iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: VII Month of publication: July 2019

DOI: http://doi.org/10.22214/ijraset.2019.7057

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VII, July 2019- Available at www.ijraset.com

Analysing the Structure, Dynamics and Contents of Social Networks

Rohini Lokhande
Department of Information Technology, Thakur College of Engineering and Technology, Mumbai, India

Abstract: Social media permits its users to have numerous accounts in diverse social networking podiums. Popular social networking sites have billions of users that possess similar interests. The aim of this paper is to determine the identity of users and map those identities with those having similar interests. The developed system can be utilized for business intelligence that would help in creating a database of all users that possess specific interests. In this study, a system is developed wherein a user's core interest is recognized in well-known social networking sites (i.e., Twitter and Plurk) by means of latent Dirichlet allocation (LDA) algorithm and Kullback–Leibler divergence method. In addition, in order to identify users across the abovementioned social networking sites, LDA will cogitate social network structure and article content of users. LDA initially establishes a user's core interest. Next, it calculates similarity between target users. The results in this study prove to be quite promising, which reveal that LDA is an effective algorithm to map users across various platforms and efficient when compared with other methods.
Keywords: LDA, Twitter, Plurk, Social Network, KL Divergence

I. INTRODUCTION

Facebook, Twitter, LinkedIn, Plurk, etc. are some of the most prominent social networking sites that have billions of users who socialize with each other. Socializing in such sites means sharing of knowledge, ideas, and interests with the like-minded people. Enormous information is generated through such sites, and handling such information is a crucial factor so that business intelligence (BI) can be managed appropriately. For appropriate management of BI, user identification and mapping identified users from numerous sites are vital.

Social networking can be defined as the usage of internet-based programs to connect with acquaintances, relatives, colleagues, and clienteles. Social networking sites can be used for not only social purposes but also business purposes. Social networking sites are one of the key areas for marketers or salespersons trying to engross users [1–3].

A social network consists of individuals known as nodes that are connected by a particular type of interdependency such as comradeship, similarity, common hobby or activity, and rapports of philosophies, learning, or stature.

Fig. 1 Structure of social network

Social network analyses relationships through network theory that contains nodes, edges, links, or connections [4]. Within a network, individual actors are called nodes, and relationships between the actors are called ties [5]. Ties can be of several types between nodes. It has been observed that social networking sites function at various levels (i.e., from family level to the level of nations) and portray an indispensable part in regulating the way how problems are solved, organizations are administered, and the extent to which an individual flourishes in attaining his/her goal. In other words, a social network is a map of specified ties.

Accordingly, the nodes to which an individual is coupled are called social contacts of that individual. The value that an individual obtains from social network is called social capital.

Plenty of information and functions are present in social networks. User activities on social networks have turned out to be convoluted and impulsive. Determining core interests of users from superfluous data is one of the main objectives of this paper. In order to determine the core interest of a user, it is necessary to first find target user's core circle of friends. Figure 3 shows structure of matching users. There are two users, i.e., *User X* and *User Y*, who have friends in various social networking sites (e.g., *User X* can be *User* X_1, *User* X_4, *User* X_3, and *User* X_5 and *User Y* can be *User* Y_1, *User* Y_2, *User* Y_3, *User* Y_4, *User* Y_5, and *User* Y_6). If both these users be a member of same individual, then there would be analogous circle of friends, but the only problem would be that the circle of friends will not be of same size.

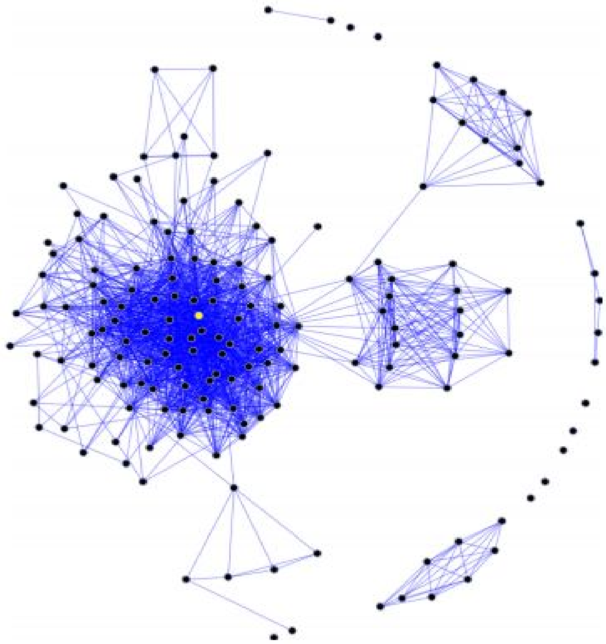

Fig. 2 Example of a social network diagram: The node with the highest betweenness centrality is marked in yellow

Finding similar circle of friends is not that easy because we are not aware of all identities of users. Therefore, in order to determine core interests of a user, the following assumptions have to be taken into account:

1) The more frequent is the communication between two users, the more closer they are.
2) There could be a possibility that similar circle of friends subsists in disparate social networks.

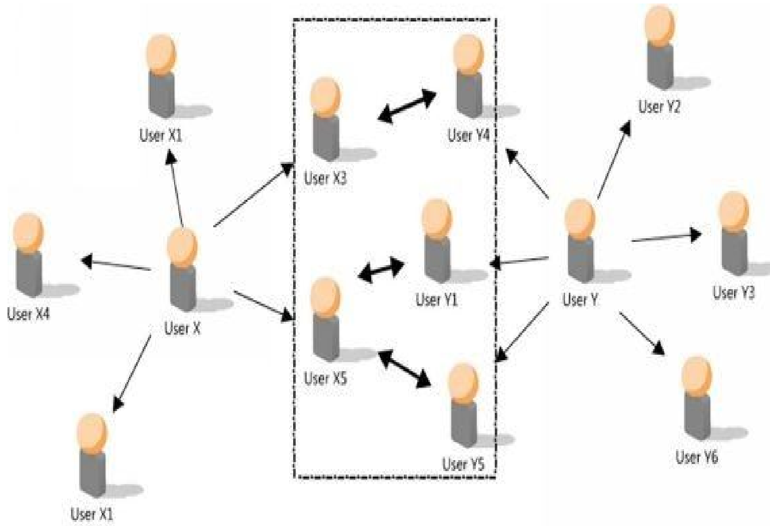

Fig. 3 Structure of matching users

The remainder of this paper is organized as follows. Section II describes about the review of literature that is carried out in this domain. Section III focuses on the proposed system and the algorithm used in this study. Sections IV and V illustrates on the implementation details of the proposed system along with results and discussion. Section VI concludes the study with scope for future.

II. REVIEW OF LITERATURE

A great deal of literature review has been performed under this domain. Some of them are as follows. Zafarani Reza and Huan Liu [6] successfully showed that MOBIUS is helpful in recognizing users through social networking sites. The authors too created a path for scrutinizing and mining across social networking sites, which in turn lead to the establishment of innovative online services across sites. However, they were unable to analyze possibilities and discover features indigenous to specific sites beyond those constricted to usernames and incorporating them into MOBIUS for future needs. Stephen Paul Marsh [7] investigated philosophies of trust in diverse circumstances and developed an official narrative of its use with distributed and intelligent representatives. However, it was observed that Marsh's model was convoluted, predominantly conjectural in nature, and strenuous during implementation. Abdul-Rahman Alfarez and Stephen Hailes [8] propositioned a model for strengthening trust in computer-generated communities on the basis of undeviating proficiencies and repute. Nonetheless, the proposed model was ad-hoc in nature, which restrains the applicability of the model in far-reaching scope. Schillo Michael, Petra Funk, and Michael Rovatsos [9] developed a trust model for situations where result of interaction is Boolean (i.e., good or bad) between trust relationship of two agents. A major disadvantage of their trust model was that they failed to contemplate on the degrees of satisfaction. Esfandiari Babak, and Sanjay Chandrasekharan [10] proposed two one-to-one trust acquisition mechanisms in their trust model. The first trust acquisition mechanism is based on observation. The authors also proposed the usage of Bayesian networks, and in order to accomplish trust acquisition, they suggested Bayesian learning. However, the proposed mechanism was unable to make a distinction between distrust and lack of knowledge about trust. Yu Bin and Munindar P. Singh [11] specified that information that is stored by an agent about direct interactions is a collection of values that replicate quality of these interactions. Moreover, they specified that only the most recent experiences with each concrete partner are taken into account for calculations. However, the model failed to merge direct information with witness information. Mui Lik, Mojdeh Mohtashemi, and Ari Halberstadt [12] put forward a computational model grounded on sociological and biotic understanding. Their model not only calculated an agent's trust score but also reputation score. However, the authors failed to observe the consequences of duplicity in their model. Nie Yuanping, Yan Jia, Shudong Li, Xiang Zhu, Aiping Li, and Bin Zhou [13] proposed a dynamic core interests mapping (DCIM) algorithm that takes into account not only a user's social network structure but also the article content of a user to relate users across several social networking platforms. However, the authors were unable to improve the accuracy of user's core topic analyses. Kumar Shamanth, Reza Zafarani, and Huan Liu [14] insinuated a realistic approach to scrutinize migration patterns in social networking sites. They discovered patterns provided insights, which helped the authors in understanding social networking sites and estimating their attractiveness to improve BI and generating revenue by retaining users. However, they were unable to obtain the mapping of users across different social media sites.

Moreover, the authors were unable to determine if a user has moved to another site. Liu Siyuan, Shuhui Wang, Feida Zhu, Jinbo Zhang, and Ramayya Krishnan [15] proposed a solution named HYDRA—a framework that consists of the following steps: exhibiting mixed behavior by long-term behavior distribution investigation and multiresolution progressive information matching, forming physical steadiness graph to calculate high-order physical steadiness on users' core social structures throughout dissimilar social networking platforms, realizing mapping function through multiobjective optimization comprising supervised learning as well as cross platform structure consistency maximization.

However, the authors considered only two real datasets (i.e., five popular Chinese social networks and two popular English social networks). More real datasets would be considered, which would have improved the accuracy of the proposed system.

Cao Gaofeng, and Li Kuang [16] conducted experiments and the results of the experiments demonstrated the usefulness of extraction of core users and proved that ~20% core users permit recommender systems to attain >90% accuracy of top-N recommendation. However, the authors were unable to determine more approaches to define core users from different aspects (e.g., fusing similar relationships and trust relationships). Moreover, the authors were unable to find relatively stable approach to generate core users so as to reduce the frequency of updates.

Considering the abovementioned research gap limitations, this research work aims to propose and develop a unique approach to analyze the structure, dynamic, and content of social networking sites.

Firstly, trust degree and interest similarity between all pairs of users are calculated and sorted from highest value to the lowest. Secondly, two strategies are used to select core users. This first strategy is to select users who appear the most in all other user's nearest neighbor list. The second strategy is to select a user who has the highest weight of location in all other users' nearest neighbor list. Thirdly, the effect of extracted core users in recommendation is validated.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VII, July 2019- Available at www.ijraset.com

III. PROPOSED SYSTEM

Figure 4 shows the block diagram of the proposed system that is used for identifying core users that possess similar interests in social networking sites (i.e., Twitter and Plurk).

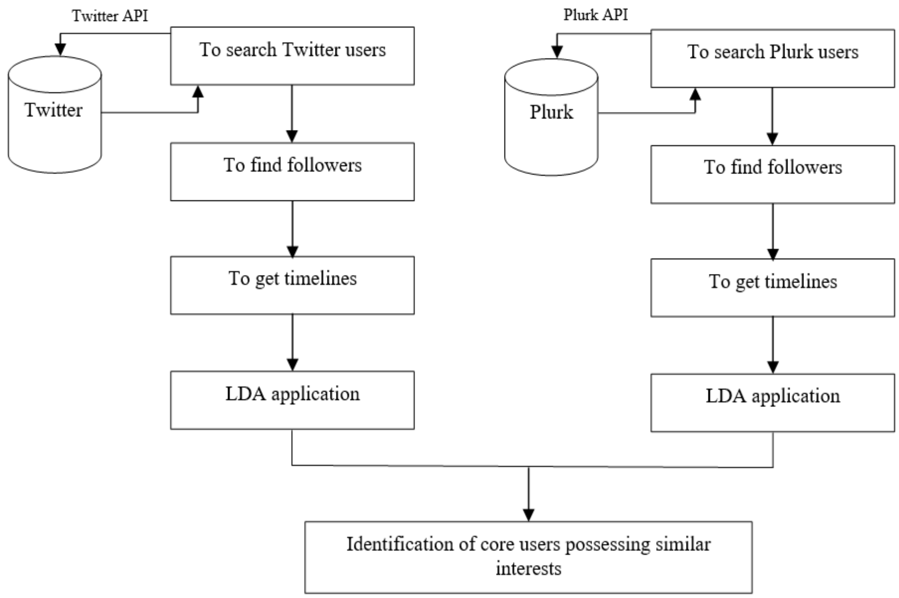

Fig. 4 Block diagram of proposed system

In this study, two social networking sites have been considered (i.e., Twitter and Plurk). In Twitter, there are as sequence of steps that need to be followed for identifying core users that possess similar interest. The first step involves searching Twitter users. In order to search Twitter users, we use Twitter application programming interface (API). For instance, Twitter API is able to provide names of 1000 Twitter users. The next step involves reading the timelines of those 1000 Twitter users and identifying their followings. In other words, for a particular Twitter user, we try to determine the total number of followers that user has. Once the total number of followers have been identified, the next step is to read the timelines of those followers. Once the timelines of the followers have been read, the last step is to apply LDA algorithm that provides a list of core users that possess similar interests. Similarly, for Plurk, the first step involves searching Plurk users. In order to search Plurk users, we use Plurk API. For instance, Plurk API is able to provide names of 1000 Plurk users. The next step involves reading the timelines of those 1000 Plurk users and identifying their followings. In other words, for a particular Plurk user, we try to determine the total number of followers that user has. Once the total number of followers have been identified, the next step is to read the timelines of those followers. Once the timelines of the followers have been read, the last step is to apply LDA algorithm that provides a list of core users that possess similar interests. Finally, a comparison can be made between the two social networking sites that identify core users possessing similar interests.

A. Latent Dirichlet Algorithm

Blei et al. [17] proposed LDA that can be used for topic modeling. It is used for understanding topics across documents (humans vs. machines). The main intention of using LDA is mainly to detect underlying topics in text documents. LDA can be used for a variety of purposes like sentiment analysis, object localization for images, automatic harmonic analysis for music, and bioinformatics. In LDA, the first assumption is that documents with similar topics will use similar groups of words. LDA suggests that words carry strong semantic information and documents discussing similar topics will use similar words. Latent topics are therefore discovered by identifying groups in the corpus that will currently occur together within documents. In addition, the second assumption is about document definitions/modeling. For instance, documents are probability distributions over latent topics, and topics are probability distributions over words. This means that according to LDA every different document contains a number of topics, each topic has a distribution of words associated with it. Note that in LDA, probability distributions are used instead of strict word frequencies. So, while other bag-of-words models may focus on the most frequently occurring words in a document, in this study, a holistic approach is used wherein a lot of concentration is given on the distribution of words across topics.

Figure 5 shows several documents (i.e., Doc1 and Doc2) that comprise distribution of topics.

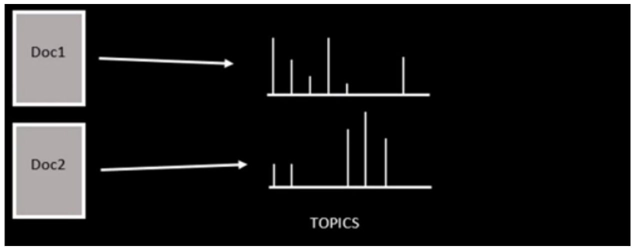

Fig. 5 Documents (Doc1 and Doc2) comprising distribution of topics

Figure 6 shows several topics (i.e., Topic A and Topic B) that comprise distribution of words.

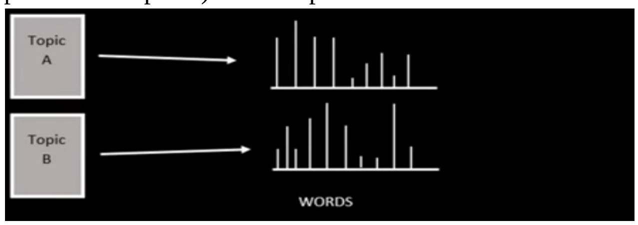

Fig. 6 Topics (Topic A and Topic B) comprising distribution of words

B. Generative Process

LDA assumes that new documents are created in the following manner:

1) Determine the number of words in a document.
2) Choose a topic mixture for the document over a fixed set of topics (i.e., 20% topic A, 30% topic B, and 50% topic C).
3) Generate words in the document by:
a) First, select a topic based on the document's multinomial distribution.
b) Second, select a word based on the topic's multinomial distribution.

C. LDA as a Topic Model

LDA is a topic model that generates topics based on word frequency from a set of documents. It is specifically useful for finding reasonably accurate mixtures of topics within a given document.

Steps involved in performing LDA:

1) Create a collection of documents from news articles.
2) Each document represents a news article.
3) Data cleaning is the next step:
a) *Tokenizing:* Converting a document to its atomic elements.
b) *Stopping:* Removing meaningless words.
c) *Stemming:* Merging words that are equivalent in meaning.

LDA assigns a random topic to each word in the corpus of documents that have been provided. It starts by randomly assigning topics.

IV. IMPLEMENTATION

A. Installation of Twitter Library

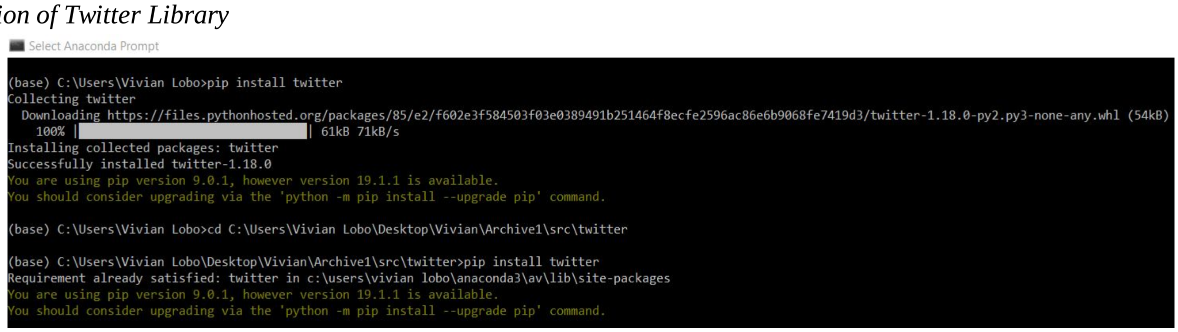

Fig. 7 Twitter library installation

B. Installation of Tweepy Library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>pip install tweepy
Collecting tweepy
  Downloading https://files.pythonhosted.org/packages/d5/5f/daac4b4e9b30d7d2a6fdd16a880ff79f27918fe388e4dfc1983dec3a9876/tweepy-3.7.0-py2.py3-none-any.whl
Requirement already satisfied: requests>=2.11.1 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from tweepy)
Requirement already satisfied: PySocks>=1.5.7 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from tweepy)
Requirement already satisfied: six>=1.10.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from tweepy)
Collecting requests-oauthlib>=0.7.0 (from tweepy)
  Downloading https://files.pythonhosted.org/packages/c2/e2/9fd03d55ffb70fe51f587f20bcf407a6927eb121de86928b34d162f0b1ac/requests_oauthlib-1.2.0-py2.py3-none-any.whl
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests>=2.11.1->tweepy)
Requirement already satisfied: idna<2.7,>=2.5 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests>=2.11.1->tweepy)
Requirement already satisfied: urllib3<1.23,>=1.21.1 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests>=2.11.1->tweepy)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests>=2.11.1->tweepy)
Collecting oauthlib>=3.0.0 (from requests-oauthlib>=0.7.0->tweepy)
  Downloading https://files.pythonhosted.org/packages/16/95/699466b05b72b94a41f662dc9edf87fda4289e3602ecd42d27fcaddf7b56/oauthlib-3.0.1-py2.py3-none-any.whl (142kB)
    100% |████████████████████████████████| 143kB 302kB/s
Installing collected packages: oauthlib, requests-oauthlib, tweepy
Successfully installed oauthlib-3.0.1 requests-oauthlib-1.2.0 tweepy-3.7.0
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 8 Tweepy library installation

C. Installation of LDA Library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>install lda
'install' is not recognized as an internal or external command,
operable program or batch file.

(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>pip install lda
Collecting lda
  Downloading https://files.pythonhosted.org/packages/fa/fe/98cec39e45411a83430978ea2ad21a47468875013e3c3512a9aef9afc46f/lda-1.1.0-cp36-cp36m-win_amd64.whl (334kB)
    100% |████████████████████████████████| 337kB 501kB/s
Requirement already satisfied: numpy<2.0,>=1.13.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from lda)
Collecting pbr<4,>=0.6 (from lda)
  Downloading https://files.pythonhosted.org/packages/0c/5d/b077dbf309993d52c1d71e6bf6fe443a8029ea215135ebbe0b1b10e7aefc/pbr-3.1.1-py2.py3-none-any.whl (99kB)
    100% |████████████████████████████████| 102kB 2.9MB/s
Installing collected packages: pbr, lda
Successfully installed lda-1.1.0 pbr-3.1.1
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 9 LDA library installation

D. Installation of NLTK Library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>pip install nltk
Requirement already satisfied: nltk in c:\users\vivian lobo\anaconda3\av\lib\site-packages
Requirement already satisfied: six in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from nltk)
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 10 NLTK library installation

E. Installation of stop-words Library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>pip install stop-words
Collecting stop-words
  Using cached https://files.pythonhosted.org/packages/1c/cb/d58290804b7a4c5daa42abbbe2a93c477ae53e45541b1825e86f0dfaaf63/stop-words-2018.7.23.tar.gz
Building wheels for collected packages: stop-words
  Running setup.py bdist_wheel for stop-words ... done
  Stored in directory: C:\Users\Vivian Lobo\AppData\Local\pip\Cache\wheels\75\37\6a\2b295e03bd07290f0da95c3adb9a74ba95fbc333aa8b0c7c78
Successfully built stop-words
Installing collected packages: stop-words
Successfully installed stop-words-2018.7.23
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 11 Stop-words library installation

F. Installation of Gensim Library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>pip install gensim
Collecting gensim
  Downloading https://files.pythonhosted.org/packages/81/80/858ef502e80baa6384b75fd5c89f01074b791a13b830487f9e25bdce50ec/gensim-3.7.3.tar.gz (23.4MB)
    100% |████████████████████████████████| 23.4MB 352kB/s
Requirement already satisfied: numpy>=1.11.3 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from gensim)
Requirement already satisfied: scipy>=0.18.1 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from gensim)
Requirement already satisfied: six>=1.5.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from gensim)
Collecting smart_open>=1.7.0 (from gensim)
  Downloading https://files.pythonhosted.org/packages/37/c0/25d19badc495428dec6a4bf7782de617ee0246a9211af75b302a2681dea7/smart_open-1.8.4.tar.gz (63kB)
    100% |████████████████████████████████| 71kB 1.9MB/s
Requirement already satisfied: boto>=2.32 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from smart_open>=1.7.0->gensim)
Requirement already satisfied: requests in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from smart_open>=1.7.0->gensim)
Collecting boto3 (from smart_open>=1.7.0->gensim)
  Downloading https://files.pythonhosted.org/packages/36/cf/ae45d1ca1e4d5aea1e6ec260906ae27ef53185ac986940f23ddee47b5364/boto3-1.9.162-py2.py3-none-any.whl (128kB)
    100% |████████████████████████████████| 133kB 608kB/s
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->smart_open>=1.7.0->gensim)
Requirement already satisfied: idna<2.7,>=2.5 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->smart_open>=1.7.0->gensim)
Requirement already satisfied: urllib3<1.23,>=1.21.1 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->smart_open>=1.7.0->gensim)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->smart_open>=1.7.0->gensim)
Collecting botocore<1.13.0,>=1.12.162 (from boto3->smart_open>=1.7.0->gensim)
  Downloading https://files.pythonhosted.org/packages/9c/70/519c9fce0af131989042906d3584f0b105588de211cf6a060ed018c98b10/botocore-1.12.162-py2.py3-none-any.whl (5.5MB)
    100% |████████████████████████████████| 5.5MB 724kB/s
Collecting jmespath<1.0.0,>=0.7.1 (from boto3->smart_open>=1.7.0->gensim)
  Downloading https://files.pythonhosted.org/packages/83/94/7179c3832a6d45b266ddb2aac329e101367fbdb11f425f13771d27f225bb/jmespath-0.9.4-py2.py3-none-any.whl
Collecting s3transfer<0.3.0,>=0.2.0 (from boto3->smart_open>=1.7.0->gensim)
  Downloading https://files.pythonhosted.org/packages/16/8a/1fc3dba0c4923c2a76e1ff0d52b305c44606da63f718d14d3231e21c51b0/s3transfer-0.2.1-py2.py3-none-any.whl (70kB)
    100% |████████████████████████████████| 71kB 1.9MB/s
Requirement already satisfied: docutils>=0.10 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from botocore<1.13.0,>=1.12.162->boto3->smart_open>=1.7.0->gensim)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1; python_version >= "2.7" in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from botocore<1.13.0,>=1.12.162->boto3->smart_open>=1.7.0->gensim)
Building wheels for collected packages: gensim, smart-open
  Running setup.py bdist_wheel for gensim ... done
  Stored in directory: C:\Users\Vivian Lobo\AppData\Local\pip\Cache\wheels\73\6b\89\bb14fd56b74774a39a771a12f525a6a14c2c2692d3084ad048
  Running setup.py bdist_wheel for smart-open ... done
  Stored in directory: C:\Users\Vivian Lobo\AppData\Local\pip\Cache\wheels\5f\ea\fb\5b1a947b369724063b2617011f1540c44eb00e28c3d2ca8692
Successfully built gensim smart-open
Installing collected packages: jmespath, botocore, s3transfer, boto3, smart-open, gensim
Successfully installed boto3-1.9.162 botocore-1.12.162 gensim-3.7.3 jmespath-0.9.4 s3transfer-0.2.1 smart-open-1.8.4
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 12 Gensim library installation

G. Installation of Pyldavis Library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>pip install pyldavis
Collecting pyldavis
  Downloading https://files.pythonhosted.org/packages/a5/3a/af82e070a8a96e13217c8f362f9a73e82d61ac8fff3a2561946a97f96266/pyLDAvis-2.1.2.tar.gz (1.6MB)
    100% |████████████████████████████████| 1.6MB 433kB/s
Requirement already satisfied: wheel>=0.23.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Requirement already satisfied: numpy>=1.9.2 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Requirement already satisfied: scipy>=0.18.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Requirement already satisfied: pandas>=0.17.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Requirement already satisfied: joblib>=0.8.4 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Requirement already satisfied: jinja2>=2.7.2 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Requirement already satisfied: numexpr in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Requirement already satisfied: pytest in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pyldavis)
Collecting future (from pyldavis)
  Downloading https://files.pythonhosted.org/packages/90/52/e20466b85000a181e1e144fd8305caf2cf475e2f9674e797b222f8105f5f/future-0.17.1.tar.gz (829kB)
    100% |████████████████████████████████| 829kB 682kB/s
Collecting funcy (from pyldavis)
  Downloading https://files.pythonhosted.org/packages/b3/23/d1f90f4e2af5f9d4921ab3797e33cf0503e3f130dd390a812f3bf59ce9ea/funcy-1.12-py2.py3-none-any.whl
Requirement already satisfied: python-dateutil>=2 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pandas>=0.17.0->pyldavis)
Requirement already satisfied: pytz>=2011k in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pandas>=0.17.0->pyldavis)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from jinja2>=2.7.2->pyldavis)
Requirement already satisfied: py>=1.5.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pytest->pyldavis)
Requirement already satisfied: six>=1.10.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pytest->pyldavis)
Requirement already satisfied: setuptools in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pytest->pyldavis)
Requirement already satisfied: attrs>=17.2.0 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pytest->pyldavis)
Requirement already satisfied: pluggy<0.7,>=0.5 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pytest->pyldavis)
Requirement already satisfied: colorama in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from pytest->pyldavis)
Building wheels for collected packages: pyldavis, future
  Running setup.py bdist_wheel for pyldavis ... done
  Stored in directory: C:\Users\Vivian Lobo\AppData\Local\pip\Cache\wheels\98\71\24\513a99e58bb6b8465bae4d2d5e9dba8f0bef8179e3051ac414
  Running setup.py bdist_wheel for future ... done
  Stored in directory: C:\Users\Vivian Lobo\AppData\Local\pip\Cache\wheels\0c\61\d2\d6b7317325828fbb39ee6ad559dbe4664d0896da4721bf379e
Successfully built pyldavis future
Installing collected packages: future, funcy, pyldavis
Successfully installed funcy-1.12 future-0.17.1 pyldavis-2.1.2
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 13 Pyldavis library installation

H. Installation Of ipython Library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\twitter>pip install IPython
Requirement already satisfied: IPython in c:\users\vivian lobo\anaconda3\av\lib\site-packages
Requirement already satisfied: setuptools>=18.5 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: jedi>=0.10 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: decorator in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: pickleshare in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: simplegeneric>0.8 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: traitlets>=4.2 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: prompt_toolkit<2.0.0,>=1.0.4 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: pygments in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: colorama in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from IPython)
Requirement already satisfied: parso==0.1.* in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from jedi>=0.10->IPython)
Requirement already satisfied: ipython_genutils in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from traitlets>=4.2->IPython)
Requirement already satisfied: six in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from traitlets>=4.2->IPython)
Requirement already satisfied: wcwidth in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from prompt_toolkit<2.0.0,>=1.0.4->IPython)
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 14 IPython library installation

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VII, July 2019- Available at www.ijraset.com

1) *For Twitter: To search Twitter users and find followers of a Twitter user*

TABLE I
Twitter Users and Their Respective Followers

realDonaldTrump *Donald J. Trump*	RepMcCaul	Marcorubio
Jim_Jordan	ONEinAmerica	**mike_pence *Mike Pence***
MariaBartiromo	Chicagosmayor	Alyssafarah
VP	SecArmy	RyanAFournier
GOPChairwoman	Smartwomen	SecondLady
Parscale	SenatorShaheen	VP
PressSec	Jacquelyn_M	TrumpInaugural
TuckerCarlson	Surabees	Seanspicer
JesseBWatters	DBohigian	BetsyDeVos
WhiteHouse	**FLOTUS *Melania Trump***	Parscale
Scavino45	WhiteHouse	CLewandowski_
Trump *The Trump Organization*	WhiteHouseHstry	MELANIATRUMP
MarALago	BarackObama	**LaraLeaTrump *Lara Trump***
MichaelBreed	realDonaldTrump	TimMurtaugh
TrumpStore	VP	JoeTalkShow
MELANIATRUMP	SecondLady	realBenjiIrby
GMatTRUMPNY	POTUS	TomFitton
mdamelincourt	**EricTrump *Eric Trump***	Deneenborelli
ExecutivePour	PamelaBrownCNN	TPUSA
JodieWidaseck	EmersonPolling	_BlairEllis
AlbemarleEstate	Andrew_M_Nelson	SaraCarterDC
TrumpPanama	HillSchoolYA	ShannonBream
DonaldJTrumpJr *Donald Trump Jr.*	ColbyCovMMA	MarkHarrisNC9
allahpundit	AviBerkow	**kimguilfoyle *Kimberly Guilfoyle***
JeffLandry	BilldeBlasio	WalidPhares
alx	LonnekeEngel	Smerconish
Doranimated	VAKruta	CuomoPrimeTime
RichSementa	Mchooyah	Yahoolifestyle
dougstafford	**TeamTrump *Official Team Trump***	AaronLeuerCEO
RobManess	Kimguilfoyle	JanitaKan
DevinNunes	Thehannahjane	RajaFlores
AndrewCMcCarthy	RNCLatinos	Alyssafarah
SecBernhardt	GaryCoby	UN_Spokesperson
IvankaTrump *Ivanka Trump*	RNCResearch	TimRunsHisMouth
theGESsummit	AZachParkinson	
	MattWolking	
	ErinMPerrine	
	TrumpWarRoom	

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VII, July 2019- Available at www.ijraset.com

2) For Twitter: To get timelines of the followers

TABLE II
Twitter Users and Their Respective Followers Through Their Timelines

Twitter users		Followers
DonaldJTrumpJr	1.	allahpundit
	2.	JeffLandry
	3.	alx
	4.	Doranimated
	5.	RichSementa
	6.	dougstafford
	7.	RobManess
	8.	DevinNunes
	9.	AndrewCMcCarthy
	10.	SecBernhardt
EricTrump	1.	PamelaBrownCNN
	2.	EmersonPolling
	3.	Andrew_M_Nelson
	4.	HillSchoolYA
	5.	ColbyCovMMA
	6.	AviBerkow
	7.	BilldeBlasio
	8.	LonnekeEngel
	9.	VAKruta
	10.	mchooyah
FLOTUS	1.	WhiteHouse
	2.	WhiteHouseHstry
	3.	BarackObama
	4.	realDonaldTrump
	5.	VP
	6.	SecondLady
	7.	POTUS
IvankaTrump	1.	theGESsummit
	2.	RepMcCaul
	3.	ONEinAmerica
	4.	chicagosmayor
	5.	SecArmy
	6.	Smartwomen
	7.	SenatorShaheen
	8.	Jacquelyn_M
	9.	Surabees
	10.	DBohigian
kimguilfoyle	1.	WalidPhares
	2.	smerconish
	3.	CuomoPrimeTime
	4.	yahoolifestyle
	5.	AaronLeuerCEO
	6.	JanitaKan
	7.	RajaFlores
	8.	Alyssafarah
	9.	UN_Spokesperson
	10.	TimRunsHisMouth
LaraLeaTrump	1.	TimMurtaugh
	2.	JoeTalkShow
	3.	realBenjiIrby
	4.	TomFitton
	5.	Deneenborelli
	6.	TPUSA
	7.	_BlairEllis
	8.	SaraCarterDC
	9.	ShannonBream
	10.	MarkHarrisNC9
mike_pence	1.	Alyssafarah
	2.	RyanAFournier
	3.	SecondLady
	4.	VP
	5.	TrumpInaugural
	6.	seanspicer
	7.	BetsyDeVos
	8.	parscale
	9.	CLewandowski_
	10.	MELANIATRUMP
POTUS	1.	USTradeRep
	2.	realDonaldTrump
	3.	NASA
	4.	foxandfriends
	5.	ericbolling
	6.	StateDept
	7.	EPA
	8.	DHSgov
	9.	USDOL
	10.	usedgov
realDonaldTrump	1.	Jim_Jordan
	2.	MariaBartiromo
	3.	VP
	4.	GOPChairwoman
	5.	parscale
	6.	PressSec
	7.	TuckerCarlson
	8.	JesseBWatters
	9.	WhiteHouse
	10.	Scavino45
TeamTrump	1.	kimguilfoyle
	2.	thehannahjane
	3.	RNCLatinos
	4.	GaryCoby
	5.	RNCResearch
	6.	AZachParkinson
	7.	MattWolking
	8.	ErinMPerrine
	9.	TrumpWarRoom
	10.	marcorubio
TiffanyATrump	1.	Femail
	2.	DailyCaller
	3.	DailyMailUK
	4.	MailOnline
	5.	DailyMail
	6.	SecondLady
	7.	AviBerkow
	8.	RealWalkAway
	9.	usminority
	10.	greggutfeld
Trump	1.	MarALago
	2.	MichaelBreed
	3.	TrumpStore
	4.	MELANIATRUMP
	5.	GMatTRUMPNY
	6.	mdamelincourt
	7.	ExecutivePour
	8.	JodieWidaseck
	9.	AlbemarleEstate
	10.	TrumpPanama

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VII, July 2019- Available at www.ijraset.com

3) For Twitter: LDA application

TABLE III

Twitter Users and Their Respective Followers after LDA Application

Sr. no	Twitter users and their respective followers
1.	DonaldJTrumpJr.csv
	['user', 'DonaldJTrumpJr', 'allahpundit', 'JeffLandry', 'alx', 'Doranimated', 'RichSementa', 'dougstafford', 'RobManess', 'DevinNunes', 'AndrewCMcCarthy', 'SecBernhardt']
2.	EricTrump.csv
	['user', 'EricTrump', 'PamelaBrownCNN', 'EmersonPolling', 'Andrew_M_Nelson', 'HillSchoolYA', 'ColbyCovMMA', 'AviBerkow', 'BilldeBlasio', 'LonnekeEngel', 'VAKruta', 'mchooyah']
3.	FLOTUS.csv
	['user', 'FLOTUS', 'WhiteHouse', 'WhiteHouseHstry', 'BarackObama', 'realDonaldTrump', 'VP', 'SecondLady', 'POTUS']

a) Users discussing about same topic

TABLE IV

Twitter Users and Their Followers Discussing About Same Topic with Probabilities

Sr. no.	Twitter user and their followers	Probabilities
1.	'kimguilfoyle', 'TimRunsHisMouth'	-0.9214469282793404
2.	'IvankaTrump', 'Surabees'	-0.8315991287755492
3.	'DonaldJTrumpJr', 'dougstafford'	-0.3566544373516506
4.	'kimguilfoyle', 'WalidPhares'	-0.30050967549318747
5.	'kimguilfoyle', 'RajaFlores'	-0.2557175640279278
6.	'DonaldJTrumpJr', 'DevinNunes'	-0.24967235811028976
7.	'mike_pence', 'RyanAFournier'	-0.19863694962683967
8.	'mike_pence', 'VP'	-0.07791243918480027
9.	'realDonaldTrump', 'Scavino45'	0.08059212085693612
10.	'TrumpBabyUK', 'amvetsupport'	0.10202766196804715
11.	'Trump', 'GMatTRUMPNY'	0.21545873699029228
12.	'kimguilfoyle', 'JanitaKan'	0.23366878690650783
13.	'DonaldJTrumpJr', 'Doranimated'	0.25839186866215075
14.	'LaraLeaTrump', 'JoeTalkShow'	0.268561785101905
15.	'FLOTUS', 'VP'	0.26873412413292036
16.	'IvankaTrump', 'SenatorShaheen'	0.29909631125111175
17.	'kimguilfoyle', 'smerconish'	0.29951136683952484
18.	'FLOTUS', 'WhiteHouseHstry'	0.3091136692568364
19.	'IvankaTrump', 'SecArmy'	0.3331523223724417
20.	'IvankaTrump', 'theGESsummit'	0.3459069390773432

4) Installation of plurk_oauth library

```
(base) C:\Users\Vivian Lobo\Desktop\Vivian\Archive1\src\plurk>pip install plurk_oauth
Collecting plurk_oauth
  Downloading https://files.pythonhosted.org/packages/a8/62/7df71a6cf6038a30759d2b33aa5c77fcaa7ee8da0af795ebc6d4a9e1eef5/plurk_oauth-0.9.2-py3-none-any.whl
Collecting oauth2 (from plurk_oauth)
  Downloading https://files.pythonhosted.org/packages/a0/6f/86db603912ecd04109af952c38bc08928886cf0e34c723481fa7db98b4b5/oauth2-1.9.0.post1-py2.py3-none-any.whl
Requirement already satisfied: nose in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from plurk_oauth)
Requirement already satisfied: requests in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from plurk_oauth)
Collecting httplib2 (from oauth2->plurk_oauth)
  Downloading https://files.pythonhosted.org/packages/e8/b3/b34037575d6d75ff8dcfcf75315f56befbe409952be9f95c9b8cc9ee0499/httplib2-0.12.3-py3-none-any.whl (94kB)
    100% |████████████████████████████████| 102kB 82kB/s
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->plurk_oauth)
Requirement already satisfied: idna<2.7,>=2.5 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->plurk_oauth)
Requirement already satisfied: urllib3<1.23,>=1.21.1 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->plurk_oauth)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\vivian lobo\anaconda3\av\lib\site-packages (from requests->plurk_oauth)
Installing collected packages: httplib2, oauth2, plurk-oauth
Successfully installed httplib2-0.12.3 oauth2-1.9.0.post1 plurk-oauth-0.9.2
You are using pip version 9.0.1, however version 19.1.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.
```

Fig. 15 Plurk_oauth library installation

5) For Plurk: To search Plurk users and find followers of a Plurk user

TABLE V

Plurk Users with their respective IDs, Count, and Names of Friends

Plurk ID	Plurk users	Count of friends	IDs & names of friends
5079107	Trumpa	found 0 friends	
3604685	Trumpeeet	found 8 friends	3977232 cat-bat!
			3307343 ♥; AHG
			4283097 clupper
			3395776 Evelyn
			4098887 hanzhi
			4174104 honeydukes06
			4038644 SUPERgirl
4689322	trump806	found 2 friends	3395776 Evelyn
14119959	TrumpAndConquer	found 0 friends	
14070064	Trumptyte	found 0 friends	
11510640	Trumpcondos	found 9 friends	6412150 Omii_pLm
			13585687 A1astrology
			13814610 Maryasu
			13811608 acosmetic85
			3977232 cat-bat!
			3307343 ♥; AHG
			4283097 clupper
			3395776 Evelyn

6) For Plurk: To get timelines of the followers

TABLE VI
Plurk IDs of Users That Follow a Plurk User Through Timelines

Sr. no.	Plurk ID
1.	11510640
2.	13811608
3.	3604685
4.	4038644
5.	4428202
6.	3948291

7) For Plurk: LDA application

TABLE VII
Plurk IDs and Their Respective Followers Through LDA Application

Sr. no	Plurk ID and their respective followers
1.	11510640
	['user', '11510640', '13811608']
2.	3604685
	['user', '3604685', '4038644']
3.	4428202
	['user', '4428202', '3948291']
4.	6155566
	['user', '6155566', '3415748']
5.	7205924
	['user', '7205924', '5453739']
6.	7448468
	['user', '7448468', '3779245']

a) Users discussing about same topic:

TABLE VIII
Plurk IDs and Their Respective Follower IDs Discussing About Same Topic Along With Their Probabilities

Sr. no.	Plurk IDs and their followers	Probabilities
1.	'6155566', '3415748'	-0.6641188575103365
2.	'4428202', '3948291'	0.07728450444334724
3.	'7448468', '3779245'	1.6823931966488148
4.	'3604685', '4038644'	2.0133803945872986
5.	'7205924', '5453739'	3.131833195469799
6.	'11510640', '13811608'	6.17876217974276

V. RESULTS AND DISCUSSION

Figure 16 shows graphical representation of core users possessing similar interests with probabilities on y-axis. As evident from the figure, Twitter usernames are plotted on x-axis such as CLewandowski, realDonanldTrump, Scavino45, DanScavino, FoxNews, seanspicer, KeithSchiller45, amongst others. Users discussing about the same topic are shown by lines of various colors.

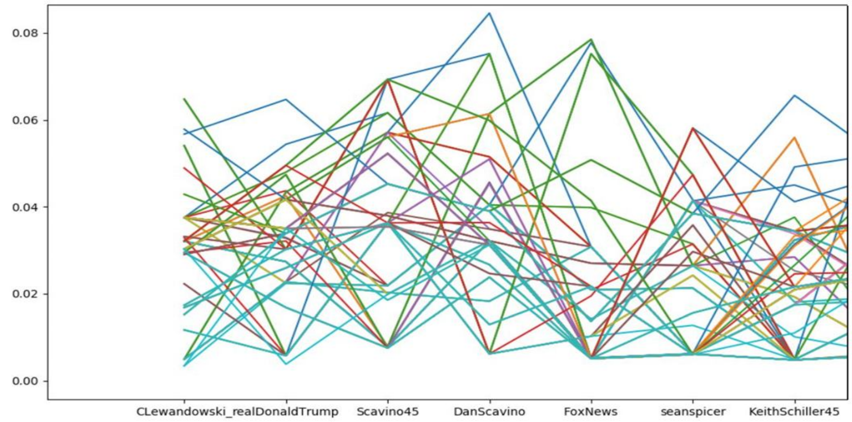

Fig. 16 Graphical representation of core users possessing similar interests with probabilities on *y*-axis

Figure 17 shows graphical representation of core users possessing similar interests with probabilities on x-axis. As evident from the figure, Twitter usernames are plotted on y-axis such as CLewandowski, realDonanldTrump, Scavino45, FDRLST, Heminator, EricTrump, Discovery, BenLecomteSwim, amongst others. Users discussing about the same topic are shown by lines of various colors.

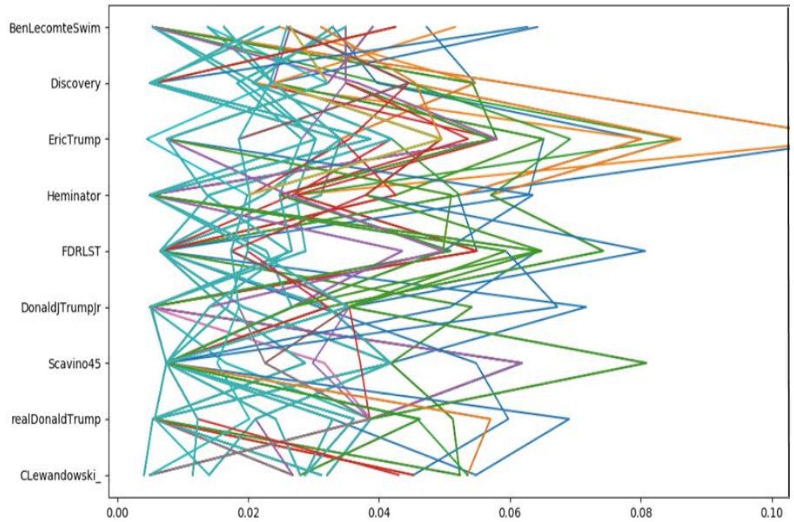

Fig. 17 Graphical representation of core users possessing similar interests with probabilities on *x*-axis

VI. CONCLUSION AND FUTURE SCOPE

The developed system proves to be an applicable technique for analyzing the dynamics, structure, and content of two well-known social networking sites, i.e., Twitter and Plurk. The system helps in understanding and identifying core users that possess similar interests in the abovementioned social networking sites through the application of LDA algorithm and KL divergence technique. Identifying core users possessing similar interests will help social analysts to a large extent to understand the patterns of core users and their interests so that suitable decisions can be taken by them in near future. While this research work has demonstrated the potential of efficiently analyzing the dynamics, structure, and content of two well-known social networking sites, i.e., Twitter and Plurk, many opportunities for extending the scope of this research work remain. In future, we plan to use some other algorithms such as DCIM algorithm and JS technique. We also plan to implement the developed system keeping in mind some other factors such as we will consider only those Tweets and Plurk messages that are tweeted by popular users, which means only those Tweets and Plurk messages will be considered for analyzing the dynamics, structure, and content of Twitter and Plurk users who have maximum number of followers. Also, in the near future, we plan to fetch Tweets and Plurk messages for a particular Twitter and Plurk user on the basis of location so that we can efficiently analyze the dynamics, structure, and content of Twitter and Plurk. Moreover, we would include a few more social networking sites such as Facebook, Instagram, and Netlog for analysis purposes.

REFERENCES

[1] "What is the purpose of social networking?" [Online] Available: https://brainly.in/question/4013138 [Accessed on June 17, 2019].
[2] "EduRev Infinity" [Online] Available: https://edurev.in/question/526480/Write-a-short-note-on-social-networking--Persona of Social Networking in Computing and Informatics Era [Accessed on June 17, 2019].
[3] "The definition of social networking", [Online] Available: https://zeltser.com/definition-of-social-networking/ [Accessed on June 17, 2019].
[4] B. Tejas, M. Rajkumar, K. Sushan, and K. Chandrasekaran, "Trust management in ad-hoc networks: A social network-based approach," Network and Complex Systems, vol. 1, no. 1, pp. 24–32, 2011.
[5] L. Zhou, W. P. Wu, and X. Luo, "Internationalization and the performance of born-global SMEs: The mediating role of social networks," Journal of International Business Studies, vol. 38, no. 4, pp. 673–690, 2007.
[6] R. Zafarani and H. Liu, "Connecting users across social media sites: a behavioral-modeling approach," In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 41–49, 2013.
[7] S. P. Marsh, "Formalising trust as a computational concept", University of Stirling, 1994.
[8] A. Abdul-Rahman and S. Hailes, "Supporting trust in virtual communities," In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, IEEE, pp. 1–9, 2000.
[9] M. Schillo, P. Funk, M. Rovatsos "Using trust for detecting deceitful agents in artificial societies," Applied Artificial Intelligence, vol. 14, no. 8, pp. 825–848, 2000.
[10] B. Esfandiari and S. Chandrasekharan "On how agents make friends: Mechanisms for trust acquisition," In Proceedings of the fourth workshop on deception, fraud and trust in agent societies, vol. 222, p. 19, 2001.
[11] B. Yu and M. P. Singh, "Distributed reputation management for electronic commerce," Computational Intelligence, vol. 18, no. 4, pp. 535–549, 2002.
[12] L. Mui, M. Mohtashemi, and A. Halberstadt, "A computational model of trust and reputation," In Proceedings of the 35th Annual Hawaii International Conference on System Sciences, IEEE, pp. 2431–2439, 2002.
[13] Y. Nie, Y. Jia, S. Li, X. Zhu, A. Li, and B. Zhou, "Identifying users across social networks based on dynamic core interests", Neurocomputing, vol. 210, pp. 107–115, 2016.
[14] S. Kumar, R. Zafarani, and H. Liu, "Understanding user migration patterns in social media," In Twenty-Fifth AAAI Conference on Artificial Intelligence, pp. 1204–1209, 2011.
[15] S. Liu, S. Wang, F. Zhu, J. Zhang, and R. Krishnan, "Hydra: Large-scale social identity linkage via heterogeneous behavior modeling," In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, ACM, pp. 51–62, 2014.
[16] G. Cao and L. Kuang, "Identifying core users based on trust relationships and interest similarity in recommender system," In 2016 IEEE International Conference on Web Services (ICWS), IEEE, pp. 284–291, 2016.
[17] D. M. Blei and A. Y. Ng, and M. I. Jordan "Latent Dirichlet allocation," Journal of Machine Learning Research, pp. 993–1022, 2003.