

```
In [25]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
In [32]: !pip install plotly==4.7.1
!wget https://github.com/plotly/orca/releases/download/v1.2.1/orca-1.2.1-x86_64.AppImage -O /usr/local/bin/orca
!chmod +x /usr/local/bin/orca
!apt-get install xvfb libgtk2.0-0 libgconf-2-4
```

```
Collecting plotly==4.7.1
  Downloading https://files.pythonhosted.org/packages/d7/78/eb6cbe96c8379c54819592bb228c58ed7386fcc60a55eca7db99432fdf14/plotly-4.7.1-py2.py3-none-any.whl (11.5MB)
    |████████████████████████████████████████| 11.5MB 6.0MB/s
Requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.7/dist-packages (from plotly==4.7.1) (1.3.3)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from plotly==4.7.1) (1.15.0)
Installing collected packages: plotly
  Found existing installation: plotly 4.4.1
  Uninstalling plotly-4.4.1:
    Successfully uninstalled plotly-4.4.1
Successfully installed plotly-4.7.1
```

```
--2021-04-12 16:08:34-- https://github.com/plotly/orca/releases/download/v1.2.1/orca-1.2.1-x86_64.AppImage
Resolving github.com (github.com)... 140.82.113.4
Connecting to github.com (github.com)|140.82.113.4|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://github-releases.githubusercontent.com/99037241/9dc3a580-286a-11e9-8a21-4312b7c8a512?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWNJYAX4CSVEH53A%2F20210412%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20210412T160834Z&X-Amz-Expires=300&X-Amz-Signature=1d09771015a60288dac06b88a3eb0e7cfdd25ccef4398c30da40af8e2778826e&X-Amz-SignedHeaders=host&actor_id=0&key_id=0&repo_id=99037241&response-content-disposition=attachment%3B%20filename%3Dorca-1.2.1-x86_64.AppImage&response-content-type=application%2Foctet-stream [following]
--2021-04-12 16:08:34-- https://github-releases.githubusercontent.com/99037241/9dc3a580-286a-11e9-8a21-4312b7c8a512?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWNJYAX4CSVEH53A%2F20210412%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20210412T160834Z&X-Amz-Expires=300&X-Amz-Signature=1d09771015a60288dac06b88a3eb0e7cfdd25ccef4398c30da40af8e2778826e&X-Amz-SignedHeaders=host&actor_id=0&key_id=0&repo_id=99037241&response-content-disposition=attachment%3B%20filename%3Dorca-1.2.1-x86_64.AppImage&response-content-type=application%2Foctet-stream
Resolving github-releases.githubusercontent.com (github-releases.githubusercontent.com)... 185.199.110.154, 185.199.108.154, 185.199.111.154, ...
Connecting to github-releases.githubusercontent.com (github-releases.githubusercontent.com)|185.199.110.154|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 51607939 (49M) [application/octet-stream]
Saving to: '/usr/local/bin/orca'
```

```
/usr/local/bin/orca 100%[=====] 49.22M 91.3MB/s in 0.5s
```

```
2021-04-12 16:08:35 (91.3 MB/s) - '/usr/local/bin/orca' saved [51607939/51607939]
```

```
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  gconf-service gconf-service-backend gconf2-common libdbus-glib-1-2
  libgail-common libgail18 libgtk2.0-bin libgtk2.0-common
Suggested packages:
  gvfs
The following NEW packages will be installed:
  gconf-service gconf-service-backend gconf2-common libdbus-glib-1-2
  libgail-common libgail18 libgconf-2-4 libgtk2.0-0 libgtk2.0-bin
  libgtk2.0-common xvfb
0 upgraded, 11 newly installed, 0 to remove and 31 not upgraded.
Need to get 3,715 kB of archives.
After this operation, 17.2 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic/main amd64 libdbus-glib-1-2 amd64 0.110-2 [58.3 kB]
Get:2 http://archive.ubuntu.com/ubuntu bionic/universe amd64 gconf2-common all 3.2.6-4ubuntu1 [700 kB]
Get:3 http://archive.ubuntu.com/ubuntu bionic/universe amd64 libgconf-2-4 amd64 3.2.6-4ubuntu1 [84.8 kB]
Get:4 http://archive.ubuntu.com/ubuntu bionic/universe amd64 gconf-service-backend amd64 3.2.6-4ubuntu1 [58.1 kB]
Get:5 http://archive.ubuntu.com/ubuntu bionic/universe amd64 gconf-service amd64 3.2.6-4ubuntu1 [2,036 B]
Get:6 http://archive.ubuntu.com/ubuntu bionic/main amd64 libgtk2.0-common all 2.24.32-1ubuntu1 [125 kB]
Get:7 http://archive.ubuntu.com/ubuntu bionic/main amd64 libgtk2.0-0 amd64 2.24.32-1ubuntu1 [1,769 kB]
Get:8 http://archive.ubuntu.com/ubuntu bionic/main amd64 libgail18 amd64 2.24.32-1ubuntu1 [14.2 kB]
Get:9 http://archive.ubuntu.com/ubuntu bionic/main amd64 libgail-common amd64 2.24.32-1ubuntu1 [112 kB]
Get:10 http://archive.ubuntu.com/ubuntu bionic/main amd64 libgtk2.0-bin amd64 2.24.32-1ubuntu1 [7,536 B]
Get:11 http://archive.ubuntu.com/ubuntu bionic-updates/universe amd64 xvfb amd64 2:1.19.6-1ubuntu4.8 [784 kB]
Fetched 3,715 kB in 1s (3,038 kB/s)
Selecting previously unselected package libdbus-glib-1-2:amd64.
(Reading database ... 160983 files and directories currently installed.)
Preparing to unpack .../00-libdbus-glib-1-2_0.110-2_amd64.deb ...
Unpacking libdbus-glib-1-2:amd64 (0.110-2) ...
Selecting previously unselected package gconf2-common.
Preparing to unpack .../01-gconf2-common_3.2.6-4ubuntu1_all.deb ...
Unpacking gconf2-common (3.2.6-4ubuntu1) ...
Selecting previously unselected package libgconf-2-4:amd64.
Preparing to unpack .../02-libgconf-2-4_3.2.6-4ubuntu1_amd64.deb ...
Unpacking libgconf-2-4:amd64 (3.2.6-4ubuntu1) ...
Selecting previously unselected package gconf-service-backend.
Preparing to unpack .../03-gconf-service-backend_3.2.6-4ubuntu1_amd64.deb ...
Unpacking gconf-service-backend (3.2.6-4ubuntu1) ...
Selecting previously unselected package gconf-service.
Preparing to unpack .../04-gconf-service_3.2.6-4ubuntu1_amd64.deb ...
Unpacking gconf-service (3.2.6-4ubuntu1) ...
Selecting previously unselected package libgtk2.0-common.
Preparing to unpack .../05-libgtk2.0-common_2.24.32-1ubuntu1_all.deb ...
Unpacking libgtk2.0-common (2.24.32-1ubuntu1) ...
```

```

Selecting previously unselected package libgtk2.0-0:amd64.
Preparing to unpack .../06-libgtk2.0-0_2.24.32-1ubuntu1_amd64.deb ...
Unpacking libgtk2.0-0:amd64 (2.24.32-1ubuntu1) ...
Selecting previously unselected package libgail18:amd64.
Preparing to unpack .../07-libgail18_2.24.32-1ubuntu1_amd64.deb ...
Unpacking libgail18:amd64 (2.24.32-1ubuntu1) ...
Selecting previously unselected package libgail-common:amd64.
Preparing to unpack .../08-libgail-common_2.24.32-1ubuntu1_amd64.deb ...
Unpacking libgail-common:amd64 (2.24.32-1ubuntu1) ...
Selecting previously unselected package libgtk2.0-bin.
Preparing to unpack .../09-libgtk2.0-bin_2.24.32-1ubuntu1_amd64.deb ...
Unpacking libgtk2.0-bin (2.24.32-1ubuntu1) ...
Selecting previously unselected package xvfb.
Preparing to unpack .../10-xvfb_2%3a1.19.6-1ubuntu4.8_amd64.deb ...
Unpacking xvfb (2:1.19.6-1ubuntu4.8) ...
Setting up gconf2-common (3.2.6-4ubuntu1) ...

Creating config file /etc/gconf/2/path with new version
Setting up libgtk2.0-common (2.24.32-1ubuntu1) ...
Setting up libdbus-glib-1-2:amd64 (0.110-2) ...
Setting up xvfb (2:1.19.6-1ubuntu4.8) ...
Setting up libgconf-2-4:amd64 (3.2.6-4ubuntu1) ...
Setting up libgtk2.0-0:amd64 (2.24.32-1ubuntu1) ...
Setting up libgail18:amd64 (2.24.32-1ubuntu1) ...
Setting up libgail-common:amd64 (2.24.32-1ubuntu1) ...
Setting up libgtk2.0-bin (2.24.32-1ubuntu1) ...
Setting up gconf-service-backend (3.2.6-4ubuntu1) ...
Setting up gconf-service (3.2.6-4ubuntu1) ...
Processing triggers for libc-bin (2.27-3ubuntu1.2) ...
/sbin/ldconfig.real: /usr/local/lib/python3.7/dist-packages/ideep4py/lib/libmkldnn.so.0 is not a symbolic link

Processing triggers for man-db (2.8.3-2ubuntu0.1) ...

```

```

In [26]: !pip install fsspec
import dask.dataframe
import tqdm

```

```

Requirement already satisfied: fsspec in /usr/local/lib/python3.7/dist-packages (0.9.0)
Requirement already satisfied: importlib-metadata; python_version < "3.8" in /usr/local/lib/python3.7/dist-packages (from fsspec) (3.8.1)
Requirement already satisfied: typing-extensions>=3.6.4; python_version < "3.8" in /usr/local/lib/python3.7/dist-packages (from importlib-metadata; python_version < "3.8") (3.7.4.3)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata; python_version < "3.8") (3.4.1)

```

New dataset creation

```
In [27]: def generateNewDataset(size,name,df):  
        df_pd = df[(df["Source"]<=size) | (df["Destination"]<=size)].persist()  
        df_pd = df_pd.compute()  
        df_pd.to_csv(path+project+"/Datasets/"+name,index=False)
```

```
In [ ]: path = "/content/drive/My Drive/"  
        project = "2_TwitterFollowGraph"  
  
        df = dask.dataframe.read_csv(path+project+"/Datasets/Twitter/twitter-2010.txt",sep = ' ')  
        df.columns = ["Destination","Source"]  
        generateNewDataset(40000,'Twitter/twitter-final.csv',df)
```

Complete dataset degree distribution

```
In [38]: def calculate_degree():  
        in_degree = {}  
        out_degree = {}  
        with open(path+project+"/Datasets/Twitter/twitter-2010.txt") as infile:  
            for line in infile:  
                data = line.split()  
                if data[1] in out_degree:  
                    out_degree[data[1]] += 1  
                else:  
                    out_degree[data[1]] = 1  
  
                if data[0] in in_degree:  
                    in_degree[data[0]] += 1  
                else:  
                    in_degree[data[0]] = 1  
        return in_degree, out_degree
```

```
In [39]: in_degree, out_degree = calculate_degree()
```

```
In [40]: def degree_distribution(degree_list):  
        degree_distribution = {}  
  
        for node in degree_list:  
            degree = degree_list[node]  
            if degree in degree_distribution:  
                degree_distribution[degree] += 1  
            else:  
                degree_distribution[degree] = 1
```

```

degree_distribution = sorted(degree_distribution.items(),key= lambda x:x[0])

degree = [value[0] for value in degree_distribution]
node_count = [value[1] for value in degree_distribution]

N = sum(node_count)
node_count = list(map(lambda x:x/N,node_count))

return degree,node_count

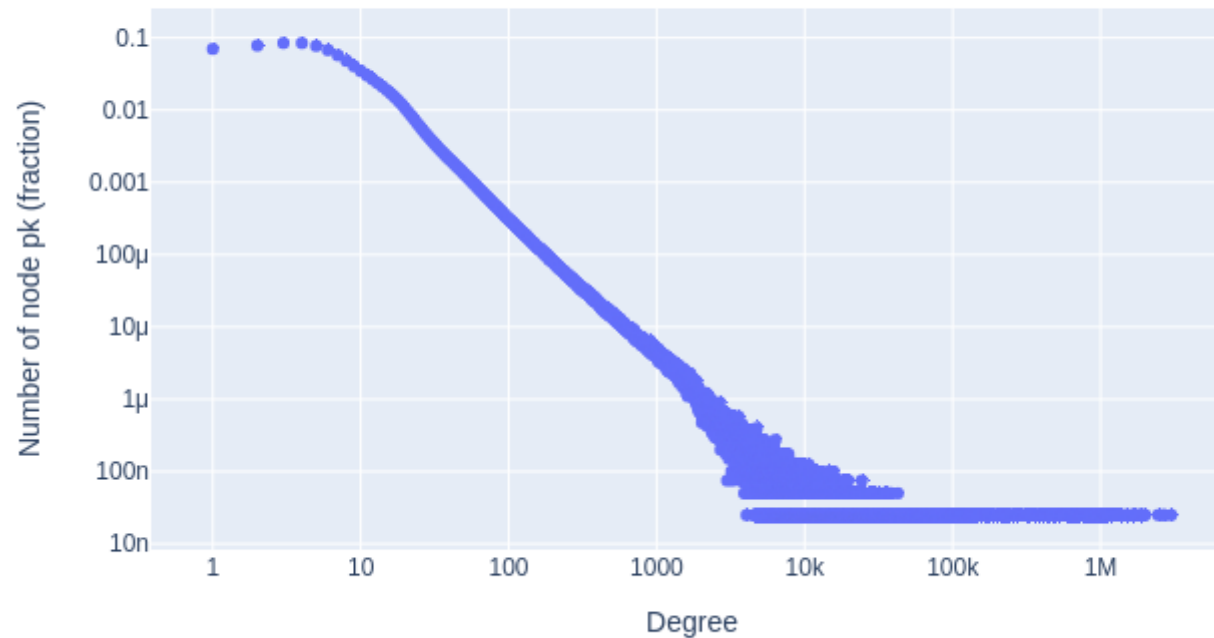
```

```

In [41]: degree,node_count = degree_distribution(in_degree)
import plotly.graph_objects as go
figure = go.Figure()
figure.add_trace(go.Scatter(x=degree,y=node_count,mode='markers',name="Mean"))
figure.update_xaxes(type="log",title_text="Degree")
figure.update_yaxes(type="log",title_text="Number of node pk (fraction)")
figure.update_layout(title="In degree distribution")
figure.show(renderer="png")

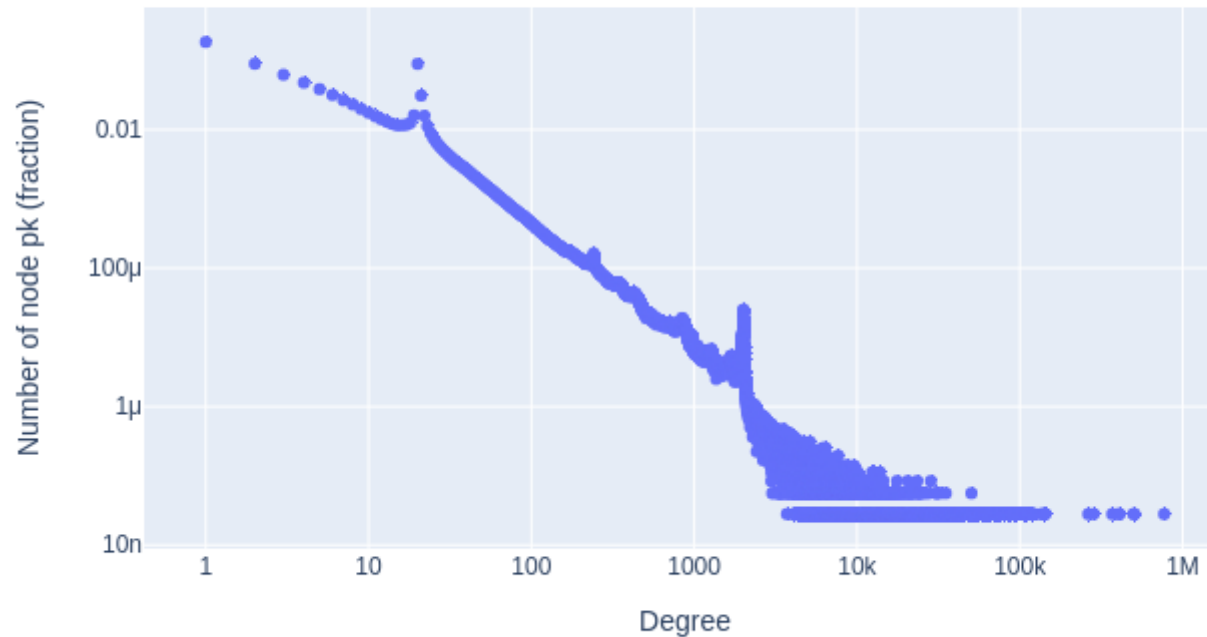
```

In degree distribution



```
In [42]: degree,node_count = degree_distribution(out_degree)
import plotly.graph_objects as go
figure = go.Figure()
figure.add_trace(go.Scatter(x=degree,y=node_count,mode='markers',name="Mean"))
figure.update_xaxes(type="log",title_text="Degree")
figure.update_yaxes(type="log",title_text="Number of node pk (fraction)")
figure.update_layout(title="Out degree distribution")
figure.show(renderer="png")
```

Out degree distribution



Twitter Ego Network creation

```
In [43]: def getHUB(degree,limit):
hubs = []
for node in degree:
    if degree[node]>limit:
        hubs.append(node)
return hubs

In [44]: def generateEgoNetwork(df_pd,hubs):
df_pd = df_pd[(df_pd["Destination"].isin(hubs))].persist()
df_pd = df_pd[(df_pd["Source"].isin(hubs))].persist()
df_pd = df_pd.compute()
df_pd.to_csv(path+project+"/Datasets/Twitter-Ego/twitter-ego.csv",index=False)
```

```
In [45]: hubs = getHUB(in_degree,50000)
```

```
In [ ]: generateEgoNetwork(df,hubs)
```