In [1]:
```
!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 5.9MB/s
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from plotly==4.7.1) (1.15.0)
Requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.7/dist-packages (from plotly==4.7.1) (1.3.3)
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 16:23:16--  https://github.com/plotly/orca/releases/download/v1.2.1/orca-1.2.1-x86_64.AppImage
Resolving github.com (github.com)... 140.82.112.4
Connecting to github.com (github.com)|140.82.112.4|: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=2021041
2T162317Z&X-Amz-Expires=300&X-Amz-Signature=23b31ef668ca1c8d667aff2078ec979c987b71fd71411b75cceebb0850931ad5&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 16:23:17--  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=20210412T162317Z&X-Amz-Expires=300&X-Amz-Signature=23b31ef668ca1c8d667aff2078ec979c987b71fd71411b75cceebb0850
931ad5&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, 185.199.110.154, ...
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  64.7MB/s    in 0.8s

2021-04-12 16:23:18 (64.7 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 (1,835 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 [2]:
```
import pandas as pd
import networkx as nx
from collections import Counter
import plotly.graph_objects as go
import numpy as np
import pandas as pd
from tqdm.autonotebook import tqdm
```

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

In [3]:
```
path = "/content/drive/My Drive/"
project_name="2_TwitterFollowGraph"
df = pd.read_csv(path+project_name+"/Datasets/Linkedin/soc-linkedin.csv",sep = ' ')
df.columns = ["Source","Destination"]
edgelist = df[["Source","Destination"]].values.tolist()
```

In [4]:
```
def createAdjacencyList(edgelist):
  adjacencyList = {}
  for edge in edgelist:
    if edge[0] in adjacencyList:
      adjacencyList[edge[0]].add(edge[1])
    else:
      adjacencyList[edge[0]] = set([edge[1]])

    if edge[1] in adjacencyList:
      adjacencyList[edge[1]].add(edge[0])
    else:
      adjacencyList[edge[1]] = set([edge[0]])

  return adjacencyList
```

In [5]:
```
adjacencyList = createAdjacencyList(edgelist)
```

In [6]:
```
def calculate_degree_distribution(adjacencyList):
  degree = {}
  for node in adjacencyList:
    degree_node = len(adjacencyList[node])

    if degree_node in degree:
      degree[degree_node] += 1
    else:
      degree[degree_node] = 1

  degree_distribution = sorted(degree.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
```

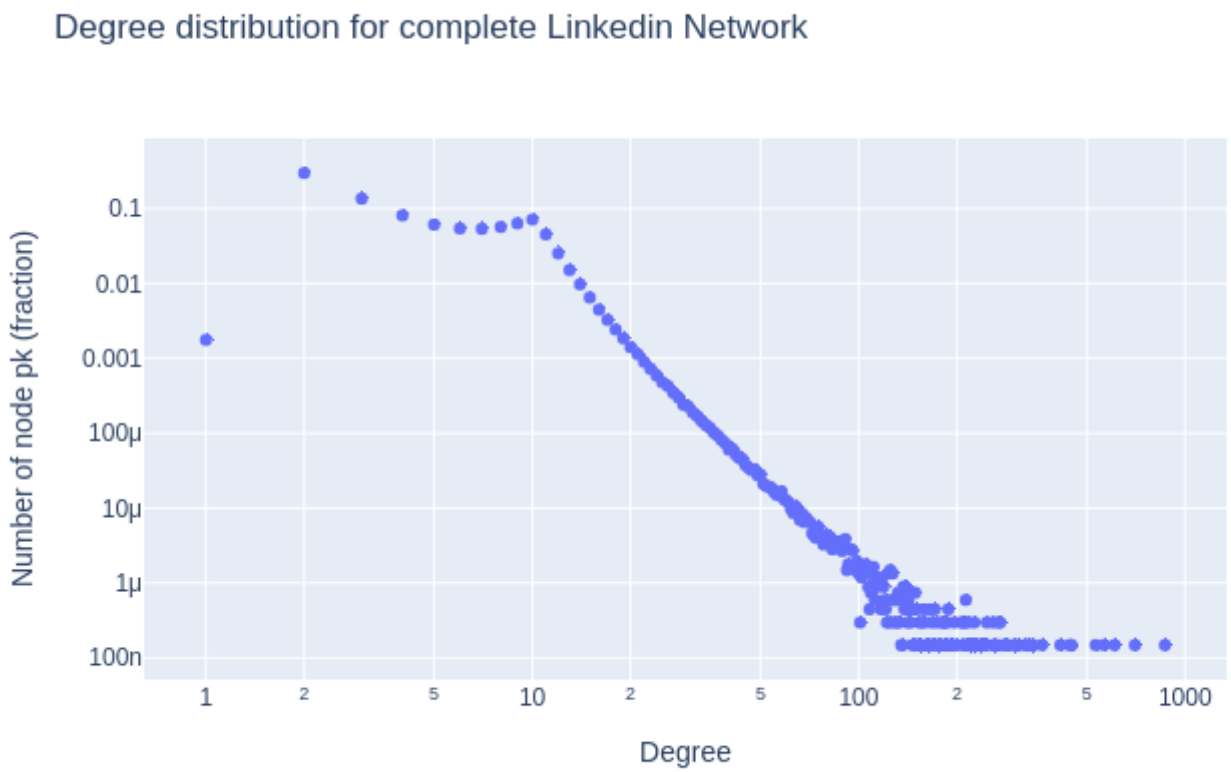
In [7]:
```
degree,node_count  = calculate_degree_distribution(adjacencyList)
```

In [8]:
```
def removedNodes(adjacencyList):
  removedNodes = []
  for node in adjacencyList:
    degree_node = len(adjacencyList[node])

    if degree_node < 10:
      removedNodes.append(node)

  return removedNodes
```

In [9]:
```
figure = go.Figure()
figure.add_trace(go.Scatter(x=degree,y=node_count,mode='markers'))
figure.update_xaxes(type="log",title_text="Degree")
figure.update_yaxes(type="log",title_text="Number of node pk (fraction)")
figure.update_layout(title="Degree distribution for complete Linkedin Network")
figure.show(renderer="png")
```



In [10]:
```
def create_graph(df,columns):
  G=nx.Graph()
  G.add_edges_from(df[columns].values.tolist())
  return G
```

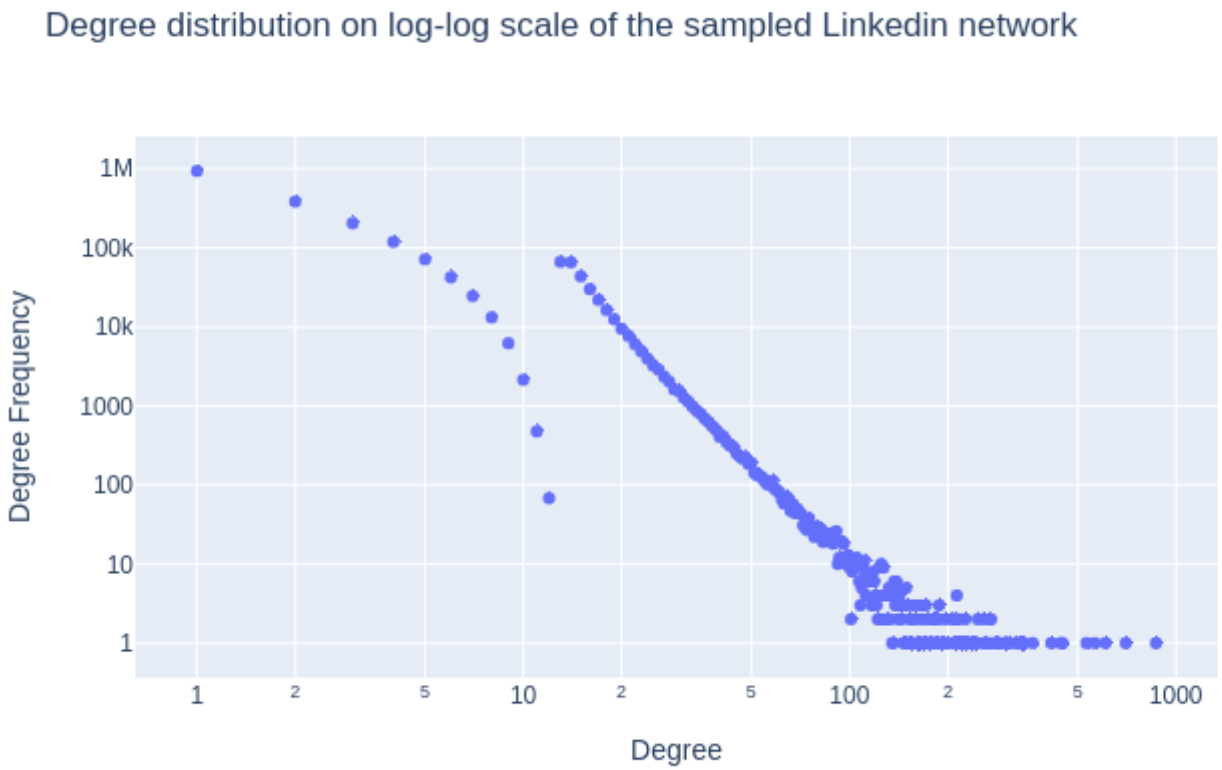
In [11]:
```
def compute_degree_distribution(G,subtitle):
  node_list=list(G.nodes)
  degree_dict={}
  for node in node_list:
    degree_dict[node]=G.degree(node)
  degree_dict_final=dict(sorted(dict(Counter(degree_dict.values())).items()))
  figure = go.Figure()
  figure.add_trace(go.Scatter(x=list(degree_dict_final),y=list(degree_dict_final.values()),mode='markers'))
  figure.update_xaxes(type="log",title_text="Degree")
  figure.update_yaxes(type="log",title_text="Degree Frequency")
  figure.update_layout(title="Degree distribution on log-log scale of the {}".format(subtitle))
  figure.show(renderer="png")
```

In [12]:
```
df_linkedin=pd.read_csv(path+project_name+"/Datasets/Linkedin/linkedin-1M.csv")
columns=df_linkedin.columns
G=create_graph(df_linkedin,columns)
```

In [13]:
```
print(G.number_of_nodes(),G.number_of_edges())
```

```
2122488 4626421
```

In [14]:
```
compute_degree_distribution(G,"sampled Linkedin network")
```



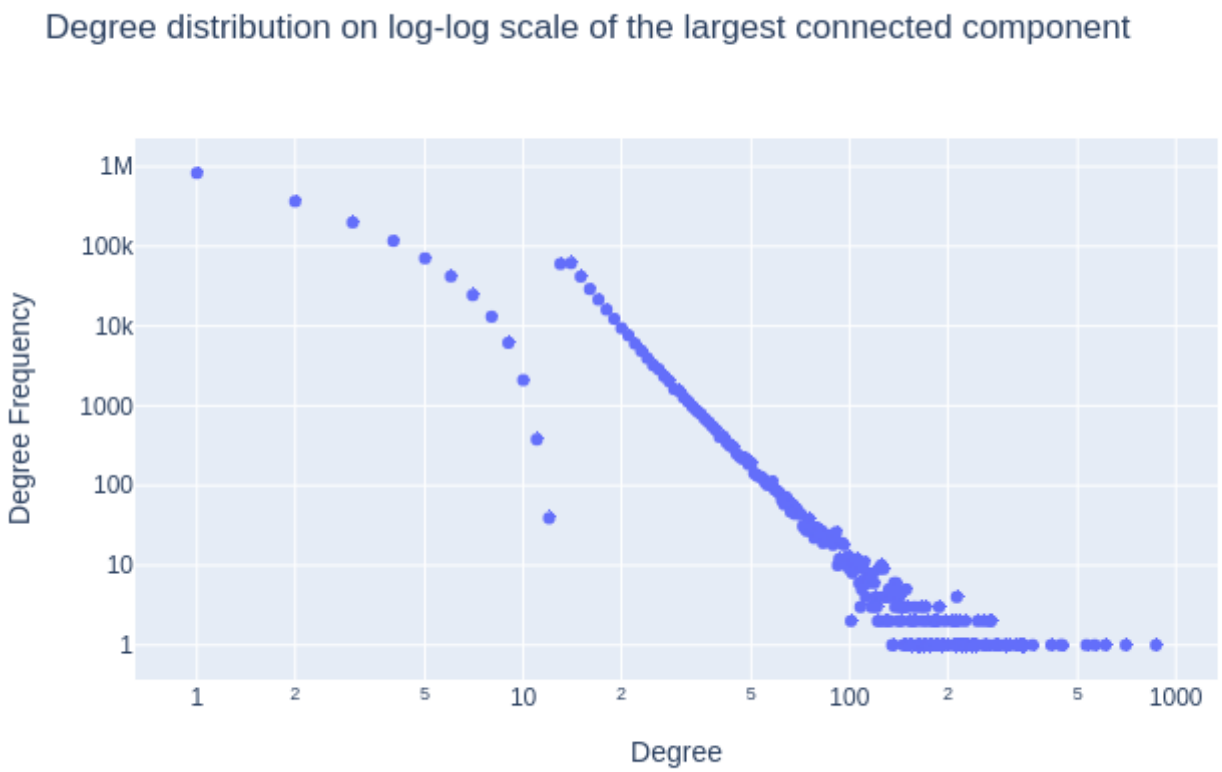
In [15]:
```
connected_components_length_list=[len(l) for l in list(nx.connected_components(G))]
connected_components_length_list.sort(reverse=True)
print("Size of top 10 connected components in the network:{}".format(connected_components_length_list[:10]))
```

```
Size of top 10 connected components in the network:[1971455, 384, 344, 329, 297, 272, 247, 242, 240, 230]
```

In [16]:
```
print("Number of connected components in the network:{}".format(len(list(nx.connected_components(G)))))
```

```
Number of connected components in the network:7704
```

In [17]:
```
subgraph_nodes = max(nx.connected_components(G),key=len)
largest_connected_component=G.subgraph(subgraph_nodes)
compute_degree_distribution(largest_connected_component,"largest connected component")
```



In [ ]:
```
print("Average clustering coefficient of largest connected component is {}".format(nx.average_clustering(largest_conne
cted_component)))
print("Degree Assortavity Coeffecient of largest connected component is {}".format(nx.degree_assortativity_coefficient
(largest_connected_component)))
```

```
Average clustering coefficient of largest connected component is 0.2673942721125565
Degree Assortavity Coeffecient of largest connected component is -0.07320491676041042
```