Advanced Cartographic Techniques for Making Sense of Big Dynamic Data

Alexandre Sorokine [a], *, Steven Fernandez [b]

[a] Oak Ridge National Laboratory, Oak Ridge, TN, U.S.A., *SorokinA@ornl.gov*
[b] Consultant 2, Almeria Analytics, Rio Rancho, NM U.S.A., *fernandezsj01@aol.com*

* Corresponding author

Abstract:

Now Big Data has become a ubiquity in almost every area of business and research. Billions of cameras, sensors, Internet-of-Things devices constantly stream their measurements that need to be preserved and analyzed. Present day technology is able to address the difficulties of transmitting, storing, and accessing large volumes of information. However, the need for understanding big data still remains the major challenge. Automated understanding of the big data streams using machine learning and other artificial intelligence algorithms is the goal of many of the current studies. Nevertheless, the results of the automated analysis have to be presented to humans in simple and understandable form and this is where visualization methods come into play. Since most of the big data records have geospatial and temporal reference there is a demand for developing cartographic techniques that would be effective at portraying big data streams for a variety of applications and visual environments.

For effective big data portrayal, such factors as a high level of noise and limitations of human cognition have to be taken into account. Preprocessing may be needed for filtering out of noise and unimportant data points, emphasizing salient features and processes, detecting critical events, and other enhancements. Resulting cartographic visuals have to be crafted so that useful information is not lost in the noise. Here we discuss experimental cartographic techniques that we have developed for dynamic visualization of the data coming from the U.S. electric grid. The visualizations are intended for providing control center operators with up-to-date (down to a second) actionable information on the current state of the grid. Based on this information operators should be able to perform control actions to prevent the development of blackouts and other emergency conditions. Our preliminary results are demonstrated using simulated data.

U.S. electric grid is a sophisticated network system that consists of thousands of transformers, generators, and millions of loads interconnected through high voltage transmission lines and local distribution systems. To keep the electric grid running and to prevent outages, a balance must be kept between the power production and power demand. This goal might be aided by equipping the electric grid with a large number of sensors such as the synchrophasors, frequency disturbance recorders, and others. These monitoring data must be analyzed and presented to grid operators and other decision makers in a timely and concise manner. As an outcome, the operators should be able to receive a clear and up-to-date picture of the current state of the grid including the geographic context of the events.

Command and control in the US electric grid is a challenging task. First, the grid is a very large system with about 50 thousand substations and lots of different types of other equipment. Second, this is a very tightly interconnected system with processes in one part of the system may be affecting operation of the other part of the system. The opposite is possible too, when nearby equipment may not be closely coupled. Third, the system requires real-time operation and control with reaction times on the order of seconds. This creates the requirement of system-wide real-time situational awareness that cannot be achieved at the utility level and is aimed at improving reliability and preventing major disruptions like 2003 blackout. The other aim is to prepare for the demands of transitioning of the U.S. electric grid infrastructure to new technologies like renewable generation, smart grid controls, energy storage that are expected to come online in the coming decade.

Main components of the electric grid are generation, transmission, distribution, and load. Generation and distribution are connected via transmission lines to electric buses that are located at substations. The sensors can measure a variety of parameters such as the status of the line, frequency, voltage, and other and are installed on the buses at substations. Data are typically collected from sensors at 60 measurements per second, timestamped, and georeferenced by substation. Frequency is a very important parameter that can convey lots of information about the state of the grid. The higher frequency indicates excess of generation and lower frequency indicates the lack of generation or excessive demand even though this relation is nonlinear. The primary goal of grid operators is to balance generation and demand. In geographic terms electric grid can be seen as a network with busses as nodes and transmission lines as links.

Such large amount of dynamic spatio-temporal data coming from the sensors creates a need for an animated map. However, it has been demonstrated in cartographic research that animated maps may produce a high cognitive load

on users and limit their ability to comprehend the situation. A data stream with 60 new maps per second is well beyond human cognitive abilities. To address this problem, we are proposing a solution that involves a combination of several computational, visualization, and cartographic techniques.

First, the stream of sensor data that comes at 2-12 TB/hr is processed by an event detection code that is able to distinguish between normal and abnormal operation of the system. In case when abnormal operation is detected, the state of the grid is matched against a database of disruption scenarios. The database of disruption scenarios was created by running a simulation of the grid under various stress conditions defined as the number and locations of the failed buses. Each disruption scenario is identified by the set of failed buses. The matching algorithm finds 5-6 of the closest matches that are visualized. Operators have to understand what is going in the grid and take mitigation actions such as dispatching repair crews, rerouting power transmission and overriding automatic controls if necessary.

General requirements for the visualization system include fast operation (the data should be visualized within 2-4 seconds after collections), the visuals should be suitable for large display walls and be readable from a distance. The state of the grid should be presented in the geographic context, i.e, on an easy to recognize background map and can be overlaid with other geographic information like extreme weather events. The most important requirement for the visuals is the ability of operators to discern important patterns in the electric grid data.

To address this requirement, we have designed a map of a disruption scenario with the following goals: (1) the map should be able to convey a general pattern of the spatial distribution of frequency deviations rather than providing means to read deviations at specific substations, (2) the operators should be able to distinguish between the scenarios with small differences, (3) areas of positive and negative frequency deviations should be clearly visible, and (4) the design should be suitable for large display walls, desktops, and mobile devices.

The substation symbol for the map was designed to show bivariate data as a combination of an indicator of positive and negative deviations and the magnitude of deviations. To show deviation magnitude, we use a diverging color ramp created with the ColorBrewer. Selection of the color was motivated not so much by the aesthetics but by the need for high contrast with the background on a display wall.

The first difficulty developing color scale was the uneven distribution of frequency deviations not only among different scenarios but even within a single scenario for positive and negative deviations. As a result, it is not possible to use linear or logarithmic color scale because they will hide the majority of deviations and will not be effective at showing the general pattern. To address this problem, we have selected 10 most representative scenarios and combined their values of frequency deviations in a single pool. Then we have used Jenks (or natural breaks) optimization on the positive and negative deviation separately. Even though Jenks optimization has been originally developed for choropleth maps, the method is well applicable for the map of substations because the substations form dense enough network to fill most of the space almost completely.
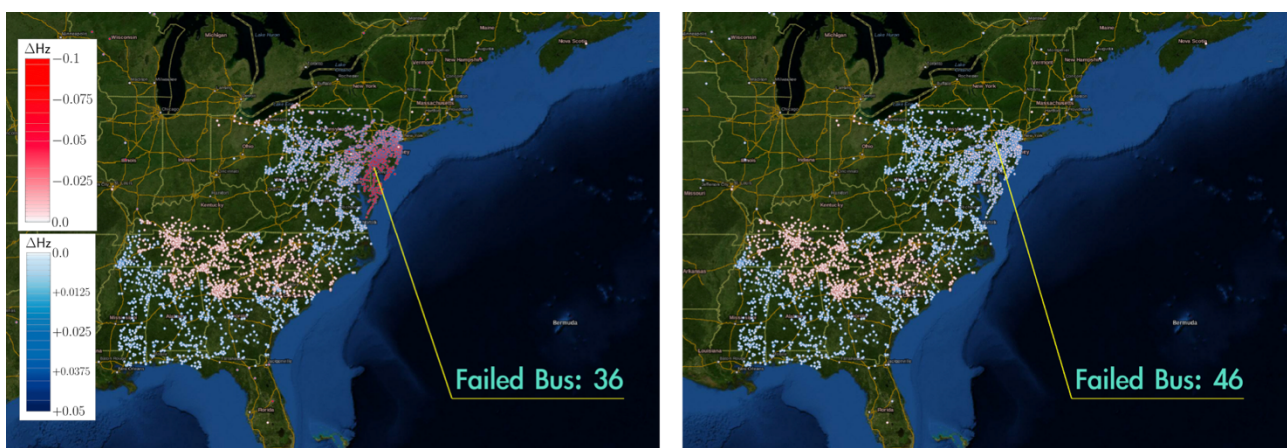

Figure 1. Two examples of disruption scenarios portrayed using proposed visualization technique.

The maps for two disruption scenarios are shown on Fig. 1. They show a selection of about 10,000 substations of the Eastern Interconnect. The map of the disruption scenario #36 demonstrates the ability of the proposed technique to clearly show the regions of positive deviation, negative deviations of low magnitude and negative deviations of high magnitude. Scenario #46 provides an example of a scenarios with deviations of low magnitude for both positive and negative parts of the spectrum. However, our approach still allows the user to discern regions of positive and negative deviations.

We have demonstrated application of advanced cartographic approaches to visualization of the U.S. electric grid. To address the challenges of high cognitive load of animated maps, we have applied an approach to combine data preprocessing, event detection, and several advanced cartographic techniques. In the future, the proposed methodology can be generalized to other types of dynamic big data.

Acknowledgements and Copyright