International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 46 • 2016

Structuring Circle and Lists in Social Media by Identifying Effective Attributes using Principal Component Analysis and R

Nisha Sethi[a], Harshpreet Singh[b] and Deep Mann[c]

[a]*School of Computer Applications, Lovely Professional University, Punjab, India*
[b,c]*School of Computer Science and Engineering, Lovely Professional University, Punjab, India*

Abstract: Big Data refers to large amount of data generated from heterogeneous and autonomous resources. Now days, social media is a great source of data which aims in providing information about the individual and their respective behavior. The user attributes of user profile in social data also provide a structure to form various social networks. The work carried out in this paper focuses on analyzing the social data for finding attributes which are effective in making circles and lists in social media. Principal Component Analysis is mathematical procedure to reduce number of dimensions of dataset by maintaining its original variability. This approach is utilized for isolating the attributes effective in making circles. The paper aims in providing the essential components using various dimensions such as first name, last name, school, place, location, birthday, degree, class, education etc. which aids in building the social network. Interpretation and validation of the proposed methodology will be plotted by scree test in R programming environment.

Keywords: Social Structure, Circle/Lists, Principal Component Analysis, Scree test, Social media, Big Data, Social Data.

1. INTRODUCTION

Big Data refers to large, complex and growing volume of data sets with evolving relationships from heterogeneous and autonomous sources. Over 2.5 quintillion bytes of data are created every day and 90% of the present data has been created in the last two years [1] [2].

Big Data is generated from number of data intensive application like online discussions, Flicker (public picture sharing site), sensors, online shopping sites, social media giants (facebook, twitter, LinkedIn, YouTube, Google+ and more), scientific data analysis, mobile devices and more [3] [4].

Each day Google has 1 billion above queries, Twitter has 250 million above tweets, facebook has 800 million above updates, and YouTube has more than 4 billion views per day. Nowadays data is produced in zettabytes (10^{21} bytes). Data produced by social media sites have major contribution towards Big Data [5].

There are many social media sites, such as Google+, facebook, Twitter generating enormous data[6]. Social networking sites have hundreds of millions of users which are increasing rapidly. Users in social networking sites join network, create profiles and relationships with any users of same social network. Social networks are tool for connecting people, mirroring real-life relationships and building societies. Users of social media maintain profile information like name, location, education, birthday, class, school, hometown, and much more[7].

Social Network is powerful means of sharing, organizing and finding contents and contacts[8]. Social networks revolve around users and users group friends using the mechanism of circles in Google+ and lists in facebook [9]. Social networking sites allow users to categorize their friends manually into their social groups, circles and in their social lists by providing recommendations[10]. Manual categorization of friends in lists and circles is time consuming and lengthy [11]. Circles can be used for content filtering, privacy and sharing data between its users [12]. Circles can be used to significantly increase the relevance of what is shown to the user for advertising and marketing of products[13].

Data is collected from social networking sites about user through different attributes like "tag", "comments", "like", "status", "photos" and "video" which is referred as social media data[14][15]. This data is the basis for creating models of the relationships between users [16].

As the number of components increases with the information provided by the user, it is difficult to keep all components for building social circles[17]. This paper aims at identifying the primary components depending on which the social circles can be build and the users in a circle or list can be recommended. For building social circles we firstly investigate the components which can be included for the formation of social circles.

Principal Component Analysis (PCA) [18] is analyzed and used as it has the ability to transform number of correlated variables into smaller number of uncorrelated variables called principal components. The goal is to reduce the number of variables and not changing the original description of the data [19]. PCA is equivalent in finding direction axis which have maximum variance and using these new directions to define new basis for constructing social circles

Many researchers have focused on structuring the circles with node formation based on the similar attributes [20]. In our previous work we investigated the approach for the formation of circles using the principal components[21]. The user can himself choose the component or can be guided by the system to build circles on particular components.

The paper is organized as follows: Section 2 provides an overview of the proposed methodology which extends our previous work. Section 3 discusses the experimental setup for the methodology. Section 4 illustrates the result computed using the R tool and Section 5 discusses the interpretation of the result. Section 6 concludes the result with future research directions.

2. PROPOSED WORK

In social media, user attributes are used as profile information which help user to form social circles[22]. User add friends into circles and lists by categorizing them using profile information and making it easy to filter the friends. The user in circle has some particular number of attributes, which help them to make circle. The paper proposes an approach on finding attributes which effectively contribute in making the circle using PCA.

Consider n as numbers of attributes which define a user in social media for creating circles. PCA helps to identify the attributes which are meaningful and reduce dimensionality of data. With PCA the complexity of data can be reduced and need less number of plots to analyze. PCA is mathematical procedure to reduce number of dimensions (attributes) of the dataset but maintaining its original variability. Instead of working on all n-dimensions, first perform PCA on original data (n-dimensions) and then use only first few components say PC_1, PC_2, ..., PC_m in analysis where $m < n$. These reduced dimensions can be used to focus on making circle.

Circle is made with users and users have number of attributes. Focusing on all the attributes is a lengthy and costly computation. With reduced attributes which can be identified using PCA, analysis can be easy as the focus is on those attributes which are efficient in making circles.

We extend the previous mathematical study by implementation of PCA for finding effective attribute using R programming [23]. R programming is an environment and programming language used mainly for statistical computation and graphical representations. It is widely used for data mining, data analysis and statistical inferences. R is equipped with number of techniques used for classification, clustering, and performing statistics function on data. R can be extended with various add-on packages for exploratory analysis.

3. EXPERIMENTAL SETUP

PCA is the distribution of variation of the multivariate dataset across components in such a way to find patterns that may not observe using other analysis and graphics technique. It is a procedure that transforms number of correlated variables into smaller number of uncorrelated variables called principal components. The goal is to reduce the number of variables but retain original meaning of the data. The entire process of identifying effective attributes in structuring social circles in social media can be represented in the following steps.

A. Preparing Data

The initial phase of analysis of social media data is collection of data. The data for analysis is collected from profile information of users and their circles. In social media data, profile information of user has many dimensions such as name, school, place, location, birthday, degree, class, school, name, hometown etc.

Consider the dataset with 224 attributes which are used for building the profile of the user in Google+ and facebook [24]. The dataset consists of number of circles, their users and attributes of the users. Attribute value is either '1' or '0'. Attribute value '1' means attribute consists of a value and attribute value '0' means attribute does not contain any value. The attribute values are anonymized for privacy concerns. Profile information of a user is represented in tree structure as given in Figure 1 which consists of the attributes as described in the dataset.

(a) Cleaning of data: The attributes which have zero users are removed from dataset. After cleaning dataset dimension will be reduced.

(b) Scaling of data: Scaling of data is to standardize and centralize the data. Scaled data can be obtained by subtracting mean from each measurement which will result data with zero mean.

B. Calculate the Covariance Matrix between users of Circle and Attributes of Profile Information

A *c*ovariance can be used to measure how much two variables change together; covariance can be positive or negative. In PCA covariance is positive covariance, close to zero. Variance is special case of covariance, when variables are similar. In PCA variance are maximized and covariance are minimized.

C. Compute Eigen Values and Eigen Vector from the Covariance Matrix

The highest Eigen value is first principal component (PC1), the second highest is second principal component (PC2), and so on.

D. Compute Principal Components

The principal components are Eigen vectors of covariance matrix, and are in decreasing order of 'importance'. The first Eigen vector is more meaning full as compared to other Eigen vectors.

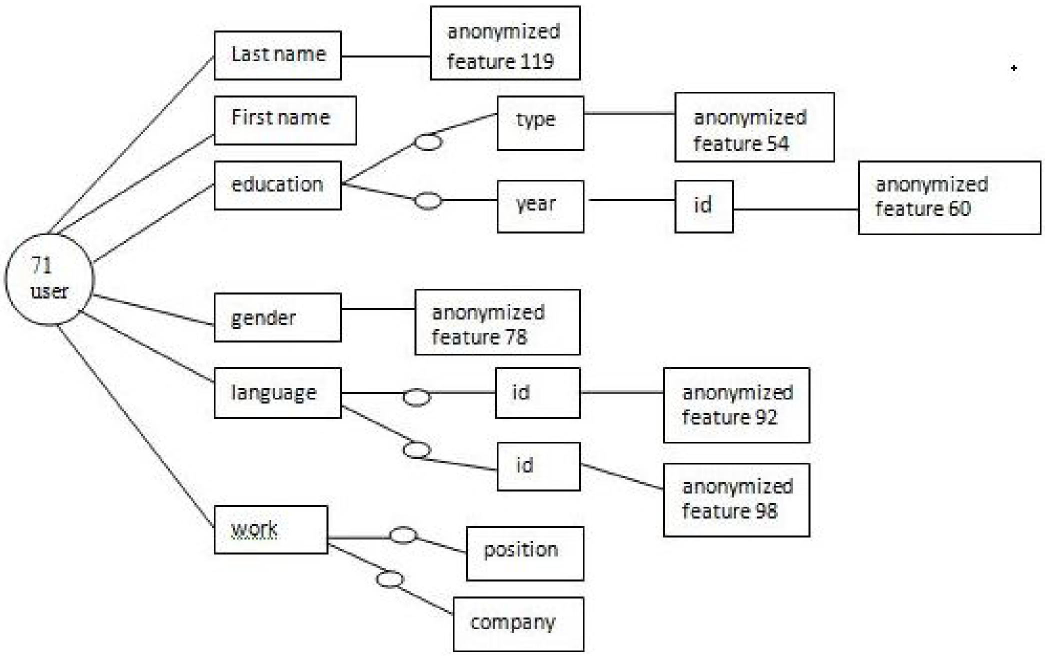


Figure 1: User71 attribute tree

Table 1
Attribute present in user 71 and attribute name

Attribute Present	*Attribute Name*
1	Lastname:anonymizedfeature119
0	First name
1	Education:type:anonymizedfeature54
1	Education:year:id:anonymizedfeature60
1	Gender:anonymizedfeature78
1	Language:id:anonymizedfeature92
1	Language:id:anonymizedfeature98
0	Work: position
0	Work: company

E. Result Interpretation

The results of PCA are plotted in R programming. The plots show the attributes for which users of circle0 are highly correlated.

4. RESULTS AND FINDINGS

The step wise results of the various identifying attribute phases are as follows:

A. Preparing Data

The dataset consists of 24 circles. Every circle is formed from at least one user and ranging till 133 users as given in Table 2. Let us take an example from dataset circle1 contains only one user and circle15 contains 133 users. The same user can be in different circles or in one circle. User258 in dataset is present in circle4 and circle16.Circles tend to overlap is they comprise of users with similar id. As in the dataset circle 1 and 16, 8 and 20 overlap each other as shown in Table 3.

The circle8 present in dataset contains only one user that is user282 and another circle20 has many users, one user among circle20 is user282 as shown in Table 3. The users which are not connected in form of friends or not having a common attribute cannot form the circle. The user should be connected to each other directly or indirectly. The study shows PCA reduce the dimensionality of social data.

The dataset consists of 224 attributes for each user, making it difficult to work on all 224 attributes in formation of circle. In order to analyze the data PCA provides a way to find meaningful attributes which have effective contribution in making circles.

Table 2
Circle Names with their User Count

Circle name	*No. of users in circle*
Circle 0	20
Circle 1	1
Circle 2	9
Circle 3	3
Circle 4	17
Circle 5	1
Circle 6	20
Circle 7	2
Circle 8	1
Circle 9	10
Circle 10	4
Circle 11	30
Circle 12	1
Circle 13	5
Circle 14	2
Circle 15	133
Circle 16	32
Circle 17	9
Circle 18	1
Circle 19	13
Circle 20	6
Circle 21	1
Circle 22	1
Circle 23	3

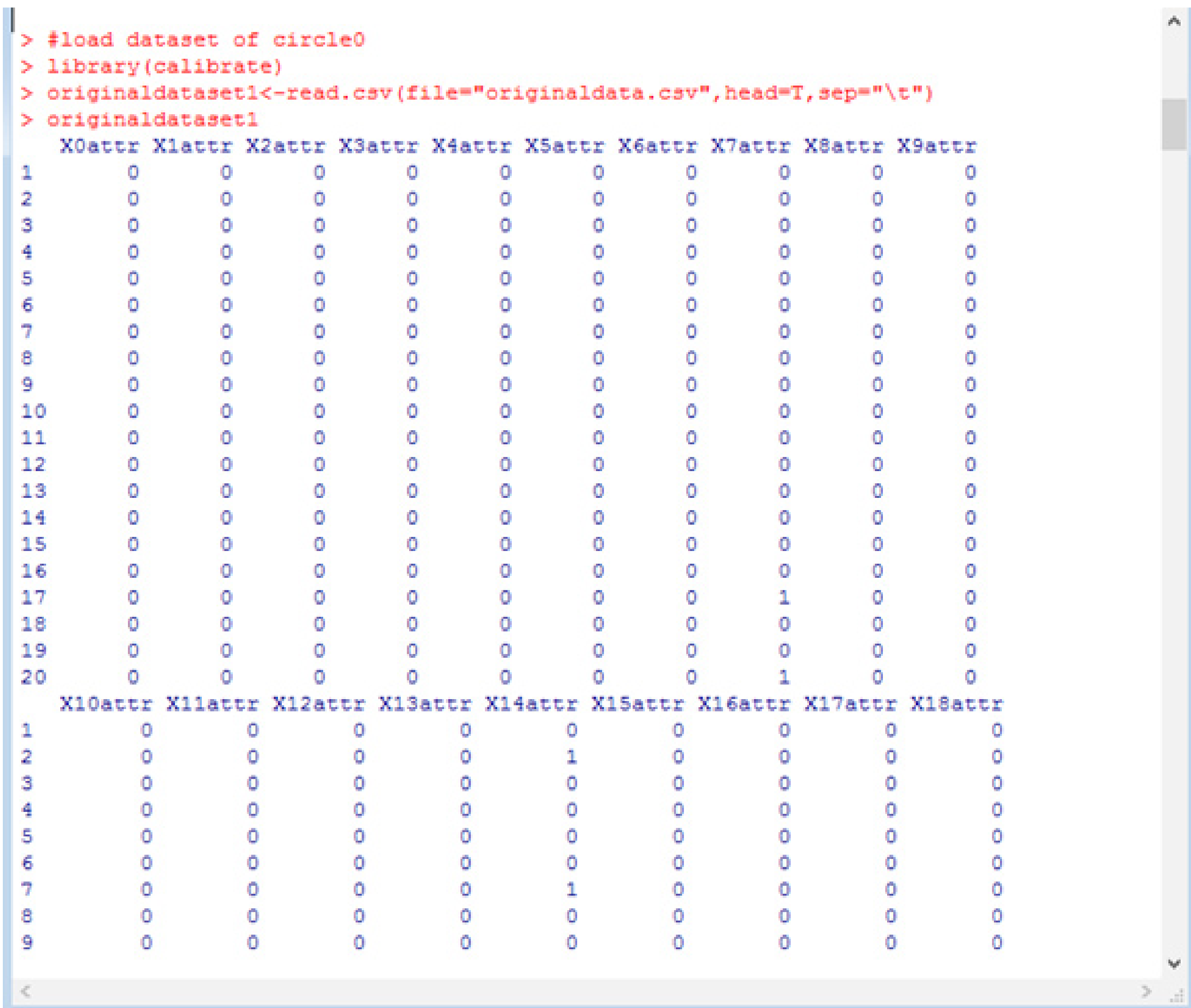
Considering the above dataset, Let us consider circle0 for our analysis. Circle0 consists of 20 users with the user ids as user71, user215, user54, user61, user298, user229, user81, user253, user193, user97, user264, user29, user132, user110, user163, user259, user183, user334, user245, user222. Implementation of PCA is done in R programming. The users along with their attributes are given in Figure.2.

As it can be observed, the attributes form columns and users form rows of circle0 dataset. The first row are the attribute values of user71, second row are attribute values of user215 and last 20^{th} row are attribute values of user 222.

Table 3
Circle with their Users

Circle name	*User ids of circles*
Circle 4	125,344,295,257,55,122,223,59,268,280,84,156, 258,236,250,239
Circle 16	251,94,330,5,34,299,254,24,180,194,281,101,266,135,197, 173,36,9,85,57,37,258,309,80,139, 202,187,249,58,127,48,92
Circle 1	173
Circle 8	282
Circle 20	244,282,262,293,220,174

The dataset contains 224 attributes for each 20 users of circle0. The cleansing of data is removal of those attributes which have zero number of users in circle0. After cleansing 73 attributes of profile information of circle0 remains. Cleaned dataset is given in Figure 3.



```
> #load dataset of circle0
> library(calibrate)
> originaldataset1<-read.csv(file="originaldata.csv",head=T,sep="\t")
> originaldataset1
   X0attr X1attr X2attr X3attr X4attr X5attr X6attr X7attr X8attr X9attr
1       0      0      0      0      0      0      0      0      0      0
2       0      0      0      0      0      0      0      0      0      0
3       0      0      0      0      0      0      0      0      0      0
4       0      0      0      0      0      0      0      0      0      0
5       0      0      0      0      0      0      0      0      0      0
6       0      0      0      0      0      0      0      0      0      0
7       0      0      0      0      0      0      0      0      0      0
8       0      0      0      0      0      0      0      0      0      0
9       0      0      0      0      0      0      0      0      0      0
10      0      0      0      0      0      0      0      0      0      0
11      0      0      0      0      0      0      0      0      0      0
12      0      0      0      0      0      0      0      0      0      0
13      0      0      0      0      0      0      0      0      0      0
14      0      0      0      0      0      0      0      0      0      0
15      0      0      0      0      0      0      0      0      0      0
16      0      0      0      0      0      0      0      0      0      0
17      0      0      0      0      0      0      0      1      0      0
18      0      0      0      0      0      0      0      0      0      0
19      0      0      0      0      0      0      0      0      0      0
20      0      0      0      0      0      0      0      1      0      0
   X10attr X11attr X12attr X13attr X14attr X15attr X16attr X17attr X18attr
1        0       0       0       0       0       0       0       0       0
2        0       0       0       0       1       0       0       0       0
3        0       0       0       0       0       0       0       0       0
4        0       0       0       0       0       0       0       0       0
5        0       0       0       0       0       0       0       0       0
6        0       0       0       0       0       0       0       0       0
7        0       0       0       0       1       0       0       0       0
8        0       0       0       0       0       0       0       0       0
9        0       0       0       0       0       0       0       0       0
```

Figure 2: Social Media Dataset

```
> #load data in R
> dataset1<-read.csv(file="dataset1.csv",head=T,sep="\t")
> dataset1
   X7attr X14attr X23attr X32attr X43attr X50attr X52attr X53attr X54attr
1       0       0       0       0       0       0       0       1       1
2       0       1       0       0       0       0       0       0       1
3       0       0       0       0       0       0       0       0       0
4       0       0       0       0       0       0       0       1       0
5       0       0       0       0       0       1       0       1       0
6       0       0       0       0       0       1       0       1       0
7       0       1       1       0       0       0       1       1       1
8       0       0       0       0       0       0       0       1       0
9       0       0       0       0       0       0       0       0       0
10      0       0       0       0       0       0       0       0       0
11      0       0       0       0       0       0       0       0       0
12      0       0       0       0       0       1       0       1       0
13      0       0       0       0       1       1       0       1       0
14      0       0       0       0       0       0       0       0       0
15      0       0       0       0       0       0       1       0       1
16      0       0       0       0       0       0       0       0       0
17      1       0       0       0       0       0       1       1       0
18      0       0       0       0       0       1       0       1       0
19      0       0       0       1       0       0       0       0       1
20      1       0       0       0       0       1       0       1       0
   X55attr X58attr X59attr X60attr X63attr X65attr X66attr X68attr X77attr
1        0       0       0       1       0       1       0       0       0
2        0       0       0       0       0       0       0       0       0
3        0       0       0       0       0       0       0       0       0
4        0       0       0       0       0       0       0       1       1
5        1       0       1       0       0       1       0       0       0
6        1       0       1       0       0       1       0       0       1
7        1       0       0       0       1       0       0       0       1
8        0       0       0       0       0       0       0       0       1
9        1       0       0       0       0       0       0       0       0
10       0       0       0       0       0       0       0       0       0
```

Figure 3: Cleaned dataset

As it can be observed, the number of columns is reduced after cleaning. The Figure 4 shows the dimensions of dataset before and after cleansing. From the Figure 4, it is observed that first dimension is number of users in circle and the second dimension is number of attributes. The number of attributes reduces from 224 to 73 after cleaning. After cleaning the dataset is standardizing in R**.** The command to standardized dataset in R and the scaled data is given in Figure 5.

```
> #dimension of dataset before cleaning
> dim(originaldataset)
[1]  20 224
> #dimension of dataset after cleaning
> dim(dataset1)
[1] 20 73
> 
```

Figure 4: Dimension of original dataset and cleansed dataset

As observed from above Figure.6, the mean of scaled dataset is approximately near to zero.

```
> # Scale the cleansed dataset of circle0
> standardize <- function(x) {(x - mean(x))}
> my.stddata = apply(dataset1,2,function(x) (x-mean(x)))
> my.stddata
      X7attr X14attr X23attr X32attr X43attr X50attr X52attr X53attr X54attr X55attr
 [1,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15    0.45    0.75   -0.45
 [2,]   -0.1     0.9   -0.05   -0.05   -0.05    -0.3   -0.15   -0.55    0.75   -0.45
 [3,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15   -0.55   -0.25   -0.45
 [4,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15    0.45   -0.25   -0.45
 [5,]   -0.1    -0.1   -0.05   -0.05   -0.05     0.7   -0.15    0.45   -0.25    0.55
 [6,]   -0.1    -0.1   -0.05   -0.05   -0.05     0.7   -0.15    0.45   -0.25    0.55
 [7,]   -0.1     0.9    0.95   -0.05   -0.05    -0.3    0.85    0.45    0.75    0.55
 [8,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15    0.45   -0.25   -0.45
 [9,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15   -0.55   -0.25    0.55
[10,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15   -0.55   -0.25   -0.45
[11,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15   -0.55   -0.25   -0.45
[12,]   -0.1    -0.1   -0.05   -0.05   -0.05     0.7   -0.15    0.45   -0.25    0.55
[13,]   -0.1    -0.1   -0.05   -0.05    0.95     0.7   -0.15    0.45   -0.25    0.55
[14,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15   -0.55   -0.25   -0.45
[15,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3    0.85   -0.55    0.75   -0.45
[16,]   -0.1    -0.1   -0.05   -0.05   -0.05    -0.3   -0.15   -0.55   -0.25   -0.45
[17,]    0.9    -0.1   -0.05   -0.05   -0.05    -0.3    0.85    0.45   -0.25    0.55
[18,]   -0.1    -0.1   -0.05   -0.05   -0.05     0.7   -0.15    0.45   -0.25    0.55
[19,]   -0.1    -0.1   -0.05    0.95   -0.05    -0.3   -0.15   -0.55    0.75   -0.45
[20,]    0.9    -0.1   -0.05   -0.05   -0.05     0.7   -0.15    0.45   -0.25    0.55
      X58attr X59attr X60attr X63attr X65attr X66attr X68attr X77attr X78attr
 [1,]   -0.05   -0.15    0.95   -0.15     0.7   -0.05   -0.05   -0.55    0.55
 [2,]   -0.05   -0.15   -0.05   -0.15    -0.3   -0.05   -0.05   -0.55    0.55
 [3,]   -0.05   -0.15   -0.05   -0.15    -0.3   -0.05   -0.05   -0.55    0.55
 [4,]   -0.05   -0.15   -0.05   -0.15    -0.3   -0.05    0.95    0.45   -0.45
 [5,]   -0.05    0.85   -0.05   -0.15     0.7   -0.05   -0.05   -0.55    0.55
 [6,]   -0.05    0.85   -0.05   -0.15     0.7   -0.05   -0.05    0.45   -0.45
 [7,]   -0.05   -0.15   -0.05    0.85    -0.3   -0.05   -0.05    0.45   -0.45
 [8,]   -0.05   -0.15   -0.05   -0.15    -0.3   -0.05   -0.05    0.45   -0.45
 [9,]   -0.05   -0.15   -0.05   -0.15    -0.3   -0.05   -0.05   -0.55    0.55
```

Figure 5: Scaled Dataset

```
> mean(my.stddata)
[1] -3.608801e-18
>
```

Figure 6: Mean of scaled dataset of circle0

B. Calculation of Covariance Matrix between users of Circle and Attributes of Profile Information

A covariance can be used to measure how much two variables change together; covariance can be positive or negative. Variance is special case of covariance, when variables are similar. The covariance between users of circle and their attributes is given in Figure 7 and variance of dataset is given in Figure 8. The total number of attributes of user profile is 73 after cleansing. So, the covariance matrix for circle0 will be 73x73 matrix.

```
> covariance2=cov(my.stddata)
> # calculate covariance
> covariance2=cov(my.stddata)
> covariance2
               X7attr       X14attr       X23attr       X32attr       X43attr
X7attr    0.094736842 -0.010526316 -0.005263158 -0.005263158 -0.005263158
X14attr  -0.010526316  0.094736842  0.047368421 -0.005263158 -0.005263158
X23attr  -0.005263158  0.047368421  0.050000000 -0.002631579 -0.002631579
X32attr  -0.005263158 -0.005263158 -0.002631579  0.050000000 -0.002631579
X43attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579  0.050000000
X50attr   0.021052632 -0.031578947 -0.015789474 -0.015789474  0.036842105
X52attr   0.036842105  0.036842105  0.044736842 -0.007894737 -0.007894737
X53attr   0.047368421 -0.005263158  0.023684211 -0.028947368  0.023684211
X54attr  -0.026315789  0.078947368  0.039473684  0.039473684 -0.013157895
X55attr   0.057894737  0.005263158  0.028947368 -0.023684211  0.028947368
X58attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X59attr   0.036842105 -0.015789474 -0.007894737 -0.007894737 -0.007894737
X60attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X63attr  -0.015789474  0.036842105  0.044736842  0.044736842 -0.007894737
X65attr   0.073684211 -0.031578947 -0.015789474 -0.015789474 -0.015789474
X66attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X68attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X77attr   0.047368421 -0.005263158  0.023684211  0.023684211 -0.028947368
X78attr  -0.047368421  0.005263158 -0.023684211 -0.023684211  0.028947368
X84attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X90attr   0.042105263 -0.010526316 -0.005263158 -0.005263158 -0.005263158
X91attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X92attr   0.057894737  0.005263158  0.028947368 -0.023684211  0.028947368
X93attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579  0.050000000
X94attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X98attr  -0.010526316 -0.010526316 -0.005263158 -0.005263158  0.047368421
X100attr  0.042105263 -0.010526316 -0.005263158 -0.005263158 -0.005263158
X103attr -0.005263158  0.047368421  0.050000000 -0.002631579 -0.002631579
X106attr -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X126attr  0.036842105 -0.015789474 -0.007894737 -0.007894737 -0.007894737
```

Figure 7: Covariance between users of circle and their attributes

```
> var1=var(my.stddata)
> var1
               X7attr       X14attr       X23attr       X32attr       X43attr
X7attr    0.094736842 -0.010526316 -0.005263158 -0.005263158 -0.005263158
X14attr  -0.010526316  0.094736842  0.047368421 -0.005263158 -0.005263158
X23attr  -0.005263158  0.047368421  0.050000000 -0.002631579 -0.002631579
X32attr  -0.005263158 -0.005263158 -0.002631579  0.050000000 -0.002631579
X43attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579  0.050000000
X50attr   0.021052632 -0.031578947 -0.015789474 -0.015789474  0.036842105
X52attr   0.036842105  0.036842105  0.044736842 -0.007894737 -0.007894737
X53attr   0.047368421 -0.005263158  0.023684211 -0.028947368  0.023684211
X54attr  -0.026315789  0.078947368  0.039473684  0.039473684 -0.013157895
X55attr   0.057894737  0.005263158  0.028947368 -0.023684211  0.028947368
X58attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X59attr   0.036842105 -0.015789474 -0.007894737 -0.007894737 -0.007894737
X60attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X63attr  -0.015789474  0.036842105  0.044736842  0.044736842 -0.007894737
X65attr   0.073684211 -0.031578947 -0.015789474 -0.015789474 -0.015789474
X66attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X68attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X77attr   0.047368421 -0.005263158  0.023684211  0.023684211 -0.028947368
X78attr  -0.047368421  0.005263158 -0.023684211 -0.023684211  0.028947368
X84attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X90attr   0.042105263 -0.010526316 -0.005263158 -0.005263158 -0.005263158
X91attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X92attr   0.057894737  0.005263158  0.028947368 -0.023684211  0.028947368
X93attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579  0.050000000
X94attr  -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X98attr  -0.010526316 -0.010526316 -0.005263158 -0.005263158  0.047368421
X100attr  0.042105263 -0.010526316 -0.005263158 -0.005263158 -0.005263158
X103attr -0.005263158  0.047368421  0.050000000 -0.002631579 -0.002631579
X106attr -0.005263158 -0.005263158 -0.002631579 -0.002631579 -0.002631579
X126attr  0.036842105 -0.015789474 -0.007894737 -0.007894737 -0.007894737
X127attr -0.036842105  0.015789474  0.007894737  0.007894737  0.007894737
X128attr  0.036842105  0.036842105  0.044736842 -0.007894737 -0.007894737
```

Figure 8: Variance of social dataset

```
> # Eigen values and Eigen vectors calculated from covariance matrix
> eigen<-eigen(covariance2)
> eigen
$values
 [1]  1.241971e+00  8.753604e-01  8.200766e-01  5.378652e-01  5.293768e-01
 [6]  4.967103e-01  4.539936e-01  3.666376e-01  3.185543e-01  2.333138e-01
[11]  2.214553e-01  1.875918e-01  1.351862e-01  1.220965e-01  9.333904e-02
[16]  5.041922e-02  1.676987e-02  1.507175e-02  8.038853e-17  6.618602e-17
[21]  5.520305e-17  5.428814e-17  4.468017e-17  3.600871e-17  2.687111e-17
[26]  2.548367e-17  1.936004e-17  1.625108e-17  1.617109e-17  1.334060e-17
[31]  1.310471e-17  1.084108e-17  1.036508e-17  9.816341e-18  9.535742e-18
[36]  9.254534e-18  8.856898e-18  6.618498e-18  5.564902e-18  5.235699e-18
[41]  4.980899e-18  4.870975e-18  4.179371e-18  1.923539e-18  1.881337e-18
[46]  1.169492e-18  1.101511e-18  9.713395e-19  7.197881e-33 -4.090171e-19
[51] -4.764321e-19 -7.773330e-19 -1.172619e-18 -1.560783e-18 -3.552571e-18
[56] -3.755520e-18 -3.870887e-18 -4.454883e-18 -5.417394e-18 -5.873575e-18
[61] -5.925330e-18 -6.345657e-18 -6.399160e-18 -7.163053e-18 -1.032500e-17
[66] -1.204956e-17 -1.528532e-17 -1.705940e-17 -2.448128e-17 -2.882595e-17
[71] -4.390955e-17 -7.843500e-17 -1.479479e-16

$vectors
             [,1]         [,2]          [,3]        [,4]         [,5]          [,6]
 [1,] -0.137858420  0.018017805  1.328275e-01  0.01133074 -0.102759909  0.0139591608
 [2,]  0.067473308  0.007427536  9.935037e-02  0.09493996  0.077187836  0.0898493381
 [3,]  0.005686553  0.010116842  1.213495e-01  0.06708495  0.071919642  0.0755494767
 [4,]  0.044210463  0.043537134  1.758458e-02 -0.01525419  0.007516163 -0.0360197438
 [5,] -0.041623309 -0.054202606 -6.645646e-02  0.02511333 -0.088724765  0.1716040631
 [6,] -0.358768311 -0.014870206 -1.838545e-01  0.07471618  0.009179467  0.0532066446
 [7,]  0.027164485  0.014172486  2.993277e-01  0.24557590  0.079304309  0.1382591353
 [8,] -0.362297012 -0.098696995  1.128556e-01 -0.02400458  0.202695204 -0.0077165524
 [9,]  0.196668336 -0.126492408  1.862733e-01  0.04083839  0.215260636  0.1129835027
[10,] -0.372260222 -0.034077028  7.154461e-02  0.24685733 -0.027841173  0.2020864714
[11,] -0.058157333  0.097471159 -5.945880e-02 -0.05374607  0.186560037  0.1308576551
[12,] -0.206039759 -0.065996395 -9.150093e-02  0.11348419 -0.055337302 -0.1803770546
[13,]  0.016345130 -0.177692477  3.345153e-02 -0.15103152  0.116440143 -0.0061174360
[14,]  0.118536450  0.053889375  1.748210e-01  0.16401489  0.093552298  0.1048010773
[15,] -0.289804041 -0.232010991  1.176036e-01  0.01862422  0.021052514 -0.2579341953
[16,] -0.058157333  0.097471159 -5.945880e-02 -0.05374607  0.186560037  0.1308576551
```

Figure 9: Eigen values and Eigen vectors of circle0

```
> #show diagonal values of variance of circle0 cleansed dataset
> diag(variance)
    X7attr    X14attr    X23attr    X32attr    X43attr    X50attr    X52attr    X53attr    X54attr
0.09473684 0.09473684 0.05000000 0.05000000 0.05000000 0.22105263 0.13421053 0.26052632 0.19736842
   X55attr    X58attr    X59attr    X60attr    X63attr    X65attr    X66attr    X68attr    X77attr
0.26052632 0.05000000 0.13421053 0.05000000 0.13421053 0.22105263 0.05000000 0.05000000 0.26052632
   X78attr    X84attr    X90attr    X91attr    X92attr    X93attr    X94attr    X98attr   X100attr
0.26052632 0.05000000 0.09473684 0.05000000 0.26052632 0.05000000 0.05000000 0.09473684 0.09473684
  X103attr   X106attr   X126attr   X127attr   X128attr   X129attr   X133attr   X134attr   X138attr
0.05000000 0.05000000 0.13421053 0.13421053 0.13421053 0.09473684 0.09473684 0.05000000 0.13421053
  X139attr   X141attr   X144attr   X148attr   X149attr   X152attr   X153attr   X156attr   X160attr
0.05000000 0.13421053 0.05000000 0.05000000 0.05000000 0.05000000 0.05000000 0.09473684 0.09473684
  X164attr   X165attr   X169attr   X171attr   X172attr   X173attr   X174attr   X175attr   X179attr
0.05000000 0.09473684 0.09473684 0.09473684 0.05000000 0.09473684 0.05000000 0.05000000 0.05000000
  X181attr   X185attr   X191attr   X192attr   X195attr   X200attr   X201attr   X202attr   X206attr
0.09473684 0.09473684 0.05000000 0.05000000 0.05000000 0.09473684 0.05000000 0.05000000 0.09473684
  X207attr   X210attr   X211attr   X212attr   X213attr   X214attr   X215attr   X216attr   X217attr
0.05000000 0.05000000 0.05000000 0.05000000 0.05000000 0.05000000 0.09473684 0.05000000 0.09473684
  X220attr
0.05000000
> #sum of diagonal values of variance of circle0 dataset
> sum(diag(variance))
[1] 6.715789
> # calculate sum of eigen values of circle0 dataset
> sum(Eigenvalues)
[1] 6.715789
> 
```

Figure 10: Sum of diagonal variance and Eigen values

C. Computation of Eigen Values and Eigen Vectors from Covariance Matrix

After finding covariance matrix, the eigenvalues are computed. The Eigen value and Eigen vectors of all attribute of profile information of circle0 is given in Figure 9.

As can be observed from Figure.9, the Eigenvalues and Eigen vectors are arranged in decreasing order. The first Eigen value is highest Eigen value and last Eigen value is smallest Eigen value of attributes of profile formation. The variance of dataset is equal to total of Eigen value of dataset. The variance and sum of Eigen values is given in Figure.10. Figure.10 shows that total variance of circle0 dataset is equivalent to sum of Eigen values. The attribute names along with their Eigen values and variance is given in Table 4. The attributes are arranged according to decreasing order of Eigen values.

Table 4
Variance of Component in Principle Component

Attr_no	*Attribute/ Dimension*	*Eigen values*	*Variance %*	*Cumulative %*
Attr92	92 languages;id;anonymized feature 92	1.24E+00	3.87931	3.87931
Attr53	53 education;type;anonymized feature 53	8.75E-01	3.87931	7.75862
Attr55	55 education;type;anonymized feature 55	8.20E-01	3.87931	11.63793
Attr77	77 gender;anonymized feature 77	5.38E-01	3.87931	15.51724
Attr78	78 gender;anonymized feature 78	5.29E-01	3.87931	19.39655
Attr65	65 education; year; id;anonymized feature 65	4.97E-01	3.26153	22.65808
Attr50	50 education; school; id;anonymized feature 50	4.54E-01	3.26153	25.91961
Attr54	54 education;type;anonymized feature 54	3.67E-01	2.93887	28.85848
Attr52	52 education; school; id;anonymized feature 52	3.19E-01	1.99843	30.85691
Attr59	59 education;year;id;anonymized feature59	2.33E-01	1.99843	32.85534
Attr63	63 education;year;id;anonymized feature63	2.21E-01	1.99843	34.85377
Attr126	126 locale;anonymized feature 126	1.88E-01	1.99843	36.8522
Attr127	127 locale;anonymized feature 127	1.35E-01	1.99843	38.85063
Attr128	128 location;id;anonymized feature 128	1.22E-01	1.99843	40.84906
Attr138	138 location;id;anonymized feature 137	9.33E-02	1.99843	42.84749
Attr141	141 work;employer;id;anonymized feature 140	5.04E-02	1.99843	44.84592
Attr7	7 birthday;anonymized feature 7	1.68E-02	1.41065	46.25657
Attr14	14 education;concentration;id;anonymized feature 14	1.51E-02	1.41065	47.66722

The attributes ***'92 languages;id;anonymized feature 92','53education;type;anonymized feature 53','55education;type;anonymized feature 55 ', '77 gender; anonymized feature 77', '78gender;anonymized feature 78','65 education; year; id; anonymized feature 65', '50 education; school; id; anonymized feature 50','54education;type;anonymized feature 54','52 education; school; id; anonymized feature 52'***etc have major contribution in making of the circle 0, and the attributes which have zero contribution in making this circle *are* ***'0 birthday; anonymized feature 0', '1 birthday; anonymized feature 1',' 2 birthday; anonymized feature 2',' 48 education; school; id; anonymized feature 48',' 70 education; year; id; anonymized feature 70',' 87 hometown; id; anonymized feature 87',' 86 hometown; id; anonymized feature 86'***, etc.

The study shows that first eighteen attributes explained 48% of total variability of data attributes of profile information. Portion of first four attributes that are Attr92, Attr53, Attr55, Atrr77 are 4%,4%,4%, and 4% as shown in Figure.11.

D. Apply Scree Test on Eigen values of Attributes of Profile Information

The scree test plots the eigenvalues or variances with respect to their component, the large eigenvalues/variances and small eigenvalues/variances display "break" between components. The attributes which appear before the "break" are supposed to be meaningful and those attributes which appear after the break are supposed to be unimportant and are not retained. If the scree test shows number of large breaks, in that case attributes appearing before last large break are supposed to be important. The command for plotting scree test in R is given in Figure 12.

```
> screeplot(my.prc, main="Scree Plot", type="line" )
> |
```

Figure 12: Command for Scree Plot in R

Scree plot of principal components with variance in R is shown in Figure.13. Figure represents number of big breaks before eigenvalues/variance begins to level down. The scree test shows that first five components are significant for the formation of circle.

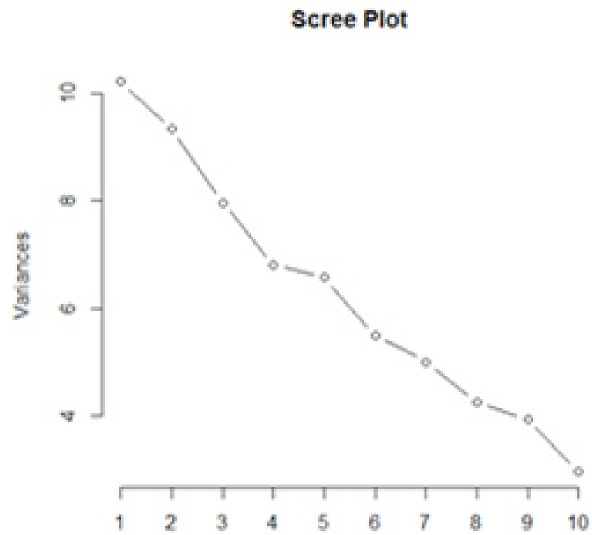


Figure 13: Scree Plot of Components according to variance

E. Computation of Principal Components

The Eigen vectors of highest Eigen values are principal components. The principal components are given in Figure.14.

According to scree test results first five components are supposed to be meaningful. The structure of first five meaningful components is given in Table 5. The Table 5 contains only those variables which contribute maximum in principle components.

Each Meaningful Component is explained as below:

Component 1 is eigenvector of first Eigen value. The main parts of first component are 54attr, 78attr, 160attr, 63attr, 127attr. Thus, this component can provide a great grouping among users *(user71, user215, user163, user81, user245, user54, user297, user193, user97, user132, user163, user259)* from the aspect of **54attr,78attr** that is '**54education; type; anonymized feature 54**' and **'78gender; anonymized feature 78'.**

Component 2 is eigenvector of 2nd Eigen value. The second component are affected by attribute **77atrr, 126attr, 169attr**.The second component provide grouping among users (user61, *user229, user81, user253, user264, user29, user110, user183, user334, user245, user* 259, user222) form aspect of **77attr,126attr** that are **'77gender; anonymized feature 77'** and **'126locale; anonymized feature 126'**.

Component 3 is Eigen vector of third Eigen value. The attr54, attr63, attr127, attr77, attr52, attr92, attr128, attr141, attr53 are main part of component 3. Third component provides grouping among users *(user81, user183, user163)* from aspect of attr52, attr128, that are '**52 education; school; id; anonymized feature 52' and'128location; id; anonymized feature 128'.**

Component 4 is Eigen vector of fourth Eigen value. The **attr63, attr78, attr127, attr160, attr52, attr128, attr55, attr141, attr215** are main part of component four. Fourth component provides grouping among users *(user81, user163, user183, user297, user229, user193, user29, user132, user222)* from aspect of '52 **education; school; id; anonymized feature 52',' 128 location; id; anonymized feature 128'** and **'55education; type; anonymized feature 55'**.

```
> loadings<-Eigenvectors
> loadings
              [,1]         [,2]          [,3]        [,4]         [,5]
 [1,] -0.137858420  0.018017805  1.328275e-01  0.01133074 -0.102759909
 [2,]  0.067473308  0.007427536  9.935037e-02  0.09493996  0.077187836
 [3,]  0.005686553  0.010116842  1.213495e-01  0.06708495  0.071919642
 [4,]  0.044210463  0.043537134  1.758458e-02 -0.01525419  0.007516163
 [5,] -0.041623309 -0.054202606 -6.645646e-02  0.02511333 -0.088724765
 [6,] -0.358768311 -0.014870206 -1.838545e-01  0.07471618  0.009179467
 [7,]  0.027164485  0.014172486  2.993277e-01  0.24557590  0.079304309
 [8,] -0.362297012 -0.098696995  1.128556e-01 -0.02400458  0.202695204
 [9,]  0.196668336 -0.126492408  1.862733e-01  0.04083839  0.215260636
[10,] -0.372260222 -0.034077028  7.154461e-02  0.24685733 -0.027841173
[11,] -0.058157333  0.097471159 -5.945880e-02 -0.05374607  0.186560037
[12,] -0.206039759 -0.065996395 -9.150093e-02  0.11348419 -0.055337302
[13,]  0.016345130 -0.177692477  3.345153e-02 -0.15103152  0.116440143
[14,]  0.118536450  0.053889375  1.748210e-01  0.16401489  0.093552298
[15,] -0.289804041 -0.232010991  1.176036e-01  0.01862422  0.021052514
[16,] -0.058157333  0.097471159 -5.945880e-02 -0.05374607  0.186560037
[17,]  0.011034363  0.041202977 -9.413846e-05 -0.04262604  0.006254821
[18,] -0.195004943  0.331647870  2.530315e-01 -0.17498682  0.059037488
[19,]  0.195004943 -0.331647870 -2.530315e-01  0.17498682 -0.059037488
[20,]  0.027983039 -0.033143910 -8.041743e-03  0.03874938 -0.102208457
[21,] -0.100109412  0.011677881  1.756530e-01  0.05617156 -0.040050328
[22,]  0.016345130 -0.177692477  3.345153e-02 -0.15103152  0.116440143
[23,] -0.218998522 -0.267684251  2.426810e-01 -0.07655207 -0.254164920
[24,] -0.041623309 -0.054202606 -6.645646e-02  0.02511333 -0.088724765
[25,] -0.052947910  0.007857636  3.356170e-02 -0.01013526 -0.033318503
[26,] -0.025278179 -0.231895083 -3.300493e-02 -0.12591819  0.027715378
[27,] -0.057901147 -0.009838995 -3.528627e-02 -0.09166588 -0.157812159
[28,]  0.005686553  0.010116842  1.213495e-01  0.06708495  0.071919642
[29,] -0.052947910  0.007857636  3.356170e-02 -0.01013526 -0.033318503
[30,] -0.116260272  0.132364489 -1.245146e-01 -0.13748984  0.080225738
[31,]  0.116260272 -0.132364489  1.245146e-01  0.13748984 -0.080225738
[32,]  0.027164485  0.014172486  2.993277e-01  0.24557590  0.079304309
[33,]  0.072193501  0.010393225  9.542837e-03  0.02349519 -0.094692294
```

Figure 14: Principal Components

Component 5 is Eigen vector of fifth Eigen value. The **attr165, attr53, attr54, attr169** are main part of component five. Fifth component provides grouping among users *(user71, user215, user297, user81, user163, user245)* from aspect of **54 education; type; anonymized feature 54**' and '**165 work; end date; anonymized feature 162**'.

Table 5
Structure of First Five Components for Circle0 Dataset

Attributes	*PC1*	*PC2*	*PC3*	*PC4*	*PC5*
Attr 54	0.1966683	–0.126492	1.8627e-01	0.04083839	0.21526
Attr 63	0.1185364	0.0538893	1.7482e-01	0.16401489	0.09355
Attr 78	0.1950049	–0.331647	–2.530e-01	0.17498682	–0.05904
Attr 127	0.1162602	–0.132364	1.2451e-01	0.13748984	–0.08023
Attr 160	0.1304261	–0.002453	1.3887e-02	0.14003915	0.019385
Attr 77	–0.195005	0.3316478	2.5303e-01	–0.1749868	0.059038
Attr 126	–0.116260	0.1323644	–1.245e-01	–0.1374898	0.080226

Attributes	*PC1*	*PC2*	*PC3*	*PC4*	*PC5*
Attr 169	–0.105319	0.1012914	8.2632e-02	0.01256075	0.179829
Attr 52	0.0271644	0.0141724	2.9932e-01	0.24557590	0.079305
Attr 92	–0.218998	–0.267684	2.4268e-01	–0.076552	–0.254165
Attr 128	0.0271644	0.0141724	2.9932e-01	0.24557590	0.079305
Attr 55	–0.372260	–0.034077	7.1544e-02	0.24685733	–0.02784
Attr 141	0.0494609	–0.029088	1.6993e-01	0.21724033	–0.09483
Attr 215	–0.087586	–0.119795	–1.707e-01	0.21454747	0.00569
Attr 53	–0.362297	–0.098697	1.1285e-01	–0.024004	0.20269
Attr 165	–0.029618	–0.243285	–7.079e-02	0.03840	0.21086

Table 6
Representation of major attributes of components with their users,'1' represents user is in group and '0' represents particular user is not in group

Users	*54attr*	*78attr*	*77attr*	*126attr*	*52attr*	*128attr*	*55attr*	*165attr*
71user	1	1	0	0	0	0	0	1
215user	1	1	0	0	0	0	0	0
54 user	0	1	0	0	0	0	0	0
61 user	0	0	1	0	0	0	0	0
297user	0	1	0	0	0	0	1	1
229user	0	0	1	0	0	0	1	0
81 user	1	0	1	0	1	1	1	0
253user	0	0	1	0	0	0	0	0
193user	0	1	0	0	0	0	1	0
97 user	0	1	0	0	0	0	0	0
264user	0	0	1	0	0	0	0	0
29 user	0	0	1	1	0	0	1	0
132user	0	1	0	0	0	0	1	0
110user	0	0	1	0	0	0	0	0
163user	1	1	0	0	1	1	0	0
259user	0	1	0	1	0	0	0	0
183user	0	0	1	0	1	1	1	0
334user	0	0	1	0	0	0	1	0
245user	1	0	1	0	0	0	0	0
222user	0	0	1	1	0	0	1	0

Before applying PCA all 224 attributes are to be considered and which attribute is meaningful in forming the circle is difficult to judge. But after the implementation of PCA only few attributes can be considered in forming the circles.

At the starting of implementation, we have 224 attributes. After the implementation of PCA on data, the numbers of attributes are reduced to only seven attributes. These eight attributes are '**54education; type; anonymized feature 54**','**78gender; anonymized feature 78', '77 gender; anonymized feature 77**', **'126locale; anonymized feature 126', '52 education; school; id; anonymized feature 52,' '128location; id; anonymized feature 128', '55 education; type; anonymized feature 55**' and **'165work; end_date; anonymized feature 162**'. These attributes are efficient in making a circle.

Firstly, PC1 and PC2 are plotted in Figure.15, PC1 is *x*-axis and PC2 is *y*-axis. Plotting shows that maximum numbers of users are plotted in y-axis direction. Secondly, PC2 and PC1 are plotted in Figure.16, PC2 is *x*-axis and PC1 is *y*-axis. Plotting shows that minimum numbers of users are in y-axis direction. The conclusion is that PC2 is axis in which majority of users vary as compared to PC1.

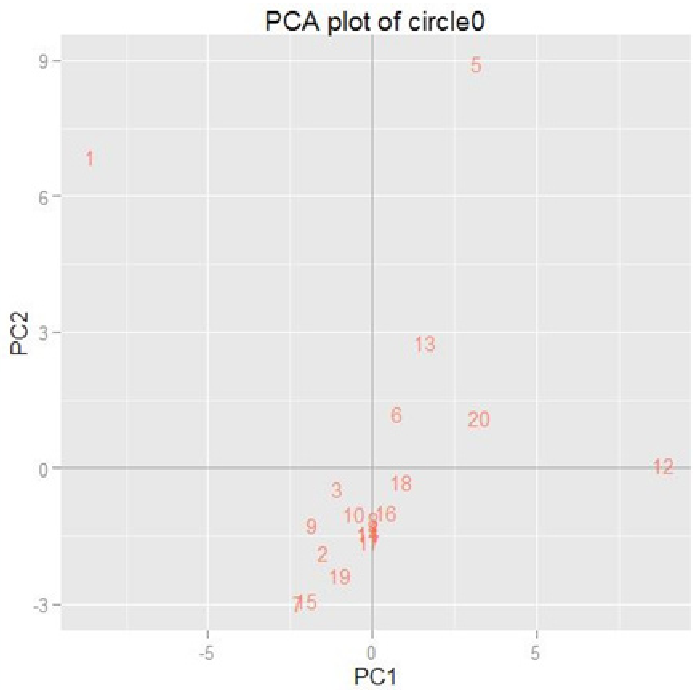


Figure 15: PCA plot of circle with PC1 as *x*-axis

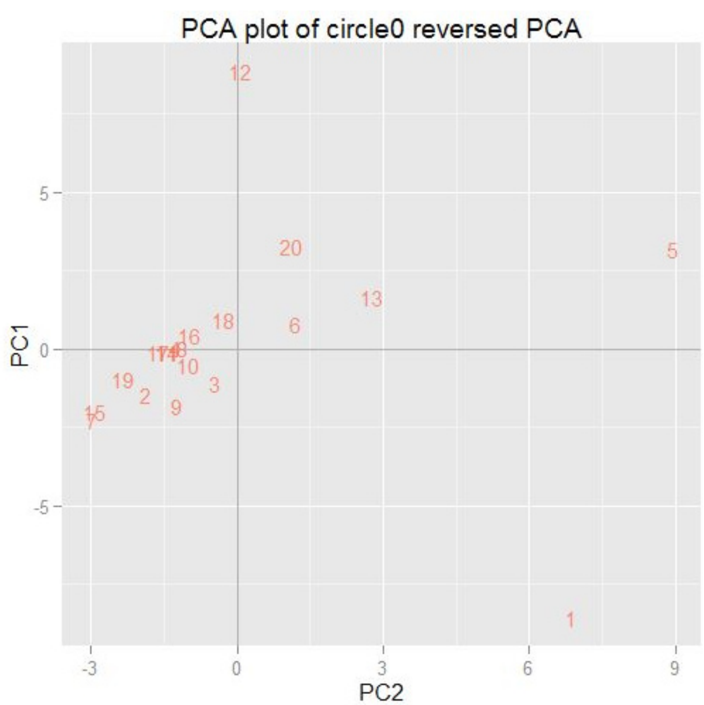


Figure 16: PCA plot of circle with PC2 as *x*-axis

5. RESULT INTERPRETATION

The PCA after implementation gives eight output attributes which are effective in structuring circle0 of social media dataset. The PC2 is more effective as compared to PC1 is observed from Figure 15 and Figure 16. The PCA on attributes of profile information can be observed in Figure 17.

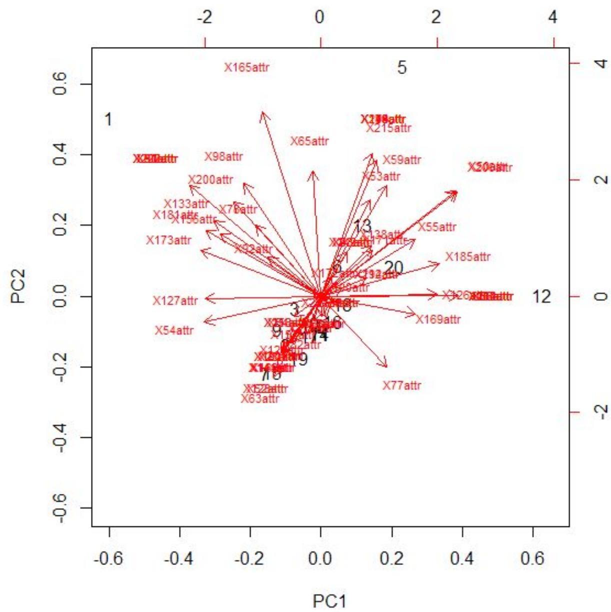


Figure 17: PCA Interpretation

From Figure 17, it is observed that fourteen users are between attr77 and attr54 which shows highly correlations among users. One user is between attr54 and atrr165, two users are between attr55 and attr126 and three users are attr165 and attr55. The maximum users are between attr77 and attr54 so **attr77 '77 gender; anonymized feature 77'** and **attr54 '54education; type; anonymized feature 54'** are effective in structuring circle0 in social media data.

6. CONCLUSION

Large data is produced by social networking sites and as new users create and access their accounts daily. The social networking sites analyze interest of online user and show advertisements on user account according to their interest.

The social networking site analyzes all attributes of users to find similarity and interest. PCA can reduced job of working on all attributes and provides only important attribute which are effective. The number of dimensions can be reduced in order to make effective circles. PCA can reduce effort and cost by reducing unimportant variables.

Future research will focus on analyzing PCA reduced attributes of circles with clustering techniques to understand the effects of PCA in reducing the effort of clustering algorithms.

REFERENCES

[1] L. Wang, J. Zhan, C. Luo, Y. Zhu, Q. Yang, Y. He, W. Gao, Z. Jia, Y. Shi, S. Zhang, C. Zheng, G. Lu, K. Zhan, X. Li, and B. Qiu, "BigDataBench: A big data benchmark suite from internet services," *Proc. - Int. Symp. High-Performance Comput. Archit.*, pp. 488–499, 2014.

[2] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data Mining with Big Data," *Knowl. Data Eng. IEEE Trans.*, vol. 26, pp. 97–107, 2014.

[3] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," *Imc*, 2007.

[4] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen, "The mobile data challenge: Big data for mobile computing research," *Proc. Work. Nokia Mob. Data Challenge, Conjunction with 10th Int. Conf. Pervasive Comput.*, pp. 1–8, 2012.

[5] R. Zafarani, M. A. Abbasi, and H. Liu, "Social Media Mining An Introduction," *Cambridge Univ. Press*, p. 382, 2014.

[6] D. J. Hughes, M. Rowe, M. Batey, and A. Lee, "A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage," *Comput. Human Behav.*, vol. 28, no. 2, pp. 561–569, 2012.

[7] M. Duggan and J. Brenner, *The demographics of social media users, 2012*, vol. 14. Pew Research Center's Internet & American Life Project Washington, DC, 2013.

[8] K. Lewis, M. Gonzalez, and J. Kaufman, "Social selection and peer influence in an online social network," *Proc. Natl. Acad. Sci.*, vol. 109, no. 1, pp. 68–72, 2012.

[9] X. Yang, Y. Guo, Y. Liu, and H. Steck, "A survey of collaborative filtering based social recommender systems," *Comput. Commun.*, vol. 41, pp. 1–10, 2014.

[10] J. Tang, J. Tang, and H. Liu, "Recommendation in social media," *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '14*, pp. 1977–1977, 2014.

[11] J. McAuley and J. Leskovec, "Learning to Discover Social Circles in Ego Networks," ... *Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[12] F. Adu-oppong, C. K. Gardiner, and P. P. Tsang, "Social Circles : Tackling Privacy in Social Networks," *Symp. Usable Priv. Secur.*, pp. 1–2, 2008.

[13] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137–146.

[14] J. Tang, X. Hu, and H. Liu, *Social recommendation: a review*, vol. 3, no. 4. 2013.

[15] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha, "Like like alike," *Proc. 20th Int. Conf. World wide web - WWW '11*, vol. 58, no. September, p. 537, 2011.

[16] H. Liu and D. Lee, "LIKE and Recommendation in Social Media," pp. 1539–1540, 2015.

[17] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media," *Bus. Horiz.*, vol. 54, no. 3, pp. 241–251, 2011.

[18] R. Bro and A. K. Smilde, "Principal component analysis," *Anal. Methods*, vol. 6, no. 9, p. 2812, 2014.

[19] K. Tran, "Dimensionality Reduction with PCA Introduction Dimensionality Reduction," 2011.

[20] J. McAuley and J. Leskovec, "Discovering social circles in ego networks," *ACM Trans. Knowl. Discov. from Data*, vol. 8, no. 1, p. 4, 2014.

[21] S. Shafi and H. Singh, "Circle / Lists In Social Media Using Principal Component," vol. 10, no. 9, pp. 23137–23151, 2015.

[22] N. Mishra, R. Schreiber, I. Stanton, and R. E. Tarjan, "Clustering Social Networks," *Lect. Notes Comput. Sci.*, vol. 4863, pp. 1–12, 2007.

[23] J. Fox and R. Andersen, "Using the R Statistical Computing Environment to Teach Social Statistics Courses," *Unpubl. Manuscr.*, no. January, pp. 1–35, 2005.

[24] J. Leskovec and A. Krevl, "{SNAP Datasets}: {Stanford} Large Network Dataset Collection." Jun-2014.