Geographical-based Friendship Network

Assignment 1d

Probability & Statistics SA 2023-2024

Social networks are systems that map the relationships and interactions between individuals. Although social networks are typically dynamic, we focus here on a static example. Geographical distance plays a crucial role in social network formation, with closer physical proximity typically leading to more frequent and stronger social ties. Additionally, geographical features, such as the density of people in an area (more in cities and less in remote mountainous regions), profoundly influence the composition and scale of these networks, shaping the diversity and connectivity of communities within a given region.

In the Assignment, you will receive questions regarding the creation of an inhabitant map and the simulation of friendships among individuals in that map. Below we give a short description of these two steps, which you need to use for answering your questions in Section 2.

Creating the Inhabitant Map
You will create **two kinds of Inhabitant Maps**:

1. *Purely Random Map*:

   First, you randomly place 1000 individuals within a unit square, $[0, 1]^2$. Each individual is represented by coordinates (x, y).

2. *Geographical based Random Map*:

   Secondly, you will create several geographical areas allowing for diverse features like circular desert(s), elliptical mountain range(s), winding river(s), and squared or rectangular cities. Each individual in this case is characterized by coordinates (x, y). The function f maps the coordinates into the corresponding geographical feature. You denote with n_f the number of features you include in your model.

   $$f(x, y) = \text{geographical feature } f \text{ characterizing } (x, y)$$

   1000 individuals are no longer distributed uniformly, but the density of people in an area depends on the specific geographical feature:

   (a) Set a positive score for each of the n_f geographical features in your map, according to the ease of living there.

   $$s_f = \text{ease in living in the geographical feature } f \tag{1}$$

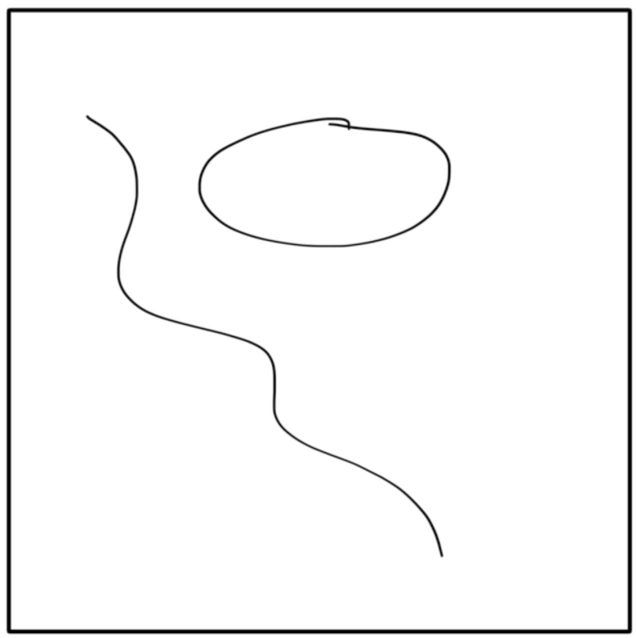
   (b) Compute the probability of an individual settling in location (x, y), as the following ratio:

   $$P[\text{random person settles in } (x, y)] = \frac{s_{f(x,y)}}{\displaystyle\sum_{(x',y')\in \text{Map}} s_{f(x',y')}} \tag{2}$$

Friendship Network
Starting from the two different maps, you will also create **friendships among the people** following different models:

1. *Starting from map 1*:

   You will model friendships via the Erdős–Rényi model.



2. *Starting from the map 2*:

   You will assign different probabilities to different friendships. The probability of a relationship between individual A and individual B is modelled as follows:

$$P(A-B) = \exp\{-\beta * \text{dist}[(x_A, y_A), (x_B, y_B)]\} \tag{3}$$

   where dist() is the Euclidean distance.

Assignment

1. **Scenario 1**:

   (a) **(20 points) Design the pseudocode for the Random Inhabitant Map and Friendship Network**

   i. **(5 points) Sampling individuals.** Write the pseudocode to randomly sample $n = 1000$ points within a unit square, $[0, 1]^2$. They will represent your individuals.

   ii. **(5 points) Design the euclDistMatrix function.** Provide pseudocode to compute the $n \times n$ matrix including the Euclidean distance between n individuals.

   iii. **(5 points) Create the distMatrixFromN function.** Develop the pseudocode that, given a number n of individuals as input, repeats the following operations 100 times:

   - sample n points in $[0, 1]^2$;
   - compute the distance matrix;
   - calculate summary information for the distance matrix (such as minimum, maximum, average, etc.);

   Return, for each of the chosen summary measures, a vector of 100 values.

   iv. **(5 points) Develop the ErdosRenyiFromP function.** Write the pseudocode to create an Erdos-Renyi Graph, given n (n. of individuals in the network) and p (probability of an edge between a couple of nodes).

   (b) **(15 points) Exploring alternative modeling approaches**

   i. **(5 points) Exploring n** For different values of n, particularly $10, 100, 1000, 10000$, explore how minimum, maximum and, mean varies across iterations.

   ii. **(10 points) Exploring p** Deepen the Erdos-Renyi Graphs for different values of p, particularly:

$$0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.9$$

   Explore the results, possibly with graphical representations.
   Some suggestions for exploring are:

   - The degree distribution, the structure of the adjacency matrix, the average path length;
   - The fact that the maximum eigenvalue of the adjacency matrix is included between the average and the maximum degree;
   - The dimension of the largest connected component and the link with the eigendecomposition of the corresponding adjacency matrix.

   (c) **(10 points) Regarding the ErdosRenyiFromP function for generating Friendship Networks**:

   i. **(5 points** Do you think that this model is a realistic model for how friendships arise? If yes, why? If not, why? Explain the main foundations of the model you are using.

   ii. **(5 points** Find some advantages and disadvantages of the model.

2. **Scenario 2**:

   (a) **(30 points) Design the pseudocode for the Geographical-based Random Inhabitant Map**

      i. **(5 points) Write the pseudo code in order to design the geographicalFeatureMap function** that outputs a $n \times n$ grid, where each point is characterized by a geographical feature. Your map should include at least: one (e.g., circular) desert (Desert), one (e.g. ellipsoidal) mountain range (Mountain), one (e.g. sinusoidal) river (River), and at least three squared or (e.g., rectangular) cities (City).

      ii. **(5 points) Develop the pseudocode for the geographicalFeatureProbability function**. It assigns a value of probability for each point in the grid, following Equations 1 and 2.

      iii. **(5 points) Design the pseudocode for the geographicalSample function**. It samples the location of n individuals according to the probabilities computed with geographicalFeatureProbability.

      iv. **(5 points) Write the pseudo code to compute the distance among the individuals** by using function euclDistMatrix.

      v. **(5 points) Create the pseudocode for the function adjMatrixFromDist** that, given a distance matrix and a value β, returns a sampled network of friendships (as an adjacency matrix). To accomplish this, do the following steps for each pair of individuals $(A - B)$:

         - sample a value u from a uniform distribution in $[0, 1]$;
         - compute the probability $P(A - B)$ in Equation 3;
         - consider A and B to be friends if $u < P(A - B)$.

      vi. **(5 points) Write the pseudocode to develop the function DegreeFromAdjMatrix** that, given an adjacency matrix created via adjMatrixFromDist, provides the mean degree of the corresponding network.

   (b) **(15 points) Exploring alternative modeling approaches**

      i. **(5 points) Create the inhabitant map for $n = 1000$.**. Compute the average, mininum, and maximum distance among all the possible individuals on the map. Then, explore the distance within each particular geographical area. Show and compare the results possibly with interesting graphical representations. Compare the results with the one found in Question 1(b)i.

      ii. **(10 points) Explore several values of β**, in particular those that allows to find a mean degree of $5, 10, 15$. You might be interested in:

         - Plotting the values of β versus the mean degree of the graph.
         - Inspecting the degree distribution, the structure of the adjacency matrix, the average path length;
         - The dimension of the largest connected component.

   (c) **(10 points) Regarding the adjMatrixFromDist function for generating Friendship Networks**:

      i. **(5 points)** What is the relationship between the geographical features and the groupings of friends in the network? Do you think this model is more or less realistic for how friendships arise with respect to the one seen before? Why?

      ii. **(5 points)** Explain some limitations of this model.

3. **(Bonus: 10 points) Multimedia presentation:** Design an animation that vividly displays all the simulation features in a movie of about 3 minutes.

Assignment Guidelines

Submission Formats

You can submit your assignment in one of the following formats:

1. **PDF + Code:** A PDF of your written answers, analysis, and results, accompanied by a separate code file (.R or .py).
2. **RMarkdown Report:** If you are working with R, you can submit an RMarkdown file (.Rmd) along with a rendered PDF or HTML. Ensure that the RMarkdown file can be knitted without errors.
3. **Jupyter Notebook:** If you are working in a Python environment, you can submit a Jupyter notebook (.ipynb) along with a rendered PDF version.

Pseudocode Guidelines

Pseudocode is a way to plan your program without the constraints of programming syntax. Here are some guidelines for writing effective pseudocode:

- **Language Agnostic:** Your pseudocode should not be specific to any programming language.
- **Structured Approach:** Use a step-by-step approach, breaking down complex tasks into simpler sub-tasks.
- **Use English:** Write in plain English, avoiding technical jargon.
- **No Syntax:** Do not worry about exact syntax, but maintain a consistent style.
- **Indentation:** Use indentation to show loops, conditionals, and other control structures.
- **Annotations:** You can add brief comments to explain the logic or any assumptions.

Best Practices

- **Documentation:** Comment your code adequately. Your markers should be able to understand your logic just by reading the comments.
- **Function Decomposition:** Break your code into smaller, manageable functions or chunks. Each function should ideally perform one task.
- **Cite Sources:** If you consult external sources or use someone else's code snippets (with modification or directly), make sure to cite them.

Academic Integrity

Remember to maintain the highest standards of academic honesty. While discussions are encouraged, the final work should be your own.