What-Why Analysis of Expert Interviews:
Analysing Geographically-Embedded Flow Data

Yalong Yang*   Sarah Goodwin†

Caulfield School of Information Technology, Monash University

ABSTRACT

In this paper, we present our analysis of five expert interviews, each from a different application domain. Such analysis is crucial to understanding the real-world scenarios of analysing geographically-embedded flow data. The results of our analysis show that similar high-level tasks were conducted in different domains. To better describe the targets of these tasks, we proposed three flow-targets for analysing geographically-embedded flow data: single flow, total flow and regional flow.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

Many domain experts and analysts around the world are interested in discovering insights and patterns related to some form of commodity flow between different geographic locations. Such activities include migration patterns, urban planning, animal movement, diesase mapping, financial trading and commuting behaviour.

Geographically-embedded flow data, often termed spatial movement data, trajectory or origin-destination (OD) data, contains geographical locations (the origin, destination and potentially mid-way points), a connection or trajectory between them and a value referring to the magnitude of the flow between the locations. Data of this type is increasing with advances in technology, yet there are still limited tools to effectively analyse and visualise such data in order to discover patterns and inform decision making. In recent years we have seen growing support and solutions for the analysis of flow data, yet to the best of our knowledge, there is no research focusing on identifying real-world tasks conducted by the domain experts.

This work was conducted as part of a 3.5 year PhD project focusing on exploring the design space of geographically-embedded flow visualisation [13, Chap. 2]. The motivation for these interviews was to understand the requirements and role of flow data in real-world applications and existing work flows across different disciplines, in order to help inform new visualisation designs [14, 15].

The contributions of this work are: 1) three high-level flow-targets for analysing geographically-embedded flow data; 2) using Munzner's framework [11] to abstract analytic tasks in interview analysis; 3) five use cases for motivating flow data visualisation design.

2 RELATED WORK

Task Taxonomy — Visualisation tasks have been grouped and classified in a number of examples e.g. [2, 9]. More recently, Brehmer and Munzner compiled these classifications into the *what-why-how* framework [6, 11] to systematically abstract work flows for visual analytics. Specific taxonomies for geographical data visualisation has also been explored. Closest to this study is the study of Roth [12], who interviewed expert interactive map users to develop a taxonomy of interaction primitives for map-based visualisation. Earlier

---

*e-mail: yalong.yang@monash.edu
†e-mail: sarah.goodwin@monash.edu

research by Andrienko *et al.* also discussed several task taxonomies for analysing spatial-temporal data [3]. Both Roth and Andrienko *et al.* distinguished tasks into two levels according to the level of data analysis: *elementary* and *synoptic*. *Elementary tasks* refer to individual elements; *Synoptic tasks* [3] (refered to as 'general' in [12]) involve the whole reference set or its subsets. This task taxonomy research inspired our detailed discussion of task targets in Sec. 6.

Expert Interviews — There are several other recent interview studies in visualisation research. Alspaugh *et al.* interviewed data analysts in a range of industrial, academic, and regulatory environments to understand exploratory activities [1]. Kandel *et al.* interviewed enterprise analysts about their overall analysis process and the organizational context [8]. Both were conducted with a broad focus, while in this paper we concentrate specifically on the analysis of geographically-embedded flow data across different domains.

3 METHODOLOGY

Recruitment — Potential interviewees were identified and recruited from diverse domains. These experts were chosen because of their knowledge and experience of having analysed geographically-embedded flow data during their professional careers (see Tab. 1).

Preparation — In order to tailor the interview questions to meet the specific needs and expertise of the interviewee and to ensure the interview time was kept to a minimum, prior to each interview each interviewee conducted a preliminary survey. This survey contained three specific sections: *consent*; *personal information*; and *related projects*. We present an overview of these details in Tab. 1.

A semi-structured interview was prepared base on the survey information for each candidate. The basic structure of the interview consisted of the following questions:

- Can you give a brief overview of the related project(s)?
- Can you explain the data in more detail?
- What are/were the initial motivations for the project(s)?
- What tasks are involved in the analysis?
- Did you use visualisation tools in your analysis? If so, which?
- Do you have suggestions on how to improve these tools?

Interview Method — Interviews were conducted via VoIP software, such as Skype or face-to-face when possible. All interviews were audio-recorded and later transcribed. A pilot interview was also carried out internally to test the structure, questions and timing.

Analysis — As each project was unique the transcriptions contained very detailed descriptions of each project and whilst the data described in each interview is similar in structure (containing geographical locations and flows between them), the data is analysed and visualised for very different reasons. An initial qualitative coding exercise, similar to [1, 12], revealed the difficulty of using this method of analysis due to diversity of the projects. As an alternative, we chose to analyse the transcripts using the *what-why-how* framework by Munzner [11] as this framework provides us with a structure to think systematically about the similarities in analysis tasks and design choices, within the large and complex design space. For this analysis, we paid particular focus on:

- **What: Data Abstraction**. Datasets are classified into: tables, networks, fields, geometry (or spatial) [11, Chap. 2].
- **Why: Task Abstraction**. Tasks are analysed with: *actions* and *targets* [11, Chap. 3].

Domain Code	Date Interviewed	Degree	Yrs Exp	Proficiency self-rating				M/F	Organisation(s)	Role	Main Tool(s)
				Prog	Stats	GIS	Vis				
Refugee (Sec. 5.1)	Dec-16	Master in GIS	16	5	5	5	5	M	UNHCR, International Organization for Migration	Information Manager	Excel, Tableau, D3
Urban (Sec. 5.2)	Sep-16	Master in Urban Planning	5	1	2	4	2	M	Urbanization Institute of Sun Yat-sen University	Urban Planner	Excel, ArcGIS, R, Python
	Dec-16†		3	3	3	4	3	F			
Flu (Sec 5.3)	May-18	PhD in Geomatics	10+	5	4	5	4	M	Uni of Melbourne, Uni of Zurich	Lecturer	QGIS, JS, Python Processing, Java, R
Energy (Sec 5.4)	Oct-16	PhD in Elec. Eng.	40+	3	4	2	3	M	SECV, ESC, Jacobs, Monash Uni	Senior Research Fellow	PSS/E
Aircraft (Sec 5.5)	Jul-18	PhD in HCI, Vis	5	5	4	4	5	M	Steria Group, Monash Uni	Research Fellow	FromDaDy, Unity, OpenGL, DirectX

Table 1: Interviewee background and expertise. Proficiency self-rating range from 1 (No Experience), 2 (Beginner), 3 (Knowledgeable), 4 (Skilled) and 5 (Expert) with clear definition in the survey (see supplementary materials). Abbreviations: UNHCR (United Nations High Commissioner for Refugees), SECV (State Electricity Commission of Victoria) and ESC (Essential Services Commission). † Date of presentation (see Sec. 5.2).

– There are three levels of *actions*: the highest refers to *consume* or *produce* information; the middle to *finding* the required information; and the bottom to query *targets* at different scopes.

– *Targets* refer to the aspect of the data that is of interest to the viewer. There are four kinds of high-level targets: all data, attributes, network data and spatial data.

A visualisation task can be described as *actions* with a *target*. Compared to other analysis methods, this framework guides the translation of domain-specific problems into abstract visualisation tasks using a multi-level typology of tasks. This typology provides abstract and flexible descriptions of tasks which allows useful comparison to be made between different application domains [6]. Due to very limited visualisation examples provided by the interviewees, we could not analyse the "How" part of the framework.

Limitations — The topics and tasks described by the interviewees might be different from those undertaken by other analysts in a similar position in the same domain. Due to our desire to gain a breadth in understanding, we had to spend time during each interview gaining knowledge of the domain. As each interview was limited to one hour we focused on understanding the high-level tasks, domain and data characteristics. This did not allow time to explore more detailed aspects of the data such as spatial scale, spatial resolution, amount, quality, uncertainty or time dimension. Therefore there is a possibility that we have missed tasks through our method of query and analysis. Further interviews and continued discussions could reveal further details and important tasks in each domain.

4 OVERVIEW

We briefly introduce the abstracted data and general tasks conducted in the five projects referenced in our interviews.

What: Data Abstraction — Geographically-embedded flow data can be considered as the combination of *spatial* data and *network* data. In all five interviews these two types of data were mentioned as being collected and analysed. Fig. 1 provides a graphic depiction of the data abstractions using icons from [11, Chap. 2].

In the case of energy analysis (see Sec. 5.4), geographic patterns in the data are secondary in current work flows, as analysts are more concerned about the network perspective of power grids. Thus, we highlight this circumstance using a dashed box in Fig. 1. Additional tabular data were sometimes used to add context (this is shown as a grey dashed box in Fig. 1). For example, analysts investigate the correlation between weather and refugee and migrant arrivals (see Sec. 5.1) and demographic data is used to inform the design of existing or new airways (see Sec. 5.5). With the exception of Refugee Movement (Sec. 5.1), all other interviewees mentioned scale hindering their ability to analyse and visualise the data.

Why: Task Abstraction — A task is described as the combination of actions and a target. Firstly, the highest level of action — ***Analyse***: from the interviews we identify two main goals: analysts *consume*
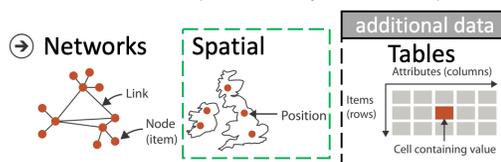

Figure 1: *What — data abstraction* for different interviews. Icons derived from [11, Chap. 2] under CC BY-4.0 Licence.

information to get better insights into the data as well as using other analytic tools to *produce* new information. *Consume* tasks mentioned in the interviews mainly related to *presenting*. This is possibly because the interviewees mainly talked about their experience of decision-making, planning and forecasting instead of generating new hypotheses. *Produce* tasks mentioned in the interviews mainly relate to *deriving* new data. For the middle level action — ***Search***: different specific actions were taken based on whether the identity and the location of the target is known or not. For the lowest level action — ***Query***: *identify*, *summarise* and *compare* are commonly mentioned in the interviews with different targets.

In terms of common *targets* of interest, the interviewees were interested in: *Trends*, *Outliers*, *Features* and *Topology*. Fig. 2 provides a graphic depiction of the task abstraction, using icons from [11, Chap. 3]. Fig. 1 and Fig. 2 are referred to throughout the following section, where we describe each individual interview in terms of problem context, data and task abstraction.

5 DETAILS

5.1 Refugee and Migrant Movement

At times of violent conflict in the world, hundreds of thousands of innocent lives can be affected. In such disastrous situations, many organisations around the world work hard to ensure people's safety. UNHCR and IOM (see Tab. 1) are two of these organisations. They both collect and analyse refugee and migrant data and produce situation reports to support evidence-based decision-making and advocacy on refugee issues.

What: Data Abstraction — Data collection and quality assurance is particularly difficult in this domain. Data collection is a manual process and there is often missing, incomplete or inaccurate data recorded. Collection involves multiple countries, languages and different standards. Data is often not detailed enough and this can cause difficulties during analysis; As stated by the interviewee, "*UNHCR and IOM conduct regular discussions with governments to attempt to better harmonise this information.*"

Important data related to refugee and migrant movement analysis is *arrival data*. This data is collected by local governments or agencies about individuals with the following attributes: *country of origin*; *country of arrival*; *intended destination*; *sex*; *age*; *whether travelling with companion(s), and if so with whom*. In the case of children, whether or not they are unaccompanied or have been
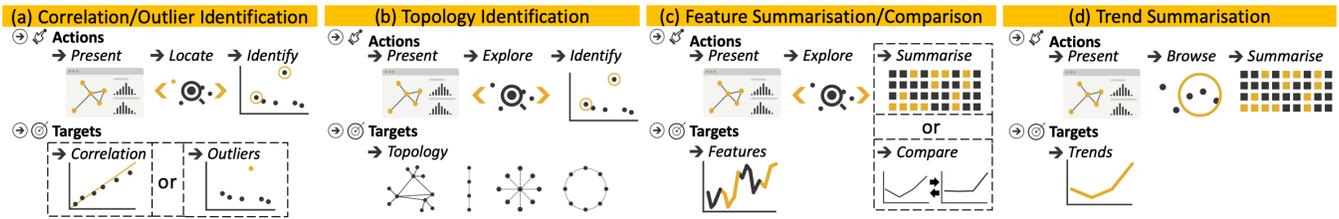
Figure 2: *Why — task abstraction* for different interviews. Icons derived from [11, Chap. 3] under CC BY-4.0 Licence.

separated from their parents. In some cases, refugees and migrants are relocated to another country after their arrival. Such individual records are collected with: *country of arrival* and *of relocation*.

The analysis is usually conducted with aggregated data rather than individual records. There are many different ways to aggregate records e.g. by OD pair, by time period or the combination of them. The analysts are also interested in estimating future arrivals. One way to achieve short term projections is to investigate the correlation between the refugee and migrant geographical locations and other factors that influence movement, such as the weather.

Why: Task Abstraction — The main motivation for this work is to more effectively align scarce resources. This aim can be broken into two distinct analytical tasks:

- Estimate the arrival. Better estimation can help governments and organisations prepare. There are different scenarios where visualisation can facilitate this task:
  - Analysts investigate if arrivals in different locations change across time (Fig. 2d: **Trend** Summarisation);
  - Analysts investigate possible factors that correlate with arrivals. For example, they found bad weather usually results in fewer arrivals. (Fig. 2a: **Correlation** Identification)
- Prioritise relief operations. Due to limited resources, the most urgent cases, like unaccompanied children, are prioritised. These cases are found through **Outlier** Identification (Fig. 2a).

Besides resource allocation, there is also a focus on identifying smuggling. For this, analytical modelling and calculations to **derive** the data are needed. Then visual analytical tools can help identify possible smuggling routes via **Outlier** Identification (Fig. 2a).

5.2 Urban Planning: a Bike Sharing Schema

Urban planning focuses on the design of the urban environment, including land, environmental resources and public infrastructures. With the rapidly increased population and size of cities, many challenges emerge and affect people's daily life. On the other hand, vast amounts of data can be collected from urban environments and activities. Such data can be used to improve urban planning practice. In this section, we describe a bike sharing schema planning project with the analysis of various datasets.

The interviewee worked on a bike sharing schema planning project for the city of Xiamen (population of ∼3.5 million). The project involved analysing the existing bike sharing schema and planning for an extension to encourage more bike sharing members and ultimately reduce car congestion in the city. Part of the analysis and findings were published in [16]. In addition, the task analysis from the interview in this section was supplemented with additional context from an oral presentation of the project (see Tab. 1).

What: Data Abstraction — One of the main datasets used in the analysis was the bike sharing useage data from the existing schema, in operation since July 2014. This data has two distinct components:

- *Docking Stations* consisting of: *geographic location*, *capacity* and *number of docked bikes* in real-time.
- *Journey records*, consisting of: *origin docking station*, *start time*, *destination docking station* and *end time*.

For context of the city and how the population move around the city, the following datasets were also used:

- Road Network: seven different types of roads in Xiamen.
- Public Transport: bus stations, routes and travel card journeys.

- Taxi Journeys: GPS tracked trajectories.
- Region Category: Living, Working and Mixed regions were derived from population and economic census data.
- Points of Interests (POIs): locations of important locations such as schools, hospitals, parks, shopping malls, industrial parks etc.

Why: Task Abstraction — Trying to encourage people to use bike sharing instead of cars can be achieved by expanding the catchment area (building new docking stations), adding more bikes to the schema (building new and increasing current docking station capacity), and/or facilitating safe cycling in the city by extending existing or constructing new bike lanes. A key question is how to choose the locations for these stations and lanes to best serve people's travel goals and fit within the constraints of the built environment.

The previous bike usage data was analysed, and revealed that the majority of previous use was for short distance travel. To explore this further, short distance travel routes in the form of OD and trajectories were extracted from previous usage data, public transport data, and taxi data. To plan for new docking stations, three types of candidates were identified through **Outlier** Identification (Fig. 2a) relating to popular origins or destinations, dense intersections of trajectories and POIs. To plan for new and extended bike lanes, the existing lane network was analysed, and missing short connections between docking stations and other facilities, e.g. POIs or public transit stations (Fig. 2a and c: **Outlier** and **Topology** Identification) were identified. These became candidates for new bike lanes.

In addition to the main aim, the interviewees discussed some other interesting findings, which conform to existing literature on bike sharing schema use in other countries, such as in London, UK [4]:

- Analysing bike usage and current bike lanes (see Fig. 2a: **Correlation** Identification), revealed that whilst many of the existing lanes were on main roads, these lanes were hardly used by the bike sharing riders. One possible reason suggested was safety concerns of riding on busy roads.
- Through analysing the topology of bicycle usage (Fig. 2c: **Topology** Identification), different patterns of use appear in different types (working, living and mixed) regions.

5.3 Flu Forecasting

Flu not only affects people's daily lives but the local economy and health services. Influenza epidemics vary substantially in size, location, timing and duration from year to year, making it a challenge to deliver timely and proportionate responses [10]. Understanding, and predicting epidemics is important to assess the risk and effectively manage the public health resources. The interviewee works with computational epidemiologists, introducing spatial tracking data into the prediction model aiming at fine-grained spatial resolution, to allow more targeted predictions for resources[1]. At the time of the interview, the project was at an early stage, and the team was focused on exploration. The data described and tasks abstracted describe the data analysis and expected tasks in later stages.

What: Data Abstraction — There are two main datasets in this project: flu notifications and urban mobility. Flu notification data contains weekly (time) updates of positively tested cases of flu by postcode (geographical location). Urban mobility data was acquired

---

[1]https://networkedsociety.unimelb.edu.au/research/projects/active/urban-flu-forecasting

from two different sources: a) Australian Bureau of Statistics (ABS) journey to work data; b) Satellite Navigation (satnav) tracking data of around 120,000 users across Australia. All data contains spatial information, while the urban mobility data also form networks. Records are investigated in an aggregated manner, usually by time and/or spatial regions.

Why: Task Abstraction — One task is to investigate "flu spread" in a spatial and temporal manner. The time-varying semantic is essential for this task along with the geographical locations of flu reports. By first presenting the change of location of flu reports visually and then connecting these locations heuristically, a network of flu spread structures can be approximated. Subsequently the topology can be analysed based on the derived geographical networks using the geographical information (Fig. 2c: Topology Identification). To introduce spatial flow data, they first investigated the spatial correlation between urban mobility and the flu spread, which involves visually comparing two geographic networks and identifying correlation between them (Fig. 2a: Correlation Identification). They now plan to integrate urban mobility mathematically into the existing statistical prediction model.

The interviewee noted that only limited visualisations have been created for this project, and whilst they store the OD data as a matrix they have presented most of their findings as bar and line charts to show changes over time. Scalability was identified as a potential issue when visualising the urban mobility data, as visual clutter can make it difficult to read useful information.

5.4 Energy Network Flows

Put simply, a power system has a network of transmission lines connecting power stations and customers. At present electricity is not economically storable. To optimise a power system, the demand from customers needs to be met at every moment [5]. Maintaining the system effectively is becoming even more challenging with embedded micro generation, such as solar panels on customer sites. In this interview, we focused on the area of power transmission system analysis, network design and planning.

What: Data Abstraction — There are two high-level components in a power system: *elements* and *transmission lines*. *Elements* is a collective name for transformers, voltage regulators, generators, etc. Different information is captured for different types of devices. Voltage and temperature are commonly measured in real time and the geographic location is recorded. *Transmission lines* are the links connecting *elements*. Along the *transmission lines*, there are two types of flows: *active* and *reactive* power flows. Current is also measured. For emergency alert and analysis, additional datasets are used, e.g weather, to assist the reasoning process.

Why: Task Abstraction — Two main problems for energy flow analysis were described during the interview: a) the operating problem, that is instantaneously adjusting elements to increase the capacity of a portion of the power system; b) the planning problem, which focuses on how to economically relieve constraints where they may occur in the future.

For both problems, different types of constraints were modelled first based on the *elements* and *transmission lines*: a) Thermal: Current in the transmission can generate too much heat, melt the element, or make the transmission line sag down to an unsafe level. b) Voltage collapse: Voltages can become outside normal operating limits, and damage elements. c) Other transient and dynamic oscillatory type behaviours that can affect the power system.

For operating problems, operators monitor the constraints in real time, and identify cases when some constraints enter an insecure margin (Fig. 2a: Outlier Identification). To relieve the constraints, operators first propose possible solutions, summarise and compare the advantages and disadvantages of each, then choose the most suitable (Fig. 2b: Feature Summarisation and Comparison).

For planning problems, the work flow is the same as demonstrated

in Fig. 2a: Outlier Identification. Experts analyse the past records to identify where constraints occur and how frequently they occur. They then modify the model to improve the compatibility for future years. This process usually combines analysing the trends of additional *tabular* data, like demography (Fig. 2d: Trends Summarisation). For extreme system collapse cases (e.g. tornado), analysis will be conducted afterwards with external *tabular* data, like weather, to understand the reasons (Fig. 2a: Correlation Identification).

5.5 Aircraft Trajectories

Aircraft trajectory data is used in both Air Traffic Control (ATC) and Management (ATM). ATC look at real-time aircraft trajectories to maintain traffic fluidity and security, while ATM analyses large amounts of past aircraft trajectories to understand past situations and improve future procedures [7].

What: Data Abstraction — Information about each aircraft is collected during each flight, including: *unique identifier*; *3D Position*, i.e. longitude, latitude and altitude; *timestamp*, *speed, direction* and *weather*, e.g. temperature, wind speed etc. As sections of the airways have to avoid residential areas due to take off and landing noise, demography data is also usually used to inform the redesign of existing airways or design new ones.

Why: Task Abstraction — For ATC, the pilots and the air traffic controllers are actively communicating with each other on a regular basis. Some emergency cases were discussed in the interview:
- Conflict between two aircrafts; i.e. if two aircraft retain their direction of flying, accidents can occur. Automatic systems are commonly used to warn of such a situation.
- Accidents happen on the runway; air traffic controllers will ask pilots to take an alternative runway for take off or landing.
- Weather; for landing in bad weather, pilots could be instructed to put on hold, and in some cases, to land at another nearby airport.

These actions involve Outlier Identification (Fig. 2a) in both trajectories and other *tabular* datasets, like weather. In some cases, other aircraft journeys may be affected, in which case urgent response proposals need to be derived. Air traffic controllers must summarise and compare different proposals to chose the most suitable (Fig. 2b: Features Summarisation and Comparison).

In ATM, the main goal is to get a better understanding of unusual historical events and to avoid similar issues occurring in the future. Initally, aircraft or route trajectory outliers are identified. Then these outliers are correlated with other information, such as the aircraft condition or bad weather, to infer possible reasons (Fig. 2a: Correlation Identification). According to the interviewee, analysts mainly concentrate on individual trajectories, however they are also interested in the overview of a large number of trajectories, for example to investigate patterns in structure (Fig. 2c: Topology Identification).

6 SUMMARY AND RESULTS

In this section we aim to summarise the common analytical tasks identified in each of the five domains. From the interviews we identified six different **Consume** tasks and a number of **Produce** tasks. The **Produce** tasks usually involve additional modelling processes (e.g. *deriving* constraints in energy network flow analysis). For completeness we described all tasks in the individual interviews in detail in Sec. 5, however, as our focus is on how to visually represent geographically-embedded flow data for analytical tasks, we concentrate on the **Consume** tasks for the rest of our analysis. The six **Consume** tasks are presented in Tab. 2, revealing common tasks in each interview. Notably, nearly all analysts were interested in *Locate → Identify → Correlation or Outliers* tasks as well as *Explore → Identify → Topology* tasks. Although this task taxonomy gives us a method to demonstrate the common themes between our interviews, it leds to over simplification of some of the tasks. When we look into the individual tasks we recognise that there is a difference in *target* when performing tasks on geographically-embedded flow data.

Consume					
Search	Locate		Explore		Browse
Query	Identify		Summarise	Compare	Identify	Summarise
Targets	Outliers	Correlation	Features		Topology (space)	Trends (time)
Graph	Fig. 1(a)		Fig. 1(b)		Fig. 1(c)	Fig. 1(d)
Domains	Refugee Urban Flu Energy Aircraft	Refugee Urban Energy Aircraft	Energy Aircraft	Energy Aircraft	Urban Flu Aircraft	Refugee Energy

Table 2: Tasks mentioned in each of the interviews.

Inspired by existing geovisualisation taxonomy research [3, 12], we distinguish tasks into two levels: *elementary* and *synoptic*. For the *Locate → Identify → Correlation or Outliers* tasks, our domain experts are either interested in a flow between two locations (e.g. energy flow between two elements or the frequency of bicycle usage between two docking stations) or the total collection of flows into or out of a location (e.g. migrations into one country or number of bikes checked out from one docking station). These can be considered as *elementary tasks*. In addition, the domain experts were also interested in the spatial patterns *Explore → Identify → Topology* (e.g. the spatial movement of flu spread). This often involves the mental calculation of multiple flows into and out of multiple locations to see the regional pattern, or topology of the data. These can be considered as *general tasks*. We therefore identify that there are three flow-targets for analytical tasks relating to geographically-embedded flow data. These distinct differences in targets from the literature are fundamental in not only understanding the different tasks used in these domains, but also for describing the design space for geographically-embedded flow visualisation. To fill this gap, we propose three flow-targets for analysing flow data:

- *Single flow (SF)*: the magnitude of flow between two locations.
- *Total flow (TF)*: the magnitude of flow in/out of a specific location.
- *Regional flow (RF)*: the magnitude of flow between or within sets of locations. i.e. considering multiple locations as a single region.

We categorise the tasks mentioned in interviews into these three flow-targets and found these targets were mentioned in almost all of the interviews (Tab. 3). Notably the six different consume tasks (particularly *features* and *correlation*) appear in nearly all the three flow-targets. This helps demonstrates the importance of this multi-level target approach. Further evidence is shown through the specific reference to these different flow-targets in our evaluation of existing works and a new design for geographically embedded flow data [14].

7 CONCLUSION

This paper analysed five expert interviews to better understand the real-world practice of analysing geographically-embedded flow data. Our analysis gives us a clearer understanding of the common tasks that analysts in different domains perform. Data and tasks discussed in the interviews were abstracted using Munzner's what-why-how framework [11]. This was useful to structure our analysis of the interviews and demonstrate common tasks described by our interviewees, but we also recognised that a more detailed classification of the targets is required to better describe the role of flow map visualisation in the domains. Thus, we propose three flow-targets for analysing geographically-embedded flow data: *single flow*, *total flow* and *regional flow*. We will continue to explore more domains and from more perspectives (e.g. scale and interactions).

ACKNOWLEDGMENTS

Domain	Query	Target	Detail Description of Target
Single Flow			
Refugee	Identify	outliers	possible smuggling routes
Urban	Identify	outliers	dense intersections of short distance travels
Urban	Identify	outliers	missing connections between docking stations
Urban	Identify	correlation	of bicycle usage and bike lanes
Flu	Identify	correlation	of urban mobility and the flu spread
Energy	Identify	outliers	insecure transmission lines
Energy	Summarise	features	newly proposed network
Energy	Compare	features	newly proposed network
Aircraft	Identify	correlation	of trajectories and emergency cases
Aircraft	Summarise	features	newly proposed network
Aircraft	Compare	features	newly proposed network
Total Flow			
Refugee	Summarise	trends	of arrivals of a locations across time
Refugee	Identify	correlation	of arrivals of a locations and other factors
Urban	Identify	outliers	popular origins and destinations
Flu	Identify	topology	of flu spreading, and found the CBD as a hub
Energy	Identify	outliers	insecure elements
Energy	Summarise	features	newly proposed network
Energy	Compare	features	newly proposed network
Energy	Summarise	trends	of demography across time
Energy	Identify	correlation	of collapse cases and other factors
Aircraft	Identify	topology	of large amount of trajectories
Regional Flow			
Urban	Identify	topology	different patterns of using bicycles in regions
Flu	Identify	topology	of flu spreading
Flu	Identify	correlation	of urban mobility and the flu spread
Energy	Summarise	features	newly proposed network
Energy	Compare	features	newly proposed network
Aircraft	Identify	topology	of large amount of trajectories

Table 3: Tasks in the interviews categorised based on the targets of single flow (SF), total flow (TF) and regional flow (RF).

REFERENCES

[1] S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst. Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices. *IEEE Trans Vis Comput Graph*, 25(1):22–31, 2019.

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proc IEEE Symp on Inform Vis*, pp. 111–117. IEEE, 2005.

[3] N. Andrienko and G. Andrienko. Tasks. In *Exploratory Analysis of Spatial and Temporal Data*, pp. 47–161. Springer, 2006.

[4] R. Beecham, J. Wood, and A. Bowerman. Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems*, 47:5–15, 2014.

[5] Q. W. Bouts, T. Dwyer, J. Dykes, B. Speckmann, S. Goodwin, N. H. Riche, S. Carpendale, and A. Liebman. Visual Encoding of Dissimilarity Data via Topology-Preserving Map Deformation. *IEEE Trans Vis Comput Graph*, 22(9):2200–2213, 2015.

[6] M. Brehmer and T. Munzner. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Trans Vis Comput Graph*, 19(12):2376–2385, 2013.

[7] M. Cordeil, T. Dwyer, and C. Hurter. Immersive solutions for future Air Traffic Control and Management. In *the 2016 ACM Companion on Interactive Surfaces and Spaces*, pp. 25–31, 2016.

[8] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Trans Vis Comput Graph*, 18(12):2917–2926, 2012.

[9] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proc AVI workshop on BEyond time and errors*, pp. 1–5. ACM, 2006.

[10] R. Moss, A. Zarebski, P. Dawson, and J. M. McCaw. Retrospective forecasting of the 2010–2014 Melbourne influenza seasons using multiple surveillance systems. *Epidemiol Infect*, 145(1):156–169, 2017.

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

[12] R. E. Roth. An Empirically-Derived Taxonomy of Interaction Primitives for Interactive Cartography and Geovisualization. *IEEE Trans Vis Comput Graph*, 19(12):2356–2365, 2013.

[13] Y. Yang. *Visualising Geographically-Embedded Origin-Destination Flows: in 2D and immersive environments*. PhD thesis, Monash, 2018.

[14] Y. Yang, T. Dwyer, S. Goodwin, and K. Marriott. Many-to-Many Geographically-Embedded Flow Visualisation: An Evaluation. *IEEE Trans Vis Comput Graph*, 23(1):411–420, 2017.

[15] Y. Yang, T. Dwyer, B. Jenny, K. Marriott, M. Cordeil, and H. Chen. Origin-Destination Flow Maps in Immersive Environments. *IEEE Trans Vis Comput Graph*, 25(1):693–703, 2019.

[16] J. Zheng, S. Chen, X. Li, T. Chen, and D. Wang. Optimization Of Bicycle Lane, Xiamen. *Planners*, (7):100–105, 2018.