Experiment No 7

Title: To implement Graph Based Clustering

Lab Objective: PO5: To demonstrate learning with Clustering.

Theory:

The fundamentals of graph machine learning are connections between entities. As graphs get immensely large, it's imperative to use metrics and algorithms to understand and get graph features. Depending on your context as well, different metrics and algorithms will prove useful and, more importantly, meaningful to your use case.

Furthermore, we can use these metrics as features in a supervised or unsupervised learning task but we have to be careful which we use because they can add as much noise as signal.

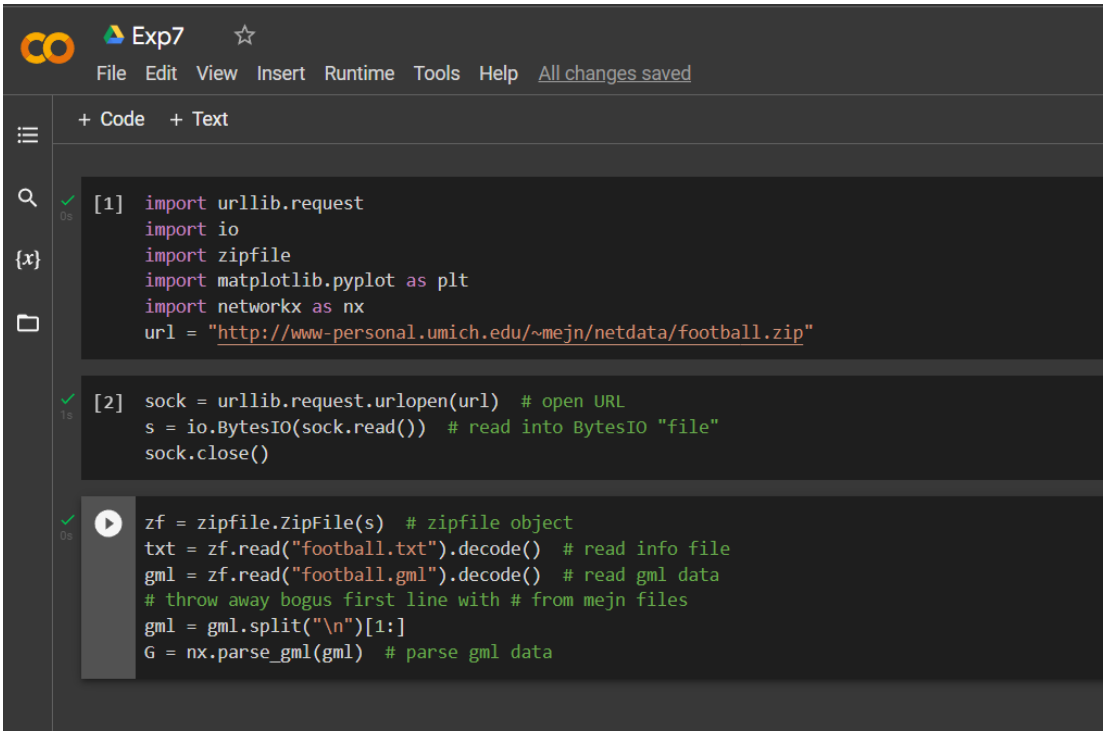
Other than many more metrics and algorithms, the depths of Graph ML cover a wide array of supervised and unsupervised learning tasks. From naïve to advanced techniques, we can use graph structure and inference to go beyond structural data. In future sections I'll cover these machine learning tasks (node, edge, and graph level) on real data.

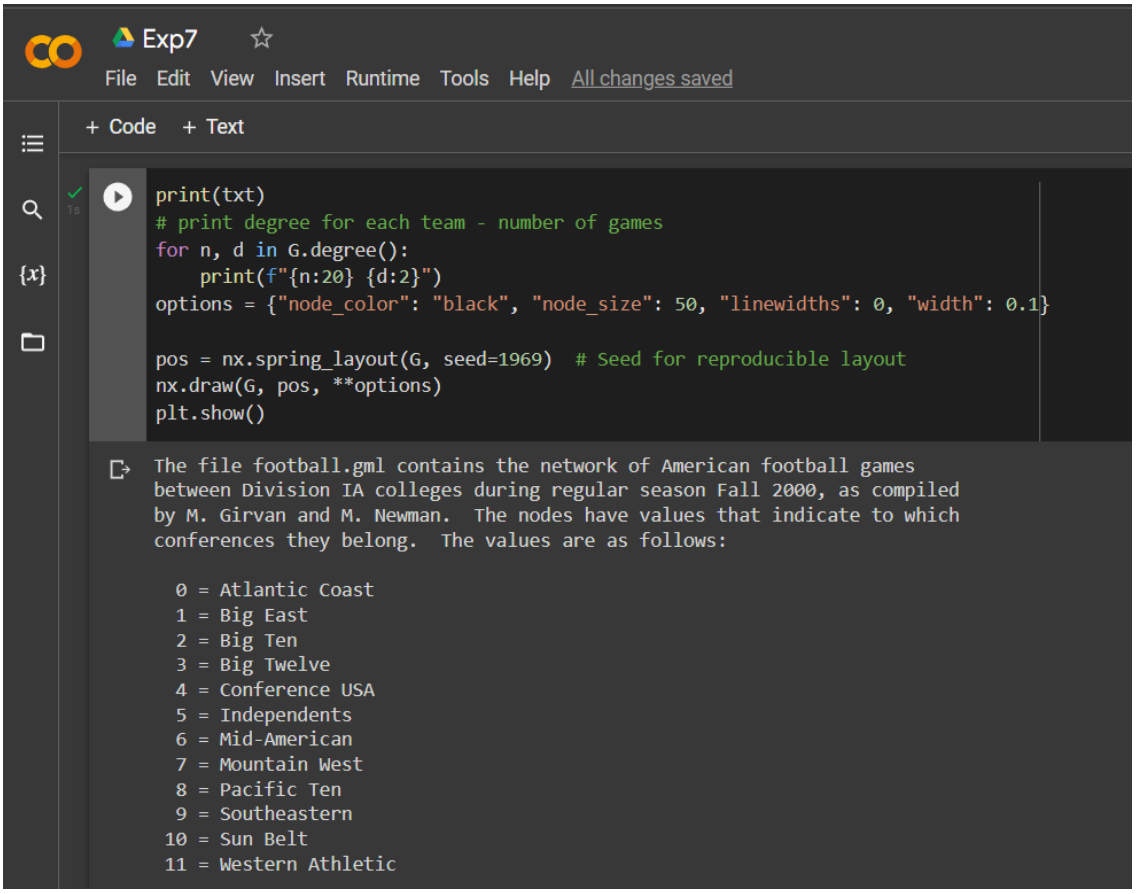
Program Code:

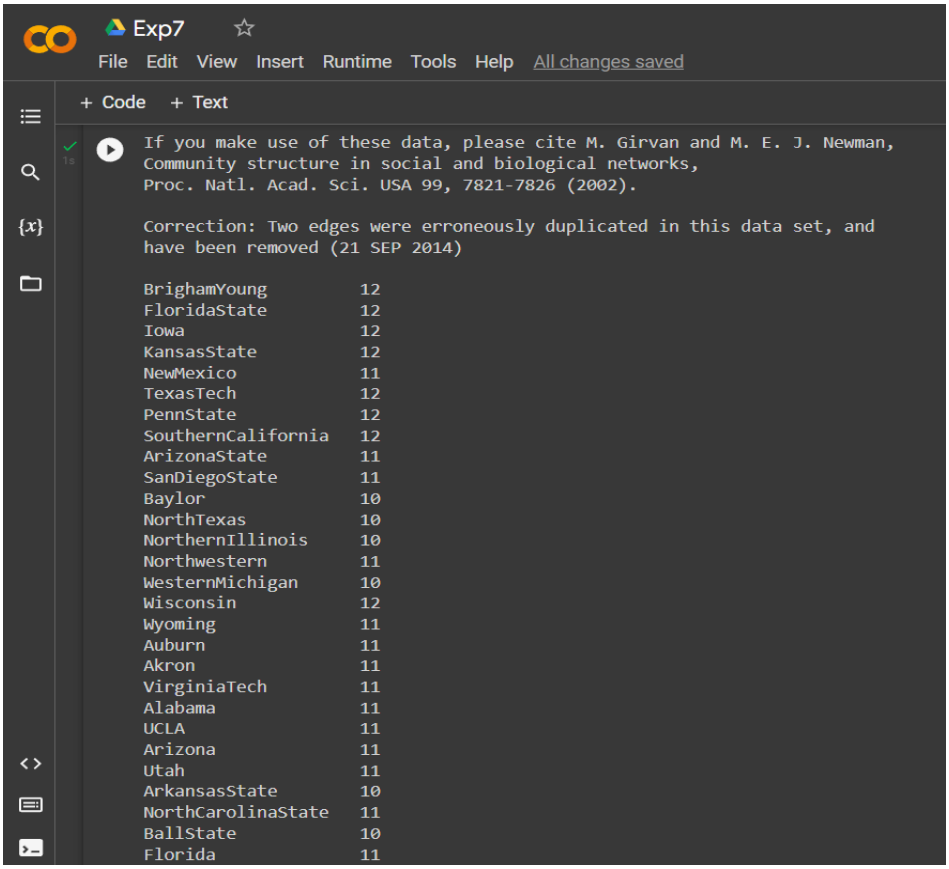
```
"""
========
Football
========
Load football network in GML format and compute some network statistcs.
Shows how to download GML graph in a zipped file, unpack it, and load
into a NetworkX graph.
Requires Internet connection to download the URL
http://www-personal.umich.edu/~mejn/netdata/football.zip
"""
import urllib.request
import io
import zipfile
import matplotlib.pyplot as plt
import networkx as nx
url = "http://www-personal.umich.edu/~mejn/netdata/football.zip"

sock = urllib.request.urlopen(url)  # open URL
s = io.BytesIO(sock.read())  # read into BytesIO "file"
sock.close()

zf = zipfile.ZipFile(s)  # zipfile object
txt = zf.read("football.txt").decode()  # read info file
gml = zf.read("football.gml").decode()  # read gml data
# throw away bogus first line with # from mejn files
gml = gml.split("\n")[1:]
G = nx.parse_gml(gml)  # parse gml data



print(txt)
# print degree for each team - number of games
for n, d in G.degree():
    print(f"{n:20} {d:2}")

options = {"node_color": "black", "node_size": 50, "linewidths": 0, "width": 0.1}
```

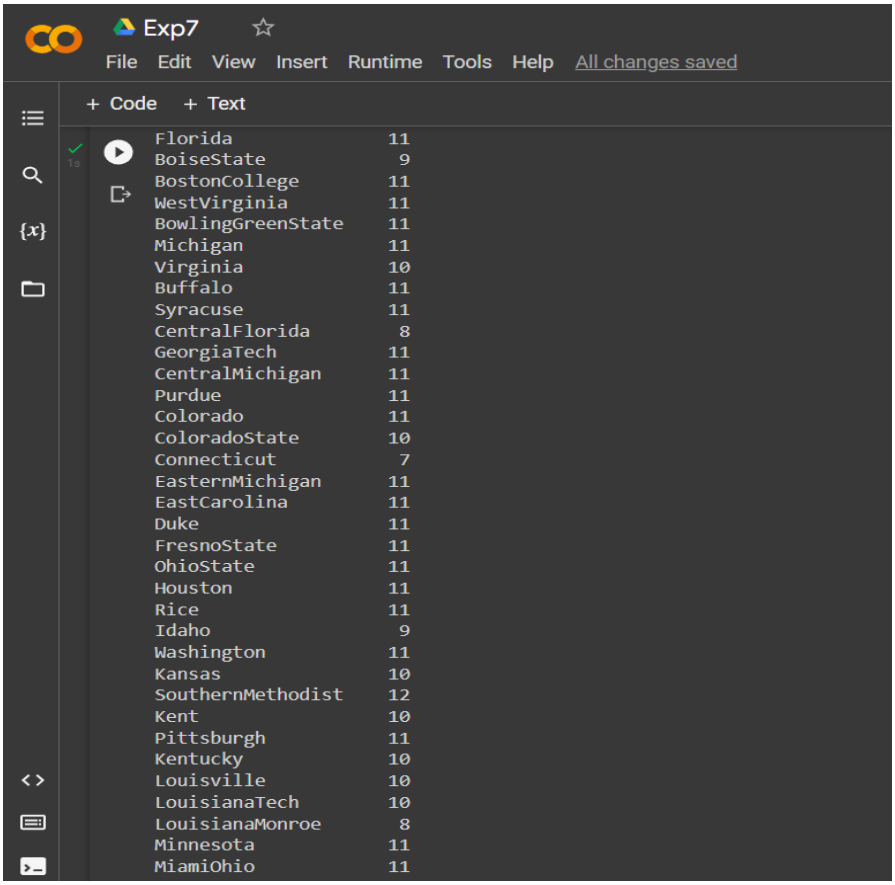
```
pos = nx.spring_layout(G, seed=1969)  # Seed for reproducible layout
nx.draw(G, pos, **options)
plt.show()
```

Implementation:






Exp7
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
If you make use of these data, please cite M. Girvan and M. E. J. Newman,
Community structure in social and biological networks,
Proc. Natl. Acad. Sci. USA 99, 7821-7826 (2002).
Correction: Two edges were erroneously duplicated in this data set, and
have been removed (21 SEP 2014)
BrighamYoung 12
FloridaState 12
Iowa 12
KansasState 12
NewMexico 11
TexasTech 12
PennState 12
SouthernCalifornia 12
ArizonaState 11
SanDiegoState 11
Baylor 10
NorthTexas 10
NorthernIllinois 10
Northwestern 11
WesternMichigan 10
Wisconsin 12
Wyoming 11
Auburn 11
Akron 11
VirginiaTech 11
Alabama 11
UCLA 11
Arizona 11
Utah 11
ArkansasState 10
NorthCarolinaState 11
BallState 10
Florida 11


Exp7
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Florida 11
BoiseState 9
BostonCollege 11
WestVirginia 11
BowlingGreenState 11
Michigan 11
Virginia 10
Buffalo 11
Syracuse 11
CentralFlorida 8
GeorgiaTech 11
CentralMichigan 11
Purdue 11
Colorado 11
ColoradoState 10
Connecticut 7
EasternMichigan 11
EastCarolina 11
Duke 11
FresnoState 11
OhioState 11
Houston 11
Rice 11
Idaho 9
Washington 11
Kansas 10
SouthernMethodist 12
Kent 10
Pittsburgh 11
Kentucky 10
Louisville 10
LouisianaTech 10
LouisianaMonroe 8
Minnesota 11
MiamiOhio 11

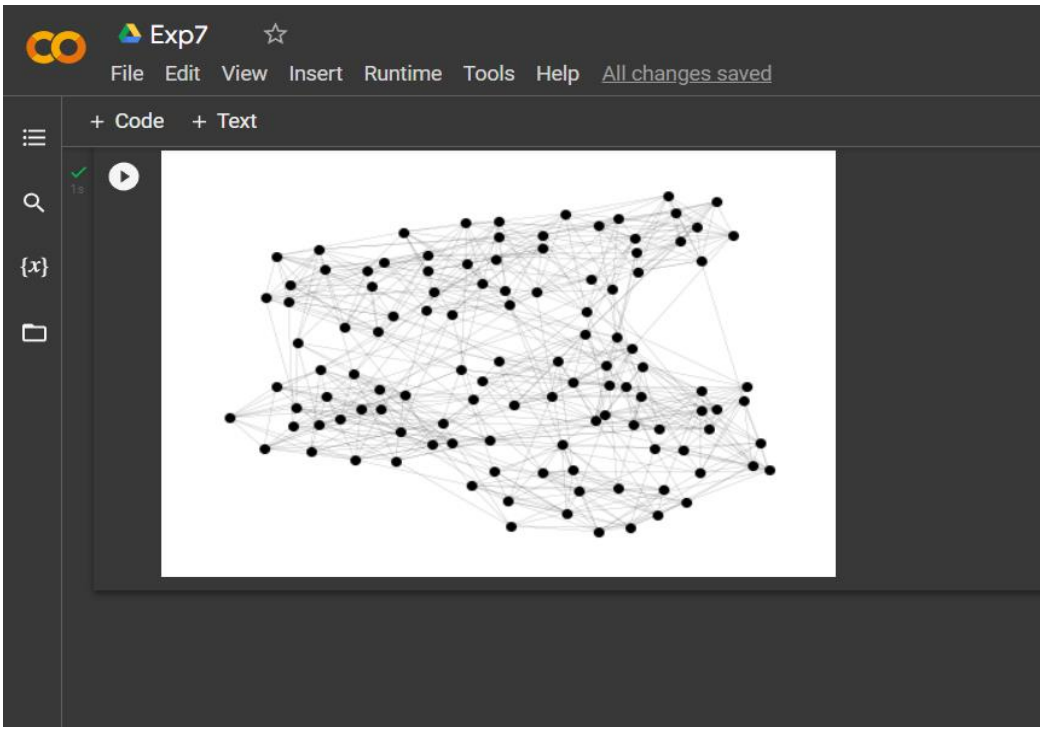
Exp7

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
Temple                11
Navy                  11
TexasA&M              11
NotreDame             11
TexasElPaso           11
Oklahoma              11
Toledo                 9
Tulane                11
Mississippi           11
Tulsa                 12
NorthCarolina         11
UtahState              9
Army                  11
Cincinnati            11
AirForce              10
Rutgers               10
Georgia               10
LouisianaState        10
LouisianaLafayette     8
Texas                 11
Marshall              10
MichiganState         11
MiamiFlorida          10
Missouri              10
Clemson               10
NevadaLasVegas        12
WakeForest            10
Indiana               11
OklahomaState         10
OregonState           10
Maryland              11
TexasChristian        11
California            11
AlabamaBirmingham     10
Arkansas              10
Hawaii                11
```

Output:



Conclusion:

Thus, in this practical we studied performed Graph Based Clustering using python programming language.