P. Wagenseller III, A. Avram, D. Langley
Faculty Advisors: Dr. Feng Wang, Dr. Haiyan Wang
School of Mathematical and Natural Sciences

Spatial-Temporal Flu Prediction Using Community Regions in Commuter Network

Introduction

Despite increased advancements of medical technology and availability of vaccines, emerging and re-emerging epidemics such as SARS, influenza A(H1N1), avian flu, Ebola, and Zika continue to pose tremendous threats to public health. Effective early detection of epidemic outbreaks with specific location information can greatly increase the ability of governmental agencies and health organization to take appropriate actions to control and treat the epidemic. In this project, we propose a novel framework for rapid and accurate epidemic outbreak detection. The key aspects of this project include:

- Introducing a new transdisciplinary framework for detection of epidemic outbreaks.
- Quantifying human mobility reported in a commuter network built from the county-to-county commuting flow.
- Applying community detection algorithms to define new community-based region delineation in the commuter network.
- Proposing new metrics to capture the impact of both physical proximity and human mobility on epidemic diffusion.
- Developing new mathematical theorems of partial differential equations to uncover mechanisms of epidemic diffusion.

CDC Dataset – Each state reports geographic spread of influenza activity as 'no activity', 'sporadic', 'local', 'regional', or 'widespread'. Influenza surveillance data collection are based on a reporting week that starts on Sunday and ends on Saturday of each week. Each surveillance participant is requested to summarize weekly data and submit it to CDC by Tuesday afternoon of the following week. CDC updates it FluView weekly each Friday. CDC defines 51 states, Puerto Rico, Guam, and U.S. Virgin Islands into 10 regions.

Commuter Network & Regions

- **Dataset** – We use the 2006-2010 US Census county-to-county commuting flows data to build two separate networks. This data set describes the commuting patterns of individuals between their residence and workplace. This provides us with a way to measure how counties and states are connected economically.
- **State Level Network** – describes commuting flows of individuals between states. Each node is a state, edges are assigned weights based on the aggregate sum of commuting flows across state lines.
- **County Level Network** – describes the communing flows of individuals between counties. Each node is a (state, county) tuple, edges are assigned weights corresponding to the flow represented in the data set.

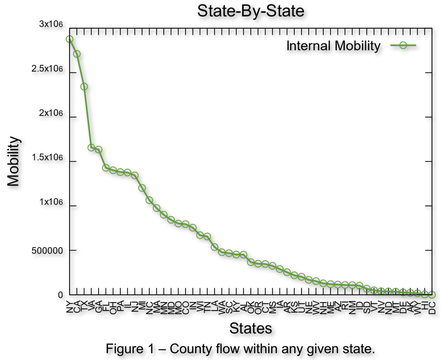

Figure 1 – County flow within any given state.

As can be seen in figure 1, commuting flows differ from state to state. As expected, states with the highest population densities have the highest commuting flows such as NY, CA, and TX. The state with the lowest commuting flow is DC. The reason for this is due to its small land mass and the fact that we removed self loops from within the networks.

CDC Regions & Community Regions

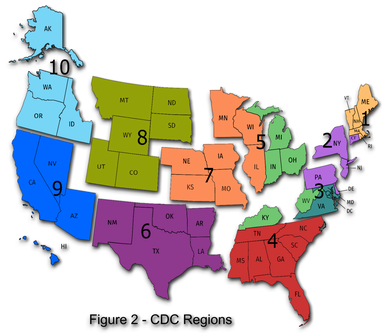

Figure 2 - CDC Regions

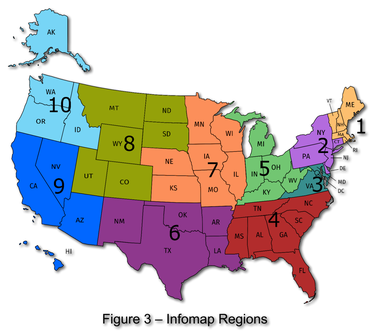

Figure 3 – Infomap Regions

Figure 2 illustrates 10 regions defined by the CDC. Figure 3 illustrate 10 regions generated by running the Infomap community detection algorithm on the state level commuter network to find cohesively connected states. Community regions 6, 8, 9, and 10 are the same as the corresponding CDC regions.

Metrics For Region Commuting Flows

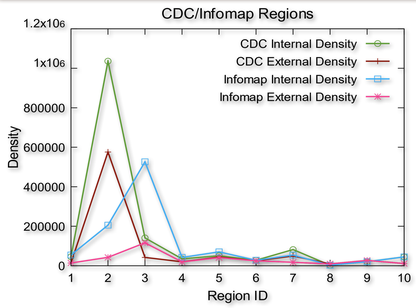

Figure 4 – Density CDC vs Infomap

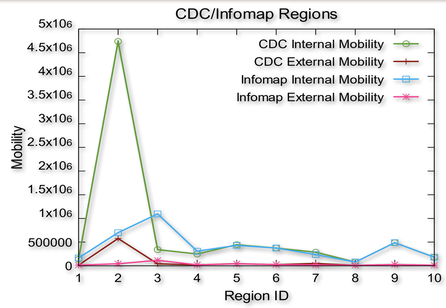

Figure 5 – Mobility CDC vs Infomap

Equation 1 - Region Density

$$regionInternalDensity = \frac{\sum_{u,v \in R, v \neq u} w(u,v)}{N(N-1)/2}$$

$$regionExternalDensity = \frac{\sum_{u \in R, v \notin R} w(u,v)}{N(N-1)/2}$$

In order to measure cohesion amongst states we define the metrics, region internal external density. Then to measure county level mobility inside and across state lines we make two separate metrics called region internal and external mobility. The key difference between these two metrics is that density is calculated on the *state level network* and mobility is calculated on the *county level network.* We calculate each of these using equation 1.

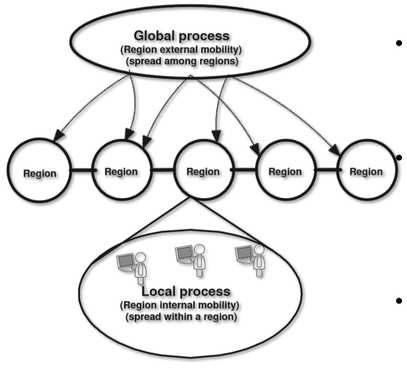

Figure 6 – Flu spreading process.

- **Flu Spreading Process** – the flu can spread as a local process or a global process.
- **Local Process** – the spread of flu within a local community, in our case a region. We capture this concept with internal mobility.
- **Global process** – the spread of flu across other regions. We capture this with external mobility.

PDE Mathematical Model

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial (a e^{-b(x)} \frac{\partial u}{\partial x})}{\partial x} + r(t) u \left[h(x) - \frac{u}{K(x)} \right], t \geq 1, x \in (l, L) \\ u(1, x) = \phi(x), x \in (l, L) \\ \frac{\partial u}{\partial x} = 0, x = l, L, t \geq 0 \end{cases}$$

- Where, $u = u(x,t)$ is the flu case density in region x at time t.
- $\frac{\partial u}{\partial t}$ corresponds to the rate of change of flu case density as time progresses.
- $ae^{-b(x)}, x > 0$ is an exponential decay term for the number of interactions between different regions.
- $\phi(x)$ is the initial flu case density.
- $r(t)$ represents the decay or growth of flu case density with respect to time t.
- $h(x)$ represents the heterogeneity of growth rate in region x.
- $\mathrm{K}(x)$ is the carrying capacity at region x which is a piecewise-defined function whose height at each x is the region internal mobility at region x.
- $\frac{\partial u}{\partial x} = 0, x = l, L, t \geq 0$ is the neumann boundary condition which means no flux of flu across the two ends.

Experimental Results

Equation 2 – MSE and MAE

$$MSE = \frac{1}{2} \sum_{j=1}^{n} (y_j - \hat{y}_j)^2$$

$$MAE = \frac{1}{2} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$

Model	MSE	MAE
CDC	0.2043	0.3403
Infomap	0.1967	0.3311
CDCMobility	0.1856	0.3382

Table 1 – Experimental Results

Given real flu data, from the 45th week of 2016 to the 20th week of 2017, 33 weeks in all, we use three weeks of flu levels as training data to predict flu levels two weeks in the future. This was done as a reflection of the fact that the CDC releases its flu data with a two-week delay. Using MATLAB, we compare the performance of the following models: 1) PDE model using CDC region definition and without the consideration of mobility (called CDC) vs. PDE model using CDC region definition and with the consideration of mobility (called CDCMobility), and 2) PDE model using CDC regions vs PDE model using Infomap region definition and without consideration of mobility (called Infomap). The errors between predicted flu levels and real flu levels are measured with Mean Squared Error and Mean Absolute Error, as defined in Equation 2. As we can see in Table 1, CDCMobility demonstrates lower error rate than CDC, and Infomap demonstrates lower error rate than CDC as well. These results justify our choices of community-based region definition and new PDE model which considers the effect of mobility on flu spreading.

Conclusion

In this project, we proposed new metrics, region internal/external density and region internal/external mobility, to measure the cohesion of a region and its mobility. We then built PDE models which consider the mobility of regions and defined new community-based regions for the model. We incorporate geo-proximity, internal/external density, and internal/external mobility of communities when modeling the flu spreading process. Empirical analysis shows that our proposed PDE modeling architecture can predict epidemic outbreak with low error.