Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol.6, No.1, June 2018, pp. 215~228
Available online at: http://pen.ius.edu.ba

Data Science: Identifying influencers in Social Networks

Srikanth Bethu[1], V Sowmya[2], B Sankara Babu[3], G Charles Babu[4], Y.Jeevan Nagendra Kumar[5]
[1,2,3]Departement of Computer Science and Engineering, GRIET, JNTU Hyderabad, Telangana, India
[4]Departement of Computer Science and Engineering, Mallareddy Engineering College, JNTU Hyderabad, Telangana, India
[5]Departement of Information Technology, GRIET, JNTU Hyderabad, Telangana, India
srikanthbethu@gmail.com, sowmyaakiran@gmail.com, bsankarababu81@gmail.com, charlesbabu26@gmail.com,
jeevannagendra@gmail.com

Article Info

Article history:

Received Jun 12th, 201x
Revised Aug 20th, 201x
Accepted Aug 26th, 201x

Keyword:

Data Analysis and Mining
Data Science
Online Social Networks
Network communication

ABSTRACT

Data science is a "concept to unify statistics, data analysis and their related methods" in order to "understand and analyze actual phenomena" with data. The common use of Online Social Networks (OSN)[2] for networking communication which authorizes real-time multimedia capturing and sharing, have led to enormous amounts of user-generated content in online, and made publicly available for analysis and mining. The efforts have been made for more privacy awareness to protect personal data against privacy threats. The principal idea in designing different marketing strategies is to identify the influencers in the network communication. The individuals influential induce “word-of-mouth” that effects in the network are responsible for causing particular action of influence that convinces their peers (followers) to perform a similar action in buying a product. Targeting these influencers usually leads to a vast spread of the information across the network. Hence it is important to identify such individuals in a network, we use centrality measures to identify assign an influence score to each user. The user with higher score is considered as a better influencer.

Corresponding Author:

First Author,
Departement of Computer Science and Engineering,
National Chung Cheng University,
GRIET, JNTU Hyderabad, Telangana 500090, India.
Email: srikanthbethu@gmail.com

1. Introduction

Now a day’s Social Networks plays a communication media in real time for the user’s interaction. They are used to share all the experiences and their personal valid opinions on various topics like news, politics, celebrities, sports, events and products. In this way online social network has become important resource for knowledge sharing and knowing. For brand communications like Fashion industry, it exhibits high potential in digital marketing for integral growth. Now it has become brand ambassador for its messages and promotions to produce awareness among audience through continuous brand advertisement activities. The existing relations in a social network are as follows:

- Similarities depending on demographic characteristics, locations or group memberships attributes of any two nodes.
- The interaction relationships like speaking; chatting refers to continuous exchange of information between all the actors or users.

Social network analysis (SNA)[1] maps the interconnectedness between actors in a network through mathematics that aims is to understand the structural relations and to explain both why their occurrence and consequences are. To know how influential an author is, in a network (Twitter in this case), and to assign score to authors in the network based on the relevancy of posts using centrality measure.
Influencer measuring on social network by conceptual method differs each one from others. The influence is nothing but who spreads the information and influence the people. Influence in through word-of-mouth [16] marketing can be used in:

- ❖ Public influence in the flow of mass communication.
- ❖ Helping business in product development by using market shares.
- ❖ Improvement of broad awareness innovation.

Centrality: It means there is no unanimity in measuring the market and its network progress. The below Fig.1 shows the working procedure ofcentrality [11][25].

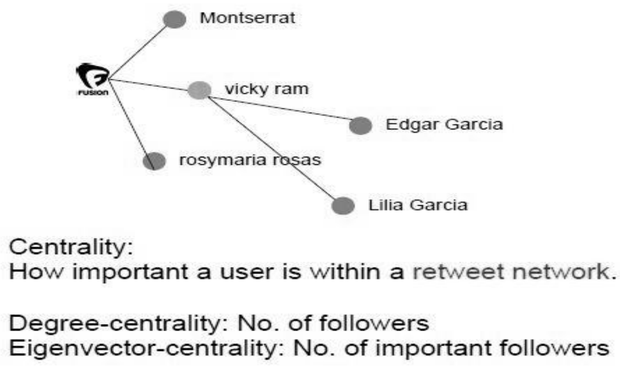


Figure 1. Centrality

The centrality explains about how to measure and quantify the "structural features" of a particular single node in a given graph, and finds an actor who plays itcentrally in the graph. The centrality score is high for a given vertex *i*, then it gives information of about hub that has more contacts and nodes. The formula can be represented in both undirected and unweighted graph [14] 'G', is given as

$$\sigma_D(i) = \sum_{j}^{N} a_{ij}$$

Here *i* is focal node value in the network,
j gives total number of other nodes in a network,
N is total number of nodes in a network,
a = adjacency matrix
if a(i,j)==1,then node *i* is connected to node *j*.
if a(i,j)==0,then node *i* is not connected to node *j*.

Mathematically, the simplest computation of closeness centrality σC can be represented as follows:

$$\sigma_C = \frac{1}{\sum_j^n d_G(i,j)}$$

where

dG(i, j) is the number of links in the geodesic distances from node i to node j.

Eigen vector centrality (EC) when compared to Direct Centrality (DC), takes into account the number of direct links and indirect contacts in the network. Eigen vector *X(i,G)* is given as follows:

$$X(i,j) = \sum_{j \in N(i)} X(i,G)$$

Any social network like Twitter, Facebook, Watsapp markets business since this sites provide real-time data for business insiders. The interconnectedness in Fig.2 shows relationships between actors to show business how important it is to find more influencers and understand their requirements.

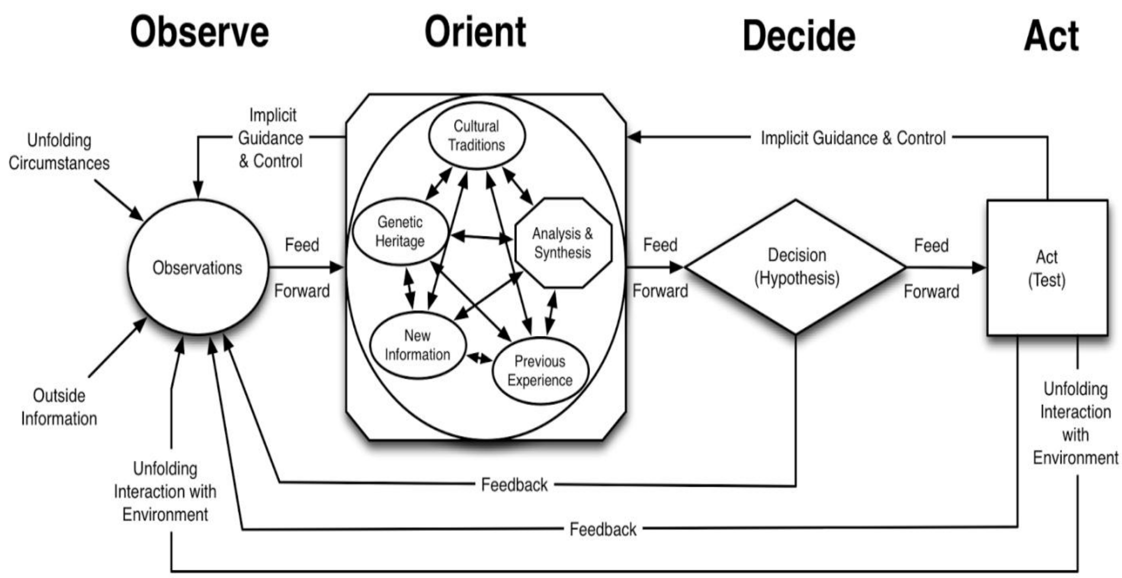


Fig .2. Centrality

From the above fig.2, we can understand the performance and evaluation of centrality depends on Observation, Orientation, Decidability and Actions.

2. Literature Survey

2.1. Twitter Data Analysis

We have observed that being a follower and account holder, user automatically receives messages posted by his/her followed accounts. In this paper we have analysed the followers and their relationships based on the total numbers of followers towards a node and its friends lie on outward degree. Comparing with other social network sites, the reciprocity is not necessary and everyone can follow any other, and no objections in it.

Twitter is viewed as pyramidal type structure because some influence accounts like movie stars, journalists, celebrities, sports personalities, will have millions of followers without any obligation. Whereas Face book is viewed as circular type structure where friends are reciprocal. In twitter '@' is used to re-tweet and '#' is used to follow.

2.2. Related Work

By using different metrics that are related to centrality measures has addressed the influencers and diffusion process considers only digraphs.

There are two types of networks that exist to determine the network link structure.

- The network is established for friendship between existing two nodes, if at least one follows the other.

- Interaction network is directed between two nodes for replying and re-tweet.

Re-tweets can be used to reinforce a message. Not surprising, mentioned users were mostly celebrities.

2.3. Over view of Social Network Analysis Technique

To identify influencers on social network sites like twitter, we have described in step wise extraction method is used to map users network.
Fig .3, shows the following steps.

a. A list of 1000 top most tweets related to fashion technology information is selected.
b. We select 80 from 100 influencing users based on the re-tweeted information, after that again we add another most influencing 80 users to gather opinion. Repeat the process until it reaches a 100 accounts after some iteration process.
c. Afterwards, the data frame is build and is translated into a CSV file.

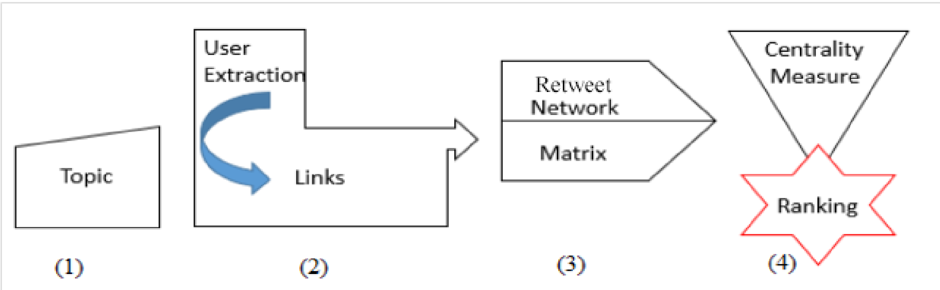


Figure 3. Data Gathering

a) Identify a list of 10 influential fashion authors
b) Extract all the users (retweeted) in the list
c) Identity the 80 most influencing users
d) Go back to 2 step to retrive information on the 80 new IDs.

Build a dataset and store as a CSV file with the obtained list.

3. Architectural Design for Social Network Analysis

3.1. Architectural Design

Architecture diagram(Fig 5) explains how data is importing directly to phython data frame using twitter streaming API to system database. Mining of data is done by phython regular expression for analysing the results and storing in a phython data frame. This architecture shows the proceeding of data in stepwise manner. Finally result is stored in CSV (comma seperated variable) file format.

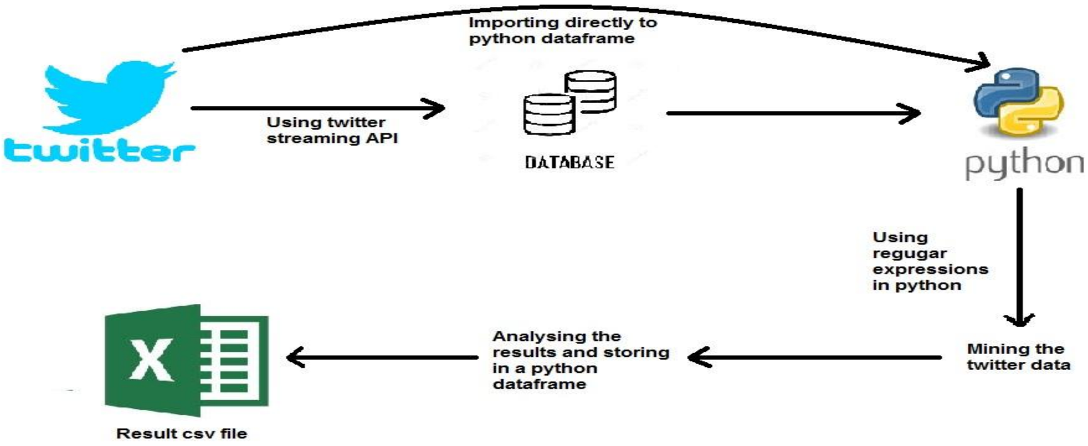


Figure 4. Architectual Design

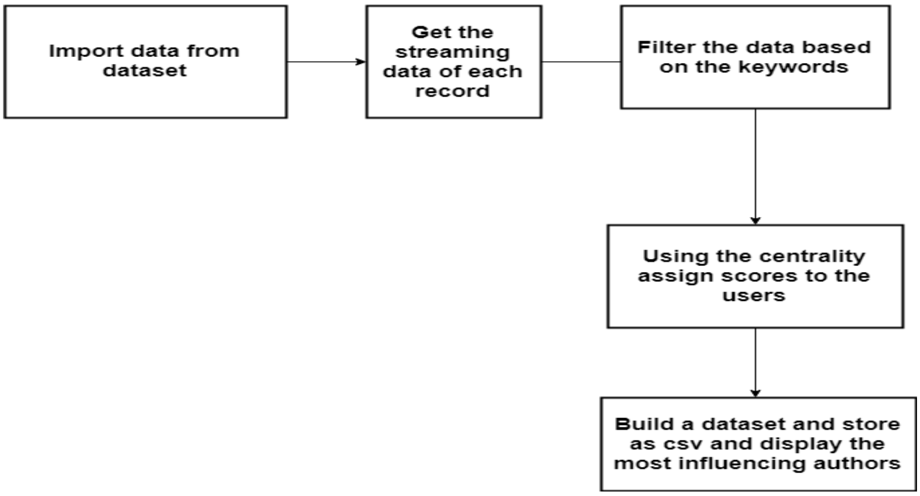


Figure 5. Data Processing

Fig.4 and Fig.5, gives the information of analysis of social data and its data processing steps for analysis.The dataset we used have a good number of features but we mostly focused on tweet_id and tweet_text (fig 6).

t_id	t_retweet	t_text
6.15E+17	1204	People tweet #FreeBree in support of the black woman who removed S
6.19E+17	1010	We made a Chrome extension to add actual Donald Trump quotes to ev
6.18E+17	444	What's this dude's name?
6.23E+17	340	A 6-year-old totally owned the Financial Times over a Minecraft error h
6.20E+17	298	Inside the moaning, dripping world of Minion porn http://t.co/dYSfCSh
6.16E+17	226	WATCH: How black women experience police violence.
6.14E+17	214	More Americans have been killed by white supremacists than Muslim e
6.18E+17	178	It seems Sepp Blatter found a way to watch #FIFAWWC final while still a
6.13E+17	121	WATCH: Science says helping others makes humans happiest.
6.15E+17	120	Haiti just won a hockey world championship http://t.co/PwC4FmOoaF I
6.16E+17	118	There are now more Spanish speakers in the U.S. than in Spain http://t.
6.18E+17	114	Nice to see U.S. elected officials have made the trek up to Canada, but J
6.21E+17	113	Former Mexican official says leaked intel may have caused #ElChapo's
6.14E+17	109	Trans woman interrupts Obama at White House LGBT reception http://t

Figure 6. Input Data set

3.2. Implementation

To reduce the interaction interfacing we use API call load data to retrieve all the tweet IDs of the given list of 100 tweets based on the fashion industry, and stored them in a variable called t (dataframe). Hence we used the function *retweet_users_of_a_tweet()* with tweet_id as argument. We stored this result in a list and passed it to *t_user_rank()*. The *t_user_rank()* function will return a dictionary of user objects.

```
def retweet_users_of_a_tweet(tweet_id):
    retweets = api.retweets(tweet_id, 100)
    return [rt.user.id for rt in retweets]
```

```
udic = t_user_rank(retweet_users_of_a_tweet(t.t_id[i])) #
follower = [udic.values()[x][0] for x in range(len(udic))]
mention = [udic.values()[x][1] for x in range(len(udic))]
score = [udic.values()[x][2] for x in range(len(udic))]
keys = udic.keys()
t_id = [t.t_id[i] for x in range(len(udic))]
```

The function *t_all_tweets()* returns all the tweets posted by the user. It takes two arguments : userid and number of pages. The returned object is passed to *t_mentions()* which will parse the tweets and finds the mention of the user(mention is count of the keywords we are interested in) and returns an integer value. The method *t_user_rank()* assigns rank to each user in the list created and frames a dictionary object.

```
for user in users:
    screen_name = api.get_user(id=user).screen_name
    follower = api.get_user(id=user).followers_count
    mention = t_mentions(user)
    udic[screen_name] = [follower, mention, (follower*mention)]
```

```
def t_mentions(user):
    tweets = t_all_tweets(user, 2) # first 2 pag
    t_text = ''
    for t in tweets:
        t_text += t.text
    return len(re.findall('(@Fasion)', t_text))
```

Implementation process has following steps

a) Set the code to twitter limitations of available GET requests.
b) Rate the every limit to 15 mints of time.
c) To avoid error messages and solve this problem, divide list 'i' in 2 and by using time-sleep( ) function with 60 as argument for 90 seconds.
d) For estimated system block after 5 GET requests, apply rule as for each 'i' if the 'i mod 5' is equal to zero, then use time-sleep( ) function.
e) The result set is obtained from each iteration concatenation of list 1 to n.

Construct the dataset:

After the result set is obtained the result is displayed one the screen and the same is copied and stored as a csv file for sharing.

Provide sufficient detail to allow the work to be reproduced. Methods already published should be indicated by a reference: only relevant modifications should be described.

4. Results Analysis

The Spyder IDE: The program is written in python 2.7 in spyder Ide.

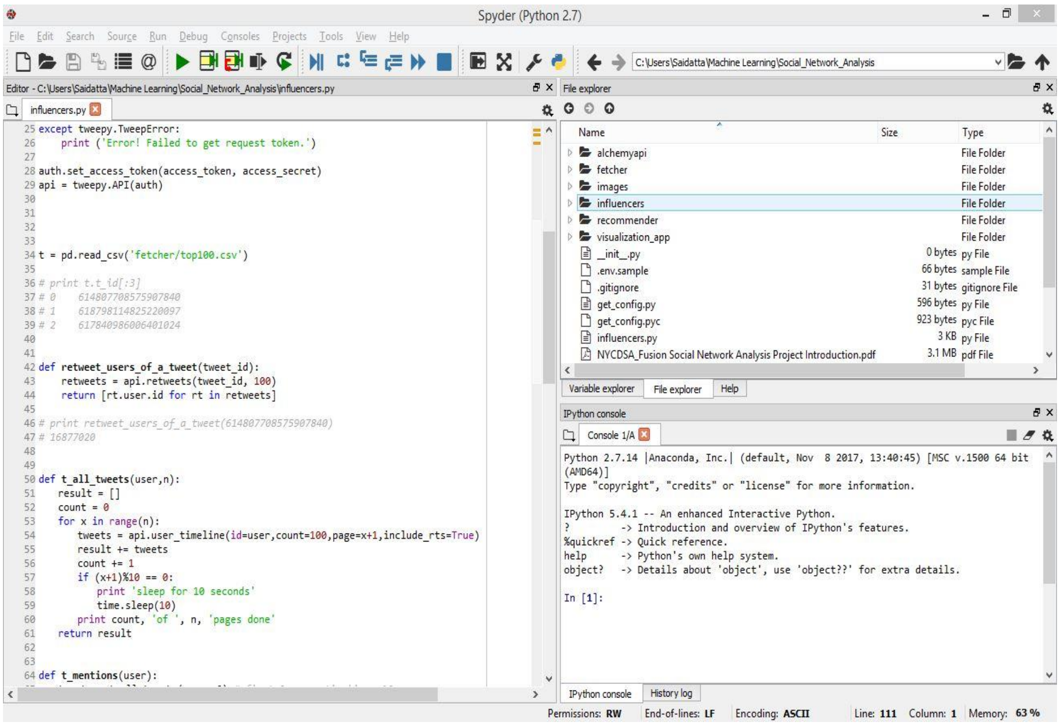


Figure 7. Spyder IDE

IPython Console: Each page analysed is requested by the program from Twiter API. The current image shows the execution of a tweet.

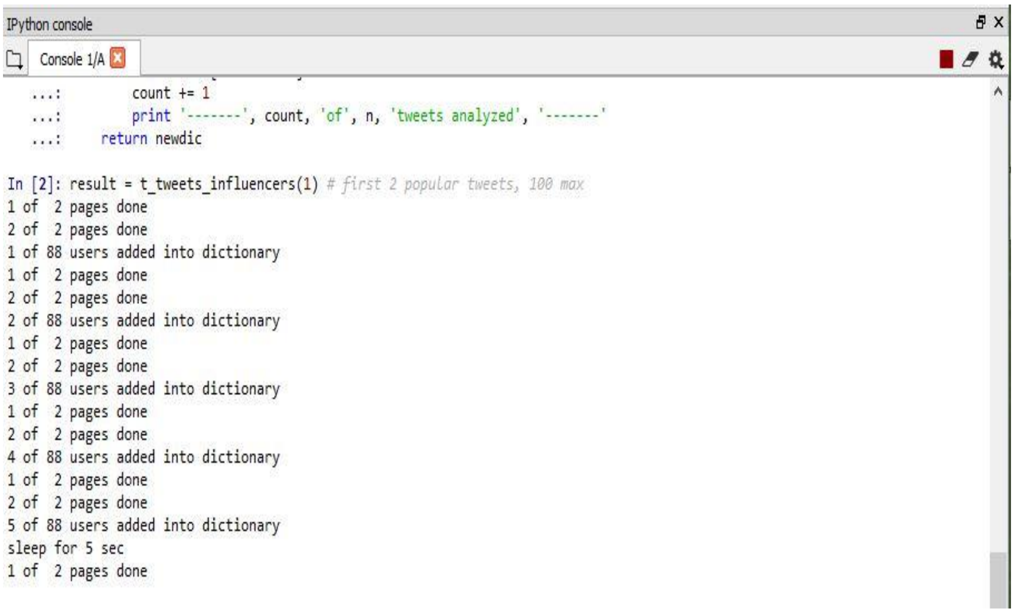


Figure 8. IPython Console

Execution in Progress: The execution is under process and the list of influencers is being added to dictionary.

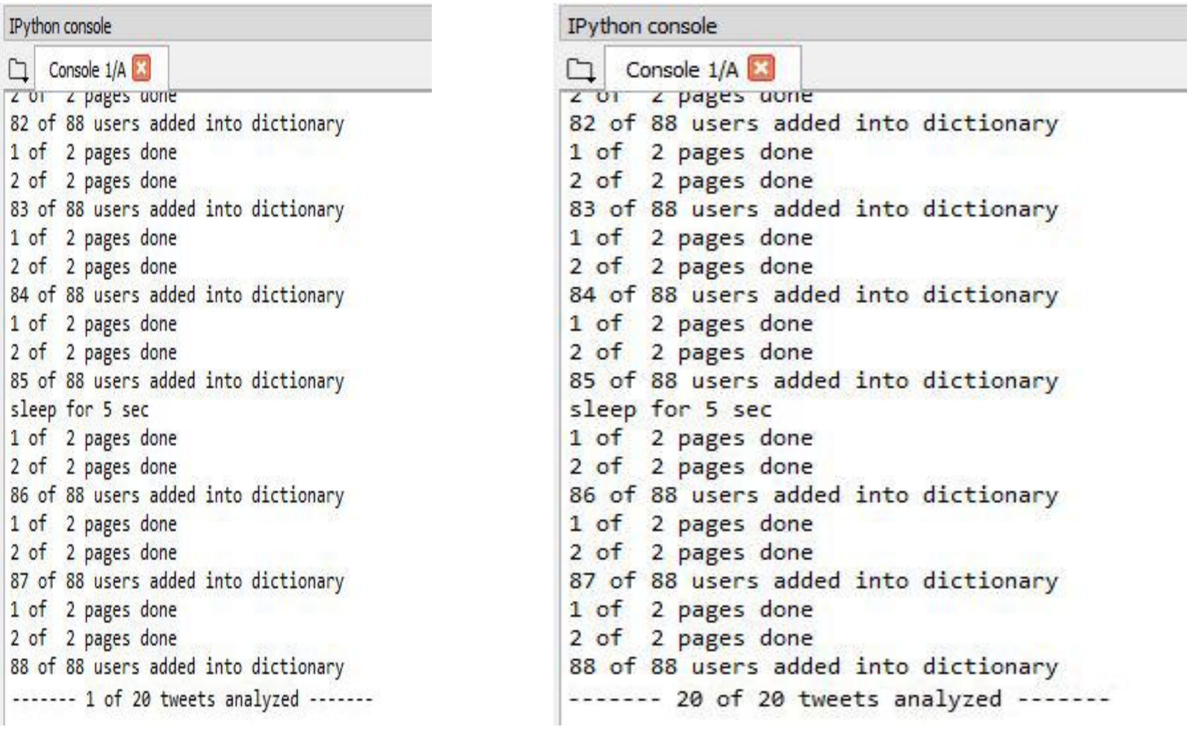


Figure 9. Execution and Tweets analyzation

Execution Completion: All the users are assigned scores and dataframe object is created.

Input dataset: This is the input feed to the program.

t - DataFrame

Index	t_date	t_favorites	t_hashtags	t_id	t_mentions	t_retweets	t_text	t_url	w_authors	w_date	w_genre	
0	2015-06-27 14:49:14	814	FreeBree	614807708575…	nan	1204	People tweet #FreeBree in…	http://t.co/ n2DtA16cRn	"John Walker"	"2015-06-27T…	story	157
1	2015-07-08 15:05:41	785	nan	618798114825…	nan	1010	We made a Chrome exten…	http://t.co/ HIaF9dt6RG	"Patrick Hogan"	"2015-07-08T…	story	162
2	2015-07-05 23:42:24	615	USA	617840986006…	nan	444	What's this dude's name?…	nan	nan	nan	nan	nan
3	2015-07-18 20:55:24	351	nan	622509999277…	nan	340	A 6-year-old totally owne…	http://t.co/ dx3Px9vrbd	"Kevin Roose"	"2015-07-12T…	story	165
4	2015-07-11 17:55:35	197	nan	619928034959…	nan	298	Inside the moaning, dri…	http://t.co/ dYSfCShSXa	"Charles Pulliam-Moor…	"2015-07-08T…	story	162
5	2015-07-01 14:55:58	149	nan	616258951219…	nan	226	WATCH: How black women …	https://t.co/ 9eSknQHHD5	nan	nan	nan	nan
6	2015-06-25 05:55:17	118	nan	613948557859…	nan	214	More Americans ha…	http://t.co/ D90pXAHfn1	"Nidhi Prakash"	"2015-06-24T…	story	155
7	2015-07-05 22:51:59	183	FIFAWWC	617828296433…	nan	178	It seems Sepp Blatter foun…	nan	nan	nan	nan	nan
8	2015-06-22 16:50:06	168	nan	613026185321…	nan	121	WATCH: Science says…	https://t.co/ FlEK3wEea5	nan	nan	nan	nan
9	2015-06-28 17:06:12	83	nan	615204562715…	nan	120	Haiti just won a hockey…	http://t.co/ PwC4FmOoaF	"Tim Rogers"	"2015-06-28T…	story	157
10	2015-07-01 05:25:20	48	nan	616115346622…	nan	118	There are now more Spanish…	http://t.co/ hR4TwMmliB	"Casey Tolan"	"2015-06-30T…	story	158
11	2015-07-05 23:27:31	149	USA	617837239314…	nan	114	Nice to see U.S. elected…	nan	nan	nan	nan	nan
12	2015-07-15 15:04:23	121	ElChapo	621334501419…	nan	113	Former Mexican offi…	https://t.co/ HzDlkSWq77	nan	nan	nan	nan
13	2015-06-24 22:01:29	90	nan	613829320822…	nan	109	Trans woman interrupts O…	http://t.co/ FSeKfFBECr	"Jorge Rivas"	"2015-06-24T…	story	156
14	2015-08-09 03:55:10	68	nan	630225784356…	vocativ	106	Hillary Clinton's me…	http://t.co/ S5uxpDOIz8	nan	nan	nan	nan
15	2015-06-26 20:40:44	110	nan	614533777159…	atlasobscura	103	The first artifact rec…	http://t.co/ tQvJcqBU1Z	nan	nan	nan	nan
16	2015-06-20 04:55:11	69	nan	612121494232…	nan	103	Mexican lawmaker wan…	http://t.co/ kVtiaIODCu	"Rafa Fernandez De…	"2015-06-19T…	story	153
17	2015-08-09 18:59:21	32	AmberMonroe	630453330825…	nan	101	Loved ones mourn #Amber…	http://t.co/ FYztYDDtMS	"Molly McArdle"	"2015-08-09T…	story	180
18	2015-07-31 20:25:27	72	nan	627213503871…	nan	97	Dear NBC, BBC, CNN, an…	http://t.co/ LhcXz13kXm	"Nidhi Prakash"	"2015-07-30T…	story	175
19	2015-07-15 21:25:39	57	nan	621430450825…	TheNextWeb	97	This Vine from the Har…	https://t.co/ ROa7MgGQOe	nan	nan	nan	nan

Format Resize Background color Column min/max OK Cancel

Figure. 10. Input Dataset

Result Object: This is the result object with all the scores assigned to users.

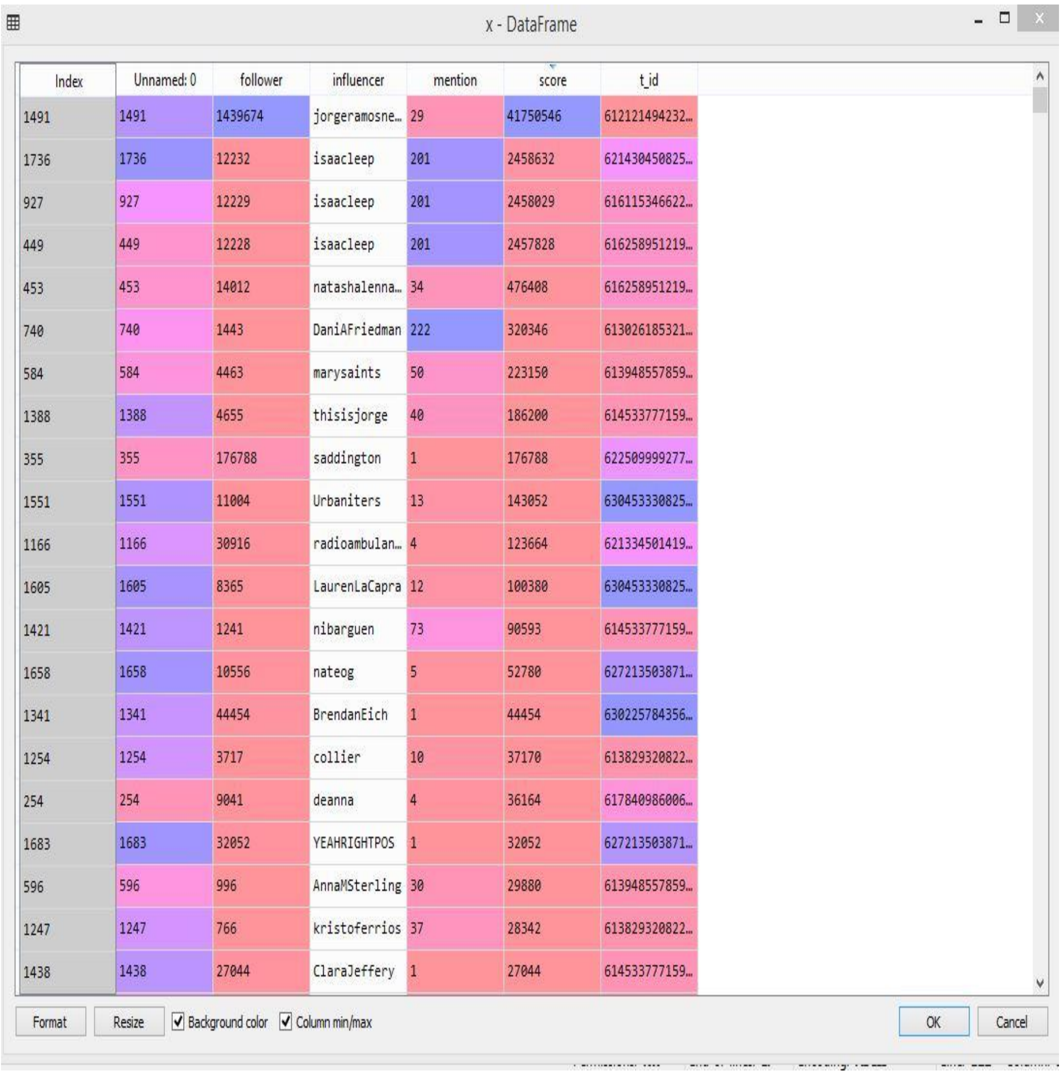
result - Dictionary (5 elements)

Key	Type	Size	Value
follower	list	88	[69, 390, 613, 96, 71, 1705, 54, 2246, 1702, 259, ...]
influencer	list	88	['erdmann_paul', 'KeKoJoNeZ', 'sugarRoyalty', 'ReyBee10', 'yona_menash ...
mention	list	88	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...]
score	list	88	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...]
t_id	list	88	[614807708575907840, 614807708575907840, 614807708575907840, 614807708 ...

OK Cancel

Figure .11. Result Object

Result DataFrame: The above object is converted to dataframe object for further process.



x - DataFrame

Index	Unnamed: 0	follower	influencer	mention	score	t_id
1491	1491	1439674	jorgeramosne…	29	41750546	612121494232…
1736	1736	12232	isaacleep	201	2458632	621430450825…
927	927	12229	isaacleep	201	2458029	616115346622…
449	449	12228	isaacleep	201	2457828	616258951219…
453	453	14012	natashalenna…	34	476408	616258951219…
740	740	1443	DaniAFriedman	222	320346	613026185321…
584	584	4463	marysaints	50	223150	613948557859…
1388	1388	4655	thisisjorge	40	186200	614533777159…
355	355	176788	saddington	1	176788	622509999277…
1551	1551	11004	Urbaniters	13	143052	630453330825…
1166	1166	30916	radioambulan…	4	123664	621334501419…
1605	1605	8365	LaurenLaCapra	12	100380	630453330825…
1421	1421	1241	nibarguen	73	90593	614533777159…
1658	1658	10556	nateog	5	52780	627213503871…
1341	1341	44454	BrendanEich	1	44454	630225784356…
1254	1254	3717	collier	10	37170	613829320822…
254	254	9041	deanna	4	36164	617840986006…
1683	1683	32052	YEAHRIGHTPOS	1	32052	627213503871…
596	596	996	AnnaMSterling	30	29880	613948557859…
1247	1247	766	kristoferrios	37	28342	613829320822…
1438	1438	27044	ClaraJeffery	1	27044	614533777159…

Format Resize Background color Column min/max OK Cancel

Figure.12. Result DataFrame

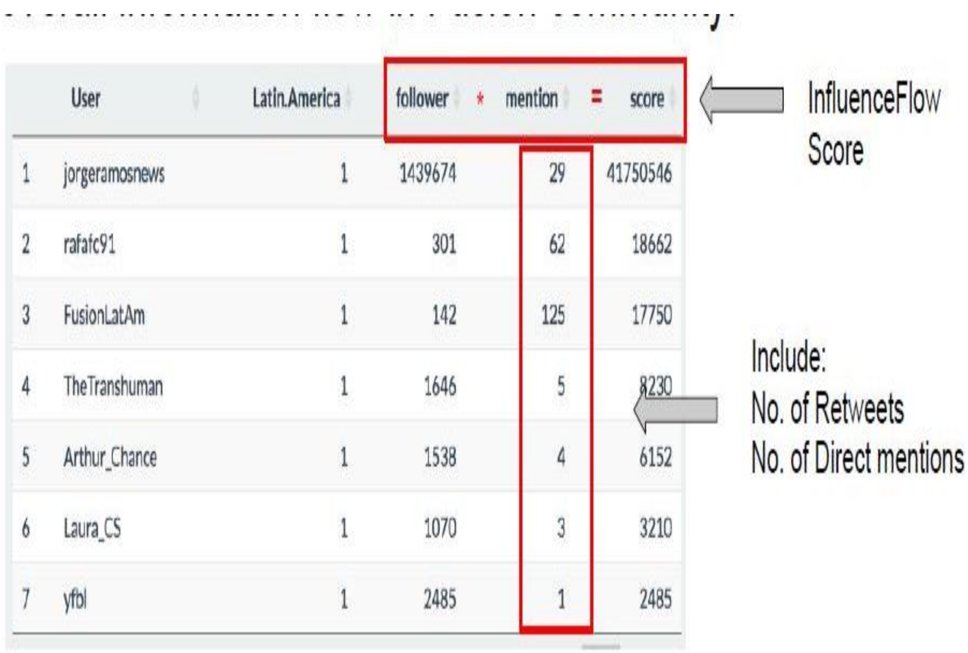
Output File: This is the result file.

follower	influencer	mention	score	t_id
1439674	jorgeramosnews	29	41750546	6.12121E+17
12232	isaacleep	201	2458632	6.2143E+17
12229	isaacleep	201	2458029	6.16115E+17
12228	isaacleep	201	2457828	6.16259E+17
14012	natashalennard	34	476408	6.16259E+17
1443	DaniAFriedman	222	320346	6.13026E+17
4463	marysaints	50	223150	6.13949E+17
4655	thisisjorge	40	186200	6.14534E+17
176788	saddington	1	176788	6.2251E+17
11004	Urbaniters	13	143052	6.30453E+17
30916	radioambulante	4	123664	6.21335E+17
8365	LaurenLaCapra	12	100380	6.30453E+17
1241	nibarguen	73	90593	6.14534E+17
10556	nateog	5	52780	6.27214E+17
44454	BrendanEich	1	44454	6.30226E+17
3717	collier	10	37170	6.13829E+17

Figure .13. Output File

Based on the information given on the Twitter and other socila networking sites uses the "fashion", "beauty", "wear" and "style" like magazines, brands, fashion designers on e-commerce websites, we have come to know and understand that about 90% of the accounts are attracted and influenced in fashion technology. Reciprocity is observed by linking their acounts mutually and surprisingly find a high value parameter as accounts having common friends.

The user with more score is considered as a better influencer in the network about a particular topic or field (Fig 15).



	User	Latin.America	follower	*	mention	=	score
1	jorgeramosnews	1	1439674		29		41750546
2	rafafc91	1	301		62		18662
3	FusionLatAm	1	142		125		17750
4	TheTranshuman	1	1646		5		8230
5	Arthur_Chance	1	1538		4		6152
6	Laura_CS	1	1070		3		3210
7	yfbl	1	2485		1		2485

Fig 14. Result Analysis

We have used Phython programming language for mapreducing and generating results. NumPy fundamental Package is used for generating multidimensional generic data. Pandas is used as datastructure tool. Tweepy is used as Twitter authentication method. JSON is used as script language for data-interchange format.

5. Conclusion

We have created homophily samples to validate our extraction method to apply centrality on given data set. During this process we have found identification of actors who are coordinated with the network has become tough problem. So interaction here needs to be more adaptive. Hence extraction method allows to rank centrality measures of influencers. In this paper we have taken only Twitter data for analysis and this concept can be extended by comparing all other social networks that influence the net browsing users. In our future research we will be producing comparative results as extension of this concept.

References

[1] G " Python for Informatics: Exploring Information" – Book by Charles Severance

[2] "Practical Data Science Cookbook" – Book by Abhijit Dasgupta, Benjamin Bengfort, Sean Patrick Murphy, and Tony Ojeda.

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

[4] L. A. Adamic. The Small World Web. In Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries (ECDL'99), Paris, France, Sep 1999.

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

[6] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of Topological Characteristics of Huge Online Social Networking Services. In Proceedings of the 16th international conference on World Wide Web (WWW'07), Banff, Canada, May 2007.

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

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

[9] A. Awan, R. A. Ferreira, S. Jagannathan, and A. Grama. Distributed uniform sampling in real-world networks. Technical Report CSD-TR-04-029, Purdue University, 2004.

[10] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06), Philadelphia, PA, Aug 2006.

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

[12] L. Becchetti, C. Castillo, D. Donato, and A. Fazzone. A Comparison of Sampling Techniques for Web Graph Characterization. In Proceedings of the Workshop on Link Analysis (LinkKDD'06), Philadelphia, PA, Aug 2006.

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

[14] 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. In Proceedings of the 9th International World Wide Web Conference (WWW'00), Amsterdam, May 2000.

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

[16] d. boyd. Friends, Friendsters, and Top 8: Writing community into being on social network sites. First Monday, 11(12), 2006.

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

[18] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-Law Relationships of the Internet Topology. In Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM'99), Cambridge, MA, Aug 1999.

[19] S. Garriss, M. Kaminsky, M. J. Freedman, B. Karp, D. Mazi`eres, and H. Yu. Re: Reliable Email. In Proceedings of the 3rd Symposium on Networked Systems Design and Implementation (NSDI'06), San Jose, CA, May 2006.

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

[21] Google Co-op. http://www.google.com/coop/. [20] M. Granovetter. The Strength of Weak Ties.

American Journal of Sociology, 78(6), 1973.

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

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

[24] J. Kleinberg. The Small-World Phenomenon: An Algorithmic Perspective. In Proceedings of the 32nd ACM Symposium on Theory of Computing (STOC'00), Portland, OR, May 2000.

[25] J. Kleinberg and S. Lawrence. The Structure of the Web. Science, 294:1849–1850, 2001.

[26] J. M. Kleinberg and R. Rubinfeld. Short paths in expander graphs. In IEEE Symposium on Foundations of Computer Science (FOCS'96), Burlington, VT, Oct 1996.

[27] R. Kumar, J. Novak, and A. Tomkins. Structure and Evolution of Online Social Networks. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06), Philadelphia, PA, Aug 2006.

[28] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the Web for Emerging Cyber-Communities. Computer Networks, 31:1481–1493, 1999.

[29] S. Lee, R. Sherwood, and B. Bhattacharjee. Cooperative peer groups in NICE. In Proceedings of the Conference on Computer Communications (INFOCOM'03), San Francisco, CA, Mar 2003.

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

[31] L. Li and D. Alderson. Diversity of graphs with highly variable connectivity. Physics Review E, 75, 2007.

[32] 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.

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

[34] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat. Systematic Topology Analysis and Generation Using Degree Correlations. In Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM'06), Pisa, Italy, August 2006.

[35] S. Milgram. The small world problem. Psychology Today, 2(60), 1967.

[36] A. Mislove, K. P. Gummadi, and P. Druschel. Exploiting social networks for Internet search. In Proceedings of the 5th Workshop on Hot Topics in Networks (HotNets-V), Irvine, CA, Nov 2006.

[37] M. Molloy and B. Reed. A critical point for random graphs with a given degree distribution. Random Structures and Algorithms, 6, 1995.

[38] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. Combinatorics, Probability and Computing, 7, 1998.

[39] R. Morselli, B. Bhattacharjee, J. Katz, and M. A. Marsh. Keychains: A Decentralized Public-Key Infrastructure. Technical Report CS-TR-4788, University of Maryland, 2006.

[40] MozillaCoop. http://www.mozilla.com.

[41] MySpace is the number one website in the U.S. according to Hitwise. HitWise Press Release, July, 11, 2006. http://www.hitwise.com/press-center/ hitwiseHS2004/social-networking-june-2006.php.

[42] M. E. J. Newman. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences (PNAS), 98:409–415, 2001. [42] M. E. J. Newman. Mixing patterns in networks. Physics Review E, 67, 2003.

BIBLIOGRAPHY OF AUTHORS

SRIKANTH BETHU has completed M.Tech in Computer Science and Engineering in 2011 at Osmania University, Hyderabad, India. He was currently working as Assistant Professor in Department of Computer Science and Engineering at Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad. He was very fond of Data Mining concepts. He has published several research papers in the areas of Data Mining, Data Engineering and Big Data. Attended and Presented several research papers in various National and International conferences.

V. SOWMYA is working as Associate Professor in Department of Computer Science and Engineering, GRIET, Hyderabad. She completed M.Tech(Software Engineering) from Aurora College of Engineering, Bhongiri. Her area of research interest includes: Text Mining, Machine Learning, Data Science and Natural Language Processing. She has 11 years of teaching experience.She has published papers in international conferences and journals.

Dr. B. SANKARABABU was a doctorate from Aacharya Nagarjuna University, Guntur, Andhrapradesh, India, and completed PhD in 2016. Currently working as a Professor in Department of Computer Science and Engineering at Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad .He has done his research on Data Mining. He has published several research papers on Data mining concepts.

G.Charles Babu , Presently working as a Professor in Dept. of CSE in Malla Reddy Engineering College(Autonomous), Secunderabad, Telangana Since 5 Years and Total Teaching experience of 20 Years. Completed B.Tech (CSE) in 1997 from KLCE, M.Tech(SE) in 1999 from JNTUH and Ph.D(Data Mining) from ANU. Published more than 50 Research Papers in Data Mining, Cloud Computing, Image Processing.

Dr. Y. Jeevan Nagendra Kumar, obtained his Ph.Din Computer Science and Engineering from Acharya Nagarjuna University, Guntur, AP in 2017 and M.Tech Computer Science Technology from Andhra University in 2005. He is working as Professor and Dean - Technology and Innovation Cell in GRIET since 2005. He has about 9 Research Papers in International / National Conferences and Journals and also attended many FDP Programs to enhance his knowledge. With his technical knowledge he guided the students in developing the useful Web applications and data mining related products.