CSE 6242 course project

Logistics for course project

Grading & Schedule:

- Proposal (7.5% of course grade) (due Mar 12, Friday)
- Proposal presentation (5%) (due Mar 12, Friday)
- Progress report (5%) (due Apr 2, Friday)
- Final poster presentation (7.5%) (due Apr 23, Friday)
- Final report (25%) (due Apr 23, Friday)

Logistics for Proposal

Grading scheme & Submission instructions:

- 60% for the literature survey
- 30% for innovation
- 10% for plan of activities
- For every ***Heilmeier question*** that's **not** mentioned, deduct 5%.
- You may consider organizing your proposal based on the Heilmeier questions (e.g., each section addresses one question)
- Your literature survey should have **at least 3** papers or book chapters per group member (outside of any required reading for the class).

Logistics for Proposal

Grading scheme & Submission instructions (continue):

- Your literature survey should have **at least 3** papers or book chapters per group member (outside of any required reading for the class).
  - Short papers, like PNAS, Nature, Science papers, count as 0.5.
  - Copying the abstract of the papers is obviously prohibited, constituting plagiarism.
  - For each paper, describe
    - (a) the main idea,
    - (b) why (or why not) it will be useful for your project, and
    - (c) its potential shortcomings, that you will try to improve upon.

Logistics for Proposal

Grading scheme & Submission instructions (continue):

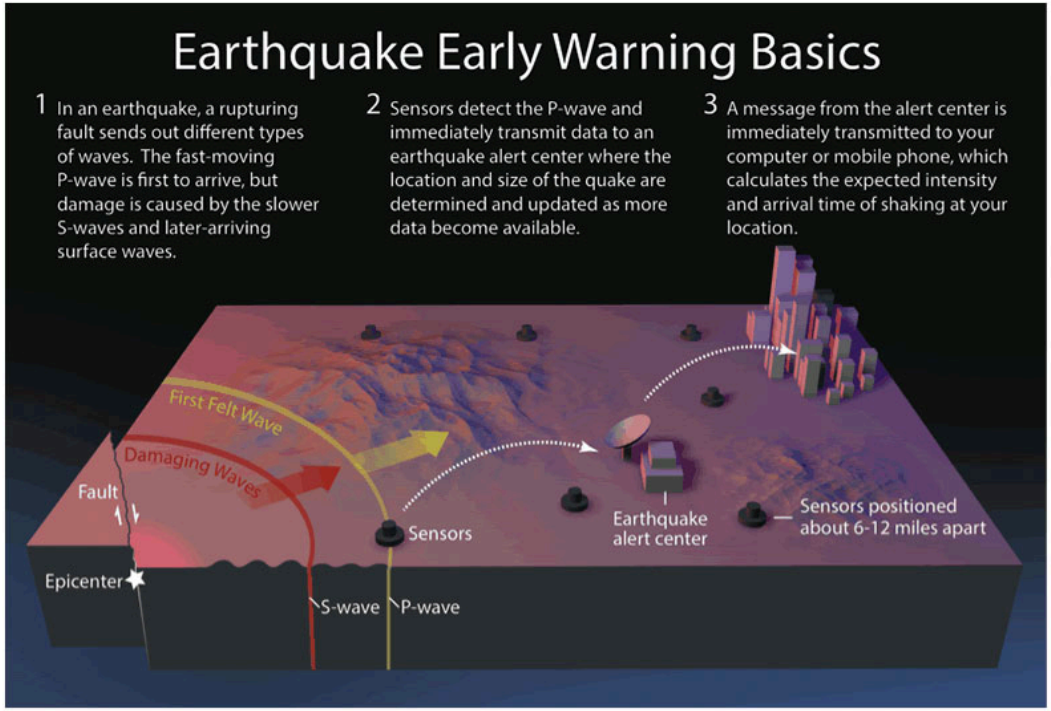
- You may use any citation style (e.g., APA, Chicago). Google Scholar supports a wide range of citation styles; it also provides BibTeX (needed if your team is using LaTeX).
- Please **make sure to cite your references** in your literature survey.
- **Clear problem definition**: give a precise formal problem definition, in addition to a jargon-free version (for Heilmeier question #1).
- Provide a **plan** of activities and time estimates, per group member. **List what each group member has done, and will do.**
- [-5% if not included] Distribution of team member effort. Can be as simple as "all team members have contributed similar amount of effort". If effort distribution is too uneven, I may assign higher scores to members who have contributed more.

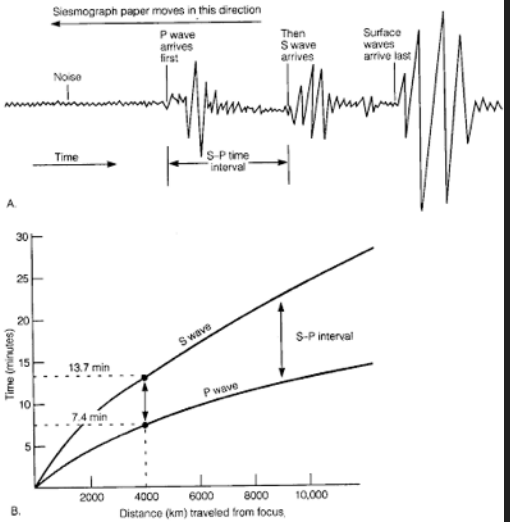
9 Heilmeier questions

- (1) What are you trying to do? Articulate your objectives using absolutely no jargon.
- (2) How is it done today; what are the limits of current practice?
- (3) What's new in your approach? Why will it be successful?
- (4) Who cares?
- (5) If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)?
- (6) What are the risks and payoffs?
- (7) How much will it cost?
- (8) How long will it take?
- (9) What are the midterm and final "exams" to check for success? How will progress be measured?

Earthquake Early Warning 101

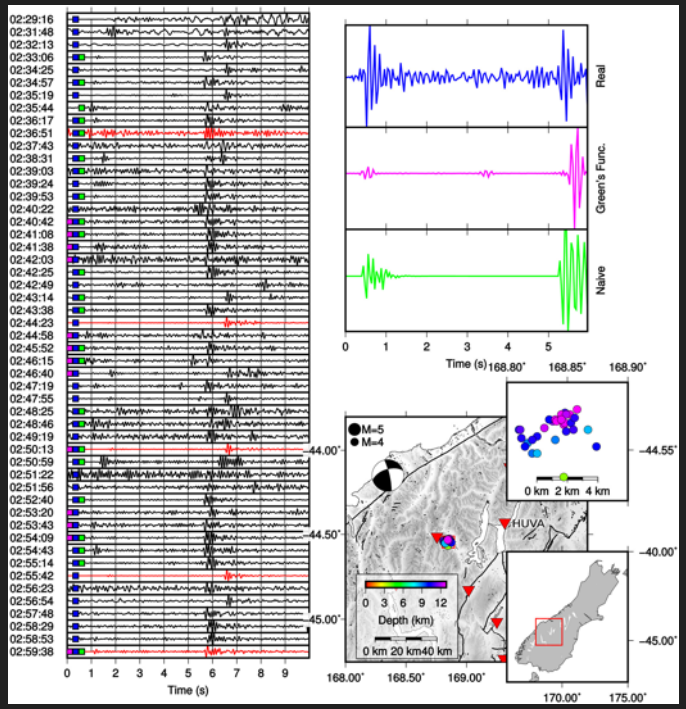
Earthquake early warning is NOT earthquake prediction!





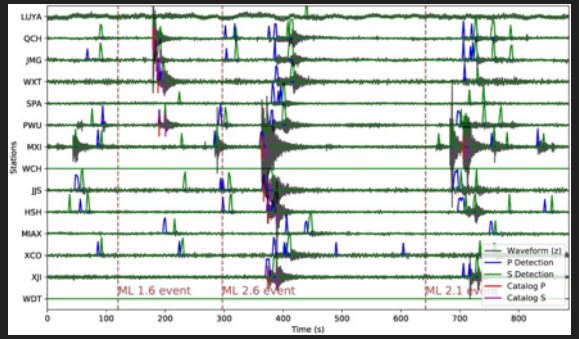
Earthquake Early Warning 101

Earthquake early warning depends on accurate and efficient automation tech.

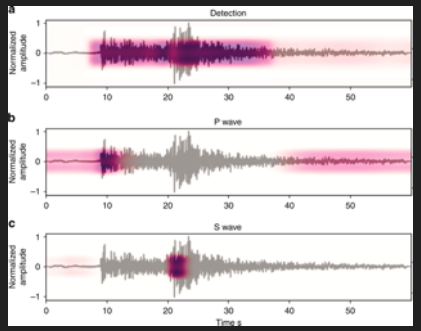


Earthquake Early Warning 101

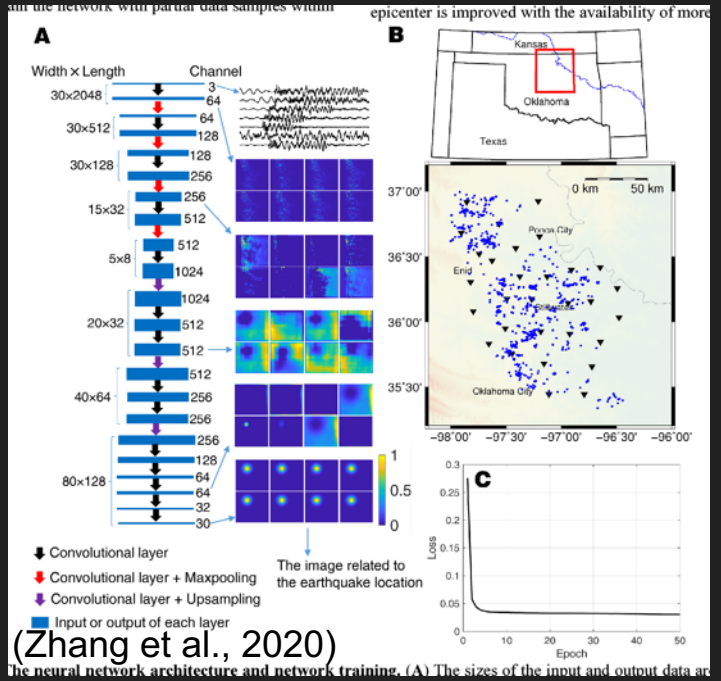
Deep learning as a great tool to automatically early-detect earthquakes



(Zhu et al., 2019)



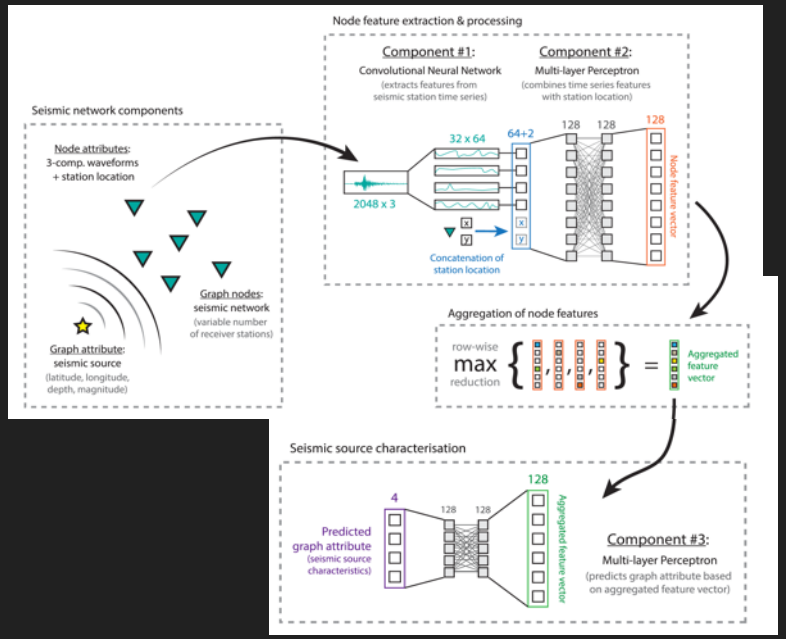
(Mousavi et al., 2020)



(Zhang et al., 2020)

Earthquake Early Warning 101

A little catch of deep learning based method

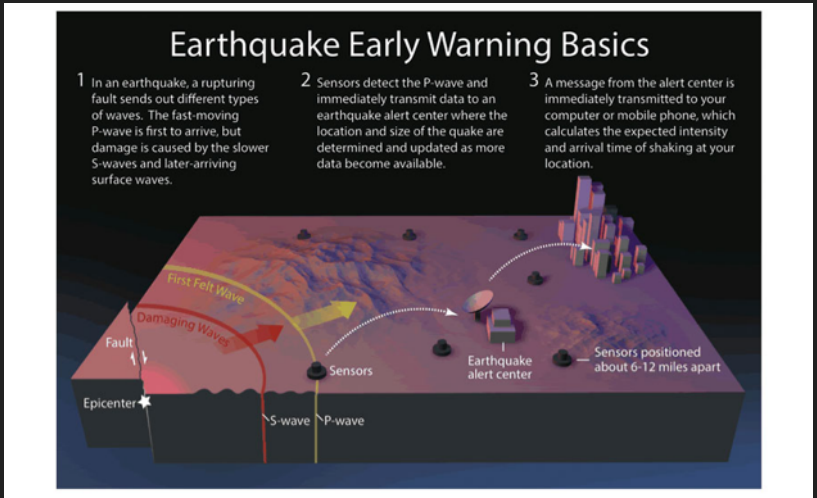


- Each station is a **node**
  - Each node carries two attributes (features):
    - A three-component seismic waveform time-series
    - Geographic location
- Nodes form **graphs**
  - Each graph carries four attributes:
    - Latitude, Longitude, Depth, Magnitude of the seismic source

The objective: Through aggregation of node attributes, predict the graph attributes

Earthquake Early Warning 101

Can we reformulate an earthquake detection problem as a link prediction and/or clustering problem?



- Each station is a **node**
  - Each node carries two attributes (features):
    - A three-component seismic waveform time-series
    - Geographic location
- Nodes form **graphs**
  - Each graph carries four attributes:
    - Latitude, Longitude, Depth, Magnitude of the seismic source

The objective: Through aggregation of node attributes, predict the graph attributes

Building blocks for this project

01 Data scraping

Input: study area (Lat. Lon.), station, earthquake catalog

Method: Obspy (Python)

Output : Time series (seismogram) in a structured way and meta data

02 Feature extraction

Input: Time series data

Method: Autoencoder(TensorFlow)

Hand-enginerred data (Sciki-learn)

Output : Features (compressed version of original data)/ compressed time series

03 Database and structure

Input: Features/compressed time series

Method: SQLite? Pandas? Hadoop?

Output : A database that works well with graph structure

04 Link and EQ prediction

Input: A database

Method: Graphical Neural Network/Clustering algorithm

Output : Updated database with

Node embeddings, linkages, and graph attribute

Building blocks for this project

05 Graph visualization

Input: Updated database

Method: D3.js? OpenGL? Three.js?

Output : A fancy, informative and interactive graph visualization of the database

06 Realtime deployment (optional)

Method: Some cloud environment?

Idea: Fetch data every hour an run through the detection scheme automatically?

Technical issues: Where to deploy? What tool to use?

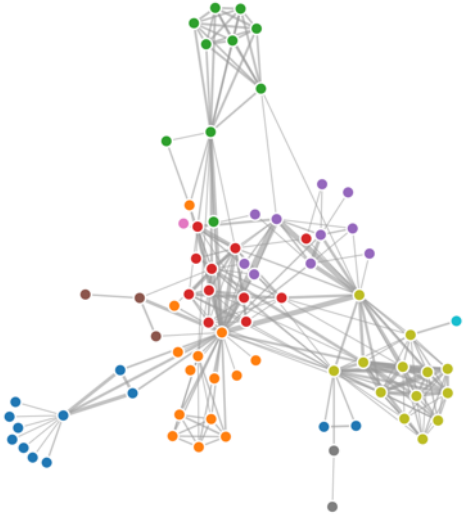
Is the data big?

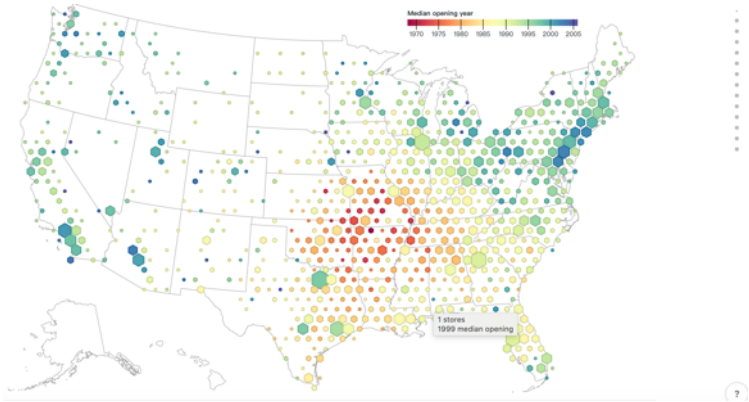
- https://www.iris.edu/app/station_monitor/#Today/PE-IUPA/map/
- In North America solely, each day there are:
- ~ 1600 publicly accessible real-time seismic stations that send continuous recordings back to data centers.
- One day of recording is ~15 Mb, so each day we are looking to process at least 24 Gb of data. This is only the ones that's publicly accessible in North America.
- Some of the special seismic network setups produce several Tb of data everyday.

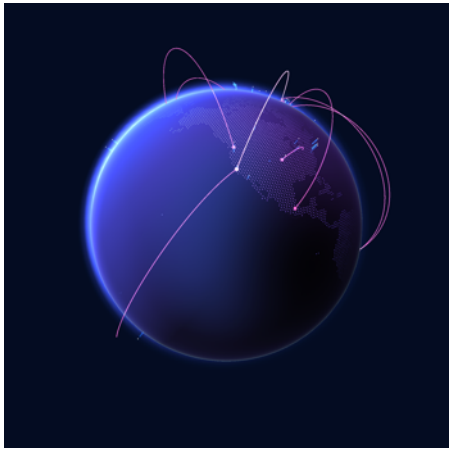
Links to datacenter

- **IRIS Seismic monitor** http://ds.iris.edu/seismon/index.phtml
- **IRIS Seismic station monitor** https://www.iris.edu/app/station_monitor/
- **USGS World earthquake catalog** https://earthquake.usgs.gov/earthquakes/map/?extent=16.46769,-137.19727&extent=55.22902,-52.82227
- **USGS earthquake hazard** https://earthquake.usgs.gov/
- **Obspy mass data downloader** https://docs.obspy.org/packages/autogen/obspy.clients.fdsn.mass_downloader.html

Graph Visualization





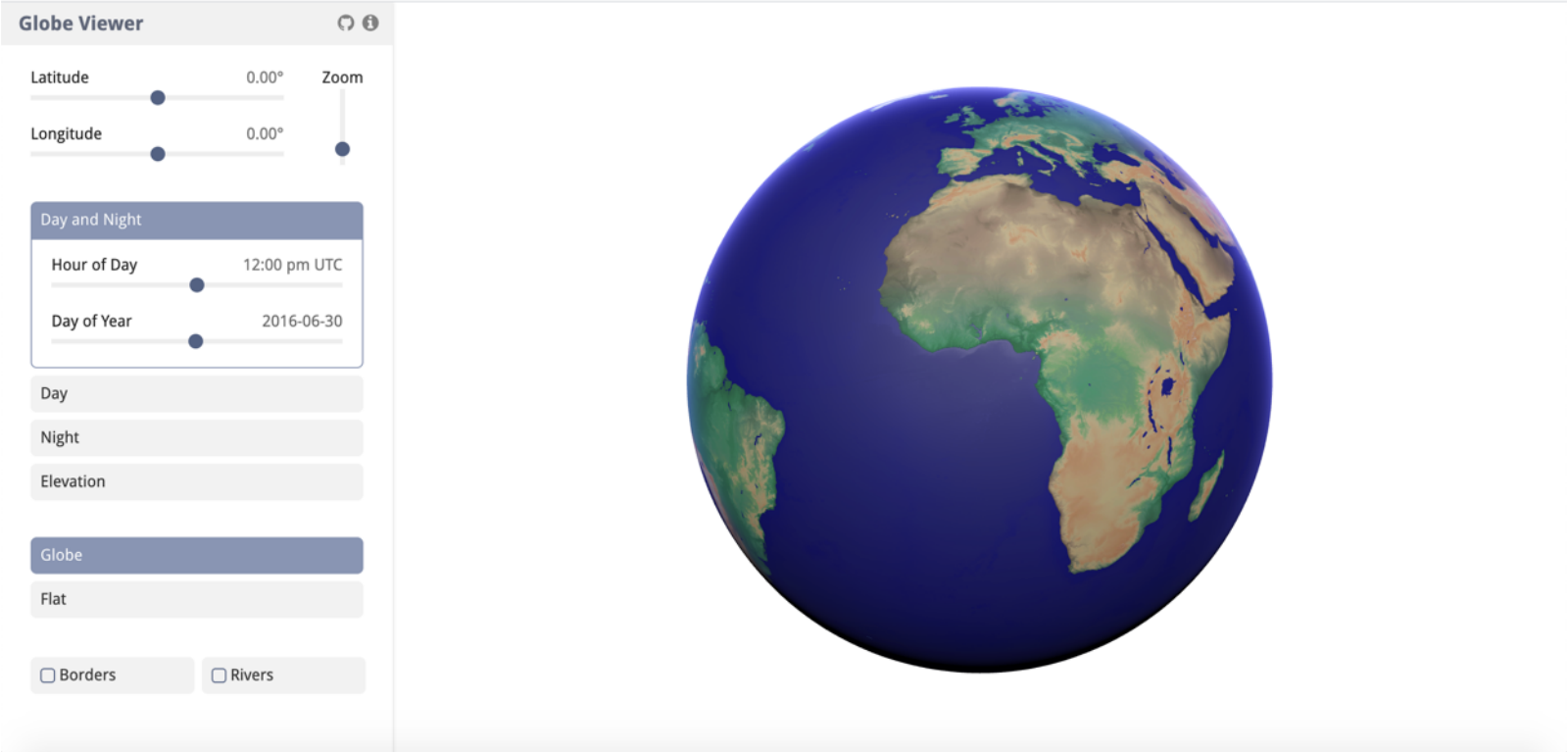


https://observablehq.com/@d3/force-directed-graph

https://observablehq.com/@d3/temporal-force-directed-graph

https://github.com/home

Graph Visualization



http://k9.github.io/globe-viewer/

Let's divide up tasks!

01 Data scraping : Lindsay

02 Feature extraction & 04 Link and EQ prediction (2 + Lindsay) :
George, Meghana

03 Database and structure (2) : Hsin-Yin, Gabriel

05 Graph visualization (2) :
Paarth, Preston

06 Realtime deployment (optional)

List of the co-working environments

Slack

GitHub

Overleaf

Google Colab

Data storage

Next meeting (two weeks from now?)

Date? Time?

Touchbase on the progress of proposal:

- literature review
- tools to use
- technical issues
- new ideas, asking for help, anything people want to talk about

Second meeting

Logistics for course project

Grading & Schedule:

- Proposal (7.5% of course grade) (due Mar 12, Friday)
- Proposal presentation (5%) (due Mar 12, Friday)
- Progress report (5%) (due Apr 2, Friday)
- Final poster presentation (7.5%) (due Apr 23, Friday)
- Final report (25%) (due Apr 23, Friday)

Proposal & proposal presentation

Overleaf : https://www.overleaf.com/project/602175e8e3fb1b1158e3ef4f

Proposal:
(1) Text : Get it done by Tuesday March 9 (?)
(2) Tools to use: ASAP (need them for presentations)
(3) Volunteers to proofread and restructure before submission (1-2 people)

Proposal presentation :
(1) a few slides

(2) 2-minute-long presentation video

Progress report (due Apr.2)

Overleaf : https://www.overleaf.com/project/602175e8e3fb1b1158e3ef4f

4-pages

- [70%] for proposed method (should be almost finished)

- [25%] for the design of upcoming experiments / evaluation

- [5%] for plan of activities (please show the old one and the revised one, along with the activities of each group member)

- Clear list of innovations: give a list of the best 2-4 ideas that your approach exhibits.

Collaboration going forward (up to Apr 2)