CLEVis: A Semantic Driven Visual Analytics System for Community Level Events

Chao Ma
Kent State University
Ye Zhao
Kent State University
Andrew Curtis
Case Western Reserve University
Farah Kamw
Concordia University
Shamal AL-Dohuki
University of Duhok
Jing Yang
UNC Charlotte
Suphanut Jamonnak
Kent State University
Ismael Ali
University of Zakho

Abstract: Community-level event (CLE) datasets, such as police reports of crime events, contain abundant semantic information of event situations and descriptions in a geospatial-temporal context. They are critical for frontline users, such as police officers and social workers, to discover and examine insights about community neighborhoods. We propose CLEVis, a neighborhood visual analytics system for CLE datasets, to help frontline users explore events for insights at community regions of interest (CROIs), namely fine-grained geographical resolutions such as small neighborhoods around local restaurants, churches, and schools. CLEVis fully utilizes semantic information by integrating automatic algorithms and interactive visualizations. The design and development of CLEVis are conducted with solid collaborations with real world community workers and social scientists. Case studies and user feedback are presented with real world datasets and applications.

Keywords: Community Level Events, Semantic Information, Visual Analytics

INTRODUCTION

A community-level event (CLE), such as a crime or a traffic accident, is usually recorded and characterized by time, location, type of event, persons involved and a semantic description. A CLE dataset, such as records of crimes/accidents from a police department, other emergency call logs (e.g., 911 calls), hospital electronic records or other mortality or morbidity surveillance, contain rich information that may help community planners, service providers, and residents monitor community functioning, identify critical events, and plan changes that can enhance people's living experience. Visually intuitive tools can help frontline users, such as police officers, health officers, and social workers in a local community, with decision making by more effectively utilizing their CLE datasets for *neighborhood visual analytics*. These frontline users are often interested in analyzing CLEs around points of interest (POIs) with a fine granularity in both space and time.

Different visual analytics (VA) approaches for spatial-temporal event data are available to reveal trends and patterns with aggregated data, such as by Zip Code, city, or county. While a few projects utilize ontology and categories of events in the analysis, there is a disconnect between these method utility and frontline users, such as police officers in the field. Through collaboration with such users it became evident that there was a need for frontline users to be able to explore community events for insights about POIs at finer geographic scales, and with ability to mine text for additional contextual detail.

To fill this need we developed CLEVis, a VA system that helps frontline users harness semantically rich CLE data. CLEVis allows users to perform a visual analysis with specific domain knowledge and ontologies, using a Community Region of Interest (CROI) approach to represent a small neighborhood area (usually less than 10 road segments) with aggregated event patterns. Automated algorithms and interactive visualizations are developed to conduct a variety of VA tasks through a process of CROI discovery and examination. CLEVis is designed, developed, and evaluated together with typical end-users and domain related academic researchers providing input and evaluation.

RELATED WORK

There have been many different approaches to visualizing spatiotemporal event data. GeoTime visualization combines geospatial and temporal displays in 3-D view [15]. Maciejewski et al. [12] discover hot spots by kernel-based aggregations. They further explore a pandemic simulation model by visualizing aggregated events over geographical areas [11]. VisAdapt [7] visually links homes to climate change models, while there are no clustering and semantic information involved. Rasheed et al. [16] visualize a graph of crime events and use it to detect hidden criminal activities. Our method handles similar spatiotemporal event data, but unlike them, we support visual analysis in a local community for studying individual POIs, streets, and buildings. Therefore, our data mining algorithms and visualizations are designed in particular for the finer levels of detail.

HotSketch [3] allows police officers to sketch a path on a map to query crime locations along the path, which is overlaid on a point-based crime visualization. Jentner et al. [6] present a visual tool for criminal intelligence analysis, where multiple clustering algorithms are used to analyze similarities among crime cases, though the ability to apply more interactively spatial querying is limited. Malik et al. [13] propose a visualization system that allows users in law enforcement to refine crime data categories with clustering and forecasting algorithms. Their algorithms are mostly based on statistical values but not semantic information, and users can't drill down to POI levels and individual events. Razip et al. [17] present a mobile visualization system for patrol officers. The system provides situation awareness of a location by visualizing previous events in the surrounding area. While all these methods provide valuable insights, CLEVis combines context and space by automatically defining regions of interest based on semantic descriptions. It allows frontline users to define and generate spatiotemporal event clusters in multiple ways, allowing for the prioritizing of intervention using different criteria.

There have also been various text mining systems that use ontologies, categories, or patterns extracted using VA operations (Please see a survey of text related visualization by Liu et al. [9]). However, most of them are not specifically developed for frontline users to analyze community events. Utilizing semantic event information through interactive keyword-based operations is identified by our users as the most preferred function. CLEVis is designed based on this requirement.

Related to CLEVis ability to query geographic text, different VA tools have been developed for exploring geo-tagged social media data [18]. Chae et al. [4] present a VA approach to discovering abnormal topics and events from various social media data sources. MacEachren et al.

[10] provide a visual monitoring tool for situational awareness of Twitter data. Thom et al. [19] detect anomalies from local event reports and global media reactions. Miranda et al. [14] extract urban pulse over seminal public spaces. Cho et al. [2] discover Twitter events with interactive visualizations. The task of the VAST 2011 Grand Challenge is to investigate potential terrorist activities and their relation to the spread of an epidemic. Among many good solutions, Bertini et al. [1] analyze Twitter data to trace epidemics. Our preliminary work using Twitter and Facebook data in a small town of Ohio reveal limitations on using / relying on these data sources for this type of community. More specifically for medium to small size towns: (1) not a lot of social media data is available; (2) only a small number of social media data items are geo-tagged; (3) local residents may not post important information online. However, there are several other text-enriched datasets for these environments and these are the data inputs for CLEVis.

REQUIRMENT ANALYSIS AND SYSTEM OVERVIEW

Real-world CLE Datasets

Two real world CLE datasets have been utilized in the CLEVis prototype including:

New Orleans 911 records: The dataset has 6,521 emergency call records from Aug. 29, 2005 to Sep. 8, 2005 at New Orleans, Louisiana. This time range covers the period when Hurricane Katrina made landfall on the morning of Aug. 29 and caused significant flooding covering 80% of the city and leading to a final death toll of 1,464 people. Each record has the textual description translated from the record of a phone conversation. It also has the accurate call time, categorical situations, victim address, and phone number. The semantic information includes situational needs such as “they're trapped”, “there is no way out”. Investigating these data not only provides a template of needs for similar events, but can also provide insights to improve current response strategies.

Police reports: The dataset includes 45,196 police reports from a small town in Ohio. The reports were generated between Jan. 2013 and Aug. 2017. Each report provides a written description about the investigation of an incident such as assault, traffic, home violence, etc. Each report also contains latitude and longitude, date and time, street address, situation category, gender, race, and age of victim. The collaboration with the local Police Department led to the design of VA tools suitable for use with their data. This application aims to improve the social and security welfare of similar sized communities across the US.

Community Users and Requirements

Target users of CLEVis are anyone involved with local events, such as police officers, social workers, analysts, or city officials. Feedback from the Chief and officers from local area police departments through multiple face-to-face meetings occurred across a period of eight months. The prototype developed from these contacts was then evaluated using the New Orleans 911 call records. Several domain experts in social and geographical studies also participated in this project through regular meetings. These experts were involved in designing neighborhood scale intervention strategies and their expertise and experience helped the design and evaluation of CLEVis.

The following specific requirements of a VA system were identified as a result of these meetings:

- The system should be intuitive to understand and easy to use by frontline personnel, such as police officers investigating the environment around a bar, or patterns of local robberies, or health officers studying where best to place Narcan kits to help reduce overdose mortalities. The system should have simple-to-use automated functionality without over burdening the users.
- The system should support fine scale VA operations in both space and time, with the ability for querying and analyzing specific locations, or geographic names (such as a bar) in a neighborhood.
- The system should fully utilize available semantic information of events, such as text descriptions about what happened and semantic categories that group events into classes (e.g. Assault, Traffic, Theft). Semantic driven interactions should be provided for users to effectively distill information of interest.
- The system should be compatible with users' domain knowledge and ontologies (such as varying words used to describe illegal drugs). Patterns discovered from the visualizations should also be described in familiar terms for the users.

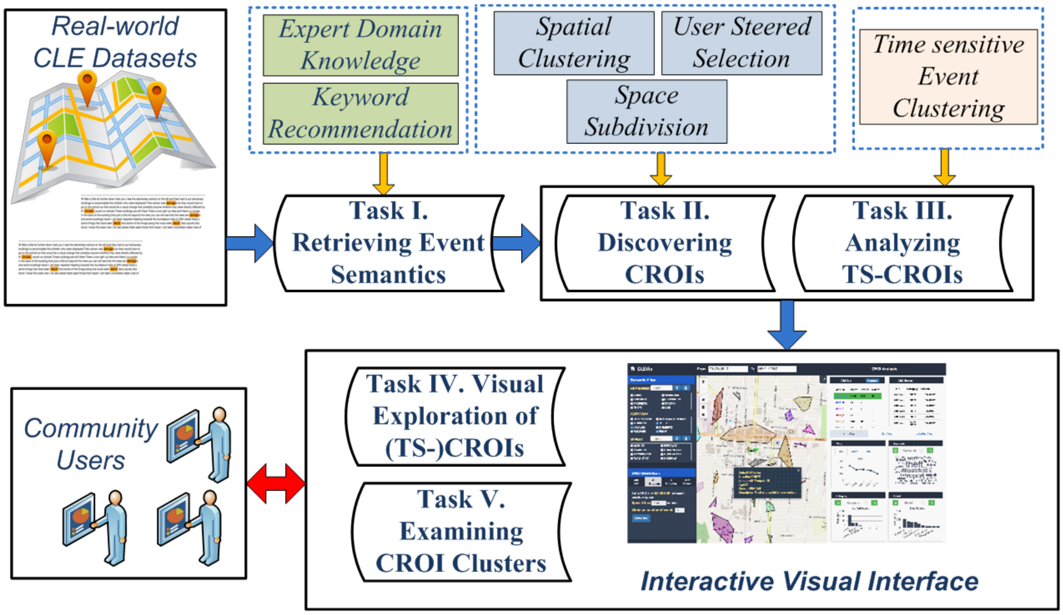


Figure 1: Processing pipeline of CLEVis system. Five VA tasks (Task I to Task V) are supported for community users.

CLEVis Design and VA Tasks

CLEVis employs automated algorithms and interactive visualizations for exploratory analysis. A Community Region of Interest (CROI) is used to represent a small neighborhood area with specific event patterns. Users are supported to retrieve events and discover CROIs with desired spatio-temporal and semantic information and examine whether they have similar or disparate patterns. Figure 1 illustrates the processing pipeline of CLEVis. The system supports a variety of VA tasks including:

Task I. Retrieving events with semantic information and domain knowledge: Semantic study utilizes text mining algorithms to automatically discover significant keywords from event semantics. Users can interactively select keywords to retrieve events of interest and construct CROIs. In addition, users can employ their domain specific "real world" ontology. For example, keywords such as "*weed", "spice", "dope", "Marijuana", "drug"* are all used to describe illegal drugs. Users can upload predefined keyword lists and then interactively select them as filter terms.

Task II. Discovering CROIs: A variety of approaches are provided for interactive CROI discovery including: (1) Density based spatial clustering; (2) Directly drawing on the map by domain knowledge and interest; and (3) Space subdivision by grids or geographical shapefiles.

Task III. Analyzing bursts of events: Event "bursts" often happen in a small area and in a short time period. They are of great interest for community workers. CLEVis uses a time sensitive event clustering algorithm to discover TS-CROIs. For example, one TS-CROI represents a neighborhood area with a set of flu events in Spring. Another TS-CROI refers to an area related to a food poisoning outbreak in the Fall. The two TS-CROIs overlap at a local school, where hygiene practices might be investigated, so in this case the temporal information of the two events is critical to separate them. To discover TS-CROIs, a 3D DBSCAN algorithm utilizing both spatial and time dimensions is used together with semantic filtering to cluster CLEs. As TS-CROIs may overlap in space, they are visualized by a 3D space-time cube view.

Task IV. Visually browsing and examining CROIs: CLEVis provides a rich set of coordinated views and interactions for CROI browsing and examination. A map view is presented where users can examine CROIs. Timelines, tag clouds, and distribution charts allow users to learn CROI details. Users can further select and compare the contents of interesting CROIs in a comparison view. Moreover, all CROIs in a city are clustered and projected to a 2D scatterplot view where similar/disparate CROIs are revealed.

Task V. Examining CROI Clusters: CROIs can be clustered by their semantic content for similarity analysis. Here, each CROI is represented by a CROI document which consists of the textual descriptions of all CLEs in this CROI. Then, the semantic similarity between a pair of CROIs are calculated by keyword (category)-based document similarity.

COMMUNITY REGIONS OF INTEREST

For a community worker, a typical task is to analyze a neighborhood space, which might be just a few blocks, houses, or buildings. This area might be a fixed, predefined region. For example, a police officer may want to analyze a series of car burglaries to discover crime-prone neighborhoods. Our major design goal is to allow users to easily detect and study such areas through CROIs. Most existing work conducts VA over coarser regions, such as zip codes, cities, or counties. This is not appropriate for many communities across the US, for example, the whole of our study town has a single zip code.

CLEVis allows for users to visually define CROIs (**Task II**) by: (1) Drawing a circular, rectangular or freestyle polygon to enclose a preferred area; (2) Loading a predefined shapefile (the typical map output from a GIS); (3) Dividing the map into grid cells; (4) Automatically discovering a region by aggregating CLEs by semantic information (e.g., where are the areas with excessive drug activities?). After identifying multiple CROIs, users can (1) visually examine the spatial, temporal, and semantic contents of the CROIs and their CLEs; (2) visualy compare CROIs to find patterns, trends and outliers; and (3) conduct semantic driven interactive explorations based on keyword and category filters.

Automatic CROI Generation

CLEs have spatial and temporal dimensions, together with semantic contents, and can be considered as a set of points. These CLE points are first filtered by semantic information, and then clustered to build multiple CROIs by density-based spatial clustering (DBSCAN) which groups together points that are geographically proximate. Here, the similarity between CLE points is computed by their spatial distance. Two parameters ε and *minPts* are set (with user input) to find clusters of a minimum number of *minPts* points, and with a maximum radius of the region of ε meters.

Time-Sensitive CROI

The aforementioned CROI generation methods cannot easily find event “bursts”, which refers to a series of related CLEs that happened within a small region in a short time period (**Task III**). A density-based spatio-temporal clustering algorithm is adopted and modified to discover TS-CROIs that represent such burst regions, where both spatial vicinity and temporal vicinity are considered. We adapt the ST-DBSCAN algorithm [20] to cluster CLEs in both spatial (x, y) and temporal (t) dimensions. A similarity between two CLE points, $\mathrm{A}(\mathrm{A}_x, \mathrm{A}_y, \mathrm{A}_t)$ and $\mathrm{B}(\mathrm{B}_x, \mathrm{B}_y, \mathrm{B}_t)$ is computed as:

$$s(\mathrm{A}, B) = \sqrt{w_1(A_x - B_x)^2 + w_1(A_y - B_y)^2 + w_2(A_t - B_t)^2}$$

Here, two coefficients $w_1$and w_2 are needed to combine space and time units into the similarity computation. Our method is different from ST-DBSCAN by defining a new way to setup the parameters, which allows for easy integration within the CLEVis system.

To improve usability, inputs available to users should be intuitive. So, we allow users to define two parameters by answering: (1) ε_s: what is the spatial size of a potential CROI in which event points can be grouped? (2) ε_t: what is the time frame (period) of a potential CROI in which event points can be clustered? Then, the similarity computation becomes

$$s(\mathrm{A}, B) = \sqrt{(\frac{(A_X - B_x)}{\varepsilon_s})^2 + (\frac{(A_y - B_y)}{\varepsilon_s})^2 + (\frac{(A_t - B_t)}{\varepsilon_t})^2}$$

This equation shows that we project both spatial distances and temporal distances to dimensionless quantities between 0 and 1, by utilizing user defined scales ε_s and ε_t. Then ε can be set as $\sqrt{3}/2$ to enclose a cube of equal edge size 1. Here only the parameter *minPts* needs to be set by users in addition to ε_s and ε_t. For example, the parameters *minPts* =5, ε_s =220 and ε_t = 6 can be explained to community users as "finding groups of at least 5 events which happened within a close distance of 220 meters, in a period of 6 months", which we have found can be understood easily by domain users.

SEMANTIC DRIVEN CLE STUDY

CLEs contain important textual insights so semantic retrieval (**Task I**) becomes a key focus of CLEVis, which differentiates the approach from other geospatial event systems. CLE text may also include slang or other informal or local descriptors. Users can visually select and filter CLEs through keywords for CROI studies due to their intuitive nature. Future systems will expand to also include phrase- or sentence-based analysis.

Keyword Recommendation

CLEVis supports keyword selection by applying text mining algorithms to the CLE textual data. First, we apply typical natural language processing methods to pre-process the data by removing stop words and tokens. Then, keywords are recommended to users through a variety of approaches:

- R1: TF (term frequency) keywords: The top TF keywords sometimes work well, but sometimes include non-useful but frequently occurring terms like "officer or vehicle".
- R2: TF-IDF (term frequency–inverse document frequency) keywords: The TF-IDF keywords sometimes may return specific names (such as popular first names Nicole, Ethan, etc.) of people or locations (such as Main, East, etc.).
- R3: LDA (Latent Dirichlet Allocation) topics with variational Bayes algorithm and TF weighted input: We apply a LDA algorithm to generate a series of topics based on the online variational Bayes model [5], where the input terms to LDA is weighted by TF. The keywords in each topic are grouped and recommended to users. For example, if using 3 topics (with top 5 keywords shown here), the police reports reveal:
    1. Topic1: home, house, violation, station, school
    2. Topic2: domestic, address, property, street, met
    3. Topic3: vehicle, theft, phone, station, door
- R4: LDA topics with variational Bayes algorithm and TF-IDF weighted input: We further use TF-IDF weights in LDA with the

variational Bayes model [5]. For the same police reports, 3 topics are identified:

1. Topic1: stay, window, employee, drug, threatening
2. Topic2: result, truck, showed, clerk, damage
3. Topic3: mother, dog, damage, fight, child

- R5: LDA topics with collapsed Gibbs sampling: We apply a different LDA based on collapsed Gibbs sampling and the TF input terms [5]. Again, 3 topics are achieved:

    1. Topic1: vehicle, speed, property, reside, park
    2. Topic2: park, theft, cite, vehicle, suspicious
    3. Topic3: home, refer, violation, house, juvenile

CLEVis recommends keywords by utilizing R1-R5 to the user in order to filter the CLEs. For the police report documents, TF and TF/IDF often fail to find keywords of interest (e.g., the names of people and streets are often extracted but not useful). Instead, the LDA topics extract meaningful keywords in groups which provide informational support to the users' decision making. For example, in R5, Topic1 relates to vehicle events and Topic3 relates to home violations. Probabilistic topic distributions over the documents are diverse since each of them often refers to one case (e.g., either theft or home violation). Therefore, the probabilities of distribution information are not used for visualization in the system.

The implementation includes several steps: (1) processing CLE data items and extracting keywords using the Natural Language Toolkit (NLTK 3.4.5); (2) computing keyword weights (e.g., TF or TF-IDF); (3) applying LDA modeling using the Stanford Topic Modeling toolkit. We set the number of topics to five by default and show the top five keywords in these topics. (4) These keywords are displayed to the user to help an iterative exploration of text content through keyword selection.

User Specified Keywords

In the real-world there is often a topic specific, or location specific, ontology. CLEVis allows users to enter new keywords at any time to integrate this information. In addition, different words may refer to the same meaning. For instance, "weed, grass, Marijuana, MJ, chronic" are synonyms for the same drug. Some of these terms may only appear a few times so they cannot be extracted by automatic methods. Therefore, we allow users to incorporate a set of ontology keywords loaded from predefined files or by their own input.

CROI Document and CROI Clustering

CROIs can be clustered to find similar and disparate patterns in event activities. We develop a new concept of CROI *document*, which represents the semantic information of CLEs in a discovered CROI. It is constructed by combining all text descriptions of CLEs in this CROI and then represented by a vector of keywords k_z, z ∈ $[1\,, Z]$, where Z is the number of keywords, which are selected by users as above. Then *document similarity* of two CROI documents (C_i, C_j) is computed by the cosine similarity:

$$\mathrm{sim}\left(C_i, C_j\right) = \frac{\sum_{z=1}^{Z} f(k_z|C_i) f(k_z|C_j)}{\sqrt{\sum_{z=1}^{Z} f(k_z|C_i)^2 \sum_{z=1}^{Z} f(k_z|C_j)^2}}$$

where $f(k_z|C_i)$ is the term frequency of k_z in C_i. Users can also compute the similarity from semantic categories. Once CROIs are clustered, they are projected to a 2D scatterplot view so that users can intuitively discover clusters and outliers. Users can also interactively adjust keywords used in the computation and examine how the changes affect CROI relationships.

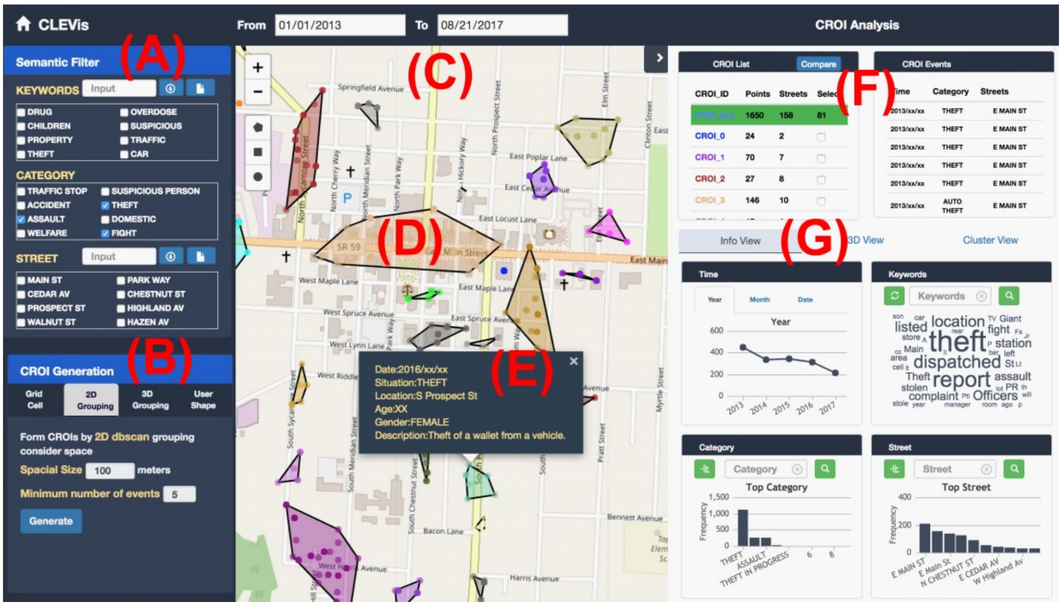


Figure 2: CLEVis interface. (A)-(B) CROI generation; (C)-(E) CROI on the map; (F)-(G) CROI exploration. The data displayed is synthetic but based on real situational events.

CLEVis System

As shown in Figure 2, the CLEVis interface consists of three views: CROI generation panel (Figure 2(A)(B)), CROI visualization (Figure 2(C)-(E)), and CROI exploration panel (Figure 2(F) and (G)). CROIs shown in Figure 2 are neighborhoods identified from police reports from 2013-2017. They are generated by selecting events in the categories of assault, theft and fight, and using ε =100 meters and *minPts* = 5 for CROI generation. For the purposes of this paper we use synthetic examples based on real world data examined in the system. Note that users can load data for other locations and time periods to our database by using a specifically designed software module which is not shown here.

CROI Generation

Semantic Filter: As shown in Figure 2(A), a user can choose from keyword recommendation groups, manual input, or loading files to define keywords. They can also select street names of event locations. A list of categories can be checked. These semantic elements can be multi-selected and combined to support joint filtering.

CROI Generation Panel: As shown in Figure 2(B), the panel supports four types of CROIs by: (1) dividing space into grid cells; (2) performing 2D automatic CROI generation; (3) conducting 3D clustering to generate TS-CROI; and (4) loading predefined shapefiles.

CROI Visualization on the Map

Figure 2(C) is the map view where different map styles and layers can be selected. A set of CROIs are visualized as regional polygons, where Figure 2(D) depicts one CROI with some event points inside. Labels of CROIs and events can be shown for further analysis (e.g., Figure 2(E)).

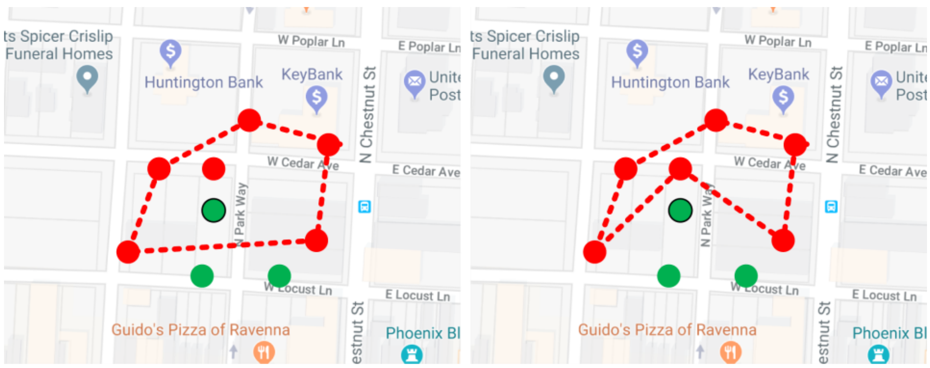


Figure 3: Drawing a CROI of multiple events. Left: the convex hull of red event points; Right: a concave hull of this CROI. The events are created only for illustration.

CROI Concave Hull: The automatically generated CROIs do not include an explicit border. Therefore, a CROI drawing method defines the region of a CROI based on the points it contains. The convex hull of a CROI (red points in Figure 3 left) may cover areas that belong to other CROIs (green points). We present a concave hull-based algorithm to find a bounding region. Concave hulls are not unique for a point set. We use an approach to get a concave hull with a given concaveness value (Figure 3 right). In the implementation, we first calculate the convex hull with a Divide and Conquer algorithm and then use a revised Gift Opening algorithm to create the hull, which peels external triangles after applying Delaunay triangulation over the CLE points. In particular, when peeling triangles from outside to inside on the convex hull, if we found one CLE point inside a triangle, then the corresponding edges are removed. This algorithm is expedient since the number of CROIs and the number of points is usually small in number.

CROI Coloring: Multiple CROIs are shown with distinct colors. Instead of randomly selecting colors, the colors are pre-selected from a pool of colors to make better visual distinction. Moreover, users can select an active CROI highlighted on the map. The other CROIs are shown with high transparency to enhance the focus on the active one.

CLE Labeling and Timeline: CROIs and their CLEs are labeled when users select them by mouse clicking or from a selection on the list view. The CLEs are shown as dot markers whose sizes and colors represent the densities of events at locations (Figure 8). To overcome the clutter problem, CLEVis only shows dots in an active CROI.

Moreover, a timeline view shows the temporal distribution of multiple CLEs at the same location (Figure 9). Users can interact with the timeline to find the details of each CLE. Here, we have tried different approaches. A bar/line chart can show the time distribution but cannot show detail contents. A story line is good for locations with a few events. However, it displays too much textual information for high intensity locations with many CLEs, which may overwhelm users. The interactive timeline allows users to study event details on demand.

CROI Visual Exploration

A CROI list is shown in Figure 2(F). The views in Figure 2(G) visually present: characteristics of the active CROI; 3D event cube of TS-CROI; and CROI clusters. These views are coordinated with the map view to conduct **Task IV** of (TS-)CROI visual exploration.

Visualization of active CROI characteristics: CLEVis presents details of the active CROI in: (1) Detail view: A list to show locations, time stamps, categories, and other information of events in the CROI (Figure 2(F) right). (2) Temporal view: A temporal bar chart to show the event distribution over time (Figure 2(G) top-left). (3) Keyword view: A word cloud to show keywords in the event descriptions (Figure 2(G) top-right). (4) Street view: A histogram of the numbers of events in different streets (Figure 2(G) bottom-right). (5) Category view: A bar char of the numbers of events in different categories (Figure 2(G) bottom-left).

Users can interactively brush and filter the information for drill-down studies. A monochromatic color scheme is used in these charts. All the charts use the same color as the CROI in the map view.

We first used colorful keywords in the cloud to depict frequency. However, users found the colors to be confusing especially given the color pallets of the CROIs employed in the map view. The monochromatic cloud instead leads to a better depiction. The CROI colors are pre-selected to distinguish CROIs.

3D Event Cube Visualization: TS-CROIs are specific CROIs which may overlap at places, because they include CLEs in the same area but within different time periods. A space-time cube view [4] visualizes them with x and y coordinates referring to spatial dimensions and z representing time. Users can rotate and zoom into the cube. All the CLEs are visualized as 3D points. TS-CROIs are shown in the same colors as in the map view.

An alternative approach to the 3D cube view is to use small multiples each representing one TS-CROI on a 2D map. As there may be many small multiples making it hard to find overlaid events. The 3D cube view has visualization strengths but may impose burden for users to interact and understand the 3D data points. We currently adopt the 3D view while we will provide both options in the system.

Visual Comparison of CROIs: Users can select multiple CROIs and compare them in detail. Figure 4 shows the interface where major street locations, categories, timelines, and keyword clouds are displayed concurrently. An alternative design for keyword clouds (which may place similar words in different places) is to use the ranked word lists with links among similar keywords. Here the clouds are used since the users prefer this familiar visualization. We will try other alternatives for comparison in future work.

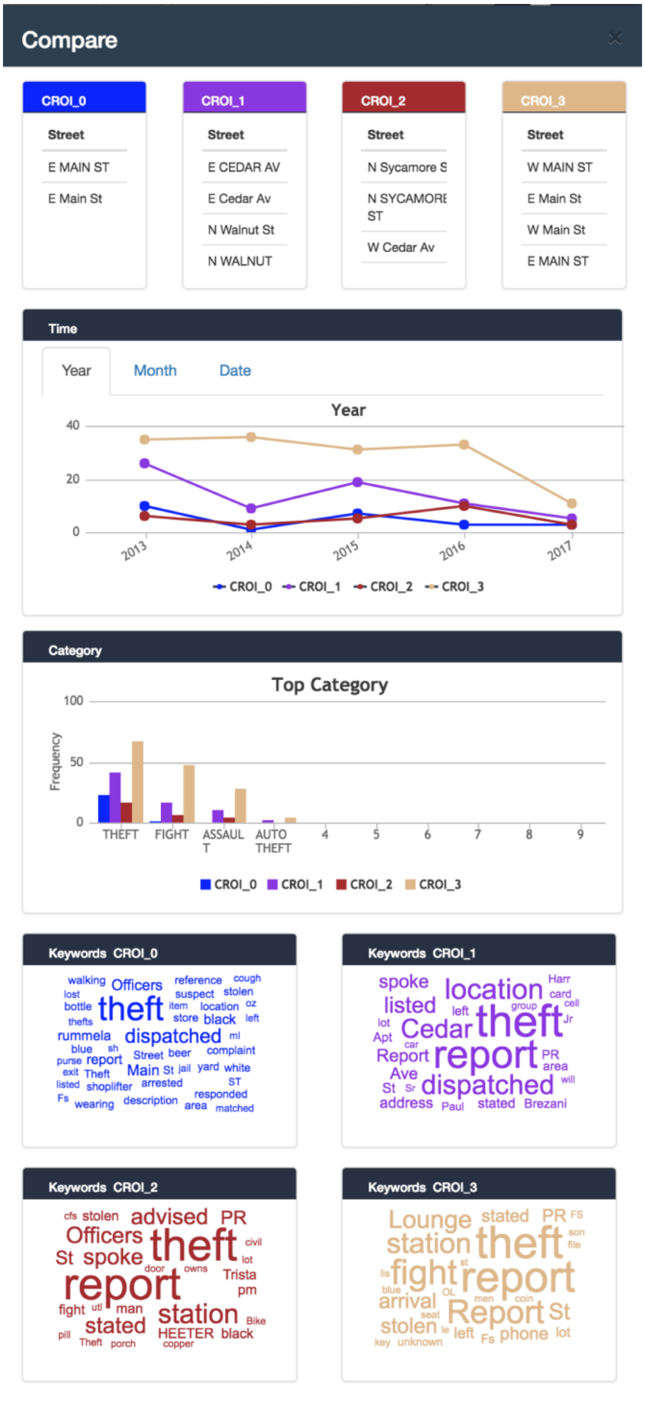


Figure 4: Compare multiple CROIs in a comparison view.

In addition, CLEVis supports users to select and investigate clusters and outliers of CROI documents for VA **Task V**. In a scatterplot view, each CROI is shown as a point and the distances among the points represent the closeness of the CROIs. In the implementation, each CROI is represented as a high dimensional vector computed by the frequencies of their top keywords. The MDS (Multi-dimensional) projection is applied to project the CROI vectors to 2D points. T-SNE and PCA can be applied as well. The points are colored to show which cluster they belong to as shown in Figure 8(d). Meanwhile, the CROIs in the map view are also colored to reflect the cluster memberships. Users can examine each CROI with interactions over the scatterplot view.

Clustering CROIs into several groups can intuitively convey "these neighborhoods have similar behavior and others do not". Such immediate findings can trigger users' interest to further investigate the details involved. This design is better than simply using the projected view without clustering, in which users need to study relationships among individual CROIs. This is the reason why we apply CROI clustering in the visual comparison together with the MDS view.

(a) (b) (c) (d)

Figure 5: Investigate 911 call records in New Orleans during Katrina landing from Aug, 29 to Sep. 8, 2005. (a) Find CROIs in New Orleans East. The top three CROIs: (b) A nursing home; (c) A senior living center; (d) A local hospital.

CASE STUDIES

We conducted several case studies with domain collaborators to illustrate how CLEVis can be employed in real world investigation.

Investigating Katrina Flooding in New Orleans

In this example New Orleans East, one of the hardest-hit areas in Hurricane Katrina, is selected resulting in the generation of CROIs automatically with ε = 200 meters and *minPts* = 5(Figure 5(a)). Further smaller areas within New Orleans East are identified with aggregated emergency calls. The top three CROIs in the CROI list are explored on the map. These include:

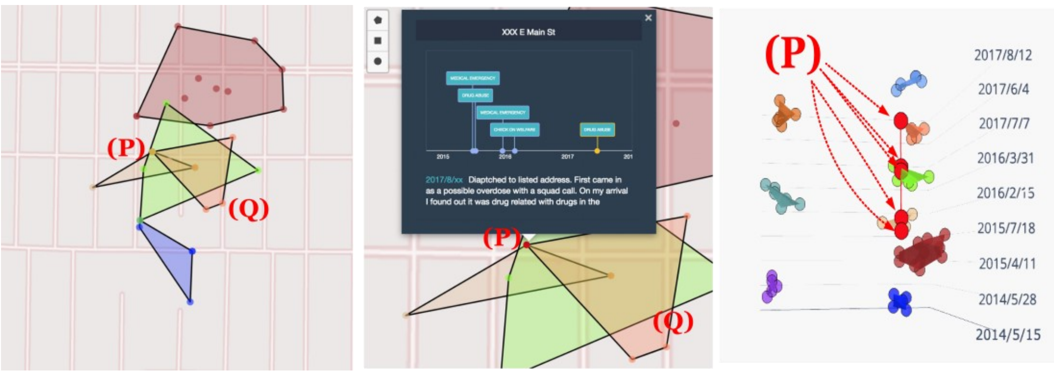
- (a) (b) (c) (d)

- Figure 5(b): A nursing home where 17 calls were made in 5 days, and where 22 residents died.
- (a) (b) (c) (d)
- Figure 5(c): A senior living center with 14 calls all from the same address.
- (a) (b) (c) (d)
- Figure 5(d): A local hospital with 10 calls. Here, one location has multiple calls depicted in the timeline view, while another label shows one call in a slightly different address. However, these calls should belong to the same hospital although they are recorded with four different addresses in the dataset. Our CROI algorithm automatically groups these calls into one CROI which facilitates easy study for the users.

These CROIs are all related to vulnerable people (seniors and patients) which would be vital for emergency services responding during an event, with potentially important linked insights that might not be gleaned from a single event-at-a-time software system.

Identifying Patterns of Overdoses

One of the initial justifications for the development of CLEVis was to find additional insights into local overdose situations. After filtering the police dataset by "drug" and "overdose", the resulting CLEs are clustered to generate TS-CROIs. Our collaborative team used different combinations of parameters and careful examination of the resulting TS-CROIs in the map view to decide whether the results are useful. As an example, Figure 6 shows the specific results with ε_s =220 meters, *minPts* = 5 and ε_t = 6 months, from which they found meaningful clusters in the downtown area.



(a) (b) (c)

Figure 6: TS-CROI generation by using keywords "drug" and "over-dose" in police report data. Here the map contents are synthetic for privacy protection. (a)

3D spatiotemporal clustering results; (b) One hot location P on several overlapped TS-CROIs has 5 different events from 2015-2017 shown in the timeline; (c) Event cube shows 5 red event points linked to P belonging to different TS-CROIs.

These examples are synthetic for privacy protection though based on real data. (a) (b) (c)

Figure 6(a) identifies overlapped TS-CROIs with at least 5 nearby events occurring within 6 months. In (a) (b) (c)

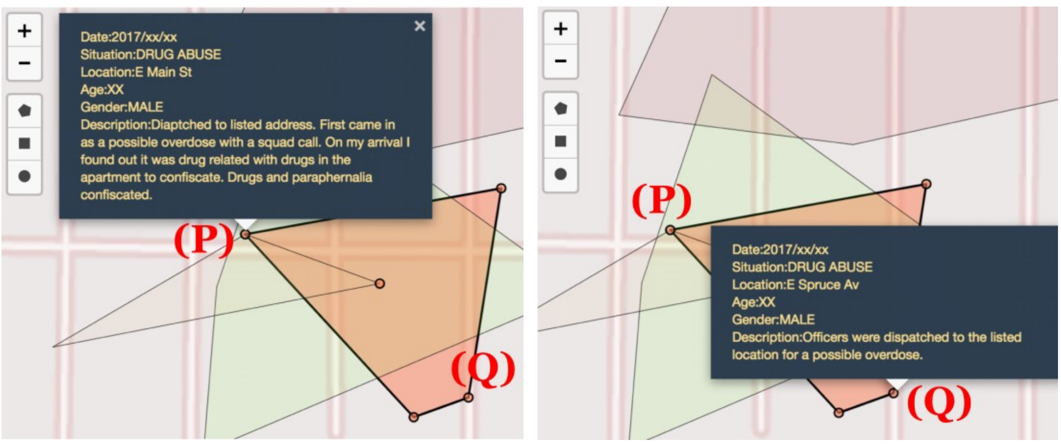
Figure 6(a), a critical location, *P*, resides on three TS-CROIs. (a) (b) (c)

Figure 6(b) displays five independent events on *P*. (a) (b) (c)

Figure 6(c) shows the 3D event cube visualization on which these five event points are connected by a red line indicating they overlap on *P*. They are grouped into three TS-CROIs representing event bursts in 2015, 2016, and 2017, respectively. Focusing on the 2017 TS-CROI, (a) (b)

Figure 7(a) shows an event of a male's drug abuse at 2017/XX/XX. *P* is grouped with other events at a nearby location *Q*. There was another drug abuse event at Q within two months as shown in (a) (b)

Figure 7(b). These two events are close in both time and space. These two events might be enough for the local health department to conduct a localized intervention.



(a) (b)

Figure 7: Study one CROI from (a) (b) (c)

Figure 6. (a) One drug abuse event at P. (b) A related event at a nearby location Q in the same year (2017).

Analyzing Crimes Throughout Town

By dividing the map into cells (CROIs), a user can compare the similarities between different locations, for example around two fast food restaurants. (a) (b) (c) (d)

Figure 8 shows an example of such an investigating. The users select the categories "theft, assault, suspicious person, fight" for the study. (a) (b) (c) (d)

Figure 8(a) is the overview of the city with each CROI cell (200 by 200 meters), where the color saturation shows the density of events. (a) (b) (c) (d)

Figure 8(b) visualizes CROI attributes. It indicates that the top event is "theft", and the top street is "Main Street".

(a) (b) (c) (d)

Figure 8: Studying crime events. (a) Grid-based CROIs with different densities of events; (2) Visualizing their characteristics; (3) CROI clusters with different colors; A and B are two CROIs belonging to the red and blue clusters, respectively; (4) CROI clusters in the scatterplot view.

(a) (b) (c) (d)

Figure 8(c)-(d) show the results after applying CROI clustering to identify four clusters. When CROI (A) is selected, (a) (b) (c) (d)

Figure 8(c) shows a label of the top crime event in it.

Reading the scatterplot (a) (b) (c) (d)

Figure 8(d) carefully, there is an outlier CROI, (B), which forms an outlier cluster in blue. The users visualize this CROI in more detail (Figure 9). A special location with overlapped events is found. The timeline view shows that this location has an excessive number of events over 4 years. Most of the events were “theft” in a local supermarket. The users identify that such events decline during winters and increase during summers. Such findings might help the police and the store to develop more seasonally effective prevention strategies.

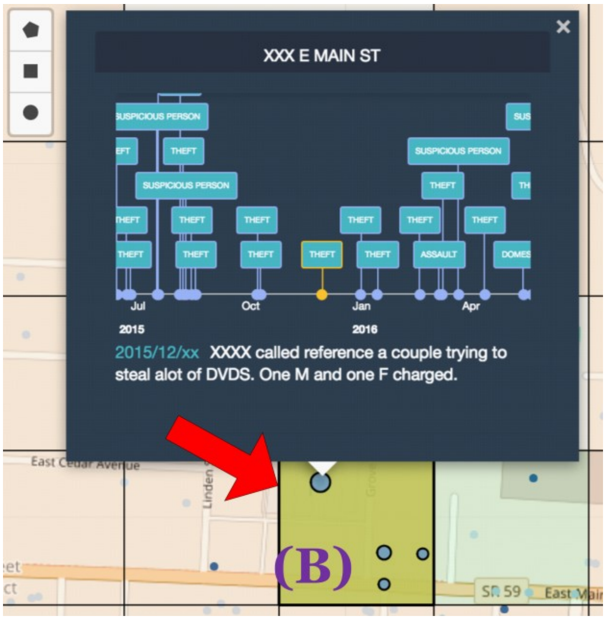


Figure 9: Study one outlier CROI (B) from (a) (b) (c) (d)

Figure 8, which identifies excessive theft events.

DISCUSSION

Design Process and Decisions

Multiple meetings over course of a year with local area police and health care workers framed the development of the system described here. Starting with an initial conceptualization of the problem, subsequent meetings focused on data issues, usability concerns, analytical tools to be employed, and design aspects. As a result of this evolving collaboration the following decisions were made.

Proof of Concept: At the beginning, after gaining major domain requirements for CLEVis, we presented a conceptual design of a visual query interface that displayed CLEs as marker points and/or heatmaps. The design was not totally successful since (1) markers were prone to being cluttered; (2) the heatmap only gave an overview without the needed contextual detail. We were told that a typical use would be to conduct a visual investigation over one or a few small, meaningful neighborhoods, and examine the aggregated activities within. As a result, we designed

CROIs as the basic operating units. Moreover, we identified that increases in events along time should be easily detected which guided our design of the TS-CROIs.

Semantic integration: As event descriptions are important, we designed the keyword based visual query to filter CLEs using TF or TF-IDF. However, this approach did not initially satisfy the users in tasks such as detecting various terms about illegal drugs. We then combined several automatic keyword recommendations. Moreover, we allowed users to load and input keywords from their domain ontology (i.e., a set of concepts and categories in their specific domain).

Visualization design: We implemented different visual metaphors at the beginning such as a parallel coordinates plot drawing the number of events on multiple categories, while each coordinate represented a CROI. However, we found such visualizations were not well accepted by the users since the limited screen size easily introduced clutter due to the many CROIs. To further improve usability, we developed views with diagrams/charts that are more familiar to non-technical users.

User Feedback

After the prototype was implemented, additional feedback from law enforcement was collected, especially with regards comparing CLEVis to other tools that might be used (e.g., Word, Excel, Google Map).
In general, the users appreciated the potential of the system to combine map and content queries, such as studying how many overdose records are within two blocks of a bar, what are their temporal changes, and what are the details of the overdose cases. They agreed that the advance of CLEVis over existing analytical software is that textual data now also becomes analyzable. They liked the ability to select keywords for additional details and follow-up information. The users commented that data input, region manipulation, and visual analyses in CLEVis are easy to understand and use. The word cloud is helpful for identifying co-occurring words and the interactive CLE detail visualization was also appreciated. The timeline view to study a series of events in a location of interest was particularly well received.

One advance that was mentioned was that CLEVis presented the means to combine multiple agency data (different police departments, or police and hospital for example). This is an ongoing problem in many jurisdictions, especially those smaller locations with less funding for analytical and data needs. CLEVis can provide a single visual system where data, as long as a geocode and text are available, can be easily combined and mined.

However, based on the feedback sessions, CLEVis also has limitations that need to be addressed in future work. First, the visual charts and map view are easy to understand by general users (e.g., police officers). However, the 3D cube view may only be used by higher level analysts since it is hard to understand especially when working in the field. In addition, the grid size selection, parameter setting, and word cloud-based interactions can still be improved in terms of ease of use. The 3D cube view may include more geo-context information by including the map and semantic information in the cube. Data input also needs work in terms of making CLEVis truly operational for multiple data generators working in the same space. This would also lead to the additional questions of confidentiality and access / editing rights.

For current target users, the datasets often do not exceed tens of thousands of CLEs, so the interactions are smooth. Data scalability will be an issue if CLEVis is implemented in much larger cities, where special data handling tools such as NanoCubes [8] may be adapted.

More work is also needed on the sematic analysis of typical data input into CLEVis. For example, there may exist false positives since a keyword can have multiple meanings. For example, users may find sometimes “drug” refers to medicines while mostly it refers to illegal ones in police reports. As a result, smarter text mining tools, possible geographically specific, need to be considered to further improve the system.

CONCLUSION AND FUTURE WORK

Small towns, just like large metropolitan areas, generate multiple spatial and textual data that are required to inform local decision making. Traditionally it has been difficult to study these records to look for patterns when resources are limited, and datasets are incompatible. One of the biggest challenges in working with complex social data is to help users conduct easy data exploration through integrated semantic and spatio-temporal queries and analysis. The described collaboration led to the conceptualization and then a product solution for this problem. CLEVis presents a VA solution that can easily be used to search for local patterns, identify trends or emerging problems, and by mining associated text, even identify possible solutions. We expect CLEVis to continue to evolve both with the current data partners, and through other collaborations. To this end, we intend to (1) incorporate more traditional data, such as census layers or deprivation indices; (2) add more visual clues to the map view for prompt convey of salient clusters; (3) develop different methods for the identification of CROIs; (4) process real time streaming data; and (5) explore how advanced semantic analysis methods could improve local area problem solving.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive comments. This work was supported in part by the U.S. NSF under Grant 1535031, Grant 1535081, and Grant 1739491.

REFERENCES

[1] E. Bertini, J. Buchmuller, F. Fischer, S. Huber, T. Lindemeier, F. Maass, F. Mansmann, T. Ramm, M. Regenscheit, C. Rohrdantz, C. Scheible, T. Schreck, S. Sellien, F. Stoffel, M. Tautzenberger, M. Zieker, and D. A. Keim (2011). Visual analytics of terrorist activities related to epidemics. *IEEE Conference on Visual Analytics Science and Technology*, 329-330.

[2] I. Cho, R. Wesslen, S. Volkova, W. Ribarsky, and W. Dou (2017). CrystalBall : A Visual Analytic System for Future Event Discovery and Analysis from Social Media Data. *IEEE Visual Analytics Science and Technology (VAST)*.

[3] A. Godwin and J. Stasko (2017). HotSketch: Drawing Police Patrol Routes among Spatiotemporal Crime Hotspots. *Proceedings of the International Conf. on Systems Sciences*, 1372-1380.

[4] J. Chae, D. Thom, H. Bosch, Y. Jang, R Maciejewski, DS Ebert, T. Ertl. (2012). Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. *IEEE Conference on Visual Analytics Science and Technology*, 143-152.

[5] M.D. Hoffman, D. M. Blei, and F. Bach. (2010). Online Learning for Latent Dirichlet Allocation. *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, 856-864.

[6] W. Jentner, D. Sacha, F. Stoffel, G. Ellis, L. Zhang, and D. A. Keim (2018). Making machine intelligence less scary for criminal analysts: reflections on designing a visual comparative case analysis tool. *Visual Computer*, 1-17.

[7] J. Johansson, T. Opach, E. Glaas, E., T.-S. Neset, C. Navarra, B.O. Linner and J.K. Rod (2017). VisAdapt: A Visualization Tool to Support Climate Change Adaptation. *IEEE Computer Graphics and Applications, 37*, 54-65.

[8] L. Lins, J. T Klosowski, and C. Scheidegger (2013). Nanocubes for Real-Time Exploration of Spatiotemporal Datasets. *IEEE Transactions on Visualization and Computer Graphics, 19*, 2456-2465.

[9] S. Liu, X. Wang, C. Collins, W. Dou, F. Ouyang, M. El-Assady, L. Jiang and D.A. Keim (2018). Bridging Text Visualization and Mining: A Task-Driven Survey. *IEEE Transactions on Visualization and Computer Graphics*, 25(7), 2482-2504.

[10] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford (2011). SensePlace2:

GeoTwitter analytics support for situational awareness. *IEEE Conference on Visual Analytics Science and Technology*, 181-190.
[11] R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. Ebert, R. Brigantic, C. D. Corley, G. A. Muller and S.W. Sanders (2011). A pandemic influenza modeling and visualization tool. *Journal of Visual Languages and Computing, 22*, 268-278.
[12] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W.S. Cleveland, S. J. Grannis, and D. S. Ebert (2010). A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics, 16(2)*, 205-220.
[13] A. Malik, R. Maciejewski, S. Towers, S. McCullough, and D. S. Ebert (2014). Proactive spatiotemporal resource allocation and predictive visual analytics for community policing and law enforcement. *IEEE Transactions on Visualization and Computer Graphics, 20*, 1863-1872.
[14] F. Miranda, H. Doraiswamy, M. Lage, K. Zhao, B. Gonçalves, L. Wilson and C. T. Silva (2017). Urban Pulse: Capturing the Rhythm of Cities. *IEEE Transactions on Visualization and Computer Graphics, 23*, 791-800.
[15] T. Kapler, W. Wright, GeoTime Information Visualization. (2005). *Information Visualization*, Vol 4, Issue 2, 136-146.
[16] A. Rasheed and U. K. Wiil (2016). A Tool for Analysis and Visualization of Criminal Networks. *Proceedings - UKSim-AMSS 17th International Conference on Computer Modelling and Simulation,* 97-102.
[17] A. M. Razip, A. Malik, S. Afzal, M. Potrawski, R. Maciejewski, Y. Jang, N. Elmqvist, and D. S. Ebert (2014). A mobile visual analytics approach for law enforcement situation awareness. *IEEE Pacific Visualization Symposium*, 169-176.
[18] T. Schreck and D. S. Keim (2013). Visual Analysis of Social Media Data. *Computer, 46*, 68-75.
[19] D. Thom, H. Bosch, S. Koch, M. Worner, and T. Ertl (2012). Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages. *IEEE Pacific Visualization Symposium*, 41-48.
[20] D. Birant and A. Kut (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data and Knowledge Engineering, 60*, 208-221.

ABOUT THE AUTHORS

Chao Ma is a PhD candidate in the CS Department, Kent State University. His research interests include urban data visualization and interactive visual analytics. Email: cma1@kent.edu

Ye Zhao is a professor in the CS Department, Kent State University. His research interests include visual analytics of urban data, deep learning visualization, medical image processing, and information visualization. Email: zhao@cs.kent.edu

Andrew Curtis is a professor in the Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University. His research interests include spatial epidemiology, context enriched spatial data collection, and fine scale mapping of risk. Email: ajc321@case.edu

Farah Kamw is an assistant professor in the CS Department, Concordia University Ann Arbor. She received her PhD degree from the Kent State University by 2019. Her research interests include visual analytics of urban trajectory data, urban accessibility, and urban navigation. Email: fkamw@kent.edu

Shamal AL-Dohuki is an assistant professor in the CS Department at the University of Duhok. He received his PhD degree from the Kent State University by 2019. His research interests include urban data management and visualization, the visual query of trajectory data, and semantic data query and analytics. Email: saldohuk@uod.ac

Jing Yang is a professor in the CS Department, University of North Carolina at Charlotte. Her research interests include visual analytics of multidimensional data, time-oriented data, text data, and urban transportation data. Email: Jing.Yang@uncc.edu

Suphanut Jamonnak is a PhD candidate in the CS Department, Kent State University. His research interests include geographical multimedia data management and visualization, visualization software design and visual analytics. Email: sjamonna@kent.edu

Ismael Ali is a lecturer in the CS Department, University of Zakho. He received his PhD degree from the Kent State University by 2018. His research interest is cognitive computing, semantic computing, and natural language analysis. Email: iali1@kent.edu