. This characteristic ensures that running the method with different parameter settings, particularly, the number of topics to derive, yields more consistent results compared to probabilistic methods like LDA [13]. Such consistency is beneficial for the effective implementation of our method for determining the optimal number of topics.

Choosing the number of topics. As the suitable number of topics is usually not known in advance, it is necessary to consider several variants of parameter setting in order to make a reasonable choice. One way is to automatically run topic modelling multiple times with setting the number of topics to consecutive values from a specified range, represent all topics so obtained by positions in a common 2D embedding space, and find out which of the results of the individual runs represents sufficiently well the distribution of the points in the embedding space, as illustrated in Fig. 6. This approach is suitable when the "vocabulary" (i.e., the total number of places or links) is not very large. Otherwise, the high dimensionality of the term weight vectors defining the topics creates a large data volume in which the topics are distributed very sparsely, so that there is little difference between their pairwise distances [74, 75, 76]. The result of projecting this distribution to low-dimensional space is therefore unreliable. Another problem that iterative running of topic modelling for obtaining different numbers of topics takes quite long time when the vocabulary is large, which may be unsuitable for interactive exploration.

We practice the following heuristic approach. As the set of places is usually not very large (due to the possibility to regulate it through the parameter of the data-driven tessellation algorithm), we use the iteration-embedding procedure to choose a suitable number of place-based topics. Visual exploration and comparison of the spatial distributions of the chosen number of topics allows us to estimate how many link-based topics may be sufficient for uncovering relevant movement patterns involved in the formation of these distributions. It is important to take into account properties of the data, which may or may not allow detection of directional patterns, i.e., pathways through sequences of places. Our experiments show that such patterns can be extracted from quasi-continuous, especially network-constrained trajectories (as in Section 4). In such a case, a reasonable number of link-based topics to obtain should be from 1.5 to 2 times more than the number of expressive, well-distinguishable place-based topics. If the data are not suitable for detection of directional patterns, as in Section 5, the number of link-bases topics can be roughly the same as the number of place-based topics.

In any case, it is not enough to consider just one result with a particular number of topics, but it is necessary to compare it with what can be obtained when the number of topics is slightly lower and when it is slightly higher. Taking a lower number of topics may eliminate uninteresting "weak" topics having low support (i.e., the number of trajectories with high weights of these topics) and/or low weights of all terms (i.e., places or links). Taking a higher number of topics may reveal additional interesting patterns deserving attention.

The desired number of topics also depends on the scale and degree of abstraction of the spatial patterns one wishes to uncover. A large number leads to fine topics combining few places or links, whereas a smaller number leads to larger spatial configurations involving more elements. We demonstrated the impact of the chosen topic number in Section 6.

Returning back to the use of 2D embedding for determining a suitable number of topics, it is necessary to note that the appearance of the topic distribution in the embedding space greatly depends on the chosen embedding algorithm and its parameter settings. We use the neighbourhood-preserving algorithm t-SNE [60], which strives to put very similar items (i.e., nearest neighbours) close to each other in the embedding space but does not care about faithful representation of larger pairwise distances. This strategy is fit for the purpose of finding groups of very similar topics that have come from different runs, while the distances between the groups are not important. The parameter "perplexity" of t-SNE regulates the number of nearest neighbours to consider for each item. For the purpose of finding dense groups of of similar topics, the perplexity should be set to a low value. We usually try out several values in the range from 5 to 15, depending on the number of runs from which we take the topics, and use the embedding with the clearest grouping.

Comparison with approach S. We juxtaposed our novel approach T with an adaptation of the common approach S. The latter involves employing 2D space embedding (projection) or clustering after selecting or defining a suitable measure of similarity between data items. In our applications, these data items represent places or transitions between places, and the similarity measure is tailored to reflect the strength of their connectedness based on co-occurrences in the same trajectories.

Our comparison revealed several advantages of approach T. Firstly, unlike space embedding, topic modelling generates tangible outcomes suitable not only for visual exploration but also for further analysis, such as application of machine learning

methods. Although clustering also yields tangible results in the form of groups of similar places or transitions, these outcomes may not be well-suited for application in other computational analysis methods, as the groups lack explicit differentiation by numeric features.

Secondly, compared to clustering, which divides the data into disjoint and unrelated groups, topic modelling provides a more nuanced characterisation by capturing not only groups of highly interconnected places or transitions but also connections between the groups.

Thirdly, while both S- and T-approaches are sensitive to parameter settings, the parameter sensitivity of the T-approach can be mitigated by applying the proposed method for selecting an optimal number of topics, as discussed earlier.

Relation to network analysis methods. In Section 4.3.3, we have mentioned that each topic obtained from the link-based representation of trajectories can be treated as a directed weighted graph, where the vertices correspond to places, edges to transition links, and edge weights are the weights of the links in the topic. While such graphs offer potential for further analysis using network analysis methods [77], it's important to note that these methods focus primarily on binary connections between graph vertices. This differs from a holistic treatment of a topic, which represents more complex relationships involving multiple places and transition links.

Alternatively, network analysis methods can be applied based on pairwise measures of place or link similarity derived from their co-occurrence in the same trajectories, such as the distance function introduced in Section 4.2. Here, places or links serve as the graph vertices, and the values of the similarity measure serve as the weights of the edges. However, since this representation captures only pairwise connections, the results of applying network analysis methods may not carry the same semantics as those of topic modelling. For instance, a network analysis method might identify a community of highly connected nodes A, B, and C. Yet, there may be no trajectory in the dataset visiting all three nodes. In contrast, a topic with high weights on nodes A, B, and C indicates that all three nodes frequently appear in the same trajectories.

Limitations. Major limitations and problematic aspects of the approach arise from two factors: first, the necessity to represent the spatial component of the data in a discrete manner, which may introduce artificial boundaries and distort patterns and, second, the inability of topic modelling to take into account the spatial neighbourhood and topological relationships between places and between links. For a topic modelling algorithm, two neighbouring places or two links with a common origin are just two distinct "terms". The algorithm may find out that they are somehow related only if they both often occur in the same trajectories. This may not be a big problem for quasi-continuous trajectories where consecutively visited places are spatial neighbours and, hence, the spatial relationships are represented well enough by the co-occurrence relationships. However, it may be quite different for episodic trajectories, as can be seen in Section 5. The consequence of the ignorance of the spatial relationships was the derivation of "excessive" topics, i.e., multiple topics with very similar spatial distributions. While their similarity is obvious to a human analyst, they are very dissimilar for the algorithm, since their cores consist of distinct groups of places or links.

We encountered a similar problem in an earlier work, where we applied topic modelling to symbolically encoded discretised time series [49]. Currently, we do not know any way to solve such problems algorithmically. What can be done is interactive aggregation of similar topics and the use of the integrated topics in the further analysis, as we did for the time series.

Relation to previous works. To a large extent, out research was inspired by prior works [50, 51], where topic modelling was applied to trajectories represented as sets of traversed streets. The aim of the analysis was to detect frequently taken routes. Unlike in these works, our primary focus is the space viewed as a system of places visited by moving entities. The aim of the analysis is to understand how the places are interlinked by the movements between them and what kinds of spatial structures are formed by the places and links. For this purpose, we consider topics as spatial distributions and strive to find interpretable patterns in these distributions.

Extending the research scope compared to the prior works, we considered different types of trajectories: quasi-continuous and episodic, network-constrained and unconstrained. We compared the approach involving topic modelling with an alternative approach involving embedding of places according to the strengths of their pairwise associations derived from co-occurrences in the same trajectories. We also proposed visualisation methods for representing topic modelling results on maps and investigated how the visual representations facilitate interpretation of the topics.

Following the prior works, our study confirms that topic modelling is a powerful analytical instrument for analysis of movement data. Dual representation of trajectories as place visits and as transitions allows considering space use from different perspectives, opening new opportunities for discovering and relating patterns of different types [78].

Directions for future research. In the current approach, topic modelling is applied independently to the places and to the links for obtaining different kinds of information. There is no formal way to match place-based and link-based topics; it can only be done through visual exploration based on noticed similarities between the spatial distributions. However, it can be taken into account that terms representing links include place identifiers; hence, topics obtained from lists of links implicitly refer to the places connected by the links. This gives an opportunity to assess the importance of the places for each of the link-based topics. A simple approach is to compute the weights of the places by applying some aggregation operator, such as maximum or sum, to the weights of the incoming and outgoing links. Besides, since the link-based topics can be treated as spatial graphs, it is also possible to assess the place importance using the existing measures of graph node centrality [79], particularly, the ones specifically devised for edge-weighted directed networks [80, 81]. While these ideas are easy to implement technically, research is required to understand how each

measure should be interpreted in the context of a link-based topic.

Our approach can be extended to analysing changes in space use over time. To do this, it is necessary to break down the sequences of place visits and transitions into subsequences corresponding to different time periods and compute the topic weights for these subsequences based on the occurring terms.

8. Conclusion

We have devised, implemented and tested an approach to uncovering and analysing patterns of utilising space as medium for movement. The core of the approach is application of topic modelling techniques to trajectories transformed to lists of visited places and lists of transitions between the places. We investigated the potential of the approach using data sets with different properties. It has demonstrated its capability to reveal several types of patterns: regions of high internal connectivity, areas where different regions interconnect, major pathways of moving through the space, activity centres and their connections to surrounding areas. We described the workflow of the movement data analysis employing our approach and developed recommendations for applying it to different kinds of movement data. Overall, we can conclude that application of topic modelling to movement data provides useful new, previously unexplored possibilities for analysing movement as a spatial phenomenon at a desired level of spatial abstraction, which can promote insights into the spatial structure of the movement and its relationships with the underlying space.

Acknowledgements

This work was supported by Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia as part of the *Lamarr Institute for Machine Learning and Artificial Intelligence* (Lamarr22B), and by EU in projects *SoBigData++* and *CrexData* (grant agreement no. 101092749).

References

[1] Andrienko, G, Andrienko, N, Bak, P, Keim, D, Wrobel, S. Visual Analytics of Movement. Springer; 2013. doi:10.1007/978-3-642-37583-5.

[2] Andrienko, N, Andrienko, G, Fuchs, G, Jankowski, P. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. Information Visualization 2016;15(2):117–153. doi:10.1177/1473871615581216.

[3] Macdonald, K, Grieco, M. Accessibility, mobility and connectivity: The changing frontiers of everyday routine. Mobilities 2007;2(1):1–14.

[4] Kindlmann, P, Burel, F. Connectivity measures: a review. Landscape ecology 2008;23:879–890.

[5] Chen, BY, Wang, Y, Wang, D, Li, Q, Lam, WH, Shaw, SL. Understanding the impacts of human mobility on accessibility using massive mobile phone tracking data. Annals of the American Association of Geographers 2018;108(4):1115–1133. doi:10.1080/24694452.2017.1411244.

[6] Li, Z, Huang, X, Ye, X, Jiang, Y, Martin, Y, Ning, H, et al. Measuring global multi-scale place connectivity using geotagged social media data. Scientific reports 2021;11(1):14694. doi:10.1038/s41598-021-94300-7.

[7] Andrienko, G, Andrienko, N, Hecker, D. Extracting Movement-based Topics for Analysis of Space Use. In: El-Assady, M, Angelini, M, editors. EuroVis Workshop on Visual Analytics (EuroVA). The Eurographics Association; 2023,doi:10.2312/eurova.20231091.

[8] Demšar, U, Buchin, K, Cagnacci, F, Safi, K, Speckmann, B, Van de Weghe, N, et al. Analysis and visualisation of movement: an interdisciplinary review. Movement Ecology 2015;3(1):1–24. doi:10.1186/s40462-015-0032-y.

[9] Andrienko, N, Andrienko, G. Spatial generalization and aggregation of massive movement data. IEEE Transactions on Visualization & Computer Graphics 2011;17(02):205–219. doi:10.1109/TVCG.2010.44.

[10] Andrienko, N, Andrienko, G, Rinzivillo, S. Exploiting spatial abstraction in predictive analytics of vehicle traffic. ISPRS International Journal of Geo-Information 2015;4(2):591–606. doi:10.3390/ijgi4020591.

[11] Andrienko, N, Andrienko, G, Stange, H, Liebig, T, Hecker, D. Visual analytics for understanding spatial situations from episodic movement data. KI - Künstliche Intelligenz 2012;26(3):241–251. doi:10.1007/s13218-012-0177-4.

[12] Vayansky, I, Kumar, SA. A review of topic modeling methods. Information Systems 2020;94:101582. doi:10.1016/j.is.2020.101582.

[13] Blei, DM, Ng, AY, Jordan, MI. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3(Jan):993–1022.

[14] Luo, M, Nie, F, Chang, X, Yang, Y, Hauptmann, A, Zheng, Q. Probabilistic non-negative matrix factorization and its robust extensions for topic modeling. In: Thirty-first AAAI conference on artificial intelligence. 2017,.

[15] Liu, L, Tang, L, Dong, W, Yao, S, Zhou, W. An overview of topic modeling and its current applications in bioinformatics. SpringerPlus 2016;5(1):1–22. doi:10.1186/s40064-016-3252-8.

[16] Wallach, HM. Topic modeling: Beyond bag-of-words. In: Proceedings of the 23rd International Conference on Machine Learning. ICML '06; New York, NY, USA: Association for Computing Machinery; 2006, p. 977–984. doi:10.1145/1143844.1143967.

[17] Wang, S, Bao, Z, Culpepper, JS, Cong, G. A survey on trajectory data management, analytics, and learning. ACM Computing Surveys 2021;54(2). doi:10.1145/3440207.

[18] Zheng, Y. Trajectory data mining: An overview. ACM Transactions on Intelligent Systems and Technology 2015;6(3). doi:10.1145/2743025.

[19] Mazimpaka, JD, Timpf, S. Trajectory data mining: A review of methods and applications. Journal of Spatial Information Science 2016;2016(13):61–99. doi:10.5311/JOSIS.2016.13.263.

[20] Andrienko, G, Andrienko, N, Chen, W, Maciejewski, R, Zhao, Y. Visual analytics of mobility and transportation: State of the art and further research directions. IEEE Transactions on Intelligent Transportation Systems 2017;18(8):2232–2249. doi:10.1109/TITS.2017.2683539.

[21] Liu, H, Chen, X, Wang, Y, Zhang, B, Chen, Y, Zhao, Y, et al. Visualization and visual analysis of vessel trajectory data: A survey. Visual Informatics 2021;5(4):1–10. doi:10.1016/j.visinf.2021.10.002.

[22] Andrienko, N, Andrienko, G, Patterson, F, Stange, H. Visual analysis of place connectedness by public transport. IEEE Transactions on Intelligent Transportation Systems 2020;21(8):3196–3208. doi:10.1109/TITS.2019.2924796.

[23] Palomo, C, Guo, Z, Silva, CT, Freire, J. Visually exploring transportation schedules. IEEE Transactions on Visualization and Computer Graphics 2016;22(1):170–179. doi:10.1109/TVCG.2015.2467592.

[24] Guo, D, Chen, J, MacEachren, AM, Liao, K. A visualization system for space-time and multivariate patterns (vis-stamp). IEEE Transactions on Visualization and Computer Graphics 2006;12(6):1461–1474. doi:10.1109/TVCG.2006.84.

[25] Wood, J, Dykes, J, Slingsby, A. Visualisation of origins, destinations and flows with OD maps. The Cartographic Journal 2010;47(2):117–129. doi:10.1179/000870410X12658023467367.

[26] Teitelbaum, CS, Hepinstall-Cymerman, J, Kidd-Weaver, A, Hernandez, SM, Altizer, S, Hall, RJ. Urban specialization reduces habitat connectivity by a highly mobile wading bird. Movement ecology 2020;8(1):1–13. doi:10.1186/s40462-020-00233-7.

[27] Fahrig, L. Effects of habitat fragmentation on biodiversity. Annual review of ecology, evolution, and systematics 2003;34(1):487–515. doi:10.1146/annurev.ecolsys.34.011802.132419.

[28] Yuan, G, Sun, P, Zhao, J, Li, D, Wang, C. A review of moving object trajectory clustering algorithms. Artificial Intelligence Review 2017;47:123–144. doi:10.1007/s10462-016-9477-7.

[29] Rinzivillo, S, Mainardi, S, Pezzoni, F, Coscia, M, Pedreschi, D, Giannotti, F. Discovering the geographical borders of human mobility. KI - Künstliche Intelligenz 2012;26(3):253–260. doi:10.1007/s13218-012-0181-8.

[30] Brilhante, IR, Berlingerio, M, Trasarti, R, Renso, C, Macedo, JAFd, Casanova, MA. Cometogether: Discovering communities of places in mobility data. In: 2012 IEEE 13th International Conference on Mobile Data Management. 2012, p. 268–273. doi:10.1109/MDM.2012.17.

[31] Fortunato, S. Community detection in graphs. Physics Reports 2010;486(3):75–174. doi:10.1016/j.physrep.2009.11.002.

[32] Vieira, VdF, Xavier, CR, Evsukoff, AG. A comparative study of overlapping community detection methods from the perspective of the structural properties. Applied Network Science 2020;5(1):51. doi:10.1007/s41109-020-00289-9.

[33] Van Der Maaten, L, Postma, EO, van den Herik, HJ, et al. Dimensionality reduction: A comparative review. Journal of Machine Learning Research 2009;10(66-71):13.

[34] Duran, BS, Odell, PL. Cluster analysis: a survey; vol. 100. Springer Science & Business Media; 2013. doi:10.1007/978-3-642-46309-9.

[35] Wenskovitch, J, Crandell, I, Ramakrishnan, N, House, L, North, C. Towards a systematic combination of dimension reduction and clustering in visual analytics. IEEE transactions on visualization and computer graphics 2017;24(1):131–141. doi:10.1109/TVCG.2017.2745258.

[36] Wenskovitch Jr, JE. Dimension reduction and clustering for interactive visual analytics. Ph.D. thesis; Virginia Tech; 2019.

[37] Huang, Z, Witschard, D, Kucher, K, Kerren, A. Va + embeddings star: A state-of-the-art report on the use of embeddings in visual analytics. Computer Graphics Forum 2023;42(3):539–571. doi:10.1111/cgf.14859.

[38] Dzemyda, G, Kurasova, O, Zilinskas, J. Multidimensional data visualization. methods and applications. Springer optimization and its applications 2013;75. doi:10.1007/978-1-4419-0236-8.

[39] Nonato, LG, Aupetit, M. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. IEEE Transactions on Visualization and Computer Graphics 2019;25(8):2650–2673. doi:10.1109/TVCG.2018.2846735.

[40] Irani, J, Pise, N, Phatak, M. Clustering techniques and the similarity measures used in clustering: A survey. International journal of computer applications 2016;134(7):9–14. doi:10.5120/ijca2016907841.

[41] Lesot, MJ, Rifqi, M, Benhadda, H. Similarity measures for binary and numerical data: a survey. International Journal of Knowledge Engineering and Soft Data Paradigms 2009;1(1):63–84. doi:10.1504/IJKESDP.2009.021985.

[42] Boriah, S, Chandola, V, Kumar, V. Similarity measures for categorical data: A comparative evaluation. In: Proceedings of the 2008 SIAM international conference on data mining. SIAM; 2008, p. 243–254.

[43] Rieck, K. Similarity measures for sequential data. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2011;1(4):296–304. doi:10.1002/widm.36.

[44] Schwering, A. Approaches to semantic similarity measurement for geospatial data: a survey. Transactions in GIS 2008;12(1):5–29. doi:10.1111/j.1467-9671.2008.01084.x.

[45] Cassisi, C, Montalto, P, Aliotta, M, Cannata, A, Pulvirenti, A, et al. Similarity measures and dimensionality reduction techniques for time series data mining. Advances in data mining knowledge discovery and applications 2012;:71–96doi:10.5772/49941.

[46] El-Assady, M, Sperrle, F, Deussen, O, Keim, D, Collins, C. Visual analytics for topic model optimization based on user-steerable speculative execution. IEEE Transactions on Visualization and Computer Graphics 2019;25(1):374–384. doi:10.1109/TVCG.2018.2864769.

[47] Chen, S, Andrienko, N, Andrienko, G, Adilova, L, Barlet, J, Kindermann, J, et al. LDA ensembles for interactive exploration and categorization of behaviors. IEEE Transactions on Visualization and Computer Graphics 2020;26(9):2775–2792. doi:10.1109/TVCG.2019.2904069.

[48] Chen, TH, Thomas, SW, Hassan, AE. A survey on the use of topic models when mining software repositories. Empirical Software Engineering 2016;21(5):1843–1919. doi:10.1007/s10664-015-9402-8.

[49] Andrienko, N, Andrienko, G, Shirato, G. Episodes and topics in multivariate temporal data. Computer Graphics Forum 2023;42(6):e14926. doi:10.1111/cgf.14926.

[50] Chu, D, Sheets, DA, Zhao, Y, Wu, Y, Yang, J, Zheng, M, et al. Visualizing hidden themes of taxi movement with semantic transformation. In: 2014 IEEE Pacific Visualization Symposium. 2014, p. 137–144. doi:10.1109/PacificVis.2014.50.

[51] Liu, H, Jin, S, Yan, Y, Tao, Y, Lin, H. Visual analytics of taxi trajectory data via topical sub-trajectories. Visual Informatics 2019;3(3):140–149. doi:10.1016/j.visinf.2019.10.002.

[52] Cleveland, WS, McGill, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association 1984;79(387):531–554. doi:10.1080/01621459.1984.10478080.

[53] Spence, I. No humble pie: The origins and usage of a statistical chart. Journal of Educational and Behavioral Statistics 2005;30. doi:10.3102/10769986030004353.

[54] Robinson, AH. The thematic maps of charles joseph minard. Imago Mundi 1967;21:95–108. URL: http://www.jstor.org/stable/1150482.

[55] Bernard, J, Steiger, M, Mittelstädt, S, Thum, S, Keim, D, Kohlhammer, J. A survey and task-based quality assessment of static 2D colormaps. In: Kao, DL, Hao, MC, Livingston, MA, Wischgoll, T, editors. Visualization and Data Analysis; vol. 9397. International Society for Optics and Photonics; SPIE; 2015, p. 93970M. doi:10.1117/12.2079841.

[56] Andrienko, G, Andrienko, N, Fuchs, G, Wood, J. Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. IEEE Transactions on Visualization and Computer Graphics 2017;23(9):2120–2136. doi:10.1109/TVCG.2016.2616404.

[57] Bernard, J, Wilhelm, N, Scherer, M, May, T, Schreck, T. TimeSeriesPaths: Projection-based explorative analysis of multivariate time series data. In: Journal of WSCG. 2012, p. 97–106.

[58] Andrienko, N, Andrienko, G. It's about time: Analytical time periodization. Computer Graphics Forum 2023;42(6):e14845. doi:10.1111/cgf.14845.

[59] Kruskal, JB. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. Psychometrika 1964;29(1):1–27. doi:10.1007/BF02289565.

[60] van der Maaten, L, Hinton, G. Visualizing data using t-SNE. Journal of Machine Learning Research 2008;9(86):2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.

[61] Ayesha, S, Hanif, MK, Talib, R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion 2020;59:44–58. doi:10.1016/j.inffus.2020.01.005.

[62] Ankerst, M, Breunig, MM, Kriegel, HP, Sander, J. Optics: Ordering points to identify the clustering structure. In: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data. SIGMOD '99; New York, NY, USA: Association for Computing Machinery; 1999, p. 49–60. doi:10.1145/304182.304187.

[63] Tobler, W. Experiments in migration mapping by computer. Cartography and Geographic Information Science 1987;14:155–163. doi:10.1559/152304087783875273.

[64] Wattenberg, M, Viégas, F, Johnson, I. How to use t-sne effectively. Distill 2016;doi:10.23915/distill.00002.

[65] Espadoto, M, Martins, RM, Kerren, A, Hirata, NST, Telea, AC. Toward a quantitative survey of dimension reduction techniques. IEEE Transactions on Visualization and Computer Graphics 2021;27(3):2153–2173. doi:10.1109/TVCG.2019.2944182.

[66] Bernard, J, Dobermann, E, Bögl, M, Röhlig, M, Vögele, A, Kohlhammer, J. Visual-interactive segmentation of multivariate time series. In: Proceedings of the EuroVis Workshop on Visual Analytics. Goslar, DEU: Eurographics Association; 2016, p. 31–35. doi:10.2312/eurova.20161121.

[67] Guo, D. Flow mapping and multivariate visualization of large spatial interaction data. IEEE Transactions on Visualization and Computer Graphics 2009;15(6):1041–1048. doi:10.1109/TVCG.2009.143.

[68] Jo Wood, JD, Slingsby, A. Visualisation of origins, destinations and flows with od maps. The Cartographic Journal 2010;47(2):117–129. doi:10.1179/000870410X12658023467367.

[69] von Landesberger, T, Brodkorb, F, Roskosch, P, Andrienko, N, Andrienko, G, Kerren, A. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. IEEE Transactions on Visualization and Computer Graphics 2016;22(1):11–20. doi:10.1109/TVCG.2015.2468111.

[70] Khaled, S. A guide to football pitch zones. https://seifkhaled.me/a-guide-to-football-pitch-zones-download-pdf/; 2023. Ac-

cessed: July 14, 2023.
[71] Andrienko, N, Andrienko, G, Fuchs, G, Jankowski, P. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. Information Visualization 2016;15(2):117–153. doi:10.1177/1473871615581216.
[72] Albalawi, R, Yeap, TH, Benyoucef, M. Using topic modeling methods for short-text data: A comparative analysis. Frontiers in Artificial Intelligence 2020;3. doi:10.3389/frai.2020.00042.
[73] Egger, R, Yu, J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts. Frontiers in Sociology 2022;7. doi:10.3389/fsoc.2022.886498.
[74] Bellman, R. Dynamic programming. Science 1966;153(3731):34–37.
[75] Aggarwal, CC, Hinneburg, A, Keim, DA. On the surprising behavior of distance metrics in high dimensional space. In: Van den Bussche, J, Vianu, V, editors. Database Theory — ICDT 2001. Berlin, Heidelberg: Springer Berlin Heidelberg; 2001, p. 420–434. doi:10.1007/3-540-44503-X_27.
[76] Chen, L. Curse of Dimensionality. Boston, MA: Springer US; 2009, p. 545–546. doi:10.1007/978-0-387-39940-9_133.
[77] Brandes, U, Erlebach, T. Network analysis: methodological foundations; vol. 3418. Springer Science & Business Media; 2005.
[78] Andrienko, N, Andrienko, G, Miksch, S, Schumann, H, Wrobel, S. A theoretical model for pattern discovery in visual analytics. Visual Informatics 2021;5(1):23–42. doi:10.1016/j.visinf.2020.12.002.
[79] Gómez, S. Centrality in Networks: Finding the Most Important Nodes. Cham: Springer International Publishing; 2019, p. 401–433. doi:10.1007/978-3-030-06222-4_8.
[80] Kaur, S, Gupta, A, Saxena, R. Identifying central nodes in directed and weighted networks. International Journal of Advanced Computer Science and Applications 2021;12(8). doi:10.14569/IJACSA.2021.01208100.
[81] Zhang, P, Wang, T, Yan, J. Pagerank centrality and algorithms for weighted, directed networks. Physica A: Statistical Mechanics and its Applications 2022;586:126438. doi:10.1016/j.physa.2021.126438.