

**Design/Practical Experience**  
**Department of Computer Science and Engineering**  
**Final Report**

**Academic Year :** 2022-23

**Semester:** I

**Date of Submission of Report:** 17/11/2022

**Name of Students:** Gaurav Kumar (B19CSE032) & Deep Patel (B19CSE028)

**Title of the Project :** Constructing a network graph of covid19 contact tracing.

**Project Category:** Summer/Winter/Semester projects with Institute faculty within or outside the department

**Name of Mentors:** Dr. Dweepobotee Brahma

**Targeted Deliverables:**

- Data cleaning
- Making networks of covid contacts on the basis of age and gender.

**Theory:**

Due to the pace and scope of its spread, the WHO had described the coronavirus disease 2019 (COVID-19) as a pandemic. Without efficient interventions, the burden of therapeutic treatments would significantly rise due to the rapidly rising number of COVID-19 cases. In order to stop transmission and allocate the limited public health resources, it is essential to identify the sources and pathways of transmission. So in order to do that we have constructed a network graph which depicts from which source a person is infected and to whom that infected person is spreading that virus. The dataset which we worked upon contains the details such as the Patients' address, age, gender, date on which covid was detected, date on which patient was released and the source from which they were infected. We can see the sample of the dataset as shown in the below figure.

PatientID	PatientID_N	District	Gender	Age	Date_detected	Date_discharge	Date_death	SourceofCovid
P-1		1 Bengaluru	M	46	10-Mar-20	27-Mar-20		traveled to US
P-2		2 Bengaluru	F	47	10-Mar-20	24-Mar-20		P-1
P-3		3 Bengaluru	F	13	10-Mar-20	27-Mar-20		P-1
P-4		4 Bengaluru	M	50	10-Mar-20	24-Mar-20		traveled to US
P-5		5 Greece	M	26	12-Mar-20	20-Mar-20		
P-6		6 Kalaburgi	M	76	29-Feb-20	09-Mar-20	10-Mar-20	traveled to Saudi Ara
P-7		7 Kalaburgi	F	45	16-Mar-20	01-Apr-20		P-6
P-8		8 Bengaluru	M	32	16-Mar-20	01-Apr-20		P-6
P-9		9 Kalaburgi	M	63	17-Mar-20	02-Apr-20		P-6
P-10		10 Bengaluru	F	20	17-Mar-20	02-Apr-20		traveled to UK, return

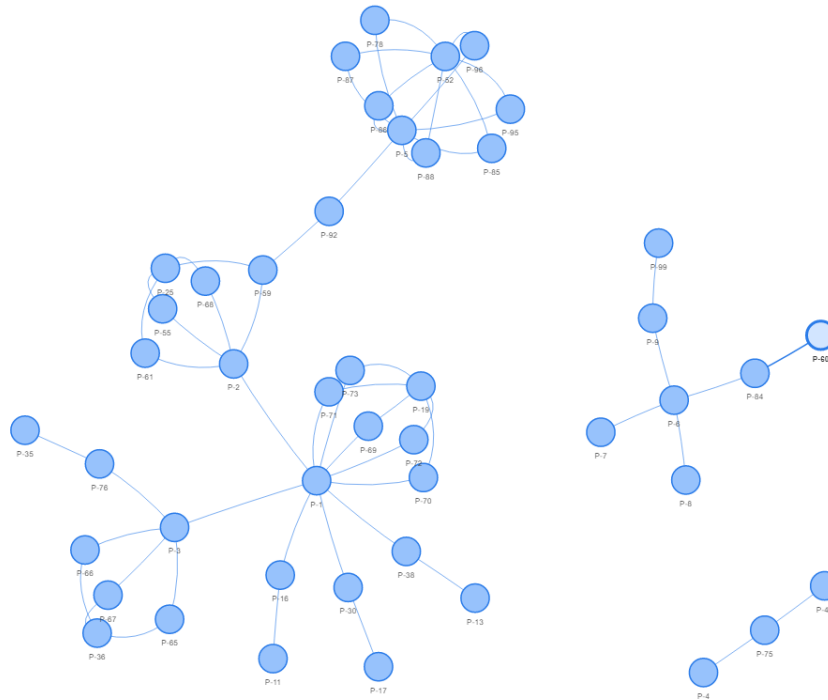
### Work Done:

Our first requirement was to do data cleaning of the given covid contract trace dataset since the data was not well structured . For this we used python programming language and **pandas** library. First of all we assigned patient\_id to those patients whose patient\_id was not mentioned. The next task was to do data cleaning on the source of covid of the infected persons. We were mainly concerned with the covid infections that came because of the contact with some other infected persons.

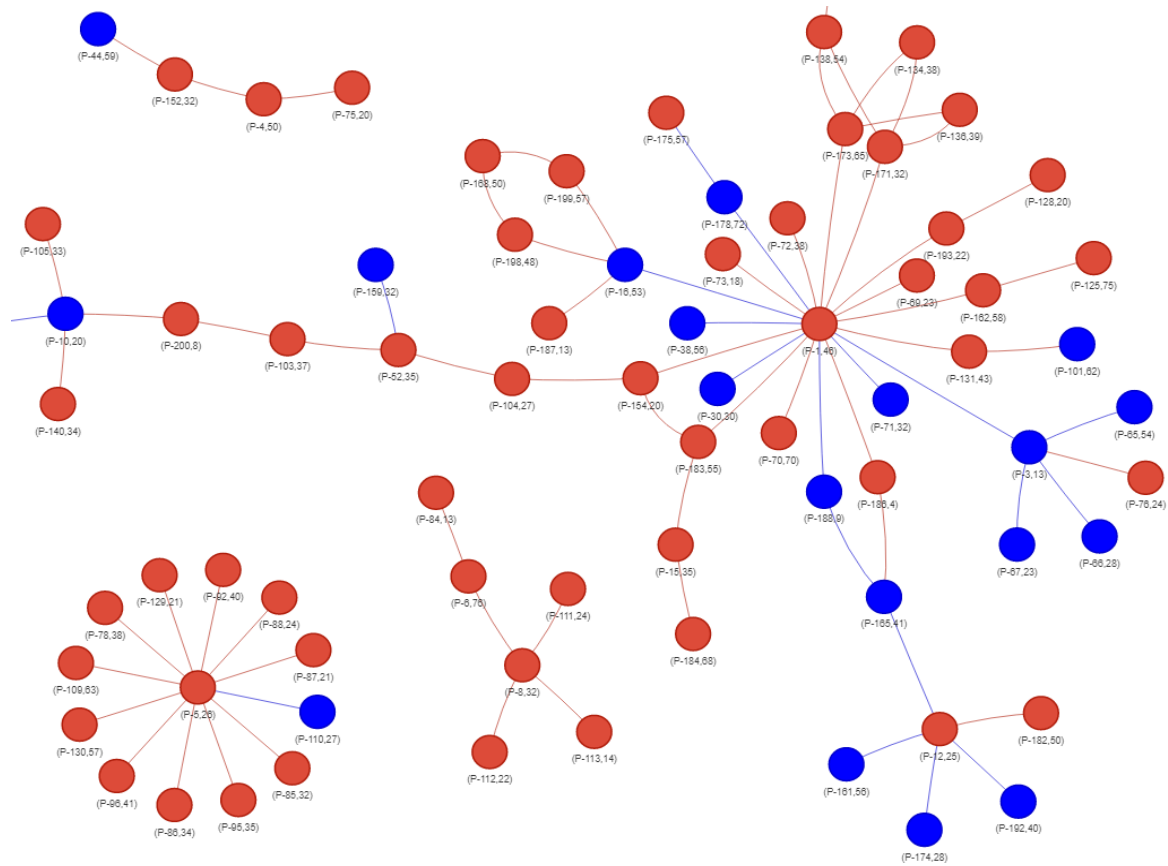
For this we have cleaned the dataset and removed those infections which didn't come from another person. In the given dataset the source of covid was given in text and from these text we extracted the id of the patient if it is there. If the source of covid doesn't contain another patient then we ignore such infections . Data after cleaning will have patient\_id and their contacts along with age as we are mainly concerned with the source of covid.

PatId    Contacts	
0	(P-2,47)    (P-1,46)
1	(P-3,13)    (P-1,46)
2	(P-7,45)    (P-6,76)
3	(P-8,32)    (P-6,76)
4	(P-9,63)    (P-6,76)

We have used the Python library **pyvis** to make networks. To start with we firstly plotted a basic network of covid infections on a small set of data. The nodes represent the patient\_id and if two patient\_id have an edge between them it means they came in contact and got infected.



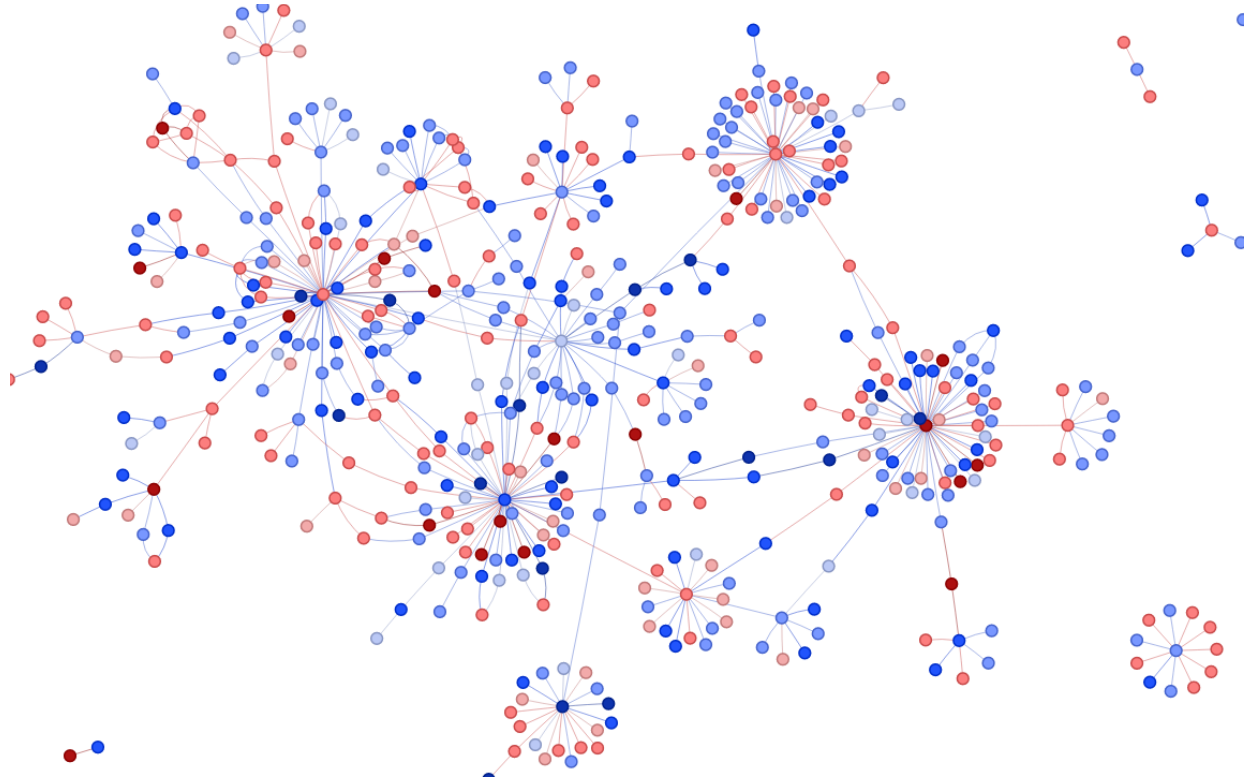
After doing basic plotting we moved to differentiate nodes on the basis of gender. We have used blue color for females and red colors for males. At the same time while plotting these we also made changes in nodes names and included age while making nodes. Blue colored nodes represent females and red colored are for males. The network now looks like:



To make graphs more differentiable, we introduced different shades of color so that we can make more distinguished nodes on the basis of age. We have used four shades of each color and divided age groups into four categories. The age groups are : Below 18, Between 18 to 30, 31 to 60 and above 60.

**Future Work:** We can further analyze the covid contacts by including another feature ‘date of infections’. It will help us to analyze how fast the covid is increasing over a period of time. We can do it on a monthly basis or weekly basis depending on the time period we are looking for. It will help to predict parameters like doubling rates and we can also make machine learning models to predict expected numbers of covid infections at a given time period.

## Results-



## Links:

Notebook link of implementation :

[https://drive.google.com/file/d/1BDhQSodJ7pcGzS2Pbo73UBG7jxdHKZFn/view?usp=share\\_link](https://drive.google.com/file/d/1BDhQSodJ7pcGzS2Pbo73UBG7jxdHKZFn/view?usp=share_link)

Link for covid\_contact dataset csv file:

[https://drive.google.com/file/d/1mPjRFy-S8FKGtv9NAu7o9GsdmreHmZ3w/view?usp=share\\_link](https://drive.google.com/file/d/1mPjRFy-S8FKGtv9NAu7o9GsdmreHmZ3w/view?usp=share_link)