Bird's-Eye - Large-Scale Visual Analytics of City Dynamics using Social Location Data

Robert Krueger[†1], Qi Han[†2], Nikolay Ivanov[2], Sanae Mahtal[2], Dennis Thom[2], Hanspeter Pfister[1], Thomas Ertl[2]

[1]Harvard University, Visual Computing Group, Cambridge, USA
[2]University of Stuttgart, Institute for Visualization and Interactive Systems, Stuttgart, Germany

Abstract

The analysis of behavioral city dynamics, such as temporal patterns of visited places and citizens' mobility routines, is an essential task for urban and transportation planning. Social media applications such as Foursquare and Twitter provide access to large-scale and up-to-date dynamic movement data that not only help to understand the social life and pulse of a city but also to maintain and improve urban infrastructure. However, the fast growth rate of this data poses challenges for conventional methods to provide up-to-date, flexible analysis. Therefore, planning authorities barely consider it. We present a system and design study to leverage social media data that assist urban and transportation planners to achieve better monitoring and analysis of city dynamics such as visited places and mobility patterns in large metropolitan areas. We conducted a goal-and-task analysis with urban planning experts. To address these goals, we designed a system with a scalable data monitoring back-end and an interactive visual analytics interface. The monitoring component uses intelligent pre-aggregation to allow dynamic queries in near real-time. The visual analytics interface leverages unsupervised learning to reveal clusters, routines, and unusual behavior in massive data, allowing to understand patterns in time and space. We evaluated our approach based on a qualitative user study with urban planning experts which demonstrates that intuitive integration of advanced analytical tools with visual interfaces is pivotal in making behavioral city dynamics accessible to practitioners. Our interviews also revealed areas for future research.

CCS Concepts

• ***Human-centered computing*** → *Geographic visualization;* • ***Information Search and Retrieval*** → *Information Filtering;*

1. Introduction

Today, more than 50% of the human population lives in cities [Nat14]. The United Nations predicts that by 2050, 66% of the world's population will be urban. The metropolitan area of Tokyo, for example, is home to around 38 mio. people. To maintain city infrastructures functional, and to keep the urban communities vibrant, ongoing monitoring, analysis, and planning is a crucial task for urban and transportation planners, as today's demands, lifestyle, and needs of the inhabitants are changing rapidly.

Having relied on manual street survey data and administration reports for a long time, tracking technology has made automated data-driven analysis possible [GR12, HL11]. City planners have now the opportunity to leverage different sensor-based data, comprising online surveys and trip records from communication and transportation authorities. In addition, data from Location-Based Social Networks (LBSN), e.g., Foursquare[‡] and Twitter[‡], contains semantic information about visited places and events. Such data sources may not be fully reliable because of noise and incompleteness, but contain valuable information about urban activities to help characterize the behavior of citizens. Although there are some visual analytic approaches that leverage these data, they are still rarely applied in actual use. This is not only due to the challenges to cope with the large data volumes and data velocities, but also because of gaps between what researchers provide and what planners need.
To address these issues, we conducted a design study and developed an expert system for planners that leverages social location data for monitoring and analysis of city dynamics. We approached it in three main stages. First, we interviewed urban planners to derive goals towards a data-driven approach for better monitoring and awareness of urban infrastructure and mobility patterns (1). From these goals, we derived analytic tasks and realized them in our system called *Bird's-Eye* (2). The system's monitoring components allow recording and storing large-scale social location data for different metropolitan areas over long time ranges. These data can be analyzed through an interactive visual interface integrating unsupervised machine learning. Geographic and temporal visualizations reveal routine behavior and outliers of visited places and mobility patterns. To evaluate our approach, we carried out qualitative user studies and collected feedback from the planning experts through structured interviews (3).

† These authors contributed equally to this work.
‡ www.foursquare.com, www.twitter.com, accessed March 10, 2019

Figure 1: *Bird's-Eye's interactive visual interface consists of multiple coordinated views. Top: A geographic map can be overlayed with social location information, encoded with spatial clusters and tagmaps. A mobility graph shows transitions between clusters (here Tokyo). Bottom: Two different timeline components help to understand temporal patterns of visited places (per category) and the transitions in-between them. Right: The sidebar allows users to enable or disable categories, steer clustering and outlier detection, and to inspect locations in detail.*

2. Related Work

We first review the state-of-the-art in data analysis for urban and transportation monitoring and planning. We then summarize existing approaches for social location mining before we address recent techniques and systems for interactive and visual analysis.

2.1. Urban and Transport Planning

Urban planning has a long history. The field was established with a conference in New York in 1898 [Eri12]. Before then, architects, public health professionals, and social workers approached the questions of city planning in very different ways [Eri12]. The architects focused on the city as built environment. Public health professionals dealt with the city's infrastructure, e.g., considering the connection to diseases [Sno55] and social conditions. And social workers wanted to improve social life of the inhabitants, focusing on fresh air, light, and public transport.

Early planning and monitoring activities mostly relied on manual reports. People used printed maps and their own observations as a basis for their decision making. Nowadays, planning activities are largely digital. Geographic information systems such as ArcGIS [§] and MPO TMS [¶] have become state-of-the-art in urban and transport planning to deal with the increasingly large amounts of data gathered with tracking technology. Zheng et al. summarize concepts, methodologies, and application in the area of urban computing [ZCWY14] and discuss frequent challenges from a computer science perspective.

Recently, a wide range of approaches have applied visual analysis techniques to the realm of urban planning and mobility analysis. For example, Miranda et al. [MDL*17] illustrate how city dynamics can be visually investigated by city planners and human behavioral analysts through multiple temporal resolutions. To this end, they design a visual exploration framework that centers around the idea of an "urban pulse", a concept that captures temporal activity variation centered around urban locations. Focusing more on spatio-temporal correlations, Senaratne et al. [SMB*18] present a system that enables city planners to investigate urban dynamics based on cellular network data. As part of their approach, they introduce space-time prisms, an extension of space-time cubes [GAA04], in which uncertainties in the movement data are highlighted as 3D volumes. Finally, the Visual Analyzer for Urban Data (VAUD) [CHW*18] by Wei Chen et al. provides a combination of spatio-temporal visualizations with an enhanced visual query model to understand urban behavioral patterns. Similar to Senaratne et al. they also focus on cellular network data to extract the movement information and cross-correlate it with other available information such as social network and real estate data. They present three case studies as well as a user study that demonstrate how their approach is effective in analyzing traffic jams or regional lifestyle characteristics.

Besides explicitly collected data, more and more geographic information becomes available via social media. This data covers a rich variety of information, such as people's movements, opinions, and needs. Compared to cellular network data, social media is usually more accessible to urban planners in terms of timeliness, legal restrictions, and willingness of data providers to cooperate. However, the more advanced technical challenges to process and filter the data for the relevant information prevents many urban and transportation planners from making use of these resources. While there is some research carried out in this domain [HZU13, CSHS12], the practical use of social media data in urban planning is still rare.

§ www.arcgis.com, accessed March 10, 2019

¶ https://www.mp-objects.com, accessed March 10, 2019

2.2. Social Location Mining

The data mining community has also discovered social media as a rich geo-referenced data source containing patterns about activities, mobility, and opinions. Most work focuses on algorithms tailored to specific data and applications. Some projects include static visualizations to communicate results. Noulas et al. analyze Foursquare data focusing on temporal and categorical patterns such as check-in counts on weekdays and weekends [NSMP11]. Melia et al. [MSZB*12] also make use of Foursquare data, but investigate activity duration and visualize the results with static sequence views. Analyzing both spatial and temporal patterns in Twitter and Foursquare, Kling et al. [KP12] use a probabilistic topic model combined with static result visualizations. Assem et al. [AXBO16] argue that regions of activities change over time, and use social location data to analyze time dependencies of activity regions in cities. Preotiuc-Pietro and Cohn also analyze check-in behavior of Foursquare users [PPC13]. In addition to showing temporal changes of visitation counts for places, they use matrix visualizations to explore people transfers between locations.

2.3. Visual Analysis of Social Location Services

Because social media data can be quite complex, it is often not sufficient to compute a result for a specific spatio-temporal resolution and degree of aggregation. Exploratory analysis is needed to gain an overview and to drill down to interesting patterns at different aggregation levels. Kisilevich et al. [KKK*10] provide a visual analytics approach to understand activity patterns in Flickr and Panoramio data. While some approaches analyze the past, Max et al. [SSH12] describe how to build a real-time event recommendation system based on social location data to predict future activities. Similar to our design study, Thom et al. [TKE16] present an expert evaluation of their system for situational awareness using microblog messages. They, however, focus mostly on analyzing message contents to aid disaster management.

Other approaches focus on the analysis of transitions between check-in information and other spatio-temporal events. For example, Andrienko et al. [AA18] developed state-transition graphs to analyze first-order activity transitions derived from traffic data. A fast adaptive graph aggregation method to visualize geographic transitions is presented by Zhou et al. [ZMT*18]. Chen et al. [CYW*16], Krueger et al. [KTE15], and Von Landesberger et al. [VLBR*16] propose approaches to reconstruct, aggregate, and analyze movement traces from geo-referenced Sina Weibo and Twitter data. Zeng et al. [ZFA*17] propose a visual analytics approach for understanding the relationship between human mobility and points of interest. Similar to Krueger et al. [KTE15], they enrich transportation data with Points of Interest (POI) for understanding relationships between human movements and activity distributions. However, their approach is limited to daily check-in distributions for certain POIs. While most of these works focus on techniques and systems, Liu presents a designs-study using movement data for better advertisement planning [LWL*17].

Our project improves on the prior art with the following contributions. First, we developed a system to monitor, store, and pre-process vast social location data. Second, Bird's-Eye interactive interface allows the exploration of a large variety of spatio-temporal check-in and mobility patterns. It uses data mining approaches to guide analysts to routine and anomalous behavior. The analysis can be carried out for recorded data or in near real-time, i.e., with only minutes of delay. Third, we provide the results of a goal and task analysis and evaluations conducted with urban planning experts.

3. Goal and Task Analysis

To understand for what goals in urban and transportation planning the analysis of social location data would be useful, we contacted three experts from the urban planning department of the University of Stuttgart. All three experts had expertise in urban planning acquired through various research projects. Two of the experts also had extensive practical experience in urban planning from working in industry prior or during their academic career. Expert A (EA) is a senior member of the department. Her research focuses on urban renovation and sustainable cities. She works closely with the local government for improving the city's infrastructure. Expert B (EB) is also a senior member. His research focus is on studying the redeveloping process of cities in Europe, development of university campuses, and digitization in urban planning. Together with the ministry for science and arts, he develops blueprints for digital citizen participation. Expert C (EC) is a second year PhD student at the same institute. Her main focus is on the development and the social aspects of so-called urban villages to inform urban planning.

3.1. Goals

We introduced our idea of leveraging social location data from different services to the experts and asked them to fill out a survey. The survey contained a lists of goals derived from literature as well as from our own experience of using social location data in visual analytic tools. We then asked the experts to rank our initial goals by importance using a Likert scale, and to add additional goals they might have. The following list shows the resulting five main goals.

> G 1: Planning with Historic and Current Data
>
> To establish a more data-driven planning environment, experts need to be able to derive up-to-date and historic social location data for different time-spans. The data should contain information about places, functional zones, and citizen activities in a city of choice as a supportive basis for city planning.

> G 2: Understanding Functional Structures and Zones
>
> Our experts would benefit from additional information of functional districts and land-use patterns of a city. This includes awareness of developments such as urban renewals, businesses, shopping, living, and food districts, as well as long-term trend analysis on how these areas change over time.

> G 3: Understanding Visiting Dynamics of Places
>
> Besides the spatial distributions of a cities infrastructure, it is important to understand how citizens are using it. A goal is to gain information on where and when certain places and zones are visited in order to react faster to changes or emergencies.

G 4: Understanding Mobility Routines

Understanding mobility patterns of citizens can inform the improvement of infrastructures such as road systems and places to better fit people's routines and needs. A goal is thus to better understand movement and mobility behavior at different temporal granularity, such as daily, weekly, or seasonal.

G 5: Increasing Situational Awareness

Not everything is following a routine. Even when the infrastructure is functional on regular days, certain events such as festivals and construction can lead to significant changes of commuter routes and visited places. These events can exceed the infrastructure's normal capacity or make areas become orphaned. Hence, another goal is to establish better near real-time monitoring to increase situational awareness and support decision making.

3.2. Tasks

We developed the project goals together with the experts and collected in-depth feedback throughout the development process. To provide guidelines for an effective visual analytic system, we broke down these higher-level domain goals into eight lower-level tasks. In essence, we fulfilled the *translator* role put forth in the design study methodology by Sedlmair et al. [SMM12], requiring a comfort level with task abstractions in computer science.

Task 1: To get access to, analyze, and plan with data (G1) is the predecessor of every other goal. At the beginning of each analysis, an important task is thus to *quickly retrieve and visualize data* for a region and time span of interest. **Task 2 and 3:** In order to support goals G2-4, analysts need to be able to *gain overview* about the data and get *details-on-demand by zooming and filtering* to reveal information at finer levels of details for urban zones, places, and citizen movements. **Task 4:** G2 and G3 both involve point-based information, i.e., the location of places (G2), how often they are visited (G3), and how this data changes over time. Therefore, being able to *analyze spatio-temporal changes and correlations* is another important tasks. **Task 5:** With respect to the large volumes of the data, an effective system also has to support spatial and temporal *data aggregation at different scales* to avoid clutter and over-plotting to keep zones, places, and movement routines visible (G2-4). **Task 6:** To support goals G2 and G3, we identify *comparisons* as an essential task as the user may want to look into similarities and differences of places and functional zones. **Task 7:** In order to analyze mobility routines, the system has to *support analysis of people transitions in space and time* with respect to different areas, places, and place categories. **Task 8:** In goal G5, events can be seen as extra-ordinary spatio-temporal happenings. It is thus important to support the *detection of outliers and anomalies* in time and in space.

Tasks 2-6 mainly deal with point-based location data (places or geo-spatial areas) for a given time span, while tasks 7 and 8 also take into account transition data, i.e., movement data of people from location to location.

4. System Overview

To support these tasks, Bird's-Eye consists of a scalable data monitoring back-end and an interactive visual analytics interface. The monitoring component records social location data and uses intelligent pre-aggregation to enable fast dynamic queries. The visual analytics interface leverages unsupervised learning to reveal clusters, movement routines, and unusual behavior in massive data.

4.1. Scalable Data Monitoring Back-end

To meet tasks 1-8, the system provides information about functional zones and places, their visiting frequencies, as well as mobility patterns between them. Our monitoring system supports these tasks with a scalable backend, allowing to record and leverage data from different APIs, and pre-aggregate and store it as a basis for fast retrieval and visualization (T1).

Data: Bird's Eye makes use of Foursquare and Twitter data. Foursquare offers a social location service allowing people to virtually check-in at different places (called venues) with their smartphones. Moreover, Foursquare can make use of the data collective as a basis for recommendation services. A Foursquare check-in date contains at least the location of a place and it's category information at different levels of detail, such as "Restaurant–Italian Restaurant–Luigi's Pizzeria". With 50 mio. monthly active users, this sums up to a large dataset, containing places and check-in frequencies in many areas of the world.
Twitter is a microblog platform with 362 mio. monthly active users, allowing them to write and share short messages about their status, observation, and opinions. Messages (tweets) at least contain a timestamp, a user and tweet ID, and textual content. Some contain precise geo-coordinates or IDs and names of places. Many Foursquare users link their check-ins on Twitter to automatically create a corresponding tweet to reach a broader audience.

Monitoring: From Foursquare's API‖ we collect information about venues (places) and consecutive checked-in users (Figure 2, A). The *Venues Search* allows retrieving venues within a given radius together with real-time check-in information. The number of such requests per hour is, however, limited to 5,000, and each response will only contain up to 50 venues. Hence, there is a trade-off between spatial, temporal, and categorical resolution. For example, cycling requests for a whole city area through the 850 venue categories would usually lead to responses with less than 50 entries and a poor temporal resolution of 17 hours per venue. To increase this resolution, our configurable polling scheme makes use of Foursquare's filters and queries for higher-level categories only. This leads to updates every 12 seconds while still covering most venues.

Unfortunately, Foursquare API responses only provide information on how many people are currently checked in at a given place. This does not allow to trace users along different places they visited to derive and show mobility information (T7). Because many users have linked their Foursquare account with their Twitter account, we overcome this issue by collecting Foursquare-related messages

‖ https://developer.foursquare.com, accessed March 10, 2019

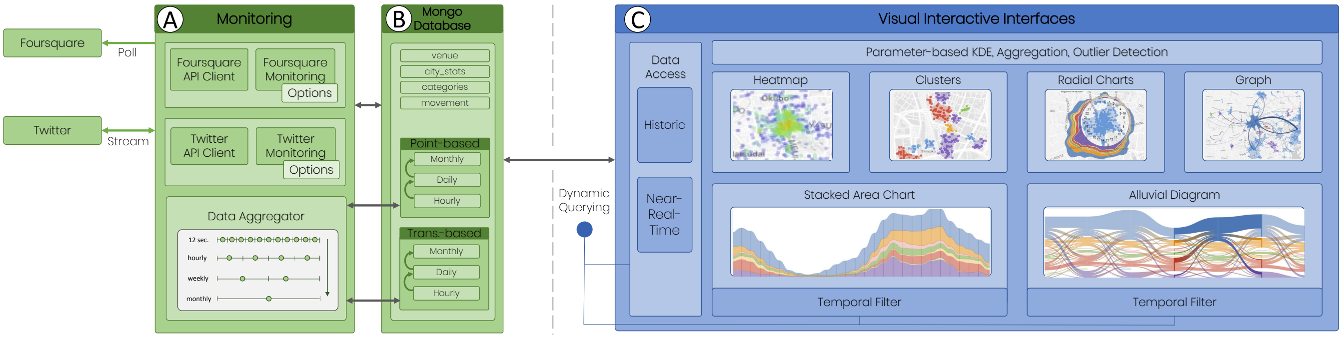


Figure 2: *Bird's-Eye's architecture consists of a monitoring component (A) that records and stores Foursquare and Twitter data in a No-SQL database (B). An autonomous aggregator routine accesses the data stored in the database in parallel and processes it into hourly, weekly, and monthly chunks for faster point-based and transition-based queries from the visual analysis components (C). The interface consists of interactively linked visualizations, including a map with different data layers, time series using stacked area chart and alluvial diagrams, as well as controllers for parameter steering for, e.g., clustering. Users can access historic or near real-time data. Once loaded, selections in the timeline trigger data requests that leverage the pre-aggregates in the database.*

from the Twitter Streaming API**. This allows linking of multiple venues and respective check-in times to a specific user, and provides our experts with information about citizens' travel patterns (T7). Although we only derive these transitions for a subset of users, they still amount to a reasonable sampling of the overall movement.

Storage: Storing and visualizing real-time streams of social media data requires fast insertion and retrieval operations on large numbers of entries. As we know the tasks the system needs to support, the queries and aggregation schemes are well defined and there is usually no need to update entries or make complicated ad-hoc joins among different properties. We, therefore, decided to use MongoDB (Figure 2, B), a document-oriented No-SQL database, instead of a relational database. Its flexibility and schema-less design support fast development iterations. Data from each poll is stored in the collection `city stats` where each document contains a timestamp, a list of venues, their respective categories, and people currently checked-in. Other collections store a list of known venues and categories, as well as users' movements with the fields userID, timestamp, previous, and current venue.

Aggregation and Retrieval: Our visual interface allows users to select an arbitrary time period they are interested in to explore places and activities in time and space at different levels of aggregation (T5). To this end, our back-end has to obtain the numbers of check-ins of any venue during the specified time. A naive method would first collect all venue statistics (names, check-in counts, category information) from the `city stats` collection that lie within this period and accumulate the numbers per venue. However, for periods that cover multiple days or weeks, this would lead to response times of several seconds or even minutes and break the interactive workflow.

Drawing on ideas from recent research [LJH13, LKS13], we developed a hierarchical pre-aggregation scheme to accelerate queries (see Figure 2). Each aggregation level is represented in one collection (monthly, daily, hourly), where each document stores the accumulated check-in numbers by venue for one respective time unit. The same process is performed for the transitional data. A pre-aggregation script operates directly on the raw data in the database without requiring programmatic integration with the interactive system components.

4.2. Interactive Visual Analytics Interface

To allow users to visually explore the data (T1) and reveal patterns, we developed an interactive visual analysis interface. It consists of multiple linked visualizations and configurable panels and tables (see Figure 1). Users can start an analysis by defining a new metropolitan area to monitor, or by querying historic data from the database (see Figure 1, main menu (top)). Accordingly, the system will load data for the pre-computed aggregation levels (e.g., weekly) and request lower level aggregations depending on the user's filter actions.

Timeline: The temporal distribution of check-ins by category (e.g., offices, shops, restaurants) (T4) can be visualized in a stacked

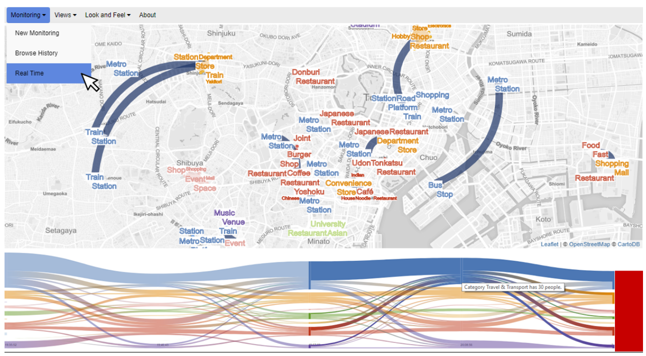


Figure 3: *The near real-time mode gives a configurable view over the last few hours with updates every 10 minutes. The full system functionality can be used to monitor events with information about check-in counts as well as mobility patterns.*

** https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter, accessed March 10, 2019

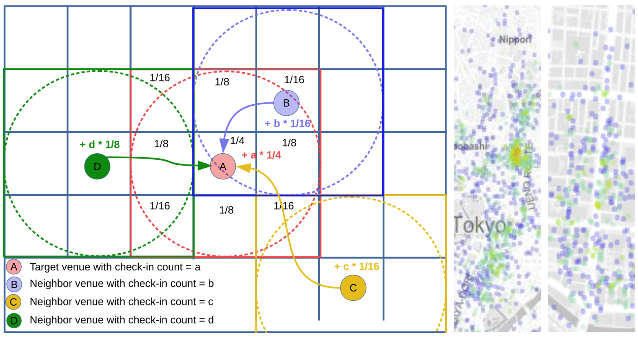


Figure 4: *Scattered Kernel Density Estimation: Venues (B,C,D) surrounding venue (A) contribute to its Gaussian splat by multiplying (A)'s grid value with the other venue's check-in counts.*

Figure 5: *Density-based clustering allows to detect functional zones (here food districts in Tokyo). On a city level, it gives an aggregated overview of the main areas (left). By zooming in, one can reveal smaller restaurant clusters (Tokyo-Akhibara, right image). To emphasize cohesiveness we compute concave hulls.*

area chart. Each ribbon maps a category and is color-coded accordingly. The visible categories can be selected and deselected. Initially, the chart renders all loaded data for several weeks or months (T2). The user can select a sub-range of the loaded time span to drill-down (T3). Each action synchronizes the visualizations in the map, supporting users to analyze spatio-temporal correlations (T4).

Map: The interactive map enables geographic analysis. It can be zoomed in and out and explored like in typical web-based GIS systems such as Google Maps (T2, T3). Multiple visualizations can be superimposed on the map. Individual places with check-in counts can be rendered as a dotted heatmap (see Figure 4, right)), or can be aggregated to clusters and tag maps, i.e. a visual representation of frequently used words directly on the map (see [TBK*12]) (T5). The user can call up radial area chart overlays (see Figure 6) to compare (T6) visiting frequencies of a cluster's places (check-ins) on an hourly, daily, or weekly basis. Users can analyze mobility patterns (T7) between clusters and investigate spatial outliers (T8).

Details and Controls: A panel (see Figure 1) allows users to request details of places and zones (T3). It contains a sorted list of venues in the current selection together with the respective number of checked-in users. A set of sliders allow users to interactively change parameters to affect the behaviors of the heatmap, clustering, outlier-detection, and graph-drawing. A color-coded list of the Foursquare top-level categories serves as a legend and can also be used to filter the data by categories. Users can select singular or multiple categories to be displayed in the timeline and map using mouse selection. A history panel allows to undo or redo the filtering operations.

4.3. Automated Analytics Components

An important component of our system consists of three interactive tools that are based on automated data mining methods. They support users in discovering hot-spots and clusters of functional zones (T2, T6) and identifying outliers at different scales (T8) that would otherwise be hidden by visual clutter and information overload. They also enable analysis of cross-correlation and cascading effects of the spatial and temporal aspects of city dynamics (T4). It was a design challenge to integrate these methods seamlessly with the visualizations and to provide a responsive and intuitive interaction experience. In the following, we will describe each analytics feature and its role in a typical workflow.

4.3.1. Check-in-based Kernel Density Estimation

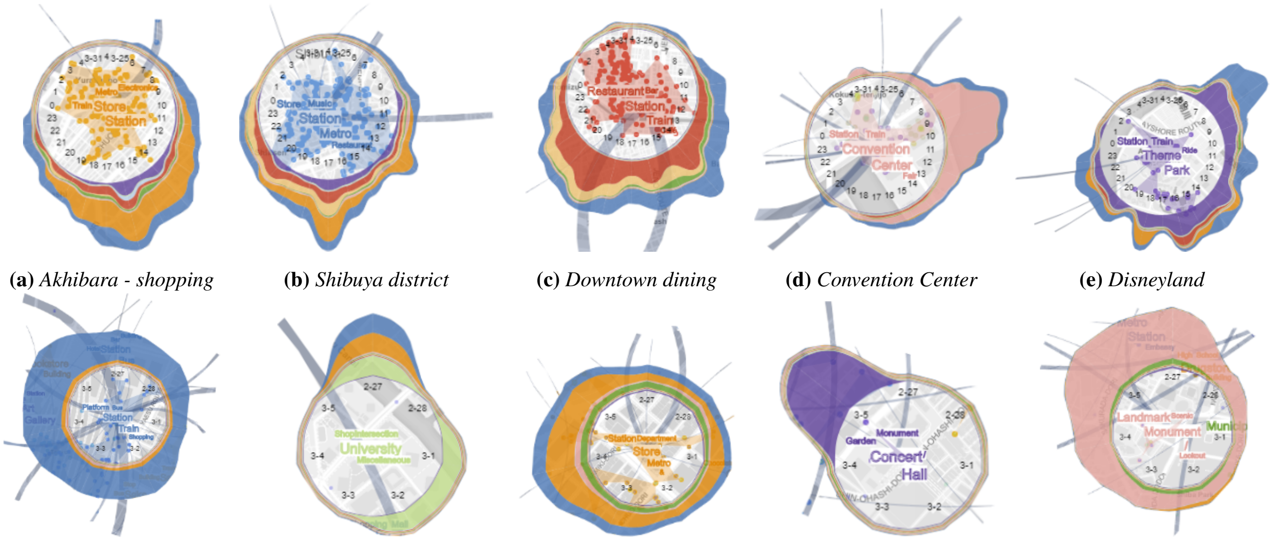
To quickly spot areas with a high number of POIs and activity counts, Bird's-Eye offers a heatmap that renders the general spatial distribution and corresponding check-in densities using a kernel-density-based splatting algorithm [MRH*10]. This is a pivotal element to achieve goals G2 and G3 and corresponding tasks (T2-6). The heatmap shows 'hotter' colors over areas of the map where more check-in activity can be observed during the selected temporal period.

In a typical splatting scheme, the map is first overlaid with a regular high-resolution grid. For each check-in, one would then apply a radial Gaussian kernel at the respective venue-location on the map. The values of subsequent splats are then accumulated within the cells. Finally, each grid cell would be assigned a color by interpolating along a color palette. Such regular splatting assumes an independent distribution of points over the grid. However, in our case, multiple points (check-ins) all sit at the same location (a venue). Also, venues in a city are located at pre-defined discrete structures (i.e., inside buildings). Visualizing a smooth density field can thus be misleading.

Instead, we use a customized splatting scheme to determine the 'heat' value. To calculate values around a given venue (A), we assign a radial Gaussian splat with fixed radius on top of it. We then iterate through all grid cells covered by the splat and test whether other venues (B,C,D) are covered by the splat radius. If we find such a venue, the value of the Gaussian at the center of that cell is multiplied with the number of check-ins for that venue and added to the final heat-value of the target venue A. Finally, the color for the venue is determined by linear mapping of the 'heat' value to a color scale.

4.3.2. Cluster-based Activity Aggregation

While the heatmap reveals busy areas, it is less helpful in uncovering the hierarchical spatial structures of cities ranging from regions to



(a) *Akhibara - shopping* **(b)** *Shibuya district* **(c)** *Downtown dining* **(d)** *Convention Center* **(e)** *Disneyland*

(f) *Tokyo Station* **(g)** *Shibaura University* **(h)** *Gina - shopping* **(i)** *Concert Hall* **(j)** *Tokyo Tower*

Figure 6: *Clusters can be overlayed with radial area charts showing hourly (top) or weekly (bottom) patterns of check-in counts of the clusters' places. (a) Akhibara, an electronic shopping mile, and (b) Shibuya, an entertainment area, are mostly visited in the afternoon. (c) Near downtown, there are some restaurant districts, most busy during lunch and dinner time. (d) The Convention center is frequented in the morning. (e) Disneyland has many visitors all day. (f) Tokyo station, the city's busy metro station. (g) Shibaura University is less visited during the weekend. This does not account for (h) Gina, an elegant shopping area, and (i) Shuntory concert mall, mostly visited on Saturday and Sunday. (j) The Tokyo tower, a landmark, is most busy during the weekend but also during weekdays. Colormap: Travel & Transport, Shop & Services, Residence, Professional & other Places, Outdoors & Recreation, Nightlife Spots, Food, Event, College & University, Arts & Entertainment.*

blocks and single buildings. Similarly, one can group a city into functional geographic zones of shopping and dining areas, or businesses and residential areas, which are often decoupled from its administrative organization. To reveal these data-driven and up-to-date substructures (T5), our aggregation needs to be adaptive temporally and spatially up to the level of detail investigated by the user. Our algorithm performs the following steps to identify and visualize diverse activity regions on the map.

First, we run DBSCAN clustering [EKS*96] to find groups of spatially related venues (visited places) that match the natural breaks in the geographic distribution. The parameters of the algorithm, which define whether points are considered as neighbors and whether a point sits in a densely distributed neighborhood, can be interactively adapted by the user. Unlike centroid-based algorithms, such as K-Means, DBSCAN can find clusters of varying shapes while ignoring overall noise, two properties that are frequently observed in our location data. To enable fast distance calculations, our implementation uses a spatial index (quadtree) to quickly retrieve all possible neighbors of a given point. We run fast recalculations of the clustering during zooming and panning (see Figure 5) to aggregate different levels of granularity.

To visually indicate the DBSCAN clusters, we determine their concave hulls using an algorithm by Emil Rosén et al. [RJB14]. The polygon for the concave hull is placed as a shaded area in a map overlay on top of the venues. We color-code the clusters by the most frequently represented category in the cluster.

For each cluster, we create a small tag cloud from the underlying venues to give a quick indication about their contents. We first create a list of all words that occur in the venue names and categories. Recurring words (e.g. 'restaurant') are counted and ordered according to their prominence, and the top five words are ultimately selected. These five words are then sized according to their relative prominence and placed using a particle-based layout described by Luboschik et al. [LSC08].

The detected clusters are especially useful as a handle to identify and select meaningful city areas, learn more about their time-dependent popularity, and allow comparisons between them (T6). Clicking on the clusters reveals a radial area chart, similar to work by Zheng et al. [ZFA*17], that shows check-in category activities over time for the venues in this cluster depending on the time-range selected in the temporal overview. Figure 6 shows different examples of these radial area charts with hourly (top row) and weekly (bottom row) aggregation.

4.3.3. Adaptive Outlier Detection

In addition to understanding the current and regular behavior of citizens, the experts also need to identify unusual and unexpected situations (T8). Our outlier detection method (T8) finds unusually high or low numbers of check-ins for top-level venue categories that deviate from the usual. In this case, the 'usual' case is calculated at each temporal aggregation level (hourly, daily, monthly) as the mean over samples per venue category of one respective unit of time. An outlier is detected if the current value deviates at least n times

Figure 7: *Zoom-adaptive clustering to explore mobility patterns to, from, and in Tokyo's Disneyland from overview (left) to detail (right).*

the standard deviation from the mean, where n is a value that can be interactively adapted by the user.

We designed the visualization of outliers using visual cues in both the temporal and spatial views. Since detected outliers can be considered among the most important pieces of information communicated to the user, we choose a visual property that supports pre-attentive perception [HE12]. In the temporal area chart, anomalous sections are marked by red boundaries around the temporal segments of the respective category ribbon. In the map view, clusters which are dominated by a category with anomalous check-in numbers are also marked with red boundaries. Mouse-hovering over the outlier-boundaries provides additional information about the underlying statistics, such as mean and standard deviation of check-in counts and user movement between categories.

4.4. Features for Mobility Analysis

In order to analyze mobility patterns of citizens (T7), we enhance our visual interface with an alluvial diagram and a graph visualization as an additional map layer.

Alluvial Diagram: To explore temporal transitions of Foursquare users between categories in time (T7), the stacked area chart (see Section 4.2) can be changed to an alluvial diagram (see Figure 8). Each category is shown as a ribbon with color corresponding to the category and thickness indicating the current popularity of the category. The temporal transition of people between categories is illustrated by splitting and merging the different ribbons along the horizontal axis. Depending on the time span in focus, the diagram shows weekly, daily, hourly, or 10-minute-based transitions, leveraging the pre-aggregation structures in the back-end (see Section 4.1).

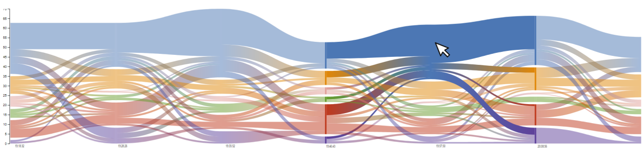


Figure 8: *The alluvial diagram shows citizens' transitions between places using color-encoded categories. Hovering over the ribbons highlights incoming and outgoing flows (here blue ribbon).*

Mobility Graph: To reveal geographic mobility patterns we overlay our map with a node-link diagram (see Figure 7). To overcome the clutter and overplotting that would occur when drawing individual transitions between places, we make use of a zoom-adaptive spatial aggregation and visualize transitions as curved lines in-between clusters. The line thickness corresponds to the relative amount of movement in that direction. The transition direction is visualized by animating the rendering from the origin to the destination point.

4.5. Example Case Study: Event Analysis

To make the contributions of Bird's-Eye more concrete, we describe a motivating case study about event analysis. According to our experts and recent urban geographic literature, analyzing *movement flows* and *behavior maps* during events and peak times is of great interest for planners to understand how infrastructure and places are accepted and utilized by citizens [Mar11], to identify shortcomings and plan future events [STW*16, MERM11] and emergency interventions [TKE16], and to possibly prevent future traffic congestion [LDJW*03, KDMN14]. For our first case study, we conducted an analysis of the public event "THE IDOLM@STER MILLION LIVE! 4thLIVE TH@NK YOU for SMILE", a three-day music concert that took place on March 10-12, 2017, in Tokyo's Nippon Budokan (Chiyoda) area.

Our scenario assumes that a city planner is interested in using Bird's-Eye to analyze the situation during events of type "Arts and entertainment", which includes sport games, music concerts, theater plays, and similar activities (see Figure 9). She starts the analysis with a general overview of the whole recorded area and step by steps zooms in to focus on Tokyo's inner city. In order to retrieve semantics about the most visited places, she turns on the clustering and corresponding tag map. The clusters show the main functional zones in the inner city. In order to get more fine-grained information but still maintain an overview, she slightly adapts the clustering parameters to detect smaller clusters. In addition, she uses the category panel to filter for arts and entertainment as well as the event category. This reveals many events in the city such as concerts, sports events,

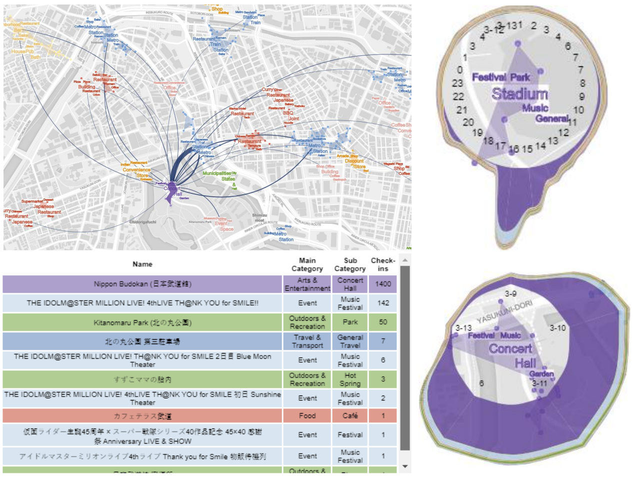


Figure 9: *In focus is a three-day music festival in Nippon Budokan, an indoor arena. The map shows the cluster (in purple) and incoming and outgoing transitions. The table gives details about places within the cluster. Radial stacked area charts show the hourly (top) and daily (bottom) visiting distribution.*

and exhibitions. To look for events that may cause higher traffic, she turns on the outlier detection. The outliers show up in the timeline and map views, respectively. She zooms in and detects a festival taking place in Nippon Budokan. The cluster is superimposed with tags such as "concert" and "music". She clicks on that cluster and gets information of all places in the interactive table (Figure 9, bottom left). To derive more information about the mobility patterns surrounding this event, she turns other categories back on and activates the mobility graph overlay. By hovering over the festivals venue cluster, which contains different stages and surrounding festival activities, the overlaying graph shows incoming and outgoing arcs to other small clusters (Figure 9, top left). This indicates Takebashi station to be a transportation hub to the event. Other arcs show transit from and to the event's parking space, restaurants, and shops. Finally, she investigates the temporal visiting distributions using the radial chart (Figure 9, right). It shows that the festival happened over the weekend with visitors arriving on Friday afternoon and leaving on Monday morning. Switching to the hourly distribution shows that the festival was most crowded between 2 and 5 pm.

5. Evaluation

To evaluate Bird's-Eye, we conducted a user study with novice users and interviews with our three domain experts.

5.1. Dataset

The data we used for the evaluation focuses on the densely populated central metropolitan area in Tokyo. Tokyo is one of the largest cities in the world, has high social media usage, and potentially covers many typical places and citizen activities. Our participants are unfamiliar with the city, allowing for an unbiased evaluation of Bird's-Eye's capabilities. The monitored area centers around Tokyo central station with a radius of 30 kilometers and covers the whole inner city as well as some surroundings. The time of the collected data ranges from February 23 to April 15, 2017, with a temporal resolution of 6 minutes. Overall we captured 98,318 venues and 487,721 transitions between them.

5.2. User Study

Before we evaluated the system with planning experts, with whom we worked closely on the system development, we conducted a user study by collecting quantitative and qualitative feedback from novice users with *'fresh eyes'* in order to evaluate UI intuitiveness and interactions and to discover and fix usability issues. We focused on insights about the usefulness and ease of use of individual features as well as the system as a whole. The study participants were required to complete both low-level tasks, for which we measured and compared quantitative task performance, and high-level tasks, for which we analyzed comments from recorded think-aloud sessions. We initially formulated the following research questions as a basis for our study design:

- Q1: Are the features of our approach intuitive and easy to use?
- Q2: Are the features useful in understanding the spatial and temporal distribution of the social location data in cities?
- Q3: Are the clustering and aggregation mechanisms helpful for the participants in discovering the mobility structure of the city?
- Q4: Can the outlier detection feature help users in identifying abnormal activity situations?

We first discuss the study design and procedure before reporting task performance results and the findings of the think-aloud study.

5.2.1. Study Design

The experiment was conducted under controlled lab conditions and isolated from any external distractions. In every session, we made an audio recording and asked the users to think aloud about observations, ideas, and intents they experienced during the session. We also made visual recordings using screen capture software.

Participants: The experiment was conducted with 10 participants, 7 female, and 3 male, all between 19 and 25 years. They were recruited by promoting the study around the University of Stuttgart campus and each was compensated with 10 Euros. Before the actual analysis session, participants were instructed to fill out a questionnaire about their age, gender, and occupational background. The questionnaire also asked about their familiarity with concepts related to our research using Likert-scales. Most of the participants indicated that they were familiar with "Data Visualization" and "Computer Science". However, most were only moderately familiar with concepts like "Machine Learning" or "Clustering". Most participants were not or slightly familiar with "Foursquare". For the study we assigned each participants to one of two groups (G1 and G2).

Preparation: The sessions began with a brief tutorial where the experimenter explained the functions of Bird's-Eye. For this tutorial we used the first week of the Tokyo data, which was then removed for the actual analysis sessions. The participants had the chance to work with the system and to ask questions for five minutes. To test if participants had achieved a basic understanding, we asked them to find a routine activity of a working day. They all completed this task in less than five minutes without major problems.

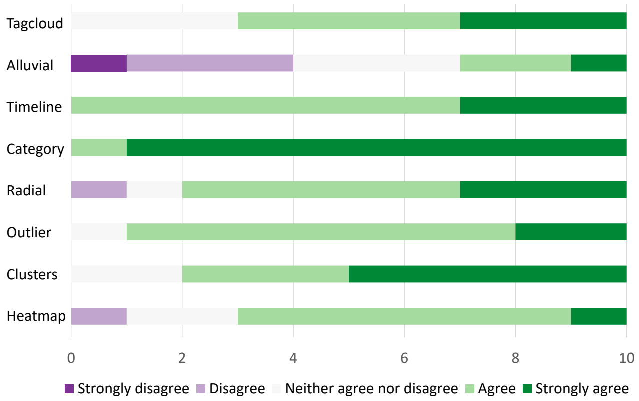


Figure 10: *Usefulness of the individual features. The chart illustrates the number of different answers to the question "The following feature was helpful in solving the tasks" from "strongly disagree" to "strongly agree."*

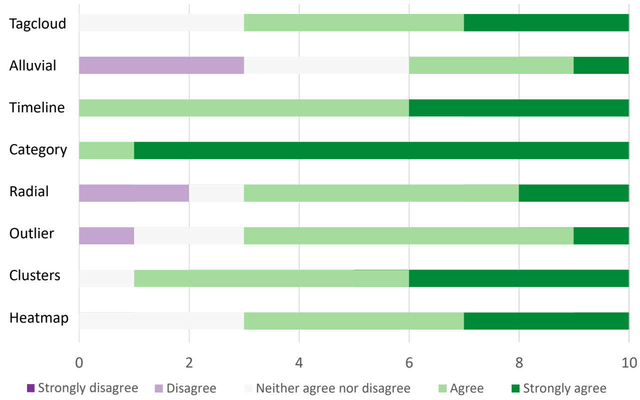


Figure 11: *Ease of use of the features. The chart illustrates the number of different answers to the question "The following feature was easy and intuitive to use" ranging from "strongly disagree" to "strongly agree."*

Analysis Sessions: In the actual analysis sessions, the participants were asked to address two groups of tasks. The first group comprised three low-level tasks, which focused on clustering and outlier detection. The first task (LT1) asked to identify three popular shopping areas on a weekend. For the second task (LT2), participants had to identify one day that stood out among three given days. We measured the task completion time for each participant. Only the participants in group G1 were provided with the clustering and outlier detection features, while group G2 had to solve the same tasks with only the basic features.

The more complex high-level tasks concerned actual scenario descriptions where participants had to freely explore the data to reach a conclusion. For the first high-level task (HT1), participants had to find an optimal place for a conference organized by the city council considering certain boundary conditions. For the second high-level task (HT2), participants were instructed to explore mobility behavior of citizens before and after a popular event and to identify possible problems, unusual patterns, and outliers. All participants managed to finish the low- and high-level tasks in the given time limit of 20 minutes each.

Wrapup: Following the analysis sessions, the participants had to complete a Likert-style questionnaire about the usefulness and ease-of-use of each individual system feature. They were also asked to provide informal feedback and comments about the system as a whole, and to provide suggestions for improvement. Generally, a study session was completed in 60 minutes (with introduction, questionnaires). After the study session, we transcribed the think-aloud records and looked through all video recording of users' interactions.

5.2.2. Results

Figures 10 and 11 show an overview of the collected user feedback. The former presents the usefulness of the features in order to solve the tasks. The latter focuses on usability (ease of use). Most features were considered easy and intuitive to use. Users especially liked the category filters and temporal overview. No participant found the clustering and the tag clouds particularly difficult to understand. Seven out of the ten participants considered the outlier detection and the radial graph feature easy to use. Feedback on the alluvial graph was diverse, with some users finding it less useful than others.

Similar answers were collected regarding the usefulness of the individual features (see Figure 10 for details). All participants agreed that the system is useful for exploring activities in cities. The fourth participant (P4) stated, "the system is really useful in detecting trends, summarizing and showing a big amount of data." P8 commented, "the system helps to explore human activities in different time and different categories." Both P5 and P6 added that the system could be very helpful for finding optimal locations for public events. Participants also suggested adding a multi-selection feature in the timeline view. P2 pointed out that the system is "still a little buggy." However, he also commented that he "had fun to work with the system." For the low-level tasks, we measured the following task completion times. For LT1, the mean completion time for group G1 was 121.2 seconds (sd = 40s), and the mean completion time for group G2 was 142.6*s* (sd = 23). For LT2, the mean completion time for group G1 was 193.4 seconds (sd = 70), and the mean completion time for group G2 was 275.2 seconds (sd = 123s).

5.3. Expert Feedback

After collecting feedback from novice users, we approached to our three planning experts (see Section 3) and conducted semi-structured interviews with them to evaluate our system. The interview comprised three phases: In *the pre-presentation phase*, we started by asking them about their educational degree, their occupation, their research focus, the application area of their research, and other stakeholders of their research projects. We then asked about their previous experience with interactive software in general, their experience with GIS systems, as well as with data visualization tools. We also asked them about typical data sources used in their domain, whether they have used social media data before, and whether they see potential in using it. During *the presentation* phase, we first gave the experts a brief introduction to Bird's-Eye. Then we showcased the analytical ability of the system using a similar use case as described in Section 4.5. Afterward, we asked the experts to explore the data with the system and to ask questions. Finally, in *the post-presentation phase*, we collected their feedback. We asked them to assess the applicability of Bird's-Eye to their work and to provide suggestions for improvements and recommendations for new features. The complete interview took about one hour.

5.3.1. Comments and Results

The expert's comments and suggestions can roughly be grouped into the following three categories.

Social media data is valuable for urban planning research. All three experts mentioned that they mainly used street survey data, government reports, and other general GIS data in their research, but no social media data. The main reason is that they feel it would be difficult to obtain, manage, filter, and aggregate such data. Furthermore, they also have concerns about obtaining the data legally, for example, to adhere to personal data privacy laws. A third major concern is that social media data only represents a subgroup of the population, which might introduce biases.

However, if these concerns can be addressed or reduced, the experts considered the crowd-sourced social media data as highly interesting for their work. EA and EB suggested that social media data could also be used complementary to conventional methods. EB said, "For example, citizen initiative projects usually have elder participants. Social media data, on the other hand, can represent the younger generation better." EC told us that social media data could be very useful in the study of normal everyday life and behavioral routines. EA also mentioned that this could be especially useful for understanding the fast development of cities in East Asia.

Bird's-Eye is very relevant for their research. All experts think that Bird's-Eye is useful and interesting for their work. EC was eager to know if we also had collected data for Beijing, as it would help her in her current project. EA and EB commented that the inter-connection and synergy effects of combining different tools in one system are most important. EA said, "The clustering, the zoom, and the temporal aspect together give a very good overview of the current state of the city." EB and EC added that the system gives a good overview of behavioral city dynamics. EC mentioned that using the system, she could easily identify the secondary city centers of Tokyo and observe the different activity patterns of them. In addition, EB and EC liked the clustering feature a lot. EB pointed out that clustering using morphological features could be useful in identifying dynamic district boundaries through human activity pattern instead of using predefined static boundaries. EC said that clustering together with the radial area chart could be very helpful in understanding the context of urban villages and in building hypotheses for understanding their development. EC also liked the outlier detection feature: "It can guide me in finding occurrences like big festivals. Then I can use the system to analyze the population behavior on those special occasions."

In addition to analysis, EA and EB saw great potential in using Bird's-Eye for estimating, monitoring, assessing, and communicating the effects of temporary events, such as urban construction efforts. EB said, "For example, there are normally estimates over where the traffics goes in case one street is blocked because of different reasons, e.g., construction projects. However, nobody knows really where the traffic goes to."

EB also highlighted that in case of controversial topics regarding urban planning the system can be a powerful tool for showing the real effects to the stakeholders. He gave one example regarding the reduction of parking spots in a city area. The typical assumption is that reducing parking spots would reduce the number of customers in the local shops. However, the contrary belief is that people feel more relaxed in areas with less traffic, which might, in turn, lead to more shopping and higher revenues. Our system could be used to analyze such changes and to reveal their effects.

Limitations and suggestions: The experts also made comments about the limitations of our approach and gave suggestions for further development. EC would like to have the ability to merge some similar categories because sometimes more detailed separation of activities does not provide more information. For example, she could image to merge the category *Food* and the category *Nightlife* for some of her analysis. Both EA and EB suggested that showing more comprehensive properties of the transitions between the clusters could help them to make more inference over the mobility pattern. In addition, using other visual encodings for the direction of movement data could be more effective. Both EA and EC wished to have additional user profile data in the system. This would allow them to analyze the activities of particular social groups among the residents.

One major concern was how representative the social media data is in general. The demographic distribution of social media users will likely be different from the demographics of the city population overall. Furthermore, the motivations for the users to submit check-ins could bring biases into the data. However, other research suggests that aggregation of larger amounts of social media data might still reflect overall activity patterns [NSMP11, KSB*16] and data can be merged and cross-validated with other sources [CDH17]. In addition, awareness of this bias can help analysts to make proper use of the data and still come to valid conclusions.

6. Conclusions

We presented a design study on how to make use of social location data to assist urban and transportation planners. We carried out a goal-and-task analysis and developed Bird's-Eye, a system to assist planning experts. We carried out a set of evaluations: user studies and expert interviews. The results demonstrate that intuitive integration of advanced analytical tools with visual interfaces is important in making behavioral city dynamics accessible to practitioners.

From expert interviews, we got several suggestions on how our system can be extended. We also would like to extend Bird's-Eye with data provenance capabilities to assist the decision-making process. In addition, we got some recommendations on how social media data can aid city and transportation planning in general. A major concern of our experts was how representative such data are considering the demographics and amount of users that participate in different areas. A challenge for future research is thus to evaluate social media data by comparing it to ground truth that could be collected through, e.g., traffic fleets or surveillance cameras.

7. Acknowledgements

This work was partially funded by the DFG SPP 1894 project Volunteered Geographic Information (VGI), the EU project Cimplex (641191), a Google Faculty Grant, and the NSF grant IIS-1447344. We would like to thank all our collaborators.

References

[AA18] ANDRIENKO N., ANDRIENKO G.: State transition graphs for semantic analysis of movement behaviours. *Information Visualization 17*, 1 (2018), 41–65. 3

[AXBO16] ASSEM H., XU L., BUDA T. S., O'SULLIVAN D.: Spatio-temporal clustering approach for detecting functional regions in cities. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)* (Nov 2016), pp. 370–377. 3

[CDH17] CEBEILLAC A., DAUDÉ É., HURAUX T.: Where? when? and how often? what can we learn about daily urban mobilities from Twitter data and Google POIs in Bangkok (Thailand) and which perspectives for dengue studies? *Netcom. Réseaux, communication et territoires*, 31-3/4 (2017), 283–308. 11

[CHW*18] CHEN W., HUANG Z., WU F., ZHU M., GUAN H., MACIEJEWSKI R.: Vaud: A visual analysis approach for exploring spatio-temporal urban data. *IEEE Transactions on Visualization & Computer Graphics*, 9 (2018), 2636–2648. 2

[CSHS12] CRANSHAW J., SCHWARTZ R., HONG J., SADEH N.: The livehoods project: Utilizing social media to understand the dynamics of a city. In *Sixth International AAAI Conference on Weblogs and Social Media* (2012). 2

[CYW*16] CHEN S., YUAN X., WANG Z., GUO C., LIANG J., WANG Z., ZHANG X. L., ZHANG J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE transactions on visualization and computer graphics 22*, 1 (2016), 270–279. 3

[EKS*96] ESTER M., KRIEGEL H.-P., SANDER J., XU X., ET AL.: A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (1996), vol. 96, pp. 226–231. 7

[Eri12] ERICKSON A.: A brief history of the birth of urban planning. 2

[GAA04] GATALSKY P., ANDRIENKO N., ANDRIENKO G.: Interactive analysis of event data using space-time cube. pp. 145– 152. 2

[GR12] GANTZ J., REINSEL D.: The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future* (2012), 1–16. 1

[HE12] HEALEY C., ENNS J.: Attention and visual memory in visualization and computer graphics. *IEEE transactions on visualization and computer graphics 18*, 7 (2012), 1170–1188. 8

[HL11] HILBERT M., LÓPEZ P.: The world's technological capacity to store, communicate, and compute information. *science 332*, 6025 (2011), 60–65. 1

[HZU13] HASAN S., ZHAN X., UKKUSURI S. V.: Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *UrbComp@KDD* (2013). 2

[KDMN14] KWOCZEK S., DI MARTINO S., NEJDL W.: Predicting and visualizing traffic congestion in the presence of planned special events. *Journal of Visual Languages & Computing 25*, 6 (2014), 973–980. 8

[KKK*10] KISILEVICH S., KRSTAJIC M., KEIM D., ANDRIENKO N., ANDRIENKO G.: Event-based analysis of people's activities and behavior using flickr and panoramio geotagged photo collections. In *Information visualisation (IV), 2010 14th international conference* (2010), IEEE, pp. 289–296. 3

[KP12] KLING F., POZDNOUKHOV A.: When a city tells a story: urban topic analysis. In *Proceedings of the 20th international conference on advances in geographic information systems* (2012), ACM, pp. 482–485. 3

[KSB*16] KRUEGER R., SUN G., BECK F., LIANG R., ERTL T.: TravelDiff: Visual comparison analytics for massive movement patterns derived from Twitter. In *Pacific Visualization Symposium (PacificVis), 2016 IEEE* (2016), IEEE, pp. 176–183. 11

[KTE15] KRUEGER R., THOM D., ERTL T.: Semantic enrichment of movement behavior with Foursquare–a visual analytics approach. *IEEE transactions on visualization and computer graphics 21*, 8 (2015), 903–915. 3

[LDJW*03] LATOSKI S. P., DUNN JR W. M., WAGENBLAST B., RANDALL J., WALKER M. D.: *Managing travel for planned special events*. Tech. rep., 2003. 8

[LJH13] LIU Z., JIANG B., HEER J.: immens: Real-time visual querying of big data. *Computer Graphics Forum 32*, 3pt4 (2013), 421–430. 5

[LKS13] LINS L., KLOSOWSKI J. T., SCHEIDEGGER C.: Nanocubes for real-time exploration of spatiotemporal datasets. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2456–2465. 5

[LSC08] LUBOSCHIK M., SCHUMANN H., CORDS H.: Particle-based labeling: Fast point-feature labeling without obscuring other visual features. *IEEE transactions on visualization and computer graphics 14*, 6 (2008), 1237–1244. 7

[LWL*17] LIU D., WENG D., LI Y., BAO J., ZHENG Y., QU H., WU Y.: Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE transactions on visualization and computer graphics 23*, 1 (2017), 1–10. 3

[Mar11] MARUŠIĆ B. G.: Analysis of patterns of spatial occupancy in urban open space using behaviour maps and gis. *Urban design international 16*, 1 (2011), 36–50. 8

[MDL*17] MIRANDA F., DORAISWAMY H., LAGE M., ZHAO K., GONÇALVES B., WILSON L., HSIEH M., SILVA C. T.: Urban pulse: Capturing the rhythm of cities. *IEEE transactions on visualization and computer graphics 23*, 1 (2017), 791–800. 2

[MERM11] MA S.-C., EGAN D., ROTHERHAM I., MA S.-M.: A framework for monitoring during the planning stage for a sports mega-event. *Journal of Sustainable Tourism 19*, 1 (2011), 79–96. 8

[MRH*10] MACIEJEWSKI R., RUDOLPH S., HAFEN R., ABUSALAH A., YAKOUT M., OUZZANI M., CLEVELAND W. S., GRANNIS S. J., EBERT D. S.: A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics 16*, 2 (2010), 205–220. 6

[MSZB*12] MELIÀ-SEGUÍ J., ZHANG R., BART E., PRICE B., BRDICZKA O.: Activity duration analysis for context-aware services using foursquare check-ins. In *Proceedings of the 2012 international workshop on Self-aware internet of things* (2012), ACM, pp. 13–18. 3

[Nat14] NATIONS U.: World urbanization prospects. 1

[NSMP11] NOULAS A., SCELLATO S., MASCOLO C., PONTIL M.: An empirical study of geographic user activity patterns in foursquare. *ICwSM 11*, 70-573 (2011), 2. 3, 11

[PPC13] PREOŢIUC-PIETRO D., COHN T.: Mining user behaviours: a study of check-in patterns in location based social networks. In *Proceedings of the 5th Annual ACM Web Science Conference* (2013), ACM, pp. 306–315. 3

[RJB14] ROSEN E., JANSSON E., BRUNDIN M.: Implementation of a fast and efficient concave hull algorithm. *Uppsala Univ* (2014). 7

[SMB*18] SENARATNE H., MUELLER M., BEHRISCH M., LALANNE F., BUSTOS-JIMÉNEZ J., SCHNEIDEWIND J., KEIM D., SCHRECK T.: Urban mobility analysis with mobile network data: A visual analytics approach. *IEEE Transactions on Intelligent Transportation Systems 19*, 5 (2018), 1537–1546. 2

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics 18*, 12 (2012), 2431–2440. 4

[Sno55] SNOW J.: *On the mode of communication of cholera*. John Churchill, 1855. 2

[SSH12] SKLAR M., SHAW B., HOGUE A.: Recommending interesting events in real-time with foursquare check-ins. In *Proceedings of the Sixth ACM Conference on Recommender Systems* (New York, NY, USA, 2012), RecSys '12, ACM, pp. 311–312. 3

[STW*16] SALAZAR N. B., TIMMERMAN C., WETS J., GATO L. G., VAN DEN BROUCKE S.: *Mega-event mobilities: A critical analysis*. Taylor & Francis, 2016. 8

[TBK*12] THOM D., BOSCH H., KOCH S., WÖRNER M., ERTL T.: Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Visualization Symposium (PacificVis), 2012 IEEE Pacific* (2012), IEEE, pp. 41–48. 6

[TKE16] THOM D., KRÜGER R., ERTL T.: Can twitter save lives? a broad-scale study on visual social media analytics for public safety. *IEEE transactions on visualization and computer graphics 22*, 7 (2016), 1816–1829. 3, 8

[VLBR*16] VON LANDESBERGER T., BRODKORB F., ROSKOSCH P., ANDRIENKO N., ANDRIENKO G., KERREN A.: Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE transactions on visualization and computer graphics 22*, 1 (2016), 11–20. 3

[ZCWY14] ZHENG Y., CAPRA L., WOLFSON O., YANG H.: Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST) 5*, 3 (2014), 38. 2

[ZFA*17] ZENG W., FU C.-W., ARISONA S. M., SCHUBIGER S., BURKHARD R., MA K.-L.: Visualizing the relationship between human mobility and points of interest. *IEEE Transactions on Intelligent Transportation Systems 18*, 8 (2017), 2271–2284. 3, 7

[ZMT*18] ZHOU Z., MENG L., TANG C., ZHAO Y., GUO Z., HU M., CHEN W.: Visual abstraction of large scale geospatial origin-destination movement data. *IEEE transactions on visualization and computer graphics* (2018). 3