



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2024

---

## **Extraction and Visually Driven Analysis of VGI for Understanding People's Behavior in Relation to Multifaceted Context**

Burghardt, Dirk ; Dunkel, Alexander ; Hauthal, Eva ; Shirato, Gota ; Andrienko, Natalia ; Andrienko, Gennady ;  
Hartmann, Maximilian ; Purves, Ross

DOI: [https://doi.org/10.1007/978-3-031-35374-1\\_12](https://doi.org/10.1007/978-3-031-35374-1_12)

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-267790>

Book Section

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Burghardt, Dirk; Dunkel, Alexander; Hauthal, Eva; Shirato, Gota; Andrienko, Natalia; Andrienko, Gennady;  
Hartmann, Maximilian; Purves, Ross (2024). Extraction and Visually Driven Analysis of VGI for Understanding  
People's Behavior in Relation to Multifaceted Context. In: Burghardt, Dirk; Demidova, Elena; Keim, Daniel A.  
Volunteered Geographic Information: Interpretation, Visualization and Social Context. Cham: Springer, 241-264.  
DOI: [https://doi.org/10.1007/978-3-031-35374-1\\_12](https://doi.org/10.1007/978-3-031-35374-1_12)

# Chapter 12

## Extraction and Visually Driven Analysis of VGI for Understanding People's Behavior in Relation to Multifaceted Context



**Dirk Burghardt, Alexander Dunkel, Eva Hauthal, Gota Shirato, Natalia Andrienko, Gennady Andrienko, Maximilian Hartmann, and Ross Purves**

**Abstract** Volunteered Geographic Information in the form of actively and passively generated spatial content offers great potential to study people's activities, emotional perceptions, and mobility behavior. Realizing this potential requires methods which take into account the specific properties of such data, for example, its heterogeneity, subjectivity, and spatial resolution but also temporal relevance and bias.

The aim of the chapter is to show how insights into human behavior can be gained from location-based social media and movement data using visual analysis methods. A conceptual behavioral model is introduced that summarizes people's reactions under the influence of one or more events. In addition, influencing factors are described using a context model, which makes it possible to analyze visitation and mobility patterns with regard to spatial, temporal, and thematic-attribute changes. Selected generic methods are presented, such as extended time curves and the co-bridge metaphor to perform comparative analysis along time axes. Furthermore, it is shown that emojis can be used as contextual indicants to analyze sentiment and emotions in relation to events and locations.

---

D. Burghardt (✉) · A. Dunkel · E. Hauthal  
Institute of Cartography, Faculty of Environmental Sciences, TU Dresden, Dresden, Germany  
e-mail: [dirk.burghardt@tu-dresden.de](mailto:dirk.burghardt@tu-dresden.de); [alexander.dunkel@tu-dresden.de](mailto:alexander.dunkel@tu-dresden.de);  
[eva.hauthal@tu-dresden.de](mailto:eva.hauthal@tu-dresden.de)

G. Shirato · N. Andrienko · G. Andrienko  
Fraunhofer Institute IAIS, Sankt Augustin, Germany  
e-mail: [gota.shirato@iais.fraunhofer.de](mailto:gota.shirato@iais.fraunhofer.de); [natalia.andrienko@iais.fraunhofer.de](mailto:natalia.andrienko@iais.fraunhofer.de);  
[gennady.andrienko@iais.fraunhofer.de](mailto:gennady.andrienko@iais.fraunhofer.de)

M. Hartmann · R. Purves  
University Zurich, Zurich, Switzerland  
e-mail: [maximilianchristoph.hartmann@geo.uzh.ch](mailto:maximilianchristoph.hartmann@geo.uzh.ch); [ross.purves@geo.uzh.ch](mailto:ross.purves@geo.uzh.ch)

Application-oriented workflows are presented for activity analysis in the field of urban and landscape planning. It is shown how location-based social media can be used to obtain information about landscape objects that are collectively perceived as valuable and worth preserving. The mobility behavior of people is analyzed using the example of multivariate time series from football data. Therefore, topic modeling and pattern analyzes were utilized to identify average positions and area of movements of the football teams.

**Keywords** Location-based social media · Football analytics · Visual analytics · Reactions · Behavior · Context · Emoji · Bias

## 12.1 Introduction

The rise of so-called Volunteered Geographic Information (VGI) has brought with it fundamental changes not only in the nature of geographic data but also its production and accessibility. These changes have implications across the board for those carrying out research requiring geographic data and most profoundly for research exploring how humans interact with and are affected by changes in their environment. The creation of user-generated spatial content is diverse, whether through the usage of a wide variety of sensors during activities in the real world or the use of social media platforms for information exchange in the digital world. Realizing the potential of VGI requires methods which take account of the specific properties of such data, for example, its heterogeneity, quality, subjectivity, spatial resolution, and temporal relevance. Of particular interest in this research are two types of information—first, information about people’s reactions and behavior and, second, modeling different types of contexts, both derived from location-based social media (LBSM) and football match data.

Location-based social media respective geosocial media are extensively used for expressing and exchanging thoughts, opinions, ideas, and feelings (publicly or within a particular group of people)—thus reactions to an event or related to a theme. A definition of reactions is given by Dunkel et al. (2019) consisting of an identifier to a referent event and four facets (spatial, temporal, thematic, and social) describing the reaction. A series of reactions from an actor (where an actor might be an individual, a group, or an organization) can be seen as a manifestation of this actor’s behavior (Luckmann 2013). Behaviors convey actions and provide information about doing and refraining. Activities in geosocial media are intentional, purposive, and subjectively meaningful, and they are usually targeted at others, i.e., they are social actions. The creation and production of such information is an expression of human behavior and as such influenced strongly by events and context. Key to any framework seeking to analyze behavior is a definition of dimensions through which space, time, thematic attributes, and events can be described. Moreover, modelling behavior and context requires methods to aggregate and generalize data across spatial and temporal scales.

The chapter is structured as follows. Section 12.2 describes the state-of-the-art research related to modeling of reactions and behavior, context modelling, use of emojis, and consideration of privacy. Section 12.3 introduces conceptual models for describing behavior and context. Section 12.4 presents generic methods for accounting for representativeness and bias in LBSM, for comparative visual analysis, and using emojis as contextual indicators. Section 12.4 presents application-related workflows for activity analyzes in the area of landscape and urban planning, as well as the exploration of people's mobility behavior in football games.

## 12.2 Related Work

Geosocial media is a special case of Volunteered Geographic Information (VGI) (Goodchild 2007), which enables users to communicate and connect with each other by sharing location-based information. Geovisual analysis of this information enables the study of events, geographic phenomena, and related human reactions for a variety of application cases, e.g., in the areas of urban and landscape planning, tourism, health, transport, or disaster management. Human responses manifest themselves within and outside of social network communication in a variety of ways as emotional expressions, opinions, and thoughts expressed or actions taken (Hauthal et al. 2019). Thus, Volunteered Geographic Information has added a new dimension to traditional geospatial data acquisition for human activity research (Li et al. 2020). Activities in Geosocial Media are intentional, purposive, and subjectively meaningful, and they are usually targeted at others, i.e., they are social actions. From a technical point of view, they can be subdivided accordingly Davis (2016) into origination (creating own, original content, e.g., tweeting, posting), acknowledgment (reactions to content, e.g., liking, favorite), associating (interaction with content, e.g., replying, commenting, mentioning, following), amplification (spreading content, e.g., retweeting, sharing), and action (moving beyond content, e.g., signing up for a newsletter, buying something, or going to a demonstration).

The research presented within this paper aims at analyzing reactions as a component of behavior, incorporating external context to better allow events and activities to be related and compared. Key to any framework seeking to analyze reactions to events is a definition of dimensions through which both reactions and events can be described. As pointed out by Teitler et al. (2008), these dimensions form the core of a description of an event and include not only ways of describing (what, who, where, when) but also explaining (how and why). Answering these questions can be seen as a way of characterizing the context of an event and associated behaviors (Dunkel et al. 2019).

As the project has been focusing on the concept of individual and collective behaviors (particularly, behaviors of social media users), it was appropriate to review the existing literature on the topic of spatial behavior analysis. The most intensive studies have been conducted in the area of sports analytics, where the researchers focus on the behaviors of players and teams in sport games. Duarte

et al. (2012) propose to consider sport teams as superorganisms. “Superorganism” is defined in sociobiology as a group of individuals self-organized by division of labor and united by a system of communication. The division of labor can be characterized by the “areas of responsibility” of the players on the pitch. The communications can be represented by graphs of interactions, e.g., passing networks. Fonseca et al. (2013) use Voronoi diagrams to study the spatial interaction behavior in terms of continuously changing players’ arrangement on the pitch. Gudmundsson and Horton (2017) published a survey of approaches that use spatiotemporal data from team sports. This includes approaches from social network analysis, which are applied to passing networks and transition networks. Specifically for football (soccer), Memmert et al. (2017) present approaches to analysis of collective organization (distribution on the pitch, maintenance of distances between players); inter-player coordination; inter-team and inter-line coordination; correspondences between team formations as an aspect of inter-team interaction; formations of tactical groups, e.g., offense and defense; etc. Many of the existing approaches are specific for sports games and can hardly be applied to other types of data. In our project, we strive to develop more general concepts and techniques, and we also strive to consider behavior dynamics rather than derive summary characteristics of particular individual or group behaviors.

Emojis play a special role in the analysis of reactions, especially feelings and opinions. Their use in geosocial media is just as popular as the use of hashtags (Bai et al. 2019; Highfield and Leaver 2016), and as a language-independent characters, they have the advantage of avoiding error-prone language processing (Hu et al. 2013; Kralj Novak et al. 2015). The use of emojis is influenced by regional and cultural context (Kejriwal et al. 2021), so emojis have great potential to characterize spatial context but also analyze activities related to places and events (Hauthal et al. 2021).

Volunteered Geographic Information is based on people’s willingness to collect and share content with the public community. It should be noted that this data is always related to individuals and could therefore be sensitive in terms of privacy and ethical issues. Olteanu et al. (2019) demands that an ethical approach must respect individual autonomy. In the case of purposeful active participation in VGI projects, this can be ensured by confirming active consent (by a declaration of informed consent). This is more critical in terms of analyzing millions of social media posts, even when users post the content publicly and agree to the terms of use for third parties to use it. Williams et al. (2017) therefore calls for user expectations to be taken into account, to perceive changed contexts, for example, in the form of “context collapse” (Crawford and Finn 2015), and to be respectful when combining potentially sensitive personal data. Through an “aggregation effect” (Solove 2013), privacy-relevant insights can be gained from different data sources without the contributing user being aware of it. An ethical research approach requires a balance between social and individual interests—the boundary of privacy is not rigid but depends on the topic, place, time, and user characteristics. Cartographers and geoscientists have a responsibility here to develop flexible methods that protect user

privacy while taking these contextual factors into account (Burghardt and Dunkel 2022).

### 12.3 Theoretical and Conceptual Foundations

In the first phase of the project, we developed a conceptual model underpinning the extraction, analysis, and visualization of *events* and *reactions* to events in location-based social media (Dunkel et al. 2019). For instance, a message or post published on a social media platform could be considered as a reaction, related to either simple events (e.g., a tweet that is observable or a single rumble of thunder) or preceding and ongoing chains of events (e.g., the Brexit). The key feature of our conceptual model and its implementation is the integration of spatial, temporal, thematic, and social dimensions combined with an explicit link between events and reactions.

The second phase of the project focused on the concept of individual and collective *behavior*. Among multiple existing definitions of the concept of behavior (Henriques and Michalski 2020), a simple definition suitable for our purposes can be “an organism’s activities in response to external or internal stimuli” (American Psychological Association n.d.). From our perspective, it is important that a behavior unfolds over time, i.e., it covers a certain time interval, unlike a singular reaction, which can be viewed as an instant event. Like a reaction, a behavior belongs to a certain *actor*, who may be an individual, a group, an organization, or even the population of a country. However, a behavior may include multiple reactions to the same or different events. A behavior may also include *movements* of the actor in space and changes of various characteristics, which can be expressed by thematic attributes. Hence, like the concept of reaction, the concept of behavior involves social (who?), temporal (when?), spatial (where?), and thematic (how?) facets. However, the temporal facet extends to a time interval, and the spatial and thematic aspects can no longer be expressed by a singular location in space and a combination of values of thematic attributes but need to be represented by a time-referenced sequence of locations (i.e., a trajectory in space) and time series of values of the thematic attributes.

Besides, as already mentioned, a behavior may include multiple reactions to one or more events, and the reactions themselves are conceptualized as events. Even more generally, a behavior may also include not only reactions but also other kinds of events in which the actor participates, for example, actions or decisions.

Hence, we conceptualize a behavior as a tuple  $B = (p, T = [t_1, t_2], S(t), A(t), E^T = \{e_1, \dots, e_N\})$ , where  $p$  is an actor,  $T$  is a time interval,  $S(t)$  is a function representing changes of the spatial position of  $p$  over time,  $A(t)$  is a function representing changes of the thematic attributes of  $p$  over time, and  $E$  is a set of events involving  $p$ , including but not limited to  $p$ ’s reactions to other events. In  $S(t)$  and  $A(t)$ , the variable  $t$  takes values from  $T$ . The times of all events in  $E^T$  are contained in  $T$  or at least overlap with  $T$ .

Since a behavior is a complex dynamic (i.e., changing over time) entity, study and understanding of behaviors require abstraction by which elementary locations, characteristics, and events that took place at different time moments are united and treated together as units, which are called *patterns*. Collins et al. (2018) proposed the following definition of a pattern: “a representation of a collection of items of any kind as an integrated whole with specific properties that are not mere compositions of properties of the constituent items.” A more formal definition of a data pattern has been later proposed by Andrienko et al. (2021b). In brief, a pattern is a combination of relationships between data items. For time-related data, the pattern-forming relationships include temporal ordering, temporal distances, and relationships between temporally arranged spatial locations (direction and distance), attribute values (equality or difference, sign and amount of difference), and sets of events (equality, inclusion, overlap, disjointedness). Such relationships can also exist between patterns, which enable joining two or more patterns into more complex patterns. For example, a pattern of increase of values of a numeric variable followed by a pattern of decrease can be joined in a single pattern of a peak in the numeric value variation.

A behavior can thus be represented as a combination of patterns reflecting changes of the spatial location (i.e., movement patterns, such as quick movement straight to the north), changes of thematic attributes (e.g., increase in frequency and duration of sport activities), and/or changes in the set of relevant events (e.g., end of exams and beginning of a holiday).

A behavior unfolding during a time interval of substantial length may need to be described in terms of complex patterns composed of temporally arranged simpler patterns, i.e., sequences of simpler patterns. Such complex structures are hard to analyze and compare. Therefore, it is practically useful to divide behaviors into parts, which may be called *episodes*. An episode is an excerpt from a behavior  $B$  having the same structure as  $B$ , i.e.,  $EP = (p, T' = [t'_1, t'_2], S(t), A(t), E^{T'} = \{e_1, \dots, e_N\})$ , but  $T' \subseteq T$ , the variable  $t$  in  $S(t)$  and  $A(t)$  takes values from  $T'$ , and  $ET' \subseteq ET$  consists of those events that existed during the interval  $T'$ . The idea behind dividing a behavior into episodes or extracting episodes from a behavior is that an episode is relatively short and can be described in terms of simpler patterns. Then, analysis of a single behavior consists of comparison of the constituent episodes, i.e., revealing similarities and differences between the patterns in the episodes, which can be followed by detecting re-occurrences of similar patterns and investigating temporal relationships between the re-occurrences. Comparative analysis of two or more different behaviors involves comparisons of their episodes, which may include, in particular, comparison of co-occurring episodes from different behaviors as well as detection of similar asynchronous episodes in different behaviors.

Any behavior and, consequently, any episode of a behavior occurs in a certain *context*, i.e., a combination of circumstances that change over time. A context can be conceptually modelled as a tuple  $C = (T, S, A^C(t, s), A^C(t), E^{CT}, B^{CT})$ , where  $T$  is a time interval,  $S$  is a part of space,  $A^C(t, s)$  represents various dynamic attributes whose values are associated with different spatial locations (e.g., weather,

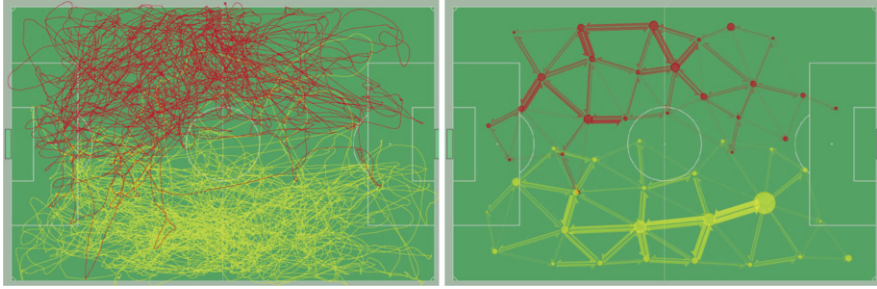
fuel prices),  $A^C(t)$  are dynamic attributes considered as global (i.e., their values refer to  $S$  as a whole),  $E^{CT}$  is a set of events that occurred during the interval  $T$  (e.g., political events or epidemic outbreaks), and  $B^{CT}$  is a set of behaviors of various actors other than the actor(s) whose behaviors are analyzed. Context-aware analysis of behaviors and their constituent episodes includes determining relationships between patterns in the behaviors/episodes in focus and patterns occurring in the context attributes, context events, and context behaviors. The analysis can be directed from the context to the behaviors (e.g., what behavior patterns occurred when the air temperature was high or after pandemic lockdowns) or from the behaviors to the context (e.g., what events or attribute development patterns preceded the occurrence of a given pattern in one or more behaviors).

Let us illustrate the concepts using an example of a football game. Each player of a team is an actor having his/her behavior happening in the course of the game, i.e.,  $T$  is the time of the game. The player's behavior includes movements of the player  $S(t)$ , his/her physical condition  $A(t)$ , and game events  $E$  involving the player, such as passes, attacks, shots on goal, goals, etc. Besides the individual behaviors of the players, there is a collective behavior of a whole team, including team movements, changes of team width, depth and relative arrangements of the players, and game events involving one or more players. The context of the team's behavior includes the behavior of the opponent team, the game events that have already happened, the weather and lighting conditions, the team's ranks, the goals set by the coach, etc. The context of the individual players' behaviors includes, additionally to what was listed above, the individual behaviors and skills of their teammates and of the opponents.

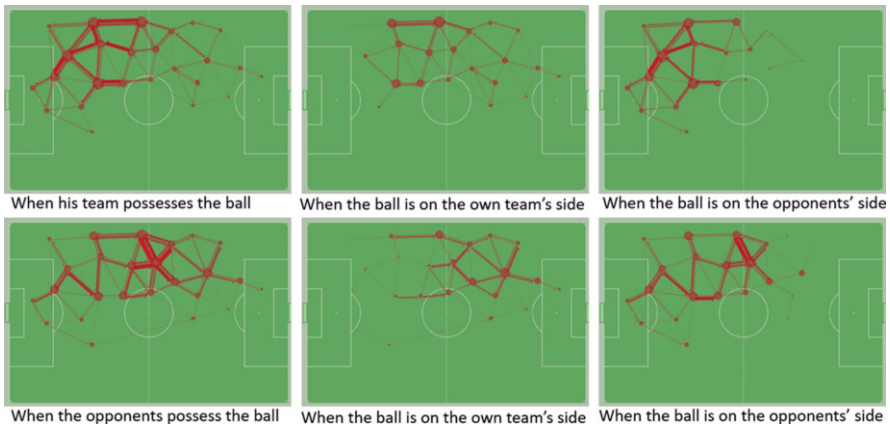
Since behaviors and their contexts are inherently complex (dynamic and multifaceted), analysis of behaviors and their relationships to contexts is a very challenging problem that cannot be tackled without simplification. Possible operations for achieving include *decomposition* (e.g., behaviors are divided into episodes), *abstraction* (e.g., unification of multiple items into patterns), *selection* (focusing on particular facets and their components, particular combinations of contextual circumstances, or particular types of patterns), and *aggregation* (representation of sets of items by summaries).

The concepts and ideas presented above are illustrated in Figs. 12.1, 12.2, 12.3, and 12.4 by example of football data. These data are more suitable for illustration purposes than data from location-based social media due to their high quality and absence of legal and ethical issues.





**Fig. 12.1** An illustration of behavior simplification by means of abstraction and aggregation. The image on the left shows the trajectories of two players during a football game. To increase the level of abstraction, the points of the trajectories were grouped and abstracted to areas on the pitch. The segments of the trajectories were transformed into transitions between the areas which, in turn, were aggregated into flows between the areas. The resulting spatial network is a simplified representation of the spatial facet of the behavior

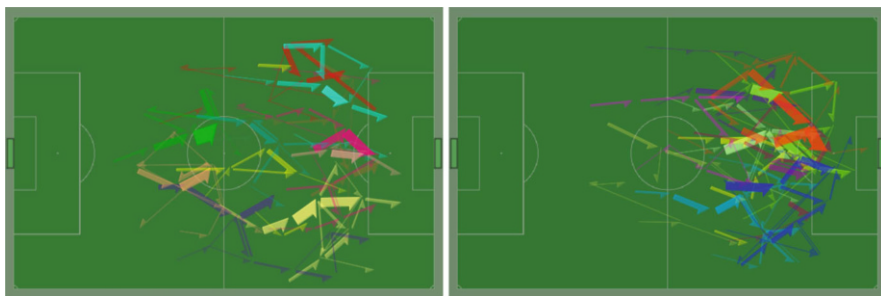


**Fig. 12.2** An illustration of the decomposition of a player's behavior during a football game into several behaviors occurring in different contexts. The behaviors are represented in the form of spatial networks resulting from abstraction and aggregation of original data, as shown in Fig. 12.1

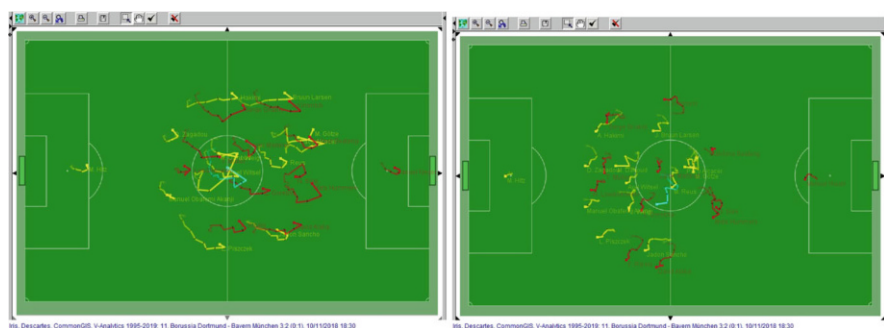
## 12.4 Generic Methods That Support Studies of Reactions and Behaviors

### 12.4.1 *Representativity and Bias in Location-Based Social Media*

In the second part of this chapter, we discuss a range of analyzes based on location-based social media. Before we describe these analyzes in more detail, it is important to consider issues related to representativity and bias with respect to such data sources. It is important to firstly remember that all analyzes are subject to these



**Fig. 12.3** Abstracted and aggregated representation of collective behavioral patterns of two teams in a selected subset of episodes. The left image corresponds to a team in possession of the ball and the right image to a defending team



**Fig. 12.4** Abstracted movement behaviors of players of two teams in two sets of episodes: first 12 seconds after gaining the ball by the yellow team (left) and after gaining the ball by the red team (right). The movements are abstracted by taking the average players' positions from all episodes at the time moments  $t + 0, t + 1, \dots, t + 12$  seconds, where  $t$  is the time moment of the ball possession change

issues—for example, if we generalize patterns found in the football data described above, it is important to consider which teams are more likely to be monitored and generate such data and to what extent these represent footballers more generally.

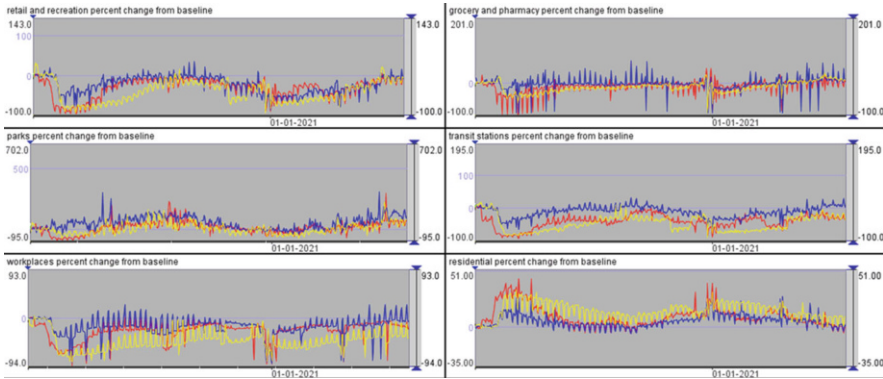
However, in location-based social media, these issues are more obvious and immediate. For example, these data are exclusively produced by people with location-enabled devices. These are not equally available within and across society, with, for example, the very old and very young being less likely to be captured by such data, and considerable variations exist in the willingness of individuals to share location-based information in different cultures and countries (Li et al. 2013; Krasnova and Veltri 2011). The first question we must ask therefore with respect to representativeness is who can, and is willing to, contribute data to platforms such as Twitter, Flickr, Instagram, and Google's location-based services. Some uses of space—for example, for play by small children—are likely to be underrepresented, especially if children are encouraged to spend time outside without their parents.

The second question that we can ask with respect to representativity relates to the ways in which these data are linked to space. Typically, we use other geographic datasets, for example, produced by national mapping agencies and/or volunteers (e.g., OpenStreetMap and GeoNames), and the content of these datasets will profoundly influence any analysis. For example, OpenStreetMap is known to have biases related to gender in terms of the categories of objects mapped (Gardner et al. 2020), while place-name density in GeoNames reflects geopolitical events (Acheson et al. 2017). A third question that we can ask with respect to these issues relates to culture and language. The underlying models used to capture, for example, emotions are often based on Western (and more specifically Anglo-Saxon) notions and assume universal emotions shared across cultures. Furthermore, many methods to analyze text focus on English as a starting point, despite clear evidence that this has limitations with respect to the ways in which we understand the world (Blasi et al. 2022).

### ***12.4.2 Methods for Comparative Analyses***

As stated earlier (Sect. 12.3), behavior is a complex, dynamic entity that can be studied and understood only with the help of simplifying abstractions and decomposition. Decomposition includes interchangeable selection of aspects and facets to put in focus and division of behaviors into episodes. Depending on the analysis goals, there can be two strategies for decomposing behaviors into episodes: partitioning of the entire behavior by dividing its time span into intervals, e.g., of a chosen fixed length and extraction of episodes with particular properties. The latter strategy is achieved by means of temporal queries, which select multiple disjoint time intervals such that the query conditions fulfil during these intervals. The pieces of behaviors contained within the selected time intervals are extracted as episodes for analysis. A set of primitives for making temporal queries has been proposed by Andrienko et al. (2021a). The query primitives enable selection of sets of time intervals containing situations with specified characteristics and, moreover, further selection of sets of intervals having certain temporal relationships to the previously selected intervals. This can be used, in particular, for considering selected episodes stepwise or for studying what happened before or after them.

Comparative visual analysis of selected aspects of behaviors can be enabled by juxtaposed representation of two or more behaviors along a time axis. This approach was taken by Chen et al. (2021) for comparison of streams of text messages published on social media by different politicians. The authors created an imaginative visual design where the flow of time is represented as a river and bridges across the river are built from significant keywords extracted from the texts. This technique has been called “co-bridges.” The visualization supports both qualitative (common and distinct keywords) and quantitative (stream volume, keyword frequencies) comparisons. Moreover, it is possible to compare two or



**Fig. 12.5** Comparison of population mobility behaviors in three countries (Germany in blue, Italy in red, and the UK in yellow) using multiple time graphs showing dynamics of different attributes

more co-bridges by juxtaposing them. These may be, for example, co-bridges corresponding to different themes of the politicians' agendas.

While alignment along a common time axis is essential for comparing behaviors as dynamic entities, the representation of behaviors depends on the type of data under analysis. One of our recent work directions was comparative analysis of behavior characteristics expressed by multiple time-variant numeric attributes, i.e., by multivariate time series. An approach is being developed using, among others, the publicly available data of COVID-19 Google Mobility Trends.<sup>1</sup>

The company Google summarizes anonymized data provided by apps such as Google Maps into statistics showing how collective mobility behaviors of people in different countries were changing throughout the COVID-19 pandemic. The data consist of daily counts of visitors to specific categories of places (e.g., grocery stores, parks, train stations, etc.) relative to baseline days before the pandemic outbreak. Hence, the population mobility behaviors in different countries are expressed by time series of six attributes.

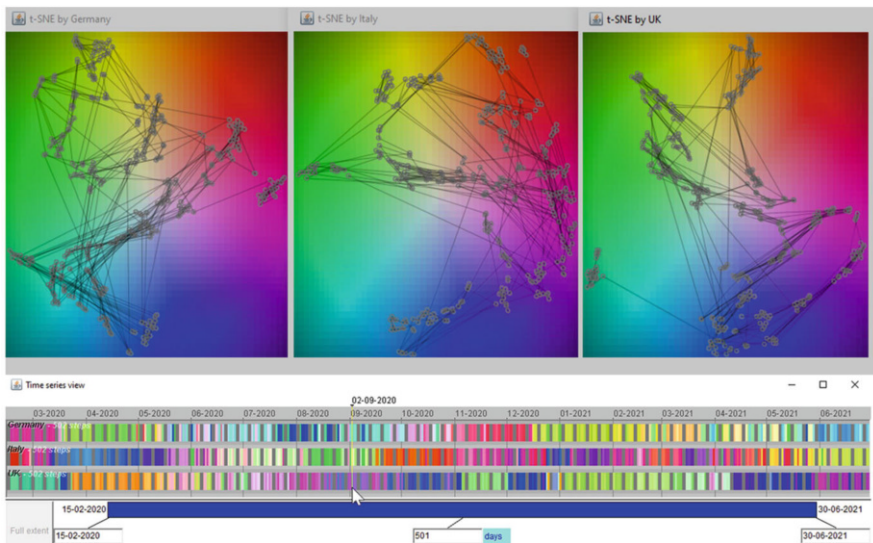
In comparing the behaviors in different countries, it is insufficient to compare the dynamics of each attribute separately from others, although this can be done relatively easily by means of usual time graphs or line charts, as shown in Fig. 12.5. This visualization does not support holistic perception of patterns of joint development of the attributes.

Consideration of temporal evolution of complex data, in particular, combinations of values of multiple attributes, is supported by the visualization technique called time curve (Bach et al. 2016); a similar method was simultaneously proposed by van den Elzen et al. (2016). The approach relies on embedding of data corresponding to different time units in a low-dimensional (typically 2D) space based on similarities between the data in terms of an appropriate similarity metric.

<sup>1</sup> <https://ourworldindata.org/covid-google-mobility-trends>.

The class of computational methods for data embedding is known as dimensionality reduction methods. Well-known examples are multidimensional scaling (MDS), Sammon's mapping, t-SNE, etc. The result of embedding (also called projection) is visualized as a scatterplot where time units are represented by points. Consecutive time units are represented by lines. Spatial arrangements of the points and lines are treated as different development patterns, such as gradual or rapid changes, stable states, oscillation, stagnation, etc. (Bach et al. 2016; van den Elzen et al. 2016).

The time curve technique is by itself poorly suited for the task of comparative analysis of two or more behaviors which, as we stated previously, calls for an opportunity to see the behaviors aligned along a common time axis. We extend the time curve technique in the following way. We paint the background of a projection plot (i.e., a plot representing a result of behavior embedding) using a continuous 2D color scale. Thereby, each point on the plot receives a specific color. The colors of the points can be transmitted to other displays, in particular, to a timeline display suitable for aligned representation of several behaviors. This generic approach is demonstrated in Fig. 12.6 using the COVID-19 mobility data for Germany, Italy, and the UK shown in Fig. 12.5.



**Fig. 12.6** Upper part: the population mobility behaviors in Germany, Italy, and the UK are represented by means of data embedding (projection) according to the time curve technique. To enable aligned representation of the behaviors along a common time axis, continuous coloring is applied to the backgrounds of the projection plots. Lower part: a joint representation of the three behaviors in a timeline display, where the horizontal dimension represents the time flow. The behaviors are represented by horizontal bars with segments painted in the colors of the corresponding points in the projection plots. In this example, the Saturdays and Sundays are filtered out (shown in the timeline view as gray segments), which enables disregarding the weekly fluctuations and focusing on long-term patterns of change

The character of the color variation in the timeline view is representative of the character of the joint development of the multiple attributes. Abrupt changes of the color tone along the timeline signify sudden or rapid changes in a behavior, bar pieces painted in very similar color shades correspond to stable states in the behavior development, and smooth changes of shades correspond to gradual transitions between states.

### ***12.4.3 Emojis to Study Sentiment, Emotion, and Context of Events***

Since geosocial media are used to state opinions, express emotions, or document experiences, they contain a lot of subjective information. The recognition of such subjective phenomena is usually done via natural language processing, which is by now quite sophisticated but can hardly recognize irony or sarcasm, for example, and is often applied limited to one or a few languages. Promising solutions have been achieved in this context with emojis, which have become extremely popular in geosocial media and are available in steadily growing numbers.

Hauthal et al. (2019) used a Twitter dataset to investigate reactions to the political event Brexit in terms of opinions and emotions, using emojis in two different approaches. In the first approach, emojis and hashtags were combined. Hashtags, established in political campaigns before the referendum, indicate which sub-topic of the overall Brexit debate is addressed in a tweet, i.e., leave or remain. A sentiment toward these topics in terms of positive or negative was detected by emojis appearing in the same tweet. For this, emojis showing a positive or negative facial expression were considered, based on the official categorization of emojis by Unicode. The combination of a hashtag and an emoji results in the rejection or support of the UK leaving the European Union. A spatial comparison of these analysis results with the actual referendum results on NUTS1 level (the highest level in the hierarchical classification used to clearly identify and classify the spatial reference units of official statistics in the Member States of the European Union) showed a higher consistency than a pure hashtag-based consideration without including emojis.

In the second approach, emojis showing faces or persons with a countenance or gesture were not only considered on a positive-negative scale but were assigned to emotional categories based on a classification according to Shaver et al. (1987), which includes love, joy, surprise, fear, sadness, and anger (see Fig. 12.7). Each category of this classification is allocated emotional terms that are most likely to be mentioned when people are asked to name those emotions. The official Unicode names of all previously used emojis were matched with these terms and assigned to the corresponding emotional categories. Emotions were then examined comparatively before and after the announcement of the Brexit referendum results, with only sadness showing a significant increase overall and fear decreasing slightly. Spatially, the increase in sadness is evident in two out of three NUTS1 regions,

positive			negative			
Love	Joy	Surprise	Anger	Sadness	Fear	Neutral

**Fig. 12.7** Classification of Unicode Emoji (List v11.0) by authors into positive, negative, and according to emotional categories based on Shaver et al. (1987)

where votes did not match the overall referendum result, and an increase in joy in five out of nine NUTS1 regions where the hopes were fulfilled.

Another use of emojis to investigate subjectivity, in this case perception, was implemented in a study by Hauthal et al. (2021). In a global Instagram dataset about sunrise and sunset, the measure typicality was applied. Typicality is a relative measure specifically tailored for geosocial media that determines how typical a particular object of interest (e.g., emoji or hashtag) is within a sub-dataset compared to the total dataset. Sub-datasets may be formed spatially, temporally, thematically, etc. Typicality is calculated by the normalized difference of two relative frequencies and returns a positive (= typical) or negative (= atypical) value. Typicality was used to identify emojis in the previously mentioned global Instagram dataset that provide information about the context of the user while observing the event. On the one hand, these emojis deliver information about activities performed and, on the other hand, also about perceived landscape features in the immediate surroundings. It was found that emojis provide more detailed information in this regard than the hashtags contained in the same dataset. Moreover, location-specific emojis were identified, which are chosen depending on the location, and match the features of the physical environment, as shown by matching them with geographic attributes. This proves that emojis are not randomly chosen but provide insights not only into the user’s situational context but also into their perception and thus appreciation of certain aspects of the environment.

These studies demonstrate the potential of emojis to provide insights into the subjectivity of geosocial media users in a relatively straightforward way. A further increase in the use of emojis as well as an increasing variety of them can still be expected, which will open up further possibilities and applications.

The conceptual framework presented in Sect. 12.3 strongly influenced both of the presented studies. In the study on Brexit, the framework led the way in allowing us to look at the reactions contained in the underlying dataset from numerous different perspectives and thus obtain in-depth results. Furthermore, the framework significantly influenced the development of the typicality measure in the second study described, as the sub-datasets required for the calculation can be formed following the four facets, which is the particularity and novelty of this measure.

## 12.5 Application-Oriented Workflows

### 12.5.1 Activity Analysis for Landscape and Urban Planning

VGI is also increasingly recognized as an important resource in the fields of landscape and urban planning, for example, to support the analysis of visitation patterns, assessing collective values, or improving human well-being through fair and equitable design of public green spaces (Ghermandi and Sinclair 2019). To this end, landscape and urban planners must first assess “what” is collectively valued, “where,” by “whom,” and “when” to understand the how and why of human behavior, as introduced in Sect. 12.3. However, the reproducibility of human behavior research is often impaired because samples, populations, and the phenomena being observed change between studies (Gruebner et al. 2017). This is particularly true for VGI and geosocial media, which are noisy, biased, difficult to fully sample, and often shared through incompletely documented and opaque application programming interfaces (APIs). In addition to these core challenges, protecting user privacy is becoming increasingly important when working with user-generated content (danah boyd and Crawford 2012). For this reason, Dunkel et al. (2023a) sought to develop a robust and transferable “workflow template,” for assessing human activities and subjective landscape values through geosocial media worldwide—without compromising user privacy.

For demonstration purposes, an event type with a strong temporal and spatial consistency was chosen that allowed for a significant reduction in the number of “incidental variables” in the study while at the same time maintaining sample volume. Sunset and sunrise were among the few events that met these criteria. In addition, improving results reproducibility ideally requires an experiment with two maximally separated datasets and “finding relationships in the same direction and of similar strength” (Laraway et al. 2019, p. 38) in both. This was difficult to implement with more “newsworthy” topics. The consistent global and long-term footprint of the sunset and sunrise offered an opportunity to maximize the sample size while also providing a basis for reproducing the results using two datasets, albeit not universally representative but independent, collected from Instagram and Flickr. Despite the narrow topic of this study, the shared expected frequencies (e.g., Flickr 300 million post counts dataset for a  $100 \times 100$  km grid, Dunkel et al. 2023b)



can also be very useful for calculating chi in studies of other phenomena at global scales, for differently sampled data, e.g., based on a different set of search terms.

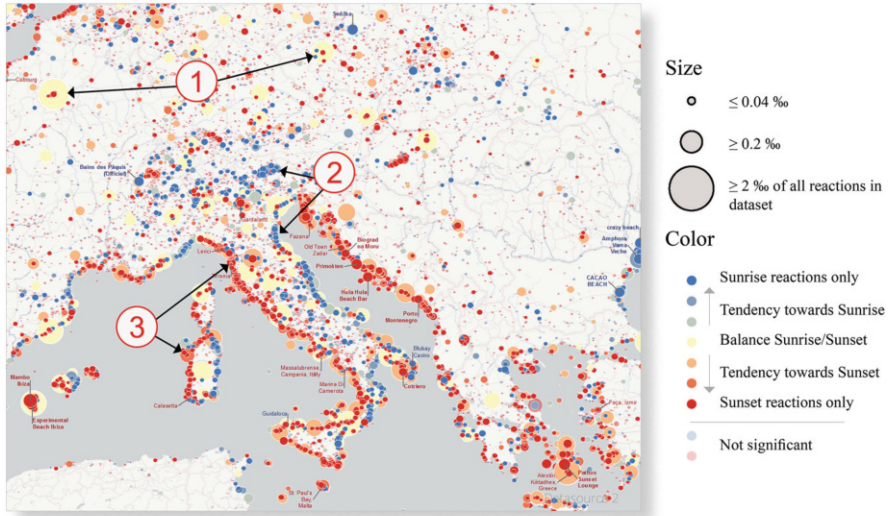
Importantly, sunset and sunrise events are entirely ephemeral yet have a profound, measurable impact on human perception and interaction with the environment. Unlike many other events, the ability to perceive sunset and sunrise is tightly bound to time, but almost completely decoupled from space. Photographs of these events function as evidence of presence in place and time. While the immediate reaction to taking a photograph of a sunset or sunrise is trivial, people will take into account all previous experiences, learned behaviors, expectations, etc. when reacting. Individual photographs therefore reflect different memorable experiences that function as artifacts of different preference contexts. The narrow thematic filter of sunset and sunrise allowed for a focused description and evaluation of these preference contexts. Based on the four-facet context model (Dunkel et al. 2019), reactions to sunset and sunrise were examined in terms of where, who, what, and when.

The study first asked whether it was possible to compare the relative importance of sunset and sunrise reactions, independent of overall visitation frequency, for different locations worldwide. Visualizing relative user frequency was critical because geosocial media tends to be skewed toward highly populated locations and cities. The goal was to provide a balanced evaluation assessment of sunset and sunrise reactions across different rural and urban regions. Several visualization methods were tested, such as based on a relative ranking method for individual locations (Fig. 12.8) and different metrics, such as user counts, post counts, or user days (Wood et al. 2013). The final workflow uses the signed chi equation, proposed by Clarke et al. (2007), to visualize over- and under-frequented with respect to these two events and aggregated using HyperLogLog for a global grid with a resolution of  $100 \times 100$  km.

Globally, sunrise events are often associated with east coasts (e.g., Italy, Sardinia in Fig. 12.8) or mountainous regions (e.g., the Alps), while sunset events are photographed on west coasts. The study also observed a strong ranking order between Flickr and Instagram reactions, despite the fact that both platforms have different user groups. In other words, we actually expected a much stronger effect of the platform on the results, and our work shows, at least for Instagram and Flickr and the selected events, that results can be reliable and reproducible across platforms. Still, for some locations, the incentives of the social media platforms themselves can have a significant impact on what gets shared and by whom. On Instagram, for instance, the Burning Man festival in Nevada ranks second worldwide for sunrise reactions. Out of a total of about 70,000 total visitors (Wikipedia, Burning Man,<sup>2</sup> 2022), 1295 ( $\pm 30$ ) users shared sunrise images on Instagram during the short period of the 2017 Burning Man festival, compared to only 54 ( $\pm 2$ ) Flickr users for the same location over a 10-year period—a pattern that can be explained by the different user groups of these platforms. Finally, the use of abstracted, estimated,

---

<sup>2</sup> [https://en.wikipedia.org/wiki/Burning\\_Man#2013\\_to\\_2019](https://en.wikipedia.org/wiki/Burning_Man#2013_to_2019).



**Fig. 12.8** Relative ranking of places based on reactions to the sunset and sunrise in Central Europe. This figure shows a preview visualization of the collected raw data that was later further aggregated to a  $100 \times 100$  km grid in Dunkel et al. (2023a). Some repeated patterns are already observable here, such as the larger cities (1) commonly featuring a balanced representation of sunset and sunrise reactions, whereas sunrise reactions dominate along eastern-facing coasts and in mountainous regions (2) and sunset reactions being predominantly shared from along western-facing coasts (3)

non-personal data based on HyperLogLog, as suggested by a scoping study (Dunkel et al. 2020), was a practically feasible solution that supports a shift toward privacy-preserving and ethically aware data analysis in human preference research. The analysis process and anonymized data are made available in a repository, allowing transparent verification, replication, and transfer to other events or datasets (data repository; Dunkel et al. 2023b).

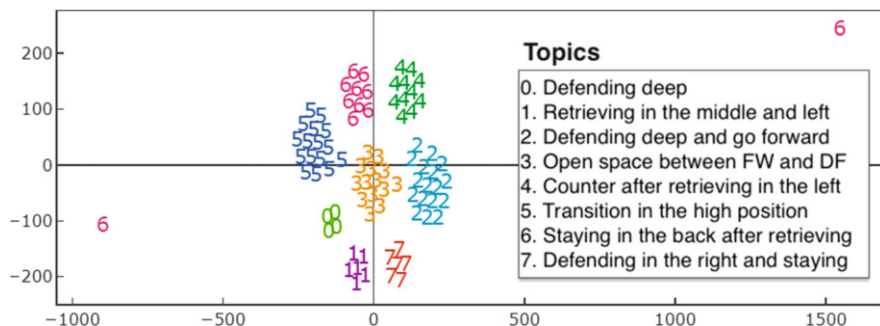
Even though individual people perceive landscapes and their attributed values differently, there are landscapes which the majority of people perceive as scenic and beautiful (Bell 2012). These prolific landscapes (e.g., Preikestolen in Norway or Wildkirchli in Switzerland) are often depicted by characteristic motif images, which are clusters of images all taken from a similar viewpoint and angle. Which landscapes become popular is driven by propagation of landscape or nature appreciation through travel guides or art from the romantic area, popularizing a selective subset of landscapes, thus not a new phenomenon. Today, tourism agencies and other influencers (e.g., celebrities, companies, movies, songs) can shape landscapes through social media promotion by planting seed images that people will try to recreate and, by doing so, form new motifs. By reaching millions of people and potentially influencing their future visiting plans, this social media-induced tourism can have drastic physical consequences on the local environment, infrastructure, and

people (Simmonds et al. 2018). Being able to monitor the spatiotemporal emergence of motifs as a proxy for induced tourism is crucial to support local decision-makers to tackle the potential increase in visitation rates to a respective landscape. In the paper by Hartmann et al. (2022), we created an operationalizable conceptual model of motifs that is able to identify, extract, and monitor prone landscapes based on geotagged social media data. More specifically, the proposed pipeline leverages creative-commons Flickr images from the YFCC100M dataset (Thomee et al. 2016) within the European Nature 2000-protected areas, which represent a network of breeding and resting sites within important landscapes for rare and threatened species. The core methodological process to identify motifs within a corpus of 2.1 million images involved two steps. Firstly, images were downsampled through spatial clustering by using Hierarchical Density-Based Clustering (HDBSCAN) (McInnes et al. 2017) since images belonging to the same motif were by definition in close proximity to one another. Secondly, with the help of the computer vision algorithm Scale-Invariant Feature Transform (SIFT) (Lowe 2004), we calculated image similarities between each image pair within a spatial HDBSCAN cluster and clustered them again based on that outcome. The results were our motifs, of which we found a total of 119 in our study sites across Europe. Analysis of the motifs revealed that 65% depict cultural elements such as castles and bridges, whereas the remaining 35% contain natural features that were biased toward coastal elements like cliffs. Ultimately, the early detection of emerging motifs and their monitoring allows the identification of locations subject to increased pressure, which enables managers to explore why sites are being visited and to take timely and appropriate actions (e.g., allocation of infrastructure such as toilets and rubbish disposals or visitor routing).

Not only descriptive textual information and emojis can be used for the analysis of geosocial media data, but it is also possible to use the image information directly. As an application for urban bicycle infrastructure planning, an object recognition algorithm based on convolutional neural networks was used to identify bicycles and potential parking spaces. The research and development work was carried out as a cooperation of a Young Research Group within the framework of the priority program VGIscience (Knura et al. 2021; Zahtila and Knura 2022). The research on object recognition was carried out in the COVMAP project (see Chap. 5); the processing of social media data and the development of methods for visual analysis were realized by the projects EVA-VGI (this chapter) and TOVIP (see Chap. 10).

### ***12.5.2 Exploring People's Mobility Behavior***

In search for a workable approach to analysis of multiple behavior episodes characterized by multivariate time series, Shirato et al. (2021) made an attempt to apply topic modelling. For this purpose, the patterns of variation of different attributes, or features, are represented by symbolic codes, which can be treated as words. The expected role of topic modelling is to reveal co-occurrences of such

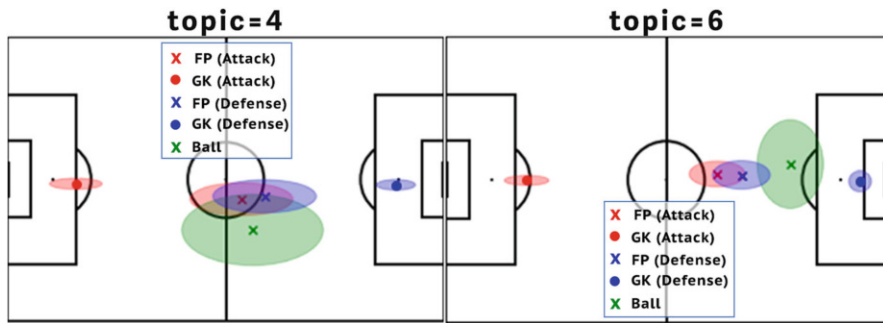


**Fig. 12.9** Episodes from a football match are represented by dots in a 2D projection space obtained by applying t-SNE to the vectors of the topic weights. The dot colors represent the dominant topics

patterns. To transform a time series of numeric values into a symbolic code, the values are compared to the median of the time series and encoded by symbols ‘-’, ‘=’, or ‘+’ depending on their relative position with respect to the median. Hence, each episode can be represented by a combination of the variation codes of the multiple variables. A method for probabilistic topic modelling, such as Latent Dirichlet Allocation (LDA), is applied to these combinations, which are treated as “texts” where the variation codes are “words.” The resulting “topics” show which variation patterns of different variables tend to occur together in the same episodes (see Fig. 12.9). The topic modelling method also assigns vectors of topic probabilities to the episodes. Using these vectors, the episodes can be clustered and/or arranged in a projection space according to similarities of the variation patterns and further explored by means of various existing methods.

The approach was tested using football data as an example. The features that were involved in the analysis reflected widths of empty spaces between team players on different levels of separation from the goal that is under attack. The topics corresponded to combinations of dynamic patterns of the changes of the widths on the different layers. To support interpretation of the results of topic modelling, the representation of the topics in the form of a table was combined with a map of the football pitch, where the behavior patterns were summarized as the average positions and areas of movement of the team centers (see Fig. 12.10).

The experiment showed that application of topic modelling to episodes characterized by multivariate (and, possibly, multifaceted) data has potential for behavior analysis. The research in this direction is worth being continued.



**Fig. 12.10** Distributions of centroids after transitions for topic 4 and topic 6. Colors represent the objects (blue, defense; red, attack; green, ball). The cross (x) means the average position of centroids, and the ellipse means the standard deviation of centroids for each topic

## 12.6 Conclusion

The research approaches presented in the chapter were designed for the visual analysis of social media posts and trajectories of football players to study people's reactions and behavior. As a starting point, a conceptual behavioral model is introduced to describe an actor in a certain period of time with regard to its spatial, thematic-attributive changes under the influence of events and external context factors. For the analysis, behavior is broken down into episodes, which refer to short periods of time and support the description with simple patterns. In order to model the external influences on behavior, a context model is also proposed, which, in addition to spatial, temporal, and attributive influences, also takes into account events that have taken place over a period of time and the behavior of other people.

Based on the conceptual model, generic methods are presented, which allow analyzing the behavior. The temporal query approach enables the investigation within in individual episodes. Time curves were used for a holistic analysis of several attributes with regard to their development over time. An extension of the time curve is proposed by using continuous coloring derived from a dimensional reduced attribute space. A method developed in the project for the visual analysis of text messages in social media data uses a "co-bridge" metaphor. Significant keywords are connected over time to enable both a quantitative comparison with regard to common and different keywords and a qualitative analysis with regard to stream volume and keyword frequencies. Furthermore, the use of emojis as contextual indicators was examined. It has been found that emojis are suitable for describing the spatial context of people in terms of perceived objects and activities taking place at these locations. Various normalization methods such as Chi-Value, "typicality," and metrics such as "user count per day" were used to account for heterogeneity and bias in social media data.

Two main areas of application were considered, on the one hand, the perception of environmental phenomena as input for questions of urban and landscape planning and, on the other hand, the analysis of the movement behavior of people on the example of football games. The study of a global phenomenon such as sunrise and sunset based on Instagram and Flickr allows conclusions to be drawn about consistency and reproducibility, as well as motives why the events were documented. Based on the four-facet context model, sunset and sunrise responses are examined for the where, who, what, and when. In addition, responses from different groups are compared with the aim of quantifying differences in behavior and information spread. As a second field of application, the movement behavior of people was analyzed using visitor statistics for selected locations during the COVID-19 pandemic and movement data from football games. Therefore, time-dependent abstracted and aggregated representations are created to compare collective behavior patterns during the pandemic in different countries in the first case and different football teams or teams in different stages of the game in the second case.

**Acknowledgments** This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within Priority Research Program 1894 *Volunteered Geographic Information: Interpretation, Visualization and Social Computing* (VGIScience, EVA-VGI, BU 2605/8-2) and the Swiss National Science Foundation (Project No 200021E-166788).

## References

- Acheson E, De Sabbata S, Purves RS (2017) A quantitative analysis of global gazetteers: patterns of coverage for common feature types. *Comput Environ Urban Syst* 64:309–320
- American Psychological Association. Behavior, n.d. <https://dictionary.apa.org/behavior>. visited on 2012-03-20
- Andrienko G, Andrienko N, Anzer G, Bauer P, Budziak G, Fuchs G, Hecker D, Weber H, Wrobel S (2021a) Constructing spaces and times for tactical analysis in football. *IEEE Trans Vis Comput Graph* 27(4):2280–2297. <https://doi.org/10.1109/TVCG.2019.2952129>
- Andrienko N, Andrienko G, Miksch S, Schumann H, Wrobel S (2021b) A theoretical model for pattern discovery in visual analytics. *Visual Inform* 5(1):23–42. ISSN 2468-502X. <https://doi.org/10.1016/j.visinf.2020.12.002>
- Bach B, Shi C, Heulot N, Madhyastha T, Grabowski T, Dragicevic P (2016) Time curves: folding time to visualize patterns of temporal evolution in data. *IEEE Trans Vis Comput Graph* 22(1):559–568. <https://doi.org/10.1109/TVCG.2015.2467851>
- Bai Q, Dan Q, Mu Z, Yang M (2019) A systematic review of emoji: current research and future perspectives. *Front Psychol* 10:2221. ISSN 1664-1078. <https://doi.org/10.3389/fpsyg.2019.02221>
- Bell S (2012) *Landscape: pattern, perception and process*. ISBN 9780203120088. <https://doi.org/10.4324/9780203120088>
- Blasi DE, Henrich J, Adamou E, Kemmerer D, Majid A (2022) Over-reliance on English hinders cognitive science. *Trends Cogn Sci* 26(12):1153–1170
- Burghardt D, Dunkel A (2022) Ethical analysis of geosocial data to balance social and individual interests. In: *Proceedings, AutoCarto 2022*

- Chen S, Andrienko N, Andrienko G, Li J, Yuan X (2021) Co-bridges: pair-wise visual connection and comparison for multi-item data streams. *IEEE Trans Vis Comput Graph* 27(2):1612–1622. <https://doi.org/10.1109/TVCG.2020.3030411>
- Clarke K, Wood J, Dykes J, Slingsby A (2007) Interactive visual exploration of a large spatio-temporal dataset: reflections on a geovisualization mashup. *IEEE Trans Vis Comput Graph* 13(6):1176–1183. <https://doi.org/10.1109/TVCG.2007.70570>
- Collins C, Andrienko N, Schreck T, Yang J, Choo J, Engelke U, Jena A, Dwyer T (2018) Guidance in the human–machine analytics process. *Visual Inform* 2(3):166–180. ISSN 2468-502X. <https://doi.org/10.1016/j.visinf.2018.09.003>
- Crawford K, Finn M (2015) The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal* 80(4):491–502
- Boyd D, Crawford K (2012) Critical questions for big data. *Inform Commun Soc* 15(5):662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Davis JD (2016) The four types of social media engagement. In: *The Union Metrics Blog*.
- Duarte R, Araújo D, Correia V, Davids K (2012) Sports teams as superorganisms. *Sports Med* 42(8):633–642
- Dunkel A, Andrienko G, Andrienko N, Burghardt D, Hauthal E, Purves R (2019) A conceptual framework for studying collective reactions to events in location-based social media. *Int J Geograph Inform Sci* 33(4):780–804. <https://doi.org/10.1080/13658816.2018.1546390>
- Dunkel A, Löchner M, Burghardt D (2020) Privacy-aware visualization of volunteered geographic information (VGI) to analyze spatial activity: a benchmark implementation. *ISPRS Int J Geo-Inf* 9 (10). ISSN 2220-9964. <https://doi.org/10.3390/ijgi9100607>
- Dunkel A, Hartmann MC, Hauthal E, Burghardt D, Purves RS (2023a) From sunrise to sunset: exploring landscape preference through global reactions to ephemeral events captured in georeferenced social media. *PLoS ONE*, 17(1). <https://doi.org/10.1371/journal.pone.0280423>
- Dunkel A, Burghardt D, Hartmann M, Ross P, Eva H (2023b) Supplementary materials for the publication “From sunrise to sunset: exploring landscape preference through global reactions to ephemeral events captured in georeferenced social media”
- Fonseca S, Milho J, Travassos B, Araújo D, Lopes A (2013) Measuring spatial interaction behavior in team sports using superimposed voronoi diagrams. *Int J Perform Anal Sport* 13(1):179–189. <https://doi.org/10.1080/24748668.2013.11868640>
- Gardner Z, Mooney P, De Sabbata S, Dowthwaite L (2020) Quantifying gendered participation in openstreetmap: responding to theories of female (under) representation in crowdsourced mapping. *GeoJournal* 85(6):1603–1620
- Ghermandi A, Sinclair M (2019) Passive crowdsourcing of social media in environmental research: a systematic map. *Global Environ Change* 55:36–47. ISSN 0959-3780. <https://doi.org/10.1016/j.gloenvcha.2019.02.003>
- Goodchild MF (2007) Citizens as voluntary sensors: spatial data infrastructure in the world of web 2.0. *Int J Spatial Data Infrastruct Res* 2(2):24–32.
- Gruebner O, Sykora M, Lowe SR, Shankardass K, Galea S, Subramanian S (2017) Big data opportunities for social behavioral and mental health research. *Soc Sci Med* 189:167–169. ISSN 0277-9536. <https://doi.org/10.1016/j.socscimed.2017.07.018>
- Gudmundsson J, Horton M (2017) Spatio-temporal analysis of team sports. *ACM Comput Surv* 50(2). ISSN 0360-0300. <https://doi.org/10.1145/3054132>
- Hartmann MC, Koblet O, Baer MF, Purves RS (2022) Automated motif identification: analysing flickr images to identify popular viewpoints in Europe’s protected areas. *J Outdoor Recreat Tourism* 37:100479. ISSN 2213-0780. <https://doi.org/10.1016/j.jort.2021.100479>
- Hauthal E, Burghardt D, Dunkel A (2019) Analyzing and visualizing emotional reactions expressed by emojis in location-based social media. *ISPRS Int J Geo-Inf* 8(3). ISSN 2220-9964. <https://doi.org/10.3390/ijgi8030113>
- Hauthal E, Dunkel A, Burghardt D (2021) Emojis as contextual indicants in location-based social media posts. *ISPRS Int J Geo-Inf* 10(6). ISSN 2220-9964. <https://doi.org/10.3390/ijgi10060407>

- Henriques G, Michalski J (2020) Defining behavior and its relationship to the science of psychology. *Integr Psychol Behav Sci* 54:328–353. <https://doi.org/10.1007/s12124-019-09504-4>
- Highfield T, Leaver T (2016) Instagrammatics and digital methods: studying visual social media, from selfies and gifs to memes and emoji. *Commun Res Practice* 2(1):47–62. <https://doi.org/10.1080/22041451.2016.1155332>
- Hu X, Tang J, Gao H, Liu H (2013) Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, New York, NY, USA, 2013. Association for Computing Machinery, pp 607–618. ISBN 9781450320351. <https://doi.org/10.1145/2488388.2488442>
- Kejriwal M, Wang Q, Li H, Wang L (2021) An empirical study of emoji usage on twitter in linguistic and national contexts. *Online Soc Netw Media* 24:100149. ISSN 2468-6964. <https://doi.org/10.1016/j.osnem.2021.100149>
- Knura M, Kluger F, Zahtila M, Schiewe J, Rosenhahn B, Burghardt D (2021) Using object detection on social media images for urban bicycle infrastructure planning: a case study of dresden. *ISPRS Int J Geo-Inf* 10(11). ISSN 2220-9964. <https://doi.org/10.3390/ijgi10110733>
- Kralj Novak P, Smailović J, Sluban B, Mozetič I (2015) Sentiment of emojis. *PLOS ONE* 10(12):1–22. <https://doi.org/10.1371/journal.pone.0144296>
- Krasnova H, Veltri NF (2011) Behind the curtains of privacy calculus on social networking sites: the study of Germany and the USA
- Laraway S, Snyckerski S, Pradhan S, Huitema BE (2019) An overview of scientific reproducibility: consideration of relevant issues for behavior science/analysis. *Perspect Behav Sci* 42(1):33–57. ISSN 25208977. <https://doi.org/10.1007/s40614-019-00193-3>
- Li J, Chen S, Chen W, Andrienko G, Andrienko N (2020) Semantics-space-time cube: a conceptual framework for systematic analysis of texts in space and time. *IEEE Trans Vis Comput Graph* 26(4):1789–1806. <https://doi.org/10.1109/TVCG.2018.2882449>
- Li L, Goodchild MF, Xu B (2013) Spatial, temporal, and socioeconomic patterns in the use of twitter and flickr. *Cartogr Geograph Inform Sci* 40(2):61–77
- Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
- Luckmann T (2013) *Theorie des sozialen Handelns*. De Gruyter, Berlin, Boston. ISBN 9783110848922. <https://doi.org/doi:10.1515/9783110848922>
- McInnes L, Healy J, Astels S (2017) hdbSCAN: hierarchical density based clustering. *J Open Source Softw* 2(11):205
- Memmert D, Lemmink KA, Sampaio J (2017) Current approaches to tactical performance analyses in soccer using position data. *Sports Med*. <https://doi.org/10.1007/s40279-016-0562-5>
- Olteanu A, Castillo C, Diaz F, Kiciman E (2019) Social data: biases, methodological pitfalls, and ethical boundaries. *Front Big Data* 2:13
- Shaver P, Schwartz J, Kirson D, O'connor C (1987) Emotion knowledge: further exploration of a prototype approach. *J Pers Soc Psychol* 52(6):1061
- Shirato G, Andrienko N, Andrienko G (2021) What are the topics in football? Extracting time-series topics from game episodes. In: 2021 IEEE Visualization Conference. <http://geoanalytics.net/and/papers/vis21poster.pdf>
- Simmonds C, McGivney A, Reilly P, Maffly B, Wilkinson T, Cannon G (2018) Crisis in our national parks: How tourists are loving nature to death. *The Guardian*. <https://www.theguardian.com/environment/2018/nov/20/national-parks-america-overcrowding-crisis-tourism-visitation-solutions>
- Solove DJ (2013) Privacy self-management and the consent dilemma' (2013). 126 *Harvard Law Review* 1880, 2012–141
- Teitler BE, Lieberman MD, Panozzo D, Sankaranarayanan J, Samet H, Sperling J (2008) Newsstand: a new view on news. In: *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '08*, New York, NY, USA, 2008. ACM, pp 18:1–18:10–18:1–18:10. ISBN 978-1-60558-323-5. <https://doi.org/10.1145/1463434.1463458>



- Thomee B, Shamma DA, Friedland G, Elizalde B, Ni K, Poland D, Borth D, Li, L-J (2016) YFCC100m. *Commun ACM* 59(2):64–73. <https://doi.org/10.1145/2812802>
- van den Elzen S, Holten D, Blaas J, van Wijk JJ (2016) Reducing snapshots to points: a visual analytics approach to dynamic network exploration. *IEEE Trans Vis Comput Graph* 22(1):1–10. <https://doi.org/10.1109/TVCG.2015.2468078>
- Williams ML, Burnap P, Sloan L (2017) Towards an ethical framework for publishing twitter data in social research: taking into account users' views, online context and algorithmic estimation. *Sociology* 51(6):1149–1168. <https://doi.org/10.1177/0038038517708140>
- Wood SA, Guerry AD, Silver JM, Lacayo M (2013) Using social media to quantify nature-based tourism and recreation. *Sci Rep* 3(1):1–7. <https://doi.org/10.1038/srep02976>
- Zahtila M, Knura M (2022) Visualizing point density on geometry objects: application in an urban area using social media VGI. *KN-J Cartogr Geograph Inform* 72(3):187–200. <https://doi.org/10.1007/s42489-022-00113-7>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

