In [1]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

In [2]:
```
!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/eb6cbe96c8379c54819592bb228c58ed7386fcc60a55eca7db99432fd
f14/plotly-4.7.1-py2.py3-none-any.whl (11.5MB)
     |████████████████████████████████| 11.5MB 262kB/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 16:31:55--  https://github.com/plotly/orca/releases/download/v1.2.1/orca-1.2.1-x86_64.AppImage
Resolving github.com (github.com)... 192.30.255.113
Connecting to github.com (github.com)|192.30.255.113|: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=2021041
2T163155Z&X-Amz-Expires=300&X-Amz-Signature=f37e2e9a0fba6ee92972fe7fb04c009af2837cb68fd76bb2e648105197c386d6&X-Amz-Si
gnedHeaders=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 16:31:55--  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-A
mz-Date=20210412T163155Z&X-Amz-Expires=300&X-Amz-Signature=f37e2e9a0fba6ee92972fe7fb04c009af2837cb68fd76bb2e648105197
c386d6&X-Amz-SignedHeaders=host&actor_id=0&key_id=0&repo_id=99037241&response-content-disposition=attachment%3B%20fil
ename%3Dorca-1.2.1-x86_64.AppImage&response-content-type=application%2Foctet-stream
Resolving github-releases.githubusercontent.com (github-releases.githubusercontent.com)... 185.199.108.154, 185.199.1
09.154, 185.199.111.154, ...
Connecting to github-releases.githubusercontent.com (github-releases.githubusercontent.com)|185.199.108.154|:443... c
onnected.
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  72.4MB/s    in 0.7s

2021-04-12 16:31:56 (72.4 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 2s (2,209 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 [3]:
```
import pandas as pd
import networkx as nx
from collections import Counter
import plotly.graph_objects as go
import numpy as np
from tqdm.autonotebook import tqdm
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:6: TqdmExperimentalWarning: Using `tqdm.autonotebook.tqd
m` in notebook mode. Use `tqdm.tqdm` instead to force console mode (e.g. in jupyter console)
```

In [4]:
```
def create_graph(df_twitter):
  G=nx.DiGraph()
  edge_list = [tuple(edge) for edge in df_twitter.values]
  for edge in edge_list:
    G.add_edge(edge[1],edge[0])
  return G
```

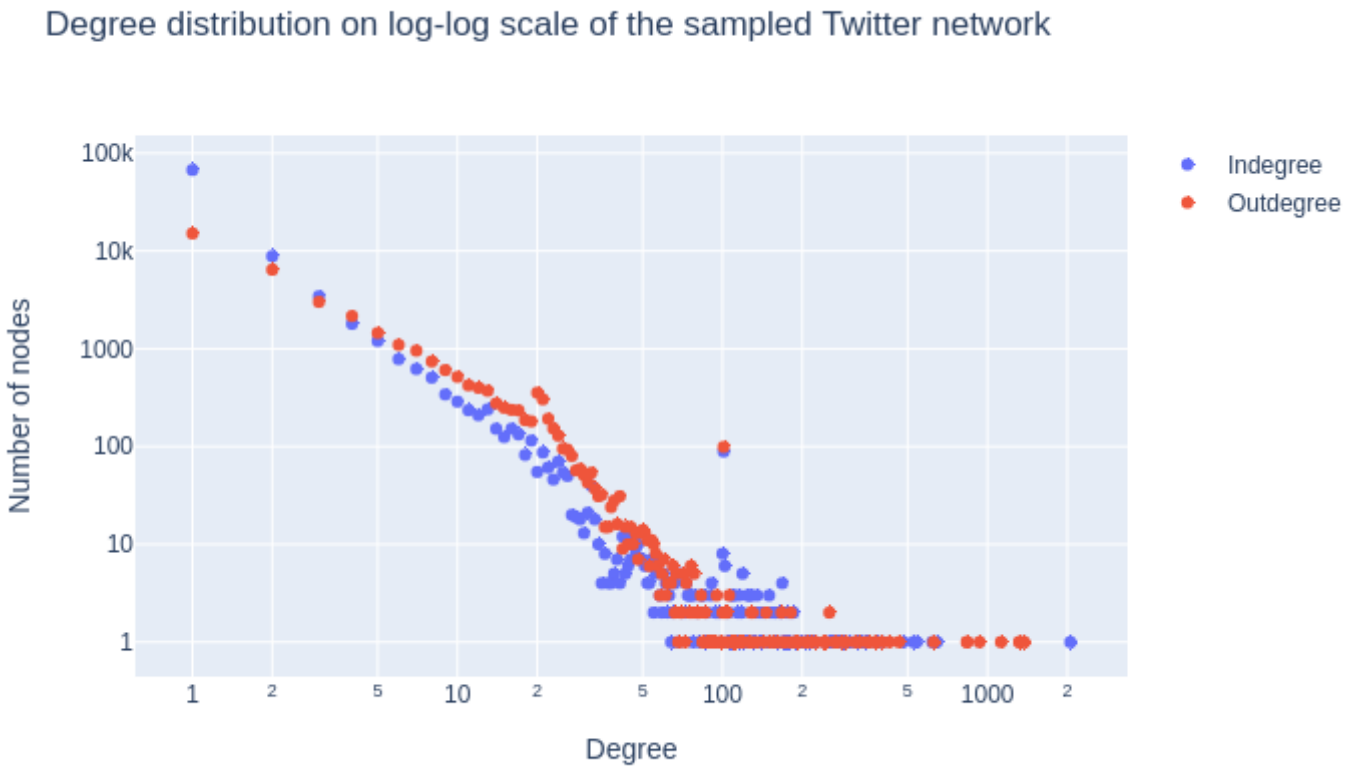
In [5]:
```
def compute_degree_distribution(G,subtitle):
  node_list=list(G.nodes)
  indegree_dict={}
  outdegree_dict={}
  for node in node_list:
    indegree_dict[node]=G.in_degree(node)
    outdegree_dict[node]=G.out_degree(node)
  indegree_dict_final=dict(sorted(dict(Counter(indegree_dict.values())).items()))
  outdegree_dict_final=dict(sorted(dict(Counter(outdegree_dict.values())).items()))
  figure = go.Figure()
  figure.add_trace(go.Scatter(x=list(indegree_dict_final),y=list(indegree_dict_final.values()),mode='markers',name="In
degree"))
  figure.add_trace(go.Scatter(x=list(outdegree_dict_final),y=list(outdegree_dict_final.values()),mode='markers',name=
"Outdegree"))
  figure.update_xaxes(type="log",title_text="Degree")
  figure.update_yaxes(type="log",title_text="Number of nodes")
  figure.update_layout(title="Degree distribution on log-log scale of the {}".format(subtitle))
  figure.show(renderer="png")
```

In [6]:
```
path = "/content/drive/My Drive/"
project_name="2_TwitterFollowGraph"
df_twitter=pd.read_csv(path+project_name+"/Datasets/Twitter/twitter-final.csv")
G=create_graph(df_twitter)
```

In [7]:
```
print(G.number_of_nodes(),G.number_of_edges())
```

```
98292 212314
```

In [8]:
```
compute_degree_distribution(G,"sampled Twitter network")
```



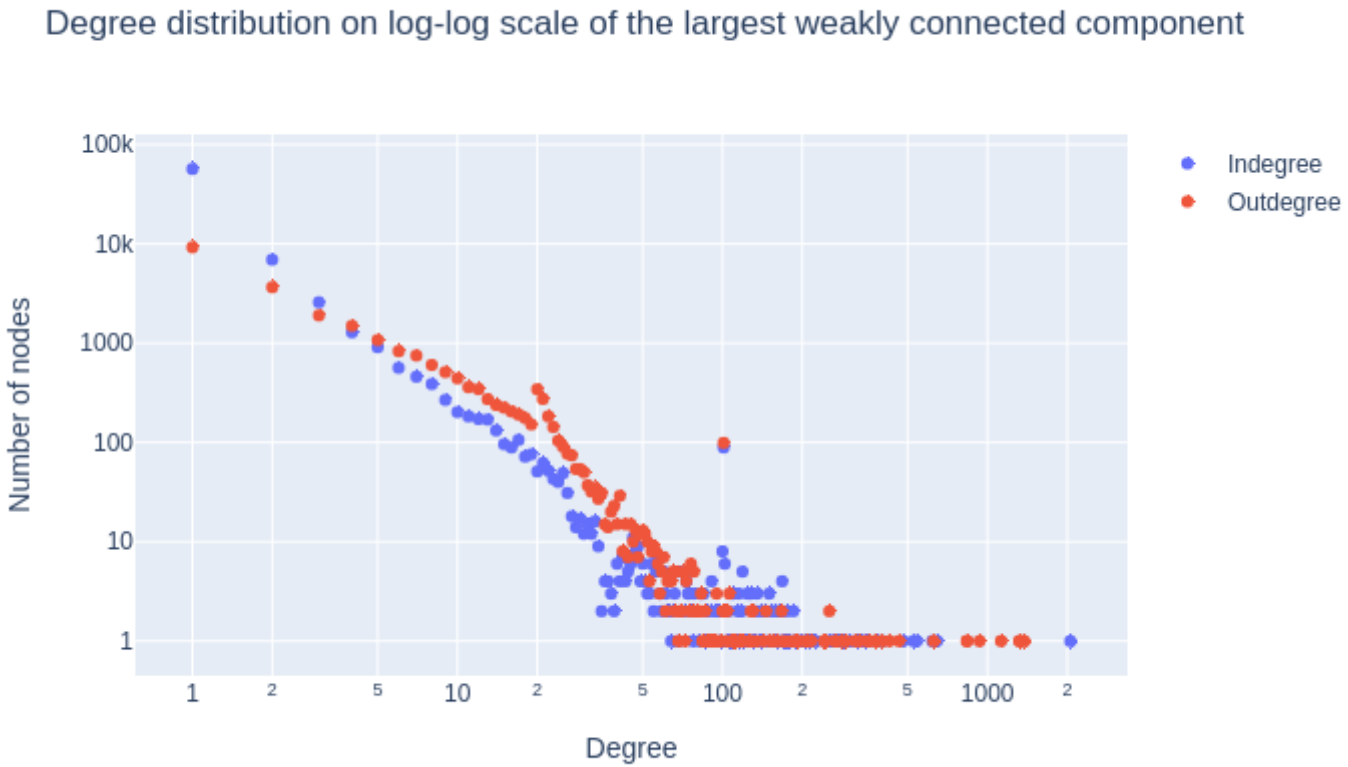
In [9]:
```
weak_list=[len(l) for l in list(nx.weakly_connected_components(G))]
weak_list.sort(reverse=True)
print("Size of top 10 weakly connected components in the network:{}".format(weak_list[:10]))
```

```
Size of top 10 weakly connected components in the network:[80074, 264, 101, 92, 83, 79, 67, 65, 50, 41]
```

In [10]:
```
print("Number of weakly connected components in the network: {}".format(len(list(nx.weakly_connected_components(G)))))
```

```
Number of weakly connected components in the network: 3851
```

In [11]:
```
subgraph_nodes_weakly = max(nx.weakly_connected_components(G),key=len)
largest_weakly_connected_component=G.subgraph(subgraph_nodes_weakly)
compute_degree_distribution(largest_weakly_connected_component,"largest weakly connected component")
```



In [ ]:
```
total_path_length=0
diameter=0
counter=0
for node in tqdm(set(subgraph_nodes_weakly)):
   distance_dict={}
   label_dict={}
   label=1
   if node not in set(list(label_dict)):
       src=node
       queue=[]
       queue.append(src)
       distance=0
       while len(queue)!=0:
           front=queue.pop(0)
           label_dict[front]=label
           neighbours=set(list(largest_weakly_connected_component.neighbors(front)))
           label_set=set(list(label_dict))
           distance_set=set(list(distance_dict))
           if len(neighbours)>0:
            for neighbour in neighbours:
                if neighbour not in label_set:
                    queue.append(neighbour)
                    if neighbour not in distance_set or distance_dict[neighbour]>distance_dict[front]+1:
                      if front not in distance_set:
                        distance_dict[neighbour]=1
                      else:
                        distance_dict[neighbour]=distance_dict[front]+1
                      counter+=1
   distance_list=list(distance_dict.values())
   total_path_length+=sum(distance_list)
   if len(distance_list)!=0:
    diameter=max(diameter,max(list(distance_dict.values())))
```

In [ ]:
```
print("Average path length of largest weakly connected component is {}".format(total_path_length/counter))
print("Average clustering coefficient of largest weakly connected component is {}".format(nx.average_clustering(larges
t_weakly_connected_component)))
print("Diameter of largest weakly connected component is {}".format(diameter))
```

```
Average path length of largest weakly connected component is 2.612595900587978
Average clustering coefficient of largest weakly connected component is 0.07628559864341032
Diameter of largest weakly connected component is 18
```

In [ ]:
```
print("Degree Assortavity Coeffecient of largest weakly connected component is {}".format(nx.degree_assortativity_coef
ficient(largest_weakly_connected_component)))
```

```
Degree Assortavity Coeffecient of largest weakly connected component is -0.08113703143196034
```

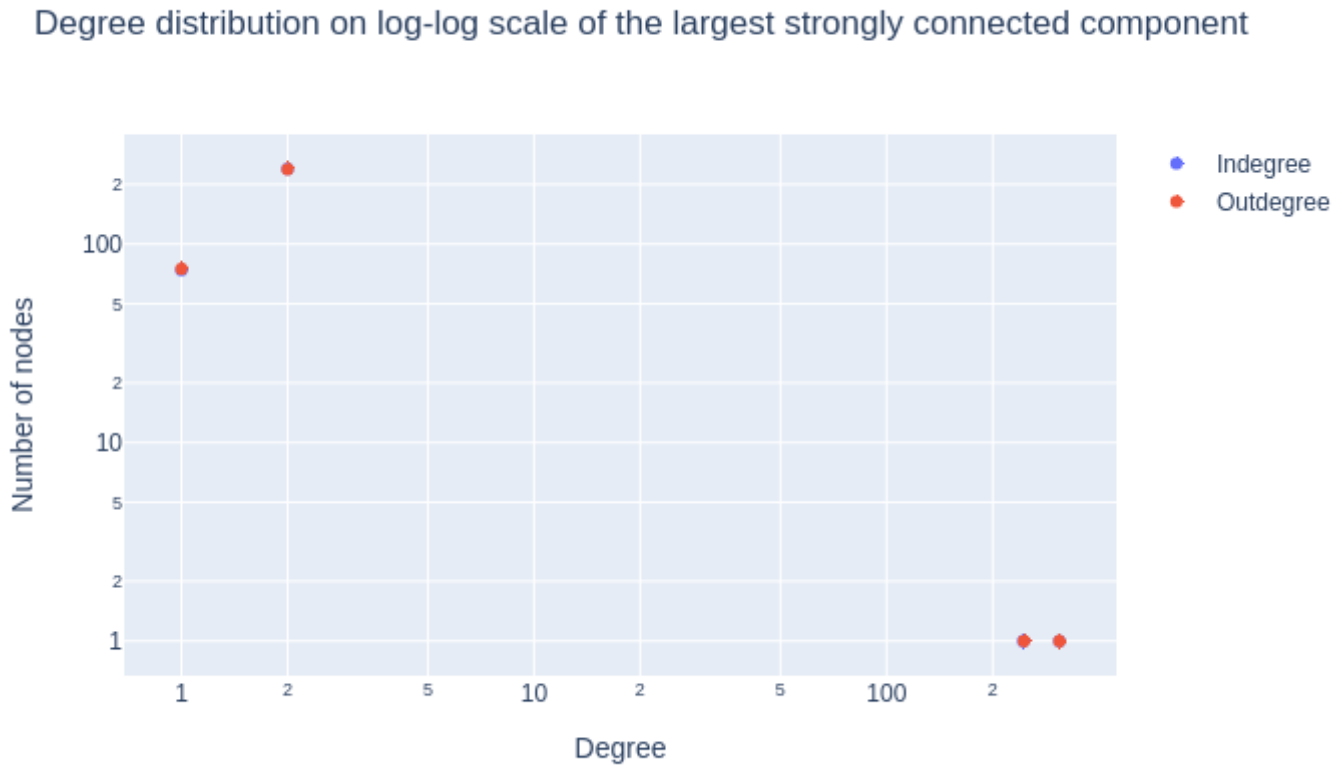
In [12]:
```
strong_list=[len(l) for l in list(nx.strongly_connected_components(G))]
strong_list.sort(reverse=True)
print("Size of top 10 strongly connected components in the network:{}".format(strong_list[:10]))
```

```
Size of top 10 strongly connected components in the network:[315, 231, 159, 143, 103, 98, 76, 75, 72, 69]
```

In [13]:
```
print("Number of strongly connected components in the network: {}".format(len(list(nx.strongly_connected_components(G
)))))
```

```
Number of strongly connected components in the network: 81273
```

In [14]:
```
subgraph_nodes_strong = max(nx.strongly_connected_components(G),key=len)
largest_strongly_connected_component=G.subgraph(subgraph_nodes_strong)
compute_degree_distribution(largest_strongly_connected_component,"largest strongly connected component")
```



In [ ]:
```
print("Average shortest path length of the largest strongly connected component is {}".format(nx.average_shortest_path
_length(largest_strongly_connected_component)))
print("Average clustering coefficient of the largest strongly connected component is {}".format(nx.average_clustering(
largest_strongly_connected_component)))
print("Diameter of the largest strongly connected component is {}".format(nx.diameter(largest_strongly_connected_compo
nent)))
```

```
Average shortest path length of the largest strongly connected component is 1.9958548175108686
Average clustering coefficient of the largest strongly connected component is 0.765120710745403
Diameter of the largest strongly connected component is 3
```

In [ ]:
```
print("Degree Assortavity Coeffecient of largest strongly connected component is {}".format(nx.degree_assortativity_co
efficient(largest_strongly_connected_component)))
```

```
Degree Assortavity Coeffecient of largest strongly connected component is -0.9711299760892993
```