Why there is three parts in this assignment: Each part fulfills one of the objectives of the class:

- **Manipulate concepts:** Getting Familiar with the technical concepts used in class, by reproducing similar arguments. Being proficient by manipulating the object to answer some small-size problem.

  You are expected to answer this question rigorously, the answer can be quite short as long as it contains all the required argument to justify your answer.

- **Experience the concepts in real case:** Being able to reproduce these concepts on real or synthetic data. Study their properties in real examples.

- **Connect the concepts to real-life:** Interpret a problem you find in light of the notions you have learned. Develop some critical eye to determine how the concepts introduced are useful in practice.

How to read this assignment : Exercise levels are indicated as follows

($\rightarrowtail$) "elementary": the answer is not strictly speaking obvious, but it fits in a single sentence, and it is an immediate application of results covered in the lectures.

  Use them as a <u>checkpoint</u>: it is strongly advised to go back to your notes if the answer to one of these questions does not come to you in a few minutes.

($\frown$) "intermediary": The answer to this question is not an immediate translation of results covered in class, it can be deduced from them with a reasonable effort.

  Use them as a <u>practice</u>: how far are you from the answer? Do you still feel uncomfortable with some of the notions? which part could you complete quickly?

($\rightarrowtail$) "tortuous": this question either requires an advanced notion, a proof that is long or inventive, or it is still open.

  Use them as an <u>inspiration</u>: can you answer any of them? does it bring you to another problem that you can answer or study further? It is recommended to work on this question only AFTER you are done with the rest!

Part A — Manipulating the Concepts

Exercise 1: Analysis of small world property in dimension $k \geq 1$ (10 pt) A good way to master the proof on augmented lattice is to manipulate it for different graphs. This exercise is a complement to the detailed proof of the case $k = 1$, found at the end of this document and in the scribing note, it will allow you to do that step by step. You may if you wish even start the exercise immediately after reading the proof in the three separate case $r < 1$, $r > 1$ and $r = 1$.

The first step is to obtain a few important properties of lattice. An infinite lattice with dimension $k \geq 1$ is the set of elements of $\mathbb{Z}^k$.

We recall the classical result that all norms are equivalent in a finite dimensional vector space. Hence, for any two norms $\|.\|_1$ and $\|.\|_2$, there always exist two constants $\alpha > 0$ and $\beta > 0$ such that

$$\forall x \in \mathbb{Z}^k, \alpha\|x\|_1 \leq \|x\|_2 \leq \beta\|x\|_1 .$$

Examples of norms defined on $\mathbb{Z}^k$ are:

- The $\mathbb{L}^1$ norm defined as: $\|x\|_{\mathbb{L}^1} = \sum_{i=1}^{k} |x_i|$.

- The $\mathbb{L}^2$ norm defined as: $\|x\|_{\mathbb{L}^2} = \sqrt{\sum_{i=1}^{k} x_i^2}$.

- More generally, the $\mathbb{L}^j$ norm for $j \geq 1$ defined as: $\|x\|_{\mathbb{L}^j} = \left(\sum_{i=1}^{k} |x_i|^j \right)^{\frac{1}{j}}$.

- The maximum norm, also called referred to as the $\mathbb{L}^\infty$ norm, which is defined as: $\|x\|_{\mathbb{L}^\infty} = \max_{i=1,2,\ldots,k} |x_i|$.

1. ($\curvearrowright$) Show that in an infinite lattice of dimension $k \geq 1$ the following statement holds:

   - $\forall u, \exists\, \alpha > 0$ and $\beta > 0$ such that: $\alpha j^{k-1} \leq \#\{v \in V \,|\, \|u - v\| = j\} \leq \beta j^{k-1}$.
   - $\forall u, \exists\, \alpha > 0$ and $\beta > 0$ such that: $\alpha j^{k} \leq \#\{v \in V \,|\, \|u - v\| \leq j\} \leq \beta j^{k}$.

2. ($\curvearrowright$) Deduce that, for a finite lattice of dimension k (with length $L - 1$, containing $N = L^k$ nodes) there exist $\alpha > 0$, $\beta > 0$ independent of N such that:

$$\alpha \sum_{j=1}^{\lfloor L/2 \rfloor} \frac{1}{j^{r-(k-1)}} \leq \sum_{v \neq u} \frac{1}{\|u - v\|^r} \leq \beta \sum_{j=1}^{L} \frac{1}{j^{r-(k-1)}} \ .$$

3. ($\rightarrowtail$) From this inequality can you briefly justify why the value $r = k$ is critical for dimension k?

4. ($\curvearrowright$) Assuming $r < k$ show that, wherever u and v are located on the lattice the probability that u is connected to v by a shortcut becomes polynomially small as N grows. In other words, there exists $\delta > 0$ and a constant $c_1 > 0$ such that: $\mathbb{P}\left[u \rightsquigarrow v\right] \leq \frac{c_1}{N^\delta}$.

5. ($\rightarrowtail$) Let us denote by I_l the set of nodes at distance at most l from the target:

$$I_l = \{\ u \in V \ |\ \|u - t\| \leq l\ \}\ ,$$

   Which one of the following is an upper bound on the probability that at least one the n first shortcuts met by the walk drawn using greedy routing connects to a node within I_l?

   $(i)\ \dfrac{2c_1 n l}{N^\delta}$ $\qquad\qquad\qquad (ii)\ \dfrac{2n l^k c_1}{N^\delta}$ $\qquad\qquad\qquad (iii)$ none of the above

6. ($\rightarrowtail$) Conclude that greedy routing needs in expectation at least a constant multiplied by $N^{\frac{k-r}{k(k+1)}}$ steps to succeed.

7. ($\curvearrowright$) We now assume $r > k$. Prove that the probability that u shortcuts has length greater than m is less than $\frac{c_3}{m^{r-k}}$. Conclude that greedy routing needs in expectation at least a constant times N^η steps for $\eta > 0$.

8. ($\rightarrowtail$) Assuming $r = k$ what can you deduce on the normalizing constant? Prove that the probability for a node in phase j to be connected to a node in phase $j' < j$ does not depend on j and becomes small slowly with N. Conclude.

Exercise 2: Extension of the small world result to an infinite lattice (5 pt) One of the limitation of the above proof is to deal frequently with normalizing constant and finite networks. In this exercise we prove that for at least two cases of the one studied above, a formulation using an infinite lattice can be drawn.

In an infinite lattice, one cannot hope to have any bound on the time to connect two arbitrarily far away nodes. On the other hand, one may hope that on an infinite lattice that starting from a node at a fixed distance D from the target, greedy routing finds a path whose length grows slowly with D.

1. ($\curvearrowright$) Assuming $r > k$, can you prove that the random biased augmented lattice (which we also called Kleinberg's model) extends naturally without modification to an infinite lattice. Why is that impossible when $r \leq k$?

2. ($\curvearrowright$) Assuming $r > k$, can you quickly justify why there exists a constant $C > 0$ and $\eta > 0$ such that, in expectation, the path found by greedy routing requires at least CD^η steps when starting from a node at distance D from the target.

We will now prove that Kleinberg model can be modified so that the proof of the critical case $r = k$ applies to an infinite lattice.

3. ($\rightarrowtail$) For an $\varepsilon > 0$ Let $f : x \mapsto \frac{1}{\ln^\varepsilon(x)}$, what is f' the derivative of f ? what is the limit $\lim_\infty f$?

4. ($\curvearrowright$) Prove that for any $\varepsilon > 0$, one can naturally extend Kleinberg model to an infinite lattice for $r = k$, assuming that the probability to have a shortcut $u \rightsquigarrow v$ is

$$\mathbb{P}\left[u \rightsquigarrow v\right] = \frac{1}{\|u - v\|^k \ln^{1+\varepsilon}(\|u - v\|)} \ .$$

5. ($\curvearrowright$) Prove that greedy routing in the above model uses at most $O(\ln^{2+\varepsilon}(D))$ steps when starting at distance D from the target.

Part B — Experiencing the Concepts

Exercise 3: Data Challenge 1 (15pt) Read and complete the first data challenge (see supplementary document). The due date for that part of the homework will be announced later.

Part C — Concepts at large

Exercise 4: Exercise 1: In a future, not so far away ... (5 pt)   Let's imagine, for the sake of a thought experiment, a planet in which live the Erdosians. The Erdosians is a strict society that made the choice of canceling all biases in social relationship between its member. Through a pervasive computer interface, occasions of social gathering (attending a course, a movie, a party, etc.) are randomly assigned. As a consequence the set of people that anyone knows at any time is drawn in a uniform random way from the population irrespective of geography and social classes.

To prove the superiority of their social organization, the Erdosians demonstrates to their visitors their "connect" application which allows any member of the society to find its shortest paths to any other. A constant monitoring over a century has shown that this distance remains remarkably small in spite of the population increase. The "connect" application made available for all members have quickly become extremely popular to keep up with friends through a multi-hop referral. One important reason why it is used is that the Erdosians, liberated from any geographical biases, have started to adopt an increasing nomadic life, making them hard to locate. Although they each nominally own a car and apartment, they rarely use it since they typically use a sharing-platform whenever they need a place to stay or a ride.

Unfortunately, a major natural disaster is responsible for a black-out of the "connect" application and its associated data, leaving each Erdosians with only a way to contact their immediate friends, that were initially cached on their cell-phone.

1. Using elements from the course, can you comment on the gravity of the situation? Assuming that it is still possible for an Erdosian to call her friend, and ask for referral, could they possibly exchange information to keep up with each other through a short chains of acquaintance?

(NB: I know you will not believe me, but as I was writing this exercise in La Guardia airport one morning, I was approached two minutes later by a teenager and her mum who asked if she could use my laptop to contact her friend after a delayed flight. When I said that I was not online but that they could use my phone for email, her answer was "Oh thank you, but it's going to be too complicated, I don't have he rnumber, I have to contact her on Facebook!". I could not help but wonder "Are Erdosians already landed? Why choose La Guardia Terminal D for a first contact?")

Exercise 5: Exercise 2: The data-set and the phone company (5 pt)   You are working with a group of students on studying the way people use cell phones to connect to each other, and the small world property of the network among them. A connection with a telephone provider allows you to get access to some anonymous data set that would turn out invaluable to answer empirically some important questions.

Unfortunately, after some concerns of public data release, your contact in the company mention that it would not possible to release the full data sets and that members of her team would simply accept to release the information about people who make call between each other very regularly (e.g., every 3 days). They mention to you that these are the most important ties of the network anyway, and that they should be sufficient for your study. Tomorrow you will meet with responsible of the team to present your project and explain how the data will be used.

1. What arguments could you use to convince them that they should provide a more complete data sets?

(This exercise was inspired by D. Easley and J. Kleinberg)