

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

Mounted at /content/drive

```
In [2]: !pip install cdlib  
import sys  
!{sys.executable} -m pip install python-igraph  
!{sys.executable} -m pip install leidenalg
```



```

Collecting shuffle-graph==1.1.1
  Downloading https://files.pythonhosted.org/packages/18/14/22153a8d389370cae50bedd58239f640afe9717085a7a4e9bc5f801886aa/shuffle_graph-1.1.1-py3-none-any.whl
Collecting ASLPAw==2.0.0
  Downloading https://files.pythonhosted.org/packages/af/a4/67dcbe51833c044c60a892e4ab5ca90a73e77317f0dcf377e2f9c6451922/ASLPAw-2.0.0-py3-none-any.whl
Collecting markov-clustering
  Downloading https://files.pythonhosted.org/packages/76/42/19e11a42fa952d35116b90577e2cde31c541ce78364a52167f852864ba29/markov_clustering-0.0.6.dev0-py3-none-any.whl
Collecting bimlpa
  Downloading https://files.pythonhosted.org/packages/11/8c/7d409d7e7bcbdd261c7f19737feaa1ef66514512c7f7d6b9a776d7aac7c33/bimlpa-0.1.2-py3-none-any.whl
Requirement already satisfied: numpy>=1.15.* in /usr/local/lib/python3.7/dist-packages (from cdlib) (1.19.5)
Requirement already satisfied: scikit-learn<0.24,>=0.21.* in /usr/local/lib/python3.7/dist-packages (from cdlib) (0.22.2.post1)
Collecting omega-index-py3
  Downloading https://files.pythonhosted.org/packages/fb/57/ab48f801af9ef0ce273fe1969df825a99f52a73ff18767c7084f60471028/omega_index_py3-0.3-py3-none-any.whl
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from pooch->cdlib) (20.9)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from pooch->cdlib) (2.23.0)
Requirement already satisfied: appdirs in /usr/local/lib/python3.7/dist-packages (from pooch->cdlib) (1.4.4)
Requirement already satisfied: decorator>=4.3.0 in /usr/local/lib/python3.7/dist-packages (from networkx>=2.4->cdlib) (4.4.2)
Collecting amply>=0.1.2
  Downloading https://files.pythonhosted.org/packages/f3/c5/dfa09dd2595a2ab2ab4e6fa7bebef9565812722e1980d04b0edce5032066/amply-0.1.4-py3-none-any.whl
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.24->cdlib) (2018.9)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas>=0.24->cdlib) (2.8.1)
Collecting pygsp
  Downloading https://files.pythonhosted.org/packages/d4/89/2f4aa73cccf12bec5179ac5d52a68b508120c838b7e5d456f5ea0c8beade/PyGSP-0.5.1-py2.py3-none-any.whl (1.8MB)
  |████████████████████████████████████████| 1.8MB 31.4MB/s
Collecting gensim==3.8.3
  Downloading https://files.pythonhosted.org/packages/5c/4e/afe2315e08a38967f8a3036bbe7e38b428e9b7a90e823a83d0d49df1adf5/gensim-3.8.3-cp37-cp37m-manylinux1_x86_64.whl (24.2MB)
  |████████████████████████████████████████| 24.2MB 1.3MB/s
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from karateclub>=1.0.0->cdlib) (1.15.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=3.0.*->cdlib) (1.3.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages (fr

```

```

om matplotlib>=3.0.*->cdlib) (2.4.7)
Requirement already satisfied: cyclopy>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib>=3.0.*->cdlib)
(0.10.0)
Collecting count-dict>=1.0.1
  Downloading https://files.pythonhosted.org/packages/b4/99/e46091f8f5f09ec4346e9b59da4f31838345878b54da0c102ac0be9c6
757/count_dict-1.1.1-py3-none-any.whl
Collecting multivalued-dict>=1.7.1
  Downloading https://files.pythonhosted.org/packages/38/d9/60639ea482acbf0d2defec1d2ab0f58f42bed65985a9e0beb5b9c6a70
887/multivalued_dict-2.0.1-py3-none-any.whl
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-packages (from scikit-learn<0.24,>=0.21.
*->cdlib) (1.0.1)
Requirement already satisfied: urllib3!=1.25.0,!<1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (fro
m requests->pooch->cdlib) (1.24.3)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->pooch->cd
lib) (2020.12.5)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->pooch->cdl
ib) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->pooch->cdlib)
(2.10)
Requirement already satisfied: docutils>=0.3 in /usr/local/lib/python3.7/dist-packages (from amply>=0.1.2->pulp>=2.1-
>cdlib) (0.16)
Requirement already satisfied: smart-open>=1.8.1 in /usr/local/lib/python3.7/dist-packages (from gensim==3.8.3->karat
eclub>=1.0.0->cdlib) (4.2.0)
Building wheels for collected packages: python-louvain, karateclub
  Building wheel for python-louvain (setup.py) ... done
  Created wheel for python-louvain: filename=python_louvain-0.14-cp37-none-any.whl size=9292 sha256=1b540414e0664113c
a23650f013469d5c2e407858cb78baf8352a0412fc21335
  Stored in directory: /root/.cache/pip/wheels/e7/8d/24/6b3a464bb23e96ecba3f68868e85721534fd8158a9cd7b426b
  Building wheel for karateclub (setup.py) ... done
  Created wheel for karateclub: filename=karateclub-1.0.24-cp37-none-any.whl size=94202 sha256=81a3ea43c99cdbe2f805c3
da9a921201b3d825075f2afc6afbe53bc2156dd716
  Stored in directory: /root/.cache/pip/wheels/ab/a4/81/a2761ff51bb1caa318e69e4a30f5d7d39c14f10efe73365279
Successfully built python-louvain karateclub
Installing collected packages: eva-lcd, python-louvain, pquality, dynetx, amply, pulp, demon, pygsp, gensim, karatecl
ub, nf1, chinese-whispers, shuffle-graph, count-dict, multivalued-dict, ASLPAW, markov-clustering, bimlpa, omega-inde
x-py3, cdlib
  Found existing installation: python-louvain 0.15
  Uninstalling python-louvain-0.15:
    Successfully uninstalled python-louvain-0.15
  Found existing installation: gensim 3.6.0
  Uninstalling gensim-3.6.0:
    Successfully uninstalled gensim-3.6.0

```

Successfully installed ASLPAw-2.0.0 amply-0.1.4 bimlpa-0.1.2 cdlib-0.2.0 chinese-whispers-0.7.4 count-dict-1.1.1 demo
n-2.0.5 dynetx-0.2.4 eva-lcd-0.1.0 gensim-3.8.3 karateclub-1.0.24 markov-clustering-0.0.6.dev0 multivalued-dict-2.0.1
nf1-0.0.3 omega-index-py3-0.3 pquality-0.0.7 pulp-2.4 pygsp-0.5.1 python-louvain-0.14 shuffle-graph-1.1.1

Collecting python-igraph

Downloading https://files.pythonhosted.org/packages/ae/12/1fbdb491d89fad8abb7aca0189978655cfdc984a380b846478f2ccdfdad8/python_igraph-0.9.1-cp37-cp37m-manylinux2010_x86_64.whl (3.2MB)

|██| 3.2MB 4.5MB/s

Collecting texttable>=1.6.2

Downloading <https://files.pythonhosted.org/packages/06/f5/46201c428aeb0e0ecfa83df66bf3e6caa29659dbac5a56ddfd83cae0d4a4/texttable-1.6.3-py2.py3-none-any.whl>

Installing collected packages: texttable, python-igraph

Successfully installed python-igraph-0.9.1 texttable-1.6.3

Collecting leidenalg

Downloading https://files.pythonhosted.org/packages/70/68/3f7c56da3e7ac576e5e090f2a7e8637f96e4a790688dbec63b938530ef15/leidenalg-0.8.3-cp37-cp37m-manylinux2010_x86_64.whl (2.4MB)

|██| 2.4MB 4.4MB/s

Requirement already satisfied: python-igraph>=0.8.0 in /usr/local/lib/python3.7/dist-packages (from leidenalg) (0.9.1)

Requirement already satisfied: texttable>=1.6.2 in /usr/local/lib/python3.7/dist-packages (from python-igraph>=0.8.0->leidenalg) (1.6.3)

Installing collected packages: leidenalg

Successfully installed leidenalg-0.8.3

```
In [3]: import pandas as pd
import networkx as nx
from networkx.algorithms.community centrality import girvan_newman
from cdlib import algorithms
import numpy as np
import pickle
from networkx.algorithms.link_analysis.pagerank_alg import pagerank
from tqdm.autonotebook import tqdm
import pickle
```

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:8: TqdmExperimentalWarning: Using `tqdm.autonotebook.tqdm` in notebook mode. Use `tqdm.tqdm` instead to force console mode (e.g. in jupyter console)

```
In [4]: import tweepy
        from PIL import Image
        import requests
        from io import BytesIO
        consumer_key = "CtmrTSUsnIRHKSgj4ktqK4NKb"
        consumer_secret = "1MNfKmpTsrJcq7WMYxOoVwEgFCK5DNMQeC43IZDkPucP0bCnZ"
        access_key = "1231240089600057344-rqYnkeVvMaB1XwfvrERfI7aF38r69L"
        access_secret = "KmRfAoexHxfkHKJaC3hwB0sgtfocy58IqdzYiWeaXSG0"
```

```
In [5]: path = "/content/drive/My Drive/"
        project_name="2_TwitterFollowGraph"
        dataframe = pd.read_csv(path+project_name+"/Datasets/Twitter/twitter-final.csv")
        columns=list(dataframe.columns)
```

```
In [6]: dataset = pd.read_csv(path+project_name+"/Datasets/Twitter/twitter-2010-ids.csv")
        node_id_list = dataset['node_id'].to_list()
        twitter_id_list = dataset['twitter_id'].to_list()
        node_id_twitter_id = {node_id_list[i]: twitter_id_list[i] for i in range(len(node_id_list))}
```

```
In [7]: def createGraph(dataframe,columns):
        edgelist = dataframe[columns].values.tolist()
        graph = nx.Graph()
        graph.add_edges_from(edgelist)
        return graph
```

```
In [8]: graph = createGraph(dataframe,columns)
```

```
In [ ]: def sortedCommunities(communities):
        communitiesSorted = sorted(communities,key=lambda x:len(x),reverse=True)
        return communitiesSorted
```

```
In [ ]: def topDegreeNodeCommunities(communities,graph,size):
        degree = graph.degree()
        newCommunityList = []
        for community in communities:
            newCommunity = sorted(community,key=lambda x:degree[x],reverse=True)
            newCommunityList.append(newCommunity[:size])
        return newCommunityList
```

```
In [ ]: def generateCommunities(graph,method):
        if method == "leiden":
            communities = algorithms.leiden(graph)
        elif method == "walktrap":
            communities = algorithms.walktrap(graph)
        elif method == "surprise_communities":
            communities = algorithms.surprise_communities(graph)
        communitiesSorted = sortedCommunities(communities.communities)[:20]
        topNodeCommunities = topDegreeNodeCommunities(communitiesSorted,graph,15) # Set third parameter to get required size
        return topNodeCommunities
```

```
In [15]: # Function to extract usernames
def get_username(twitter_id):
    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_key, access_secret)
    api = tweepy.API(auth)
    try:
        username = api.get_user(id=twitter_id)
        user_dict = username.__getattribute__("_json")
        return user_dict
    except Exception:
        pass
```



```
In [ ]: def filter_communities(communities_list):
    infile = open(path+project_name+"/PickleFiles/node_info_pickle.pkl", 'rb')
    node_map_dict = pickle.load(infile)
    infile.close()
    community_names_list=[[('NA',False)]*len(communities_list[0]) for i in range(len(communities_list))]
    for i,community in enumerate(communities_list):
        for j,member in enumerate(community):
            if member in set(list(node_map_dict)) and node_map_dict[member] is not None:
                community_names_list[i][j]=(node_map_dict[member]['name'],node_map_dict[member]['verified'],node_map_dict[member]['followers_count'])
            else:
                profile_dict=get_username(node_id_twitter_id[member])
                if profile_dict is not None:
                    community_names_list[i][j]=(profile_dict['name'],profile_dict['verified'],profile_dict['followers_count'])

    final_community_list=[]
    for community in community_names_list:
        temp_list=[]
        for member in community:
            if member[1] is True:
                temp_list.append((member[0],member[2]))
        if len(temp_list)!=0:
            final_community_list.append(temp_list)
    return final_community_list
```

```
In [9]: # Function to extract usernames
def profile_pic_fetcher(twitter_id):
    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_key, access_secret)
    api = tweepy.API(auth)
    try:
        username = api.get_user(id=twitter_id)
        user_dict = username.__getattribute__("_json")
        user_name = user_dict["name"]
        profile_pic_url=user_dict["profile_image_url_https"]
        if "default_profile_normal" not in profile_pic_url:
            profile_pic_url=profile_pic_url.replace("_normal","")
        response = requests.get(profile_pic_url)
        profile_pic = Image.open(BytesIO(response.content))
        return profile_pic,user_name
    except Exception:
        pass
```

```
In [10]: communities_dict={}
page_rank_dict={}

```

```
In [ ]: #Leiden community algorithm
communities_list=generateCommunities(graph,"leiden")
final_communities_list=filter_communities(communities_list)
communities_dict["leiden"]=final_communities_list
print(final_communities_list)
```

```
[[('CNN Breaking News', 61190864), ('ashton kutcher', 17622863), ('Ellen DeGeneres', 79150582), ('Britney Spears', 56074625), ('Twitter', 59475201), ('Barack Obama', 130091376), ('Lance Armstrong', 3225007), ('Oprah Winfrey', 43646337), ('Ryan Seacrest', 15390398), ('Demi Moore', 4565235), ('SHAQ', 15628657), ('The New York Times', 49711344), ('jim my fallon', 51868620), ('Al Gore', 3054622)], [('Leo Laporte, Chief TWiT and The Tech Guy', 532306)], [('Stephen Fry', 12634103), ('Jonathan Ross', 5044515), ('Russell Brand', 11206944), ('Phillip Schofield', 4564965), ('Chris Moyles', 3600198)], [('Selena Gomez', 64958647), ('Demi Lovato', 55285517), ('Jonas Brothers', 4827694)], [('Josh Olin', 103247)], [('Luciano Huck', 13283352)], [('Stocktwits', 877136)], [('little farma', 564661), ('Superinteressante', 4038373), ('Radio-Canada Techno', 15129)], [('Glenn Beck', 1312959), ('Brad Paisley', 4367501)]]
```

```
In [ ]: #Surprise communities algorithm
communities_list=generateCommunities(graph,"surprise_communities")
final_communities_list=filter_communities(communities_list)
communities_dict["surprise_communities"]=final_communities_list
print(final_communities_list)
```

```
[[('CNN Breaking News', 61190864), ('ashton kutcher', 17622863), ('Ellen DeGeneres', 79150582), ('Britney Spears', 56074625), ('Twitter', 59475201), ('Lance Armstrong', 3225007), ('Oprah Winfrey', 43646337), ('Ryan Seacrest', 15390398), ('Demi Moore', 4565235), ('SHAQ', 15628657), ('The New York Times', 49711344), ('jimmy fallon', 51868620), ('Al Gore', 3054622), ('Coldplay', 23301988)], [('Stephen Fry', 12634103), ('Jonathan Ross', 5044515), ('Russell Brand', 11206944), ('Phillip Schofield', 4564965), ('Chris Moyles', 3600198), ('The Unnamed Artist', 5762042), ('Jimmy Carr', 6737846), ('fearne cotton', 6890318), ('Jamie Oliver', 6516271)], [('George Mason University', 27191)], [('buten un binnen', 39868)], [('elizabeth', 4383)]]
```

```
In [ ]: #Walktrap communitites algorithm
communities_list=generateCommunities(graph,"walktrap")
final_communities_list=filter_communities(communities_list)
communities_dict["walktrap"]=final_communities_list
print(final_communities_list)
```

```
[[('CNN Breaking News', 61190864), ('Britney Spears', 56074625), ('Twitter', 59475201)], [('ashton kutcher', 17622863), ('Ellen DeGeneres', 79150582), ('Lance Armstrong', 3225007), ('Oprah Winfrey', 43646337), ('Demi Moore', 4565235), ('The New York Times', 49711344), ('jimmy fallon', 51868620), ('Al Gore', 3054622), ('@bbcclick', 2152173), ('Pen n Jillette', 1777615)], [('Selena Gomez', 64958647), ('Demi Lovato', 55285517)], [('Leo Laporte, Chief TWiT and The Tech Guy', 532306), ('John C. Dvorak', 99736), ('Alex Albrecht', 77584), ('Adam Savage', 1473338)], [('Phillip Schofield', 4564965), ('Jimmy Carr', 6737846), ('fearne cotton', 6890318)], [('Fox News', 20146730), ('Glenn Beck', 1312961)], [('Deepak Chopra', 3257759)], [('Jan Willem Alphenaar', 18699), ('buten un binnen', 39865)], [('Smashing Magazine', 963478), ('Google AdSense', 565352)]]
```

```
In [11]: def rank_on_page_rank(graph):
    pagerank_result = dict(pagerank(graph))
    pagerank_result = sorted(pagerank_result.items(),key = lambda x:x[1],reverse=True)
    return pagerank_result
```

```
In [12]: pagerank_result=rank_on_page_rank(graph)
```

```
In [13]: def filter_pagerank(pagerank_result,size):
         final_pagerank_result=[]
         for entity in tqdm(pagerank_result[:size]):
             profile_dict=get_username(node_id_twitter_id[entity[0]])
             if profile_dict is not None and profile_dict['verified']==True:
                 final_pagerank_result.append((profile_dict['name'],profile_dict['followers_count'],node_id_twitter_id[entity[0]]))
         return final_pagerank_result
```

```
In [16]: final_pagerank_result=filter_pagerank(pagerank_result.copy(),100)
         page_rank_dict["pagerank"]=final_pagerank_result
```

```
In [17]: final_pagerank_result
```

```
Out[17]: [('CNN Breaking News', 61257398, 428333),
          ('ashton kutcher', 17609356, 19058681),
          ('Ellen DeGeneres', 79110794, 15846407),
          ('Barack Obama', 130198925, 813286),
          ('Britney Spears', 56081171, 16409683),
          ('Twitter', 59529056, 783214),
          ('Oprah Winfrey', 43657140, 19397785),
          ('Lance Armstrong', 3222482, 16727535),
          ('Ryan Seacrest', 15379296, 16190898),
          ('SHAQ', 15647369, 17461978),
          ('Al Gore', 3052245, 17220934),
          ('The New York Times', 49806056, 807095),
          ('Demi Moore', 4561994, 19554706),
          ('Coldplay', 23294661, 18863815),
          ('jimmy fallon', 51870114, 15485441),
          ('NPR Politics', 3077586, 5741722),
          ('Stephen Fry', 12632770, 15439395),
          ('Google', 22958396, 20536157)]
```

```
In [ ]: filename = open(path+project_name+"/PickleFiles/communities.pkl","wb")
         pickle.dump(communities_dict, filename)
         filename.close()
```

```
In [18]: filename = open(path+project_name+"/PickleFiles/pagerank.pkl","wb")
pickle.dump(page_rank_dict, filename)
filename.close()

In [19]: for pagerankResult in final_pagerank_result:
profile_pic,username=profile_pic_fetcher(pagerankResult[2])
profile_pic.save(path+project_name+"/PageRankImages/Twitter/{}.png".format(username))
```