Supporting Story Synthesis: Bridging the Gap between Visual Analytics and Storytelling

Siming Chen, Jie Li, Gennady Andrienko, Natalia Andrienko,
Yun Wang, Phong H. Nguyen, and Cagatay Turkay

Abstract—Visual analytics usually deals with complex data and uses sophisticated algorithmic, visual, and interactive techniques supporting the analysis. Findings and results of the analysis often need to be communicated to an audience that lacks visual analytics expertise. This requires analysis outcomes to be presented in simpler ways than that are typically used in visual analytics systems. However, not only analytical visualizations may be too complex for target audiences but also the information that needs to be presented. Analysis results may consist of multiple components, which may involve multiple heterogeneous facets. Hence, there exists a gap on the path from obtaining analysis findings to communicating them, within which two main challenges lie: information complexity and display complexity. We address this problem by proposing a general framework where data analysis and result presentation are linked by story synthesis, in which the analyst creates and organises story contents. Unlike previous research, where analytic findings are represented by stored display states, we treat findings as data constructs. We focus on selecting, assembling and organizing findings for further presentation rather than on tracking analysis history and enabling dual (i.e., explorative and communicative) use of data displays. In story synthesis, findings are selected, assembled, and arranged in meaningful layouts that take into account the structure of information and inherent properties of its components. We propose a workflow for applying the proposed conceptual framework in designing visual analytics systems and demonstrate the generality of the approach by applying it to two diverse domains, social media and movement analysis.

Index Terms—Story Synthesis, Visual Analytics, Social Media, Spatio-temporal Data.

1 Introduction

Over the decades of the development of visual analytics techniques, researchers created sophisticated visual analytics tools for analysts to explore complex problems involving large amounts of data. However, when such tools and findings are demonstrated to those who lack visual analytics knowledge and skills, it is not unusual to get feedback such as *"fancy visuals, cool interactions, but what does this mean?"*. It is often hard for a general audience to understand composite and multifaceted analysis results represented in advanced visual interfaces that have been primarily designed to support analysis – leading to a gap between obtaining analytical results and presenting them in an accessible way. How to bridge this gap, i.e., how to proceed from data analysis to result communication, is the research problem we address in this paper.

The communication of information is an important capability of visualization. Recently, visual storytelling is receiving high attention in the information visualization community, where researchers develop authoring tools to create stories and provide visual support for storytelling [1]. The main foci of the storytelling-oriented research are the principles of story design [2] and tools that facilitate the design process [3], [4], [5]. Our focus is different and complementary to these: how to organize complex multifaceted information and prepare them for constructing a story.

Our research responds to the calls of the research community for an integrated and seamless "data analysis to storytelling pipeline" [6], [7]. So far, the research on the communication of visual analytics results has been mostly focusing on storing, annotating, and organizing analysis bookmarks or display states, possibly, after some simplification (see sect 3.2). However, when the recipient is only interested in seeing the analysis results but not in further exploration of the data, it may be more appropriate to communicate only the findings, i.e., extracted pieces of information, rather than the steps of the analysis that were used for extracting them. The visual displays that were used in the analysis process are not necessarily best suited for communicating these information pieces and for organizing multiple disjoint findings in a complete picture. Hence, analysts need methods and tools enabling them to collect findings in the process of analysis and to synthesize story contents and create understandable representations from these findings.

In response to this need, we propose a conceptual framework to inform the design and creation of visual analytics systems that support **story synthesis**. We extend the visual analytics workflow by enabling analysts to extract findings and to accumulate them in a dedicated workspace, which is then followed by a story synthesis phase where collected disjoint findings are arranged so as to represent explicitly essential relationships between them and thereby convey the full picture that has been constructed in the mind of the analyst in the result of the analysis. These arranged

- *S. Chen, G. Andrienko and N. Andrienko are with Fraunhofer Institute IAIS, Germany. S. Chen is also with University of Bonn, Germany. {siming.chen, gennady.andrienko, natalia.andrienko}@iais.fraunhofer.de*
- *J. Li is with Tianjin University, China. jie.li@tju.edu.cn.*
- *Y. Wang is with Microsoft Research Asia, China. wyawxy@gmail.com*
- *G. Andrienko, N. Andrienko, P.H. Phong, C. Turkay are with City, University of London, UK.{Phong.Nguyen.3, Cagatay Turkay.1}@city.ac.uk*

Manuscript received April 19, 2005; revised August 26, 2015.

findings make the *story content*, which can be given different appearances in designing the final story depending on the intended recipient, medium, and other criteria. Our research refers to the story content creation as an intermediate step between analysis and story design.

We propose a conceptual framework for story synthesis, in which analysis findings are organized into composite structures based on inherent structural facets (dimensions) of the information. The facets are used for two main purposes: arranging and aggregating. Possible ways to arrange and aggregate information depend on the nature and properties of the facets. We consider different types of facets, including discrete categories, linearly ordered values, time, and space. We discuss how information arrangements and aggregations are done based on these facet types and taking into account their properties. Organizing information includes creation of multi-perspective views and nested layouts, which can be used to provide detail on demand. Besides organizing information, story synthesis may involve creation of comparative views, illustration of findings by examples, and making annotations.

To demonstrate the proposed framework in action, we propose a workflow to inform visual analytics design for supporting story synthesis from analysis findings. We apply it to two diverse domains, social media and movement analysis. We discuss in more detail the use of social media data (specifically, geolocated Twitter messages) for analyzing people's reactions to significant events and creating stories about these reactions, which have multiple facets to deal with. Existing works dealing with social media data focus on either analysis [8] or storytelling [9]. Our framework shows creation and development of story content during and after the analysis and prior to storytelling.

Our research contribution can be outlined as follows:

- **Bridging the gap:** Addressing the problem of conveying analysis results to general audiences, we introduce a conceptual framework in which story synthesis is a necessary activity on the way from analysis to storytelling. We define the essential activities for story synthesis and propose a general approach in which information facets are exploited for organizing analysis findings.
- **Demonstrating the approach:** We demonstrate the use and generality of the proposed approach by applying it to complex multifaceted data taken from two diverse domains.

In the following section, we introduce our concepts and ideas by example. This is followed by an overview of the related work in Section 3, presentation of our framework in Section 4, and demonstration of its application to social media in Section 5. Expert evaluation of our approach is reported in Section 6. We briefly describe an application of the framework to a different kind of data in Section 7 and discuss the overall work in Section 8.

2 Problem Definition

2.1 A Motivating Example

To introduce our concepts, we use an example based on the IEEE VAST Challenge 2011, Mini Challenge 1 [10], requiring analysis of the circumstances of an epidemic outbreak in a fictive city Vastopolis. The data are geographically referenced microblog messages, some of which include keywords indicating disease symptoms, such as fever, chills, sweats, aches and pains, coughing, etc. The time span of the data is from April 30 to May 20, 2011. An analyst needs to find out when and where the outbreak started and how it developed. The analyst uses a visual analytics system providing multiple types of interactive visual displays and supporting database queries and data transformations.

The analyst extracts the messages containing relevant keywords from the database and, by observing their temporal distribution in a time histogram, determines *the time of the outbreak start: May 18* (**F1**). Using a map display and temporal queries, the analyst explores the spatial distributions of the messages in different days starting from May 18. She observes a *dense cluster in the city centre on May 18 and 19* (**F2**), an additional *cluster on the south-west on May 19* (**F3**), and high spatial diffusion of the messages and, simultaneously, *dense concentrations around hospitals on May 20* (**F4**).

To understand the differences between the central and south-western clusters, the analyst selects the corresponding subsets of messages by means of spatial queries and creates visualizations of the frequently occurring words. Seeing the differences between the frequent keyword sets, the analyst concludes that there were two different kinds of illness, respiratory disorders in the centre and digestive disorders on the south-west. The respiratory disorders appeared one day earlier than the digestive disorders. However, the shapes and relative spatial arrangement of the clusters suggest that the two diseases might have a common origin somewhere at a motorway bridge crossing a river.

The analyst extracts the subset of messages posted on May 17 (a day before the outbreak start) in the vicinity of the bridge, looks at the frequent keywords, and finds indications of a *truck crash, fire, and spilling of the truck cargo in the river* (**F5**). The analyst also looks at additional data concerning the weather and the river flow direction. The analyst concludes that the smoke from the fire contaminated the air, which was transmitted by the wind to the centre and caused the respiratory disorders, whereas the spilled substance contaminated the water in the river and caused the digestive disorders downstream along the river flow.

In the course of the analysis, the analyst has obtained a set of findings (labelled **F1**-**F5**), which include the outbreak start time, the spatial clusters and the times of their existence, the differing sets of frequent keywords associated with the clusters, the location and time of the truck crash, and the ways of spreading and temporal development of two diseases. As the next natural step, these findings need to be communicated to any interested audience. An often used method to achieve this is to present the findings in the form of a story. In order to construct a story, though, the analyst needs to first create the contents of the story based on the findings, synthesise the contents in structured ways, and then design a presentation that will eventually tell the intended story.

As a first required step of the content generation stage, the analyst needs to be able to represent these findings in an explicit form, extract them from the analytical environment, and collect them in some storage medium. A data structure suitable for representing a finding is shown in Fig. 1A (note

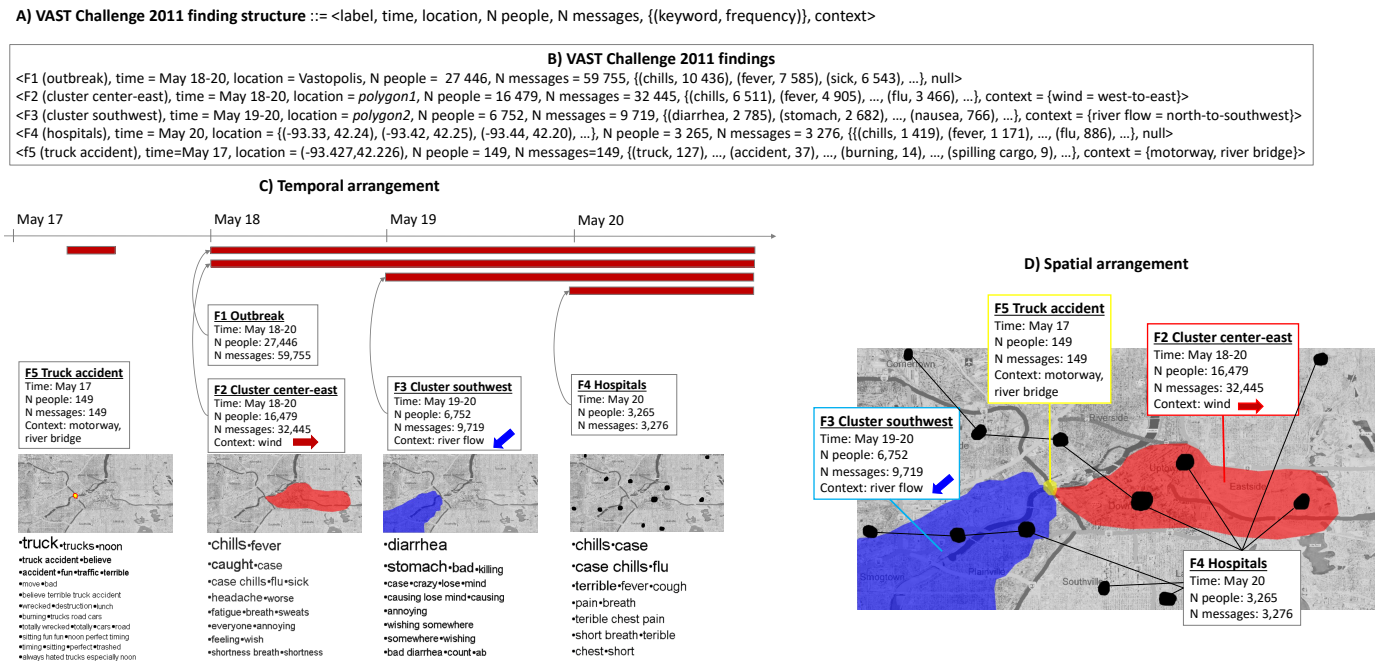

Fig. 1. The motivating example of VAST Challenge 2011 analysis outcomes: A) a structure suitable for representing the findings; B) the findings collected in a storage medium; C) the findings arranged along a time line; D) the findings arranged spatially on top of a background map.

that some fields in this structure may be undefined), and Fig. 1B shows the findings **F1-F5** represented in this format. In our framework, we refer such systematically represented findings as **story slices**.

However, just collecting multiple disjoint information pieces is insufficient for creating story contents. Story slices need to be arranged in appropriate ways revealing the relationships between the information pieces. In our example, for instance, the findings need to be organized according to the temporal and spatial relationships between them. Figures 1C and 1D show examples of such arrangements that might be created by the VAST Challenge analyst for conveying temporal and spatial relationships between the findings. Another kind of relationship the analyst may wish to reflect is the differences between the symptoms of the two diseases. For this purpose, the analyst may juxtapose the lists of the keywords corresponding to the central-eastern and south-western clusters. The analyst should be able to create and edit such arrangements in order to construct any compelling story, and they can be effectively supported in this activity by interactive visual tools.

In general, we can define *story content* as a system of story slices arranged according to relevant relationships among them, and **story synthesis** as the process of story content creation by collecting and arranging story slices.

To create a final story for communication to the intended audience, one needs to design appropriate narrative and appearance for the story content. This includes design of a suitable narrative structure, creation of suitable visual displays, selection of colours, symbols, and fonts, placement of labels, etc. The same content can be represented differently depending on the intended purpose, kind of the recipient, presentation medium, available time budget for presenting or reading the story, desired emotional impression, and other criteria. For instance, the outcomes from our example above may need to be reported to high-level health care managers, or presented to the general public in a newspaper article or on TV. For each of these presentations, the same story can be given a distinct narrative and appearance, which is structured by the creative ideas of presentation designers. The design of the story appearance is a different kind of activity than the story content creation. The focus of our work is the story synthesis activity, which creates story contents and structure and thus serves as a bridge between analysis and story design.

2.2 General Description

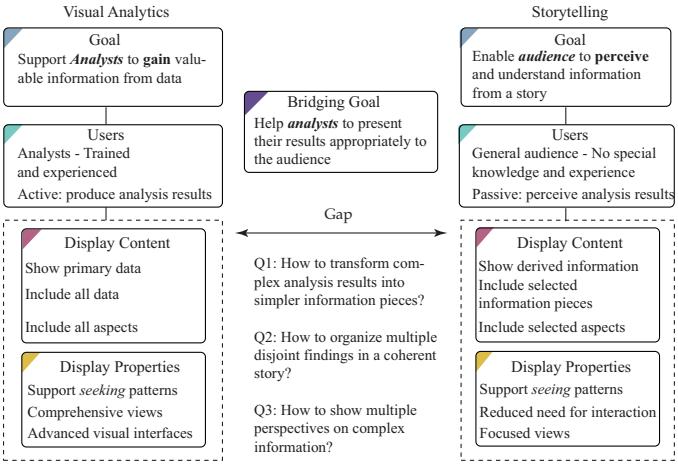

Fig. 2. Conceptualizing the gap between visual analytics and storytelling. Based on comparing the respective goals, users, display content, and display properties, the research is driven by asking the three synoptic questions.

The gap between visual analytics and communication of analysis results can be conceptualized as shown schematically in Fig. 2. Visual analytics systems are meant for

solving complex and ill-defined problems with the use of interactive visual displays and computational analysis techniques [11]. Visual analytics requires a possibility to see all aspects of complex data and explore their interrelationships, which is usually supported by multiple coordinated views and sophisticated interaction techniques. On the contrary, storytelling is meant to convey only interesting and/or important information extracted through the analysis, and this information should be presented in a simple and easily understandable way [2], [12].

In short, visual analytics and storytelling essentially differ in their purposes, target users, kind of information dealt with, and methods of presenting the information and interacting with it. Therefore, to support telling stories of visual analytics findings, there should be an intermediate step between analysis and storytelling, in which the analyst assembles and organizes information pieces to be communicated. We refer to this step as *story synthesis* (Fig. 3). Its purpose is to prepare analysis results to communication: select, assemble, summarize, transform to a suitable representation, arrange, and annotate. In a recent position article, Lee et al. [6] propose a multi-stage pipeline starting from exploring the data and curating findings to eventually forming visually shared stories. They identify "*Help People Make a Compelling Story*" and "*Make It Easier to Tell a Story*" as two high level challenges for further research and mention how *current systems fall short in collecting, organizing and structuring excerpts in producing a visual story*, pointing exactly at the gap that we aimed to address in our work.

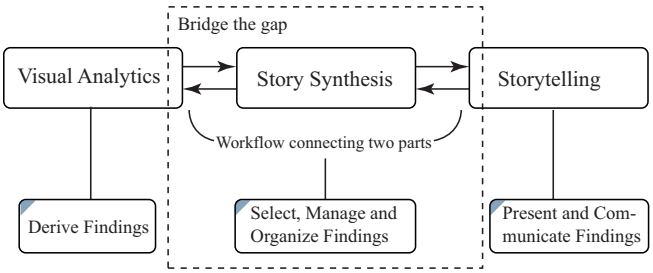

Fig. 3. Proposed framework for bridging the gap.

Unlike purely computational analysis methods, which typically produce a single result such as a model, interactive visual analysis often yields multiple findings obtained through different interactions, as in the example in Section 2.1. The mind of the analyst may contain the whole picture (i.e., a mental model of the analysis subject [13]) in which these findings are associated. However, in the external representation, the findings may be disjoint, as in Fig. 1B. To enable presenting the full picture in a story, the analyst needs to represent explicitly essential relationships between the findings. This can be achieved by arranging the finding according to these relationships, as in Fig. 1C and D.

We refer to pieces of information that need to be assembled and arranged as *story slices*. Having suitably organized story slices, the analyst or even another person can construct a story that communicates the analyst's mental model to recipients. In existing works, stories are composed from annotated screen shots or states of analytical displays [14]. While this kind of material can be useful to a story designer, it may not be well suited for communication because it inherits the complexity of visual analytics displays and also contains extra information that is not essential for conveying the story. The goal of the story synthesis stage (Fig. 3) is to transform the raw materials of analysis findings into story slices arranged according to important relationships between them. These arrangements can be used for composing the whole final story as defined by Lee et al. [6].

Summarizing our arguments, we state that creation of a story involves two kinds of activities: preparation of the content, referred to as story synthesis (Fig. 3, middle), and design of the presentation of the prepared content, referred to as storytelling (Fig. 3, right). The presentation design activity is supported by the principles and approaches developed in the research on storytelling [1], [2]. For example, choosing a suitable order for presenting story slices can be based on the work by Hullman et al. [15]. Our work mostly focuses on the content creation activities, which include generating and organizing story slices.

3 Related Work

3.1 Visual Storytelling

We investigate the works on using visual storytelling for presenting data and findings with a primary focus on ideas and approaches related to creation of story contents.

As one of the early papers that stress the importance of storytelling in visualization, Gershon and Page [12] emphasize the importance of choosing an optimal amount of information to deliver a message, and discuss how data visualizations can be arranged to generate story-like representations. Kosara and Mackinlay [16] discuss future research directions in visual storytelling with a focus on the narrative structure of the stories. Ma et al. [7] point at the limitation that visualizations used in a visual story are often created after the fact and call for an integrated process. These works highlight the need to link visual analytics and storytelling.

In their influential paper [2], Segel and Heer present a framework for the design of narrative visualizations and identify techniques in data-driven storytelling research. McKenna et al [17] define and investigate a design space for the narrative structure and introduce a number of factors that can affect the experience of the audience. This study, as well as the discussion of the visualization rhetoric techniques by Hullman et al. [18], provide useful ideas for organizing story contents. Ren et al. [19] focus on the role of annotations in data-driven storytelling and presents a design space for chart annotations. Brehmer et al. [20] present a design space for storytelling focusing on the temporal aspect alone. Bach et al. [21] introduce the concept of "data comics" and explore how the established elements and rules of the comics genre can be employed for storytelling with data.

Apart from stories, dashboards is another widely used medium for presenting data to users [22], [23]. Thus, Mckenna et al. designed a dashboard visualization for cyber security analysis [24]. Commercial tools, such as Tableau [25], Power-BI [26], Airtable [27], etc., can be used for creating shareable and customized visualizations to communicate data-based stories.

These works lay the foundations for constructing visual narratives regarding both story structure and presentation

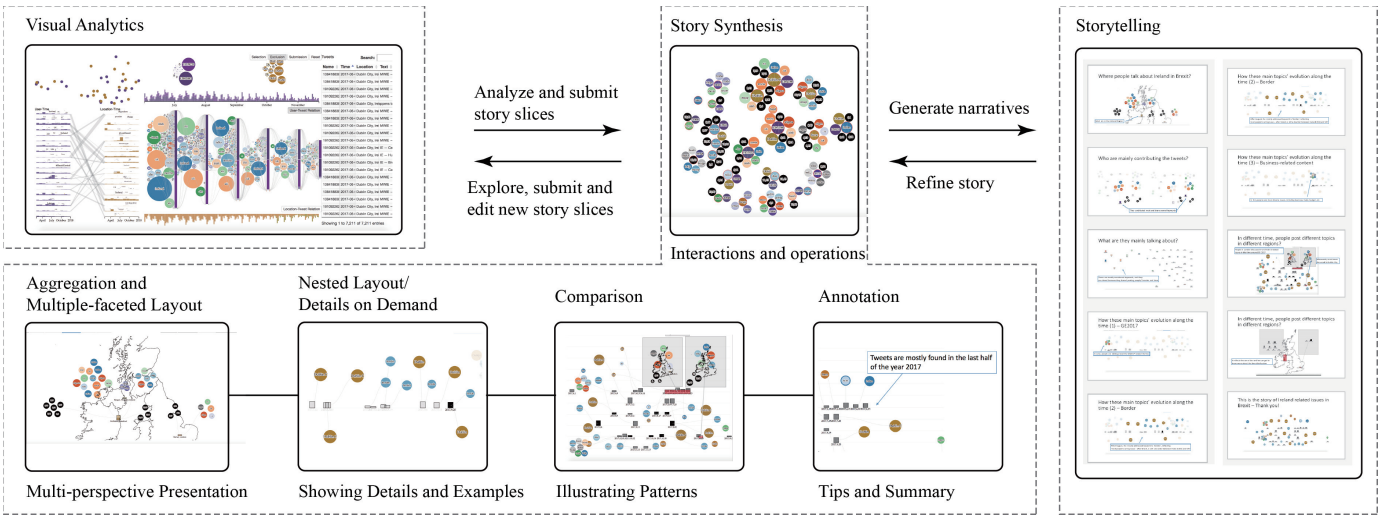

Fig. 4. The framework for integrating visual analytics, story synthesis, and storytelling is illustrated by an example of social media data.

design. Relevant to our research are the ideas concerning the organization of story contents.

3.2 Storytelling in Visual Analytics

Storytelling has also attracted significant interest in visual analytics literature. GeoTime [28] is one of the earliest visual analytics systems that support storytelling. The story is composed of texts and hyperlinks connecting to bookmarked visualizations, which may include graphical annotations. The visualizations are saved so that they can be restored in exactly the same appearance as when they were captured, allowing further exploration along with reading the story. HTVA [14] allows constructing a story by arranging thumbnails of visualization bookmarks based on their time and/or space references. Bookmarks can also be arranged manually and connected by explicit links. In these ways, story creators can convey temporal, spatial, and causal relationships between information pieces contained in visualizations. Walker et al. [29] discuss possibilities for applying storyboarding principles from film industry. For example, the *viewpoint* principle translated to presenting data states that a storyboard should provide different perspectives on the data, which may include data and display transformations.

Besides manual bookmarking, automatic capture of analytic provenance [30] is also commonly used to construct stories. SenseMap [31] automatically captures and visualizes users' actions, enabling the users to curate relevant information pieces, organize them, and communicate the analytical findings. Gratzl et al. [32] introduce an authoring tool that constructs a story based on provenance capture, adds text and drawing annotation, and plays back the story as it was originally performed. TimeLineCurator [3] focuses on the temporal aspect, facilitating the timeline creation process for journalists by automatically extracting event data from unstructured text documents and encoding them along a visual timeline. System KnowledgePearls [33] automatically generates and stores structured data describing visualization states and uses these data for finding visualizations corresponding to users' queries. Such functions could be potentially useful for story synthesis, but the authors do not consider this application for their ideas.

Visual analytics systems often include multiple interactive views, which may be too complex for communication of analytical findings to people that are not visual analytics experts. Besides, a common way to show a visualization bookmark is using a thumbnail of the entire system [28], which might be too small for a complex graphic to be understood. Therefore, display simplification and/or providing interpretation aids are required. GAV [34] provides a story mode that simplifies the interface and excludes advanced features, enabling the viewers to focus on the essential elements. A visual analytics system may allow the analyst to choose a single visual component from the entire system [29] or select a portion of the interface that emphasizes the main message [35]. HTVA [14] allows adding a simple overlay describing the visual encoding and data filtering. SenseMap [31] helps construct a story with multi-level semantics that can be flexibly presented to the audience with different backgrounds and needs.

The difference of our approach from the previous work is that we consider construction of stories not from complete visualizations [28] or their components [29] or analysis bookmarks [31] but from structured data representing findings extracted during the process of analysis. The visual representation for these data is chosen according to their structure. It may differ from the original representation used in the process of analysis; moreover, these structures are analytical artefacts that are not present in visualizations of original data, as in the example in (Fig. 1A,B). As derived information pieces, analysis findings are similar to explicitly defined concepts, which can be created in the course of analysis and organized in concept maps [36], [37]. However, a concept map may not be the most appropriate representation form for findings having complex structure and linked by different kinds of relationships.

3.3 Visual Analytics Applications

Since our illustrating examples relate to the social media and movement analysis domains, we briefly discuss the works dealing with data from these domains with a focus on pertinent information facets and storytelling.

The application of visual analytics to social media is surveyed in [38]. Dou et al. [39] apply text processing to

identify meaningful events and characterize them in terms of four facets, called "4 Ws: who, what, when, and where", which can be then interactively explored. Xu et al. [40] and Sun et al. [41] propose a river-based metaphor to analyse evolution of topics over time. WeiboEvents analyse reposting structure of a single message [42]. For multiple messages, Chen et al. propose a series of map-based visual metaphors to visualize ego-centric information diffusion and event evolution [8], [43]. Whisper visualizes information diffusion in space and time using a flower metaphor [44].

In data journalism, there are examples of using visualization of social media for storytelling [9]. For broader application areas, commercial dashboard tools [45], [46] and storylines [47] are often used. However, these approaches have limited or no support for organizing multi-faceted visual analytics results.

In visual analytics of movement [48], data are considered from different perspectives: as trajectories, spatial events, local temporal variations of movement characteristics, or evolution of overall spatial situation [49]. The perspective taken determines the relevant facets, which may include origins and destinations of trips [50], taken paths [51], locations and times of specific events [52], etc. Most approaches in this domain are designed for experts. Due to complexity and volume of movement data and patterns, visualization techniques and analysis results require efforts for understanding, especially by general audience. For example, the case study on aviation data analysis [51] revealed a story that could be interesting to broad audience if properly presented; however, approaches and tools for converting such findings to stories are missing. Observing these gaps in visual analytics research and applications motivates us to propose a framework to bridge them.

4 A Story Synthesis Framework

The proposed framework targets *story synthesis*, which is seen as an intermediate step between analysing data and presenting analysis results (Fig. 3). We discuss the activities involved in story synthesis and its links to the preceding and following steps in the overall workflow. We then present the general idea of organizing information selected for presentation according to its facets, or dimensions, and propose a workflow for applying the conceptual framework for designing visual analytics systems supporting the transition from analysis to result communication.

4.1 Story Slices

A *story slice*, as defined and exemplified in Section 2, is a structured representation of a finding or a combination of findings or, generally, an information construction obtained from original data in the course of analysis. Creation of story slices should be possible during the process of data analysis whenever the analyst observes or derives something potentially interesting. The analyst needs a workspace where to put story slices while analysing the data. The specific way of representing story slices within the workspace depends on the types of components present in the data and the structure of findings that can be extracted by means of the analytical tools. As the data may have complex structure, findings may involve several heterogeneous aspects, or facets. For example, the findings in Fig. 1A,B involve temporal and spatial references, keywords with frequencies, and attributes with numeric and textual values. Moreover, findings may be linked by various relationships, the kinds of these relationships being dependent on the kinds of facets present in the findings. These heterogeneous facets and relationships need to be reflected in a story in a sufficiently simple and easily understandable way. Building on existing works on exploring relationships among multiple facets [53], [54], we address this challenge in 1) representation and 2) organization of story slices.

4.2 Activities in Story Synthesis

We define *story synthesis* as the process of constructing story contents that includes, first, creating and assembling story slices, and, second, arranging the collected items into annotated layouts conveying relationships between the items. Story synthesis does not include the design of the final appearance of a story, but it involves choosing suitable levels of detail and perspectives for presenting the information and arranging the material in the display space according to the chosen perspectives. In brief, story synthesis defines the story content and structure but not the appearance. From this general understanding of the purpose and expected result of story synthesis, we infer more specific activities that may be undertaken for achieving this kind of result:

- **Aggregate and summarize**: Put several story slices together and represent them as a single object with properties derived by summarizing properties of the original slices. Aggregation is used for achieving the desired level of detail and also for simplification, which is important for making presented information easier to perceive and understand.
- **Embed details**: Enable recipients to see detailed information for presented aggregates on demand [2], [55].
- **Arrange**: Create a meaningful layout in which story slices and/or aggregates are positioned according to essential relationships, e.g. temporal order, relative locations in space, or semantic similarity [53].
- **Show facets**: Organize information so as to exhibit its inherent facets. It may be impossible to show all facets simultaneously in a sufficiently simple manner. A valid approach is to create several arrangements based on different facets and thereby present information from different complementary perspectives [54].
- **Annotate**: Create verbal and graphical annotations to explain the content and/or direct recipients' attention to the most important or the most interesting parts [19].

4.3 Facet-based Arrangement of Story Contents

We have earlier made several notes concerning information components, or facets:

- Facets need to be presented to story recipients to enable proper understanding of information.
- Heterogeneous facets may be hard to present simultaneously while keeping the display simple and easy to understand.

- Heterogeneous facets may be represented in complementary views providing different perspectives on the information.
- Information facets can be used for meaningful arrangement of story slices and for aggregation.

Since the main task of story synthesis is organizing multiple disjoint information pieces so that relationships between them could be perceived, we say that an arrangement is *meaningful* when it exhibits certain relationships between information pieces, such as spatial, temporal, or feature-based similarity relationships.

Arrangement and aggregation according to information facets accounts for and makes use of the inherent properties of the facets. A facet can be considered as a set (domain) elements of which may occur in data as values in data fields. Facet properties are defined based on relationships between their elements, particularly, the existence of generic relationships of ordering and/or distance. Elements may also be linked by domain-specific relationships. On this basis, the following types of facets can be distinguished:

- **Discrete entities or categories with no relationships**: There are neither inherent relationships, such as ordering or distances, between the items nor explicitly specified relationships of other kinds. Examples are people, organizations, documents, words, topics, etc.
- **Discrete entities with domain-specific relationships**: Such relationships can be specified explicitly in data or derived through analysis. Examples are scientific papers with citation relationships, paper authors linked by co-authorship relationships, social media posts with 'reply' and 'repost' relationships, etc. Such relationships are often represented by graph structures.
- **Linearly ordered elements**: There are ordering relationships between any two items. Examples are attributes expressing ranks or evaluation grades.
- **Linearly ordered elements with distances**: Apart from ordering, there are distances between items. Examples are numeric attributes.
- **Time**: Elements of time (time moments) have ordering and distance relationships between them, but time also involves cycles: daily, weekly, annual, and/or domain-specific cycles, as in astronomy or climate research. Hence, additional relationships exist between the elements based on their positions in the cycles.
- **Space**: Elements of a two- or three-dimensional space, i.e., spatial locations, have no natural ordering but have distance relationships among them. One-dimensional spaces, as in analysing traffic along a road, have both distance and ordering relationships between the locations. Partial ordering relationships exist in hierarchical spaces, such as a river with its tributaries.

The existence of inherent relationships between facet elements suggests a way to arrange story slices in a layout that is based on this facet. The principle is that **the relative positions of the story slices should be consistent with the relationships between the corresponding facet elements**, as detailed below:

- **Domain-specific relationships** can be represented using a node-link structure, where story slices are put in the nodes, and the links represent the relationships.
- **Linear ordering** is represented by arranging story slices in a linear sequence. In case of two ordered facets, slices can be arranged on a 2D plane where the dimensions correspond to the facets and the horizontal and vertical positions correspond to the order based on the respective facets.
- **Linear ordering and distances**: One may use a linear or two-dimensional (in case of two facets) layout as described above while making the distances between the slices proportional to the distances between the corresponding facet elements.
- **Time**: If temporal cycles are not relevant to the story content, temporal relationships can be represented using a linear layout as described above (e.g., Fig. 1C, Fig. 9). A singe cycle can be represented using a radial layout where the story slices are positioned along a circle according to the positions of the respective facet elements in the cycle. Two cycles (e.g., daily and weekly) can be represented using a 2D layout, as in a case of two ordered facets; each dimension corresponds to one cycle (e.g., Fig 8a).
- **Space**: To represent spatial relationships, story slices are arranged on a plane according to the relative spatial positions of the respective facet elements, so that the distances on the plane are proportional to the spatial distances. When the space is geographic, it is appropriate to use a geographic map as a background for the layout (Fig. 1D), to show the spatial context. If the distribution of slices in the space is not even, it may be useful to replace the precise positions and background map by map-like representations that preserve partial ordering and topology, such as cartograms [56] or spatial tree maps [57].
- **No inherent relationships**: A meaningful layout can be constructed based on similarity relationships established through defining a suitable similarity measure, or distance function, which is chosen or designed according to the type of information [58], [59], [60], [61]. Having similarity relationships expressed in numeric measures, the analyst can apply projection techniques, such as multidimensional scaling, for arranging story slices.

Aggregation of story slices needs to account for distance relationships regarding one or more facets, i.e., slices that are put together should refer to close positions in space and/or in time and/or have close values of numeric attributes. These kinds of aggregation can be referred to as spatial, temporal, and attribute-based aggregation. Temporal aggregation can also be done based on close positions in temporal cycles, e.g., aggregate story slices referring to morning hours of different days or to Sundays of each week.

Aggregation that does not use inherent distance relationships should be based on meaningful grouping of story slices. A possible approach is clustering according to values of multiple attributes; however, such artificially produced groups need to be given explanatory labels to make them understandable to recipients.

4.4 Links of Story Synthesis to Other Steps

The story synthesis step has two-way links to the steps of analysis, in which story slices are derived from data,

and storytelling, in which narratives are designed from the content synthesized (Fig. 3). The forward links mean that results of the analysis are used in story synthesis, the results of which, in turn, are used in storytelling. The backward link from story synthesis to analysis means that the steps can be performed iteratively. The analyst is not necessarily supposed to complete the analysis before starting story synthesis. The analyst may switch at any time from analysis to story synthesis activities to organize the story slices extracted by this moment, critically review them, and, possibly, remove some as insufficiently interesting. By manipulating the workspace content, the analyst may also get an idea how to proceed in the analysis. For example, if many story slices are very similar, the analyst may decide to seek something different. The backward link from storytelling to story synthesis means that in the process of designing the narrative the analyst may see a need in further aggregation or re-aggregation, or adding details, or creating views for side-by-side comparison, etc.

4.5 Application of the Framework

Our generic conceptual framework defines story synthesis as the process of generating story content and structure. The story content consists of story slices, which are multi-faceted information pieces (Section 4.1). The story structure is a particular organization of the story slices, including aggregates and arrangements. The framework defines the activities involved in story synthesis (Section 4.2) and proposes guidelines for facet-based arrangement of story slices (Section 4.3). This framework can be used in designing visual analytics systems that provide support for story synthesis. We propose the following workflow for applying the framework:

- **Step 1: Define the types and structures of story slices**. Depending on the data and the goals of analysis, the designer envisages the facts or patterns that can be discovered and used as story slices. The analyst defines the structure of these potential story slices, as in Fig. 1A, i.e., the inherent facets and relationships between them.
- **Step 2: Design a representation for a story slice**. According to the story slice structure, the designer chooses a fitting data structure, such as a graph or a vector, to represent extracted story slices internally in the system, as in Fig. 1B. The designer also chooses or devises a suitable visual representation of a story slice, which will be used by the analyst (e.g., as the representation of the findings F1..F5 in Fig. 1C). This representation may not necessarily be used in the final story; thus, at the storytelling stage, the designer may decide to replace it by something else that can look more appealing to the recipient.
- **Step 3: Define story synthesis support functions**. The designer defines the system functions required for constructing story slices or extracting story slices from visual displays, managing the items obtained, and supporting the story synthesis activities listed in Section 4.2.
- **Step 4: Design the visual analytics system**. This includes the design of tools enabling discovery of facts or patterns that may become story slices. Besides, the designer defines the way of supporting the transitions between the analysis and story synthesis (Section 4.4).

To demonstrate the application of the framework, we use two examples. The first example, which refers to social media analysis, is presented in more detail than the second example referring to flight analysis in aviation domain.

5 Applying the Framework to Social Media Analytics

In this example, we apply the general conceptual framework to analysing geolocated social media data and constructing stories about users' reactions to significant events happening in the world, e.g., in politics, sports, or culture (Fig. 4). Important facets of such data are the user, location, time, and message text. It is usual to represent texts as collections of keywords [39], [43]. Hence, we define a **story slice** as a tuple $< User, Location, Time, Keywords >$. Specifically, we use Twitter posts related to the *Brexit*, i.e., the plan of the United Kingdom to leave the European Union (see https://en.wikipedia.org/wiki/Brexit). We crawled 380,000 geo-tagged Twitter messages containing "Brexit" in either message texts or hashtags and posted on the territory of the UK and Ireland in 2017. The main goal of the analysis was to reveal differences in people's behaviours (i.e., reactions to Brexit-related events and their evolution) across regions and time periods. The story of Brexit is interesting to sociologists, journalists, and the general public, which motivates the use of this example for demonstrating how story synthesis can be supported in a visual analytics system.

5.1 System Functions

We translate the general requirements for supporting the story synthesis activities (Section 4.2) into a list of system functions implementing these requirements.

- **Aggregation of story slices**. We support semi-automatic aggregation based on space, time, user, and keywords.
- **Facet-based layouts**. We enable layouts according to time, location, user, and keywords.
- **Embedded details**. We enable construction of nested layouts for providing details of spatial, temporal and user distribution as well as raw Tweet messages.
- **Comparative views**. We enable analysts to select groups of findings and arrange them for comparison.
- **Iterative process**. We implement the two-way links as shown in Fig. 3 and explained in Section 4.4.
- **Annotation**. We support analysts in attaching text annotations to items (slices and aggregates), item groups, and regions in layouts.

5.2 Analysis Phase with Creation of Story Slices

Our visual analytics approach integrates topic modeling and interactive visualization focusing on the spatial, temporal, user and keyword perspectives (Fig. 6). Analysts can iteratively explore the data, extract findings, and and submit them to a dedicated workspace for story synthesis (Fig. 5). The analytical functions of the system need to be described to enable understanding of the kinds of findings that can be obtained.

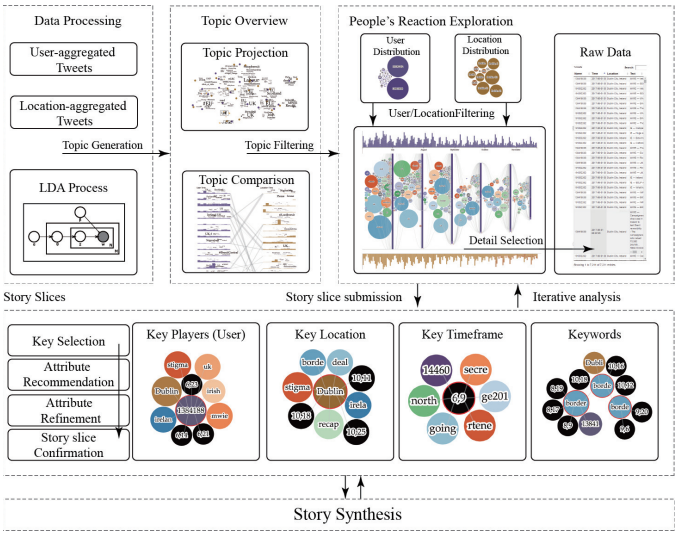

Fig. 5. Visual analytics phase. After topic modeling, analysts explore the collection of messages from four perspectives and extract story slices.

5.2.1 Topic Modeling

Topic modeling is a class of statistical techniques that discover themes from a set of texts. A topic is represented by a set of significant words that are considered as related because of the close repeating occurrences in the texts. The words are associated with weights expressing their significance in regard to the topic. We use LDA [62], which is a representative technique of topic modeling. The input of the LDA process is a set of pre-processed text documents, where each document is represented by a bag of words with their frequencies. The output is a set of extracted topics, i.e., combinations of weighted keywords, and the probabilities of the topics for each document. Topic modeling methods produce good results when documents are sufficiently long. However, social media messages are usually very short texts including only a few meaningful words; hence, it is problematic to model topics. To solve this problem, we create artificial documents by joining multiple texts based on assumptions of potential relatedness (Fig. 5-Data Processing). We use two bases for text joining (aggregation):

- User + time: We aggregate messages of the same person posted at close times, assuming that a person keeps a certain kind of reaction for some time.
- Location + time: We aggregate messages posted in neighboring locations at a close time, assuming that people living in the neighborhood may share some similarity.

We aggregate the tweets using the two approaches and generate the same number of topics based on each aggregation through LDA, following the approach proposed in [63].

5.2.2 Topic Visualization

To gain an overall understanding of the extracted topics (Fig. 5-Topic Overview), we propose two visualizations, topic projection (Fig. 6a) and comparison (Fig. 6b). The visualizations support answering the following questions: what are the topics (i.e., their semantics), how they are related to each other, and how the tweets referring to the topics are distributed over time.

For topic projection (Fig. 6a), we apply the t-SNE algorithm [64] based on the keyword weights. Purple and yellow colors in Fig. 6 correspond to the topics based on user+time and location+time aggregations, respectively. The use of colors is consistent throughout all visualizations. The projection display includes the topics represented by colored dots and the representative keywords of groups of topics that are close in the projection. The font sizes encode the frequencies.

The topic comparison view (Fig. 6b) enables comparison of topics derived in two ways, user-based and location-based. For each topic, the display contains a time histogram representing the counts of unique users or locations, respectively, by time intervals. User-based and location-based topics are linked by lines if the degree of their semantic similarity (in terms of keyword weights) exceeds a given threshold, e.g., top 20. The analyst can explore semantic and temporal similarities in user- and location-based topics.

5.2.3 Exploring People's Reactions

Analysts can select topics from topic visualization. In response, the spatial, temporal and semantic details of the selected topics are visualized (Fig. 5-People's Reaction Exploration). Analysts can also select a time period.

LDA gives the weight (probability) of each topic for each aggregated text. In response to the topic selection, the texts with high weights of the selected topics are chosen, and the corresponding sets of users and/or locations are extracted and represented in bubble charts (Fig. 6c), where each circle represents a user or a location with time while the size is proportional to the number of tweets posted by this user or at this location. Analysts can click on a circle to select details.

Fig. 6d presents a temporal display that shows how tweet posting related to the selected topics evolves over time. There are three main parts, user-tweet histogram (top), location-tweet histogram (bottom), and keyword flow (middle), sharing a common horizontal axis representing time. In the histograms, the bar lengths represent the tweet counts; each bar corresponds to one day. In the keyword flow view, tweets are aggregated by longer time intervals, such as one month, which gives more display space for presenting the keywords that occurred frequently in these intervals. Each circle corresponds to one keyword, and the size is proportional to the frequency. On the right of each keyword group, there is a bar representing the distribution of the tweets in the respective time interval over the users (in purple) or locations (in yellow), depending on the analyst's choice. The bar is divided into segments proportionally to the amounts of tweets from individual users or locations. There are light curved lines connecting bar segments to keywords to show the main themes of the users or locations. Analysts can observe temporal trends regarding the keywords in the discussions and topic-related tweeting activities of the users and locations.

Analysts can apply selection, exclusion, and filtering operations to four facets of the information dealt with: users, keywords, time, and locations. For example, after analysts remove two dominating users (Fig. 6c) and exclude dominating keywords such as *"Ireland, EU, UK", etc*, more specific keywords, such as *"business, border"*, pop up. Generally, analysts can exclude what they already know and get more

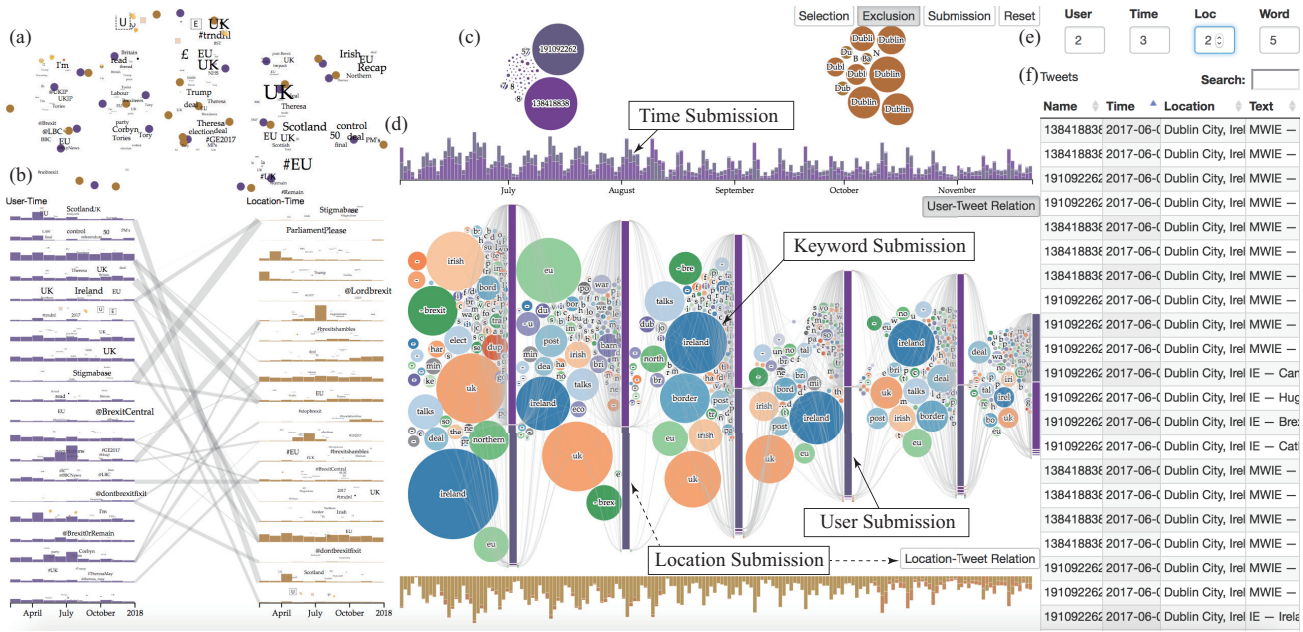

Fig. 6. Visual analytics interface of social media, including (a) topic projection view, (b) topic comparison view, (c) user/location distribution view, (d) temporal view, (e) story slice generation parameters, and (f) raw data table.

details for the remaining data. This can be done for any of the facets.

5.2.4 Generation and Refinement of Story Slices

In the course of the analysis, analysts define their findings by interacting with the visualizations and submit the findings findings to a dedicated workspace (Fig. 5-Story Slice Submission). Story slices are represented as graphs, which are constructed automatically based on analysts' interactions. Each story slice has one key and one or more attributes attached to it. The key is based on one of the information facets, depending on the analysis focus, i.e., it is a specific user, location, time interval, or keyword (Fig. 7). All information related to the key is represented as attributes, e.g., for a user, the attributes include the keywords that occurred in the user's tweets, the locations from where the tweets were posted, and the time when this happened, with the corresponding tweet counts. Story slices are generated by clicking on display elements (Fig. 6). The type of element that is clicked on defines the key of the slice (e.g., it is a user in Fig. 7a), and the system automatically extracts the related data and generates story slice attributes accordingly (Fig. 7a-b). Analysts can interactively edit slices by removing unnecessary attributes (Fig. 7b-e).

5.3 Story Synthesis Phase

5.3.1 Facet-based Layouts and Aggregation

According to the general framework, our system enables analysts to arrange story slices in different layouts based on information facets (Fig. 8). This also includes aggregation of story slices as a means of information and display simplification. Aggregation takes place when several story slices receive identical or very close positions in a layout and have a common key node. Such a group is merged in a single graph with the same key node, and the attributes are merged accordingly.

The facet chosen for a layout may correspond to either the key or attributes of the story slice. Nodes corresponding to the layout's facet receive fixed positions in the layout, and the remaining nodes are positioned in the vicinity while avoiding overlaps. Fixed nodes are shown as rectangles and associated nodes as circles. When fixing the key node, all attributes nodes are associated (Fig. 7d). When the layout's facet corresponds to two or more attributes of a story slice, the latter receives several positions in the layout. These positions are connected by lines (Fig. 7e). The examples shown in Fig. 7-d and -f correspond to the layouts presented in Fig. 8-c and -a.

For creating a time-based layout, the analyst can choose between the linear (Fig. 9) and 2D arrangement (Fig. 8a). The latter reflects the cyclic structure of time. In our use case, the rows correspond to months and the columns to days. A geographic map is automatically included as a background in location-based layouts (Fig. 8b). User-based layout arranges users according to the amounts of tweets they produced so that the most active users receive the highest prominence (Fig. 8c). In the keyword-based layout, the keywords are arranged in a 2D space by means of multi-dimensional scaling (MDS). The distance corresponds to the number of shared attributes between the keywords (Fig. 8d). Then a force-directed layout algorithm is applied to remove duplication and overlaps. In any automatically produced layout, the analyst can interactively customize the arrangement by moving the nodes.

As simplicity is important for information communication, arrangements of story slices are simplified in the following ways:

- Aggregate story slices based on one or more facets.
- Show/hide attributes. For example, if analysts primarily want to show the keywords distribution over time,

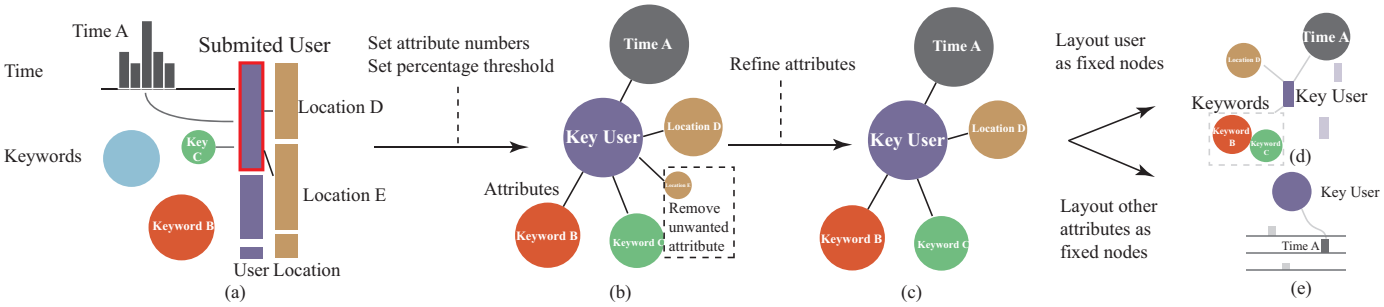

Fig. 7. An example of creating a story slice. The analyst selects an interesting user. The user's tweets were mostly posted at time A with keywords B and C from locations D and E (a). A story slice is created under parameters of the maximal number of attributes and the minimal percentage of tweets (b). It can be refined by removing unwanted attributes (c). The visual representation of the story slice depends on the chosen layout (d, e).

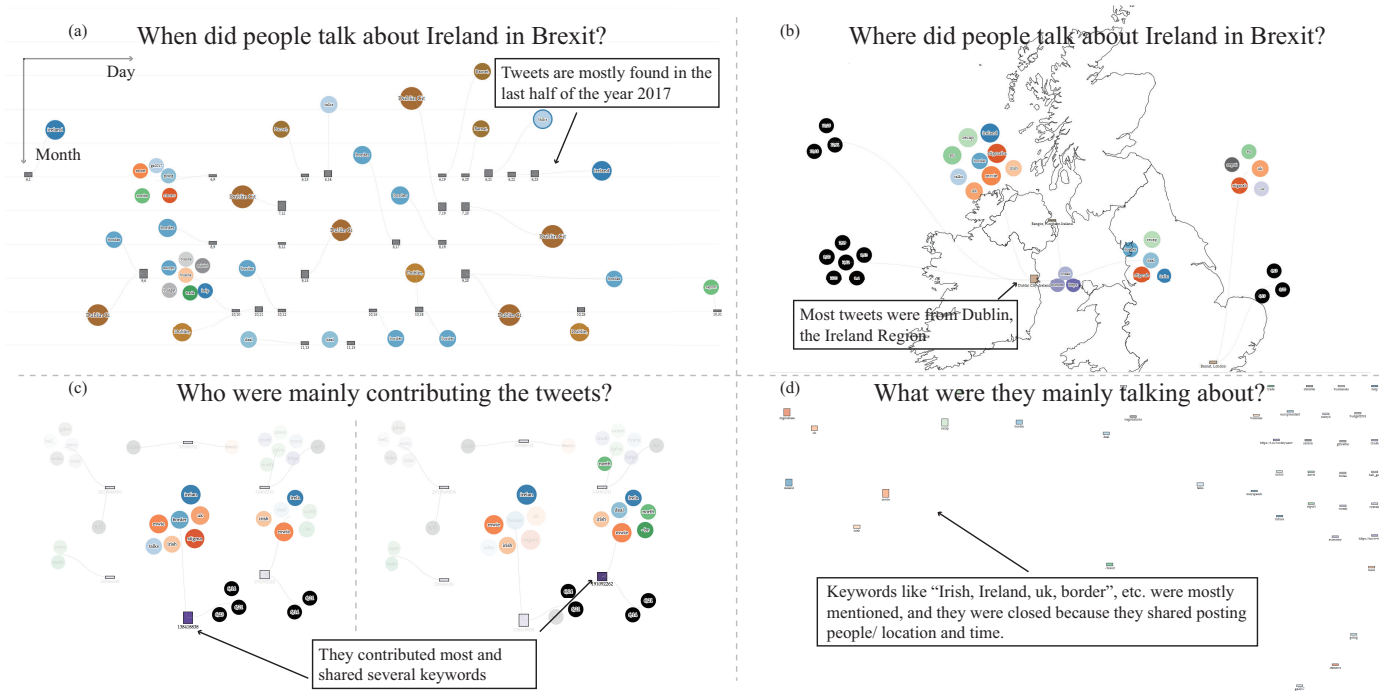

Fig. 8. Layout methods to organize story slices according to different perspectives: (a) time cycles, (b) locations, (c) users, and (d) keywords.

the display of users and locations can be switched off.

- Shift labels when two nodes are near to avoid overlapping (Fig. 9).
- Apply constrained force-directed layout technique to node positions to avoid overlapping of nodes.

5.3.2 *Details on Demand and Nested Layouts*

The principle 'details on demand' is applied in showing relationships among nodes. Explicit visual representation of all relationships would greatly increase the display complexity. To avoid this, only links from keys to attributes are shown explicitly. Other relationships can be seen through interaction with the display. When the viewer hovers on a node, this node is highlighted together with the nodes that are directly or indirectly connected and the nodes with the same key, while the remaining nodes are muted (Fig. 9). In the course of story synthesis, 'details on demand' also means that analysts can retrieve the original tweet messages by clicking on a node. Seeing the tweets, the analyst can generate explanatory annotations for inclusion in the story. Such annotations can provide more details to recipients.

Nested layouts (Fig. 10) are used for including more details for data subsets and showing them from different perspectives. When the analyst selects a subset of nodes by brushing, the system generates a sub-canvas for this subset, where the analyst can arrange the information in a specific way, which may differ from the overall layout. For example, the analyst may want to show for a specific time range where the corresponding tweets come from. The analyst selects the time range in a temporal layout. A nested canvas is created, and the analyst can position and resize it. Then the analyst chooses the location-based layout for the nested canvas. The system automatically extracts the location attributes of the selected nodes and arranges them as keys in the nested canvas. Furthermore, the analyst can highlight keyword attributes in the nested location-based layout (Fig. 10a). This will allow the recipients to see the keywords distribution over locations in the chosen time range. Another example is creation of nested views for showing the temporal distribution of tweets for selected locations (Fig. 10b). Such nested layouts can be used to combine multiple facets in one view. Creation of two or more nested

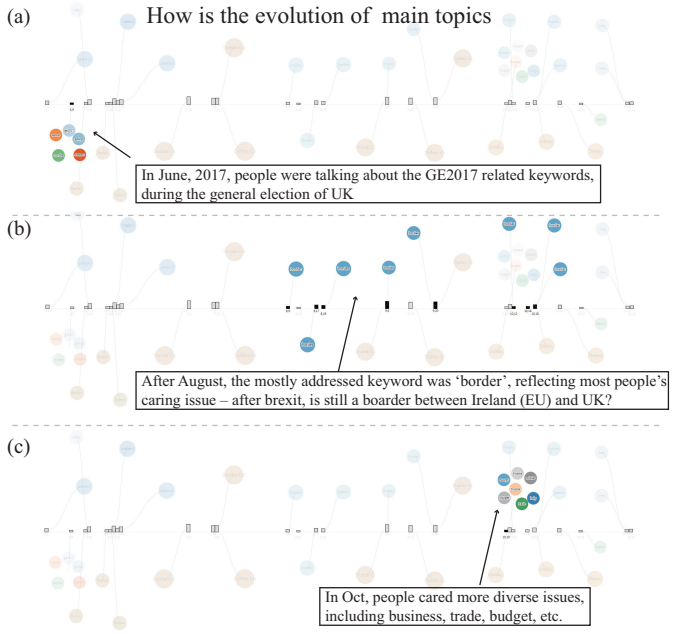

Fig. 9. Highlighting items in a time-based linear layout. Examples show highlighted keywords of (a) GE2017, (b) borders, and (c) business.

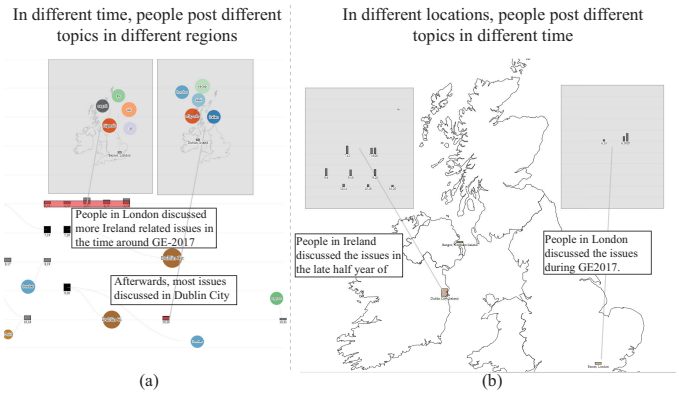

Fig. 10. Creation of nested layouts. (a) Nested layouts enable comparison of keyword distributions over locations in two time periods. (b) Nested layouts enable comparison of temporal distributions for two locations.

views can also be used for enabling comparisons.

5.3.3 Manipulation of Layouts

To adjust an arrangement of nodes to analyst's ideas for presenting the information in a story, the analyst may manipulate some arrangements by moving nodes and groups of nodes in a drag-and-drop manner. This is possible in user- and keyword-based layouts, which are not based on spatial coordinates or positions in time. Thus, the analyst can drag specific users to desired positions (Fig. 11c). In this example, the analyst has created three regions for different groups of users, which can be now compared. This demonstrates the use of layout manipulation for creating a comparison view. Another possibility is the creation of nested layouts discussed earlier (Fig. 10).

5.3.4 Annotation and Visual Linking

Annotation and visual linking are essential for storytelling. They help analysts to convey important ideas and facilitate understanding of the stories by the readers. With our system, analysts can easily add and remove annotations, put annotations in desired positions, and visually link them to selected items. For example, Figs. 10 and 11 contain multiple text boxes and links that were interactively added by the analyst.

5.3.5 Iterative Story Synthesis

We support two-way transitions between the analysis and story synthesis activities (Fig. 3). When constructing a story, the analyst can click on nodes to see the corresponding original data in the visual analytics system. When analysts feel that additional information would be good to add to the story, he/she can go back to the visual analytics system to continue exploration and extract suitable story slices. The current layout will be automatically updated when new story slices are added. In such manner, analysts can analyze data and construct stories iteratively.

5.4 An Illustrative Example of Story Synthesis

We describe a scenario of synthesizing a story about Ireland-related discussions in the corpus of tweets concerning Brexit. In the analysis process, the analyst noticed and selected a topic involving keywords related to Ireland. In response, tweets associated with this topic were retrieved for exploration. In the user distribution view, the analyst observed two outstanding Twitter users dominating others by the amounts of posted tweets (Fig. 6c). The analyst judged them as important and created story slices with these users as the keys. Then, the analyst explored the locations of the Ireland-related tweets and found that most of them were in Ireland and only a few were in London. He created story slices with these locations as the keys. He edited the slices by removing uninformative keywords, such as 'Ireland'. After excluding the dominating two people and several dominating keywords, such as 'UK' and 'EU', other users and keywords received better visibility. The analyst explored the keyword occurrences over time and created story slices for several keywords from different time periods: 'GE2017' in June, 'border' in the later months, 'business' and 'trade' in November, etc. In this way, the analyst created in total 23 story slices with 127 attributes.

In the story synthesis view, the analyst organized the slices in spatial, temporal, user-based, and keyword-based layouts and attaches annotations to selected findings (Fig. 8). He also created a temporal layout showing the evolution of keyword occurrences. From this, the analyst created views for the three earlier found periods with the prevalence of different keywords by highlighting and annotating corresponding groups of nodes (Fig. 9). Then, the analyst wanted to relate some findings to locations. He selected two time periods, June and October, and generated two nested layouts as shown in Fig. 10a. In a similar way, he created embedded temporal views in the geographic layout (Fig. 10b).

The resulting organized material for a story included four initial overviews (Fig. 8), three more detailed views

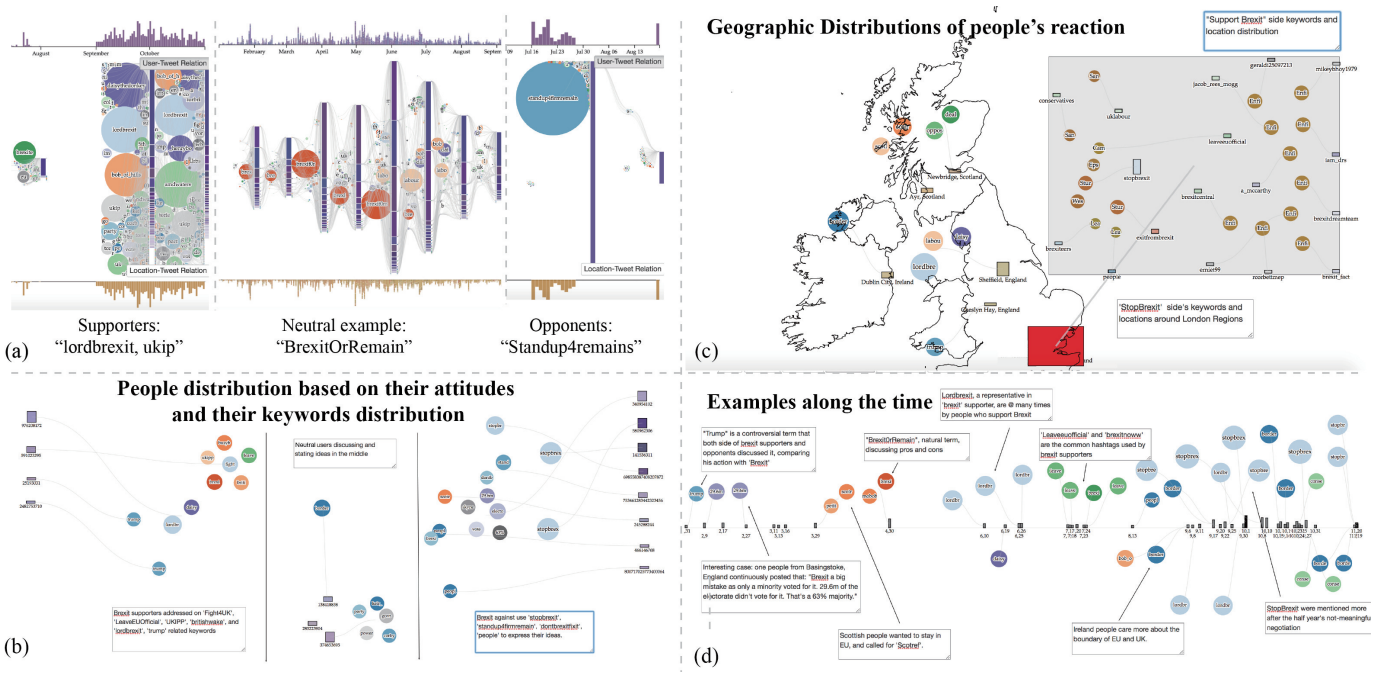

Fig. 11. Analysis and story synthesis performed by a sociology researcher. She analyzed different types of people's reactions (a) and synthesized the analysis results into a story from perspectives of users (b), locations (c) and time (d). High resolution figure is in supplementary materials.

showing trends in keyword usage (Fig. 9), and two multi-facet comparative views (Fig. 10).

6 Expert Evaluation

We collaborate with two experts from domains of sociology and political research who specialize in studying social media's reaction to events and have several years of research experience. The experts are strongly motivated to apply visual analytics in their research and spread their findings to public audience. The evaluation was conducted using semi-structured interviews. One expert worked with the system herself under our guidance. The second expert discussed a demonstration of the system operated by one of the authors. As the main goal of the evaluation was to verify the effectiveness of the proposed story synthesis framework, we asked the experts to comment more on the approach than on the features of the software system.

6.1 Expert's Experience

The expert-sociologist analyzed the Brexit data and generated her own story. We recorded the process of analysis story generation (Fig. 11). The goal of the analysis was to understand how people with different attitudes in respect to Brexit react in social media and what they mainly talk about. Starting from the topic model visualization, the expert brushed several regions in the topic projection space, including topics with representative keywords "Irish, border", "Trump, deal", "Scotland", "#stopbrexit, #brexitshambles", "UKIP" etc (Fig. 6a). During the exploration stage, she used two functions: details on demand and drill down by exclusion. She observed three main categories of people and their typical behaviors: supporters, opponents and people with a neutral opinion. She further tried to analyze the spatial, temporal and keywords distribution of these categories. For example, she identified a topic focused on the "BrexitOrRemain" keyword with a hot discussion during the first six months of 2017. She noticed that people who supported Brexit often referred to a popular user named "lordbrexit", who actively advocated Brexit and had a high social impact. By interactively drilling down to detail, she also identified people who held up the "standfirm4remain" keyword for a long time, continuously posting messages to call for anti-Brexit (Fig. 11a).

In the story synthesis process, the expert gradually filtered nodes and refined their attributes. From the initially selected 35 key nodes with 150 attributes, she removed 5 key nodes and several suggested attributes that she was not interested in, thus resulting in 30 key nodes with 128 attributes. This set was used for synthesizing a story. At the beginning of this activity, the expert stated: *"I would like to differentiate people by what they were talking about, and then examine and tell stories about the evolution in time and geographic distribution."* For this purpose she performed the following interactions. First, she switched to the User Layout panel and observed the keyword distribution for the selected users. By freely dragging the user nodes into the categories (support, against, neutral), she observed clusterings of their main keywords. Brexit supporters posted tweets with keywords *"LeaveEUOfficial, UKIP, britishawake, lordbrexit"*, while the opponents tended to use keywords *"stopbrexit, standup4firmremain, dontbrexitfixit"*, etc. (Fig. 11b). Next, she organized the story slices in a linear temporal layout and annotated examples of supporters and opponents of Brexit above and under the timeline, respectively. By citing prominent example messages such as *"Brexit is a big mistake as only a minority voted for it. 29.6m of the electorate didn't vote for it. That's a 63% majority."*, she illustrated the story of a

special keyword '29.6m' (Fig. 11d). She also examined the geographic distribution of the different people's keywords on the map. The findings of the attitude of people correspond to the overall voting patterns for regions. For the Greater London region with a complicated voting pattern, she summarized the distribution of keywords (Fig. 11c).

6.2 Experts' Feedback

We interviewed two experts for about 1.5 hours each. Both experts provided positive and constructive feedback.

The sociology expert who used the software herself commented on analysis and story synthesis: *"The interactive nature of this data visualization tool simplifies the process of deducting information for telling a coherent story. Keyword flow allows users to interact with Twitter data arranged by time, location, and frequent topic contributors. Such layouts facilitate bundling several variables in the same graph, enabling efficient patterns search. I like that I'm able to 'exclude' irrelevant keywords during the analysis, which makes it easy for me to construct the story. The consistency of the time, location, user and topics within the analysis process helps me to produce the final results."* She also had several suggestions and comments specifically to social media analysis:*"For social scientists, it will be interesting to see whether the Twitter users opinions are merely a reflection of the news report, or it is actually far from the 'mainstream' opinions. As the result of Brexit, in a similar case, Trump's being elected, were usually presented as an unexpected outcome. It will be interesting to see whether we can somehow 'predict' or observe the opinion pattern before the referendum."*

The political researcher provided comments after a demonstration of several cases, hands-on experiment, and discussion. Concerning the general idea, she commented:*"I like this concept very much. Indeed it is what I need in my current research. Usually, after conducting analytics, I have to organize findings myself. There are no suitable tools to support me to organize results. I especially like the idea of submitting results and organizing stories."* For the story synthesis phase and interactions, she pointed that *"We really need to construct our research insight from different perspectives. Especially with geographic information, my colleagues who study geopolitics will definitely need this. And also free comparison functions are needed for organizing the findings."* As political experts often study social media for understanding people reaction to government policy, she pointed out domain-specific suggestions:*"We care about the classification of users and their relationships. When we tell the story, we usually need to clarify the types of users in the final stories. If both analytics and user-facet story synthesis part can add such support, that will be great. The other issue I'm a bit concerned is the problem of tautology. The analysts should pay attention to escape that, but the current tool doesn't prevent it."*

7 Applying the Framework to Trajectory Analysis Results

To demonstrate the generality of the proposed framework, we apply it to previously obtained results from analysing aircraft trajectories [51]. For the illustrative purpose, we fulfil all steps of the suggested framework application workflow (Section 4.5) except for supporting interactive extraction of story slices from visual displays (this would require significant changes in the system that was used for the analysis; we instead construct story slices manually).

The analysis concerned the routes of aircraft approaching five airports of London during four consecutive days. The analysts identified the relationship between the route choice and the wind direction. On day 1, the airports were mostly approached from east to west, while the wind direction was from west to east. The approach direction changed to the opposite in the remaining three days, which was related to a change of the wind direction. The analysts also noticed that the change of the approach routes to one airport (Stansted) happened at a different time compared to the others. Through detailed investigation, the analysts explained this distinction by the different runway orientation in Stansted than in the other airports. The analysis also included exploration of holding loops, i.e., circular movements made by aircraft in the air when they wait for a permission to land. The highest number of loops occurred in the flights coming to the airport Heathrow. The loops were done in particular places, which were different on the first day and during the further three days.

A suitable story slice for presenting this kind of findings represents a group (cluster) of flights with similar routes using the following information facets: destination airport, flight time, landing direction, wind direction, number of loops, and location of the loops. To demonstrate the change of the approach paths over time, we aggregate the story slices representing the flight clusters in each day based on the landing and wind directions and create a layout in which the aggregated slices are arranged according to the time facet (Fig. 12b). To demonstrate the difference of Stansted from the other airports, we create a geographic layout showing the locations of the airports and incorporate in it two nested comparative arrangements with time-based layout (Fig. 12c) showing how the landing directions changed in Stansted and in the other airports. To demonstrate the high number of loops in the flights coming to Heathrow, we create a geographic arrangement showing the total numbers of flights with loops for all airports. We embed another geographic layout linked to the Heathrow airport, in which we demonstrate the relative positions of the loops with respect to the airport location (Fig. 12d). We annotate the draft views with explanatory texts.

This example demonstrates the possibility to apply the framework to a kind of information that is different from the social media data. To save technical implementation effort, we did not design a more specific visual representation for a story slice but re-used the same representation as in the social media case. As noted in Section 4.5, the visual representation of a story slice used by the analyst for constructing story contents may be replaced by another representation at the stage of storytelling. The main purpose of the story synthesis phase is to define and organize story content, and our example shows how this is done. Particularly, it demonstrates the application of the principles of facet-oriented arrangement of information presented in Section 4.3.

We can conclude that our framework is general enough at the conceptual level due to the consideration of the different kinds of information facets and corresponding approaches to arranging story slices (Sec. 4), and at the implementation level, as demonstrated by its application to

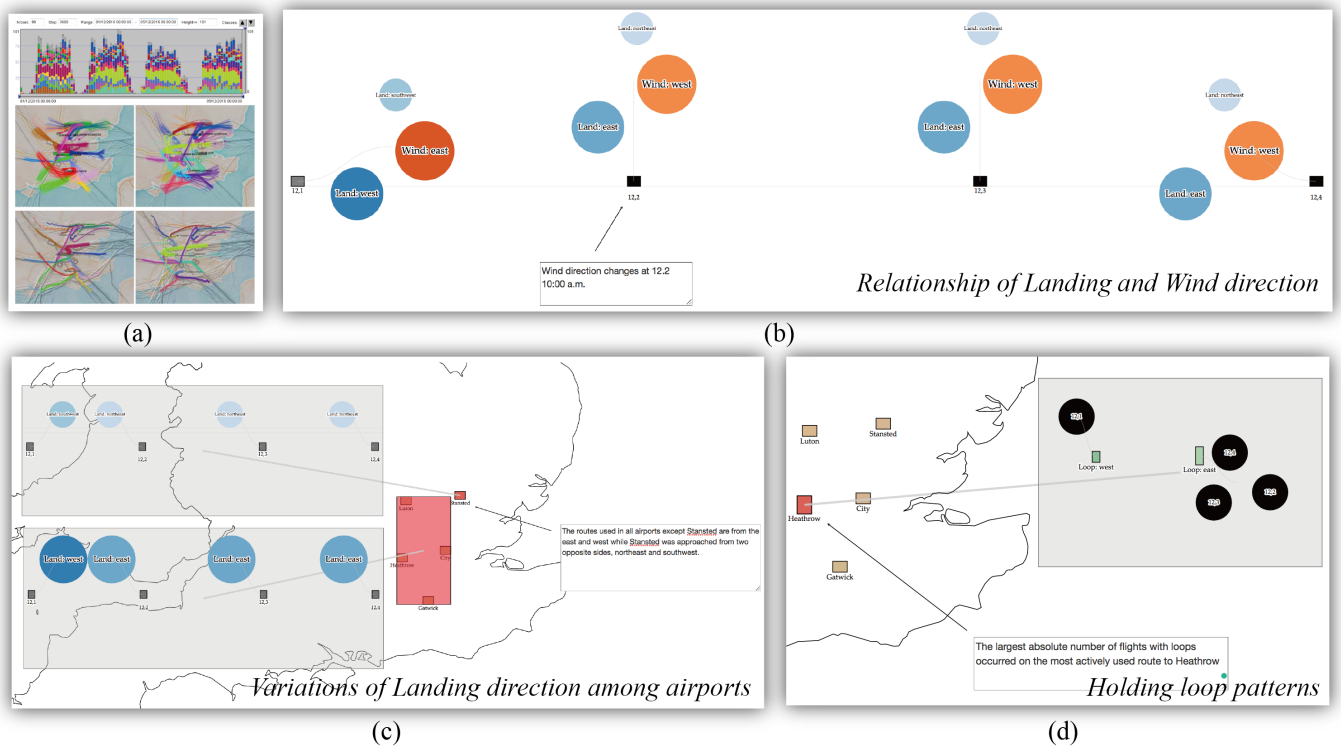

Fig. 12. The general framework applied to findings from trajectory analysis. (a) Analysis results [51]. (b-d) Synthesized draft views for illustrating the findings: relationships between landings and wind (b), differences of these relationships among the airports (c), and differences in the numbers and spatial positions of holding loops (d).

two substantially different domains.

8 Discussion and Conclusion

We developed the concept of story synthesis and the framework for story synthesis support through a practical exercise on proceeding from analyzing social media data to organizing disjoint findings in a story. This exercise stimulated and informed our reasoning on characterizing the gap between analysis and storytelling and defining the required intermediate step. Using several specific examples, we strove to reason in a general way and develop concepts and principles that would be applicable to various kinds of data and problem domains. We have thus developed the general concept of story slice as a structure involving information facets, described the possible types and inherent properties of facets, and elaborated the ideas of using these properties for meaningful arrangement of story slices. This constitutes a general framework applicable to different kinds of information. In the following, we discuss some aspects of the framework and the lessons learned during our work.

Content Preparation and Presentation Design. Story synthesis includes defining and arranging the contents of a story but not its final look and feel. The visual displays that are used in the course of story synthesis are primarily meant for the analyst rather than for the audience; therefore, they may be called "draft views". They can, in principle, be directly presented to the audience, as we did in our expert evaluation, but it would be appropriate to develop them into audience-oriented presentations by applying design principles and techniques used in storytelling [2]. We consider our research as complementary to the existing storytelling research [1].

Story Logic Guidance - Balance between Flexibility and Guidance. Story synthesis tools should support creativity, which requires the flexibility of organizing, combining and editing the story contents. At the beginning, we thought of supporting full flexibility, but understood that it entailed high complexity and lack of orientation for the analyst. We thus came to the idea of automated generation of meaningful layouts based on the properties of information facets. Still, analysts should be able to make adjustments according to their ideas. A good enhancement could be to enable (semi-)automatic generation of stories that follow established narrative structures [2], [17] while respecting the characteristics of the data. Another problem to consider is avoidance of tautologies, i.e., needless re-occurrences of the same findings, which were detected by one of the experts participating in the evaluation (Section 6.2). Such re-occurrences could be detected automatically. More generally, it would be worthwhile to conduct further research on guiding analysts towards better story organization.

Visual Analytics Requirements. In current practices, the design of visual analytics systems is often performed without proper regard to the needs for communicating analysis results. Though several papers started to consider storytelling requirements for visual analytics [28], [29], systematic guidelines for supporting storytelling in visual analytics are still lacking. We believe that it is critical to consider communication requirements early in the system design. This will help designers to define the structure of possible findings, choose a suitable representation for them,

and enable analysts to materialize their observations and thus collect findings that can then be communicated to the intended audience.

Sensemaking and Story Synthesis. Sensemaking is described as an iterative process that gradually transforms raw data into knowledge [65]. The process includes two sets of activities: the *foraging loop* (search and filter data, read and extract relevant information) and the *sensemaking loop* (organise information, build hypotheses and present them). Many visual analytics system have been developed to support the sensemaking process [66]. However, most of them focus on the foraging loop (such as the famous Jigsaw [67]), with just a few systems supporting the sensemaking loop [68], [69]. In our view, within the sensemaking loop, the final storytelling step is not straightforward because the hypotheses derived from visual analytics systems are often too complex to be understood directly. Hence, the story synthesis step is required before presenting sensemaking results to decision makers or other kinds of recipients. Our framework is designed to help synthesize information and reduce complexity in order to achieve efficient presentation of results to recipients.

Potential Bias of Story Synthesis Visualization itself may lead to cognitive biases [70]. Story synthesis process is highly dependent on analysts who conduct analysis and the ideas and opinions they want to express in the stories. Hence, in the process of story synthesis, biases can be easily introduced. While it is hardly possible to preclude conscious biases, visual analytics systems can try to decrease the risks of unconscious biases that occur when the analyst overlooks some part of information. A system can check if the analyst has explored the full data at the same level of detail and warn the analyst about a possible bias or suggest to look at unexplored parts of data.

Further Evaluation. We evaluated our framework and its implementation for social media data with involving two experts from relevant domains. It is appropriate to continue testing by applying the framework to diverse domains and involving experts and analysts with different professional backgrounds.

To summarise, we have proposed a general framework for bridging the gap between visual analytics and storytelling by introducing a story synthesis phase that extends the visual analytics workflow and connects it to storytelling. Expert evaluation showed that our framework can help analysts in organizing their findings into stories. We checked the applicability of the framework to diverse problem domains and data structures. The framework is intended to inform designers and developers of visual analytics systems who want to support analysts in preparing analysis results for presentation and communication to general audience or other recipients lacking expertise in visual analytics.

ACKNOWLEDGMENTS

This research was supported by Fraunhofer Cluster of Excellence on "Cognitive Internet Technologies", by EU in projects DiSIEM and SoBigData, by DFG (German Research Foundation) in priority research program SPP 1894 "Volunteered Geographic Information: Interpretation, Visualization and Social Computing", and by National NSFC projects (Grant numbers 61602340 and 61572348). The corresponding author is Jie Li.

REFERENCES

[1] C. Tong, R. Roberts, R. Borgo, S. Walton, R. S. Laramee, K. Wegba, A. Lu, Y. Wang, H. Qu, Q. Luo *et al.*, "Storytelling and visualization: An extended survey," *Information*, vol. 9, no. 3, p. 65, 2018.

[2] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1139–1148, 2010.

[3] J. Fulda, M. Brehmel, and T. Munzner, "Timelinecurator: Interactive authoring of visual timelines from unstructured text," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 300–309, 2016.

[4] A. Satyanarayan and J. Heer, "Authoring narrative visualizations with ellipsis," in *Computer Graphics Forum*, vol. 33, no. 3. Wiley Online Library, 2014, pp. 361–370.

[5] B. Lee, R. H. Kazi, and G. Smith, "Sketchstory: Telling more engaging stories with data through freeform sketching," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2416–2425, 2013.

[6] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale, "More than telling a story: Transforming data into visually shared stories," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 84–90, 2015.

[7] K. L. Ma, I. Liao, J. Frazier, H. Hauser, and H. N. Kostis, "Scientific storytelling using visualization," *IEEE Computer Graphics and Applications*, vol. 32, no. 1, pp. 12–19, Jan 2012.

[8] S. Chen, S. Chen, Z. Wang, J. Liang, X. Yuan, N. Cao, and Y. Wu, "D-map: Visual analysis of ego-centric information diffusion patterns in social media," in *Proc. of IEEE Visual Analytics Science and Technology*, 2016, pp. 41–50.

[9] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine, "Diamonds in the rough: Social media visual analytics for journalistic inquiry," in *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*. IEEE, 2010, pp. 115–122.

[10] G. Grinstein, M. Whiting, K. Liggett, and D. Nebesh, "IEEE VAST Challenge 2011," http://hcil.cs.umd.edu/localphp/hcil/vast11/, last accessed 01/29/2014, 2011.

[11] J. J. Thomas and K. A. Cook, Eds., *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005.

[12] N. Gershon and W. Page, "What storytelling can do for information visualization," *Communications of the ACM*, vol. 44, no. 8, pp. 31–37, 2001.

[13] N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind, "Viewing visual analytics as model building," *Computer Graphics Forum*, vol. 37, no. 6, pp. 275–299, Jan. 2018.

[14] R. Walker, A. Slingsby, J. Dykes, K. Xu, J. Wood, P. H. Nguyen, D. Stephens, B. W. Wong, and Y. Zheng, "An extensible framework for provenance in human terrain visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2139–2148, 2013.

[15] J. Hullman, S. Drucker, N. Henry Riche, B. Lee, D. Fisher, and E. Adar, "A deeper understanding of sequence in narrative visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2406–2415, Dec. 2013.

[16] R. Kosara and J. Mackinlay, "Storytelling: The next step for visualization," *Computer*, vol. 46, no. 5, pp. 44–50, 2013.

[17] S. McKenna, N. H. Riche, B. Lee, J. Boy, and M. Meyer, "Visual narrative flow: Exploring factors shaping data visualization story reading experiences," *Comput. Graph. Forum*, vol. 36, no. 3, pp. 377–387, 2017. [Online]. Available: https://doi.org/10.1111/cgf.13195

[18] J. Hullman and N. Diakopoulos, "Visualization rhetoric: Framing effects in narrative visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2231–2240, Dec. 2011.

[19] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe, "Chartaccent: Annotation for data-driven storytelling," in *2017 IEEE Pacific Visualization Symposium, PacificVis 2017, Seoul, South Korea, April 18-21, 2017*, 2017, pp. 230–239.

[20] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner, "Timelines revisited: A design space and considerations for expressive storytelling," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 9, pp. 2151–2164, 2017.

[21] B. Bach, N. H. Riche, S. Carpendale, and H. Pfister, "The emerging genre of data comics," *IEEE Computer Graphics and Applications*, vol. 37, no. 3, pp. 6–13, May 2017.

[22] M. Froese and M. Tory, "Lessons learned from designing visualization dashboards," *IEEE Computer Graphics and Applications*, vol. 36, no. 2, pp. 83–89, Mar 2016.

[23] R. Kosara, "Presentation-oriented visualization techniques," *IEEE Computer Graphics and Applications*, vol. 36, no. 1, pp. 80–85, Jan 2016.

[24] S. McKenna, D. Staheli, C. Fulcher, and M. Meyer, "Bubblenet: A cyber security dashboard for visualizing patterns," in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 281–290.

[25] Tableau, https://airtable.com/, last accessed 11/13/2018.

[26] PowerBI, "Interactive data visualization bi tools," https://powerbi.microsoft.com/en-us/, last accessed 11/13/2018.

[27] Airtable, https://airtable.com, last accessed 11/13/2018.

[28] R. Eccles, T. Kapler, R. Harper, and W. Wright, "Stories in geotime," *Information Visualization*, vol. 7, no. 1, pp. 3–17, 2008.

[29] R. Walker, L. ap Cenydd, S. Pop, H. C. Miles, C. J. Hughes, W. J. Teahan, and J. C. Roberts, "Storyboarding for visual analytics," *Information Visualization*, vol. 14, no. 1, pp. 27–50, 2015.

[30] K. Xu, S. Attfield, T. Jankun-Kelly, A. Wheat, P. H. Nguyen, and N. Selvaraj, "Analytic provenance for sensemaking: A research agenda," *IEEE computer graphics and applications*, vol. 35, no. 3, pp. 56–64, 2015.

[31] P. H. Nguyen, K. Xu, A. Bardill, B. Salman, K. Herd, and B. W. Wong, "Sensemap: Supporting browser-based online sensemaking through analytic provenance," in *Proc of IEEE Visual Analytics Science and Technology (VAST)*. IEEE, 2016, pp. 91–100.

[32] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit, "From visual exploration to storytelling and back again," *Computer Graphics Forum (EuroVis '16)*, 2016.

[33] H. Stitz, S. Gratzl, H. Piringer, T. Zichner, and M. Streit, "Knowledgepearls: Provenance-based visualization retrieval," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018.

[34] P. Lundblad and M. Jern, "Geovisual analytics and storytelling using html5," in *Information Visualisation (IV), 2013 17th International Conference*. IEEE, 2013, pp. 263–271.

[35] K. Xu, P. Nguyen, and B. Fields, "Visual analysis of streaming data with savi and sensemap," in *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2015.

[36] D. Gotz, M. X. Zhou, and V. Aggarwal, "Interactive visual synthesis of analytic knowledge," in *2006 IEEE Symposium On Visual Analytics Science And Technology*, Oct 2006, pp. 51–58.

[37] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan, "Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 340–350, 2018.

[38] S. Chen, L. Lin, and X. Yuan, "Social Media Visual Analytics," *Computer Graphics Forum*, vol. 36, no. 3, pp. 563–587, 2017.

[39] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou, "Leadline: Interactive visual analysis of text data through event identification and exploration," in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2012, pp. 93–102.

[40] P. Xu, Y. Wu, E. Wei, T. Peng, S. Liu, J. J. H. Zhu, and H. Qu, "Visual analysis of topic competition on social media," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2012–2021, 2013.

[41] G. Sun, Y. Wu, S. Liu, T.-Q. Peng, J. Zhu, and R. Liang, "Evoriver: Visual analysis of topic coopetition on social media," *IEEE transactions on visualization and computer graphics*, vol. 20, pp. 1753–1762, 2014.

[42] D. Ren, X. Zhang, Z. Wang, J. Li, and X. Yuan, "Weiboevents: A crowd sourcing weibo visual analytic system," in *Visualization Symposium (PacificVis), IEEE PacificVis Notes*, 2014, pp. 330–334.

[43] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, and X. Zhang, "E-map: A visual analytics approach for exploring significant event evolutions in social media," in *Proc. of IEEE VAST*, 2017.

[44] N. Cao, Y. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu, "Whisper: Tracing the spatiotemporal process of information diffusion in real time," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2649–2658, 2012.

[45] "Agorapulse," https://www.agorapulse.com, accessed 2018-08-17.

[46] "Brandwatch analytics," https://www.brandwatch.com/brandwatch-analytics/, accessed 2018-08-17.

[47] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu, "Storyflow: Tracking the evolution of stories," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2436–2445, 2013.

[48] N. Andrienko and G. Andrienko, "Visual analytics of movement: an overview of methods, tools, and procedures," *Information Visualization*, vol. 12, no. 1, pp. 3–24, 2013.

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

[50] J. Wood, J. Dykes, and A. Slingsby, "Visualization of origins, destinations and flows with od maps," *The Cartographic Journal*, vol. 47, no. 2, pp. 117–129, 2010.

[51] G. Andrienko, N. Andrienko, G. Fuchs, and J. M. C. Garcia, "Clustering trajectories by relevant parts for air traffic analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 34–44, Jan 2018.

[52] G. Andrienko, N. Andrienko, and M. Heurich, "An event-based conceptual model for context-aware movement analysis," *International Journal of Geographical Information Science*, vol. 25, no. 9, pp. 1347–1370, 2011.

[53] J. Zhao, C. Collins, F. Chevalier, and R. Balakrishnan, "Interactive exploration of implicit and explicit relations in faceted datasets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2080–2089, Dec 2013.

[54] M. Dörk, N. H. Riche, G. Ramos, and S. Dumais, "Pivotpaths: Strolling through faceted information spaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2709–2718, Dec 2012.

[55] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Visual Languages, 1996. Proceedings., IEEE Symposium on*. IEEE, 1996, pp. 336–343.

[56] W. Tobler, "Thirty five years of computer cartograms," *Annals of the Association of American Geographers*, vol. 94, no. 1, pp. 58–73, 2004.

[57] J. Wood and J. Dykes, "Spatially ordered treemaps," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1348–1355, Nov 2008.

[58] E. Baikousi, G. Rogkakos, and P. Vassiliadis, "Similarity measures for multidimensional data," in *2011 IEEE 27th International Conference on Data Engineering*, April 2011, pp. 171–182.

[59] M. Alamuri, B. R. Surampudi, and A. Negi, "A survey of distance/similarity measures for categorical data," in *2014 International Joint Conference on Neural Networks (IJCNN)*, July 2014, pp. 1907–1914.

[60] S. seok Choi and S. hyuk Cha, "A survey of binary similarity and distance measures," *Journal of Systemics, Cybernetics and Informatics*, pp. 43–48, 2010.

[61] M. K. Vijaymeena and K. K, "A survey on similarity measures in text mining," *Machine Learning and Applications: An International Journal*, vol. 3, pp. 19–28, 2016.

[62] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[63] J. Li, S. Chen, W. Chen, G. Andrienko, and N. Andrienko, "Semantics-space-time cube: A conceptual framework for systematic analysis of texts in space and time," *IEEE Transactions on Visualization & Computer Graphics*, 2018.

[64] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[65] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," pp. 2–4, 2005. [Online]. Available: https://analysis.mitre.org/proceedings/Final_Papers_Files/206_Camera_Ready_Paper.pdf

[66] G. Ellis and F. Mansmann, "Mastering the information age solving problems with visual analytics," in *Eurographics*, vol. 2, 2010, p. 5.

[67] J. Stasko, C. Görg, and Z. Liu, "Jigsaw: supporting investigative analysis through interactive visualization," *Information visualization*, vol. 7, no. 2, pp. 118–132, 2008.

[68] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort, "The sandbox for analysis: concepts and methods," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 801–810.

[69] P. H. Nguyen, K. Xu, R. Walker, and B. W. Wong, "Schemaline: Timeline visualization for sensemaking," in *2014 18th International Conference on Information Visualisation*. IEEE, 2014, pp. 225–233.

[70] G. Ellis, *Cognitive Biases in Visualizations*. Springer, 2018.

Siming Chen is a research scientist at Fraunhofer Institute IAIS and a PostDoc researcher of University of Bonn in Germany. He got his PhD from Peking University. His research interests include visual analytics of social media, cyber security and spatial temporal data. He published several papers in IEEE VIS, IEEE TVCG, EuroVis, etc. More information can be found at http://simingchen.me.

Jie Li is an assistant professor of the school of computer software at Tianjin University. His current research interests include visualization and visual analytics, in particular, for environmental sciences, public security, and social media. He has served as the program chair of VINCI 2017 and published about 20 papers in visualization conferences and journals, including TVCG, IEEE VIS, Journal of Computer, etc.

Gennady Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Gennady Andrienko was a paper chair of *IEEE VAST* conference (2015–2016) and associate editor of *IEEE Transactions on Visualization and Computer Graphics* (2012–2016), *Information Visualization* and *International Journal of Cartography*.

Natalia Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Results of her research have been published in two monographs "*Exploratory Analysis of Spatial and Temporal Data: a Systematic Approach*" (Springer 2006) and "*Visual Analytics of Movement*" (Springer 2013). Natalia Andrienko is an associate editor of *IEEE Transactions on Visualization and Computer Graphics*.

Yun Wang is an associate researcher at Microsoft Research. She obtained her Ph.D. in Computer Science and Engineering from the Hong Kong University of Science and Technology and her B.Eng. from Fudan University. Her research interests are in data storytelling and visual data analytics. She has published papers extensively in IEEE VIS, IEEE TVCG, ACM CHI, etc. For more information, please visit http://www.cse.ust.hk/ywangch/.

Phong H. Nguyen is a Research Associate at the giCentre, City, University of London. His research mainly focuses on the design and application of interactive visualizations to make sense of complex datasets, with a special interest in analytic provenance, logs and general temporal categorical data. He has published papers in high-impact journals including IEEE TVCG, InfoVis, VAST, CG&A and IVS. Phong holds a PhD in Visual Analytics from Middlesex University, London, UK.

Cagatay Turkay is a Senior Lecturer in Applied Data Science at the giCentre at the Computer Science Department of City, University of London conducting research in visual analytics. He serves as a committee member for several conferences including InfoVis and EuroVis, and part of the organising committee for IEEE VIS on 2017 and 2018. He is currently a guest editor for *IEEE Computer Graphics and Applications*, and an editorial board member for the *Machine Learning and Knowledge Extraction* journal.