CUSP-GX 8153 Complex Urban Systems: HW2

Qingyang Zhao, qz2567

March 19, 2024

Note: The code of Q2, Q8 and Q9 are hard to show completely. So there is a ipynb file which named 'Code' including solutions about the three questions above. Please read that 'Code' file.

1 Q1

Consider a small dataset of four points $D =$ *(0,2),(1,0),(2,2),(6,4)*. Compute the equation of the linear regression, the total explained variance, the total variance, and the coefficient of determination. Please detail your steps– do not use a built-in package, rather write down the algebra.

Answer:

set the formula is $y = ax + b$. And we know these four points. So, we can get $x_m ean = \frac{0+1+2+6}{4} = 2.25$ and $y_m ean = \frac{2+0+2+4}{4} = 2$. according to this, we can get the

$$a = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{40}{83} \approx 0.482$$

,

$$b = \bar{y} - a\bar{x} \approx 0.916$$

.

Finally, after computing the a and b, we can get the linear is

$$y = 0.482 \times x + 0.916$$

.

the total explained variance: $TEV = \sum(\hat{y_i} - \bar{y})^2 = (0.916 - 2)^2 + (1.398 - 2)^2 + (1.88 - 2)^2 + (3.808 - 2)^2 \approx 4.82$, the total variance: $TV = \sum(y_i - \bar{y})^2 = (2-2)^2 + (0-2)^2 + (2-2)^2 + (4-2)^2 = 8$

We know that the coefficient of determination:

$$R^2 = \frac{TEV}{TV} = \frac{\sum(\hat{y_i} - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 0.6025$$

2 Q2

From https://ghsl.jrc.ec.europa.eu/ghs_stat_ucdb2015mt_r2019a.php, download the associated csv file and perform a scaling analysis between the total population in 2015 and the urban center extension for the following nations: USA, Canada, Poland, and Russia. In your analysis, please exclude any city with less than 100 km^2 in area. Does the analysis match your expectations? Please explain your findings.

Answer:

According to the requires, after processing, we can get a dataframe of 216 rows $\times$ 4 columns finally. Then I do a scaling analysis between the total population in 2015 and the urban center extension, see the plot on fig1. I find the plot is not very linear and hard to research. So I do another plot on Log scale, see fig2. The x label is population and the y label is area in the two plots. After analyzing, I think it match my expectations.If we evaluate the proportion of suitable urban construction areas in these four countries, it is obvious that the United States is ahead of the other three countries and has a higher tendency for urban area expansion. Therefore, under the same urban population size, the urban areas of the United States are higher than those of the other three countries. The higher slopes of the fitting lines at the scatter points representing the United States in the figure also illustrate this.

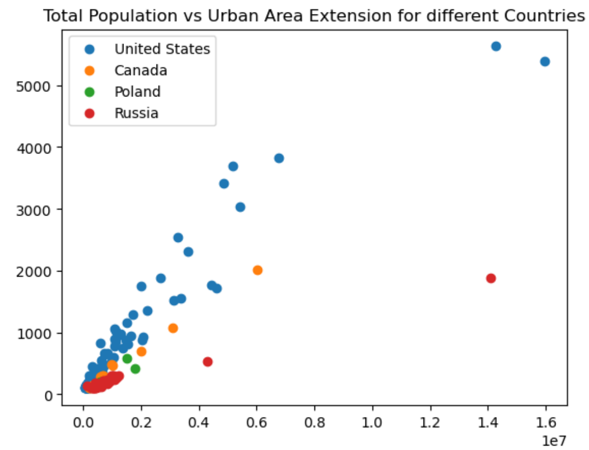

Figure 1: total population vs urban area expansion

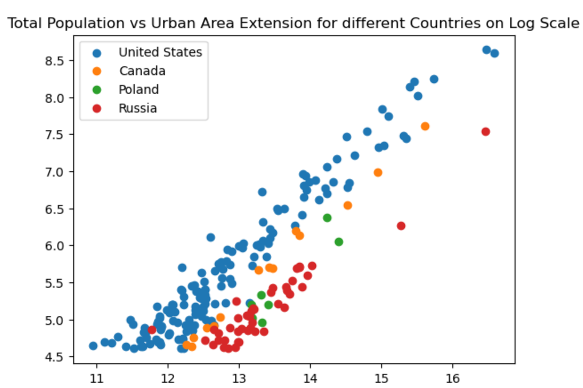

Figure 2: total population vs urban area expansion on log scale

3 Q3

Consider the amorphous settlement model in Part 1 of Chapter 3. Assuming that the distance traveled by a person per day is on average 2km and that the physical distance used for defining an interaction is about 1m, what is the expected total number of interactions per capita in Brooklyn (pull data from Wikipedia on population and area)? Does the number make sense to you based on your daily experience? If not, why?

Answer:

According to the Wikipedia, the total area of Brooklyn is about $250km^2$ and the land area is about $183.4km^2$. The total population is 2,736,074, the density of population is 14,917/km^2.

We know that the function is $k = \frac{a_0 l}{A}(N-1)$ on the part of amorphous settlement model. According to the assume and data in the question, I think the interaction area of single person is about $1 \times 2000 = 2000m^2$, so the total number of interaction per capita in BK is about 30 according to $0.002 \times 14917 \approx 30$. It means one person will interact with 30 other people everyday in BK.

Actually, I don't think it make sense to me. Because I usually meet and interact with friends and some acquaintances. There may be many interactions with same groups of people and so do they. I don't think it make a lot of sense because we may interact with each other a long time and especially we talk while walking. Therefore, I think I don't have real interactions with different people near 30 times everyday.

4 Q4

Consider again the amorphous settlement model in Part 1 of Chapter 3, but this time assume that the cost of transport has some economy of scaling so that more people living in a city will help abate the cost of transportation. Specifically, assume that c_{T0} is not a constant, rather it is equal to $(b/N)^a$ where a and b are two positive numbers. What are the implications for the relationship between A and N? Can it become superlinear? if so, for which values of a and b? Please detail your steps.

Answer:

$Y = G\frac{N}{A} = c_{T0}R$. Due to $R \sim A^{\frac{1}{2}}$, $G\frac{N}{A} = c_{T0}A^{\frac{1}{2}}$. So

$$[A(N)]^{\frac{3}{2}} = \frac{G}{c_{T0}}N$$

. Just like formula (3.12) which is written in the book. After getting that

$$A(N) = [\frac{G}{c_{T0}}N]^{\frac{2}{3}}$$

. According to the assume in the question, $c_{T0} = (b/N)^a$. So

$$A(N) = [\frac{G}{(b/N)^a}N]^{\frac{2}{3}}$$

.

$$A(N) = (\frac{G}{b^a})^{\frac{2}{3}}[N^{a+1}]^{\frac{2}{3}}$$

.

It's easy to see, the function $A(N) \sim N^{a+1}$. And because a is positive, this function become superlinear, and it is for a.

5 Q5

Consider the network depicted in Fig. 3, compute the adjacency matrix, degree of all nodes (and mean degree), between centrality of all nodes, and clustering coefficients of all nodes (and mean clustering coefficient).

Answer:

According to the Fig.1, we can get:

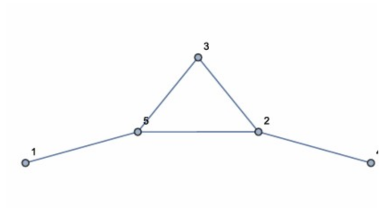

Figure 3: Small network for the computation of salient metrics.

1.adjacency matrix:

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

2.degrees:

$$d_1 = 1, d_2 = 3, d_3 = 2, d_4 = 1, d_5 = 3$$

,

$$k = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \\ 3 \end{bmatrix}$$

$$degree_mean = 2$$

3.between centrality: The shortest paths are $1-5, 1-5-2, 1-5-3, 1-5-2-4, 5-3, 5-2, 5-2-4, 3-2, 3-2-4, 2-4$. Shortest paths passing through node 1 have 0, so the BC is 0; Shortest paths passing through node 2 have 3 and they are $4-2-5, 3-2-4, 1-5-2-4$, total number of shortest paths without ending in node 2 is 6, so the BC is $\frac{1}{2}$; Shortest paths passing through node 3 have 0, so the BC is 0; Shortest paths passing through node 4 have 0, so the BC is 0; Shortest paths passing through node 5 have 3 and they are $1-5-2, 1-5-3, 1-5-2-4$, total number of shortest paths without ending in node 5 is 6, so the BC is $\frac{1}{2}$.

4.clustering coefficients:

$$z = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$c_i = \frac{2z_i}{k_i(k_i - 1)}$$

$$c = \begin{bmatrix} 0 \\ \frac{1}{3} \\ 1 \\ 0 \\ \frac{1}{3} \end{bmatrix}$$

6 Q6

Please read carefully https://royalsocietypublishing.org/doi/full/10.1098/rsif.2013.0789 and answer the following questions. How are the networks of human interactions defined? Why should the average degree increase with the population size of a city? What is the difference between reciprocal and nonreciprocal networks? Why is the scaling exponent for the degree of the reciprocal network lower? What are the main limitations of the study?

Answer:

1.How are the networks of human interactions defined?

Networks of human interactions refer to the structures that represent how individuals in a society communicate and interact with each other. These networks can be constructed based on various forms of communication, such as phone calls which is referred in the paper, emails, social media interactions, face-to-face interactions, etc.

2.Why should the average degree increase with the population size of a city?

There is a superlinear scaling of social connectivity with city size according to the paper and it has research supporting about Portugal's mobile phone data. In my opinions, larger populations provide more opportunities for individuals to interact with a greater number of people. And denser populations lead to increased socioeconomic quantities, social connectivity and networking, as people are more likely to encounter each other frequently in public spaces, workplaces, social events, etc.

3.What is the difference between reciprocal and nonreciprocal networks?

The difference between them is the mutuality of the networks. In a reciprocal network, if node A is connected to node B, then node B is also connected to node A. it means that interactions between nodes are mutual. However, in a nonreciprocal network, connections between nodes may not be mutual.

4.Why is the scaling exponent for the degree of the reciprocal network lower?

In reciprocal networks, the scaling exponent is typically lower compared to nonreciprocal networks. Because reciprocal networks tend to exhibit more localized interactions and clustering. In other words, individuals in reciprocal networks are more likely to form connections within their immediate social circles or communities, resulting in a lower average degree compared to nonreciprocal networks where connections may be more dispersed.

5.What are the main limitations of the study?

The main limitation of this study is that it is not generalizable. The study focused on communication networks in only two European countries. This may limit the generalizability of the findings to other regions and cultures around the world. Furthermore, this study may have overlooked differences in communication patterns and network structures across different socioeconomic backgrounds, technological levels, geographical locations, and time periods. To address this limitation, future research should aim to integrate data from different populations and regions, and continue to refine the research direction and research objects (such as people's interactions with their relatives on social media platforms in country A and country B).) to gain a more complete understanding of the relationship between city size and human interaction.

7 Q7

Consider the urban scaling model discussed in class, which extends the amorphous settlement model. Change the second hypothesis regarding the network area, assuming that individuals are not adjusting their distance with the city area and the population. Rather, assume that urban dwellers will want to always be a fixed distance d_0 apart: what are the consequences on the network area, how will that scale with population?

Answer:

Network areas will no longer scale proportionally to population size or city area. Instead, it will remain relatively constant, determined by d_0. Furthermore, since the network area no longer depends on population size, its scaling behavior with population will change. Instead of exhibiting superlinear scaling, it may exhibit linear scaling.

8 Q8

Consider the classical (compartmental) SIR model discussed in class for the following parameters $\beta = 0.2$, $\gamma = 0.1$, and N = 10,000. Simulate the model and plot the three variables from t = 0 to 200 for the case in which there is only one infected subject at the beginning of the infection and nobody is immune to the disease. Compute the reproductive number and the effective reproductive number. Would you expect the disease to be epidemic? If so, does this match the numerical results? What do you think is a disease with a similar behavior? Please compute analytically the peak number of

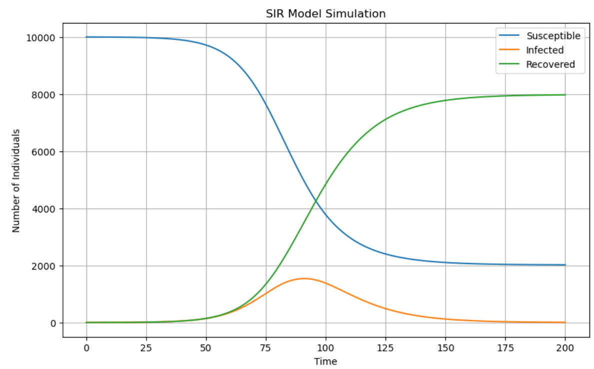

Figure 4: SIR model about Q8

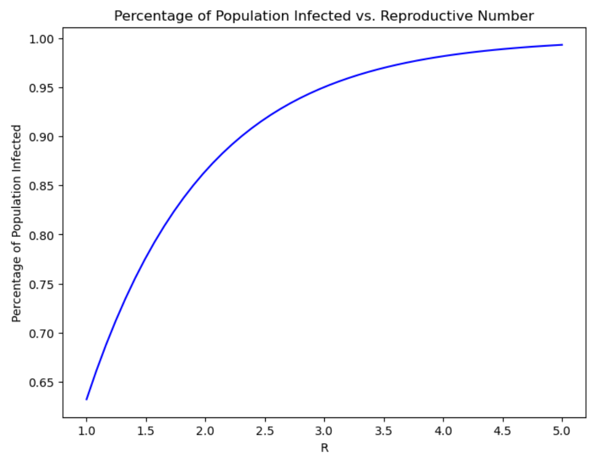

Figure 5: Percentage of Population Infected vs. Reproductive Number

infections and compare with simulations. Also, please compute the steady-state number of susceptible individuals by solving the associated implicit equation and compare with numerical results.

Answer:

I use the SIR model to compute and some of them on the code file. Present the results and plot Fig.4.

the reproductive number: $\frac{\beta}{\gamma} = 2$ the effective reproductive number: $\frac{S_0\beta}{N\gamma} = 1.9998$

the disease is epidemic because $1.9998 > 1$, the disease increase over time and it has a peak. This match my result. I think the viral cold is with a similar behavior. The peak number of infections is about 1538 and it matches the simulations.

$$\frac{dI}{dt} = 0, S = \frac{\gamma N}{\beta}$$

The steady-state number of susceptible individuals is about 5000.

9 Q9

Consider again an SIR model, compute the fraction of infected individuals at steady-state as a function of the reproductive number based on equation 3.33 in the textbook. You should generate a figure similar to the right panel of figure 3.17.

Answer:

According to the equation 3.33:

$$S_\infty = S_0 e^{-R(1-\frac{S_\infty}{N})}$$

, We can generate the Fig.5.

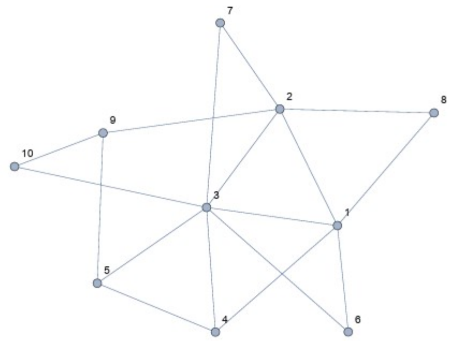

Figure 6: Small network for the computation of searchability.

10 Q10

Consider the network depicted in Fig. 6, compute the searchability of the path 1→2→3→4.

Answer:

For the path 1→2→3→4. $H_x = log_2 k_1 + log_2(k_2 - 1) + log_2(k_3 - 1) = log_2 5 + log_2 4 + log_2 6 = 6.907$. So the searchability of the path is 6.907.