CSE 515 – MULTIMEDIA AND WEB DATABASES
PHASE 1 REPORT

Group Members
Uma Sampath Mallampalli
Sravan Garipalli
Mohan Maddula
Pranay Sai Dadi
Anvesh Reddy Koppela

Abstract

This project is about analyzing spatiotemporal epidemic data to evaluate the impact of an infectious disease on different states in US. Analyzing this data during the early stages of an outbreak of a disease helps us to identify patterns among different states, make accurate predictions regarding the rate of transmissibility of a disease and take necessary intervention measures to alleviate the impact of the disease. The project starts with normalizing and quantizing the given epidemic simulation files and constructing epidemic word files to study the transmission of the disease. Secondly, to measure the impact of an outbreak of a disease in a particular state to its neighboring states, epidemic average and difference word files are generated by obtaining the neighboring state information from the connectivity graph. Lastly, the simulation data is visualized in the form of heat map highlight the states with minimum and maximum window strengths.

Key Words – Spatiotemporal, Epidemics, Normalization, Quantization, Window length, Shift length, Resolution, Connectivity Graph, HeatMap.

Introduction

This project deals with spatiotemporal epidemiological simulation data obtained from simulation software. Analysis on epidemic simulation datasets helps us to identify the rate of transmission of an epidemic disease in a defined population. It also helps us to make accurate predictions about the impact of an outbreak of a disease in particular state to its neighboring states. Moreover, after the analysis necessary intervention measures can be taken to alleviate the effect of a disease.

Terminology

- Epidemiological – The science that studies the pattern, and effects of disease outbreak in defined population [1].
- Spatiotemporal data – Data belonging to both space and time.
- Epidemic is defined as the number of new cases of a particular diseases in a given duration surpassing the expected rate [2].
- Normalization – The process of transforming the data to a specified range usually between [0 1] and [-1 1].
- Quantization – The process of mapping a large set of input values to small set of values either by rounding or categorizing the data into different levels.
- Window length – Window length specifies the number of consecutive simulations considered in epidemic simulation file for constructing the epidemic word.
- Shift length – Shift length specifies the number of simulations to be skipped before considering the next window.
- Resolution – It specifies the number of levels into which the data has to be quantized.
- Connectivity Graph – This graph represents all the states that are 1 hop distance away from the current state.
- Metadata – The data that is used to represent the data.

Goal Description

The project has three tasks to be fulfilled.

Task 1 is to implement a program that accepts the directory for epidemic simulation data files, window length, shift length and resolution as input parameters from the user and generates an epidemic word dictionary files for the given set of epidemic simulation data files. The epidemic word file has to be generated after normalizing and quantizing the normalized values into different levels derived from a Gaussian distribution.

Normalization -:

Normalize the data file values between 0.0 to 1.0. Since, we are dealing with different data files with different range of values, all the data files have to be transformed to the same range to extract meaningful information. Data normalization is also required to nullify the influence of wide range of values on small range of values. For example, if one feature has a range of values between [0 1] and another feature with range [0 1000] a small variation in feature 2 will have more influence than big variation in feature 1 [3]. In this project we have used max – min normalization to transform the data value into the range of [0 1].

Quantization -:
Quantization is used to map a large set of input values to a smaller set of values [4]. In this project we have used Gaussian distribution to divide the input data range into r different levels. The level representatives for each band are extracted from a Gaussian distribution with mean 0 and standard deviation 0.25. In this project, quantization is used is used to categorize the disease into different severity levels. If the input resolution value r is more the disease has more levels of severity.
By quantizing the data, we can identify the severity level of the disease of a state at any particular iteration. Identifying the severity level of the disease, we can take necessary intervention measures to alleviate the effect of the disease.

Epidemic Word file -:
The epidemic word file is generated based on the window length w, shift length s, resolution r. The epidemic word file is used to track the transmission of a disease. If the dataset iterations are considered analogous to days, depending on the window length, the epidemic word files determines the number of simulations observed. Shift length signifies the time period between two consecutive observations.

- If the window length and shift length are large, the user want to track the progress of the disease less frequently.
- If the window length and the shift length are small, the user want to track the progress of the disease more frequently.
- If shift length is more than the window length, the user want to skip certain simulation and track the disease progress.

Task 2 is to implement a program that reads a connectivity graph and an epidemic word file and alpha value as input parameters from the user to compute the average and difference word files. The average and difference words are generated by considering the impact (alpha value) of 1-hop neighbors of the state. The user is given a provision to determine the impact of neighboring states on the current state through alpha value.

Epidemic Average file -: The epidemic average word file is used to determine the impact of the disease on the neighboring states and the current state. Depending on the values in the average file, we can accurately determine the necessary intervention methods such as closing the transportation service between the current state and its neighboring states.
The level of impact is determined the alpha value.

- If the value of alpha is more, the neighboring states are given less importance in determining the impact of the disease.
- If the value of alpha is less, the neighboring states are given more importance in determining the impact of the disease.
- If the current state does not have any neighboring states, the impact of the disease is determined by the alpha value.

Epidemic Difference file -: The difference file is used to measure the residual severity of the disease in

the current state from its neighboring states. This file

- If the value in difference file is more, the disease has not yet spread across its neighboring states or the neighboring has not influenced the current state.
- If the value in difference file is less, the disease has spread across its neighboring states or the neighboring states are impacting the current state and necessary intervention measures have to be taken.

Task 3 is to visualize the epidemic simulation file in the form of a heat map and highlight the highest and lowest strengths with respect to either epidemic word file, epidemic average word file or epidemic difference word file depending on the user choice.

Assumptions -:

- All the input simulation files are assumed to be in the same directory and the file naming convention is assumed to be same as the naming convention used in the sample dataset provided (1.csv, 2.csv).
- The simulation files have uniform meta data structure in all the files. The number of rows and columns in each file are also assumed to be the same and the states are also assumed to be in alphabetical order.
- The Gaussian levels are divided according to the resolution representing highest band length at the bottom and lowest band length at the top.
- The simulation iterations at end of the file are ignored if the number of iterations are does not fit the window size provided by the user.
- In task 1 the time(t) for idx (f, s, t) the following assumptions were made
  - File f is represented as the number associated with the file name.
  - State is represented as a number between 1 – 51 that represents the 51 states of US.
  - Iteration in the simulation file is considered as a representative for time.
  - For example, if a word is derived from the simulation file is 5.csv and the state is AL and the window starts from 16 iteration, idx (f, s, t) is represented as (5, 2, 16).
- For task 2, the connectivity graph is assumed to be provided in csv format and program computes the average and difference files by using the intermediate results obtained from task1.
- For task 3, in the heat map visualization, if there is more than one window with the highest and lowest window strengths, the window with first highest and lowest window strengths are visualized.
- Since one task is dependent on another tasks, all the tasks are assumed to be executed sequentially in the same order specified in the project description.
- All the input values are entered according to the specification in the project description.

Description of Implementation

A main script file is implement that accept all the inputs from the user required for the application. After accepting all the inputs from the user, this program calls individual matlab functions implemented for each task in the project.

Task 1

A matlab program is implemented that accepts the directory for epidemic simulation data files, window length, shift length and resolution as input parameters and performs all the below tasks for each of the simulation file present in the directory. As an output this task generates epidemic word dictionary files. The input and out for this task are mentioned below.

- **Input**: directory dir, window length w, shift length s, resolution r
- **Output**: epidemic word files generated in dir/output directory

Approaches Identified

- In the first approach, the normalization and quantization for the given data set is calculated separately, using two different scans on the enter dataset one for normalization and another for quantization. In the first scan, the data is normalized between 0 and 1 and once the normalized values are obtained another scan is used to quantize the data into predefined levels computed from Gaussian band. Epidemic word file is generated after normalizing and quantizing the data.
- In the second approach, the normalization and quantization are performed in a single scan. Since we only require quantized values we can first calculate the normalized value for the current entry and then assign it to the appropriate level in the same scan. Epidemic word file is generated after normalizing and quantizing the data.

I have used the second approach to reduce the intermediate meta data stored in the memory. This also have an advantage that both normalization and quantization can be performed in single scan.

Normalization

In this task the data entries in the simulation data file is normalized to a range between 0 and 1. Min – Max normalization is used to normalize the data within the range specified. Min – Max normalization distributes the data equally within the range 0 and 1. The formula used for Min -Max normalization is [5]

$$X' = a + \frac{(X - X_{min})(b - a)}{X_{max} - X_{min}}$$

- X' is the new data value after normalization
- X is the data value before normalization
- Xmin is the minimum value in the given epidemic simulation file
- Xmax is the maximum value in the given epidemic simulation file
- a is the minimum value in the new range
- b is the maximum value in the new range

In this project we are normalization the data values between 0 and 1 therefore after substituting the values of a and b in the above equation the formula becomes [2]

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Quantization

In this task the normalized values are categorized into different levels depending on the resolution r. The level length for each band is derived from a Gaussian band with mean 0 and standard deviation 0.25. Gaussian distribution is used because all the data entries in the simulation files does not have equal importance. Some values are treated as importance and some are not. Moreover, the data is not distributed linearly based on importance.

Since, the simulation files represents the severity of the disease, intervention measures are not required if the disease is below a certain limit. If the severity of the disease reaches above a certain value, then necessary measures have to be taken to alleviate the effect of the disease. A convenient way of categorizing the data into different levels is to extract the bands from a Gaussian distribution and transform the data with their level representatives that fall under respective bands.

Lengths for each level are obtained from the formula.

$$length_i = \frac{\int_{(i-1)/r}^{i/r} Gaussian_{(\mu=0.0,\sigma=0.25)}(x) \ \ \delta x}{\int_0^1 Gaussian_{(\mu=0.0,\sigma=0.25)}(x) \ \ \delta x}$$

The Gaussian distribution is given by the function [2]

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

This Gaussian distribution is used to obtain the levels in the data. Each level is represented using center of the band. All the entries in the data distribution are represented using these level representatives.

Code Approach

A matlab code is implemented to compute both normalization and quantization that requires single scan on the entire data to transform it into a new representation. For each entry,

- The normalized value is calculated based on the formula mentioned.
- A linear scan is performed on the different levels obtained from the Gaussian band and the representative for the level the current entry falls into is assigned to entry.

Epidemic Word File Generation

This file is generated by considering the window length, shift length and resolution entered by the user. The records in this word file are represented using the structure (idx, win) where (idx) refers to the set (file, state, time) and (win) is represented by the consecutive simulation entries in the epidemic simulation word file.

- File in idx is represented using the filename without file extension for example 1 for 1.csv file.
- State is represented using the number between 1 – 51 for the 51 states in US.
- Time is represented by the starting iteration value of the window.

Approaches identified

- In the first approach, I have stored the actual file name (1.csv, 2.csv), state (AK, AL) and iteration number for the window idx (f, s, t). This has recorded low in performance and created problems in the next tasks.
- In the second approach, I have made additional assumptions about the data that were mentioned in the assumptions section. I have represented files with filename without the file extension, and state with the number between 1 – 51 (each number representing a separate state).

I have followed the second approach because, it has improved the performance and reduce the preprocessing required in the further tasks.

Code Approach

Matlab code is implemented to generate the epidemic word file after normalizing and quantizing the data. The following approach is used to generate the word file. As a prerequisite the state representatives, levels and the level representatives are calculated and stored in memory.

- For each file in the epidemic simulation dataset directory,
  - Store the file name, number of rows, number of columns in the memory.
  - Separate the meta data and the simulation entries.
  - Normalize and quantize the data entries as mentioned above.
  - Perform a single scan on each column of the simulation entries and generate the epidemic word only if the number of residual entries (after advancing forward of shift length's') is more than or equal to the window length.
  - For each word generated obtain the filename, state and iteration stored in memory and build a matrix with the metadata information idx (f, s, t) along with window values.
  - Generate csv file for the matrix using matlab functions and store it in a directory with name as filename_epidemic_word_file.csv.

Task 2

A matlab program is implemented that accepts the connectivity graph and alpha value as input and performs all the below tasks for each of the epidemic word file generated as part of task 1. As an output this task generates epidemic average and difference word dictionary files.

The input and out for this task are mentioned below.

- **Input**: connectivity graph, alpha value
- **Output**: epidemic average files and epidemic difference files in dir/average and dir/difference directory

Epidemic Average Word File

This file is generated by considering the impact of 1 hop neighbors for each state in epidemic word file.

The impact of the neighboring states is determined by the value of alpha.

- If the value of alpha is less, the neighboring states are given more importance
- If the value of alpha is more, the current state is give more importance in measuring the impact of the disease.

The formula used for generating words in epidemic average word file is

$$\vec{win}_{avg,i} = \left(\alpha \times \vec{win}_i\right) + \left((1-\alpha) \times AVG\left\{\vec{win}_j \mid \left(s_j \in 1HN(G, idx_i.s)\right) \wedge \left(idx_j = \langle idx_i.f, s_j, idx_i.t\rangle\right)\right\}\right)$$

- $Win_{avg,I}$ – Represents the average window in the epidemic average word file.
- Win_i – Represents the window in the epidemic word file.
- α – Represents the impact value entered by the user $0 \leq \alpha \leq 1$
- $AVG\left\{\vec{win}_j \mid \left(s_j \in 1HN(G, idx_i.s)\right) \wedge \left(idx_j = \langle idx_i.f, s_j, idx_i.t\rangle\right)\right\}$ Represents the average window computed from the 1 hop neighboring states obtained from connectivity graph.

The records in this word file are represented using the structure (idx, win) where (idx) refers to the set (file, state, time) and (win) is represented by the average value computed for the window in the neighboring states.

- File in idx is represented using the filename without file extension for example 1 for 1.csv file.
- State is represented using the number between 1 – 51 for the 51 states in US.
- Time is represented by the starting iteration value of the window.

Example

If the epidemic word files contains the following words in the form of (file, state, iteration, window)
(1, 1, 1, (1,2,3)) (1, 1, 2, (4,5,6)) (1, 1, 3, (7,8,9))
(1, 2, 1, (1,2,3)) (1, 2, 2, (4,5,6)) (1, 2, 3, (7,8,9))
(1, 3, 1, (1,2,3)) (1, 3, 2, (4,5,6)) (1, 3, 3, (7,8,9))
(1, 4, 1, (1,2,3)) (1, 4, 2, (4,5,6)) (1, 4, 3, (7,8,9))
(1, 5, 1, (1,2,3)) (1, 5, 2, (4,5,6)) (1, 5, 3, (7,8,9))

If the state 1 has 3 and 5 as its neighboring states. Then the average windows for 3 and 5 are computed as
((1,2,3) (4,5,6) (7,8,9)) + ((1,2,3) (4,5,6) (7,8,9)) / 2 = ((1,2,3) (4,5,6) (7,8,9))

If the value of alpha is 0.8 average window is computed as

0.8 * ((1,2,3) (4,5,6) (7,8,9)) + 0.2 * ((1,2,3) (4,5,6) (7,8,9)) = ((1,2,3) (4,5,6) (7,8,9))
Therefore, the word (1, 1, 1, (1,2,3)) (1, 1, 2, (4,5,6)) (1, 1, 3, (7,8,9)) is placed in epidemic average word file.

Epidemic Difference Word File

This file is generated to identify whether the neighboring states are impacted by a disease outbreak in the current state.

- High values in difference word file indicate that the disease has not spread across its neighboring states. Therefore, intervention measure should be taken only in the current state.
- Low values in difference word file signifies that the disease has spread across its neighboring states. Therefore, intervention measure should be taken both in the current state as well as neighboring states.

The formula used for generating words in epidemic difference word file is

$$\vec{win}_{diff,i} = \left(\vec{win}_i - \left(AVG\left\{\vec{win}_j \mid (s_j \in 1HN(G, idx_i.s)) \wedge (idx_j = \langle idx_i.f, s_j, idx_i.t\rangle)\right\}\right)\right) \div \vec{win}_i$$

- Win $_{diff,I}$ – Represents the difference window in the epidemic difference word file.
- Win$_i$ – Represents the window in the epidemic word file.
- $AVG\left\{\vec{win}_j \mid (s_j \in 1HN(G, idx_i.s)) \wedge (idx_j = \langle idx_i.f, s_j, idx_i.t\rangle)\right\}$ Represents the average window computed from the 1 hop neighboring states obtained from connectivity graph.

Example

If the epidemic word files contains the following words in the form of (file, state, iteration, window)
(1, 1, 1, (1,2,3)) (1, 1, 2, (4,5,6)) (1, 1, 3, (7,8,9))
(1, 2, 1, (1,2,3)) (1, 2, 2, (4,5,6)) (1, 2, 3, (7,8,9))
(1, 3, 1, (1,2,3)) (1, 3, 2, (4,5,6)) (1, 3, 3, (7,8,9))
(1, 4, 1, (1,2,3)) (1, 4, 2, (4,5,6)) (1, 4, 3, (7,8,9))
(1, 5, 1, (1,2,3)) (1, 5, 2, (4,5,6)) (1, 5, 3, (7,8,9))

If the state 1 has 3 and 5 as its neighboring states. Then the average windows for 3 and 5 are computed as
((1,2,3) (4,5,6) (7,8,9)) + ((1,2,3) (4,5,6) (7,8,9)) / 2 = ((1,2,3) (4,5,6) (7,8,9))

If the value of alpha is 0.8 average window is computed as

((1,2,3) (4,5,6) (7,8,9)) - ((1,2,3) (4,5,6) (7,8,9)) / ((1,2,3) (4,5,6) (7,8,9)) = ((0,0,0) (0,0,0) (0,0,0))

Therefore, the word (1, 1, 1, (0,0,0)) (1, 1, 2, (0,0,0)) (1, 1, 3, (0,0,0)) is placed in epidemic difference word file.

The records in this word file are represented using the structure (idx, win) where (idx) refers to the set (file, state, time) and (win) is represented by the relative difference between the current state and its neighboring states.

- File in idx is represented using the filename without file extension for example 1 for 1.csv file.

- State is represented using the number between 1 – 51 for the 51 states in US.
- Time is represented by the starting iteration value of the window.

Approaches Identified

- In the first approach, for each word in the epidemic word file, the following steps are performed.
  - The neighboring states for the current state are obtained from connectivity graph.
  - A linear scan is performed on the entire epidemic word file to identify the windows with respective state and iteration of the current word in consideration.
  - Average value for the identified windows is computed and the average and difference word files are generated from the results.
  - This procedure involves linear scan on the entire data for each word in the epidemic word file. Therefore, if the epidemic word file contains N entries then the run time complexity of this approach is N^2.
  - For example, if the average for the word (f, s, t, window) = (1, AL, 5, window) has to be calculated, identify the neighboring states for the state AL from the connectivity graph. If AL has 3 neighboring states, then all the words (f, s, t, window) = (1, s, 5, window) has to be determined by doing linear scan on the entire file. Here's' refers to the state id of neighboring states.
- In the second approach, preprocessing of the epidemic word file is done to reduce the time complexity of the algorithm. In this approach I have used the advantage of exploring the structure of the epidemic word file. In the epidemic word file generated from task 1, for any given window length and shift length, the number of epidemic words for each state will be same. Using this information
  - Matrix operations in matlab can be used to perform bulk operations.
  - The two dimensional epidemic word file is transformed to a three dimensional space where third dimension is represented with the state and the remaining two dimensions are used to represent the window.
  - For each state, neighboring states are obtained from the connectivity graph and average computation is performed on all the states at once.
  - For example, if the following structure is identified in the epidemic word file (f,s,t,window)
    (1, 1, 1, window) (1, 1, 2, window) (1, 1, 3, window)
    (1, 2, 1, window) (1, 2, 2, window) (1, 2, 3, window)
    (1, 3, 1, window) (1, 3, 2, window) (1, 3, 3, window)
    (1, 4, 1, window) (1, 4, 2, window) (1, 4, 3, window)
    (1, 5, 1, window) (1, 5, 2, window) (1, 5, 3, window)
  - In the above example, the epidemic word file is represented in three dimensional format as state id x (dimension of two the matrix corresponding to that state)
    1 x (3x2)
    (1, window) (2, window) (3, window)
    2 x (3x2)
    (1, window) (2, window) (3, window)
    3 x (3x2)

(1, window) (2, window) (3, window)
4 x (3x2)
(1, window) (2, window) (3, window)
5 x (3x2)
(1, window) (2, window) (3, window)

- If we want to compute the average for the state id 1 and it has 3, 4, 5 as its neighboring states then we can simply do element wise addition for the matrices corresponding to the states 3, 4, 5 in the three dimensional representation.

I have used the second approach to compute the average and difference files because of low time complexity of the algorithm.

Code Approach

Matlab code is implemented to generate the epidemic average and difference word file after generating the epidemic word files. As a prerequisite for computing the average and difference files, the connectivity graph and alpha value is loaded into the memory. For each epidemic word file generated in the task 1

- Transform the epidemic word file into a three dimensional space with state id as the third dimension.
- For each state in the epidemic word file, compute its neighboring states from the connectivity graph.
- If the current state has neighboring states, compute the average of the windows corresponding to the neighboring states and apply the formulas mentioned above to compute the average and difference word files.
- Once the average and difference words are computed for all the states, generate the csv files for filename_epidemic_word_file_avg.csv and filename_epidemic_word_file_diff.csv in average and difference folders.

Task 3

Matlab code is implemented to visualize the epidemic simulation file in the form of heatmap and highlight the maximum and minimum windows strengths corresponding to either epidemic word file or epidemic average word file or epidemic difference word file depending on user's choice.

Code Approach

The program accepts the word file name to be visualized as an input parameter and generates a heatmap for the epidemic word files and highlight the maximum and minimum word strengths depending on user choice.

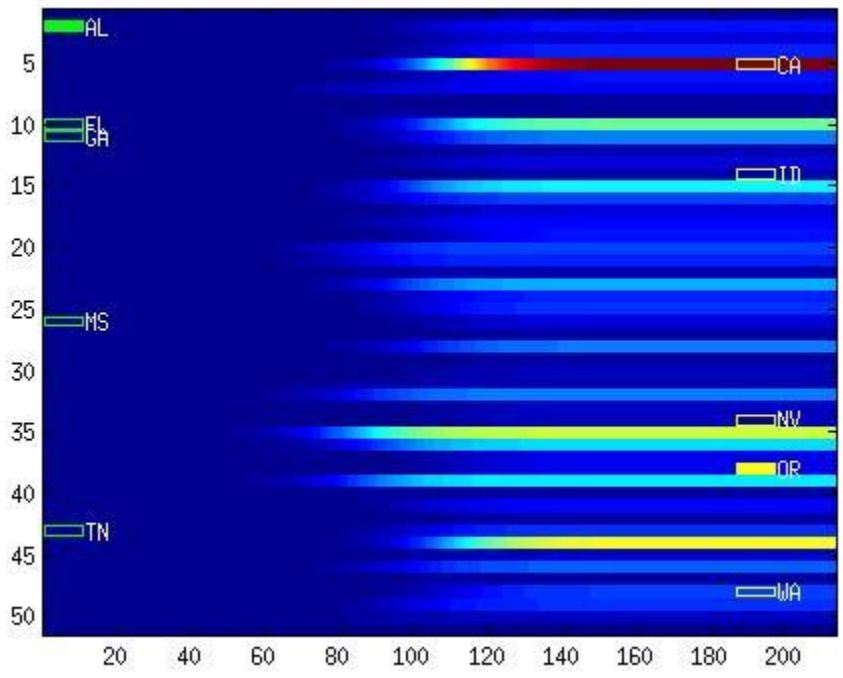
- The simulation file requested by the user is visualized in the form of heatmap.
- The program computes the window strengths for all the words in the file requested by the user. The word strengths are calculated using 2-norm.
- First occurrence of minimum and maximum window strengths and their corresponding states are identified in the requested file.
- The maximum and minimum windows highlighted using a rectangle, whose bottom left point

corresponds to the state and starting iteration of the window.

- Neighboring states for the states corresponding to the maximum and minimum window strengths are identified from the connectivity graph and visualized.
- The following color coding is used to represent the maximum and minimum window strengths
  - Maximum window strength is represented using a rectangle filled with yellow color.
  - Minimum window strength is represented using a rectangle filled green color.
  - Neighboring states corresponding to either minimum or maximum window strengths are represented using rectangles with border colors same as the colors used for maximum and minimum.
- The state codes corresponding to the window are show from a lookup table.

Sample output for this task is shown below.

- X axis represents the iterations in the simulation word file (ranges from 1 to 213).
- Y axis represents the states in the summation word file (ranges from 1 to 51).
- The rectangle filled with yellow color represents the window with highest strength.
- The rectangle filled with green color represents the window with lowest strength.
- The rectangles with green or yellow border represents the neighboring states for the minimum and maximum window strengths respectively.



In the above graph

- OR is the state with highest strength and CA, ID, NV, WA are its neighboring states.
- AL is the state with minimum strength and FL, GA, MS, TN are its neighboring states.

Interface Specification

- **Main.m** – This matlab script file takes all the inputs required for the application from the user and executes the tasks in the sequential order according to the specification.
- **generateEpidemicWordFile(datasetDir, w, h, r)** - This matlab function is used by task 1 which takes directory for epidemic simulation files (datasetDir), Window length (w), Shift length (h), resolution (r) as input parameters and normalize and quantize the data and generate the epidemic word files as output.
- **generateAvgAndDiffWordFiles(graphPath, wordfilePath, alphaValue)** - This matlab function is used by task 2 and takes the fully qualified path for the connectivity graph (graphPath), the directory for epidemic word files (wordfilePath) and the alpha value (alphaValue) as input parameters and compute the average and difference word files.
- **generateHeatMap(originalFile, filedata, windowSize, connectivityGraph, stateIndexes)** – This matlab function is used by task 3 and takes epidemic simulation file (originalFile), the epidemic word file or epidemic average file or epidemic difference file data (filedata), the window size (windowSize), connectivity graph (connectivityGraph) as input paramters and generate the heat map depending on user choice.
- **drawRectangle(connectivityGraph, choice, stateId, iteration, windowSize, stateIndexes)** - This function is used by task 3 to draw the windows with highest and lowest strengths in the form of rectangle. It takes connectivity graph (connectivityGraph), choice (min or max), state id corresponding to the highest or lowest (stateId), iteration id corresponding to the highest and lowest (iteration), window size (windowSize) as input parameters and highlight the data on the heatmap.

System Requirements

The programs are coded in the following environment and it requires the following minimum softwares.

- Matlab2013b.
- Linux 64 bit Operating System

Sample Execution Steps

- Run Main.m matlab script file
- Enter the directory location for epidemic word file without any quotation marks for the string. /home/user/Sample
- Enter the window length – 10
- Enter the shift length – 3
- Enter the resolution – 20
- Enter the complete qualified path (without quotation marks) for connectivity graph in csv format – /home/user/Sample/LocationMatrix.csv
- Enter the value for alpha $0 \leq \alpha \leq 1$
- Enter the filename to visualize the heatmap without the file extension – 1 for 1.csv
- Enter the choice to view the maximum and minimum window strengths

After entering all the values, the epidemic word files, average files, difference files are generated in the directories output, average, difference respectively. The program also visualizes the highest and lowest

window strengths.

Related Work

Title: Understanding and predicting the global spread of emergent infectious diseases
Author: Dirk Brockmann
Published In: Science Direct
Idea: The paper combines the theoretical insights from nonlinear dynamics, stochastic process and complex network theory to design efficient mitigation strategies for public health. The author also used sophisticated computer simulations to understand and predict the disease spread on a global scale.

Conclusion

In this phase we have studied and analyzed the rate of transmission of an epidemic disease, during its outbreak in a specific state. Patterns among neighboring states and their impact are also identified in this phase. By comparing the epidemic word files with epidemic average and difference files we can identify necessary intervention methods to reduce the impact of the disease.

References

[1] http://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/1-what-epidemiology

[2] Principles of Epidemiology, Second Edition. Atlanta, Georgia: Centers for Disease Control and Prevention

[3] http://www.researchgate.net/post/When_and_why_do_we_need_data_normalization

[4] http://en.wikipedia.org/wiki/Quantization_(signal_processing)

[5] http://en.wikipedia.org/wiki/Normalization_(statistics)

[6] Understanding and predicting the global spread of emergent infectious diseases by Dirk Brockmann published in Science Direct.

[7] http://stackoverflow.com/

Appendix

The approach for this phase is discussed among the group and the effectiveness of each approach is validated by the entire team. However, since this phase is an individual project and all the modules are developed by me.