

Big Data in Psychological Research

Color Figures

Edited by Sang Eun Woo, Louis Tay, and Robert W. Proctor

Chapter 7: "Human-Guided Visual Analytics for Big Data"

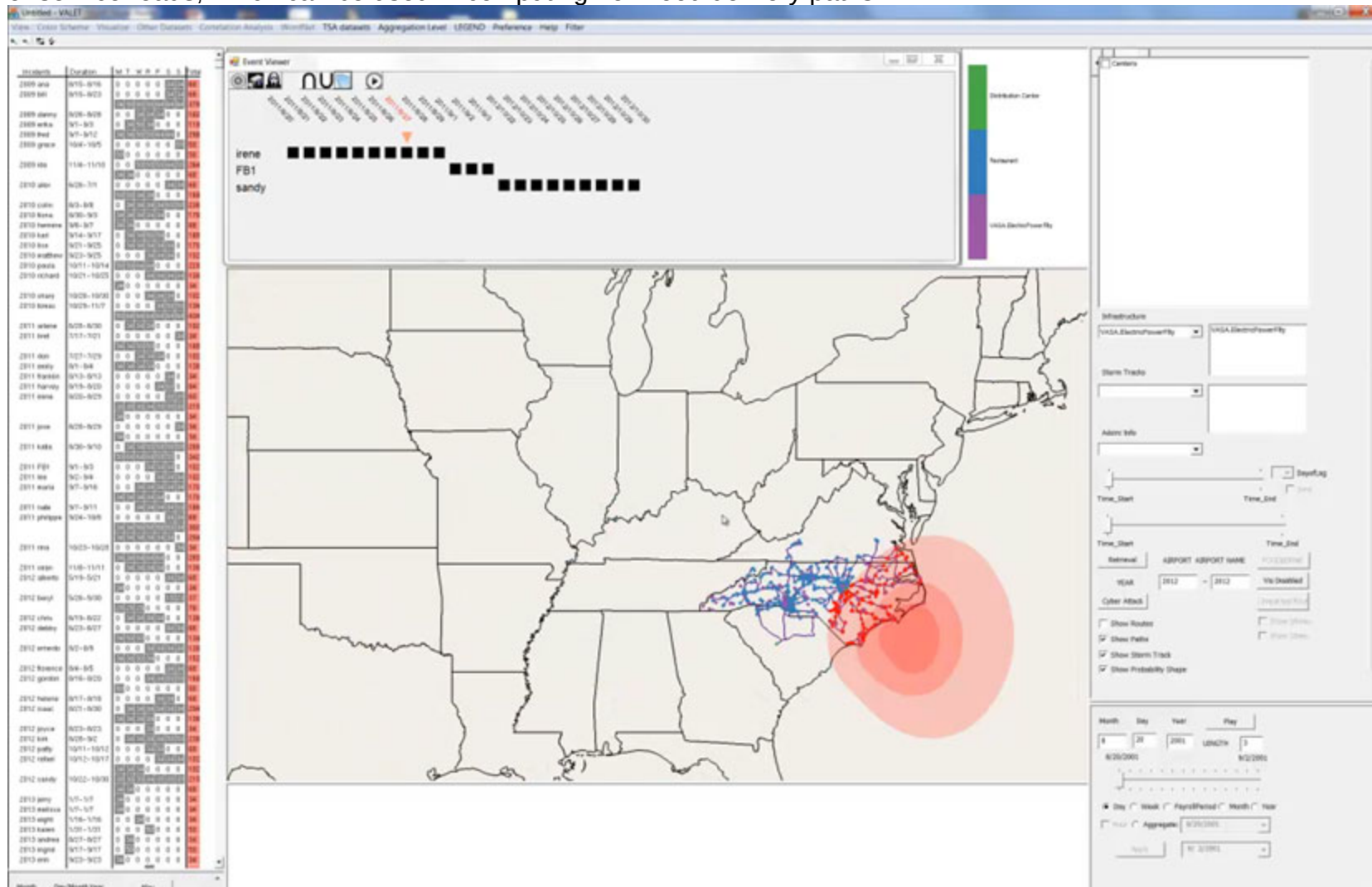
by Morteza Karimzadeh, Jieqiong Zhao, Guizhen Wang, Luke S. Snyder, and David S. Ebert

Figures 1–10

The figures provided in this book chapter are in black and white, as per APA production specifications, but they were created in systems that use color. For broader context and more information, the color versions of these figures are provided on this website. © 2020 American Psychological Association.

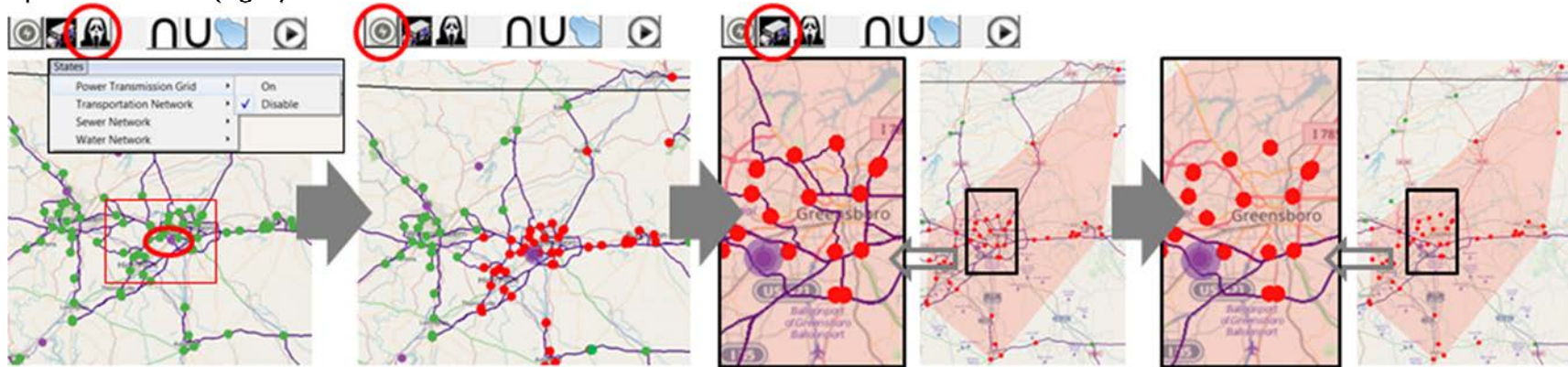
- **Figure 1.** VASA system overview, with calendar view on the left, event view on the top, map (geographical) view in the center, and advanced filtering and querying panels on the right (Ko et al., 2014). This figure shows the simulation of the landfall of Hurricane Irene in North Carolina. In this simulation, power generation units are hit by up to 34-knot winds. The hurricane proxy estimates the impacted restaurants and distribution centers. The system also allows identifying the power outage areas and out-

of-service roads, which can be used in computing new food delivery paths.



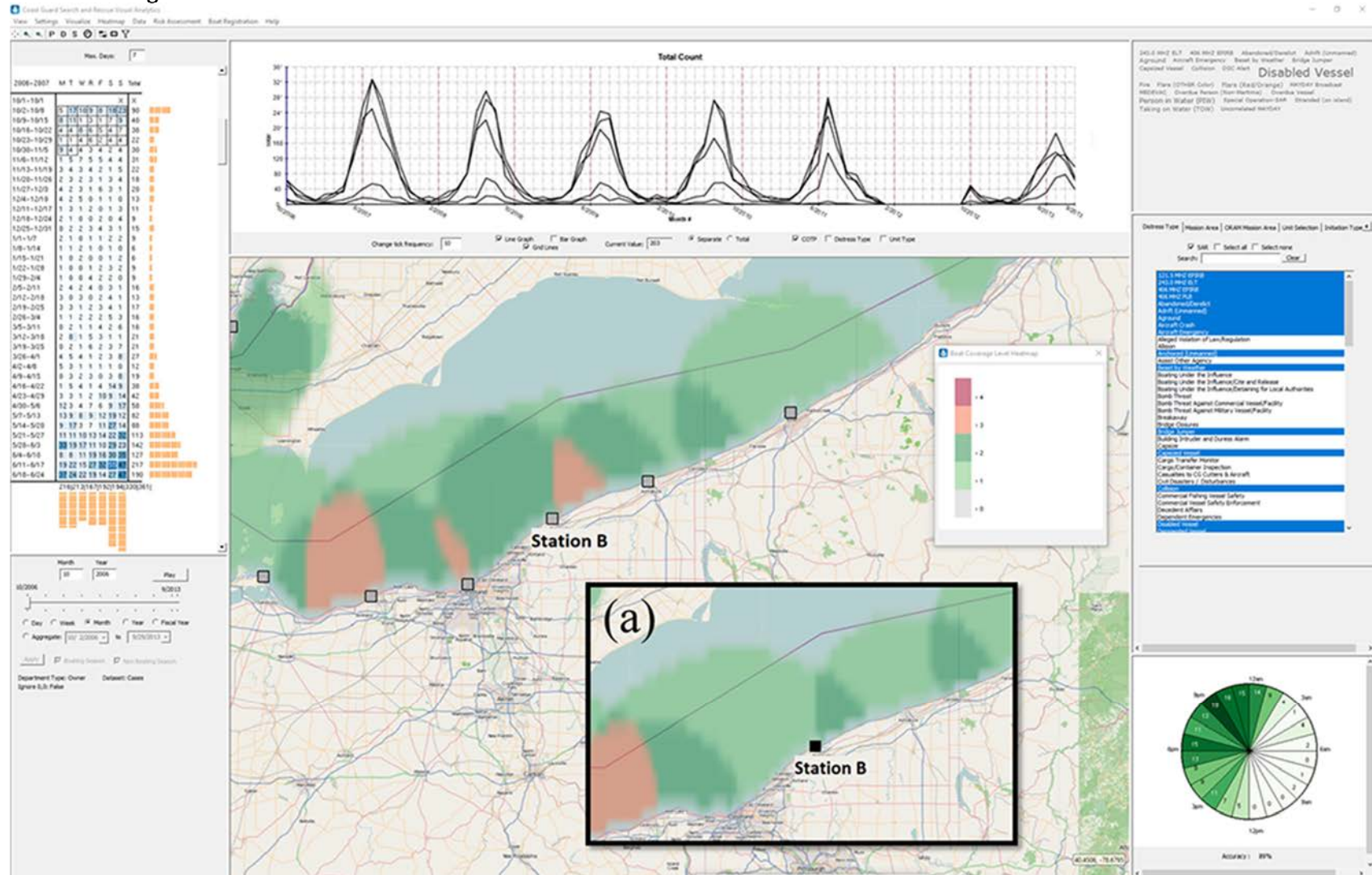
- Figure 2.** An example of a cyberattack simulation using the VASA system. The analyst selects the option to disable a power plant to simulate a cyberattack on that plant and selects a region shown using the red rectangle to indicate the area of interest (left). One main plant (purple dot) falls within this rectangle (second left). The integrated models estimate the affected subsidiary plants and workstations (red dots; second right). Finally, network-disabled regions are represented by a polygon and along with

updated routes (right).



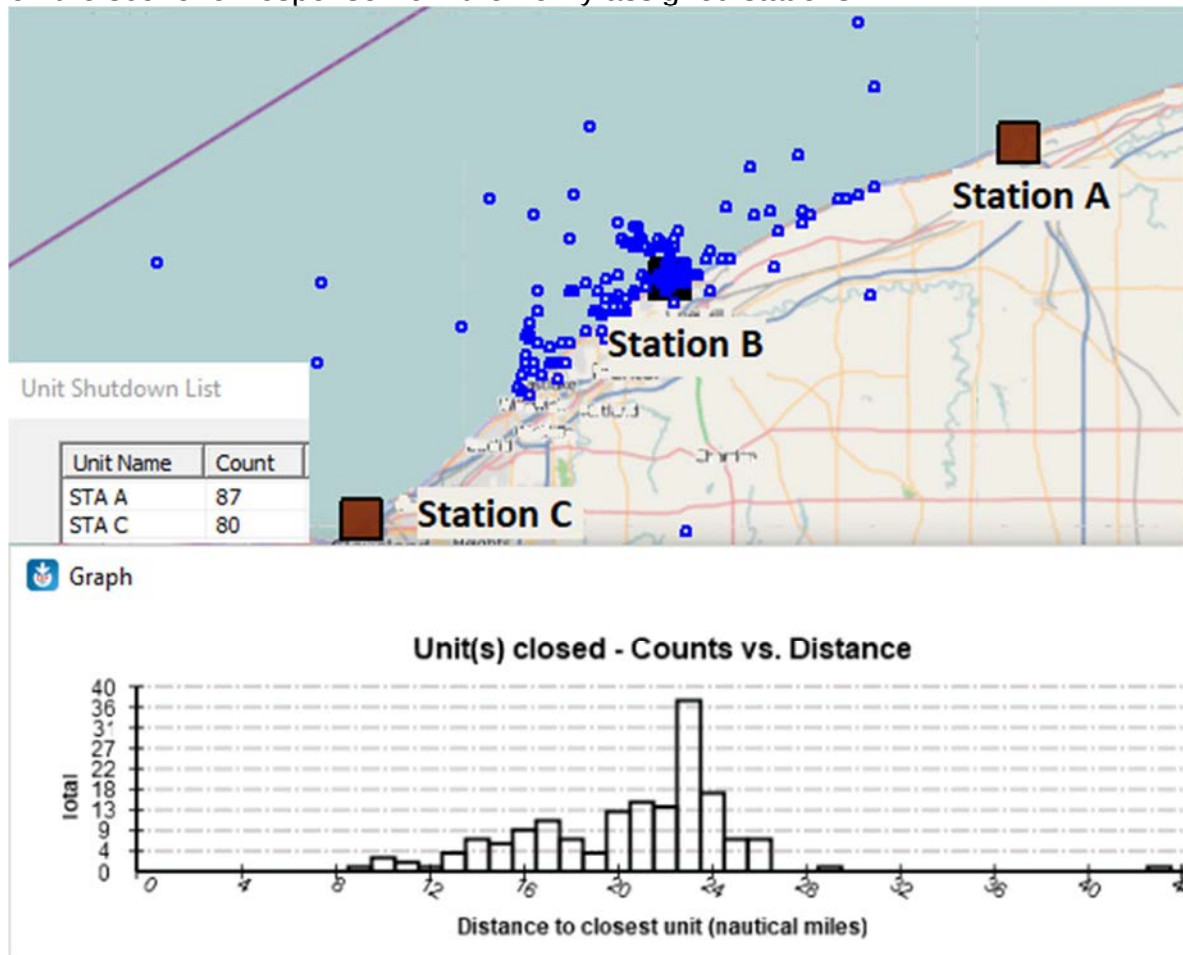
- Figure 3.** cgSARVA user interface. The calendar view at the top left counts incidents in the day-of-week layout to assist temporal analysis of incidents, the time slider at the bottom left allows users to interactively filter data in a variety of temporal granularities, the time series view at the top center counts incidents by the user-specified granularity (e.g., by year, month, or week), the clock view at the bottom right visualizes incident counts per hour. The view at the top right shows the keywords of incident reports. The view in the middle right allows users to filter data by multiple attributes (e.g., incident types or rescue stations). The map view at the bottom center shows the spatial analysis of incidents and rescue resources. Here, the map view visualizes the water area safety by the number of available boats. The subfigure (a) visualizes the water area safety with the assumption that Station B is closed and its boats are out-of-operation. The colors show boat coverage level, with green showing low and red showing higher

boat coverage levels.



- Figure 4.** Distance risk assessment in cgSARVA when a station is hypothetically closed (Malik et al., 2011). In this case, a simulation is run assuming that Station B is shut down. As a result, all cases in Station B between 2006–2013 would be handled by the other nearest stations: 87 cases in Station A and 80 in Station C (shown in Unit Shutdown List). In the map view, blue dots are Station B's cases. The histogram shows the frequency of case distances for the Coast Guard to have rescue resources

on the scene for response from the newly assigned stations.



- **Figure 5.** SMART's main user interface summarizing hurricane-related Twitter and Instagram posts for Hurricane Florence. The topic lens (center) helps analysts identify different categories of posts (e.g., weather, Hurricane Florence, Safety) according to user-defined classifiers and machine learning models, content lenses (three lenses placed on the map) showing different human-generated first-hand reports in the north (e.g., "WHAT Hurricane?!?!") and the south (e.g., "It's getting windy out here!"), and

The screenshot displays the 'VACCINE' (Visual Analysis and Classification of Network Information) tool interface. The central map shows a geographic area with a circular 'Vaccine' overlay divided into segments: Safety, Weather, Rain Job, and others. The interface includes a left sidebar with filters and controls, a top status bar, and a right sidebar with a message table and a 'Safe' status indicator.

Left Sidebar:

- Status Panel:**
 - Following filters:
 - Fluence classifier
 - Filtering
- Control Panel:**
 - Filter:
 - Keywords:
 - Start date:
 - End date:
 - Visual:
 - Spatial Filter
 - Visualizations:
 - Tweet Tooltip
 - Content Lens
 - Tag Map
 - Heat Map
 - Tweet Aging
 - Tweet Classifier
 - Tweet Clusters
 - Cluster Lenses
- View:**
 - Topic Model View
 - Geo-message Table
 - Tweet Time Chart
- Privacy:**
 - Show User Id
- Auto refresh Tables:**
 - Auto Refresh
- Trends:**
 - Google Trends
 - CNN Trends
- Search Places:**
 - Search
 - Remove
- Reset Settings:**
 - Reset All
 - Log out

Top Status Bar:

Time Monitor: 2018-09-14 13:55:44 — 2018-09-14 14:05:32

Right Sidebar:

Safe (check icon)

hurricane (check icon)

hope stay (check icon)

wait drink (check icon)

hurricane party (check icon)

florence (check icon)

humidity (check icon)

mph rain wind (check icon)

current weather (check icon)

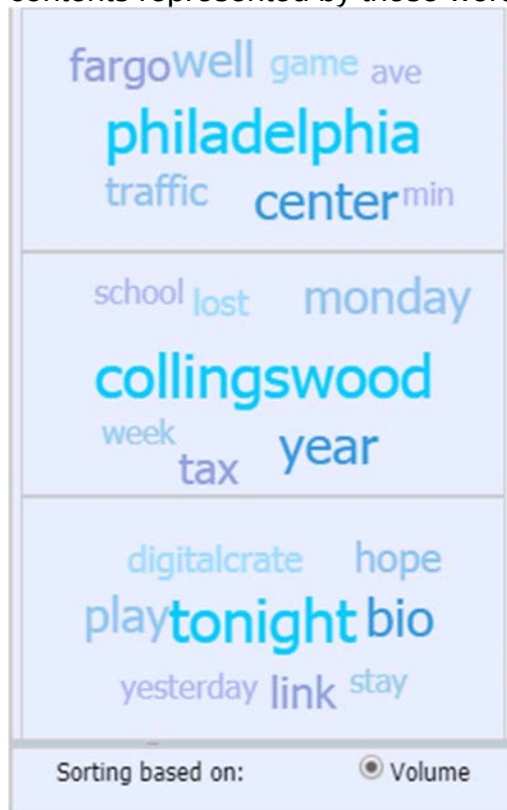
Sorting based on: Volume

Message Table:

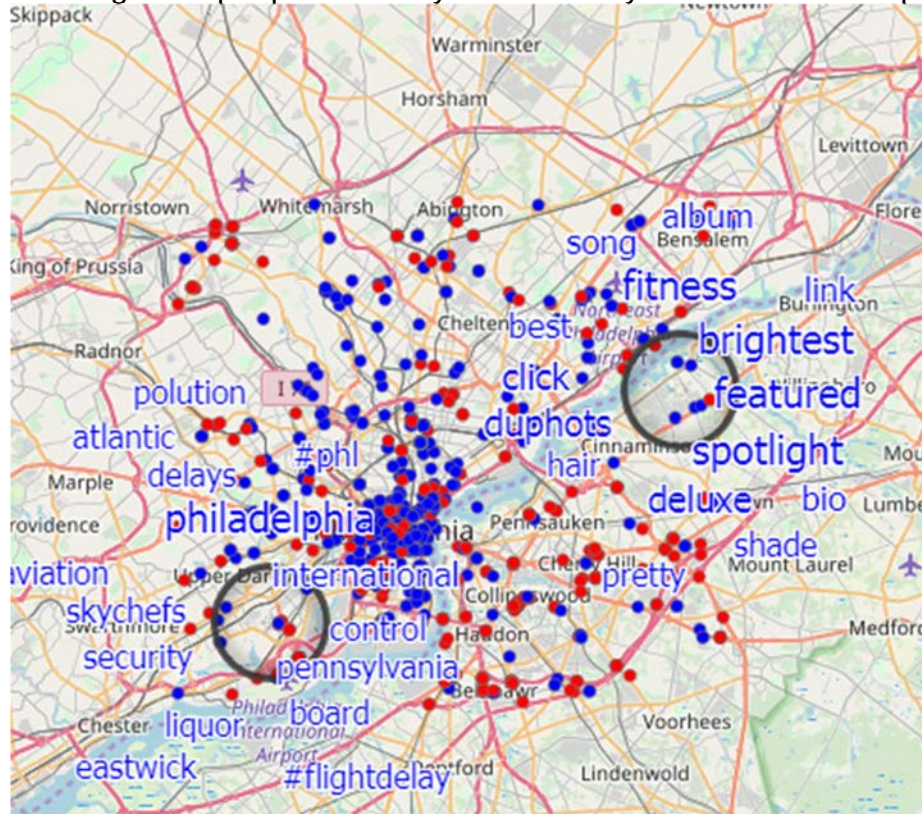
User Name	Creation Date	Tweets Content
26C8N9yhe	18-09-14 13:57:47 EDT	Flash Flood Warning including Wilton NC, Goldsboro NC, Mount Olive NC until 8:00 PM EDT https://t.co/213A0yqTf
A8P8N8Z8	18-09-14 13:57:46 EDT	Florence needs blow some a backyard to my window fr
0Kw0Y0yag	18-09-14 13:57:38 EDT	@JohnBathkLive @Hwms12Nq Florence?
CRPE04X0	18-09-14 13:57:30 EDT	the hurricane is coming tonight & I'm scared now 🙏
a5owAd2P	18-09-14 13:57:27 EDT	There have been 188 individual calls for rescue in Craven County, per Amber Parks, public spokeswoman with Craven Co. https://t.co/CT8H77z9b
VU87JUL8	18-09-14 13:57:25 EDT	oh nah, the wind is whipping my wig a little too ferocious for me, I'm going home.
	18-09-14	We're ehing! Read about our latest yjob opening here.

- Figure 6.** Topic view, showing words in each detected topic using Latent Dirichlet allocation (LDA). The size and color of words show their frequency in each topic. Users can click on any word to narrow down the visualization in other views to topical

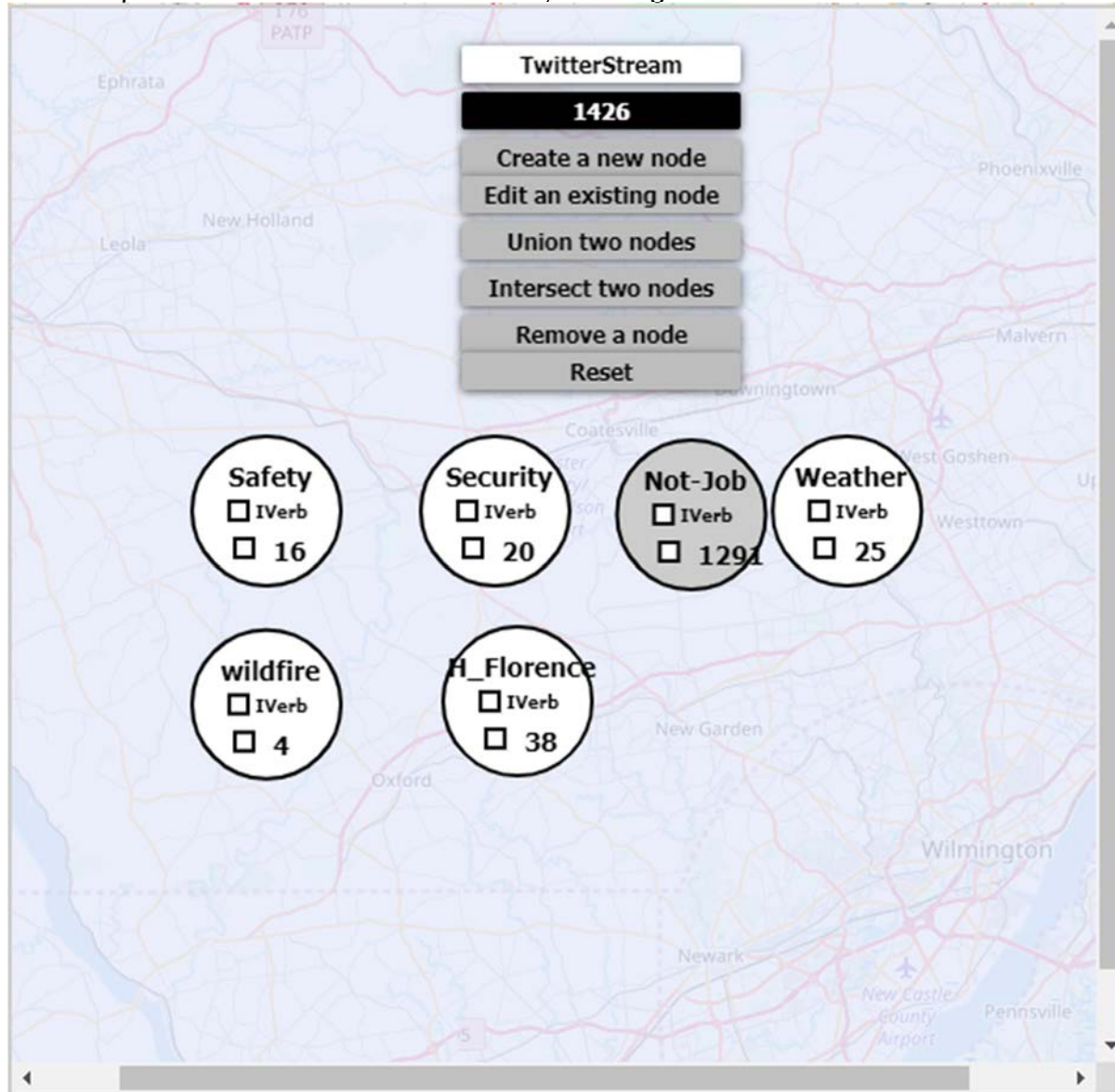
contents represented by those words.



- **Figure 7.** Content lens of SMART, combining spatial aggregation with Latent Dirichlet allocation (LDA) for topic modeling, showing the top topics and keywords for any area of interest represented by user-placed circles on the map view.

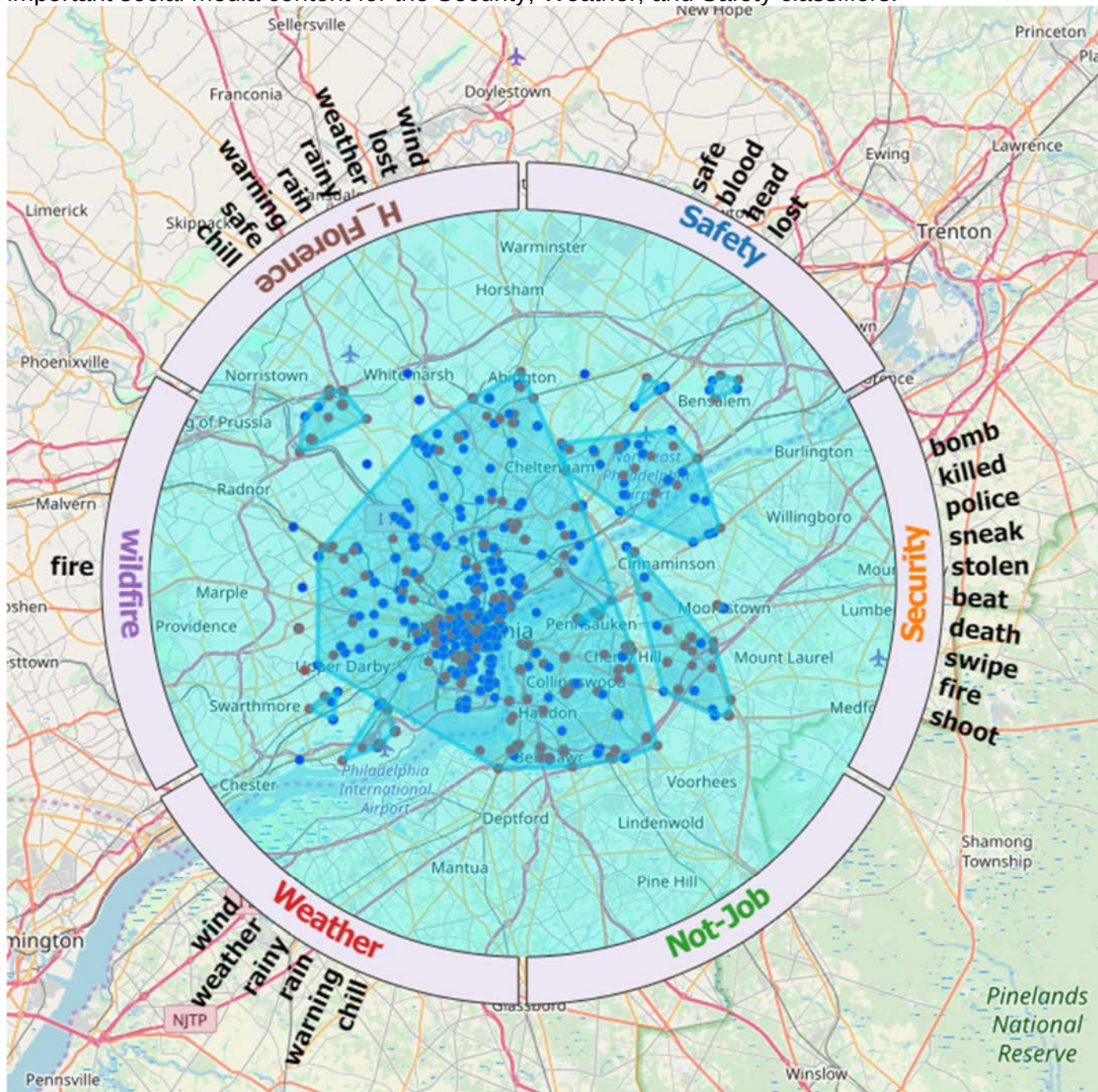


- **Figure 8.** Creating new classifiers in SMART. Users can associate certain keywords with different categories for further filtering down of posts related to those classifiers, creating a union or intersection of different classifiers.



- **Figure 9.** Cluster lens in SMART, using spatial aggregation and topic modeling to show important keywords for each category (resulting from a user-defined text classifier) on any area of interest. The cluster lens in this example signifies the presence of

important social media content for the Security, Weather, and Safety classifiers.



- Figure 10.** The overview of the organizational performance evaluation visual analytics application. (a) A control panel includes options for selecting a temporal range and filtering by incident types (e.g. dispatched vs. self-initiated). (b) A matrix visual representation shows the crime offense categories in rows and individual officers in columns. The colored blue cells demonstrate the product of [number of incidents that were responded to by an officer in one crime offense category] and the [weight assigned to the particular crime offense category by the system user]. Users can change the weight for each category interactively according to the perceived importance and policies for each organization. The red cells in the top headings show the total score of an officer, and the red cells on the left show the total score of a crime offense category. Darker colors mean higher values. (c) A weight table lists the crime offense categories and the corresponding manually assigned weights. The weights usually reflect the priorities of an organization.

