VisMiFlow: Visual Analytics to Support Citizen Migration Understanding Over Time and Space

Andreas Scheidl[1] and Roger A. Leite[1] and Silvia Miksch[1]

[1]TU Wien, Vienna, Austria.

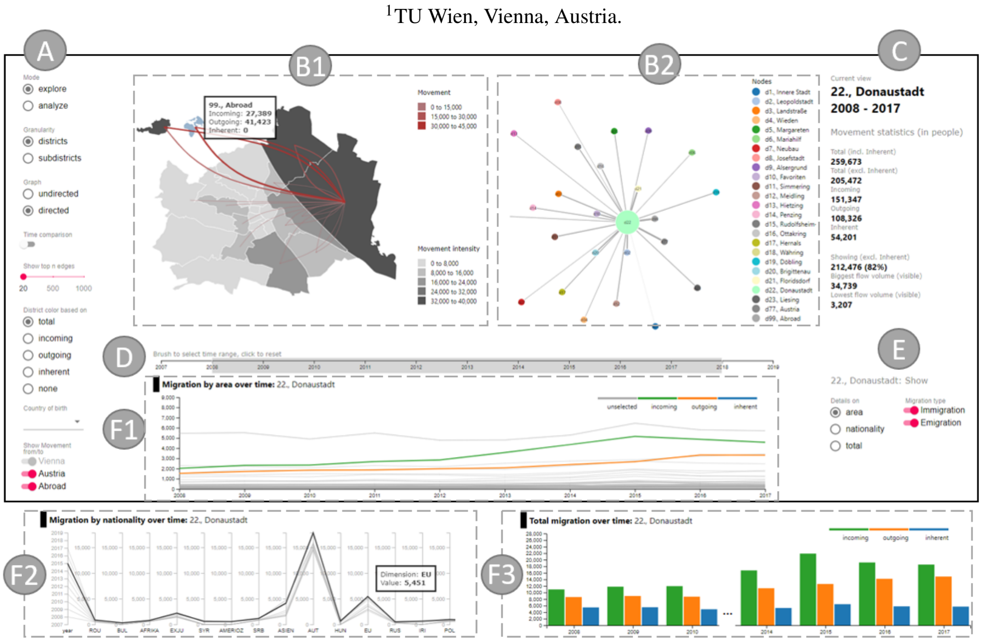

Figure 1: *Screenshot of MiFlow (Migration Flow). (A) General Filter. (B) Main Visualization: (B1) Exploration mode and (B2) Analysis mode. (C) Statistics. (D) Timeline. (E) Special Filter. (F) Times Series Data: (F1) Migration by area, (F2) Migration by country of birth, and (F3) Total migration. Each component is described in Section 4.*

Abstract

Multivariate networks are complex data structures, which are ubiquitous in many application domains. Driven by a real-world problem, namely the movement behavior of citizens in Vienna, we designed and implemented a Visual Analytics (VA) approach to ease citizen behavior analyses over time and space. We used a dataset of citizens' movement behavior to, from, or within Vienna from 2007 to 2018, provided by Vienna's city. To tackle the complexity of time, space, and other moving people's attributes, we follow a data-user-tasks design approach to support urban developers. We qualitatively evaluated our VA approach with five experts coming from the field of VA and one non-expert. The evaluation illustrated the importance of task-specific visualization and interaction techniques to support users' decision-making and insights. We elaborate on our findings and suggest potential future works to the field.

CCS Concepts

• ***Data*** $\rightarrow$ *time-oriented, multivariate, geospacial, flow events;*

1. Introduction

Open government data initiatives worldwide offer a vast amount of data on various topics. Complex datasets make it possible to construct more accurate models of real-world situations. These models enable people to acquire a deeper understanding of the relations between the data's facts. For these reasons, it is getting harder and harder to retrieve meaning or even finding important insights within this fast-growing complexity and the sheer amount of data. Since data visualizations tools support people to solve tasks in a very effective and illustrative way, taking insights from complex data relations is also getting more difficult. The key is to interact with the data using visual representation. This empowers users to explore and understand it.

Vienna's city provides the dataset on which our approach is based, precisely the department MA 23 ("Wirtschaft, Arbeit und Statistik") [MA2]. It contains statistics and movement data of citizens of Vienna gathered by Statistik Austria [Sta]. Our approach's key challenges are (1) designing appropriate visualization and interaction techniques to provide insight into a complex multivariate, spatial network that changes over time and (2) evaluating our VA approach properly. The design process is carried out based on the three cornerstones of the design triangle [MA14], *data*, *users*, and *tasks*. We implemented it using web technology, precisely, a combination of React [Rea] and D3.js [D3J]. Thus, our main contributions are:

- We tackle a real-world problem of city migration movements and iteratively designed a VA approach to support and improve decision-making on city planning.
- We evaluated our approach in a task-based qualitative user study with five domain experts and one non-expert collecting strengths and weaknesses.

2. Related Work

Movement Domain: Ko et al. [KAW*14] explore multivariate networks within the logistic domain. Its visual representations named as Petal and Thread allow network exploration and are focused on anomaly detection of movement data. Adrienko et al. [AAB*13] analyze different movement data visualization techniques and provides open challenges. They conclude that little has been done concerning multi-scaling space and time, which we tackle in our work by allowing district and sub-districts analysis. Chen et al. [CGW15] introduce the basic concept and pipeline of traffic data visualization. This survey discuss different data processing techniques and summarizes the state of the art for depicting the spatial, temporal, categorical, and numerical features of traffic data.

Network Topology: visualizing the topology and the changes in a network is an important task. Detailed studies of the structural change of a network can be found in Kerren et al. [KPW14] as well as Nobre et al. [NSML19]. Due to these networks' size, structure, and complexity, it is almost impossible to visualize everything at once while providing the *insight* we strive for. Therefore, an interactive approach is needed to let the user explore different angles of the data. Kerren et al. [KPW14] categorize different interaction techniques on three levels.

Design Principles: Jenny et al. [JSM*18] identify *design principles* to foster readability and avoid visual clutter. Based on user studies, they analyze the effectiveness of different design choices of the edges. Curved lines are preferable to straight lines because the viewers had a lower error rate when answering the questions in maps with curved lines. While sharp lines to some extend outperform arrowheads [HIvF11], the following disadvantages of sharp lines are: Long lines present a smaller gradient than short lines, therefore, they are resulting in different ambiguous gradients; gradients of thin lines are hard to see; and the direction of incoming flows is hard to be identified due to the very thin end or outgoing edge. Another approach to minimize visual clutter is presented by Holten [Hol06]. Edge bundling reduces visual clutter when applying to a geographic map layout (see Flight Connections in [Fli]). In contrast to the discussed methods, *OD-Maps* [WDS10] does not use the map as the main visualization component but abstracts the areas into a grid of equally sized cells. It offers an alternative view of extensive datasets, but the exact location, size, and shape properties are lost due to the grid's abstraction. Boyandin et al. [BBBL] present *Flowstrates*, a unique interactive approach to analyze the change over time of the flow volume. It uses two separate maps to display the origin's geographic location and destination.

3. Problem Definition

In this section, we define the requirements for our prototype by characterizing the cornerstones of the *Data-Users-Tasks-Design triangle* described in the work of Miksch et al. [MA14].

Data. The dataset consists of real-world movement of people who immigrated *to*, emigrated *from*, or changed their residency *within* Vienna from the years 2007 to 2018. The source and target regions in the movement data are divided into 250 sub-districts of Vienna and two regions for movement from/to the *rest of Austria* and from/to *Abroad*. Furthermore, the movement data contains information about the moving people. **Users.** The targeted group consists of employees of Vienna, respectively, the department of the city development. **Tasks.** The high-level tasks, which reflect the motivation of the users [AMST11] for our approach, are: (i) Understanding the migration *to* and *in* Vienna, (ii) analyzing changes over time, (iii) evaluating the impact of events (e.g., refugee crisis), (iv) analyzing specific data features, and (v) finding patterns (over time). These five tasks are refined into Requirements in Section 3.1.

3.1. Requirements

The findings in defining the cornerstones of the data-user-tasks design triangle leads us to the main requirements of our approach:

R1: District and sub-districts geo distribution. (i,iii) It is required to visualize the data within the different scale of spatial representation of Vienna: district and sub-districts. This leads to many possible advantages, like observing the *distance* of movement and *neighbourhood* relations. **R2: Detailed view of districts. (i,iv)** The user can select a specific district or sub-district to see details, visualizing every aspect of the data at once. Therefore, to avoid distraction by the visual clutter, we allow for *zooming* into the data. The zoom feature enables nodes' detailed analysis showing data aspects that the overview may not present. **R3: Movement. (i,iii,iv)** The system shows the movement data. The migration flow, represented by the edges connecting the nodes in the network, has to be presented visually. This feature enables the user to observe patterns of movement and the intensity of the relations between the areas. **R4: Temporal analysis. (i,ii,iii,iv,v)** The user can select a specific time period. This feature enables the user to observe the migration flow in a particular time frame and may correlate it to events. It plays a significant role in analyzing the city's migration flow. **R5: Immigration rate. (i,iii,v)** Observing the immigration of a specific district enables the user to understand where people are moving to the selected district. This may gain *insight* about the *composition* of a district. **R6: Emigration rate. (i,iii,v)** Observing the emigration of a specific district enables the user to understand where they are moving. This feature may gain *insight* to certain events that may have caused intense emigration or correlation patterns with other districts. **R7: Internal migration. (i,iv,v)** Observing the migration flow within a selected district enables the user to understand a district's stability. If there was no movement at all, this attribute might indicate people's satisfaction within a district. **R8: Geopolitical entities. (i,iii,iv)** The system enables to filter by geopolitical entities. E.g., observing the migration flow of certain *countries of birth* allows *insight* into various attributes about the districts, like clustering or diversity.

4. Design and Implementation

To fulfill the defined requirements and tasks, we chose various VA techniques to leverage their strengths. Each of the selected elements is linked to the requirements that they satisfy. In this sections, we are referring to labels A to F in Figure 1.

4.1. Composition and Layout

The visualization layout comprises various views, which fulfill a specific need to gain insight into the complex migration flows and their different data attributes. Each of these sections is part of a grid layout to arrange the various components.

General Filter (A) allows selecting the *mode* and filtering the migration data. It also allows the selections and filters control the *granularity* for showing either districts or sub-districts; the *information encoding* of the nodes (color/size) and edges (direction/limit); the *countries of birth* of the migration data; and the *number of edges* (shown in the main visualization components).

Main Visualization (B), shows one of two available Node-Link-Layouts, depending on the *mode* the user has selected. In the **Exploration Mode** (B1), the migration data is presented as a geographical map of Vienna. It has many advantages over, e.g., a tabular representation like an adjacency matrix. The user benefits from the geographical attributes, observing actual distances, neighborhood relations, and geographical clusters. Another advantage over a tabular layout is that a geographical representation is very well known to humans. We opt for an arrowhead at the target's endpoint. The exact position from which an edge starts or terminates is a crucial visual aspect. In (sub-)districts, where a lot of edges originate or terminate, visual clutter can be too high other edges may cover the arrowheads. Therefore, we introduce a circular buffer, offset from the district's center. According to the line's angle (based on [JSM*18]), the circle's border represents the position on which edges are starting from or terminating at. In the **Analysis Mode** (B2) the user can observe the relationship between districts disregarding their geographical position, displayed as a Force-Directed-Node-Link-Layout. This mode avoids potential visual clutter from the Exploration Mode and supports different tasks. Three use cases are: (1) finding the migration correlation between sub-districts, (2) identification of most important immigration node for people born in the same country, and (3) the clustering of nodes. Instead of a node-link approach, we considered using a combination of semantic substrate design [AS07] and treemap. However, the big amount of sub-districts overloaded the method.

Statistics View (C) shows a summary of the currently displayed mode as well as statistical values about the data selection and different aspects of the data aggregation.

Timeline (D) acts as selection tool for the desired time period. The user may choose a single year or a longer period ranging from 2007 to 2018. To fulfill requirements **R2** and **R4**, it is necessary to enable the user to *drill down* into the data by selecting a specific period of interest.

Special Filter Board (E) allows further filtering of the time series data regarding the selected (sub-)district's immigration or emigration. Based on the selected district or sub-district, the user can show and hide data of the following migration types: immigration, emigration, and inherent migration.

Time Series Data Component (F1-3) allows exploring the movement of a selected period of interest and shows the change over time for various aspects. It consists of three different time series visualizations that alternate between exploring and analyzing: (F1) migration by area over time, (F2) migration by country of birth over time, and (F3) total migration over time. **Migration by area over time (F1)** shows the temporal relation of the movement to or from the other areas as a multiple line graph. For each selection related district a line is plotted along with the time instants on the x-Axis, showing the change over time of the flow amount for this relation y-Axis. The *color coding* is: blue encodes the inherent (migration within the selected area), red encodes the emigration (outgoing), and green encodes the immigration (incoming). **Migration by country of birth over time (F2)** helps to understand the cultural diversity of an area of interest and may support the user, e.g., developing tailored integration strategies. A *parallel coordinates graph* satisfies the migration flow's multivariate attributes. The dimensions of this parallel coordinates graph are the different countries or geopolitical entities and the temporal dimension. It can also be used to understand events, like the refugee crises in 2015 and 2016. **Total migration over time (F3).** The third time-series visualization shows a simple grouped bar chart, displaying the sum of people who immigrated, people who emigrated, and people who moved within the region for each time instant. In this view, the user can compare each migration type's volume within a year or observe trends over time.

Interaction is a vital component of VA [Shn96]. We already introduced interactions on the *Visual-structure level* (Selection of areas) and the *Data level* (Filtering the data). The following methods describe interactions on the *View level*. **Cross-highlighting.** It is important to cross-reference certain entities (nodes and edges) to preserve the mental model throughout the whole visualization. We chose to apply the concept of cross-highlighting to the main visualization components as well as to the time-series visualization. **Tooltip.** Every view consists of many visual entities (nodes, edges, bars) that have different attributes attached to them. If the user is interested in one entity, the attached attributes should be displayed by *Hovering*. **Zoom.** The main visualization components show a lot of complex information. Zooming into the diagrams changes the viewpoint and enlarges the interesting area in a *"details-on-demand"* [Shn03] fashion.

5. Evaluation

In the evaluation phase, the implemented visualization prototype was evaluated against the requirements defined in Section 3.1. The evaluation of techniques in the field VA is a challenging task. Heuristic evaluation is a common approach to identifying problems in the usability of a user interface. The evaluators rate the UI based on a defined set of established usability principles. The works of Zuk et al. [ZC06] and Forsell et al. [FJ10] argue that heuristic evaluation produces useful results even if the evaluators are less experienced in the domain. ICE-T [WAM*19] also shows positive results from evaluations oriented to visualization experts as evaluators. Moreover, visualization experts have critical technical value. According to these scientific findings, we decided to use visualization experts as evaluators.

The visualization is assessed by five visualization experts and one UI expert, who work on a series of questions. They had to answer if the defined tasks (T1 to T11 is listed in the supplementary material) can be achieved, which reflect the main requirements as shown in table 1. The evaluation is divided into five parts: (1) background assessment, (2) brief introduction, (3) prototype familiarization (10 min), (4) task solving, and (5) prototype review (see table 2). The average time to conduct the evaluation was two hours. Each evaluator was able to solve all the given tasks.

Task	Requirements	Task	Requirements
T1	R1, R2, R4, R5	T7	R1, R2, R3, R8
T2	R1, R2, R4, R5, R8	T8	R5, R6, R7, R8
T3	R8	T9	R1, R2, R3
T4	R1, R2, R3, R8	T10	R2, R3, R5, R8
T5	R1, R2, R3, R4	T11	R3, R4, R5
T6	R1, R2, R4, R5, R8		

Table 1: *Task-Requirement mapping. The tasks T1 - T11 are explained in the supplementary material.*

Impression	E1	E2	E3	E4	E5	E6	AVG score
Usability	3	5	5	4	4	4	4.17
Features	5	5	5	4.5	4	3	4.42
Performance	4	4	4	4.5	4.5	5	4.33
Overall	5	5	5	5	4	3	4.50

Table 2: *The evaluators' rating impressions from 1: not usable to 5: consistent and intuitive of the prototype. The prototype was rated to be useful, relatively easy to use, rich in features, and fast.*

5.1. Identified strengths

Throughout the evaluation sessions, the evaluators highlighted strengths of the chosen design.

Interactivity. The combination of filtering, selecting, hovering, and cross-highlighting supported the evaluators to gain insights. Three evaluators explicitly mentioned the option to change the number of edges as an important tool to control the map's visual clutter and readability. **Map design.** The evaluators highlighted the efficiency of analyzing network flows with this visualization design paired with the interaction techniques. Furthermore, all evaluators mentioned the familiarity which lies within the utilization of geographic maps. **Force layout.** In contrast to the *Map layout*, this visualization was utilized heavily by only two evaluators. The other four evaluators took the *Force directed layout* as a valuable addition to the map but mentioned that the map layout is adequate enough to do an intense analysis of migration patterns or even cluster identification.**Time series data component.** The evaluators utilized all three time-series data components. Depending on the task, they explored and analyzed the relationships between districts in the line chart, the change over time of geopolitical entity movement in the parallel coordinates graph, and the change over time of the bar chart's total movement intensity.

5.2. Identified Weaknesses and Future Work

By combining our approach‘s insights and the evaluators' discussion, we define potential features to leverage the analysis process.

Data extension. Due to the network's multivariate aspects, it is possible to extend the data by adding more attributes to either nodes or edges. One interesting example for *extending node‘s attributes* could be cross-referencing the migration data with the districts' housing prices. Other attributes could be the densities of public transport, educational or health care institutions. Moreover, **visual map overlays** could support the extra dimensions understanding.

Data aggregation. Adding a lot of attributes to the network increases complexity. One approach to diminish this complexity is to apply *Dimensionality Reduction techniques*. Time-Curves [BSH*16] is a lucid example applying such a technique. *Multi-dimensional scaling (MDS)* or *Principal component analysis* could reduce the data dimensions for further analysis, as described in the work of Aigner et al. [AMST11]. The results of *Dimensionality Reduction* could be used for the *Force directed graph layout* or feed data to new visualizations.

Saving filter settings, defining presets. The option to save a specific filter setting and refer to it later on in a comparison task can speed up the analysis and exploration process. There could also be predefined settings to show migration patterns for specific events, e.g., the 2015 refugee crisis.

6. Conclusion

We presented a VA approach to explore and analyze a spatial, multivariate network over time. The target audiences in this work were the city developers. Applying the design triangle's structured methodology, we efficiently identified the fundamental requirements that guided our approach‘s design, development, and evaluation. This combination proved to be very effective and amplified the expressiveness of the data. Six evaluators were able to gain insight into the internal and external migration of residents in Vienna. Moreover, we discussed and suggested future works based on the identified strengths and weaknesses.

7. Acknowledgement

Thanks to "SmartCT" at TU Wien and the Austrian Science Fund (FWF), grant P31419-N31 (KnoVA).

References

[AAB*13] ANDRIENKO G., ANDRIENKO N., BAK P., KEIM D., WROBEL S.: *Visual analytics of movement*. Springer Science & Business Media, 2013. 2

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*. Springer London, 2011. 2, 4

[AS07] ARIS A., SHNEIDERMAN B.: Designing semantic substrates for visual network exploration. *Information Visualization 6*, 4 (2007), 281–300. 3

[BBBL] BOYANDIN I., BERTINI E., BAK P., LALANNE D.: Flowstrates: An approach for visual exploration of temporal origin-destination data. *Computer Graphics Forum 30*, 3, 971–980. 2

[BSH*16] BACH B., SHI C., HEULOT N., MADHYASTHA T., GRABOWSKI T., DRAGICEVIC P.: Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 559–568. doi:10.1109/TVCG.2015.2467851. 4

[CGW15] CHEN W., GUO F., WANG F.-Y.: A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems 16*, 6 (2015), 2970–2984. 2

[D3J] d3.js. https://d3js.org/ accessed November 21, 2019. 2

[FJ10] FORSELL C., JOHANSSON J.: An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces* (2010), pp. 199–206. 4

[Fli] Flight path connection visualization. https://bl.ocks.org/sjengle/2e58e83685f6d854aa40c7bc546aeb24 accessed November 29, 2019. 2

[HIvF11] HOLTEN D., ISENBERG P., VAN WIJK J. J., FEKETE J.: An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *2011 IEEE Pacific Visualization Symposium* (March 2011), pp. 195–202. doi:10.1109/PACIFICVIS.2011.5742390. 2

[Hol06] HOLTEN D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE transactions on visualization and computer graphics 12* (09 2006), 741–8. doi:10.1109/TVCG.2006.147. 2

[JSM*18] JENNY B., STEPHEN D. M., MUEHLENHAUS I., MARSTON B. E., SHARMA R., ZHANG E., JENNY H.: Design principles for origin-destination flow maps. *Cartography and Geographic Information Science 45*, 1 (2018), 62–75. doi:10.1080/15230406.2016.1262280. 2, 3

[KAW*14] KO S., AFZAL S., WALTON S., YANG Y., CHAE J., MALIK A., JANG Y., CHEN M., EBERT D.: Analyzing high-dimensional multivariate network links with integrated anomaly detection, highlighting and exploration. In *2014 IEEE conference on visual analytics science and technology (VAST)* (2014), IEEE, pp. 83–92. 2

[KPW14] KERREN A., PURCHASE H. C., WARD M. O.: *Multivariate Network Visualization: Dagstuhl Seminar #13201, Dagstuhl Castle, Germany, May 12-17, 2013, Revised Discussions*. Springer International Publishing, 2014. 2

[MA2] Ma 23 - wirtschaft, arbeit und statistik. https://www.wien.gv.at/kontakte/ma23/ accessed November 21, 2019. 2

[MA14] MIKSCH S., AIGNER W.: A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics, Special Section on Visual Analytics 38* (2014), 286–290. 2

[NSML19] NOBRE C., STREIT M., MEYER M., LEX A.: The state of the art in visualizing multivariate networks. *Computer Graphics Forum (EuroVis '19) 38* (2019), 807–832. doi:10.1111/cgf.13728. 2

[Rea] React, a javascript library for building user interfaces. https://reactjs.org/ accessed March 12, 2020. 2

[Shn96] SHNEIDERMAN B.: The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), 336–343. 3

[Shn03] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 2003, pp. 364–371. 4

[Sta] Austria statistik. https://www.statistik.at/ accessed November 21, 2019. 2

[WAM*19] WALL E., AGNIHOTRI M., MATZEN L., DIVIS K., HAASS M., ENDERT A., STASKO J.: A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 491–500. doi:10.1109/TVCG.2018.2865146. 4

[WDS10] WOOD J., DYKES J., SLINGSBY A.: Visualisation of origins, destinations and flows with od maps. *The Cartographic Journal 47*, 2 (2010), 117–129. 2

[ZC06] ZUK T., CARPENDALE S.: Theoretical analysis of uncertainty visualizations. In *Visualization and data analysis 2006* (2006), vol. 6060, International Society for Optics and Photonics, p. 606007. 4