SOCIAL NETWORK ANALYSIS USING PYTHON AND NETWORKX

[1]shivam shah, [2] Sumitra Menaria
[1]student B.TECH 4TH year, [2] Assistant Professor,
[1]Department Of Information Technology,
[1] Parul institute of engineering and technology(parul university),vadodara,india

Abstract : social media have received more attention now days .social site are the most visiting sites now days it contain large amount of data about user. Social network analysis is process of analyzing the social structure through the network and graph theory .it means all the relationship between the social entities are represented in form of graph. In this research paper we will study about the how the graph is represented and it tells about user. For this research paper we are using the python language and for representation of graph we are using the open source library called networkx.and the data I am using the twitter I am using the real time data from twitter that will get using the tweepy which allow to extract information twitter in real time and allow to store in json formet and we can perform analysis in that json formet. Social network analysis is now days shown as key to the morden sociology.

***IndexTerms* - python, graph theory,network,sociology,social network analysis,tweepy,json.**

I. INTRODUCTION

Social network analysis is also call as SNA is a process of analyzing the social structure through use of network and a graph theory. the graph characterize the network structure in terms of node and edges .node means individual actor and edge or link means the relationship between each actor example of social network analysis includes social media is meme spread, information circulation,collobration graph. These types of network are visualized as graph in which the nodes are represented as points and ties are represented as ties.

Visual representation of social network is important to understand about the network data and convey the result of analysis. There are numbers of methods for visualization data produced by social network analysis presented. many of social network analysis software have modules for network visualization exploration of data is done through displaying nodes and ties in various layouts and contributing colors, size and various advance properties of node visual representation of the network may be powerful method for conveying complex information but we should take care in interpreting node and graph properties in visual display alone. As they may misprint structure captured through properties analysis

Especially when using the social network analysis for facilitative change different approaches of participatory network mapping have may become useful here participants provide network data by actually mapping out the network which also include collection of some actors and its relationship.

In this research paper we well give information about how we can represent the social media data in form of graph using the python programming language and module name network x.

II. RELATED WORK

For the batter understanding network library and forming the all data in graph we have to understand some concept of graph theory.

Graph theory concepts :

- Graph: graph is G=(V,E) collection of edges and vertices where v is collection of vertices and e is collection of edges
- Digraph: it is also collection of D(V,A) where v is collection of vertices and a is collection of arcs a is made of pairs of element from v

Distributions:

- Bridge: an individual whose weak ties fill a structural hole provide ;inks between two indivisuals or clusterit will also includes shortest route when longer one is unfeasible due to high risk of message delivery faliur.
- Centrality: centrality refers to matrices that aim to qualify the influence of a particular node with a particular network.
- Density: the proportion of direct ties in network relative to total no of possible.
- Distance :the minimum no of ties required to connect two network.
- Structural holes :the absence of ties between two part of network.
- Tie strength: defined by the linear combination of time, intensity, intimacy and reciprocity.

Segmentation:

- Clustering coefficient: a measure of likelihood that two associate of nodes are associate a higher clustering coefficient indicates a grater cliquishness.
- Cohesion: the degree to which actor are connected directly to each other by cohesive bonds.

Social network potential is also useful it is nothing but the numeric coefficient derived through algorithm to represent the size of both social network and ability to influence the network.

The snp coefficient have two primary function:

- The classification based on social network potential
- Weighting of respondents in quontative market research

III. GATEING FAMILIER WITH NETWORKX AND GRAPH IN PYTHON

3.1. INTRODUCTION TO NETWORK X AND ITS FUNCTION

Networkx is open source python library for studying graph and network.Networkx is a python package for creation, manipulation and study of structure, dynamics and function of complex network

For installation of network library in python command is pip install network

program :

import networkx as nx #import network librerery in python G=nx.graph()# create the graph and now graph is empty

#add node G.add_node(1)
G.add_nodes_from([2,3])#we can also list node of argument

#add edges G.add_edges(1,2)

E=(2,3)

G.add_edges(e)

There are multiple graph function and its operation this table will give information about what are the function and its operations.

Function	Function opretation
Subgraph(n,gbunch)	Induced sub graph view of g nodes in nbunches
union(g1,g2)	Union of two graph
Disjoint_union(g1,g2)	Graph union assuming all nodes are different
Crastesian_product(g1,g2)	Returns crastriean products of graph
COMPOSE(G1,G2)	Combine the graph identifying nodes comman to both
Complement(g)	Graph complement
Create_empty_copy(g)	Create empty copy of same graph
Convert_to_directed(g)	Returns directed graph of g
Convert_to_undirected(g)	Convert undirected graph of g

Table 3. 1.networkx function and its operation

In this table we weil show information about what type of graph is access by which network class.

Graph type	Networkx class
Undirected simple	Graph
Directed simple	Digraph
With self loop	Graph,digraph
With parallel edges	Multigraph,multipledigraph

Table 3.2.graph types and its networkx class

3.2 . APPLY NETWORKX ON DATASET AND IDENTIFY RESULT

In this we will take generic dataset and we will do some manipulation so it can be some insighted into graph in form of edge list .and edge list is list of Tuple contain vertices defining every edge .

The dataset we will be looking at come from airlines industry .it has some basic information about airlines routes. There is a source of journey and destination there are also some column indicating arrival and departpture time for each journey. As you can imagine dataset lends itself beautifully to be analyzed as graph. Imagine a few cities are connected by routes.

Program:

```
Import pandas as pd
    Import numpy

Data= pd.read_csv("airlines.csv")
data['date'] = pd.to_datetime(data[['year', 'month', 'day']])
data = data.drop(columns = ['year', 'month', 'day'])
FG = nx.from_pandas_edgelist(data, source='origin', target='dest', edge_attr=True)
Fg.nodes()
Fg.edges()
nx.draw_networkx(FG, with_labels=True)
```

output:

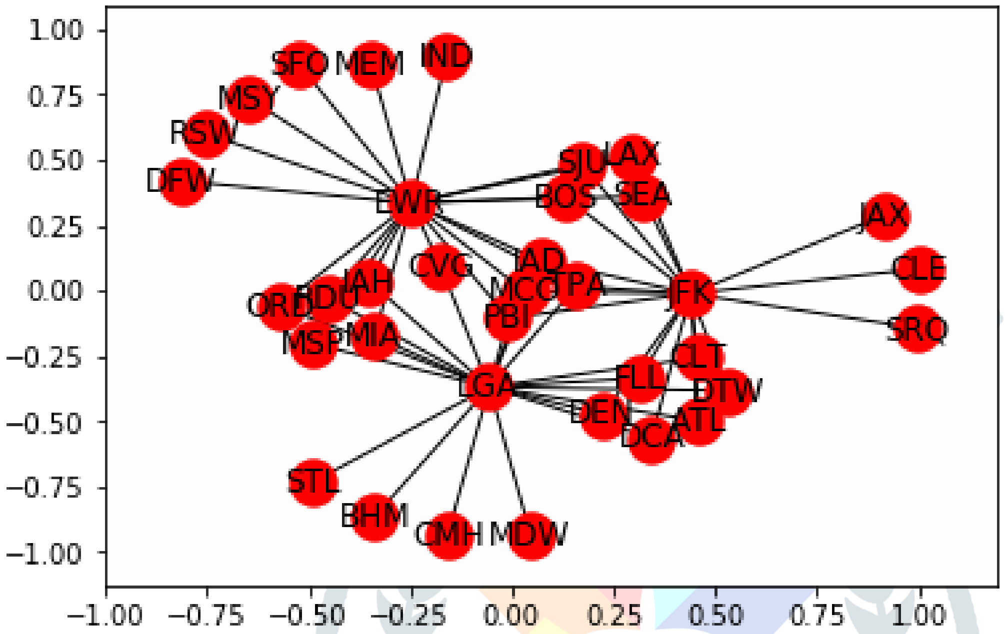

Fig 1.NETWORKX GRAPH

IV. APPLICATION

- **Marketing Analytics –** Graphs can be used to figure out the most influential peoples in a Social Network. Advertisers and Marketers can estimate the biggest bang for the marketing buck by routing their message through the most influential people in a Social Network.
- **Banking Transactions –** Graphs can be used for find unusual patterns helping in mitigating fraudulent transactions. There are examples where Terrorist activity has been detected by analyzing the flow of money across interconnected Banking networks.
- **Supply Chain –** Graphs help in identifying optimum routes for your delivery trucks and in identifying locations for warehouses and delivery centers.
- **Pharma –** Pharmacy companies can optimize the routes of the salesman using Graph theory. It helps in cutting costs and reducing the travel time for salesman.
- **Telecom –** Telecom companies typically use Graphs to understand the quantity and location of Cell towers to ensure maximum coverage.

V. CONCLUSION:

From the research paper we will conclude about how to use open source networkx library for network analysis as it will be valuable tool set for any data scientist .network analysis help us to solve common data science problem and vizulizating them at much greater scale .by social network analysis allow us to know and help us to know how different peoples are related to each other like message spread, meme spread, information circulation,. It also allow us to know who is the central edge allow us to show all the information in form of graph .

REFERENCES

[1] Stanford WebBase Project.http://www-diglib.stanford.edu/~testbed/doc2/WebBase.

[2] L. A. Adamic. The Small World Web. InProceedingsof the Third European Conference on Research andAdvanced Technology for Digital Libraries (ECDL'99),Paris, France, Sep 1999.

[3] L. A. Adamic, O. Buyukkokten, and E. Adar. A socialnetwork caught in the Web.First Monday, 8(6), 2003.

[4] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong.Analysis of Topological Characteristics of Huge OnlineSocial Networking Services. InProceedings of the 16thinternational conference on World Wide Web(WWW'07), Banff, Canada, May 2007.

[5] R. Albert, H. Jeong, and A.-L. B´arab´asi. TheDiameter of the World Wide Web.Nature, 401:130,1999

[6] L. A. N. Amaral, A. Scala, M. Barth´el´emy, and H. E.Stanley. Classes of small-world networks.Proceedingsof the National Academy of Sciences (PNAS),97:11149–11152, 2000.

[7] A. Awan, R. A. Ferreira, S. Jagannathan, andA. Grama. Distributed uniform sampling in real-worldnetworks. Technical Report CSD- TR-04-029, PurdueUniversity, 2004.

[8] L. Backstrom, D. Huttenlocher, J. Kleinberg, andX. Lan. Group Formation in Large Social Networks:Membership, Growth, and Evolution. InProceedingsof the 12th ACM SIGKDD International Conferenceon Knowledge Discovery and Data Mining (KDD'06),Philadelphia, PA, Aug 2006.

[9] A.-L. B´arab´asi and R. Albert. Emergence of Scaling inRandom Networks.Science, 286:509–512, 1999

[10] L. Becchetti, C. Castillo, D. Donato, and A. Fazzone.A Comparison ofSampling Techniques for WebGraph Characterization. InProceedings of theWorkshop on Link Analysis (LinkKDD'06),Philadelphia, PA, Aug 2006.

[11] V. Braitenberg and A. Sch¨uz.Anatomy of a Cortex:Statistics and Geometry. Springer-Verlag, Berlin, 1991.

[12] A. Broder, R. Kumar, F. Maghoul, P. Raghavan,S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener.Graph Structure in the Web: Experiments and Models. InProceedings of the 9th International WorldWide Web Conference (WWW'00), Amsterdam, May2000.

[13] A. Clauset, C. R. Shalizi, and M. E. J. Newman.Power-law distributions in empirical data, Jun 2007.http://arxiv.org/abs/0706.1062v1

[14] d. boyd. Friends, Friendsters, and Top 8: Writingcommunity into being on social network sites.FirstMonday, 11(12),2006.

[15] P. Erd¨os and A. R´enyi. On Random Graphs I.Publicationes Mathematicae Debrecen, 5:290–297,1959.

[16] M. Faloutsos, P. Faloutsos, and C. Faloutsos. OnPower-Law Relationships of the Internet Topology. InProceedings of the Annual Conference of the ACMSpecial Interest Group on Data Communication(SIGCOMM'99), Cambridge, MA, Aug 1999.

[17] S. Garriss, M. Kaminsky, M. J. Freedman, B. Karp,D. Mazi`eres, and H. Yu.Re: Reliable Email. InProceedings of the 3rd Symposium on NetworkedSystems Design and Implementation (NSDI'06), SanJose, CA, May 2006.

[18] M. Girvan and M. E. J. Newman. Communitystructure in social and biological networks.Proceedings of the National Academy of Sciences(PNAS), 99:7821–7826, 2002.

[19] Google Co-op.http://www.google.com/coop/.

[20] M. Granovetter. The Strength of Weak Ties.American Journal of Sociology, 78(6), 1973

.[21] J. Kleinberg. Authoritative Sources in a HyperlinkedEnvironment.Journal of the ACM, 46:604–632, 1999.

[22] J. Kleinberg. Navigation in a Small World.Nature,406:845–845, 2000.

[23] J. Kleinberg. The Small-World Phenomenon: AnAlgorithmic Perspective. InProceedings of the 32ndACM Symposium on Theory of Computing(STOC'00), Portland, OR, May 2000.

[24] J. Kleinberg and S. Lawrence. The Structure of theWeb.Science, 294:1849–1850, 2001.

[25] J. M. Kleinberg and R. Rubinfeld. Short paths inexpander graphs. InIEEE Symposium on Foundationsof Computer Science (FOCS'96), Burlington, VT, Oct1996.

[26] R. Kumar, J. Novak, and A. Tomkins. Structure andEvolution of Online Social Networks. InProceedings ofthe 12th ACM SIGKDD International Conference onKnowledge Discovery and Data Mining (KDD'06),Philadelphia, PA, Aug 2006.

[27] R. Kumar, P. Raghavan, S. Rajagopalan, andA. Tomkins. Trawling the Web for EmergingCyber-Communities.Computer Networks,31:1481–1493, 1999.

[28] S. Lee, R. Sherwood, and B. Bhattacharjee.Cooperative peer groups in NICE. InProceedings ofthe Conference on Computer Communications(INFOCOM'03), San Francisco, CA, Mar 2003.

[29] S. H. Lee, P.-J. Kim, and H. Jeong. Statisticalproperties of sampled networks.Physical Review E,73, 2006.

[30] L. Li and D. Alderson. Diversity of graphs with highlyvariable connectivity.Physics Review E, 75, 2007.

[31] L. Li, D. Alderson, J. C. Doyle, and W. Willinger.Towards a Theory of Scale-Free Graphs: Definitions,Properties, and Implications.Internet Mathematics,2(4):431–523, 2006.

[32] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan,and A. Tomkins. Geographic Routing in SocialNetworks.Proceedings of the National Academy ofSciences (PNAS), 102(33):11623–11628, 2005.