CUSP-GX 8153 Complex Urban Systems: Second midterm exam. The exam is open notes and open textbook. No communication is allowed with other students: this should be original work. The exam should be typewritten to receive full credit.

1. (**Schelling model**, 40 points) Consider the Shelling model in your textbook for a population of two agent types (1 and 2), with the following setup: i) solid walls (no agent can move through walls); ii) eight-neighbor stencil of interactions (each agent interacts with any agent in the immediate vicinity, including those on the diagonals); iii) asynchronous protocol with each agent being able to move on the eight-neighbor stencil; and iv) simple majority rule for motion (an agent will move if there are more neighbors different than it as compared to those of the same type – excluding the agent itself. For example, if the agent is of type 1 and it is close to 6 neighbors, 3 of type 1 and 3 of type 2, then it will not move; instead, if 2 neighbors are of type 1 and 4 are of type 2, the agent will move). Please write a computer code to simulate the model for a $5 \times 5$ grid and address the following questions.

   - Simulate the model 1,000 steps for 100 initial conditions, such that the probability that a cell is occupied by agent type 1 is 0.25, occupied by agent type 2 is 0.25, and empty 0.5.
   - Compute the Moran's $I$ for each step and for each initial condition, using the "Queen" distance to construct spatial weight, that is, two agents that are within neighboring cells (even diagonally) have a spatial weight of 1, and 0 otherwise.
   - For one realization, plot the system configuration for three consecutive step (pick them such that they are different) and report the Moran's $I$ for each. Please comment on the plots.
   - Plot the mean and the 95% confidence interval as a function of time, using your 100 realizations. Please comment on the plots.

2. (**Gibrat's law**, 30 points) Utilizing the provided document dataset containing population data for cities (metropolitan and micropolitan statistical areas) and states, spanning from 2010 to 2023, please address the following questions.

   - Calculate the growth rates for cities and states for each of the provided year.
   - Create two plots depicting the relationship between growth rate and the logarithm of population size, one for cities and one for states (include all years in the same plot, that is, each year needs to be displaced with its own point). Utilize distinct colors and markers to differentiate data points by state.
   - Perform a linear regression on the plotted data and compute statistics and confidence intervals. Your dependent variable is the growth rate for each year of each city/state, and the independent variable is the corresponding

population size (logarithmically transformed). Note that you should be working with 10,296 sample points for cities and 637 for states.

3. (**Epidemic spreading**, 30 points) Access the COVID-19 pandemic data from the repository at https://github.com/CSSEGISandData/COVID-19 and retrieve the data for New York County (Manhattan), Kings County (Brooklyn), and Los Angeles County, for all available days. Note that the data represent the cumulative daily new confirmed cases. Please consider an SIR (susceptible-infected-recovered) model of the form

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = -\beta S(t)I(t), \tag{1}$$

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = \beta S(t)I(t) - \gamma I(t), \tag{2}$$

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = \gamma I(t), \tag{3}$$

$$S(t) + I(t) + R(t) = 1, \tag{4}$$

where the parameters $\beta$ and $\gamma$ are infection rate and recovery rate, to be estimated from the data. Please write a code to simulate the model as a function of these two parameters $\beta$ and $\gamma$, as well as the initial condition $I_0$, representing the initial fraction of infected individuals. Assume the populations of Manhattan, Brooklyn, and Los Angeles to be 1.6M, 2.6M, and 9.7M, respectively, and hypothesize that $R_0 = 0$, so that the initial fraction of recovered individuals is zero.

For each of the three counties (Manhattan, Brooklyn, and Los Angeles), please address the following questions.

- Find $\beta$, $\gamma$, and $I_0$ that minimize the mean squared error of the model with respect to the cumulative confirmed cases from January 22nd, 2020 to March 9th, 2023. Note that you will need to compute the number of new cumulative confirmed case by introducing a new equation to the system $\frac{dT(t)}{dt} = \beta S(t)I(t)$ , based on the premise that transitioning out of the susceptible category means becoming infected. Report the minimum mean squared error, along with the optimal values of $\beta$, $\gamma$, and $I_0$.
- Plot the time evolution of $S(t), I(t)$, and $R(t)$ for the optimal model. Also, depict the time evolution of the predicted values and the actual values of the cumulative new confirmed cases.
- What is the basic reproduction number $\mathcal{R}_0 = \frac{\beta}{\gamma}$ according to your findings? Please comment on your findings.