

Visualizing Contagion Dynamics over Massive Networks

Students: Caroline Jareb and Simrat Saini
Mentors: Henning Mortveit, Mandy Wilson, Samarth Swarup, Stefan Hoops, Jiangzhuo Chen

Background

Our goal is to examine dynamics for contagion processes over massive networks (the US population) by visualizing disease transmission data generated from epidemic simulations developed at the Biocomplexity Institute.

EpiHiper and EpiSims37 are two of these simulators. They use complex agent-based models to simulate transmissions over social contact networks throughout the US. In each tick (e.g., day), an individual can change state (e.g., Susceptible, Infected, Recovered) based on the state of the environment.

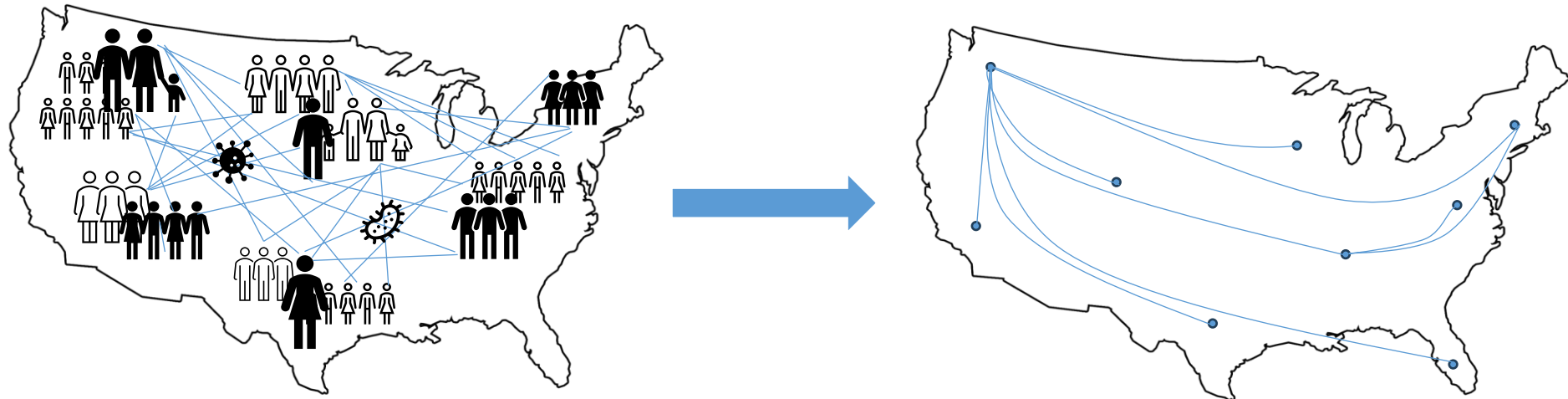


Figure 1: Our goal to simplify visualization of complex transmission networks from simulated epidemics

The Problem: These networks are massive and contain millions of nodes and edges. To effectively examine the dynamics of these networks through visualization, we can change how nodes and edges are represented. We can cluster individuals into fewer nodes and bundle edges into clearer paths.

Methods

Embedding Maps

Members of the population (P) are mapped to locations in an embedding set (M). Individuals can be mapped to the state or county their residence is in, or to the nearest city, urban area, or airport to their residence using only the n largest cities or urban areas by population and n largest airports by passenger traffic.

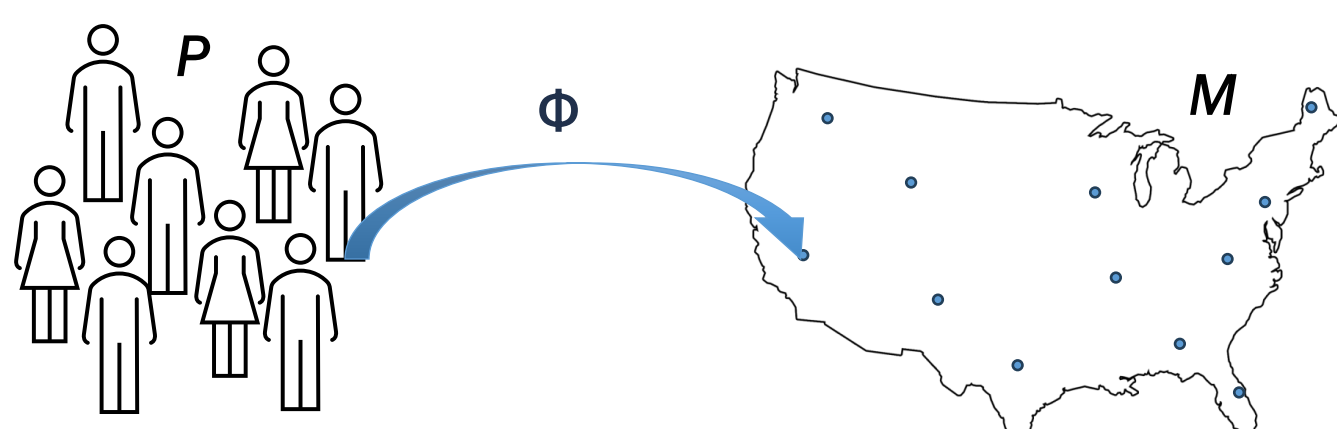


Figure 2: Illustration of Embedding

Edge Bundling

To extract patterns from the massive set of edges, we considered Kernel Density Edge Bundling (KDEB), but we ultimately decided *Force Directed Edge Bundling* (FDEB) seemed more suited for these spatially embedded networks. In FDEB, each edge is subdivided and a force is applied to each interior point (Figure 3). These forces attract other edges based on their compatibility (angle, scale, position, visibility) (Figure 4) and create bundles [1]. In KDEB, edges are bundled by their spatial density, creating white space where edges are less dense [2].

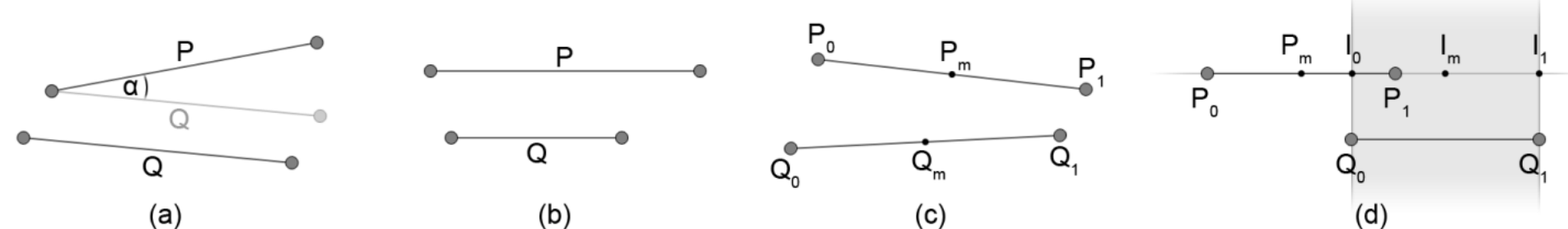


Figure 4 Edge compatibility measures (a) angle compatibility, (b) scale compatibility, (c) position compatibility, and (d) visibility compatibility. Figure from [1].

First Arrivals

Another aspect of contagion dynamics are *first arrival* times, or the first time a disease appears in a region. The first instance of an infection in each region is found in the data set and then plotted onto a map (we chose to focus on state-level dynamics). Additionally, a first arrival standard deviation plot highlighting the variance of arrival times across several replicates of the same initial seed set can also be generated.

Rolling Averages

Rolling averages offer insight into contagion dynamics that are more representative than day-to-day counts. The average of the current day plus the previous six days are computed for each day in each region and plotted onto a map (once again, we chose to focus on state-level dynamics).

n -day Rolling Average Formula for Day x in Region y : $\bar{x}_{y,n} = \frac{\sum_{x-n+1}^x \text{Case Count}_y}{n}$

References

- [1] D. Holten and J. J. Van Wijk, "Force-Directed Edge Bundling for Graph Visualization," Computer Graphics Forum, vol. 28, no. 3, pp. 983–990, 2009, doi: 10.1111/j.1467-8659.2009.01450.x.
[2] C. Hurter, O. Ersoy, and A. C. Telea, "Graph bundling by Kernel Density Estimation," in Computer Graphics Forum, Jun. 2012, p. pp 865. doi: 10.1111/j.1467-8659.2012.03079.x.

Results

We developed a preprocessing pipeline to create embedding maps and format the simulation output data to support different visualization types.

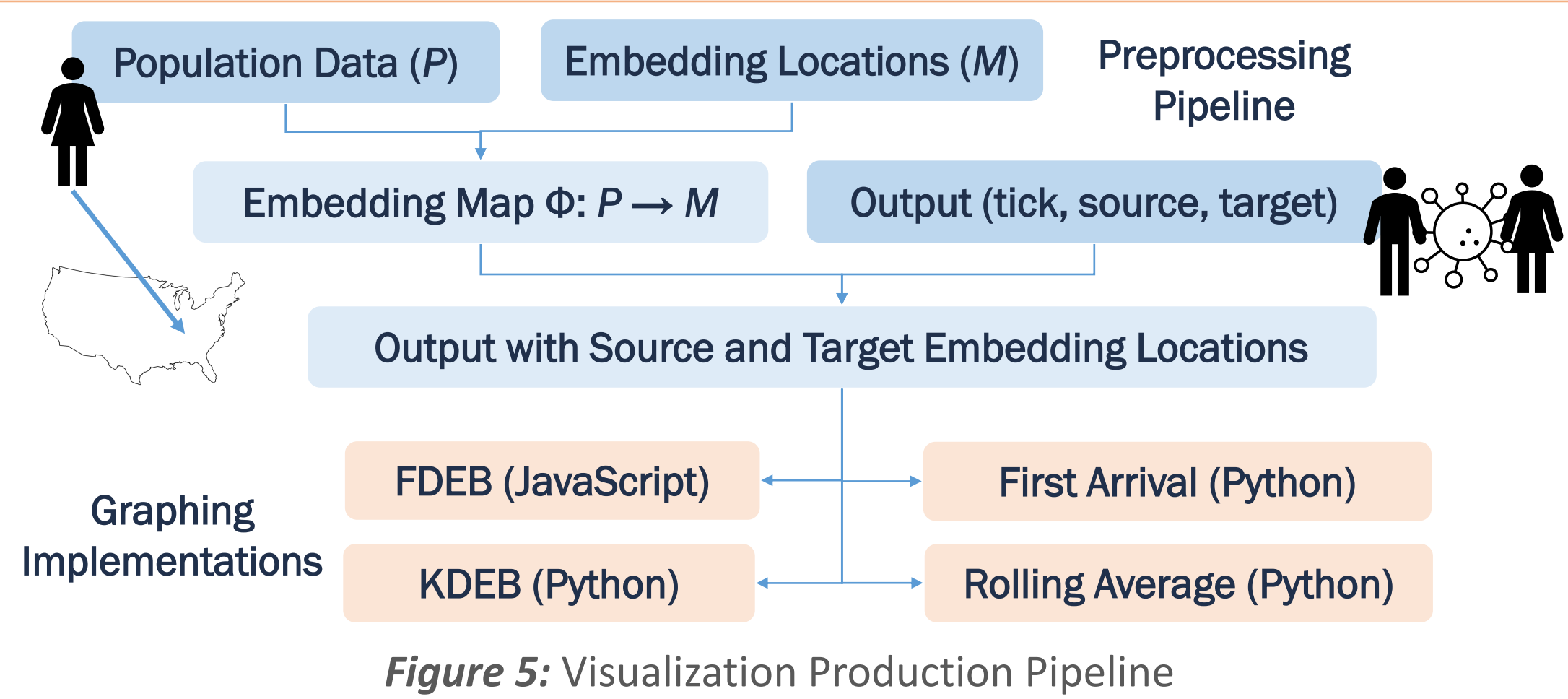


Figure 5: Visualization Production Pipeline

Sample Visualizations

EpiHiper COVID-19 Simulation originating at LAX airport

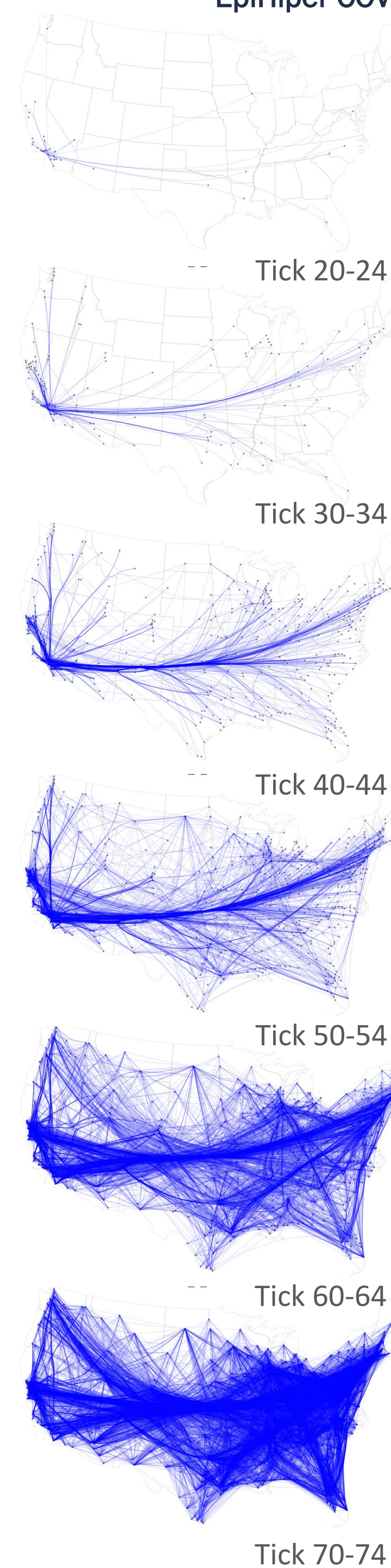


Figure 6: Force Directed Edge Bundling within tick windows

Figure 7: Kernel Density Edge Bundling Tick 40-44

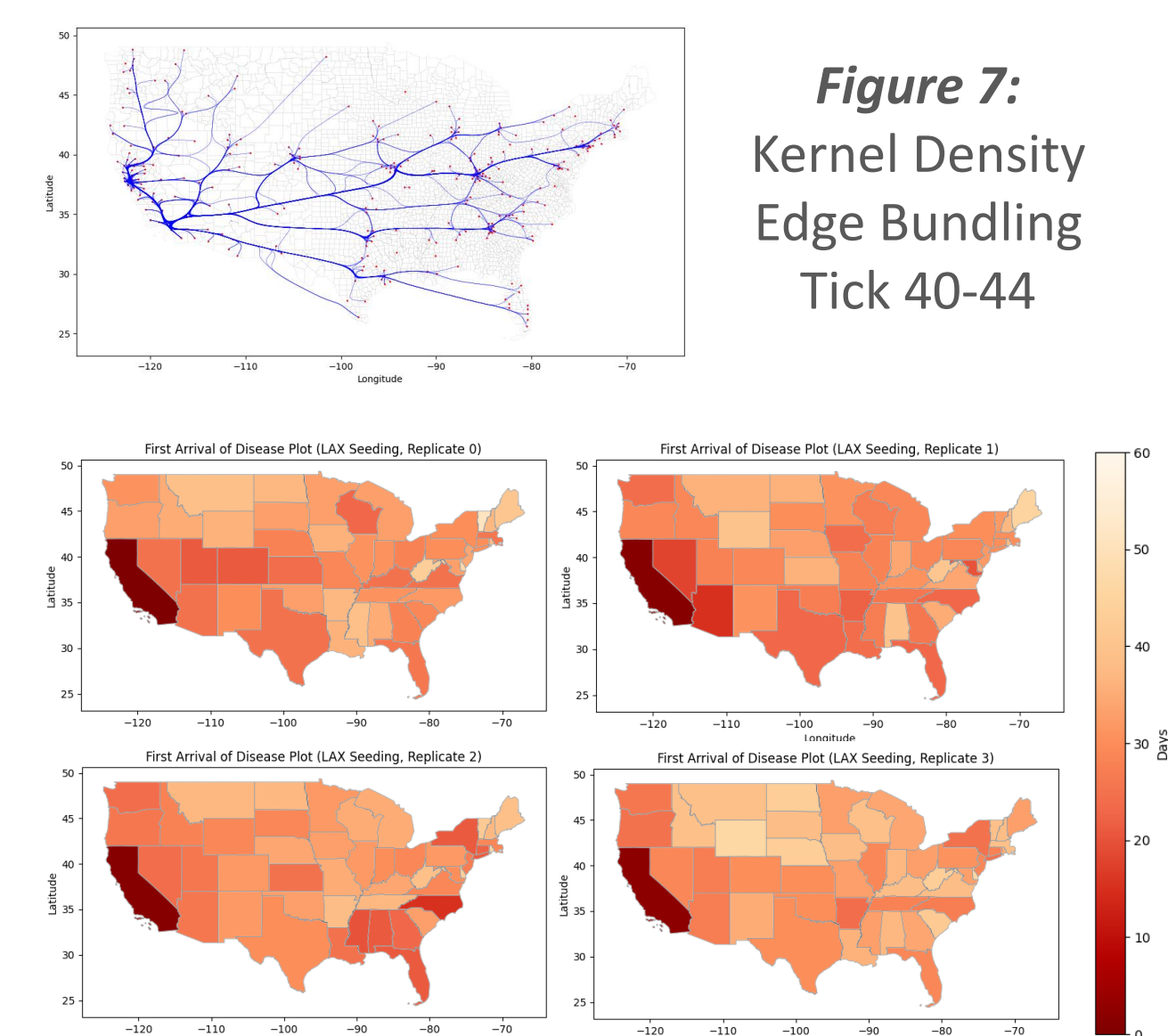


Figure 8: First Arrival Plots for 4 Replicates

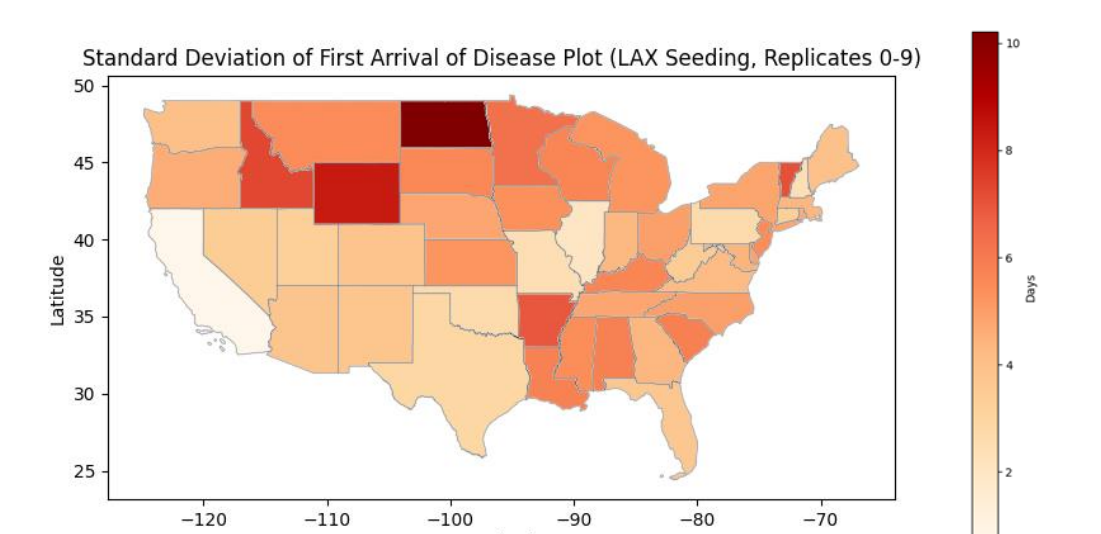


Figure 9: Standard Deviation First Arrival Plot

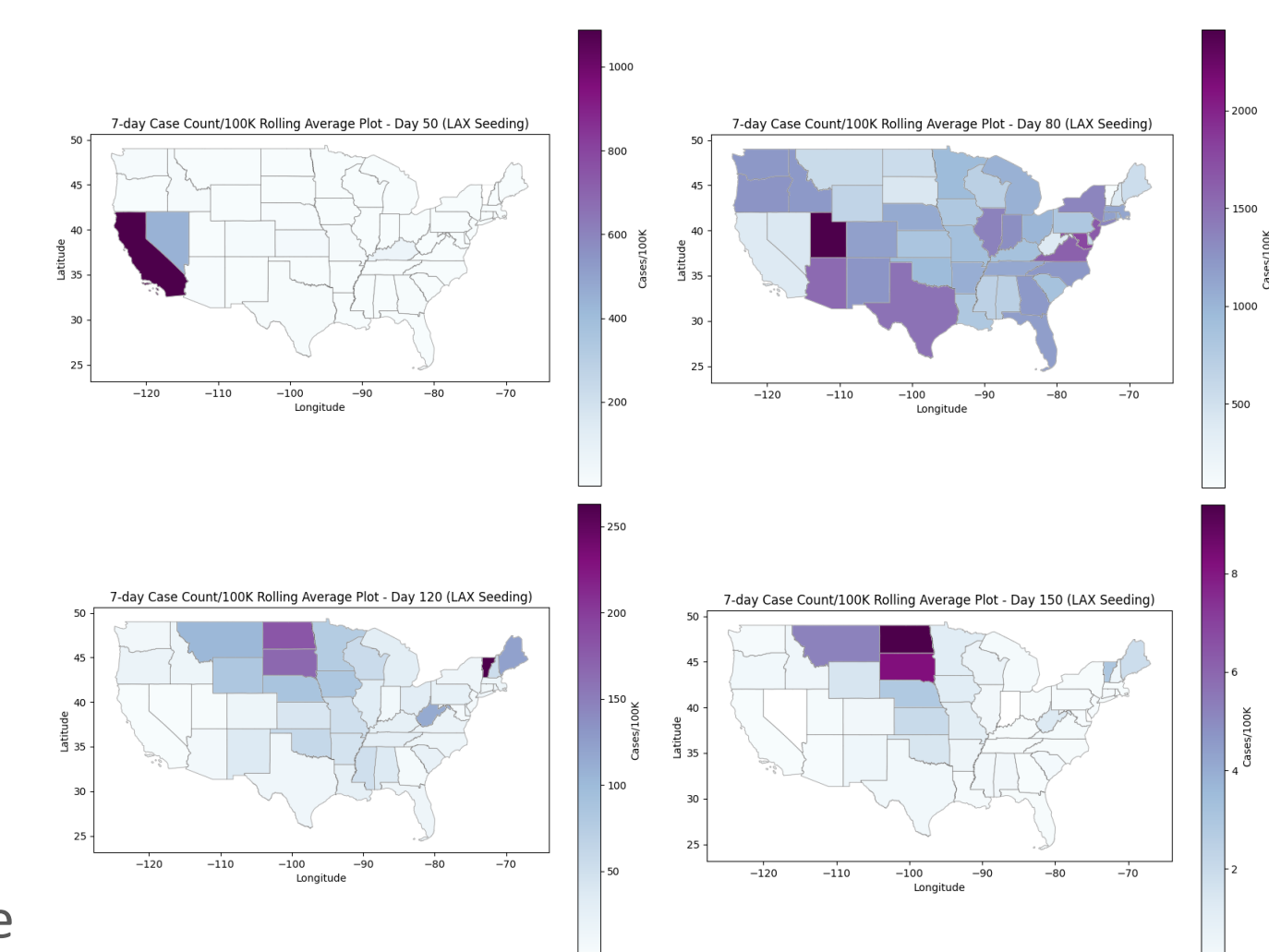


Figure 10: Rolling Average Plots

GitHub: https://github.com/NSSAC/epidemics_sensitivity_2024

Discussion and Future Work

Finding an embedding with the fewest nodes that are properly distributed across the population of the United States proved to be difficult. Clustering cities/urban areas further could potentially improve this. Our current FDEB technique does not include edge weights (transmission counts) and directionality. A GPU-based implementation may include these and be more efficient. Interactive capabilities could be added to be able to examine dynamics over time or in smaller regions. Our visualizations should be validated against raw simulation outputs but proves challenging with large networks. This pipeline could potentially be integrated into the EXCEEDS application to create these visualizations as simulations are run.