EvoFlows: an Interactive Approach for Visualizing Spatial and Temporal Trends in Origin-Destination Data

Erick Cuenca* ICTEAM, UCLouvain
Frédéric Docquier† LISER; IRES, UCLouvain
Siegfried Nijssen‡ ICTEAM, UCLouvain
Pierre Schaus§ ICTEAM, UCLouvain

ABSTRACT

Origin-Destination (OD) datasets describe the movement of entities between different geographic locations over time, such as human migration, movement of animals or diseases, traffic movement, etc. Visualizing the spatio-temporal patterns underlying OD databases is a challenging task since it involves the study of flows between spatial locations (origins and destinations) in a temporal context. Resulting from a collaborative effort of computer scientists and social scientists, our new visualization tool, EvoFlows, combines and synchronizes two complementary views. The first one relies on a system for visualizing time series data, MultiStream, and highlights the temporal dimension of OD data. It depicts inflows and outflows as aggregated stacked time series. Relying on flow maps, the second one highlights the geographic dimension and magnitude of flows at a given time-stamp. Our approach allows the temporal and spatial exploration of the flows at different levels of detail through multiple interaction techniques, visual components, and synchronized animations. The practical usability is illustrated by analyzing data on refugee migration over a period of 59 years.

Index Terms: Visualization tool—Origin-Destination data—User Interfaces—

1 INTRODUCTION

Researchers in social and medical sciences increasingly use multidimensional databases involving spatial, temporal and other aspects. To highlight a few examples, annual data on trade in goods and services are available by product, by country of origin and by country of destination over a long period. Annual data on foreign direct investments are also available per year and per country pair. Demographers, sociologists and economists developed databases documenting international migration flows and stocks by country pair, by education level, by gender, and by period. Epidemiologists use data on population movements to study the propagation of diseases across regions and countries. Visualizing the spatial and temporal patterns underlying these databases is a challenging task.

This paper focuses on three-dimensional data sets involving an origin and a destination (i.e., a dyad of spatial entities), as well as the time component. It proposes a new approach to visualize the trends in dyadic flows between spatial entities (see Andrienko *et al.* [4, 7]). Our tool is designed to rapidly and visually extract the maximum amount of information from large databases to address questions such as: What are the main dyads of countries involved? What are the main sources of variation over time at the extensive margin (i.e., emergence of new dyads) or at the intensive margin (i.e., greatest growth rates)?

Flow maps [23, 26, 32, 37] are widely used to represent dyadic flows. Maps are used to connect origin and destination locations whereas arrows are used to highlight the direction of the flows; the thickness of the arrow represents the magnitude of the flow. In general, this approach is appropriate for scrutinizing spatial patterns, but disregards the time series dimension that is frequently available. Previous approaches [11, 41] use an abstract temporal representation to deal with this limitation; however, they do not show the flow evolution over long periods and raise scalability issues when the number of dyads is large.

In order to fill this gap, we propose *EvoFlows* (Fig. 1), a user-friendly interactive visualization tool that relies on a combination of *MultiStream* [17] and Flow maps for exploring both the temporal and spatial dimensions in origin-destination data. *MultiStream* is an approach that was proposed in earlier work for the visualisation of time series; in this work we will show that the characteristics of this approach also make it useful in the context of visualizing origin-destination data. To illustrate the performance of our approach, we present a real-world application of *EvoFlows* relying on annual flows of refugee migrants. *EvoFlows* is available at `https://erickedu85.github.io/app/evoflows/` and a demonstration video is available at `https://youtu.be/7z89f-5NMvM`.

2 RELATED WORK

This section surveys existing visualization tools for depicting origin-destination flow data. We focus on the spatio-temporal aspects.

2.1 Visualizing Origin-Destination Flows

There are three known approaches for representing origin-destination flow data: Flow Maps [23, 26, 32, 37], OD-Matrices [30, 38], and OD-Maps [39, 40].

Flow Maps is the widely-used technique to visualize OD data. It connects flows by means of an arrow (straight or curved) commonly on a 2D map, where the line width is proportional to the magnitude of the flow and the arrow depicts its direction. Since Flow Maps use a cartographic representation, they naturally reveal the spatial characteristics of the flows. In a dynamic context, scalability issues make comparisons difficult. When the number of dyads is large, the crossing of lines causes visual clutter or occlusion. To partially overcome this problem, several techniques have been proposed: edge filtering [35, 37], edge bundling [10, 19, 28, 32, 33], spatial aggregation (Fig. 2) [1, 23, 24], and 3D Flow Map [42].

Edge filtering is a method where only flows with magnitudes greater than a threshold value are presented. For instance, Stephen *et al.* [35] use filtering interactions to represent the most important U.S. migration flows at the county-to-county level. Another approach to reduce visual clutter is the edge bundling technique. This method groups and merges spatially close line flows. For instance, Phan *et al.* [32] implement a hierarchical clustering method to bundling flows from a single-origin to their destinations in a tree way; this strategy, however, delivers mixed results when the number of dyads is very large. In addition, a major drawback with the edge bundling is that magnitudes and directions of the flows are not exactly represented. Considering spatial aggregation techniques, the regionalization procedure [22] is often used to build regions by agglomerating contiguous spatial locations. For instance, Guo [23] combines this method with a graph-clustering algorithm to group

*e-mail: erick.cuencapauta@uclouvain.be
†e-mail: frederic.docquier@liser.lu
‡e-mail: siegfried.nijssen@uclouvain.be
§e-mail: pierre.schaus@uclouvain.be

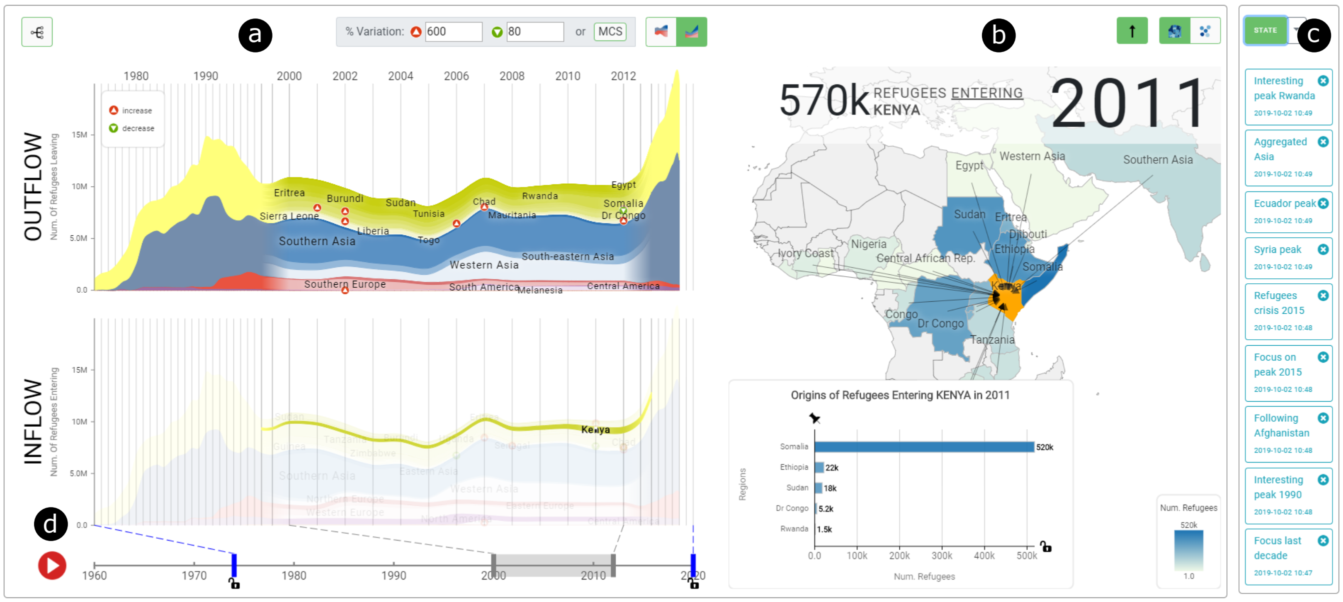

Figure 1: An overview of the *EvoFlows* user interface. (a) The temporal view shows two *MultiStream* [17] views depicting the evolution of outflows/inflows over time. The flows are colored according to the continent where they belong (i.e., Africa in yellow, Asia in blue, Europe in red, America in purple, and Oceania in orange). (b) The spatial view shows origin-destination locations at different levels of details where arrows depict flow directions and the colors of the regions depict the flow magnitude. (c) The snapshot panel displays a list of snapshots that convey interest moments for analysis. (d) A play button allows the user to activate animated transitions to explore temporal and spatial changes in the flows.

different locations into larger regions. However, due to the computational complexity of regionalization, this approach is not suitable for interactive systems. Guo *et al.* [24] propose a further approach that uses a kernel-based density method to extract massive flow patterns. Andrienko *et al.* [1] propose a space-partitioning method based on a Voronoi diagram to define suitable places; however, this method only shows flows between adjacent places and, therefore, loses spatial information of the individuals flows. Recently, Yang *et al.* [42] study the use of a third dimension in an immersive environment to increase readability of flows by mapping the magnitudes in this additional dimension. However, the use of a 3D space to encode abstract information should be used carefully [31].

On the other hand, OD-Matrices use a 2D matrix where the rows and columns represent the origins and destinations of the flows, and the cells contain the magnitude of the flows between locations pairs encoded in a color scheme. Since OD-Matrices use a table to represent flows, they are more scalable than Flow Maps. Nevertheless, identifying the geographical context is difficult [39], as is recognizing the magnitudes of the flows depicted in colors. One technique to partially alleviate those problems is to re-order rows and columns according to spatial locations [30] but again, visual clutters arise when dealing with large datasets.

Another technique to enhance the spatial perception of OD-Matrices is through OD-Maps. This method divides the geographic space into a 2D matrix nested at two levels. The first level represents the origin locations while the second level is embedded in each location as a small matrix that represents the information of the destination flow by color coding. Due to the spatial division that this technique performs, it is difficult to visually represent large datasets of flow data [5].

Hence, the above techniques are mostly designed to characterize the spatial dimension of the OD data flows at a specific moment, but are difficult to combine with the inherent temporal context.

2.2 Visualizing Temporal Changes on Flows

New challenges arise when the temporal evolution of flows is taken into account. Flow Maps and OD-Matrices represent flows of a specific time or an aggregated time interval (i.e., individual time steps are aggregated over longer periods). Usually, a discrete slider allows users to navigate through these periods. Recent approaches, such as MapTrix [41] and Flowstrates [11], add a separate view to depict time series of flow magnitudes. These systems are based on three visual components. A geographical map shows the origins, another shows the destinations, and a heatmap-matrix represent time series of magnitudes. Lines connect the two maps with the corresponding row in the matrix. The main difference between these two approaches is that in Flowstrates each row of the matrix represents a flow, instead, MapTrix gives an aggregation of the flow in each cell of the matrix. Therefore, MapTrix scales better to represent the directions of the flows. Nevertheless, the high level of abstraction provided by the heatmap makes it difficult to visualize precisely the magnitude of the flow at each time step. The analysis of flows over long periods also reduces the readability.

Aigner *et al.* [2] discuss several visualization techniques dealing with time-oriented data. For instance, a widely used approach to represent the evolution of flows over time consists in using stacked graphs [14, 26]. Flows representing the evolution of a quantitative value (e.g., flow magnitude) are stacked over time. However, the visual output quality deteriorates as the number of flows increases.

In this work, we build on an alternative stacking approach, called MultiStream [17], which overcomes this problem by organizing flows into a hierarchical structure that permits a dynamic aggregation of flows and interacting with them in a focus+context technique [16, 20] using a set of interactive views. Therefore, the MultiStream system addresses scalability issues to represent the temporal dimension in large sets of flows.

Our goal in this work is to extract spatio-temporal information of the flows and make it as readable as possible, trying to overcome the limitations of the previous approaches.

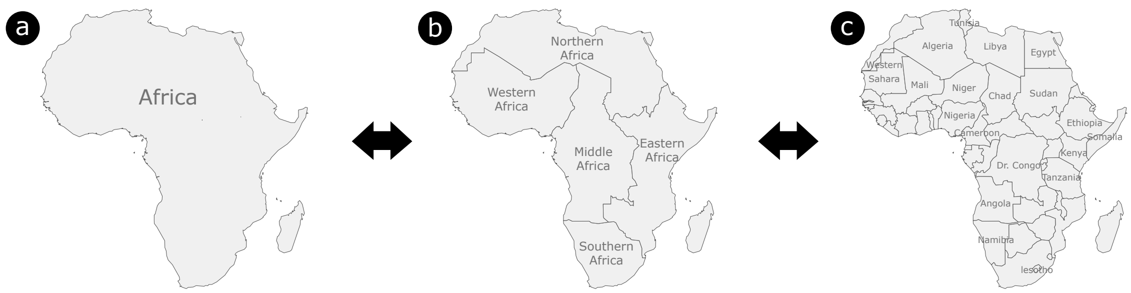

Figure 2: Example of spatial aggregation/disaggregation of the African continent according to administrative units. (a) A high-level of abstraction in a hierarchical structure. Low-level division of the hierarchy representing regions (b) and countries (c). The arrows show the generalization phases of locations.

2.3 Visualizing Flows Generalization

It is necessary to use appropriate data abstraction methods due to the scalability and readability challenges of visualizing spatial and temporal changes in the flow databases. Data abstraction is the process of reducing details of the data while retaining the essential characteristics in a simplified representation of the whole [18]. Aggregation methods are used to group flows and, therefore, reduce the number of them to represent.

Table 1 presents a comparison of different approaches based on Flow Maps considering the temporal and/or spatial aggregations they supports. We can note that spatial generalization is supported by the first three systems. Guo [23], Andrienko *et al.* [1], and Guo *et al.* [24] use techniques such as regionalization to group regions based on flow structures. These approaches provides dedicated geographical maps to interact with the aggregated flows. Regarding the temporal generalization, MapTrix [41] and Flowstrates [11] are the only systems that supplies a strategy to aggregate the time series of the flow magnitudes. Both use a heatmap-matrix representation combined with interactions techniques to navigate through it.

To summarize, most of the systems focus on the spatial aspect rather than the temporal evolution of the OD data. To the best of our knowledge, there are no visual approaches dealing with the spatial-temporal context simultaneously at different levels of aggregation.

Table 1: Comparison of various approaches based on Flow Maps and the ability to aggregate temporal or/and spatial dimensions of origin-destination data flows.

Approach	**Temporal Aggregation**	**Spatial Aggregation**
Guo [23]		x
Andrienko *et al.* [1]		x
Guo *et al.* [24]		x
Flowstrates [11]	x	
MapTrix [41]	x	
EvoFlows	x	x

3 Visualization Requirements

In this section, we describe the requirements to effectively analyze origin-destination flow data and their underlying spatial-temporal context. The following list is the result of discussions with scholars in Economics, Demography and Computer Science.

[R1] ***Visualize and explore temporal variations.*** – The first requirement is to depict the evolution of the magnitude of the flows over time in order to identify time trends, peaks and valleys, and variations across periods. In addition, we must take into account both directions of flow values: inflow (entries) and outflow (exits). For instance, peak periods in outflow could reveal important time intervals to focus the user's attention.

[R2] ***Visualize spatial distributions.*** – The second requirement is to represent precisely the spatial position of each origin and each destination, and to depict the value and the direction of each flow in a specific period. For example, showing the destination flows for a given origin could help to understand the role of neighboring countries, or to highlight more general geographic patterns.

[R3] ***Scalability.*** – The third requirement is to enable the handling of large datasets of origin-destination flows, and to support their analysis over long intervals of time. For instance, a user could be interested in analyzing a dataset about people migration between countries (about 200 x 200 flows) over the last fifty years or so.

[R4] ***Highlighting and captioning.*** – Since the number of flows can be large over a long period, mechanisms are needed to automatically highlight specific moments (e.g., peaks) and to aggregate data that is less relevant. Hence, the fourth requirement is to allow users to emphasize moments in time, to remove unneeded detail, and to take snapshots of certain moments.

[R5] ***Synchronization and animation.*** – The fifth requirement is to provide synchronization between the different visual components (i.e., the actions in a component that are also reflected in the others involved). In addition, it must support animated transitions when an action occurs (mouse hovering, filtering, zooming, etc.) and thus preserve the mental map of the whole visualization [8].

The above list attempts to achieve the common tasks in the visual analysis of flow data. Our intention is to promote the use of *EvoFlows* by the general public, as well as specialized users.

4 Proposed Technique

This section describes our visualization design and functionalities. Our exploration process follows the principle of the *visual information seeking mantra* proposed by Shneiderman: overview first, zoom and filter, then details on demand [34].

4.1 Design Rationale Summary

Based on the requirement analysis, we propose *EvoFlows* (Fig. 1), a new visualization tool that facilitates the analysis of origin-destination data flows. It is composed of several interactive and synchronized components. The *temporal* view (Fig. 1a) is based on *MultiStream* and depicts outflow and inflow evolution over time. It allows the user to define the aggregation/disaggregation level of the flows **[R3]**. It also automatically detects and highlights important periods (e.g., peaks, steady increase in flow magnitudes) **[R1, R4]**. The *spatial* view (Fig. 1b) is base on Flow Maps and shows geographical locations at different levels of detail **[R3]**, transmitting the direction of each flow in a specific period **[R2]**. The *snapshot* panel (Fig. 1c) allows adding and loading snapshots of the current configuration for analysis and communication tasks **[R4]**. All these components are synchronized with each other and support visual animations **[R5]**. They are described in detail below.

4.2 Temporal View

Visualizing many data flows over time as an unordered stacked graph strongly limits the discovery of interesting visual patterns **[R1, R3]**. MultiStream [17] was proposed as an improvement of stacked graphs to allow organizing the data according to a hierarchical structure providing aggregation/disaggregation mechanisms **[R4]**. We propose to reuse the MultiStream approach for visualizing flows. Indeed, origin-destination data flows can naturally be aggregated using high-level administrative divisions (e.g., countries, sub-regions, regions, and continents). We must take into account that inflow (entries) and outflow (exits) magnitudes can evolve differently over time. For instance, the flow of people leaving Afghanistan is probably different from the flow of people entering. To better adapt to this behavior, we use two multistreams. Fig. 3 shows the *temporal view* and the two multistreams that depict the outflow and inflow evolution worldwide in a synchronized way (Fig. 3(a,b)).

4.2.1 Aggregation and Disaggregation of Flows

MultiStream also provides a hierarchy manager that controls the level of spatial aggregation/disaggregation (Fig. 3d), which is itself tightly coupled with a time-controller (Fig. 3c). The hierarchy manager (Fig. 3d) defines two levels. The coarse one is depicted with a blue line; the fine one is depicted with a green line. This component allows the user to perform aggregation/disaggregation of flows interactively. Performing any of these actions, the levels of abstraction are updated in the temporal and spatial view (Sect. 4.3). For instance, thanks to the MultiStream approach, we can visualize the African flows at a sub-regional level (e.g., Eastern, Middle Eastern, etc.), whereas other flows are described at the country level. This functionality offers flexibility in the level of detail that the flows show i.e., flexible temporal aggregation **[R1, R3]**. To avoid overloading the visual space, we link only one hierarchy manager for both multistreams **[R5]**.

4.2.2 Temporal Filter

The time-controller (Fig. 3c) defines the period of time over which the stacked graph relies on the fine aggregation level rather than on the coarse one. This corresponds to the gray region and the two vertical blue lines on the time axis. Based on a brushing&linking technique [9], it allows moving or expanding the time-controller to update both multistreams **[R5]**, thereby offering flexibility to focus on a specific period of interest **[R1]**. This tool facilitates the analysis over long periods of time **[R3]** using several interaction techniques such as zooming [29], or focus+context [16, 20].

4.2.3 Color Coding

Categorical color coding is used to distinguish between different branches in the hierarchy. For instance, Fig. 3(a,b) shows the evolution of refugee outflow/inflow by country where flows are colored according to their continent (e.g., Africa in yellow, Asia in blue, and so forth). This helps to preserve the mental map of flow evolution and to enhance comparison tasks **[R1]**.

4.2.4 Highlight Temporal Changes

It is often difficult to determine small or sudden variations in the magnitudes of the time series of flows. Since the MultiStream method stack layers on top of each other, at the end of this process, the upper layers frequently suffer from readability due to the number of layers under them. One solution to this problem is to use a vertical rule that indicates the exact value of a single flow at each time step [17]. We extend this technique and show the rule for all flows, detailing information on the variation in the magnitude of the flows. Fig. 3e shows contextual information of the main flows including glyphs that transmit changes in the flow (e.g., increase, decrease, or stationary) according to the previous year. We can see that for the year 1999 the flow of refugees from Central America increased, while that of Eastern Europe decreased with respect to the year 1998. In addition, we observe that the Southern Europe flow remains first. This facilitates the tasks of comparison between flows **[R1, R4]**. This ruler is shown when users hover the cursor over the temporal grid of the MultiStreams.

4.3 Spatial View

The purpose of the *spatial view* (Fig. 4) is to interact with the geospatial context of the flows. This view is based on a Flow Map representation to show the geographical locations of the origins and destinations, the flow directions, and the flow values. Standard Flow Maps connect each flow by an arrow where the thickness represents the value. However, visual clutter increases when show large set of flows. The larger number of lines that connect places disturb perception to highlight patterns. Therefore, an aggregation technique is required to alleviate this problem and thus improve scalability capacities **[R3]**.

4.3.1 Spatial Aggregation

We take advantage of the nature of the flows to be organized within a hierarchy [17] and, following this structure, we perform a spatial aggregation of their locations. This strategy allows us to i) interact tightly with the *temporal view* and the *hierarchy manager* to control the level of aggregation **[R1, R5]**, ii) depict spatial locations at different levels of detail **[R2]**, iii) increase scalability by not using a regionalization method whose computational cost is high for a dynamic application **[R3]**, and iv) reduce the number of flows to show and improve exploration tasks **[R1, R2, R3, R4]**.

The spatial aggregation process is carried out in a straightforward way by grouping the flows according to their position in the hierarchy. Therefore, the values of the flows in the upper levels are equal to the sum of their children (Fig. 2). This process is done for all elements in the structure. Interactions techniques are available in the *hierarchy manager* to perform aggregation/disaggregation actions. It is this component that manages the levels of abstraction in which the flows are depicted in the temporal and spatial view.

Fig. 5 shows the outflow of refugees from Africa in 2016 at three different levels of abstraction. In Fig. 5a, outflows are aggregated at the continent level (i.e., America, Europe, Asia, and Oceania). A portion of the hierarchical structure is shown on the left, where the green line transmits the flows shown on the Flow-Map. To divide the flows of this continental level, a tooltip is providing in the hierarchy manager when users hover the mouse over a node (on the left of Fig. 5a). Fig. 5b shows a Flop-Map after disaggregation. Notice how the continents where split into regions following the process described above. Green line in the hierarchy manager now crosses flows at region levels. Finally, Fig. 5c shows outflows at a country level and their respective levels in the hierarchy on the left.

Thanks to this flexibility, the user can compare spatial flows at all levels of the hierarchical structure (i.e., to all possible variables between countries, sub-regions, regions, and continents), and thus reveal local or external patterns.

4.3.2 Depict Flows

To convey the magnitudes and directions of the flows, we propose two representations: a point-map and a choropleth-map [6]. Considering the point-map, circles of equal area are positioned at the centroid of the region/aggregated-region. A continuous two-hue color palette is used to colored the circles in relation to the magnitude. The sequential color scheme varies between light green for low values and dark blue for high values. We prefer this representation to that of the bubble-map, where the area of the circle is proportional to the magnitude, to avoid overlapping circles. Fig. 4 shows

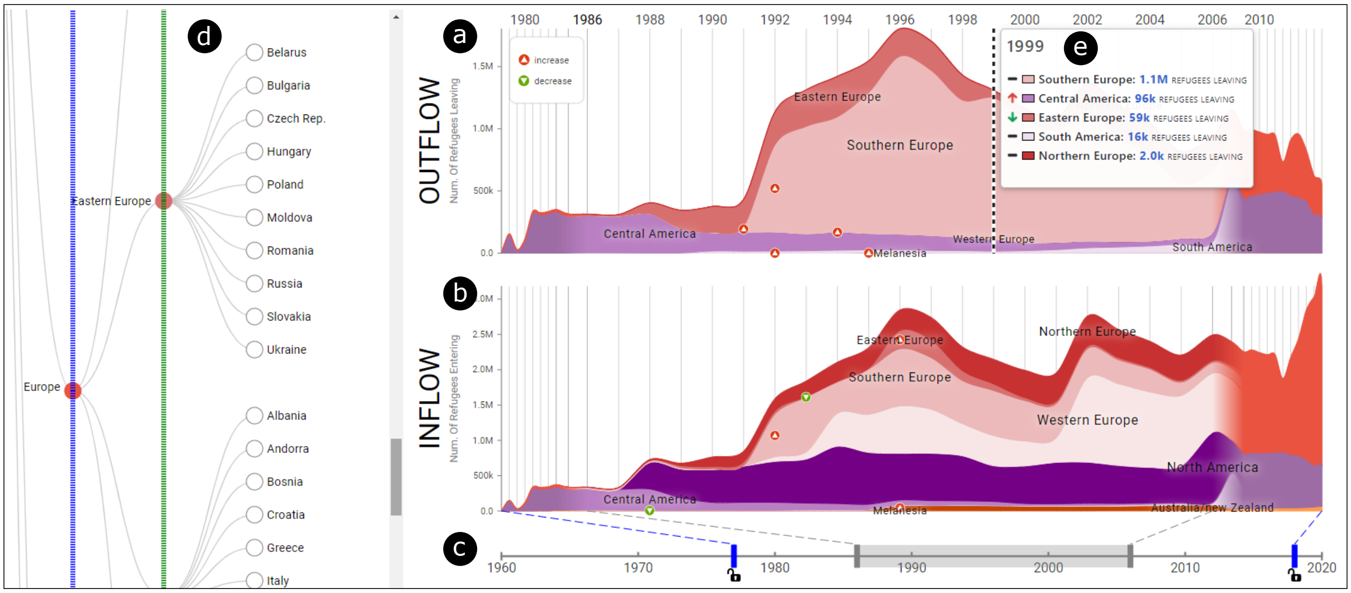

Figure 3: The temporal view. (a)(b) Two multistreams [17] depict the (a) outflows and (b) inflows evolution over time in a hierarchical organization. (c) The time-controller defines the periods of time in which the multistreams show a certain hierarchy level. (d) The hierarchy manager allows navigating through the hierarchy in both multistreams and the spatial view. (e)

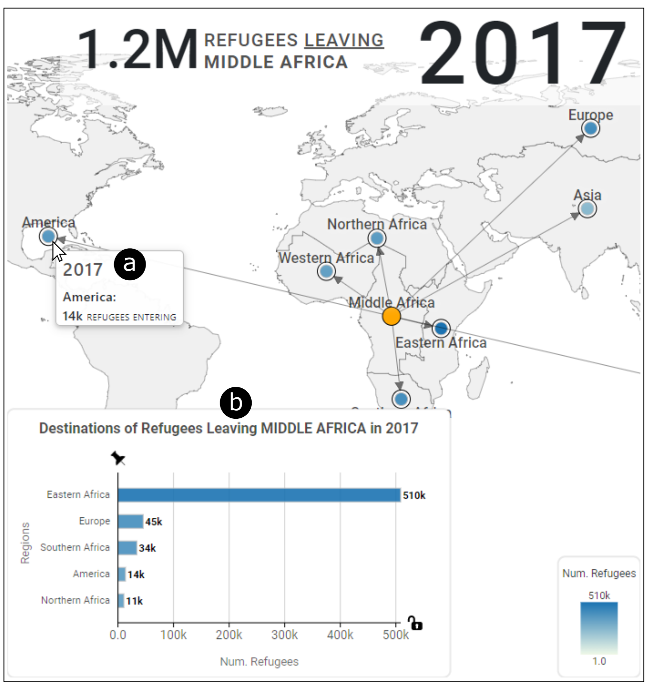

Figure 4: The spatial view showing the refugee flows leaving Middle-Africa in 2017 using a point map representation. A two-hue color scheme colors the circles in relation to the magnitude. (a) A tooltip box shows textual information of a selected flow. (b) A bar chart shows the top-k of flow magnitudes improving the recognition and comparison of values.

a point map illustrating the refugee flows leaving Middle-Africa in 2017. We can note that the flows are aggregated at the continent level (e.g., America, Europe, and so forth.), except in the African continent where the flows are aggregated at the region level (e.g., Northern, Western, etc.). The origin location is shown in orange and for each destination, flow values are depicted by a circle whose color is proportional to its value. In addition, arrows can optionally be added to represent flow directions and to strengthen the visual link between origin-destination pairs. To avoid lines crossing, we show flows for one origin or destination at a time.

The expressiveness of a visualization is improved by adding contextual information (e.g., labeling). In a Flow map, due to the constraints related to the number and predefined position of locations, it is difficult to set a label for all geographic regions. Therefore, labeling is not a minor tasks, since the establishment of labels for all locations can lead to cluttering and overlapping issues. Our approach uses a brute-force labeling algorithm [14] to find the best place to set the region name and avoid overlap at each zooming and dragging interaction **[R5]**. Furthermore, we show the flow arrow outside the circle and the label in the opposite direction. This careful design improves visual perception avoiding the overlapping between circle, label, and arrow of the same flow **[R3]**. For instance, look at the arrangement of arrows and labels in Fig. 4. The labels of the regions north of Middle Africa are positioned above the circles (e.g., Western and Northern Africa, Asia, Europe, etc.); while the regions that are to the south, are shown below the circles (e.g., Eastern Africa).

In Fig. 4, we identify a high density of refugees in neighboring countries (e.g., Eastern Africa). Depending on the size of the circles, some magnitudes are difficult to analyze. Therefore, the values can also be represented into the geographical areas using a choropleth map. In this representation, the same palette color of the point map is used. For instance, Fig. 1a shows a choropleth representing the entering refugee flows into Kenya in 2007. As can be observed, the main countries of origin are the close neighbors. However, even if the choropleth map reveals geographic patterns, deducing relative differences between values can be difficult. In addition, larger regions tend to have a greater weight on the visual map. To overcome these issues, a bar chart representation (Fig. 4b) shows the top-k flow magnitudes for the selected period.

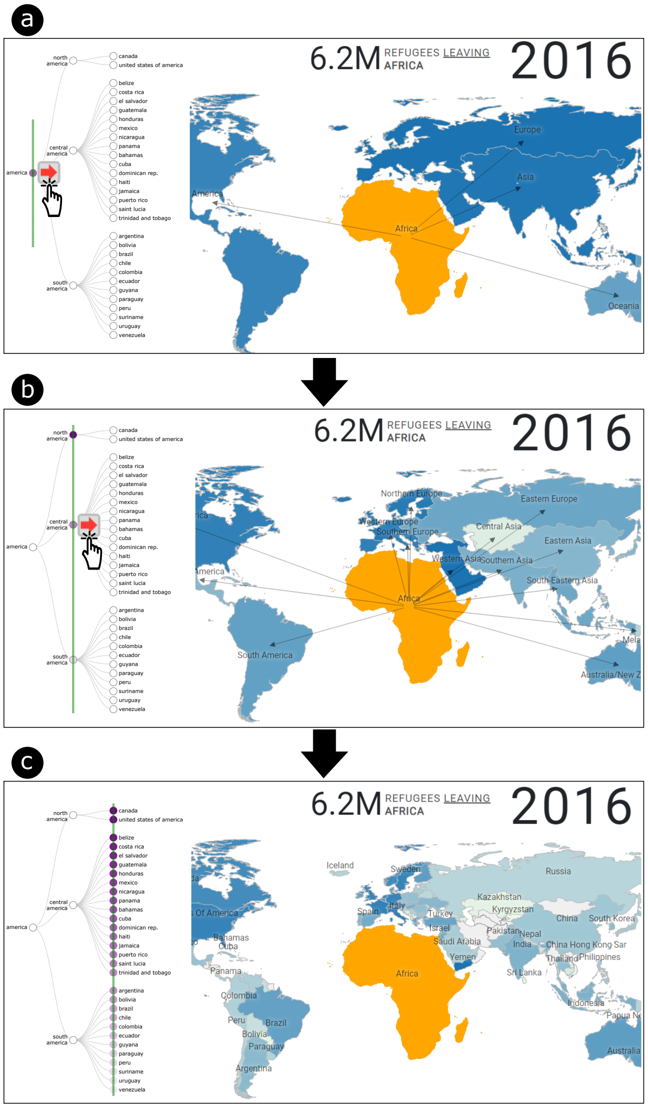

Figure 5: Spatial disaggregation of refugees leaving from Africa. (a) Outflows are shown in a high level of detail. (b) Splitting high level outflows into their children according to the hierarchy. (c) Outflows are shown at a detail level according to the hierarchy. Arrows are hidden to avoid clutter.

4.3.3 Highlight Flow Magnitudes

The purpose of the point-map and the choropleth-map is to reveal spatial patterns at a glance. However, the color abstraction used in these maps does not accurately represent the value of the flow, because the coding of quantitative values is not accurate in a color scheme. To overcome this issue, a *tooltip box* convey textual information (i.e., date, region name, and the flow magnitude). It is shown by demand while the mouse hovers a location (Fig. 4a); however, this strategy only shows the magnitude of a single flow. A bar chart representation is used to improve readability and reveal accurate values. Bar charts use the position and length which are better visual channels for exploring and comparison tasks [15] **[R2, R4]**. Fig. 4b shows the flow magnitudes in descending order in a horizontal bar chart. Thanks to this chart we can see clearly flow magnitudes and compare between them, which is not possible with a standard Flow Map.

For instance, in Fig. 4, it is difficult to discern the exact values of refugee flows to Europe and America, since the colors are close. However, using the bar chart (Fig. 4b) we can see that Europe has 45K refugees, which is more than triple the number of refugees in America with 14K. In addition, the descending order of the bars allows comparison between them. We can see that Northern Africa (11K) and America (14K) receive almost the same number of refugees from Middle Africa in 2017.

4.4 Snapshot Panel

Exploring the data is an interactive task and discovering parameters in different views to highlight patterns can take some time. In a context of analysis, users may find interesting moments that get their attention (e.g., peaks), they probably want to save those moments for further analysis in the future. The snapshot panel (Fig. 1c) provides mechanisms to freeze and record the parameters of *EvoFlows*, and thus create static narrative moments. They are listed in a reverse-chronological order of entry, the newest above. The user can then navigate easily in the snapshot list to restore those parameters in just one click in order to create a storytelling of the data **[R4]** [13]. Fig. 1c shows this panel displaying a list of snapshots. Standard features are provided, such as loading, saving and deleting.

5 Interactions

In this sections we describe the different interactions between the visual components proposed in *EvoFlows*. Interaction techniques aim to help users in exploration and analysis tasks.

5.1 Visualizing Spatio-Temporal Changes

The temporal and the spatial views are synchronized with the joint objectives of navigating along the flows at different levels of detail **[R1, R2, R3]** and conserving the mental map during the exploration process **[R5]**. Moving the mouse over a flow in the temporal view generates the following updates: i) the ruler is positioned over the flow at a desired time-step, ii) the color of the selected flow is fully saturated while the opacity decreases in the others, and iii) all the component in the spatial view are updated according to the selected time-step (i.e., the flow-map and the bar chart depicting the magnitudes). By clicking on a flow in a desired time step, the flow remains fixed on that date, allowing the user to interact/navigate on the map. The possible actions are: toggle between a point and choropleth map, hide/show the arrows of the flows to alleviate cluttering, and perform a zooming in/out which updates the labels of locations taking care that they do not overlap. Clicking again on the flow in the temporal view deactivates the selection. Hence, the user can freely navigate back and forth in time and explore spatial patterns on the map in a coordinated manner.

Fig. 1a shows the Kenya inflow highlighted in 2011. As it is an inflow, the map view (Fig. 1b) depicts the entering flows to Kenya in that year. We can notice a strong local pattern, i.e., a high number of refugees entering from neighboring countries (e.g., Somalia, Ethiopia, Dr. Congo, and Sudan). This technique is adapted to analyze a precise moment of the time series of the flows. However, it has some limits, such as the loss of temporal context on the map, or the difficulty of selecting with the cursor a small flow in the temporal view. In order to overcome these issues, some approaches propose the use of animated transitions between visual components that preserve the context [13, 21, 27, 36]. We consider this technique to assist users link the temporal and the spatial view and, therefore, reduce cognitive load.

5.2 Animated Transitions

Heer *et al.* [27] demonstrate the effectiveness of animated transition to preserve the data context between several graphics such as bar charts, pie charts, and scatter plots [27]. The study of Griffing *et* *al.* [21] reveals that animations are more effective to convey and discover geographical patterns than static small-multiples. A qualitative study of Boyadin *et al.* [12] observe that animation helps to recognize geographically local events and changes between subsequent years. Recently, Brehmer *et al.* [13] use this technique in a design of a space for narrative visualizations in a context of storytelling. In our work, we consider the use of animated transitions to enhance visual perception of temporal changes on the map and preserving the context.

To ensure the effectiveness of animations in a visual component, there are three principal points to consider [13]: i) the initial states of the visual elements in the component, ii) the animation considering those elements as a whole or separately, iii) the trajectory of the animation, and iv) the control of the transition speed. The process we follow to address these considerations are listed below:

1. Considering the first point, the initial state of the animation is when the user, with the cursor, selects a time-step by clicking on it. This action fixes the temporal and spatial view for that selected time-step; e.g., the inflow of Kenya is fixed in 2011 (Fig. 1a) as well as the spatial components that show flow the positions and magnitudes for that year (Fig. 1b). This state is the same as described above, when the user freely explore time series of flows. Once the flow is fixed at the desired time-step, a play button appears at the bottom-left of the temporal view (Fig. 1d). This button allows the user to continue with the animation transition process.

2. Regarding the second point, we perform a synchronized transition, i.e., all visual elements begin transitions at the same time. Therefore, when the user clicks on the play button, the rule above the selected flow begins to move to the next time-step, and the flow-map and the bar chart are also updated. This is a loop process performed for each time-step over the period selected by the time-controller (Fig. 3).

3. The third consideration is about the trajectory of the animation. In the temporal view, animation moves the ruler in a linear transition. In the spatial view, sudden changes in the magnitudes and positions of the flows increase the cognitive load. To alleviate this problem, the flow arrows are animated from their origin to their destination location. The bar chart can improve the animation transition in two ways, by setting the maximum value on the horizontal axis and filtering the elements to show on the vertical axis. Fig. 6a shows the horizontal axis locked at its maximum value according to the period selected in the time-controller. On the other hand, the vertical axis can filter a subset of flows to represent. Fig. 6b shows this functionality that allows users to select the flows.

4. Concerning the last point. We implement a set of buttons to control the speed of the animation when it is executed. In addition, it allows the user to start, pause, and resume the animation transition process.

Our animation transition process aims to explore the spatio-temporal context of the flows through the visual components of *EvoFlows* **[R1, R2, R3, R5]**. This technique improves the analysis, especially when flows changes suddenly, with a fixed layout of the bar chart through this process.

5.3 Exploring Events

To assist users, we automatically highlight important changes in flow trend. To graphically represent such events, a visual mark is placed in relation to the horizontal axis of MultiStreams.

We propose two methods to get insight of time series of flows. The first with the objective of showing sudden changes in flows (e.g., when a peak or pit occurs) **[R4]**. To achieve this, we calculate the percentage of variation between consecutive time steps and show the variations that exceeds a given threshold. For instance, the outflow in Fig. 3a shows several icons over time and highlights periods at which the number of refugees increases significantly in percentage compared to the previous year. Observe the flow of Southern Europe in 1992, a red icon represents an increases in 500%. The second method shows a steady increase in the magnitude of the flow over a period. Such events may attract the user's attention for a further analysis. We are looking for an algorithm that does not decrease the performance of *EvoFlows* **[R5]** (ideally in O(n)). Therefore, we use the Maximum Sum Contiguous Subarray method. This algorithm looks for the magnitudes of time series in a time window and maintains the maximum sum of contiguous segments.

These methods facilitate analysis when flows are observed over long periods **[R1, R3, R4]**. The top of the temporal view (Fig. 1a) provides input mechanisms for these methods.

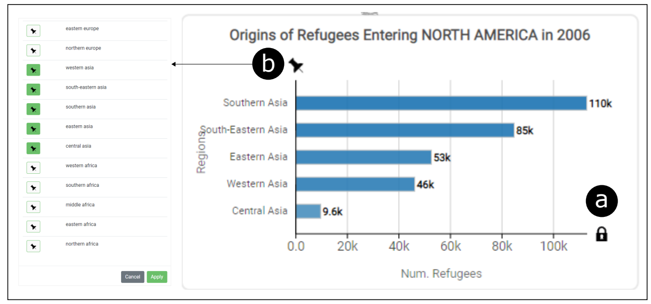

Figure 6: A bar chart allows to fix the layout. (a) Horizontal axis can be remains locked a the maximum value. (b) Vertical allows to pick and filtering a subset of flows.

6 Discussion

In this section, we discuss the facilities of *EvoFlows* to analyze and efficiently explore the spatio-temporal dimensions of time series flows. We compare our approach with other approaches that support temporal and spatial aggregations.

6.1 Visualizing Temporal Aggregations

Exploring large time series of flows is challenge due to some issues such as scalability, limit screen size, etc. Several previous approaches use a dedicated temporal view to depict flow evolutions over time. For instance, Flowstrates [11] and MapTrix [41] use a heatmap representation to represent flow evolutions. However, the main issue of these techniques is the scalability. The heatmap do not depict precisely the flow magnitude; therefore, it does not support multiple time series of flows. In addition, flow databases are often long, that is, they cover too much historical data (e.g., migration record 50 years ago). A common issue in the visualization of this data is the limited screen space.

In this research, we rely on the MultiStream approach [17] to overcome these difficulties. MultiStream organizes time series of flows in a hierarchical structure to provide aggregation and disaggregation throughout time series and, therefore, depict them at different levels of abstraction. In order to explore and analysis long time series, the MultiStream uses several interaction techniques such as focus+context, fisheye, and zoom to focus on a time segment for further analysis.

6.2 Visualizing Spatial Aggregations

Grouping spatial locations reduces visual clutter in a Flow-Map representation. However, this involves dealing with some drawbacks, such as: depicting the directions of the flows, magnitudes and locations. To overcome these problems, spatial aggregations are used. These techniques groupe the flows between origins and destinations with the objective of representing fewer flows.

Aggregate spatial local, reduce the visual clutter in a visualization. Dealing with Flow-Map entails some drawbacks such as: depicting flow directions, magnitudes and locations. In order to overcome these issues, spatial aggregations is used. This technique group flows between origins and destinations in order to depict less flows.

Previous approaches [1,23,24] use the flow structures to bundling regions using, as for example the regionalization technique. However, they do not accurately show all the characteristics of the flows (directions and magnitudes). In addition, the execution time of this techniques is not adapted for dynamic visualizations.

EvoFlows is based on the nature of the series of flows that will be organized in a hierarchical structure to perform a spatial aggregation following this structure. This technique allows closely interact with the temporal and context of flows. In addition, due to the use of a hierarchy, flows can be aggregated in different levels of detail; thus, a Flow-Map can represent flows locations at different levels. This feature is not provided by any other previous approach.

7 Application Example

To illustrate the concepts and the performance of *EvoFlows*, we focus here on dyadic flows of refugee migrants over the last 59 years. Refugee migration has become a topical issue in many industrialized countries. The number of asylum applications lodged in 2015 in EU Member States exceeded 1.3 million. This placed migration policy in the forefront of the global policy debate, as is often the case after each immigration peak. We show here how our tool can be used to put the recent refugee crisis into perspective and highlight the geographic patterns of short-term and long-term movements.

7.1 Data on Refugee Migration

For our application, we use the *UNHCR* database on global refugee movements.[1] This data set documents the annual flows of refugee migrants from 195 origin to 195 destination countries over the period 1960-2018. Each entry contains the following details: the year of observation, the country of origin, the country of destination, and the recorded number of refugee migrants.

7.2 Visual Analysis

Fig. 7 focuses on the temporal dimension and depicts the long-running trends in worldwide outflows and inflows of refugees. Countries are grouped by region or by continent. This kind of organization forms a hierarchical structure, where the first level is composed of 5 continents (Africa, Asia, Europe, America, and Oceania), the second level is composed of regions (e.g., Northern Africa, Central Asia, Eastern Europe, etc.), and the third level is composed of 195 countries. Categorical color coding is used to represent the main 5 continents (Africa in yellow, Asia in blue, Europe in red, America in purple, and Oceania in orange). The thickness of a layer features the magnitude of the annual outflow or inflow of refugees.

The top panel of Fig. 7 depicts the evolution of annual outflows at the aggregate regional level (e.g., Eastern Africa, Middle Africa, etc.) except for the Asian continent (in blue) for which the country level disaggregation is used (e.g., Afghanistan, Iran, Pakistan, Syria, etc.). The figure shows that the worldwide stock of refugees started to increase in the early seventies and reached a peak in the early nineties. Outflows from Asia, from Afghanistan in particular, and from sub-Saharan Africa are governing the trends. After 1994, this number began to decrease until the year 2000, and remained stable until 2010. The recent refugee crisis appears after 2014. Until 2010, the exodus from Afghanistan is the most remarkable fact, but in

[1]http://popstats.unhcr.org/en/persons$_{o}f_{c}oncern$

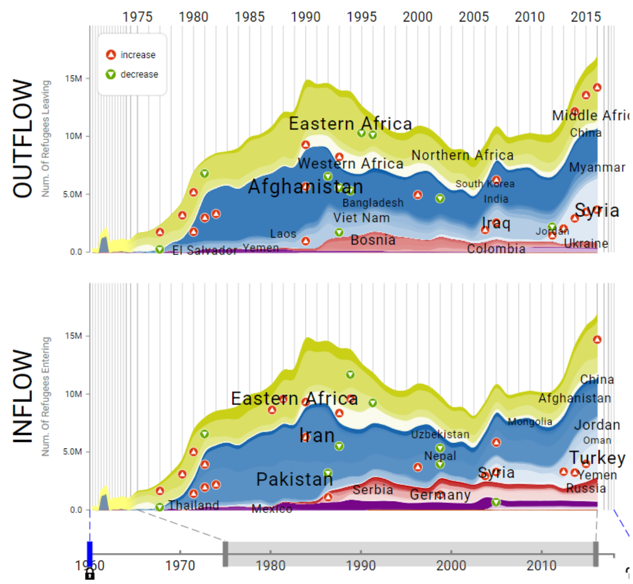

Figure 7: Temporal evolution of annual refugee stocks. Focus on the 1975 - 2015 period.

2012 the number of refugees leaving Syria makes the flow of this country surpasses that of Afghanistan. The red icons are helpful to spot migration surges. For example, a 500k increase in the number of refugees from Syria in 2012 compared to the previous year, the same phenomenon occurs until 2018.

As each outflow generates an inflow of an identical size, the bottom panel highlights the main destinations of refugee migrants. We find that Iran, Pakistan and (to a lesser extent) India hosted most Asian refugees between the mid-seventies and the mid-nineties. The bottom panel also demonstrates that North America and Europe hosted a small proportion of the total stock of refugees. This observation holds true in the recent years despite the fact that the Syrian exodus has been perceived as a massive refugee crisis. In sum, Fig. 7 provides important insights in the sources of refugee flows as well as in out- and in-migration peaks. The map provides complementary information on these patterns.

Fig. 8 shows the refugee outflows from Syria using choropleth maps for the years 2012 and 2018. A high level of spatial abstraction (i.e., regions) is shown in a1 and b1, while a detailed spatial level (i.e., countries) is depicted in a2 and b2. As pointed out before, there is a first peak in the outflow in 2012, which reached 730K people. The map in Fig. 8(a1) reflects the strong geographic concentration of Syrian refugees. Most of them moved to a neighboring country such as Turkey, Jordan or Iraq, but they also moved to EU regions such Western Europe. Fig. 8(a2) reveals that within the Western Europe region, the countries where Syrian refugees moved are Germany and Sweden (note the dark color of these countries). Similarly, Fig. 8(b1) shows a high spatial comparable information for the year 2018. In that year, the number of Syrian refugees reached 6.7M. The local spatial distribution is similar to that of 2012 (see Fig. 8(a1)). A large majority of refugees moved to the neighboring countries. However, Fig. 8(b2) reveals that a non-negligible fraction of them decided to move longer distances (e.g., to Germany, Austria, Spain, Sudan, Morocco or Nigeria). As an interesting point, we can say that the Syrian refugees move more towards the Eastern countries, probably due to the political problems with the countries of the West (e.g., Afghanistan and Libya). This behavior is evident by comparing Fig. 8(a2) y (b2).

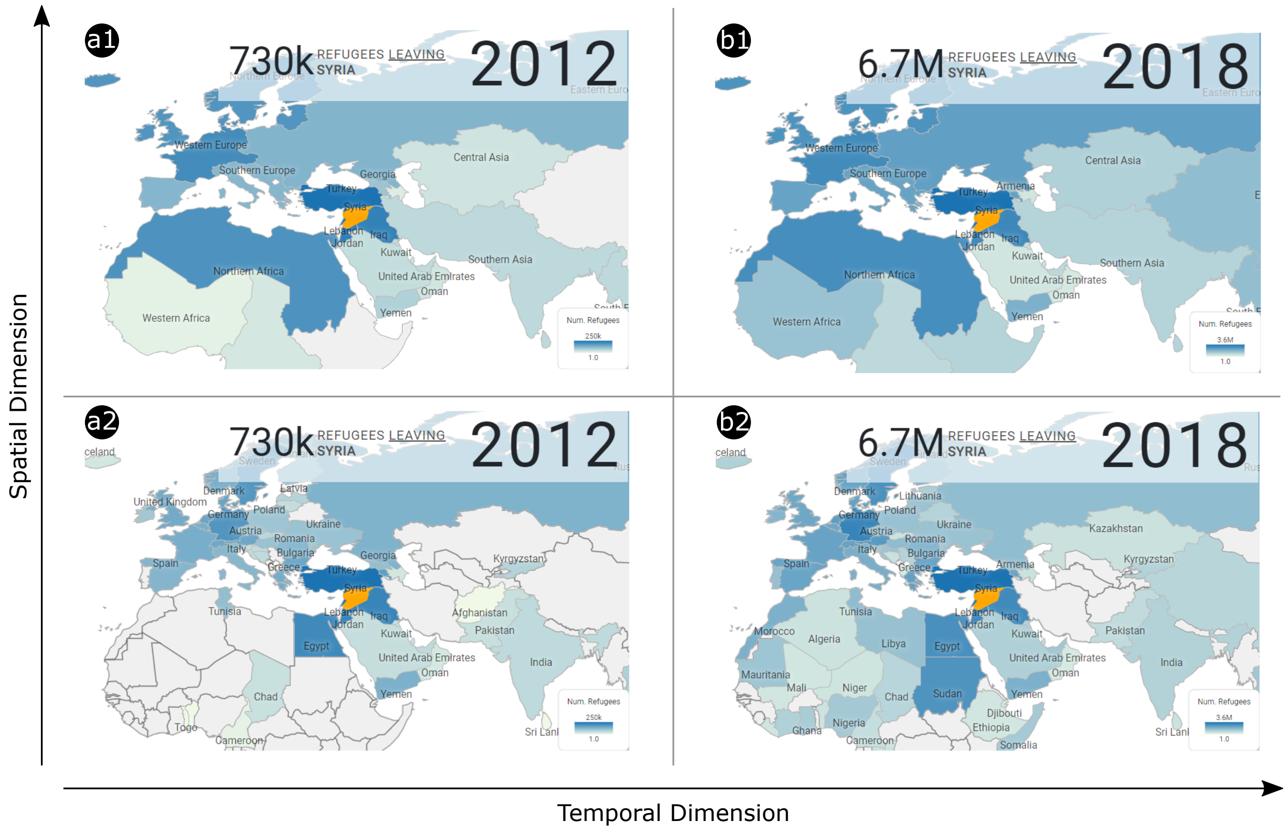

Figure 8: Temporal and spatial comparison of refugees leaving Syria in 2012 and 2018.

8 Conclusion and Future Work

In this paper, we present *EvoFlows*, an interactive visualization tool that allows to highlight and explore the spatio-temporal patterns underlying origin-destination flow data. *EvoFlows* is based on three interactive components: (i) the spatial view that uses maps to represents the geographic features of the data, (ii) the temporal view that uses a stacked approach to represent the temporal dimension at different spatial scales, and (iii) the snapshot panel that allows to save/load a snapshot to enhance communication tasks. Our work enhances temporal and spatial exploration through tailored animated transitions between visual components. The tool also eases the exploration task by highlighting remarkable periods in the evolution of flows. Moreover, *EvoFlows* was tested using a real-world dataset that contained 195x195 refugee movements worldwide, overcoming the problem of scalability present in previous approaches.

As future work, we plan to study three aspects: (i) test *EvoFlows* for other application domains like animal movements or network traffic, (ii) conduct a formal user study compared with similar approaches, and (iii) include advanced change point detection algorithms [3,25] to automatically discover interesting configuration snapshots.

References

[1] N. Adrienko and G. Adrienko. Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2):205–219, 2010.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 2011.

[3] S. Aminikhanghahi and D. J. Cook. A Survey of Methods for Time Series Change Point Detection. *Knowledge and Information Systems*, 51(2):339–367, 2017.

[4] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer, 2013.

[5] G. Andrienko, N. Andrienko, G. Fuchs, and J. Wood. Revealing Patterns and trends of Mass Mobility Through Spatial and Temporal Abstraction of Origin-Destination Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2120–2136, 2016.

[6] G. Andrienko, N. Andrienko, and A. Savinov. Choropleth Maps: Classification Revisited. In *Proceedings of the International Cartographic Conference (ICA)*, pp. 1209–1219, 2001.

[7] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, 2006.

[8] B. Bach, E. Pietriga, and J.-D. Fekete. GraphDiaries: Animated Transitions and Temporal Navigation for Dynamic Networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, 2013.

[9] R. A. Becker and W. S. Cleveland. Brushing Scatterplots. *Technometrics*, 29(2):127–142, 1987.

[10] R. Bourqui, D. Ienco, A. Sallaberry, and P. Poncelet. Multilayer Graph Edge Bundling. In *Proceedings of the Pacific Visualization Symposium (PacificVis)*, pp. 184–188. IEEE, 2016.

[11] 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.

[12] I. Boyandin, E. Bertini, and D. Lalanne. A Qualitative Study on the Exploration of Temporal Changes in Flow Maps with Animation and Small-Multiples. *Computer Graphics Forum*, 31(3pt2):1005–1014, 2012.

[13] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Timelines Revisited: A Design Space and Considerations for Expressive Story-

telling. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2151–2164, 2016.

[14] L. Byron and M. Wattenberg. Stacked Graphs - Geometry & Aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1245–1252, 2008.

[15] W. S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[16] A. Cockburn, A. K. Karlson, and B. B. Bederson. A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. *ACM Computing Surveys*, 41(1):2:1–2:31, 2009.

[17] E. Cuenca, A. Sallaberry, F. Y. Wang, and P. Poncelet. MultiStream: A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3160–3173, 2018.

[18] Q. Cui, M. Ward, E. Rundensteiner, and J. Yang. Measuring Data Abstraction Quality in Multiresolution Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):709–716, 2006.

[19] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li. Geometry-Based Edge Clustering for Graph Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1277–1284, 2008.

[20] G. W. Furnas. Generalized Fisheye Views. In *Proceedings of the Conference on Human Factors in Computing System (CHI)*, pp. 16–23. ACM, 1986.

[21] A. L. Griffin, A. M. MacEachren, F. Hardisty, E. Steiner, and B. Li. A Comparison of Animated Maps with Static Small-Multiple Maps for Visually Identifying Space-Time Clusters. *Annals of the Association of American Geographers*, 96(4):740–753, 2006.

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

[23] D. Guo. Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1041–1048, 2009.

[24] D. Guo and X. Zhu. Origin-Destination Flow Data Smoothing and Mapping. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2043–2052, 2014.

[25] V. Guralnik and J. Srivastava. Event Detection from Time Series Data. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pp. 33–42. ACM, 1999.

[26] R. L. Harris. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, 2000.

[27] J. Heer and G. Robertson. Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007.

[28] D. Holten and J. J. Van Wijk. Force-Directed Edge Bundling for Graph Visualization. *Computer graphics forum*, 28(3):983–990, 2009.

[29] D. A. Keim. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

[30] D. F. Marble, Z. Gou, L. Liu, and J. Saunders. Recent Advances in the Exploratory Analysis of Interregional Flows in Space and Time. *Innovations in GIS*, 4:75–88, 1997.

[31] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[32] D. Phan, L. Xiao, R. Yeh, and P. Hanrahan. Flow map layout. In *Proceedings of the IEEE Symposium on Information Visualization*, pp. 219–224. IEEE, 2005.

[33] S. Pupyrev, L. Nachmanson, S. Bereg, and A. E. Holroyd. Edge Routing with Ordered Bundles. In *Proceedings of the International Symposium in Graph Drawing (GD)*, pp. 136–147. Springer, 2011.

[34] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the Symposium on Visual Languages (VL)*, pp. 336–343. IEEE, 1996.

[35] D. M. Stephen and B. Jenny. Automated layout of origin–destination flow maps: U.S. county-to-county migration 2009–2013. *Journal of Maps*, 13(1):46–55, 2017.

[36] W. Thompson and S. Lavin. Automatic Generation of Animated Migration Maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 33(2):17–28, 1996.

[37] W. R. Tobler. Experiments in Migration Mapping by Computer. *The American Cartographer*, 14(2):155–163, 1987.

[38] A. M. Voorhees. A General Theory of Traffic Movement. *Transportation*, 40(6):1105–1116, 2013.

[39] J. Wood, J. Dykes, and A. Slingsby. Visualisation of Origins, Destinations and Flows with OD Maps. *The Cartographic Journal*, 47(2):117–129, 2010.

[40] J. Wood, A. Slingsby, and J. Dykes. Visualizing the Dynamics of London's Bicycle-Hire Scheme. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 46(4):239–251, 2011.

[41] Y. Yang, T. Dwyer, S. Goodwin, and K. Marriott. Many-to-Many Geographically-Embedded Flow Visualisation: An Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):411–420, 2017.

[42] Y. Yang, T. Dwyer, B. Jenny, K. Marriott, M. Cordeil, and H. Chen. Origin-Destination Flow Maps in Immersive Environments. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):693–703, 2018.