SCALING SOCIAL NETWORKS AND FINDING CORRELATION BETWEEN INFLUENZA & WEATHER VARIABLES ALONGSIDE DATA AGGREGATION

Zachery Taylor Morris Mentor: Dr. Jiangzhuo Chen

Task 1: Write a program in C++ to scale a given social network

Assumptions/Background:

- That the network is given in an extended CSV format file.
- That the file starts with a JSON string header, a CSV header, followed by rows of bidirectional edges.
- That a node is represented by an ID number and an edge is represented by two ID numbers on the same row of the file but in adjacent columns.

What the Program Does:

- Makes a network that is *K* (specified by the user) times larger, by making a large network consisting of *K* copies of the given network, and rewiring *f* fraction (specified by the user) of randomly so the copies can become connected.

Algorithm/Steps:

1. Using command line parameters, read network edges from a file and store the undirected edge representation in a "edge" object then store these edge objects in a vector.
   a) "edge" contains both IDs (nodes) and their attributes.
   b) Make original graph non-bidirectional.
2. Make *K* copies and put them in a vector to form a large network.
   a) Have a helper function which creates new IDs (nodes) based on the beginning and ending interval of original IDs.
   b) Add original IDs to large vector.
3. Iteratively choose two edges randomly and swap IDs between them, effectively rewiring the edges between nodes.
   a) Generate random numbers with an interval equal to amount of edges to use as indexes in large vector.
   b) Take first two random numbers, get edges at these indexes, and swap the IDs.
4. Write the two header lines and both directions of each node to a file.

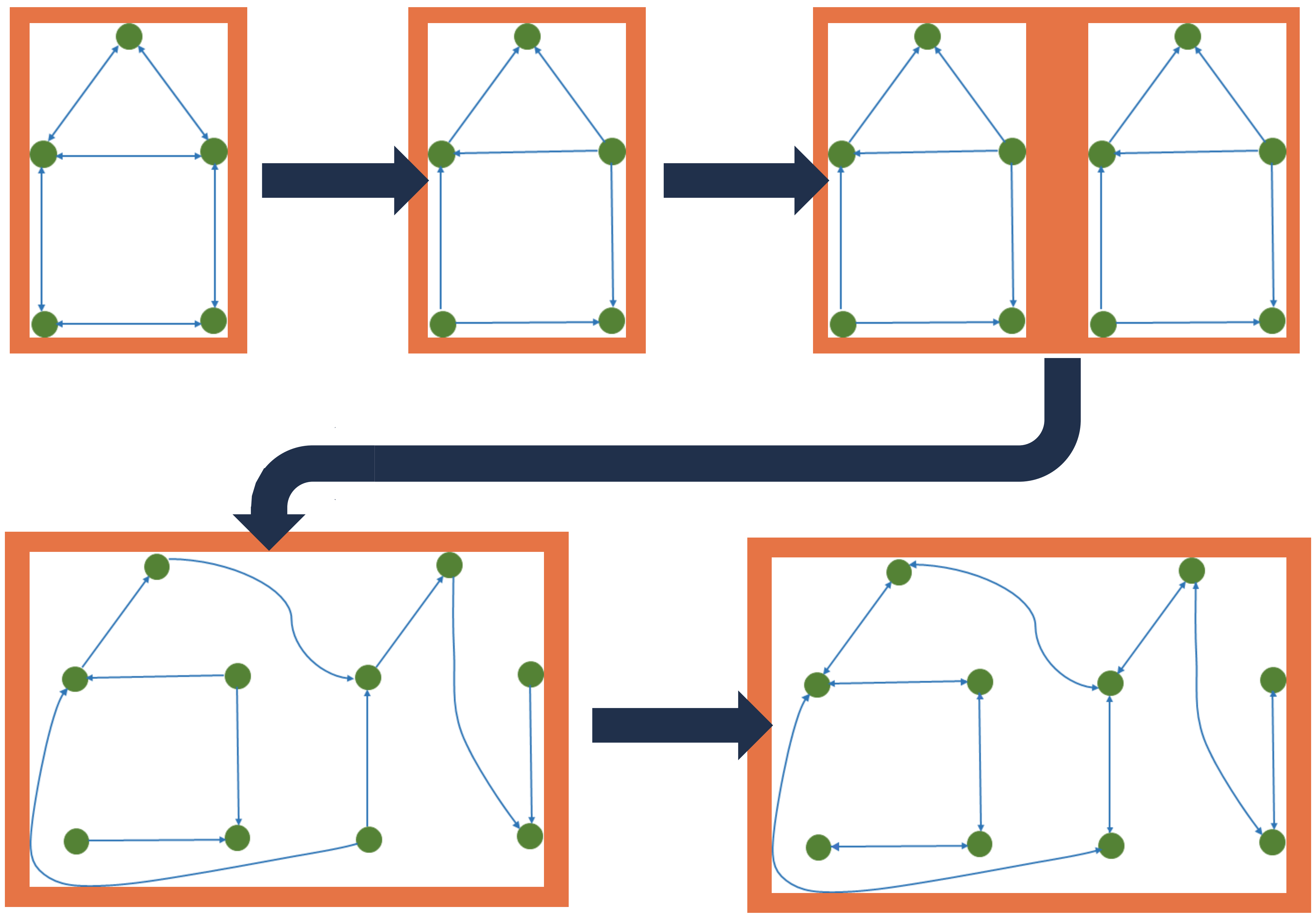


Figure 1: Depicts the transformation of the graph social network in "Task 1"

Task 2: Write a program in Python using Pandas that extracts and aggregates climate data

Assumptions/Background:

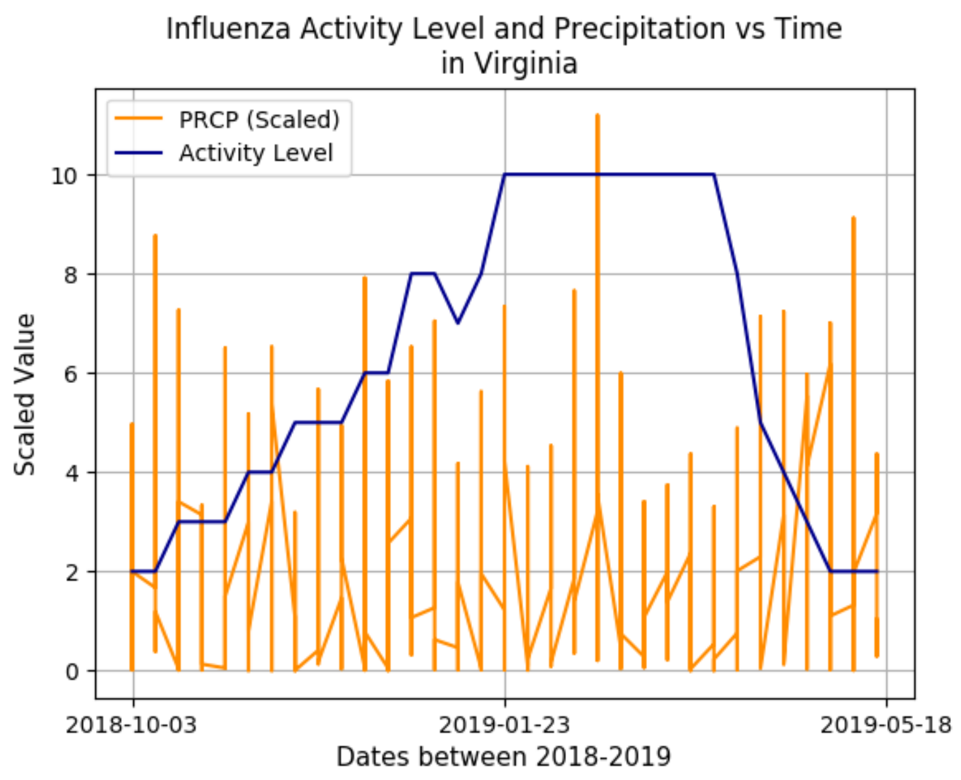
- FIPS code is a code that represents locations in the US. A state has a two-digit code. A county as a five-digit code where the first two digits represent the FIPS code of the state where the county belongs.
- Data in yyyy.csv is daily and by location. This file also has eight columns location_id, date, variable, value, MFLAG, QFLAG, SFLAG, and observation_time. We ignore the last four columns.

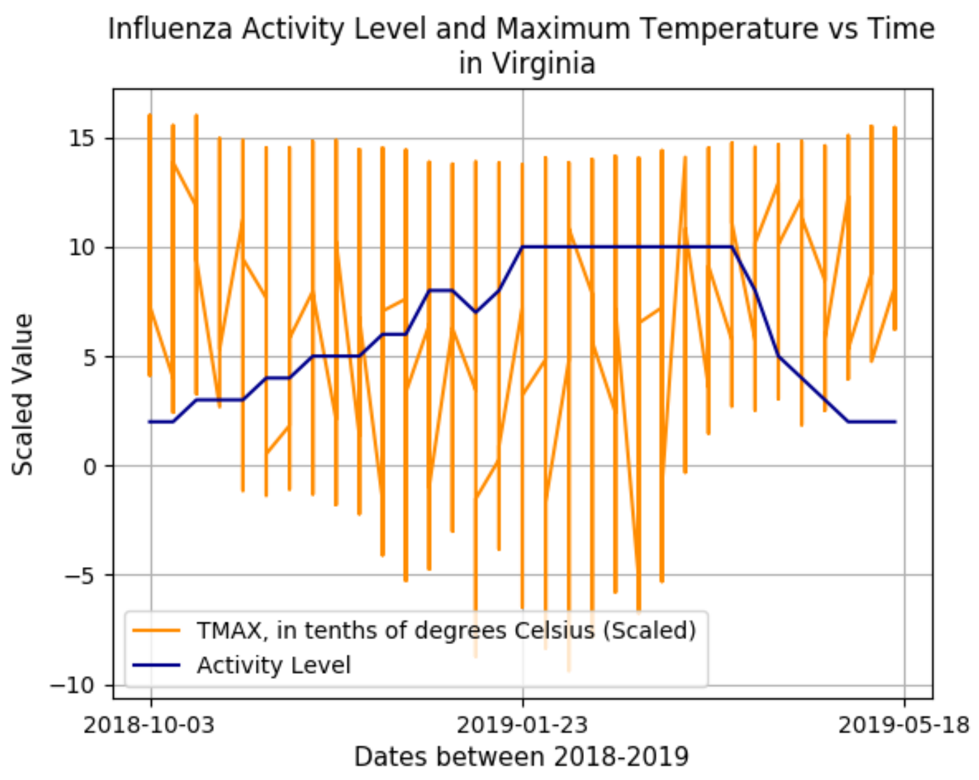
What the Program Does:

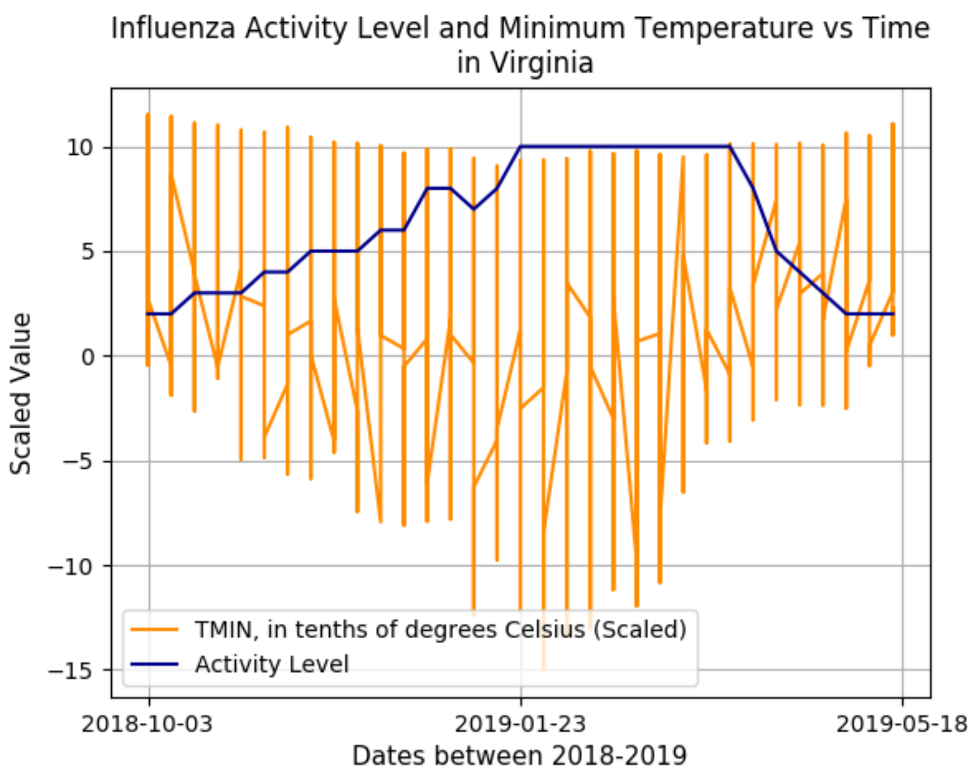
- Extracts data of a specified climatic variable for US counties and states while aggerating the data to spatial and/or temporal resolution (specified by the user).

Algorithm/Steps:

1. Implement command line parameters.
   a) *raw_data*: one-year daily observation data file, yyyy.csv
   b) *variable*: climatic variable, PRCP|TMAX|TMIN|TOBS
   c) *locations*: mapping from location ID to state or county FIPS code, county_fips.csv|state_fips.csv
   d) *spatial*: spatial resolution of output data, state|county
   e) *temporal*: temporal resolution of output data, daily|weekly
   f) *output*: the output file, output_data.csv
2. To aggerate the weather data to state or county resolution, the average of the climatic values are taken over all the locations in the state or county.
3. To aggregate the weather data to weekly resolution, we take the average of the climatic values over all days from Sunday to Saturday and label a week with its Wednesday date.
4. Output the aggerated data based on the inputted command line parameters in a csv file format. There are four columns: date, area, variable, and value.







Task 3: Visualize how influenza indicator variable varies with climatic variable

Influenza Indicator Variable with PRCP:

- At peaks of the influenza activity level, the precipitation also peaks.

Influenza Indicator Variable with TMAX:

- At peaks of the influenza activity level, the temperature max is lower than expected, it is otherwise mostly consistent.

Influenza Indicator Variable with TMIN:

- At peaks of the influenza activity level, the temperature minimum is also at its peak.

Computing *for* Global Challenges

UNIVERSITY *of* VIRGINIA
BIOCOMPLEXITY INSTITUTE