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
```

```
Requirement already satisfied: plotly==4.7.1 in /usr/local/lib/python3.7/dist-packages (4.7.1)
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)
--2021-04-12 16:40:17--  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=2021041
2T164017Z&X-Amz-Expires=300&X-Amz-Signature=456f91dfda71c2cd17fa00745fdb5f50871acbe0cb0f41e72cab054a73ba6883&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 16:40:17--  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=20210412T164017Z&X-Amz-Expires=300&X-Amz-Signature=456f91dfda71c2cd17fa00745fdb5f50871acbe0cb0f41e72cab054a73
ba6883&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  77.1MB/s    in 0.6s

2021-04-12 16:40:18 (77.1 MB/s) - ‘/usr/local/bin/orca’ saved [51607939/51607939]

Reading package lists... Done
Building dependency tree
Reading state information... Done
libgtk2.0-0 is already the newest version (2.24.32-1ubuntu1).
libgconf-2-4 is already the newest version (3.2.6-4ubuntu1).
xvfb is already the newest version (2:1.19.6-1ubuntu4.8).
0 upgraded, 0 newly installed, 0 to remove and 31 not upgraded.
```

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]:
```
def create_graph(df):
  G=nx.Graph()
  edge_list = [tuple(edge) for edge in df.values]
  for edge in edge_list:
    G.add_edge(edge[1],edge[0])
  return G
```

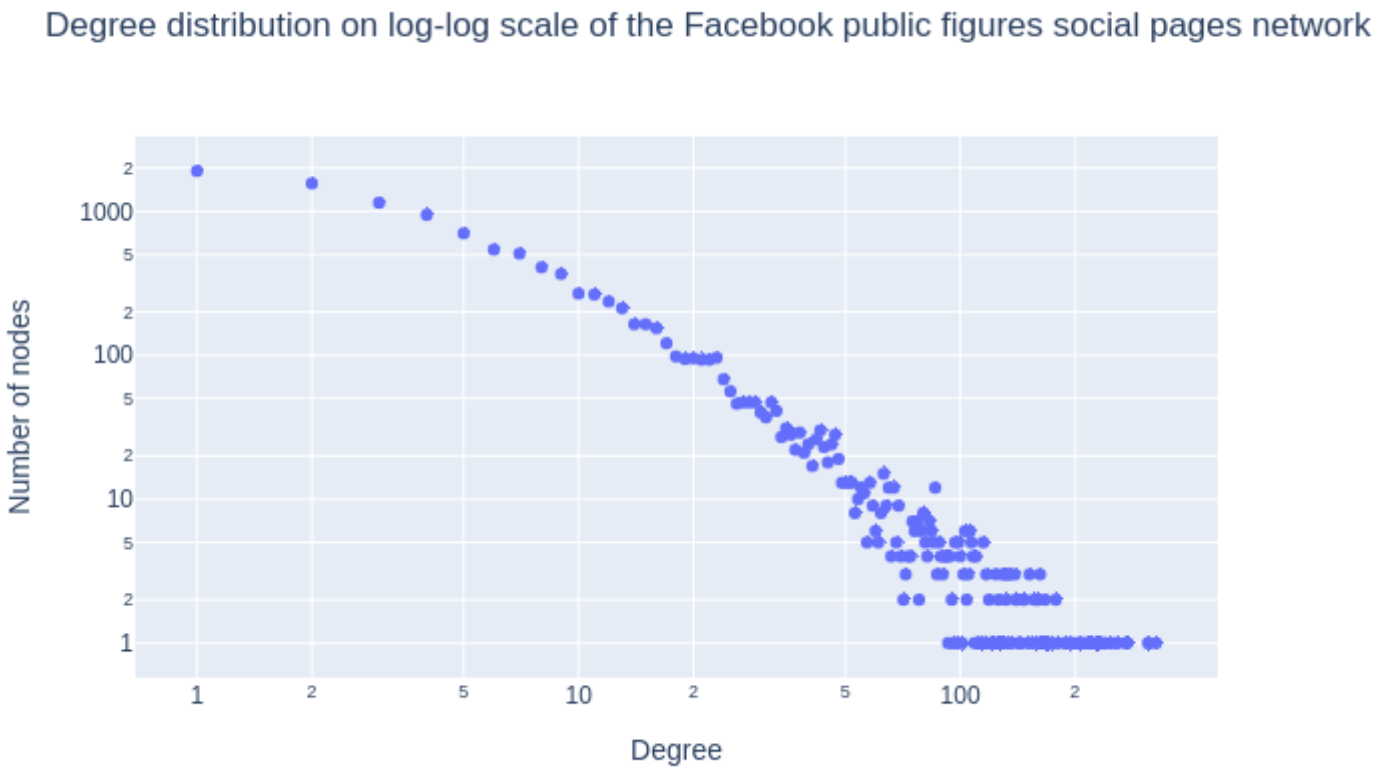
In [4]:
```
def compute_degree_distribution(G,subtitle):
  node_list=list(G.nodes)
  degree_dict={}
  for node in node_list:
    degree_dict[node]=G.degree(node)
  degree_dict_final=dict(sorted(dict(Counter(degree_dict.values())).items()))
  figure = go.Figure()
  figure.add_trace(go.Scatter(x=list(degree_dict_final),y=list(degree_dict_final.values()),mode='markers'))
  figure.update_xaxes(type="log",title_text="Degree")
  figure.update_yaxes(type="log",title_text="Number of nodes")
  figure.update_layout(title="Degree distribution on log-log scale of the {}".format(subtitle))
  figure.show(renderer="png")
```

In [5]:
```
path = "/content/drive/My Drive/"
project_name="2_TwitterFollowGraph"
df_facebook=pd.read_csv(path+project_name+"/Datasets/Facebook/public_figure_edges.csv")
G=create_graph(df_facebook)
```

In [6]:
```
print(G.number_of_nodes(),G.number_of_edges())
```

```
11565 67114
```

In [7]:
```
compute_degree_distribution(G,"Facebook public figures social pages network")
```



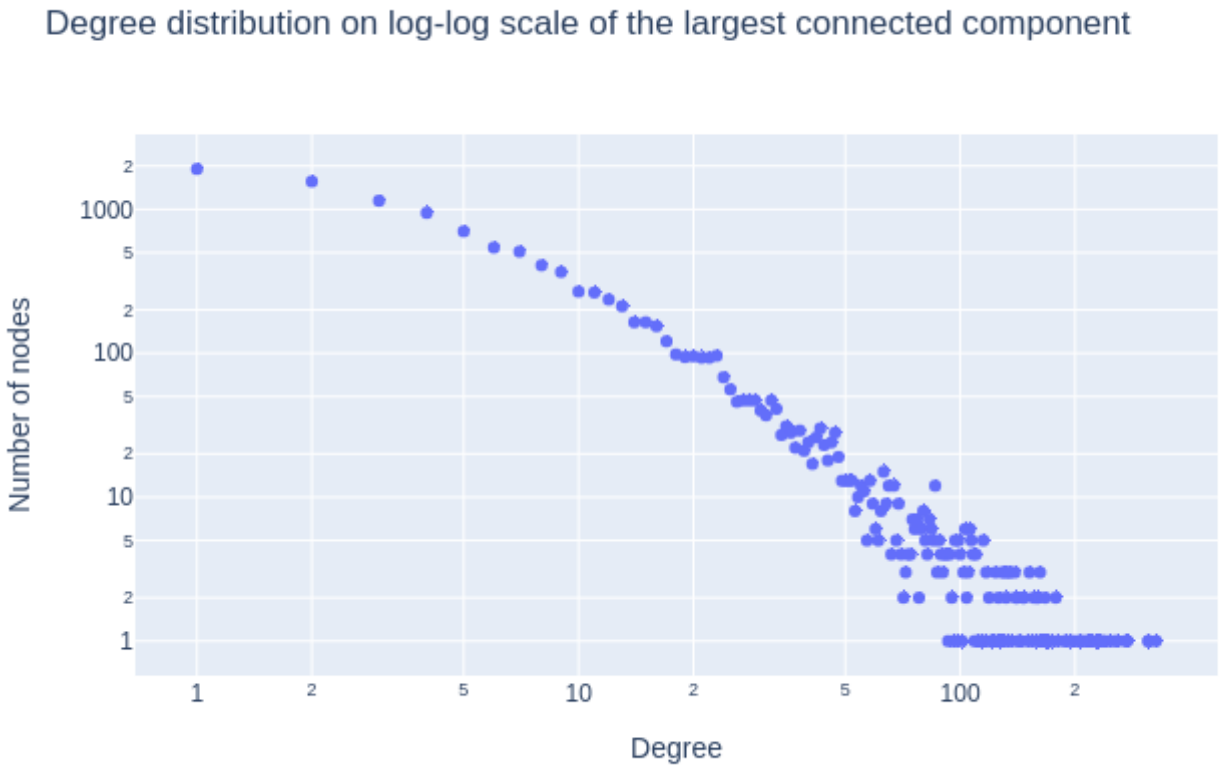
In [8]:
```
connected_components_length_list=[len(l) for l in list(nx.connected_components(G))]
connected_components_length_list.sort(reverse=True)
print("Size of top 10 connected components in the network:{}".format(connected_components_length_list[:10]))
```

```
Size of top 10 connected components in the network:[11565]
```

In [9]:
```
print("Number of connected components in the network:{}".format(len(list(nx.connected_components(G)))))
```

```
Number of connected components in the network:1
```

In [10]:
```
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_connected_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 0.17934725117824554
Degree Assortavity Coeffecient of largest connected component is 0.202161548290631
```

In [ ]:
```
print("Average path length of largest connected component is {}".format(nx.average_shortest_path_length(largest_connected_component)))
print("Diameter of largest connected component is {}".format(nx.diameter(largest_connected_component)))
```

```
Average path length of largest connected component is 4.622979301417417
Diameter of largest connected component is 15
```