City Research Online

City St George's, University of London

Citation: Von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G. & Kerren, A. (2016). MobilityGraphs: Visual Analysis of Mass Mobility Dynamics via Spatio-Temporal Graphs and Clustering. IEEE Transactions on Visualization and Computer Graphics, 22(1), pp. 11-20. doi: 10.1109/tvcg.2015.2468111

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://city-ref-test.eprints-hosting.org/id/eprint/12849/

Link to published version: https://doi.org/10.1109/tvcg.2015.2468111

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

MobilityGraphs: Visual Analysis of Mass Mobility Dynamics via Spatio-Temporal Graphs and Clustering

Tatiana von Landesberger, Felix Brodkorb, Philipp Roskosch, Natalia Andrienko,
Gennady Andrienko, and Andreas Kerren, *Senior Member, IEEE*

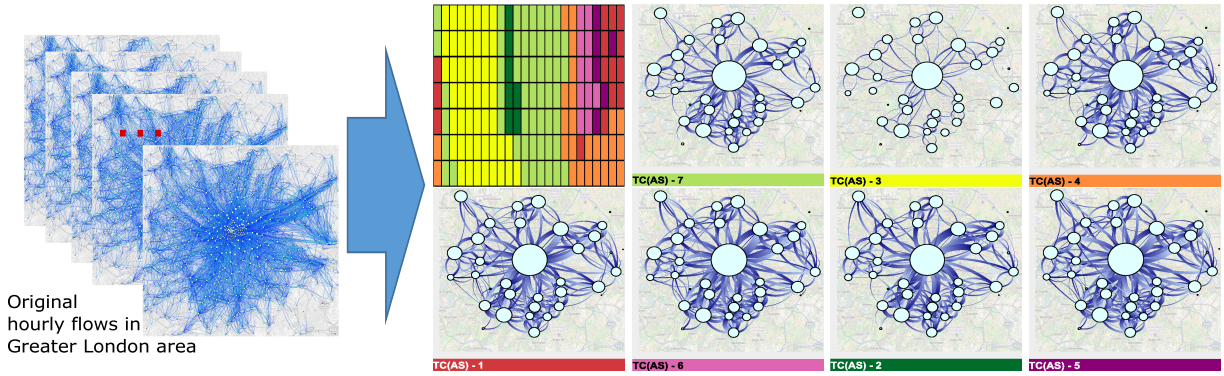


Fig. 1. Hourly flows of residential Twitter users in Greater London area over a week are spatially and temporally simplified to gain an overview of mobility dynamics. A calendar view shows the temporal cluster distribution over a week. Each map displays movements between spatial aggregates in a temporal cluster. Movement direction is represented by a color gradient from dark to light blue.

Abstract—Learning more about people mobility is an important task for official decision makers and urban planners. Mobility data sets characterize the variation of the presence of people in different places over time as well as movements (or flows) of people between the places. The analysis of mobility data is challenging due to the need to analyze and compare spatial situations (i.e., presence and flows of people at certain time moments) and to gain an understanding of the spatio-temporal changes (variations of situations over time). Traditional flow visualizations usually fail due to massive clutter. Modern approaches offer limited support for investigating the complex variation of the movements over longer time periods.

We propose a visual analytics methodology that solves these issues by combined spatial and temporal simplifications. We have developed a graph-based method, called MobilityGraphs, which reveals movement patterns that were occluded in flow maps. Our method enables the visual representation of the spatio-temporal variation of movements for long time series of spatial situations originally containing a large number of intersecting flows. The interactive system supports data exploration from various perspectives and at various levels of detail by interactive setting of clustering parameters. The feasibility and utility of our approach was tested on aggregated mobility data derived from a set of geolocated Twitter posts within the Greater London city area and mobile phone call data records in Abidjan, Ivory Coast. We could show that MobilityGraphs support the identification of regular daily and weekly movement patterns of resident population.

Index Terms—Visual analytics, movement data, networks, graphs, temporal aggregation, spatial aggregation, flows, clustering

1 Introduction

An understanding of the movement behavior of people and its dynamics is crucial for official decision makers and urban planners, among others. Mobility data sets give information about the presence of people in different places at certain points in time as well as about movements (or flows) of people between the places. Movement data can be obtained from various sources—such as traffic sensors, public transportation usage records—or reconstructed from many kinds of data sources that contain people's locations at different times despite not being collected with the purpose of tracking people's movements. Examples of the latter type of data include records about the use of mobile phones collected for billing purposes, and geographically referenced contents appearing in social media: Twitter, Flickr, YouTube, Foursquare, etc. In particular, Twitter data sets have gained prominence lately due to their wide availability and large scale. Even though, geographically referenced tweets only cover a small percentage of all tweets, the absolute amount of them is very large and thus provides a sufficiently representative sample [4]. The individual data is often pre-processed (e.g., summarized) for gaining time-varying flows of multiple people (mass movement data). We refer to them as 'movement data' or 'time-varying flow data'.

Movement data consists of records reporting the number of people in an area in a short time interval, so-called *time step* (e.g., people in the City of London between 9am and 10am). The data also contains the amount of movements (i.e., flow magnitudes) between areas during specific time periods (e.g., hourly movements between city suburbs). The data is usually collected over longer time periods, thus it is time-varying.

The goal of movement analysis is to understand how people move over the territory throughout a certain time period (e.g., movements in London during the weekly cycle). The analysis can focus on typical or extraordinary movements. Typical movements are often analyzed by urban planners for predicting common flows. Extraordinary

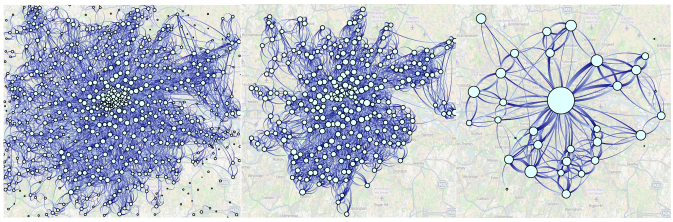
- *T. von Landesberger, F. Brodkorb and P. Roskosch are with Technical University of Darmstadt, Germany. E-mail: {tatiana.von-landesberger|felix.brodkorb}@gris.tu-darmstadt.de.*
- *Natalia Andrienko and Gennady Andrienko are with Fraunhofer Institute IAIS, Bonn, Germany and with City University, London, UK. E-mail: {natalia|gennady}.andrienko@iais.fraunhofer.de.*
- *Andreas Kerren is with Linnaeus University, Växjö, Sweden. E-mail: kerren@acm.org.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx xxx 2015; date of current version xx xxx 2015.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

movements occur in specific occasions (e.g., demonstrations or bomb attacks). They are important for emergency management. We focus on analyzing typical movements addressing the following questions:

- How do the *spatial situations* (i.e., the spatial distributions of the flow magnitudes) at particular time moments within a regularly week cycle (e.g., working days in the morning) look like? How big are the flows between urban areas at a certain point in time? Which areas have many people at a certain time moment? Which areas have high/low flows at a certain point in time?
- What is the *temporal variation* of the flows? How do spatial situations change over time (e.g., over a typical week)? What is the difference between the spatial situations in two time moments? Which flows increase/decrease?

Traditional visual analysis methods for assessing this type of data usually fail due to massive clutter and/or the inability to investigate the complex variation of the movements over time. A widely employed flow map visualization gets extremely cluttered when there are flows between non-neighboring regions (see Fig. 2(a)). Even filtering by flow magnitude or increasing edge transparency does not solve this problem (see Fig. 2(b)). Moreover, visualization of temporal data variation using animation or small multiples do not scale [7, 31, 41]. Clustering and aggregation-based visual analytics approaches mostly focus on spatial or temporal components of the flows. They have limited suitability for dense flows with many time moments (see Sect. 2).



(a) Original flows (b) Filtered flows (c) Aggregated flows

Fig. 2. Afternoon movements of the London residents (see Sect. 5.1). Original flow maps are cluttered. Filtered flow map (> 200 movements) is still cluttered. Spatial simplification (c) clearly shows important flows and their magnitudes (edge width, see also Sect. 4).

We propose a visual analytics methodology that addresses the two above-mentioned analytical questions. Our approach relies on spatial and temporal simplification that uses a spatio-temporal *mobility graph* as a representation of movement data. Here, a discrete set of places (geographical regions) with flows between them is considered as a weighted directed graph. Edge weights are determined by flow magnitudes. Our graph is dynamic, as spatial situations vary over time.

The mobility graph is spatially simplified by a specifically developed spatial graph clustering algorithm. The temporal simplification is achieved through clustering of time steps based on the similarity of the respective spatial situations. The results of the spatial and temporal simplifications are presented in several coordinated views for comprehensive data exploration. Based on our observations, these methods notably reduce the visualized data size and thus help to mitigate clutter problems as well as to visually represent the temporal variation of movement for a large number of time steps. Moreover, the proposed fast clustering algorithm allows for interactive change of parameters and thus supports interactive exploration of movements at various degrees of abstraction and from different perspectives. This assists in gaining a deep understanding of spatial situations and their changes.

Our approach can be used for analyzing any flow (i.e., collective movement) data over time, such as population migration, commuting, or flows of attendees of public events. The feasibility and usefulness of our approach has been tested on two real-world flow data sets. The first data set concerns the resident Twitter user mobility in Greater London area. The analysis of hourly movements throughout the weekly time cycle has revealed typical movement patterns such as strong city-centric movement of London residents in the morning, quiet movement periods during the working hours, and correspondence of movement patterns to the rail network topology. Most of the uncovered patterns could not be perceived from flow map displays. These results confirm the analytical power of our methodology even when using Twitter data approximating the population movement. The second real-world data set was produced from call data records over two weeks in Abidjan, Ivory Coast. The analysis of hourly movements identified similar but slightly less prominent routine mobility patterns with differentiation of morning and evening flows. We found that the city has a polycentric structure, where one center is the location of businesses and jobs and the other two have a different role.

The remainder of this paper is organized as follows. Section 2 discusses the related works. Then, we present our visual analytics approach together with design decisions in Section 4. Section 5 demonstrates the feasibility of the proposed methods on two real-world datasets. Section 6 discusses our approach. We conclude our paper with an outline of future work in Section 7.

2 Related Work

As our approach covers several research fields. We subdivide the related work into two groups: (1) visualization approaches for time-varying flow data and (2) an overview of techniques for the visual analysis of flows represented as geolocated dynamic graphs.

Visualization of time-varying flow data Much work has been done in the area of time-oriented data [2]. As for flow data, the existing visualization methods have been recently surveyed by Boyandin [13] and Andrienko [4]. There are two main classes of techniques for the visual representation of flow situations: flow map [36, 43] and origin-destination matrix (OD matrix) [52, 53].

The common approach to representing time-varying flow data is to visualize the *time series of flow magnitudes* associated with the links on a non-cartographic time series display, e.g., as it is done by Boyandin et al. [14]. However, such a visualization does not show the spatial patterns of the flows, i.e., the spatial flow situations. Bremm et al. [15] propose another kind of non-cartographic visualization of flows, where locations are represented by distinct colors. The flows between the places are represented by colored bands connecting segments of neighboring bars. When the time series are long, several complementary techniques for selecting representative time moments can be applied [48]. An obvious limitation of this method is scalability w.r.t. the number of places.

Time-variant spatial flow situations can be represented by *multiple maps* arranged either temporally in a map animation or spatially in a small multiple maps display [3]. Alternatively, one can use different operations on the space-time cube to analyze the flows [8]. *Flow maps* show the spatial context of flow data in an undistorted way but often suffer from numerous intersections, occlusions, and visual clutter. The problems of flow mapping are comprehensively discussed by Guo and Zhou [28]. Moreover, map animation may be ineffective [31, 44], because the user cannot memorize and mentally compare multiple spatial situations. In small multiples, a limited number of spatial situations can be shown simultaneously. Hence, this approach is not suitable for long time series. Therefore, simplification of individual maps or reduction of time moments has been proposed.

Reducing or simplifying the spatial data can be done by filtering (see Fig. 2(b)), when only flows with magnitudes above a chosen threshold are shown [27, 39, 43], or by reducing the opacity of minor flows [53]. Beecham and Wood [11] use an interactive based approach where a user can visually select spatial and temporal areas for filtering or highlighting to analyse different attributes of (parts of) the data. Van den Elzen and van Wijk [45] visualize only flows between a few user-selected locations. Obviously, these approaches may hide possibly relevant information. Another approach is edge bundling, i.e., merging or grouping spatially close flows [16, 22, 33, 54]. The edge bundling approaches work well only for showing flows from one origin or to one destination, or in special cases, such as prevalence of

radial movements from and to one central location [22]. Besides, map users can wrongly interpret the flow lines resulting from edge bundling as the routes of the movement.

Locations and flows can be *aggregated* for reducing the size of the represented data, while still being able to uncover general patterns of flows. Kisilevich et al. [35] describe state-of-the-art approaches of clustering for geospatial data. Guo [26, 27] applies spatially constrained hierarchic graph partitioning techniques to group places into larger units and then visualizes aggregated flows between these larger units. Guo and Zhu [28] point to the modifiable areal unit problem (MAUP) [38]. They propose kernel-based density estimation, which also normalizes the flow magnitudes. Filtering still needs to be applied for reducing the intersections of the resulting aggregated flows, i.e., only the most important flows are shown on the map. These approaches are suitable for spatial simplification, however they have a relatively high computational complexity. Moreover, they do not consider time-dependence of flows. Gao et al. apply the hierarchic clustering of Guo et al. [27] to time-dependent flows [25]. They cluster each time moment individually. This creates problems with automatically determining cluster correspondence over time and thus makes it difficult to identify changes between time moments. Correspondence needs to be found manually. Zhu and Guo [56] use hierarchic clustering of flows (not places) to analyze movement data. As each time moment needs to be inspected, these two approaches do not scale for large time periods.

Clustering of spatial situations over time [4, 5] can be used to reduce the number of distinct situations that need to be shown. Clustering puts together similar flow situations corresponding to different time moments. The individual spatial situations grouped into the clusters are then summarized by computing the mean or median flows. The summarized situations representing different clusters can be shown on small multiple flow maps. However, the problem of visual clutter and over-plotting of flow symbols on the flow maps was not solved. In the examples occurring in the previous works [4, 5], the flow maps were not too cluttered owing to either a small number of distinct locations or the absence of flow intersections (i.e., there were only flows between neighboring places).

Time-varying flows as geolocated dynamic graphs Movements between regions can be represented as a graph, where regions are graph nodes and flows between them are treated as weighted directed edges. The node position in the graphs is fixed to its geographic location. Time-dependence of the flows creates dynamic geolocated graphs. So, the assessment of movement data is closely related to the visualization of dynamic networks with geolocation constraints on node positions. The available methods for visualization of (potentially multivariate) dynamic graphs are comprehensively overviewed in recent surveys [6, 10, 29, 50, 51]. Here, we concentrate on works related to time-varying flow visualization, i.e., dynamic graphs.

Animation in dynamic graph visualization is often connected with the problem of creating a stable graph layout for supporting the users' mental map [7], such as foresighted layouts [20]. In flow visualization, however, the node positions are fixed. Nevertheless, animation can still pose a challenge for the user, due to high cognitive load, when perceiving changes between consecutive time steps [7, 44]. This problem is especially severe when many time moments or large/dense graphs need to be analyzed. Approaches mitigating this problem include reducing the number of time steps and highlighting of changes between consecutive graphs [9], usage of smooth edge bundling [34], drawing of clustered graphs or filtered graphs [1, 24]. These approaches still suffer from animation perception problems.

Timeline visualization of dynamic graphs shows all graphs in a sequence. For large graphs and/or a large number of time moments, a scalability challenge emerges. In the network visualization community, this is often mitigated by suitable node layouts [18]. This is however not applicable to geolocated graphs, as the understanding of geographic distribution is diminished. Alternatively, one can cluster time moments by similarity of corresponding graphs. This is analogical to the clustering of flow maps (see above). In juxtaposed (small multiples) views, the space for individual graph visualization is very limited. Even when clustering time moments, large and dense graphs often get cluttered in small multiples views. This can be further diminished by graph aggregation (clustering of graph nodes). It aggregates compact graph structures (e.g., communities) or nodes with similar attributes (e.g., persons of similar age) into one aggregated node and thus reduce the size of the graph for visualization [42, 47]. These algorithms exploit the graph structure, but often disregard geographic closeness of the nodes. This may result in clusters which are "scattered" in geographic space.

Superimposed visualization draws graphs on top of each other in a 2.5D display [21]. This strategy, however, suffers from occlusion, especially for dense graphs and many time moments.

Integrated approaches use specialized visual designs to show the time dimension. For example, Reitz [40] shows the time evolution of edge weights by colored sections within edges. While this enables the assessment of edge weight evolution directly on edges, it poorly performs for intersecting edges, as in the case of geolocated flows. It also does not show how the overall graph structure evolves over time.

Matrix-based approaches show dynamic graphs as adjacency matrices (i.e., OD matrices), where intra cell timelines such as sparklines [17, 55] show the dynamics of edge and node weights. Owing to the limited space, they are limited to short time periods. For large graphs, graph aggregation is needed [17]. Moreover, only the local changes of the edges and/or nodes are visible, but not the evolution of the overall graph structure.

3 Definitions

We consider time-varying flow data over a time interval $T = \{t_1, \ldots, t_\tau\}$. And we have a discrete set of spatial locations, a.k.a. places $P = \{p_1, \ldots, p_N\}$. Then, we define the number of people at place p_a in time step t_i as *presence count* $w_a(t_i)$ of this place.

There is a set of moving objects (such as people) that can move from any location to any other location. The routes of their movements are unknown, only the origins and destinations of the trips. For any ordered pair of places (p_a, p_b), all trips originating in p_a and ending in p_b in a time step t_i are called flow $f_{a,b}(t_i)$. A flow $f_{a,b}(t_i)$ from p_a to p_b in time step t_i exists only, if there was at least one trip originating in p_a and ending in p_b during t_i. We say that there is a directed link $e_{a,b}(t_i)$. If there exist a link $e_{a,b}(t_i)$ or $e_{b,a}(t_i)$, then the places p_a and p_b are *flow-connected*. The number of trips aggregated in a flow is called *flow magnitude* $m_{a,b}(t_i)$.

The time-varying flow data set can be formally represented as $< P, L, T, W(P,T), M(L,T) >$, where P is the set of places, $L \subset P \times P$ is the set of links, T is the set of time steps, $W(P,T)$ is the function that assigns presence count to the places p_a in time step t_i, and $M(L,T)$ is the function that assigns a certain non-negative value of flow magnitude to each pair (e, t_I), $e \in L$, and $t_i \in T$.

For each time step $t_i \in T$ and the whole set of links L, the function M gives a set of flow magnitudes $M(L,t_i)$, which is called *spatial flow situation* (or shorter: *spatial situation*) at time step t_i and is denoted with $S(t_i)$ or S_i.

A *simplified situation* (i.e., a *simplified spatial situation* or *aggregated situation*) AS_i, in time step t_i, is a situation, where similar places are grouped into *spatial aggregates* (i.e., regions) $R_l = \{\bigcup_{m \in l} P_m\}$, and the flows between them are aggregated accordingly into *aggregated flows AF* with magnitudes summarized according to the used clustering algorithm. The creation of spatial aggregates, e.g., by using a clustering algorithm is called *spatial simplification.*

A *temporal cluster* TC_k, $k \in \{1, \ldots, K\}$ is a group of possibly similar (simplified) spatial situations in various time steps $TC_k(S) = \{\bigcup_{i \in k} S_i\}$ or $TC_k(AS) = \{\bigcup_{i \in k} AS_i\}$. The creation of temporal clusters, e.g., by using a particular clustering algorithm is called *temporal simplification*. Temporal simplification can be done on the original situations $TC(S)$ or on simplified spatial situations $TC(AS)$.

4 Approach

In our approach, we address the following two visualization problems:

(1) Representation of the spatial situations significantly reducing the visual clutter while conveying important movement patterns.
(2) Representation of the temporal variation without the need to consider and compare a great number of time steps.

We propose a combination of spatial and temporal simplification with interactive exploration of the resulting data.

- *Spatial simplification* reduces visual clutter by the aggregation of regions using a special graph-based spatial clustering suitable for time-varying flows (see Sect. 4.1.1).
- *Temporal simplification* replaces the great number of spatial situations by a much smaller number of spatial situations that need to be viewed and compared by the analyst. We cluster the time steps by the similarity of the spatial situations (see Sect. 4.1.2).
- *Visual exploration* offers a set of interactive graph and geography-based views for the exploration of spatial situations and their changes over time. These views were specifically designed for the exploration of time-varying flow data with simplification (see Sect. 4.2). In particular, we have developed techniques supporting comparisons of spatial situations. Our approach also supports interactive adjustment of clustering parameters for the exploration of temporal flows from various perspectives and at various levels of detail.

4.1 Spatial and Temporal Simplification

Our approach combines spatial aggregation (i.e., aggregating nodes into regions and aggregating flows between places into flows between regions) and temporal clustering (i.e., grouping similar spatial situations from several time steps) for reducing the data size, reducing data clutter in the visualization, and simultaneously supporting an abstract grasp of characteristic features of the mobility behavior. The main challenge in the simplification was to develop appropriate clustering methods for time-varying flows and to combine the two simplifications.

While there are various ways of combining spatial and temporal simplification (see Sect. 6), we considered two main approaches:

(1) *Spatial aggregation preceding temporal clustering:* Spatial situations in all time steps are spatially aggregated first. This substantially reduces the data size for the subsequent temporal clustering, which uses flow magnitudes for calculating situation similarity ($<< N^2$).
(2) *Temporal clustering preceding spatial aggregation:* Temporal clustering based on original time-varying flows is performed before spatial aggregation. The temporal clustering reduces the number of time steps, and the subsequent spatial aggregation reduces data clutter in the visualization. The main problem of this approach is the determination of the similarity of spatial situations. The common approach using flows between all places [4] leads to high-dimensional data sets (up to N^2 flow magnitudes – dimensions). It leads to unstable results and long calculations. Thus, a dimensionality reduction is needed.

We investigated both types of simplification combinations with various ways of determining similarity of spatial situations in the temporal clustering. We decided to use the first option, because the second option requires feature reduction, esp. using general dimensionality reduction methods, whose results do not relate to the geographic context and therefore are not easy to interpret. In contrast, in the first option, spatial simplification is well understandable and easy to represent visually within the geographic context.

4.1.1 Spatial Simplification

We employ a specially developed method for aggregating places and aggregating flows between individual places into flows between the spatial aggregates (i.e., regions, or clusters of places). Our algorithm groups spatially close places with high flow magnitudes into regions. This aggregation greatly simplifies the graph structure. Inevitably, it hides the patterns of local movements between the places within the regions; however, it captures and prominently represents larger-scale movement patterns, which could not be seen otherwise (see Fig. 2(c) as an example).

Our spatial aggregation method extends the "Density-Based Spatial Clustering of Applications with Noise" (DBScan) [23] with additional features: (a) it considers both spatial closeness and strength of flows between places (i.e., flow strength) as well as (b) applicability to time-dependent data. Thus, our algorithm has been designed to work on time-dependent flows. The reasoning behind this approach as well as the algorithm's details are described below.

Algorithm rationale We decided to use a density-based clustering, because (a) it is fast ($O(|P|\dot{l}og(|P|))$ according to [23]), (b) it does not require pre-setting of the number of clusters, (c) it is able to detect arbitrary shaped clusters (in our case spatial aggregates, or regions) as well as outliers, and (d) it uses easily comprehensible parameters (e.g., spatial closeness, flow magnitudes). We note that DBScan can in some cases lead to slightly different results depending on the node order [23]. To mitigate this drawback, we propose to use a pre-defined node order (e.g., according to presence count).

Spatial closeness and flow magnitudes: Our algorithm computes spatial aggregates (in this section also referred to as regions for simplicity) using both the density of places and their flow magnitudes. By aggregating only spatially close places that are connected by a high flow magnitudes, we prevent cases where locations would become one aggregate (i.e., region) that are close but have (nearly) no connection to each other. An example for this case would be an airport that usually does not have a significant flow with its surrounding areas. Too small or insufficiently flow-connected spatial aggregates are treated as outliers, thus they are filtered out.

Suitability for time-variant data: As we deal with time-varying flows, the algorithm needs to ensure comparability of spatial aggregates across time steps. If we used spatial simplification on each spatial situation individually [25], it would be difficult to automatically compare the resulting spatial aggregates across time steps, since there is no inherent region correspondence. Therefore, we propose to apply the algorithm to a so-called "supergraph" [20], which combines all time steps. In this way, we derive spatial aggregates in individual time steps. Although the use of a supergraph may hide some small regions appearing in few time steps, it offers a general overview of regions and, most importantly, it provides region correspondence.

Algorithm The algorithm starts with building the supergraph combining all time steps. It then performs an extended DBScan-like algorithm including filtering out of outliers.

The *supergraph* (or *superflow*) $< P, L, T, \overline{W}(P,T), \overline{M}(L,T) >$ is composed of all places and links of all spatial situations in the data set. Its flow magnitudes $\overline{M}(L,T)$ and presence counts $\overline{W}(P,T)$ are averages of all spatial situations.

The algorithm extends DBScan with specific condition of enclosing a place in a region (i.e., spatial aggregate). A place is added to a region if two criteria are met: (1) it is spatially close **and** (2) has sufficiently high flows with the region. We developed and implemented two variants of the aggregation criteria: using absolute and relative flow magnitudes. The absolute criterion filters out nodes that are close to each other but have a non-significant flow between them. The relative flow criterion can be used to prevent nodes from being excluded from a cluster, if those nodes are small but have a large relative flow compared to their size.

(1) A place p_a is considered spatially close to a region R, if its orthodromic distance to the nearest point in the region is lower than a user-defined threshold, i.e., $min(d(p_a, p_R)) \leq th_{dist}$, $p_R \in R$.
(2) A place p_a has a sufficiently high flow, if its flow strength exceeds a user-defined threshold th_{conn}: $mc(p_a, R) \geq th_{conn}$.

We offer two ways of calculating flow strength:

- *Absolute Flow Strength*: Absolute flows are calculated by summing up the magnitudes of all flows $\overline{M}_{a,v}$ and $\overline{M}_{v,a}$ between p_a

and all places with flows to or from this place and the cluster, i.e., its flow-based neighbors in cluster $v \in (n(p_a) \cap R)$.

$$mc_{abs}(p_a, R) = \sum_{v \in (n(p_i) \cap R)} \overline{M}_{i,v} + \sum_{v \in (n(p_a) \cap R)} \overline{M}_{v,i}$$

- *Relative Flow Strength:* The flow strength of relative flows is calculated similar to the absolute flow. However for each link, we relate the magnitude of the link to the presence count in the source of the flow.

$$mc_{rel}(p_i, R) = \sum_{v \in (n(p_i) \cap R)} \frac{\overline{M}_{i,v}}{\overline{W}_i} + \sum_{v \in (n(p_i) \cap R)} \frac{\overline{M}_{v,i}}{\overline{W}_v}$$

During the procedure, we also filter out newly created spatial aggregates which are outliers (i.e., they are either too small or have few connections). All links leading to/from this outlier are also filtered out. We provide the user with three filtering options, depending on size, absolute, and relative flow strength.

- *Size Filter:* It filters out aggregates with presence count $\overline{W}(R)$ smaller than a user-adjustable threshold th_{magn}.
- *Absolute Flow Filter:* This filter uses the total sum of all flows to/from the region. When analyzing movement behavior, one might think of places irrelevant when only few people leave or enter this region.
- *Relative Flow Filter:* Relative flow relates the number of people entering/leaving a region to its presence count. This can be used to keep aggregates that are rather small but have a high fluctuation relative to their size.

Pseudo-code for our spatial simplification method is shown in Algorithm 1 with the following notation: RR is the set of all (resulting) spatial aggregates R, P is the set of all places, $pop(P)$ takes a (random or top-ordered) node from P and removes it from the collection, $n(p_i)$ gets all places that have a flow from or to p_i, and $filter(R)$ is a user-chosen cluster filtering method to eliminate outlier clusters.

Algorithm 1 Algorithm for spatial simplification

```
RR ← ∅
while P ≠ ∅ do
  p1 ← pop(P)
  R ← {pi}
  N ← n(pi)
  while N ≠ ∅ do
    pj ← pop(N)
    if d(pj,R) ≤ thdist and mc(pj,R) ≥ thconn then
      R ← R ∪ p2
      N ← N ∪ {n(pj) \ P}
      P ← P \ {pj}
    end if
  end while
  if ¬filter(R) then
    RR ← RR ∪ RC
  end if
end while
```

The *simplified spatial situations* AS_i in each time step $t_i \in T$ are extracted from the original situations S_i by aggregating their places and flows according to the clusters $R_l \in RR, R_l = \bigcup P_u$ identified by the above-mentioned algorithm. These simplified spatial situations are used as an input to the following temporal simplification (i.e., creation of temporal clusters) described in the next section.

4.1.2 Temporal Simplification

Temporal simplification groups the time steps by the similarity of the corresponding (previously simplified) spatial situations. The clustering can be performed by a clustering algorithm, such as k-means [30]. The main challenge is to determine a suitable similarity function for comparing situations that results in meaningful clustering outputs.

The similarity of spatial situation is commonly determined based on so-called *feature vectors* that uniquely characterize a spatial situation by a vector of numbers (i.e., features). A similarity of spatial situations is then calculated as a similarity of their corresponding feature vectors (e.g., Euclidean vector distance). In this respect, the challenge is to find a suitable feature vector for the spatial situations.

A way of describing a spatial situation for determining similarities is to use a set of flow magnitudes as a feature vector, as each spatial situation is uniquely characterized by its flow magnitudes. However, movements form very dense relationships, which lead to a high number of flows (quadratic to the number of places, e.g., 21,727 links between 606 places). Hence, the dimensionality of this feature vector is extremely high, leading to two kinds of problems: first, the clustering takes much time, and second—more importantly—the clustering results in such cases are known to be unstable and unreliable [12, 37].

For the above-mentioned reason, we perform temporal clustering on the spatially simplified situations. Spatial simplification groups and aggregates places and flows, thus decreasing the situation's size. This reduces the dimensionality of the feature vectors composed of the corresponding flow magnitudes. In our case, the reduction in dimensionality is significant: initial 606 places with 21,727 links are aggregated into 42 clusters with only 394 aggregate links between them. Using the magnitudes of these links as the features describing the spatial situations reduces the dimensionality to 1.8% of the original 21,727 dimensions (i.e., features). This mitigates the instability and scalability problem. Moreover, the clustering results show to be meaningful (see Sect. 5). Note, alternative feature vectors have not proven suitable for our case (see Sect. 6).

4.2 Visual Data Exploration

The results of this spatio-temporal simplification are shown in several interlinked views. They enable a deep exploration of the clustering results as well as investigation of the impact of the results depending on the choice of clustering parameters. The views combine abstract graph visualization with map-based views. Interactive features show details on demand.

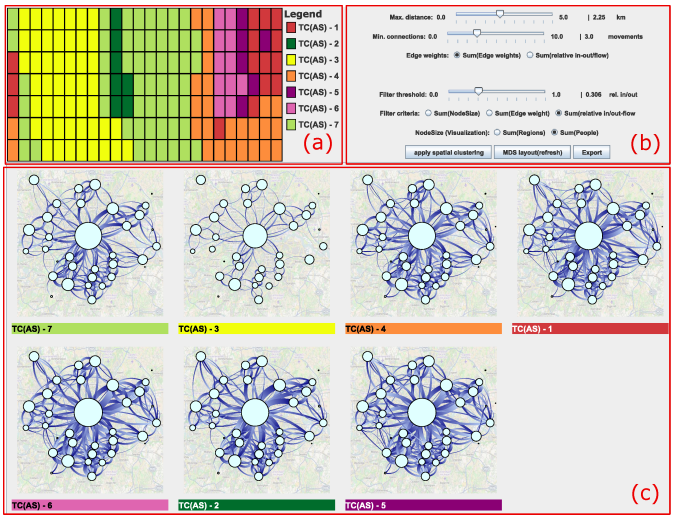


Fig. 3. Cluster result window with several views contained: (a) calendar view of temporal clusters, (b) interactive parameter setting, and (c) cluster overview.

Overview of Spatio-Temporal Clusters The user starts the data analysis in the cluster result window (see Fig. 3). It enables interactive adjustment of clustering parameters (b) and shows the result of the spatio-temporal clustering: both the time distribution of the clusters (a) and the overview of clusters (c).

Parameter adjustment: The user can employ the interactive features for steering the spatio-temporal simplification. By varying the parameters, she can examine the impact of the parameters on the resulting simplified situations. Parameter variation also allows the user to vary the degree of abstraction and schematization in representing the movement and thereby generalize or refine the knowledge gained. In addition, the tool supports finding suitable parameters. We calculate several quality values (e.g., within region flows [27] and average distance [30]) for regularly sampled combinations of connectivity and distance thresholds. The quality values are shown in a heatmap with the two thresholds as rows and columns (see Sect. 5 and Fig. 7). The user chooses suitable parameters and fine tunes them in the main window.

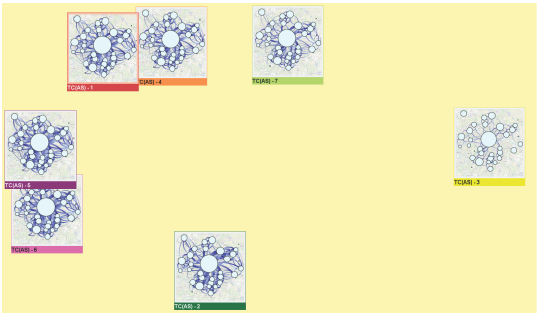


Fig. 4. The MDS layout shows similarity of spatio-temporal clusters by positional closeness.

Calendar view of temporal clusters: This view shows in which time steps the temporal clusters occur (see Fig. 3(a)). Here, colored cells represent clusters. In our case, the time steps are wrapped into days (rows) and hours (columns) similar to the calendar view of van Wijk et al. [46]. It allows to emphasize the weekly pattern of the temporal clusters.

Cluster overview: This view shows small multiples of the spatially simplified temporal clusters, called *cluster thumbnails* (see Fig. 3(c)).

Cluster Thumbnails A thumbnail is color-coded according to the temporal clusters and shows the average graph of a spatio-temporal simplification. Here, the node position is geographically determined. It is shown in the center of the included geographic regions. The underlying map supports a geographic understanding of the flow graph. In order to enhance the contrast between the map and the graph, the map saturation can be optionally reduced (e.g., to light grayscale). Node size is determined either by the average number of people in the cluster's regions or by the number of aggregated regions, as chosen by the user. The edge width is determined by the average aggregate flow between the two regions. After experimenting with various edge direction designs proposed by [32], we decided to use both color gradient (from dark to light) and the edge position to encode direction (edges to the right of the node-to-node line are outgoing). This design reduces clutter caused by edge arrows and allows to encode edge weights, which is difficult to achieve with tapered edges, for instance.

The position of cluster thumbnails on the screen can be twofold. First, the so-called sequential positioning lays the thumbnails next to each other in one or several rows (see Fig. 3(c)). The thumbnails are ordered according to their occurrence in temporal clusters. Second, the so-called distance-based positioning uses multidimensional scaling (MDS) [19] to position the thumbnails according to their similarity (see Fig. 4). The similarity function is user-chosen from a set of graph edit distance functions with varying composition (e.g., edge weights, edge existence, node existence, node weight). We chose to employ the two screen layouts as they both have complementary advantages: sequential positioning is overlap-free but disregards time cluster similarity. In contrast, the distance-based positioning also shows how the time clusters relate to each other.

Visual comparison of cluster thumbnails offers the user the possibility to examine commonalities and differences between spatial situations. The user can gain a broad understanding of the differences between those. However, the exact differences need to be examined in more detail. In this case, the user can click on two thumbnails for seeing their detailed comparison in a so-called *difference view*.

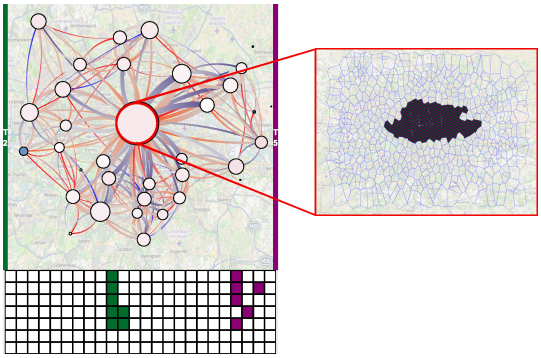


Fig. 5. The difference graph view displays differences between two spatial situations. Blue indicates decrease, white no change, and red an increase. A user-selected spatial aggregate is shown on a map.

Difference View The user can examine differences between user-selected situations in a detailed view (see Fig. 5). The compared situations are indicated by coloring of the left and right border. The time steps of the two situations are highlighted in the calendar view on the bottom. It enables the user to examine when the two compared situations occur. For example, whether both are in the same week days and similar hours, or one is in the morning and the other in the evening.

The graph representation shows the difference of the two spatial situations: both the spatial distribution of persons and flow magnitudes. Node size is determined by the number of people in the aggregate (i.e., region) in the left situation. The change of the relative size of the region is color-coded (from blue via white to red). We use relative changes (percentage change with respect to the left aggregate) as it results in more reasonable difference indication. The alternative absolute change overemphasizes large aggregates. In contrast, relative change takes into account the aggregate's size. If an aggregate (dis)appears between the two situations, it is colored yellow for appearing and black for disappearing. Edges are color-coded according to the relative differences in the flow magnitudes between the two situations. It uses the same color scheme as the nodes.

The node position is determined by its geographic location (center of the included geographic regions), and a change can be examined in detail in the linked map view (see Fig. 5 on the right). The map highlights the regions included in each of the two compared situation (by their original color) and in both situations (using color weaving). The user can thus explore commonalities and differences in spatial clusters in the context of the original geographic positions.

5 Use Cases

To test the feasibility of the devised methodology, we apply it to two real-world data sets: resident Twitter user mobility in Greater London area (Sect. 5.1) and mobile call user mobility patterns in Abidjan area (Sect. 5.2). Due to space limitations, we describe the first use case in more detail and concentrate on the main results in the second use case. For both cases, description of data source and data pre-processing as well as high resolution pictures are provided in the annex and as supplementary material. All documents are also provided at website www.gris.tu-darmstadt.de/research/vissearch/projects/MobilityGraphs.

5.1 Mobility in Greater London Area

We wish to investigate the regular weekly mobility patterns of London residents based on movement data reconstructed from geographically referenced Twitter messages posted on the territory of Greater London during a period of one year. The resident data set consists of 15,246,565 geo-referenced tweets posted by 40,246 distinct Twitter users during the time period from Nov 5, 2012 till Oct 24, 2013. We

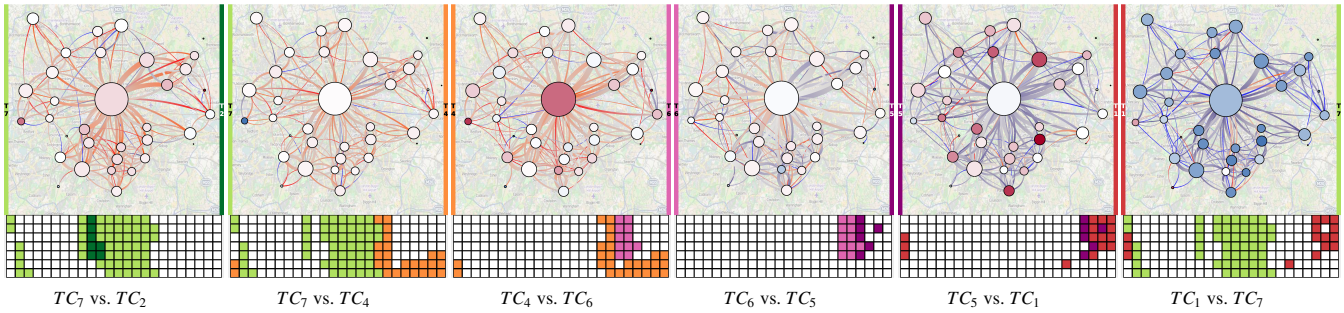


Fig. 6. Pairwise comparisons of simplified temporal clusters $TC(AS)_i$ (or simply TC_i) of mobility in the Greater London area.

obtain mass movement data from records of individuals by spatially aggregating the individuals' movement data into cells by an irregular grid (see Fig. 5 on the right). The data is also temporally aggregated into hourly intervals within the weekly time cycle (7 days $\times$ 24 hours). Thus, we receive time series of transitions between pairs of (not necessarily neighbored) cells. As this type of data is event-based—the location is stored when a posting event occurs—the resulting movement data is episodic [5], and the path between two tweets may not be determined. We have 21,727 time-varying links between the 606 places in 168 time steps.

Spatial and Temporal Simplification We first perform spatial simplification of the movement graph and analyze the quality of results for various parameter settings (see Fig. 7). We analyze a combination of three criteria as they offer a comprehensive assessment: total flows within regions (should be high), number of clusters (should be intermediate) and average distance between regions (should be high). We see that with the increasing connectivity and distance threshold, the results first improve and then get worse. This is natural as a too low thresholds lead to too small regions with low connectivity (red setting). Vice versa, large values lead to large regions disregarding internal structures (yellow setting). So, we choose a balanced setting (orange) with the following parameters: max. distance $th_{dist} = 2.25$ km, min. nr. connections $th_{conn} = 3$, and relative change in people presence threshold $= 0.3$. As a result, 302 out of the 606 places (about 50%) are united in 42 clusters, and the remaining 304 places treated as "noise".

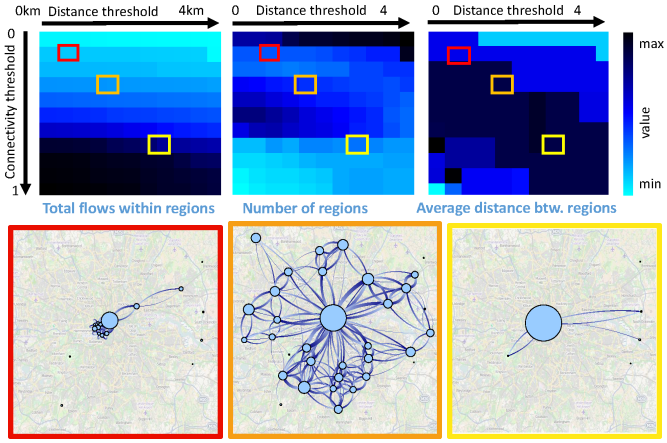


Fig. 7. Top: quality values for various combinations of connectivity and distance thresholds. Bottom: results for selected settings (see colored rectangles). Red: $th_{conn} = 1$, $th_{dist} = 0.7$, orange: $th_{conn} = 3$, $th_{dist} = 2.25$ and yellow: $th_{conn} = 6$, $th_{dist} = 3.5$. All use relative flow filter of 0.3.

We then perform temporal clustering using k-means applied to the spatially clustered flows. We tried different values of k and finally choose $k = 7$, which gives a well interpretable temporal pattern (see Fig. 1), where night, morning, midday, and evening of the working days are well differentiated. This differentiation corresponds to people's everyday mobility, which includes movements to the work or study places in the morning, followed by business time, and movements back to the home places in the afternoon and evening. The clustering also differentiates between the week days from Monday to Friday and the weekend, which could be expected.

Exploration of the Overall Flow Topology We begin our analysis with the investigation of the overall topology of the collective mobility behavior in London, which is reflected in the structures of the aggregated graphs (see Fig. 1). We see that the city has a prominently mono-centric structure with a large central region. It is characterized by the highest people's activity and highest internal mobility. Spatially, the mobility patterns correspond to the topology of the railway network in London. The central region plays the key role as the area of highest mobility and interconnectedness between places, which also attracts people from the periphery. Many people go to the center for work in the morning (around 9 o'clock) or move through the center for reaching their destinations. In the evening of working days (hours 18-19), working people move back from the center to the periphery, whereas some people move to the center (see Fig. 5). Probable purposes of these movements are social activities or cultural events.

Comparing Spatial Situations at Different Times Figure 1 shows an overview of all spatio-temporal clusters. The yellow cluster $TC(AS)$-3 stands out as it differs significantly from other clusters. The presence of moving people and the movement flows are expectedly very low. This cluster unites the night hours (see calendar view in Fig. 1). The data shows that the quiet night time begins and ends later on the weekend than on the week days. Other clusters differ less, thus we proceed with detailed difference analysis (see Fig. 6).

Time cluster $TC(AS)$-7 (abbr. TC_7 in the following; light green) chronologically both precedes and follows the morning time cluster TC_2 (dark green). Their comparison shows that the morning cluster has notably stronger flows from the periphery to the center. However, the presence of people in the nodes is slightly higher in the morning hours than before and after that. Observing the high center-directed flows in the morning and knowing that many people usually go to work or study at this time, we can conclude that many Twitter users either work/study in the center or go through the center to other places.

Next, we compare the light green cluster TC_7, which covers the day time (from 10-11 to 15-16 o'clock), to the afternoon cluster TC_4 (orange; before 18 o'clock). The presence of people and the center-directed flows mostly decrease, whereas outward movements slightly increase. Notably higher increase of the center-outward flows occurs in the following TC_6 (pink; weekdays 18-19 o'clock). Based on our background knowledge about the typical mobility behaviors of people, we can relate these outward flows to people returning home from their work or study places, evidently, located in the central region. Hours 18-19 are also characterized by more movements between the peripheral regions. Interestingly, we also see an increase of flows from the periphery towards the center. Two explanations are possible: some people may move to the center for leisure and social activities, or some people may go through the center to reach their home places.

In the following hour (TC_5; dark magenta), the flows to and from the center decrease, as well as the presence of moving people in the

nodes; however, there is still much movement between peripheral nodes. Further decrease of flows, including peripheral ones, is observed in the late evening TC_1 (red).

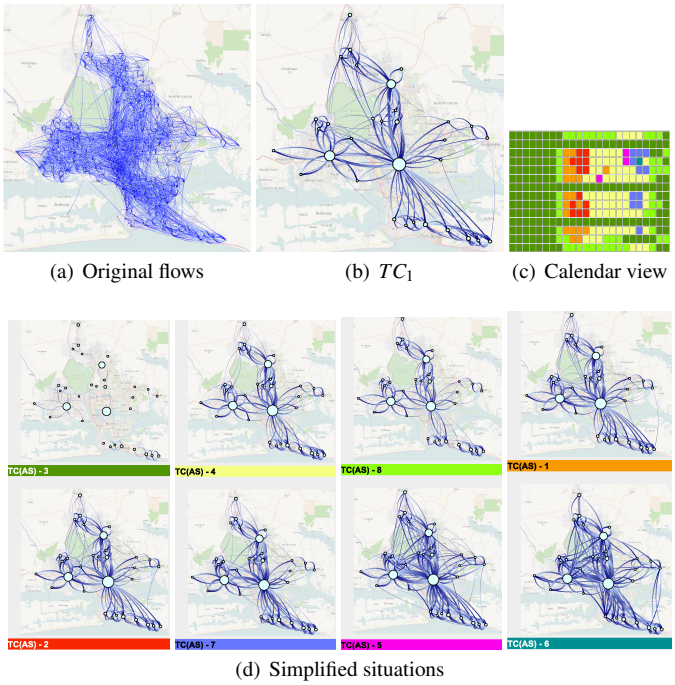
Finally, the comparison of TC_7 with TC_1 shows that people mostly stay or move within the regions represented by the nodes rather then move between the regions. This behavior is characteristic for this time of the day—early night and midday hours, also on the weekend, as well as for early morning hours of the week days.

In sum, the changes of the mass mobility behavior over an average week correspond well to our background knowledge about the general mobility behavior of people. What we learned specifically for London are the typical times of usual movements and their topology. We observed a division between the mostly residential places on the periphery and mostly public places in the center (places of work, study, leisure, etc.). Such a division is not a unique feature of London; there are many other cities where the geographic center is also the center of activities. However, this is not a universal pattern. There are other cities, where the center mostly includes tourist attractions, whereas home and work places of residents are located on the periphery.

5.2 Mobility in Abidjan Area

The second study analyzes anonymized mobile phone use records of 50,000 users in Abidjan area, Ivory Coast over two weeks from April 9 to April 22, 2012. The positions are determined by antennas of the mobile phone network. We use the hourly flows between the antenna areas. The time series has 336 moments (14 days $\times$ 24 hours). Unlike in the London case study, we did not transform the data to the weekly cycle. The mobility graph consists of the 386 nodes corresponding to the antenna areas (cells) and 55,832 links.

A detailed view (see Fig. 8(a)) shows that the mobility topology in Abidjan corresponds to the topology and geography of the city, where several districts are separated by waters and an intrusion of a national park (Parc National du Banco). Like the city itself, the mobility graph has a shape of a butterfly, whose body and four wings correspond to the central part of the city and four peripheral parts to the north, south, east, and west of the center.



(a) Original flows (b) TC_1 (c) Calendar view

(d) Simplified situations

Fig. 8. Mobility in Abidjan area. (a) original flows, (b) temporal cluster TC_1, (c) calendar view on temporal clusters, and (d) overview of situations.

We perform a spatio-temporal simplification with parameters: distance threshold = 2.0 km, connection threshold = 1.0 and filter of 20 people. The result was clustered by k-means with $k = 8$. Figure 8(d) shows the results. They preserve the "butterfly shape", while presenting it in a more schematized way (see Fig. 8(b)). Three large nodes are prominent, one close to the center and the others to the north and west of it. These large nodes are connected by flows with surrounding regions. This allows us to conclude that, unlike London, Abidjan has a polycentric structure, with three major areas of activity. The northern and eastern nodes are strongly linked to the central one. The southern "wing" of the "butterfly" is formed by the shoreline part of the city. The flows within this part are aligned along the shoreline, i.e., correspond to the underlying geography, and there are also strong links to the central part. Similar alignments, but along the major roads, can be seen on the north and on the east of the territory.

Temporal variation During the two weeks, there were 3 days (2, 7, and 11), for which data are missing. The time clustering has put the intervals from these days together with the night time intervals, when the flows were low, into time cluster TC_3—dark green, see Fig. 8(c). The three days with missing data are clearly seen as fully green rows in the calendar view. For the remaining days, a repeating (with small variations) daily pattern can be seen. The first day, April 9, 2012, Monday, differs from the other week days. We looked for possible reasons and found that it was the Easter Monday—a public holiday in Ivory Coast. Its temporal pattern is similar to that of the last day—Sunday (and data is missing for the Sunday of the first week). The time patterns of the Saturdays also differ from those of the week days.

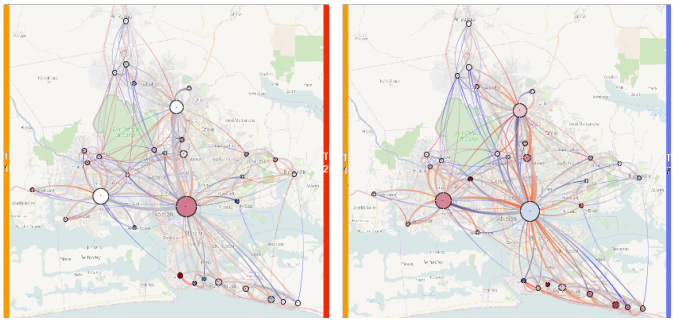


Fig. 9. Comparison of temporal clusters in Abidjan area. Left: Morning hours (TC_1 vs. TC_2). Right: Morning TC_1 vs. evening TC_7.

The temporal clusters show a few outliers (see Fig. 8(c) and 8(d)). Cluster TC_6 (steel blue), which occurred only in hour 19 on day 4, has much higher flows on the whole territory, except for the southern part, than in cluster TC_7, which occurred in the same hours on the other days. Cluster TC_5 (magenta), also with high flows, occurred twice in hour 17 (days 3 and 4) and once in hour 13 (day 6, which was Saturday). Compared to TC_6, the flows to the central region were higher and the flows from the central regions lower. Besides, there is an intrusion of time cluster TC_1, which is specific to mornings, in hour 14 of day 5 (Friday). Unfortunately, we have no background knowledge about the life in Abidjan for trying to explain these temporal outliers.

In the week days, people begin to move in hour 7 (TC_8; light green), and the movement grows in the next hour or two (TC_1; orange). This is obvious from the overview display (see Fig. 8(d)). However, the further changes throughout a working day require pairwise comparison (see Fig. 9).

In late morning hours (TC_2; red), compared to the earlier hours (TC_1; orange), the calling activity of people in the center increases, while the flows to and from the center to the other areas, as well as the flows within the peripheral parts mostly preserve or increase. In the midday and afternoon cluster (TC_4; yellow), the flows decrease throughout the whole territory. In the evening (TC_7; violet), the overall mobility increases, while the outgoing flows from the three centers increase more than the incoming flows. In two week days (days 3 and 4), time cluster TC_5 (magenta) occurs prior to TC_7. TC_5 is characterized by much higher flows than TC_7. In late evening hours (TC_8;

light green), the mobility decreases and gets especially low in the night (TC_3; dark green). Hence, the observed variation of mobility over a working day corresponds to what can be expected based on the general knowledge of the human daily behavior.

The comparison of the morning time cluster (TC_1; orange) with the evening time cluster (TC_7; violet) in Fig. 9 shows the prevalence of incoming flows to the central region in the morning and the prevalence of outgoing flows from the central region in the evening. However, for the other two centers (on the north and west), there are no direction-related differences but only quantitative ones: more movements and within-region activities in the evening than in the morning. Based on this observation, we see that the central region is where the jobs and business activities are mostly located, while the other two centers play a different role, possibly, shopping and/or leisure.

6 Discussion

Our approach uses a combination of spatial and temporal data simplification. In general, there are three variants: spatial simplification preceding temporal simplification and vice versa as well as simultaneous spatio-temporal simplification. The last method is challenging both from the algorithmic and visualization perspective. It is not clear how to visualize spatial situations with individual spatial aggregates appearing in varying time intervals. Therefore, we focus on the two combined variants in this paper. Both variants rely on a suitable description of spatial situations by feature vectors for calculating similarity in temporal clustering. A natural choice is to use a vector of flow magnitudes. For large data sets, this vector is very high dimensional (about $22,000$ dimensions for our data set) and thus leads to unreliable and unstable results. Therefore, we use an initial spatial simplification for data reduction.

We analyzed also alternative methods for data reduction. For example, the use of feature vectors associated with the places such as number of people present in an area. This would lead to 606 dimensions in the London case and 386 dimensions in the Abidjan case.. However, these features do not represent the movements. Thus, they are not suitable for grouping the time steps according to the similarity of the flow patterns. As the flows form spatio-temporal graphs, we hoped to use graph-based features for clustering [49], in particular, the degree, centrality, and clustering measures of the nodes. None of the graph centralities of nodes reflect movement direction. Respectively, they cannot differentiate the flows in the morning and evening hours of the week days. Moreover, the use of edge-related measures does not decrease the dimensionality of the data in comparison to using the original flow magnitudes.

The presented spatial and temporal simplification of flow data employs an extended version of the DBScan algorithm (spatial) combined with k-means (temporal). Our version includes both spatial similarity and flow magnitudes of places as well as includes special features for dealing with time-dependent flows. This leads to consistent clusters across time. As our clustering extends DBScan, it is quite fast (about $1s$ for London data set with 168 time steps, each with up to 606 places and $21,727$ links between places and less then 2 min compared to 15 min for the migration data set used in [27]). Our variant has similar limitations to the original DBScan algorithm, e.g., it cannot find clusters of varying densities and results in "flat clusters" (i.e., no hierarchy is produced). We mitigate the "flat" cluster structure by interactive semantic zoom into clusters of interest (see Figure 10) and interactive setting of clustering parameters for analyzing the data on various scales. The algorithm relies on a set of parameters, which cannot be optimized automatically. We support finding suitable parameters by algorithmic analysis of the resulting clusters by a set of cluster quality criteria (e.g, total flows inside clusters or average distance within clusters).

Our approach employs temporal simplification for handling data sets with hundreds of time steps. For larger data sets—e.g., very long time periods, data with multiple time cycles, or places spread around the globe—additional extensions to our approach would be needed.

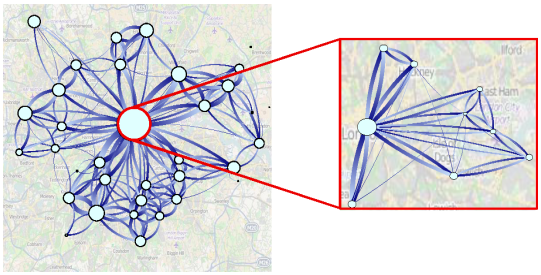


Fig. 10. Interactive change of level of detail for spatial aggregates. An aggregate (red circle) can be semantically zoomed into.

7 Conclusion and Future Work

In this paper, we presented an approach for the visual exploration of time-varying flow data, especially dynamic mass movements. Our approach combines spatial and temporal clustering for dealing with a large number of time steps and a high number of flows between places. Spatial clustering offers consistent clusters across time-steps, which enables the analysis of temporal variations in the data. An important difference to the previous approaches [27, 25] is that the spatial clustering takes time-variation into consideration and thus constructs consistent regions. Our approach supports interactive parameter setting, so that we can observe different whilst consistent spatial patterns and thus gain a more comprehensive understanding of the spatial structure of the mobility. Moreover, our approach enables a general overview of spatial situations across time steps for interrelated flows at all stages of the interactive data exploration, while previous approaches [45] focus on details or on user-selected data parts.

We have tested our approach on dynamic mass movements of residents in the Greater London city area and in Abidjan area. We specially focused on routine behavior. Interestingly, there are obvious similarities between the spatial and temporal mobility patterns in London and Abidjan: adherence to the daily and weekly cycles, centre-directed movements in the morning and outward-directed movements in the evenings, and correspondence of the flow topology to the geography of the underlying territory, particularly, to the transportation networks and natural barriers. It is also notable that these similar patterns have been revealed despite the different kinds of data used: georeferenced tweets for London and mobile phone use records for Abidjan. For finding anomalies in the data, or for other types of data sets, special preprocessing (such as data cleaning) and special visualizations (highlighting anomalies) would be of advantage.

Acknowledgements

The authors are thankful to M. Laux for his support with respect to the software development. This research has been partially supported by the EU-funded projects GRACeFUL and SoBigData.

References

[1] J. Abello, S. Hadlak, H. Schumann, and H.-J. Schulz. A modular degree-of-interest specification for the visual analysis of large dynamic networks. *IEEE Trans. on Visualization and Computer Graphics*, 20(3):337–350, Mar. 2014.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer, 2011.

[3] G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Kisilevich, and S. Wrobel. A conceptual framework and taxonomy of techniques for analyzing movement. *J. of Visual Languages & Computing*, 22(3):213–232, 2011.

[4] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual analytics of movement*. Springer Science & Business Media, 2013.

[5] N. Andrienko, G. Andrienko, H. Stange, T. Liebig, and D. Hecker. Visual analytics for understanding spatial situations from episodic movement data. *KI-Künstliche Intelligenz*, 26(3):241–251, 2012.

[6] D. Archambault, J. Abello, J. Kennedy, S. Kobourov, K.-L. Ma, S. Miksch, C. Muelder, and A. C. Telea. Temporal multivariate net-

works. In *Multivariate Network Visualization*, volume 8380 of *LNCS*, pages 151–174. Springer, 2014.

[7] D. Archambault, H. C. Purchase, and B. Pinaud. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Trans. on Visualization and Computer Graphics*, 17(4):539–552, 2011.

[8] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A review of temporal data visualizations based on space-time cube operations. In *Eurographics Conference on Visualization*, 2014.

[9] B. Bach, E. Pietriga, and J.-D. Fekete. GraphDiaries: Animated transitions and temporal navigation for dynamic networks. *IEEE Trans. on Visualization and Computer Graphics*, 20(5):740–754, 2014.

[10] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. In *Eurographics Conference on Visualization – State of The Art Reports*, 2014.

[11] R. Beecham and J. Wood. Characterising group-cycling journeys using interactive graphics. *Transportation Research Part C: Emerging Technologies*, 47:194–206, 2014.

[12] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is nearest neighbor meaningful? In *Database Theory*, volume 1540 of *LNCS*, pages 217–235. Springer, 1999.

[13] I. Boyandin. *Visualization of Temporal Origin-Destination data*. PhD thesis, University of Fribourg, Switzerland, 2013.

[14] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne. Flowstrates: An approach for visual exploration of temporal origin-destination data. *Computer Graphics Forum*, 30(3):971–980, 2011.

[15] S. Bremm, G. Andrienko, N. Andrienko, T. Schreck, and T. v. Landesberger. Interactive analysis of object group changes over time. In *Int. Workshop on Visual Analytics*, pages 41–44, 2011.

[16] K. Buchin, B. Speckmann, and K. Verbeek. Flow map layout via spiral trees. *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2536–2544, Dec 2011.

[17] M. Burch, B. Schmidt, and D. Weiskopf. A matrix-based visualization for exploring dynamic compound digraphs. In *Information Visualisation*, pages 66–73, July 2013.

[18] M. Burch, C. Vehlow, F. Beck, S. Diehl, and D. Weiskopf. Parallel edge splatting for scalable dynamic graph visualization. *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2344–2353, Dec. 2011.

[19] T. Cox and M. Cox. *Multidimensional Scaling*, volume 1. CRC Press, 2001.

[20] S. Diehl, C. Görg, and A. Kerren. Preserving the mental map using foresighted layout. In *Joint Eurographics – IEEE TCVG Symposium on Visualization*, pages 175–184. Springer, 2001.

[21] T. Dwyer and P. Eades. Visualising a fund manager flow graph with columns and worms. In *Int. Conference on Information Visualisation*, pages 147–152, 2002.

[22] O. Ersoy, C. Hurter, F. Paulovich, G. Cantareiro, and A. Telea. Skeleton-based edge bundling for graph visualization. *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2364–2373, Dec 2011.

[23] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *International Conference on Knowledge Discovery and Data Mining*, pages 226–231, 1996.

[24] Y. Frishman and A. Tal. Dynamic drawing of clustered graphs. In *IEEE Symposium on Information Visualization*, pages 191–198. IEEE, 2004.

[25] S. Gao, Y. Liu, Y. Wang, and X. Ma. Discovering spatial interaction communities from mobile phone data. *T. in GIS*, 17(3):463–481, 2013.

[26] D. Guo. Regionalization with dynamically constrained agglomerative clustering and partitioning (redcap). *International J. of Geographical Information Science*, 22(7):801–823, 2008.

[27] D. Guo. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Trans. on Visualization and Computer Graphics*, 15(6):1041–1048, 2009.

[28] D. Guo and X. Zhu. Origin-destination flow data smoothing and mapping. *IEEE Trans. on Visualization and Computer Graphics*, 20(12):2043–2052, Dec 2014.

[29] S. Hadlak, H. Schumann, and H.-J. Schulz. A survey of multi-faceted graph visualization. In *Eurographics Conference on Visualization – State of The Art Reports*, 2015.

[30] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2nd edition, 2006.

[31] M. Harrower. The cognitive limits of animated maps. *Cartographica*, 42(4):349–357, 2007.

[32] D. Holten, P. Isenberg, J. van Wijk, and J. Fekete. An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *IEEE Pacific Visualization Symposium*, pages 195–202, March 2011.

[33] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, 28(3):983–990, 2009.

[34] C. Hurter, O. Ersoy, S. Fabrikant, T. Klein, and A. Telea. Bundled visualization of dynamic graph and trail data. *IEEE Trans. on Visualization and Computer Graphics*, 20(8):1141–1157, Aug 2014.

[35] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo. *Spatio-temporal clustering*. Springer, 2009.

[36] M.-J. Kraak and F. Ormeling. *Cartography: Visualization of Spatial Data*. Guilford Press, 2011.

[37] H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Trans. on Knowledge Discovery from Data*, 3(1):1–58, 2009.

[38] S. Openshaw. *The Modifiable Area Unit Problem*. Geo Books, 1984.

[39] A. Rae. From spatial interaction data to spatial interaction information? geovisualisation and spatial structures of migration from the 2001 UK census. *Comp., Environment and Urban Sys.*, 33(3):161–178, 2009.

[40] F. Reitz. A framework for an ego-centered and time-aware visualization of relations in arbitrary data repositories. *CoRR*, abs/1009.5183, 2010.

[41] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *IEEE Trans. on Visualization and Computer Graphics*, 14(6):1325–1332, Nov 2008.

[42] S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.

[43] W. R. Tobler. Experiments in migration mapping by computer. *The American Cartographer*, 14(2):155–163, 1987.

[44] B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: Can it facilitate? *Int. J. of Human-Computer Studies*, 57(4):247–262, 2002.

[45] S. van den Elzen and J. van Wijk. Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *IEEE Trans. on Visualization and Computer Graphics*, 20(12):2310–2319, Dec 2014.

[46] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *IEEE Symposium on Information Visualization*, pages 4–9. IEEE, 1999.

[47] C. Vehlow, F. Beck, and D. Weiskopf. The state of the art in visualizing group structures in graphs. In *Eurographics Conference on Visualization – State of The Art Reports*, 2015.

[48] T. von Landesberger, S. Bremm, N. Andrienko, G. Andrienko, and M. Tekusova. Visual analytics methods for categoric spatio-temporal data. In *IEEE Conference on Visual Analytics Science and Technology*, pages 183–192. IEEE, 2012.

[49] T. von Landesberger, M. Görner, and T. Schreck. Visual analysis of graphs with multiple connected components. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 155–162, 2009.

[50] T. Von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. J. van Wijk, J.-D. Fekete, and D. W. Fellner. Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer graphics forum*, 30(6):1719–1749, 2011.

[51] A. Wolff. *Graph drawing and cartography*, page 697736. CRC Press, 2013.

[52] J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with OD maps. *The Cartographic J.*, 47(2):117–129, 2010.

[53] J. Wood, A. Slingsby, and J. Dykes. Visualizing the dynamics of London's bicycle-hire scheme. *Cartographica: The Int. J. for Geographic Information and Geovisualization*, 46(4):239–251, 2011.

[54] L. Xiao, R. Yeh, and P. Hanrahan. Flow map layout. In *IEEE Symposium on Information Visualization*, pages 219–224, Oct 2005.

[55] J. S. Yi, N. Elmqvist, and S. Lee. TimeMatrix: Analyzing temporal social networks using interactive matrix-based visualizations. *Int. J. of Human–Computer Interaction*, 26(11-12):1031–1051, 2010.

[56] X. Zhu and D. Guo. Mapping large spatial flow data with hierarchical clustering. *T. in GIS*, 18(3):421–435, 2014.