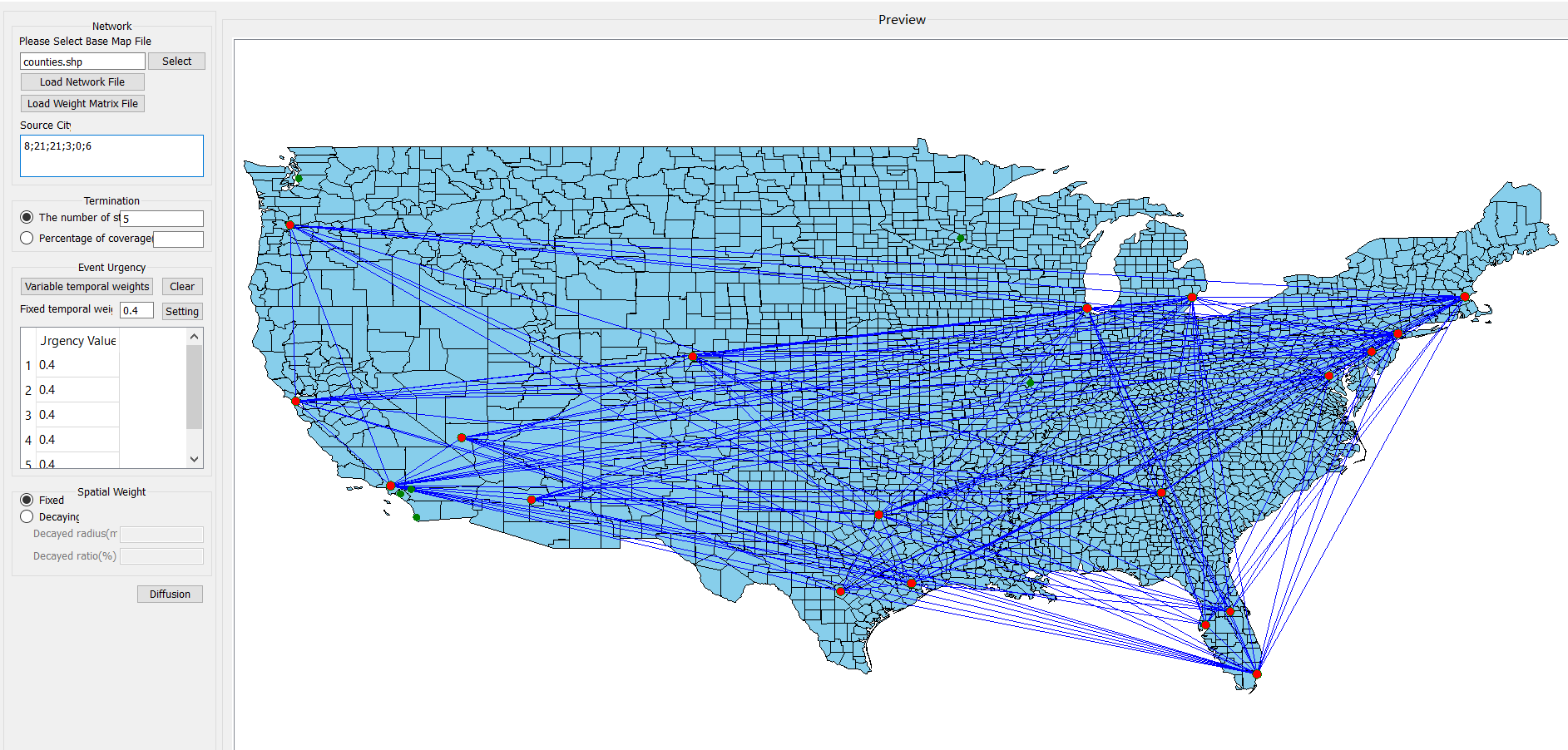
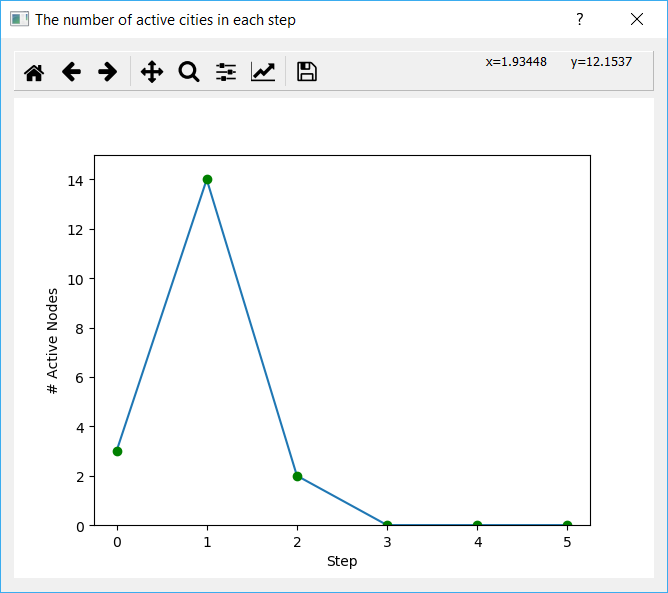
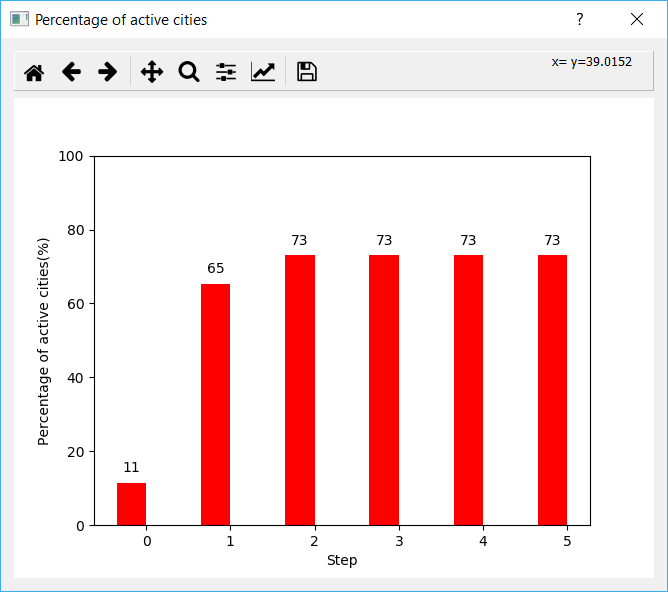
**Preparing inputs files for Information Diffusion Among Cities (City Level Diffusion) in the Social Network Simulator software**

1. There are two must-have input files to run the city level diffusion. **a)** City Network file (txt file) which contains the cities in the network and their geographic coordinates. It also has the information of the edges of the network. **b)** Weight Matrix file (txt file) which is the corresponding matrix file to city network. Each row contains a set of weights that express the city’s probability of being retweeted by other cities. The row is sorted by CITY\_ID in the city file.
2. To get the City Network, the originally raw file should have the information of city name, city ID and city XY coordinates (longitude and latitude). Then by running the *GenearteCityNetwork.py* under the *DataProcess* folder, one can get the City Network file as the result.
3. To get the Weight Matrix, one should have all the collected tweets files for all the cities. Sometimes, these files are originally in .xls format and somehow corrupted. By manually transforming .xls files into .csv files, these files then can be read. **a)** Go the *DataProcess* folder and run *merge\_csv\_files.py* to merge these files into on so that it can be processed together. **b)** with the result of Step a), run *ExtractUserLocation.py* to extract the location information of users**. c)** As the result only has *location name* instead of *location ID* that are used in the City Network, *CITY\_TO\_ID.py* is run to match the city name and ID and generates the *user\_locationid.csv* as the result*.* **d)** Finally, with *user\_locationid.csv,* the weight matrix can be generated by running *ExtractMatrix.py.*

**Results of running for Information Diffusion Among Cities (City Level Diffusion) in the Social Network Simulator software**

1. Following the instruction of User Manual, the settings are configured from the picture below for experiment:  
   
2. The results are in following pictures: One is the number of active nodes at each step, and the other is the accumulative percentage of coverage at every step. If there are green nodes on the map, that means these nodes were inactive.

As the picture indicates, the diffusion started from 3 cities which are around 11% of total cities in the network. After step 1 diffusion, there were additional 14 cities that became active which means they had participated in the topic. Step 2 witnessed another 2 cities that had adopted the topic, and in total there were already 73% of cities that had been activated. After Step 3, the diffusion continued to step 5 as the setting suggested but none of other cities had been influenced. Thus, step 3,4 and 5 showed 0 newly active node.