Dr. Eick

COSC 3337 *"Data Science I"* Fall 2022
Group Project (group size 4 students)
POIMAGIC: an Early Warning Systems for Streaming Spatial Events

Last Updated: November 5, 2022, 2p
Group Project is due: Friday, November 22, 11:59p
Responsible TAs: Raunak Sarbajna

Learning Objectives:
1. Spatial Data Analysis
2. Density Estimation
3. Hotspot Discovery
4. Change Analysis
5. Visualization of Spatial Data
6. Software Reuse

The goal of the project is to create an Early Warning System called POIMAGIC for streaming spatial events based on Earthquake hotspots and then to study if the developed system can be reused to detect Earthquakes in a different region, with different amounts of available data.

The total data in each dataset will be subdivided into 12 batches which contain the locations of events for a time interval of 2 months; e.g. one batch might contain the locations of earthquakes which happened in September till October of 2020 and your system will use a sliding window approach to create input data for the POIMAGIC system. Since the earthquake dataset is subdivided into 12 batches for time intervals September 2020, October 2020,…,August 2022 and, assuming a window size of 3 batches, your system will create hotspots/warnings for the citizen who live inside the hotspots for batch1+2+3, batch 2+3+4, …,batch 10+11+12. We also assume a hotspot is a contiguous polygon[1] in a 2D longitude-latitude space for which the event density of

[1] However, you might create hotspots of "simpler" shapes, instead of polygons; e.g. rectangular hotspots or hotspots which are contiguous regions of grid cells.

points inside the polygon is above a user-defined density threshold. Your system should create two kinds of hotspots:

a. Small, very hot spots whose density is above a “high” density threshold d1
b. Large, more regional hotspots whose density if a above a “medium high” density threshold d2; d1>d2.

Finding proper density thresholds d1 and d2 to create those two kinds of hotspots is a problem you need to solve in this project. After you succeeded in having a system which creates hotspots based on a density threshold and selected proper density thresholds d1 and d2 for your system, —assuming 12 batches and window size is 3—POIMAGIC will create an animation of 10 images depicting hotspots over the same observation area. As the last task, you will summarize how the hotspots in these two 10 image animations—one for d1 and one for d2— change over time. If you develop software to assist with this kind of change analysis, you will get ***extra credit*** for that. Moreover, data visualization—particularly visualizing the hotspots on a map—plays an important role for the project. Extra credit will be given for developing techniques which visualize how the hotspots evolve/change over time.

The streaming Earthquake dataset is taken from the ANSS Comprehensive Earthquake Catalog (ComCat) API, which contains earthquake source parameters (e.g. hypocenters, magnitudes, phase picks and amplitudes) and other products (e.g. moment tensor solutions, macroseismic information, tectonic summaries, maps) produced by contributing seismic networks. You can explore the data here: https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php.

The dataset has 5 attributes. Their range is given in brackets.

1. *Time*. In UTC. Format: YYYY-MM-DDTHH:MM:SS.000Z. Indicate the date and time when the earthquake initiates rupture, which is known as the "origin" time. Note that large earthquakes can continue rupturing for many 10's of seconds.
2. *Latitude*. [-90.0, 90.0] Decimal degrees latitude. Negative values for southern latitudes. Coordinates are given in the WGS84 reference frame
3. *Longitude*. [-180.0, 180.0] Decimal degrees longitude. Negative values for western longitudes. Coordinates are given in the WGS84 reference frame.
4. *Depth*. [0, 1000] Depth of the event in kilometers.
5. *Mag*. [-1.0, 10.0] The magnitude for the event. Commonly a moment magnitude that is based on the scalar seismic-moment of an earthquake determined by calculation of the seismic moment-tensor that best accounts for the character of the seismic waves generated by the earthquake.

Finally, after you have developed your system for the earthquake dataset, you will study how easy or difficult it is to reuse your system for another Early Warning system for the Europe region for the period 2016 ~ 2018! You will receive ***extra credits*** for identifying any major earthquake within that time period, purely using POIMAGIC.

The required data can be found in the *Files* tab of the *Datasets and Code* channel in the COSC-3337 Teams page. earthquake_contiguous_usa_12batch.zip is the US dataset for your primary task, and Earthquake_2016_2018_all_eu.zip is the Europe wide dataset.

You must present your findings in a report of 8 ~ 10 pages, excluding references and appendices. In your report, you must clearly state which members contributed to the project, and what were their contributions. Each group must submit only one report. The report must also contain clearly labelled imagery of the hotspots and the change-over-time. Your report should be submitted in PDF format and formatted using the ACM camera-ready templates available at http://www.acm.org/publications/proceedings-template. You may use MS Word/Google Docs/LaTeX/etc. for your report. Additionally, you must also submit your entire, *working* code as a zipped folder.