Homework 5

Gabriel CAO

1 SUBJECT OF THE HOMEWORK

A data set of intercontacts in a conference is provided. Each line of the file follow the format : *id1 id2 start_time end_time* where *id1* and *id2* are the identifiers of the person that started a conversation and *start_time* and *end_time* are respectively the time when the two person started the conversation and the time when the two people ended the conversation. Analyse the data set and describe the methodology used to obtain the result for the following questions :

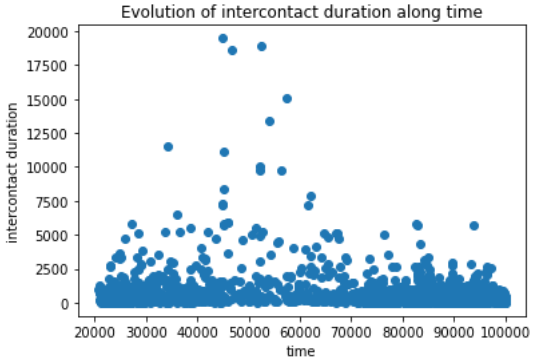
- Question 1 How are intercontacts durations arranges along time ?
- Question 2 What is the average intercontact duration, what about the variance ?
- Question 3 What is the intercontact duration distribution ?
- Question 4 Does the global intercontact duration distribution differ from the individual distribution?
- Question 5 Is there a time dependency for intercontact duration ?
- Question 6 What else can you deduce from the data set ?

2 DATA ANALYSIS

2.1 Question 1

We would like to view the evolution of intercontacts durations along time. To do this, we execute the following algorithm :

- Extracting *start_time* and *end_time*
- Compute the intercontact duration with the two extracted data
- For each intercontact duration, create the point with the coordinate x equal to *start_time* and y equal to intercontact duration.



Except for a few cases, the intercontact durations are the same. Some intercontacts durations are high and are located in the middle of the conference.

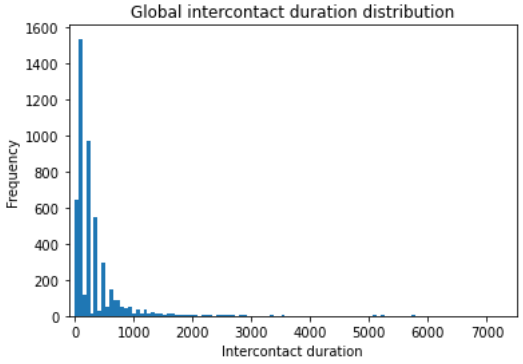
2.2 Question 2

To compute the average intercontact duration, we retrieve the intercontact durations that we extracted, we sum it all and we divide the sum by the number of intercontact duration. The average intercontact duration is 400.

To compute the variance, for each intercontact duration, we remove the average from the intercontact duration and we square the result then we sum it all and we divide it by the number of intercontact duration. The result is 751310.32.

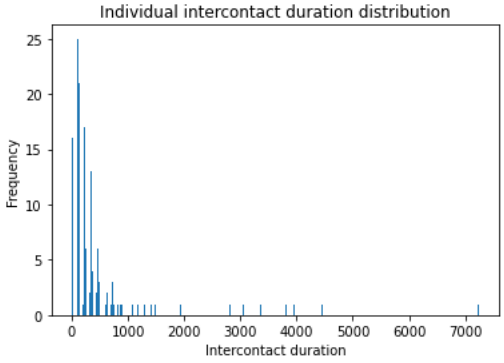
2.3 Question 3

To build the intercontact duration distribution, we extract all intercontact duration and we build an histogram with an histogram bin equal to 300.



2.4 Question 4

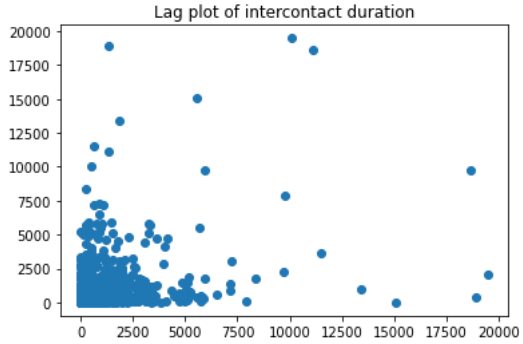
To compare the global intercontact duration distribution with an individual intercontact duration, we build an histogram for a random individual in the conference and by putting the two histograms on the same graph, we can see if there are any differences.



On the graph, the two distributions are following the same pattern, the global duration distribution doesn't differ from the individual duration distribution.

2.5 Question 5

To see if there is any time dependency for intercontact duration, we build a lag plot, we should look for any form in the plot that show the evolution of the intercontact duration a time T along the intercontact duration at time T - 1.



We can't see a pattern in the graph, we conclude that there is no time dependency.

2.6 Question 6

Nothing else was deduce from the data set