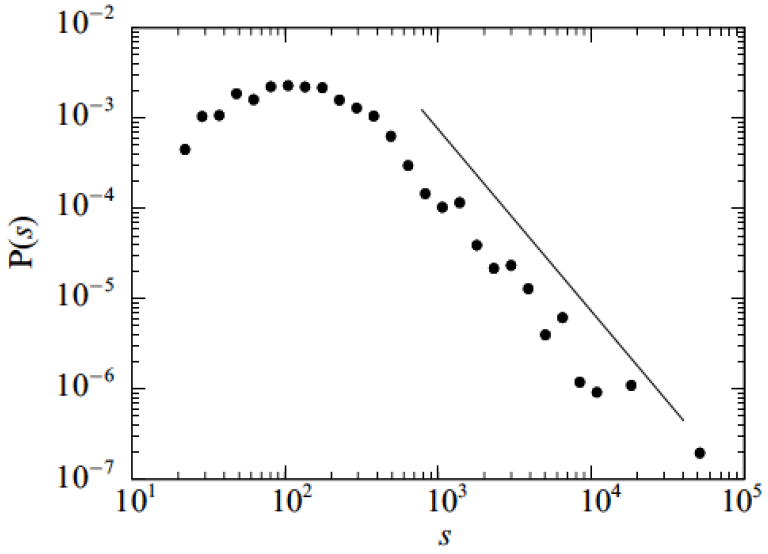
Assignment 2 part 2 (due 10/09/2020)

Question 1: Compare Trip Distributions of Data vs. Model as a Weighted Network (60 pts)

Build two undirected weighted graphs with the data and a trip distribution model of your choice (developed in Assignment 2 part 1). The goal is to analyze and compare their network properties.

1) For both data and model, calculate the degree distributions of degree strengths or weighted degree (similar to Figure 5B below, extracted from the paper: the structure of interurban traffic: a weighted network analysis presented in Lecture 7) (10pts)



2) Compare values of the two networks respect to: maximum degree K_{max}, average degree <k>, standard deviation of the average degree, σ_k, the average clustering coefficient <CC> and average shortest path <L> (30 pts)

Network	K_{min}	K_{max}	<k>	σ_k	<CC>	<L>
US commuting Data						
Model US commuting Data						

3) First, let's calculate the normalization constant C of the power law degree distribution, with exponent $\gamma > 1$ (5 pts)

If:

$$P(k) = Ck^{-\gamma} \qquad k = [K_{\min}, \infty)$$, and

$$\int_{K_{\min}}^{\infty} P(k)dk = 1$$

Complete the integration to show that:

$$C = (\gamma - 1)K_{\min}^{\gamma - 1}$$

4) Second, let's estimate the maximum degree (K_{max}) of a power law network, by solving the integral (5pts)

$$\int_{K_{\max}}^{\infty} P(k)dk \approx \frac{1}{N}$$

Show that:

$$K_{\max} = K_{\min} N^{\frac{1}{\gamma - 1}}$$

5) Plot a power law function for the degree distribution:
$P(k) = C\ k^{-\gamma}$ fixing K_{min} like in the commuting US network, and selecting $\gamma > 1$ of your choice that K_{max} is close to the data (10pts)

Question 2 Question III: Exercise for Project Preparation (20 pts)

We will prepare the project in 4 parts developed in the assignments. In this assignment we start with part I, the first step is to make your selection on area and data. Keeping in mind that for the purposes of the class you need to adapt your expectations and interests to data available.

Part I: Search and Selection:
In urban planning and policy, a 'wicked problem' is a problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize.

While data science informed policy does not bring the ultimate solution, however it can help. Identify an area of your interest to do a data science research project for this class. Examples are vulnerability of transportation infrastructures, housing crisis, traffic prediction, epidemic spreading, emergency responses, adoption of sustainable technologies, among others.

Motivate your research project answering the three questions below :

- What is the problem?

- Why should we care?
- What do you want to do applying spatial analysis? (identify your data sources, ask in piazza if you do not know where to search for it)

Part II: Data Science Story (now or guidelines later)
Select a paper related with the selected dataset keep. It does not have to be exactly the same data. It is only important that you read and refer here something published about an analysis of the similar kind of data you chose. After you select a data set of your interest, I can help you to find an appropriate article for you to write about in your project written report.

Part II: Data Analysis (guidelines later)

Part III: Visualization (guidelines later)

Submit the html and the ipynb files with the solution(no zipped)