Marketing Analytics
Professor Doctor Swagato Chatterjee
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur
Social Network Analysis and Excel Dashboards (Contd.)
Lecture 63

Hello everybody. Welcome to Marketing Analytics course. This is Doctor Swagato Chatterjee from VGSOM IIT Kharagpur and we are in week 12. We are discussing Social Network Analysis. In the last two video we have discussed about what social network analysis is and where it can be applied and in this particular session I will solve one case study to see that how this particular social network analysis can be applied in the real life situation.

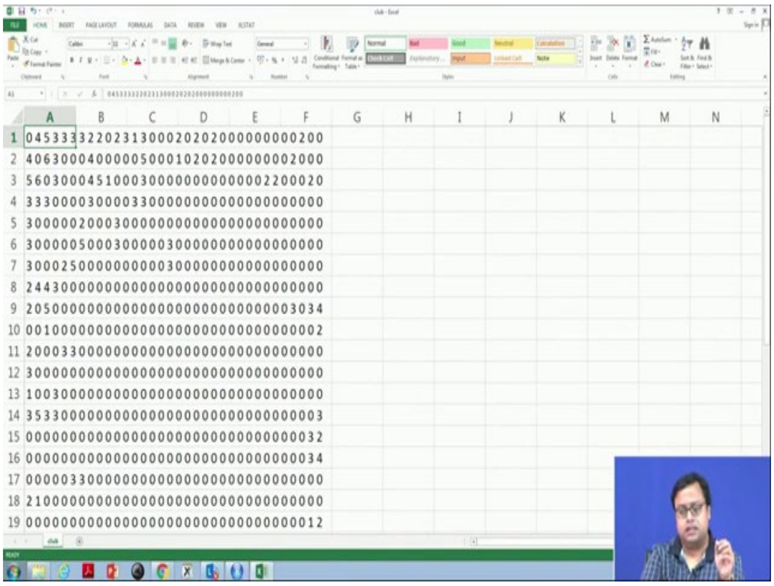
Now this particular case study was about a club. A club is where there are lots of people who actually take subscription and they use various facilities, they play games so this basically entertainment is something that is major focus. Now often times entertainment does not happen alone. It happens in a group.

So it happens in a group means that if a set of people you form, when you subscribe for a club and start going to that club, after some time you become a part of that particular social group. And whatever decisions you take, whether you arrange a party or whether you will arrange a competition or whether you will do this and that you, or you will participate in some of the club events, are generally a collective decision.

So if that is a collective decision then sometimes which club will you choose? Or whether you will be switching from this club to some other club is also sometimes is a collective decision. So this particular club all of a sudden, some club all of a sudden found that their organizations members are actually leaving the club in huge chunk. So it is a churn but unlike the churn in telecom or unlike the churn in let us say bank and etc where churning is more of a personal decision they know that this is a collective decision.

So they have to find out that who are the influencing persons in this social network who are driving this particular churn whom I can actually connect and then I can incentivize them or I can create kind of service recovery measurements or anti-churn measurement or switching barrier measures so that I can stop them from churning. So that was the major goal. So they collected the data and the data looks like this.

(Refer Slide Time: 02:50)



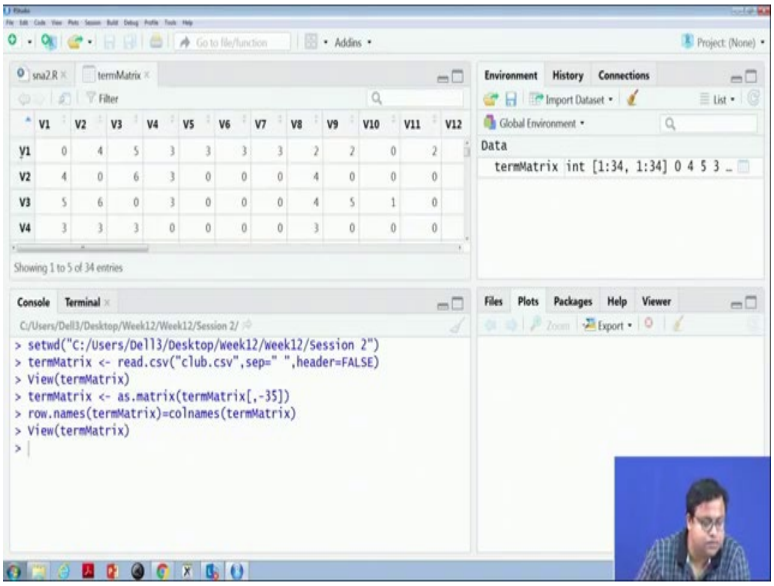
```
1  0453333220231300020202000000000200
2  4063000400000500010202000000002000
3  5603000451000300000000000002200020
4  3330000300003300000000000000000000
5  3000002000300000000000000000000000
6  3000005000300000300000000000000000
7  3000250000000000300000000000000000
8  2440000000000000000000000000000000
9  2050000000000000000000000000003034
10 0010000000000000000000000000000002
11 2000330000000000000000000000000000
12 3000000000000000000000000000000000
13 1003000000000000000000000000000000
14 3533000000000000000000000000000003
15 0000000000000000000000000000000032
16 0000000000000000000000000000000034
17 0000330000000000000000000000000000
18 2100000000000000000000000000000000
19 0000000000000000000000000000000012
```

So this is basically the text data, where there are basically, if I am not wrong 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 16, 17 around 30-40, if I think this is half, so around 30, 34-32 people were there and you see that the diagonal elements are all 0. So 0, 0, 0, 0 all the diagonal elements are 0 and this 4 means high strength, 5 means low strength and this is bi-directional.

So actually not strength, they try to find out how many people has talked with how many people. So he just called the members and asked that in the last 1 month, you have met or talked with which members and how many times? This much they have got the information. So how many times you have talked with which member was the information. And from that information they tried to find out because all of these decisions are taken, in a small time frame.

So historically whom you are more connected with, I do not care. In the recent times how many times you have talked with whom is something that is more important to understand. It can be also understood to find out that let us say if there is mass bunking in your class who has organized that mass bunking you can find out from this kind of information. Who is the person, who should be the class CR? Can be found out form this kind of information.

(Refer Slide Time: 04:27)

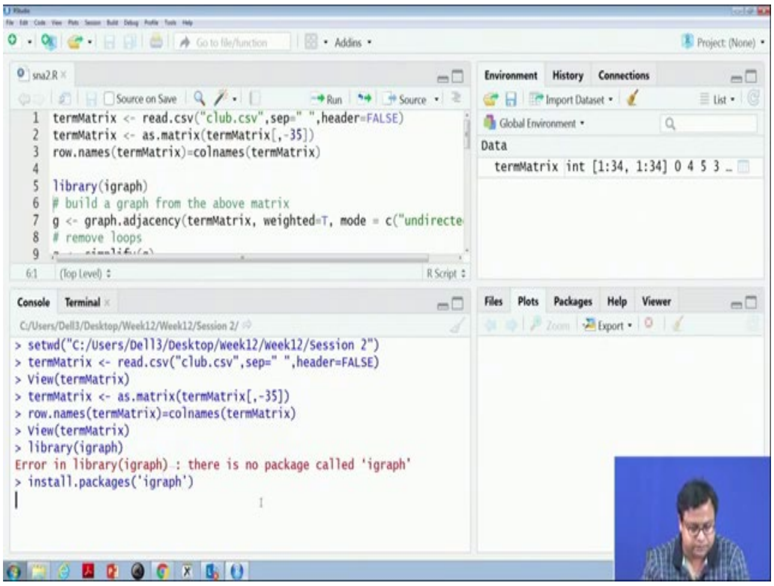


	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
V1	0	4	5	3	3	3	3	2	2	0	2	
V2	4	0	6	3	0	0	0	4	0	0	0	
V3	5	6	0	3	0	0	0	4	5	1	0	
V4	3	3	3	0	0	0	0	3	0	0	0	

Showing 1 to 5 of 34 entries

```
Data
  termMatrix int [1:34, 1:34] 0 4 5 3 …
```

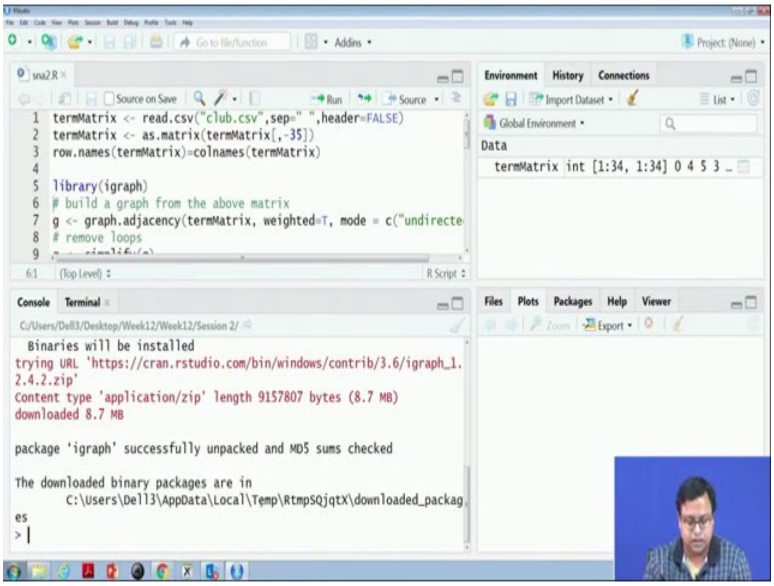
```
C:/Users/Dell3/Desktop/Week12/Week12/Session 2/
> setwd("C:/Users/Dell3/Desktop/Week12/Week12/Session 2")
> termMatrix <- read.csv("club.csv",sep=" ",header=FALSE)
> View(termMatrix)
> termMatrix <- as.matrix(termMatrix[,-35])
> row.names(termMatrix)=colnames(termMatrix)
> View(termMatrix)
> 
```



```
termMatrix <- read.csv("club.csv",sep=" ",header=FALSE)
termMatrix <- as.matrix(termMatrix[,-35])
row.names(termMatrix)=colnames(termMatrix)

library(igraph)
# build a graph from the above matrix
g <- graph.adjacency(termMatrix, weighted=T, mode = c("undirecte
# remove loops
```

```
Data
  termMatrix int [1:34, 1:34] 0 4 5 3 …
```

```
C:/Users/Dell3/Desktop/Week12/Week12/Session 2/
> setwd("C:/Users/Dell3/Desktop/Week12/Week12/Session 2")
> termMatrix <- read.csv("club.csv",sep=" ",header=FALSE)
> View(termMatrix)
> termMatrix <- as.matrix(termMatrix[,-35])
> row.names(termMatrix)=colnames(termMatrix)
> View(termMatrix)
> library(igraph)
Error in library(igraph) : there is no package called 'igraph'
> install.packages('igraph')
```



```r
termMatrix <- read.csv("club.csv",sep=" ",header=FALSE)
termMatrix <- as.matrix(termMatrix[,-35])
row.names(termMatrix)=colnames(termMatrix)

library(igraph)
# build a graph from the above matrix
g <- graph.adjacency(termMatrix, weighted=T, mode = c("undirecte
# remove loops
```

```
Data
  termMatrix int [1:34, 1:34] 0 4 5 3 …
```

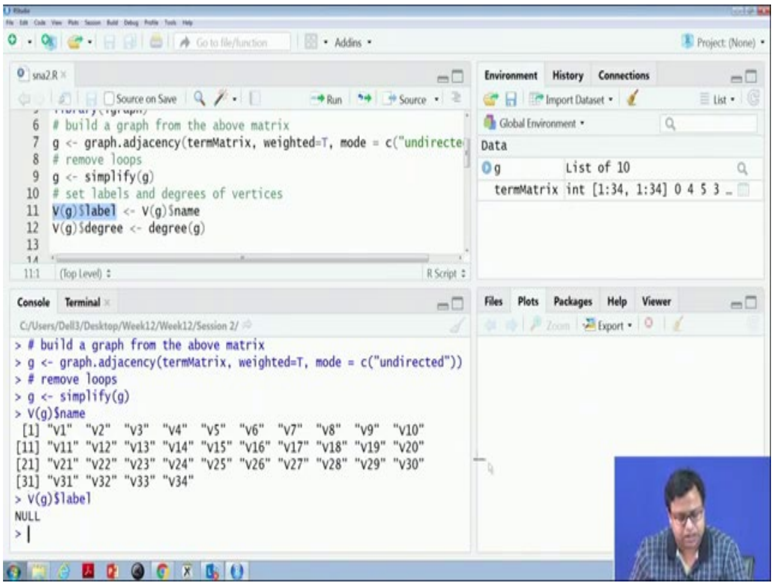
```
  Binaries will be installed
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/igraph_1.
2.4.2.zip'
Content type 'application/zip' length 9157807 bytes (8.7 MB)
downloaded 8.7 MB

package 'igraph' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Dell3\AppData\Local\Temp\RtmpSQjqtX\downloaded_packag
es
> |
```

So the first job is to set working directory to source file location and then I will just read csv, see the separator is space and there was no header. So header is equal to false. So when I do that, that gives me blank file with 34 observations as I told and 35 variables, 34 and 35 so last one is NA actually.

I do not know how that has been read so I will remove that, so 35 observations so term matrix is minus 35 that is why. The last one is, so 34 by 34, and then row names of term matrix and column names of term matrix are same. So if this is V1, V2 to V34, this will also be V1, V2 to V34. So that is something I am ensuring here.

Then I am calling a library called igraph. Now igraph package is not there. So I have to install this package. Now, I can install it here itself. I can say install dot packages and then igraph, igraph, Fair enough, it has been downloaded. I will call this one, so yeah, so igraph is connected now.

(Refer Slide Time: 06:14)



```r
# build a graph from the above matrix
g <- graph.adjacency(termMatrix, weighted=T, mode = c("undirecte
# remove loops
g <- simplify(g)
# set labels and degrees of vertices
V(g)$label <- V(g)$name
V(g)$degree <- degree(g)
```

```
Data
g           List of 10
termMatrix  int [1:34, 1:34] 0 4 5 3 …
```

```
C:/Users/Dell3/Desktop/Week12/Week12/Session 2/
> # build a graph from the above matrix
> g <- graph.adjacency(termMatrix, weighted=T, mode = c("undirected"))
> # remove loops
> g <- simplify(g)
> V(g)$name
 [1] "V1"  "V2"  "V3"  "V4"  "V5"  "V6"  "V7"  "V8"  "V9"  "V10"
[11] "V11" "V12" "V13" "V14" "V15" "V16" "V17" "V18" "V19" "V20"
[21] "V21" "V22" "V23" "V24" "V25" "V26" "V27" "V28" "V29" "V30"
[31] "V31" "V32" "V33" "V34"
> V(g)$label
NULL
> |
```

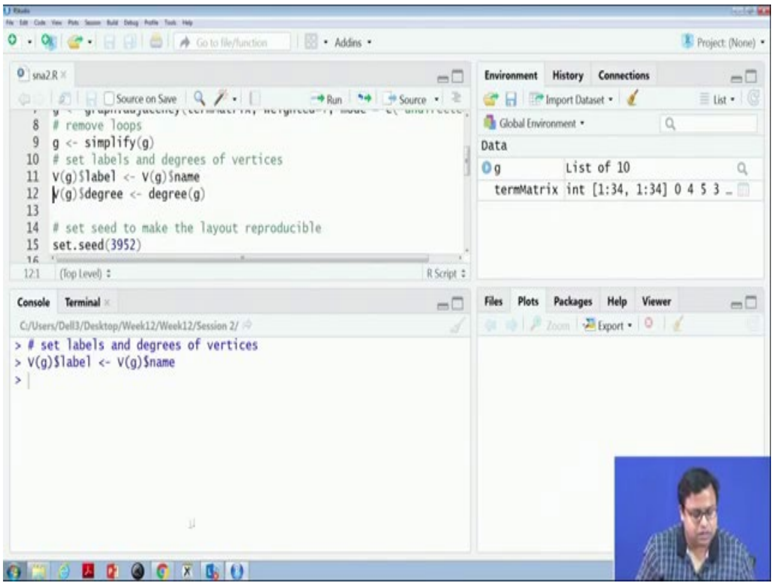
Now I will create the graph. So g is equal to graph adjacent C and the term matrix is the input, weighted is equal to true that means 3 and 5 should be considered as weights and mode is undirected because there was no direction. So I created this and all the properties have been saved in this g.

Now I want to simplify the loops. If a is connected to b and b is connected to a and there is no, means unidirectional, sorry it is bidirectional thing then only the lower triangular matrix is enough because it is bidirectional. My connection with you and your connection with me are same thing here and my connection strength with you and your connection strength with me is also same.

If it is bidirectional and the width is same in both the directions, then only one triangle, the lower triangle or upper triangle is enough. So that is what we are doing here. g is equal to, simplify g means we are reducing the number of nodes in the, number of paths in the, in the graph that we have created. And then I am saying that V(g)_level= V(g)_name

So what is V(g)_name? These are basically the names of the nodes V1, V2 to V34. I was saying that the labels, labels means whatever should be written in the node when I click the pictures, right now there are null, there is no picture so I will say that this is a V1 to V34.

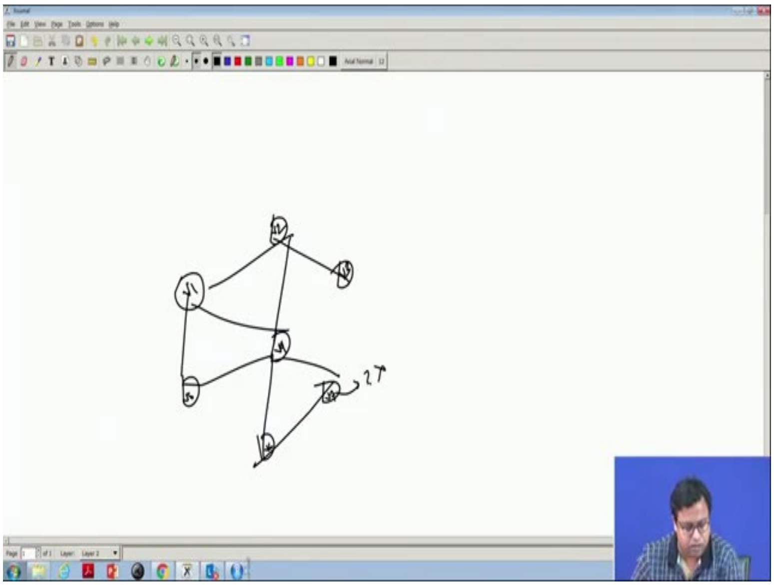
(Refer Slide Time: 08:03)



```
# remove loops
g <- simplify(g)
# set labels and degrees of vertices
V(g)$label <- V(g)$name
V(g)$degree <- degree(g)

# set seed to make the layout reproducible
set.seed(3952)
```

```
g             List of 10
termMatrix    int [1:34, 1:34] 0 4 5 3 …
```

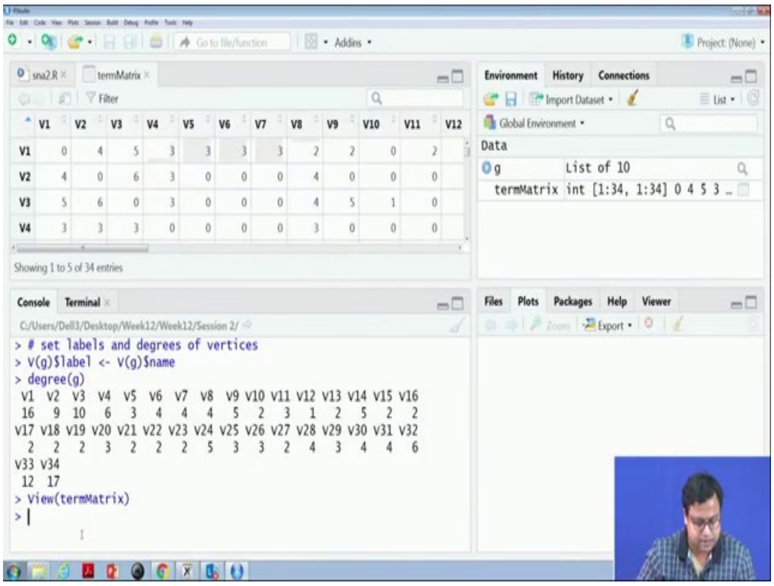
```
C:/Users/Dell3/Desktop/Week12/Week12/Session 2/
> # set labels and degrees of vertices
> V(g)$label <- V(g)$name
> 
```



So once I do that when I draw this particular thing, particular graph it will look like this unless I do that it will not look like this. So let us try to make a case, what I am trying to say that now let us say this is one. If you don't do that it will just create these nodes. It will not create any, any names. It will not give any names here.

Now after running this, this will be V1, this will be V2, this is called labels. These labels will be there. Otherwise these labels will, these values will be not there. So that is what I am doing. I am ensuring it. So, and similarly V of the g degree is degree of g. What is degree of g?

(Refer Slide Time: 08:53)



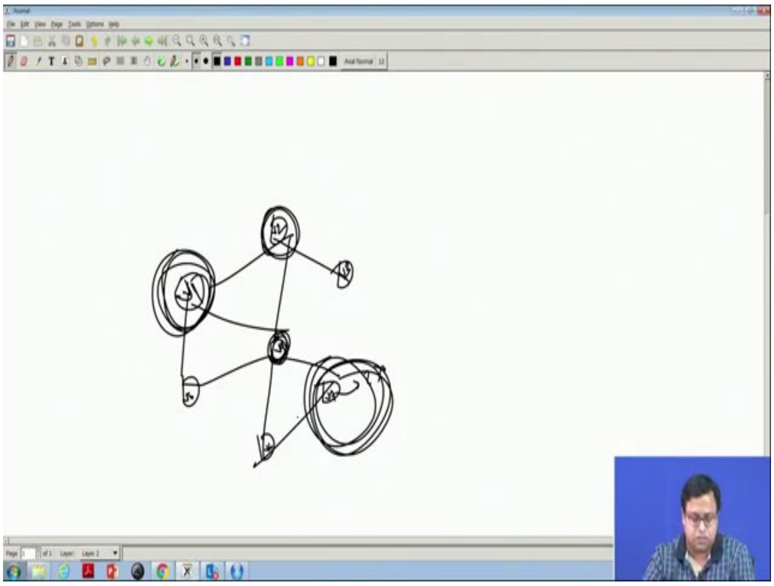
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
V1	0	4	5	3	3	3	3	2	2	0	2	
V2	4	0	6	3	0	0	0	4	0	0	0	
V3	5	6	0	3	0	0	0	4	5	1	0	
V4	3	3	3	0	0	0	0	3	0	0	0	

Showing 1 to 5 of 34 entries

Data
g List of 10
termMatrix int [1:34, 1:34] 0 4 5 3 ...

C:/Users/Dell3/Desktop/Week12/Week12/Session 2/

```
> # set labels and degrees of vertices
> V(g)$label <- V(g)$name
> degree(g)
 V1  V2  V3  V4  V5  V6  V7  V8  V9 V10 V11 V12 V13 V14 V15 V16 
 16   9  10   6   3   4   4   4   5   2   3   1   2   5   2   2 
V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 
  2   2   2   3   2   2   2   5   3   3   2   4   3   4   4   6 
V33 V34 
 12  17 
> View(termMatrix)
> |
```



So each connection they are trying to find out how much is the degree. So this is V1 16, V2 9, V3 10 and etc so V g degree is also something that I am trying to say that how fat or how small this particular nodes will be.

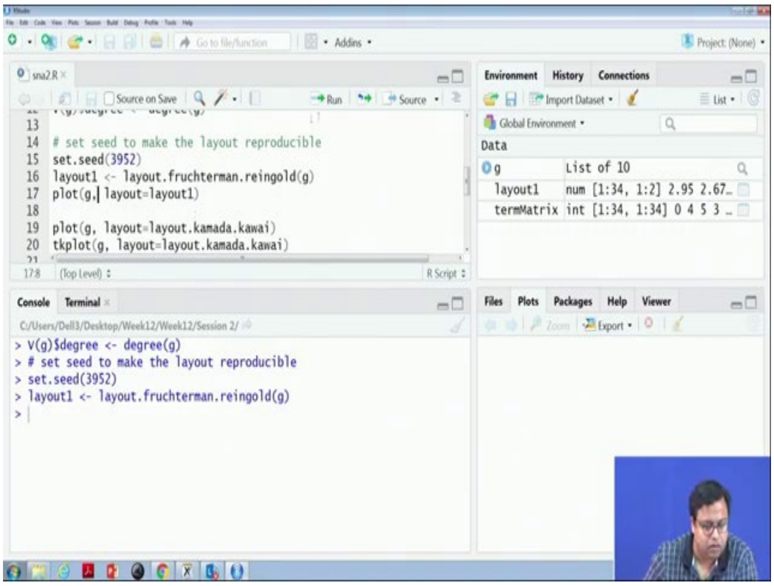
So V1 has 16 connections total. Total connection strength is 16 so that is why it is 16. V2 has, how will I know that? If we just check the term Matrix V1 4 plus 5, 9 plus 3 plus 3 plus 3 okay so this is actually more than 16. Just let me check what this value is.

So term Matrix, sum of term Matrix and then the first row is 42, sum(termMatrix[1,])=42, okay then that is not the value so term Matrix greater than 0 is 16, sum(termMatrix[1,] >0)=16, okay so how many guys he is connected with actually? V1 is connected with, so that value is 16.

So V1 is connected with 16 guys out of these 34 guys. V2 is connected with 9 guys. V3 is connected with 10 guys. So the more the connection is the fatter will this look like. So this will be bigger if V1 is 16. This will be smaller; this will be of this size. This will be of this size if their connection is much higher.

So depending on the number of connections they have, the nodes will be bigger or smaller. So that is something that we are ensuring here. So by saying, by running this line I am ensuring that. Then I will plot this.

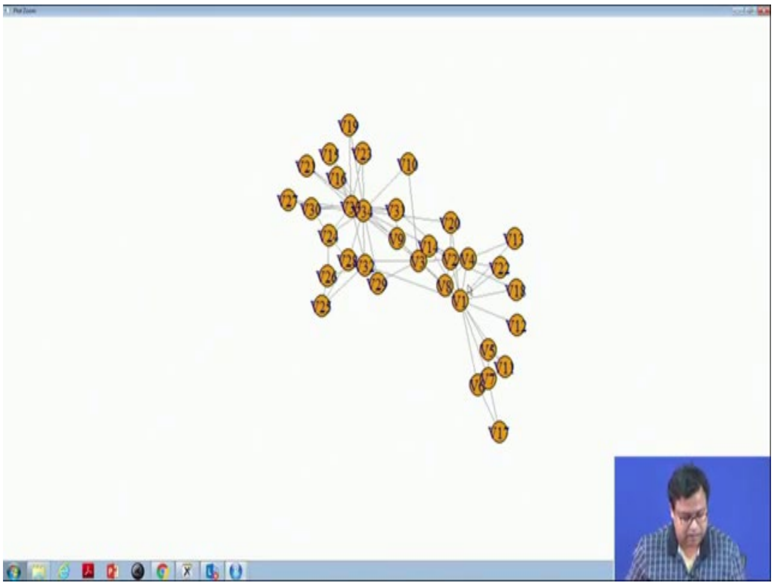
(Refer Slide Time: 10:39)



```
# set seed to make the layout reproducible
set.seed(3952)
layout1 <- layout.fruchterman.reingold(g)
plot(g, layout=layout1)

plot(g, layout=layout.kamada.kawai)
tkplot(g, layout=layout.kamada.kawai)
```

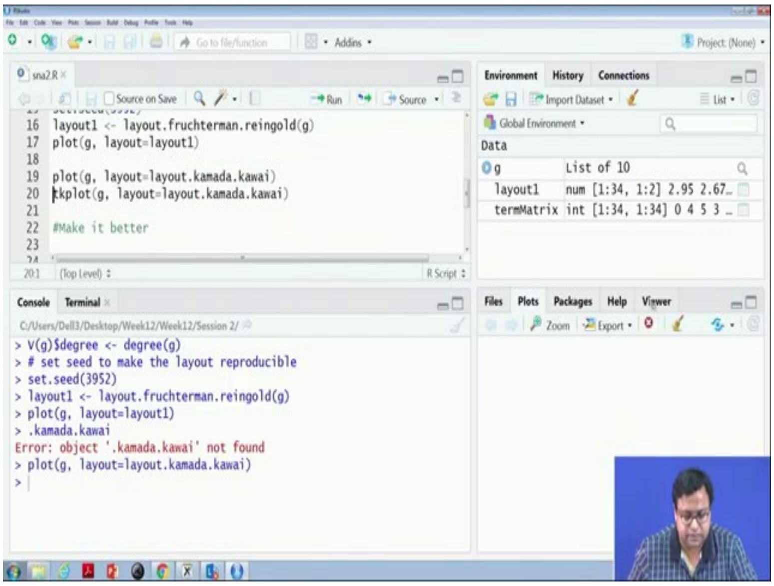
```
Data
g           List of 10
layout1     num [1:34, 1:2] 2.95 2.67…
termMatrix  int [1:34, 1:34] 0 4 5 3 …
```

```
C:/Users/Dell3/Desktop/Week12/Week12/Session 2/
> V(g)$degree <- degree(g)
> # set seed to make the layout reproducible
> set.seed(3952)
> layout1 <- layout.fruchterman.reingold(g)
> 
```



So some spreadsheet and then layout 1 is, there are different methods of laying out. So I am using the fruchterman reingold layouting so if I run this and then plot the layout I get this kind of a view. So it is saying clearly that if you check carefully that these two guys are more central, V3 and V4 and then comes probably, even I do not know this V14, this person is also V2, V14 these people are also connected but not, I cannot say that who is the major person till now. I cannot say that in a better way. So I will try to get another plot.

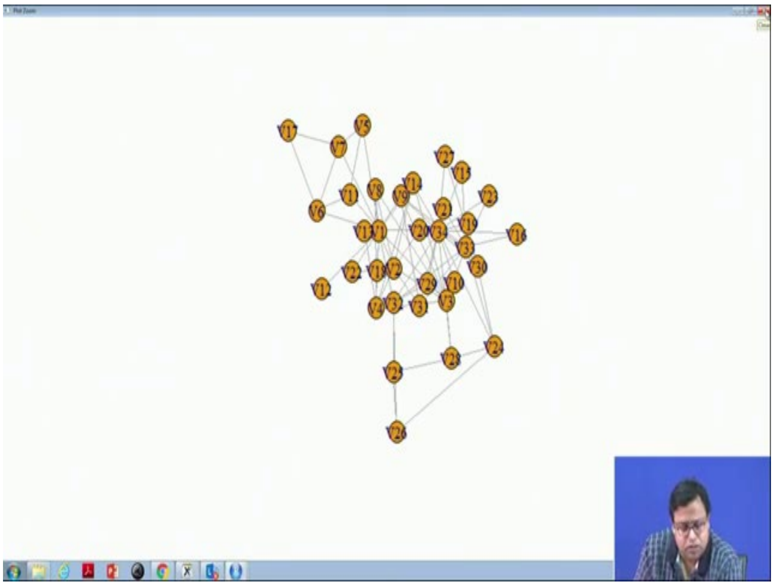
(Refer Slide Time: 11:31)



```
layout1 <- layout.fruchterman.reingold(g)
plot(g, layout=layout1)

plot(g, layout=layout.kamada.kawai)
tkplot(g, layout=layout.kamada.kawai)

#Make it better

```

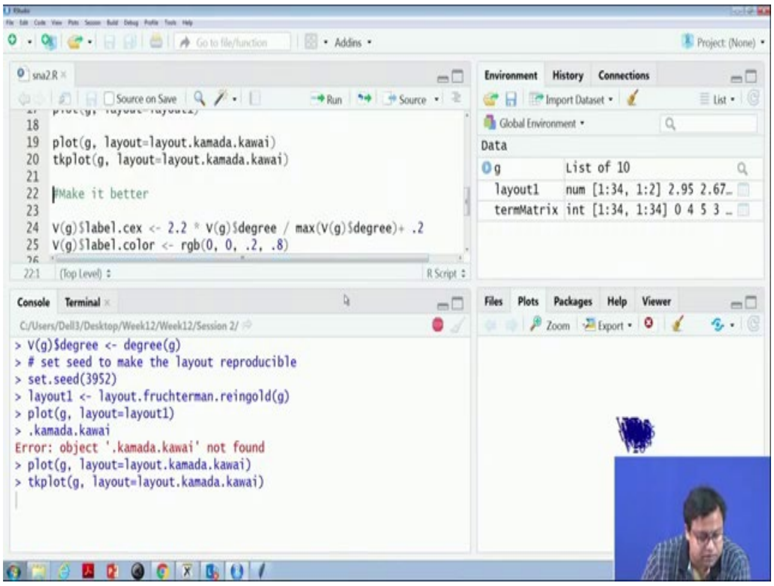
Data
g List of 10
layout1 num [1:34, 1:2] 2.95 2.67…
termMatrix int [1:34, 1:34] 0 4 5 3 …

```
C:/Users/Dell3/Desktop/Week12/Week12/Session 2/
> V(g)$degree <- degree(g)
> # set seed to make the layout reproducible
> set.seed(3952)
> layout1 <- layout.fruchterman.reingold(g)
> plot(g, layout=layout1)
> .kamada.kawai
Error: object '.kamada.kawai' not found
> plot(g, layout=layout.kamada.kawai)
> 
```



So let us say if I remove this and I plot kamada kawai plotting, kamada kawai plotting looks like this. The same thing they have plotted with a different tension and different visuals. This is further not better. Okay I cannot say who is the person who is responsible.

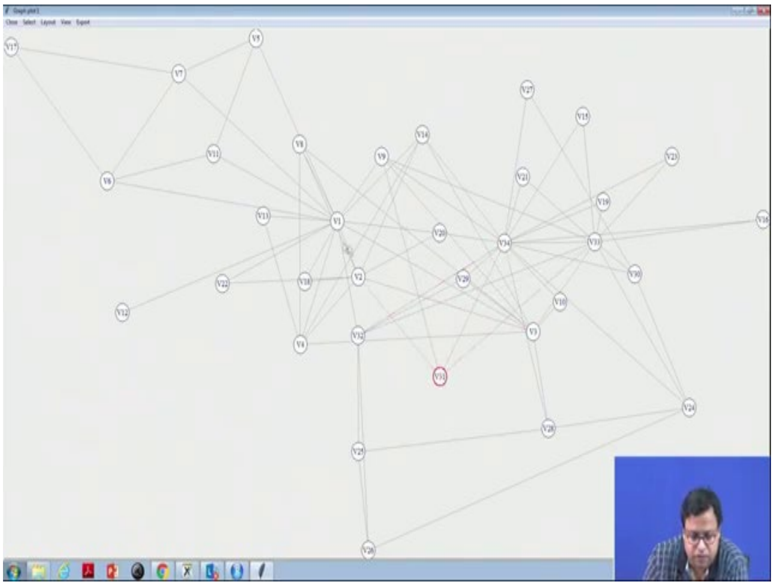
(Refer Slide Time: 11:56)

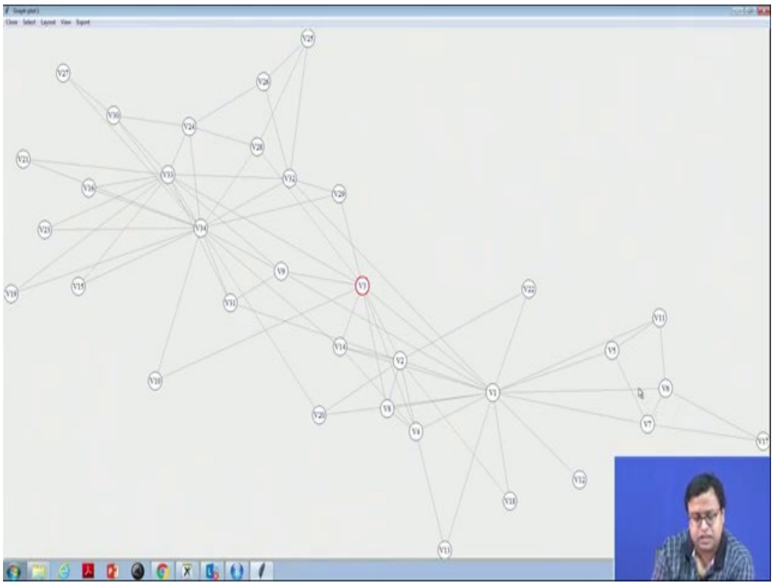


```
plot(g, layout=layout.kamada.kawai)
tkplot(g, layout=layout.kamada.kawai)

#Make it better

V(g)$label.cex <- 2.2 * V(g)$degree / max(V(g)$degree)+ .2
V(g)$label.color <- rgb(0, 0, .2, .8)
```

```
> V(g)$degree <- degree(g)
> # set seed to make the layout reproducible
> set.seed(3952)
> layout1 <- layout.fruchterman.reingold(g)
> plot(g, layout=layout1)
> .kamada.kawai
Error: object '.kamada.kawai' not found
> plot(g, layout=layout.kamada.kawai)
> tkplot(g, layout=layout.kamada.kawai)
```

Data		
g	List of 10	
layout1	num [1:34, 1:2] 2.95 2.67…	
termMatrix	int [1:34, 1:34] 0 4 5 3 …	





So I can do a tkplot with kamada kawai which is a interactive plot which looks like this, and here what I can do is I can select all the vertices and right click on any one of the vertices and vertex color I can change it to white. Now everybody is same, and then I can select all the vertices and vertex size is, let us say I make it a little bit bigger. Okay so it becomes smaller actually. So select all the vertices and then vertex size I will make it a little bit bigger.

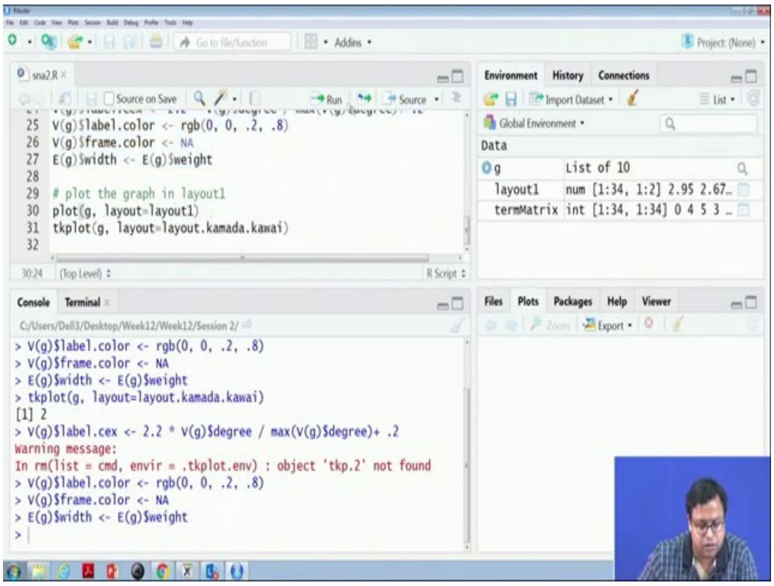
So now I can see this, fair enough? Now what will I do? I will fit view, fit the screen. Now if you can see carefully that see, this person V13 is, V34 is most connected and V1 is also connected, and V20 is a person who is in-between, not so much, V29 is also not so much, V31 okay not so much, so V34 and V1 and this is the two persons who creates most amount of connections. So probably these are the two persons who have made this kind of a problem.

Now if I just say that all edges and right click on the edge and edge width will be, okay edge width are same here, so why do not I try out another one, like let us say I try out this, this one which is fruchterman reingold as I was doing.

Here also I can say that 34 is most connected, fair enough, 33 is less connected than 34, and here I can, so I can specifically see there are two groups of people. One group of people are this, who are more connected with 33 and 34, another group of people is this which is more connected with V1 and in between probably there is V3, not V3, V3 and V9 are the person who are lying in between these two groups.

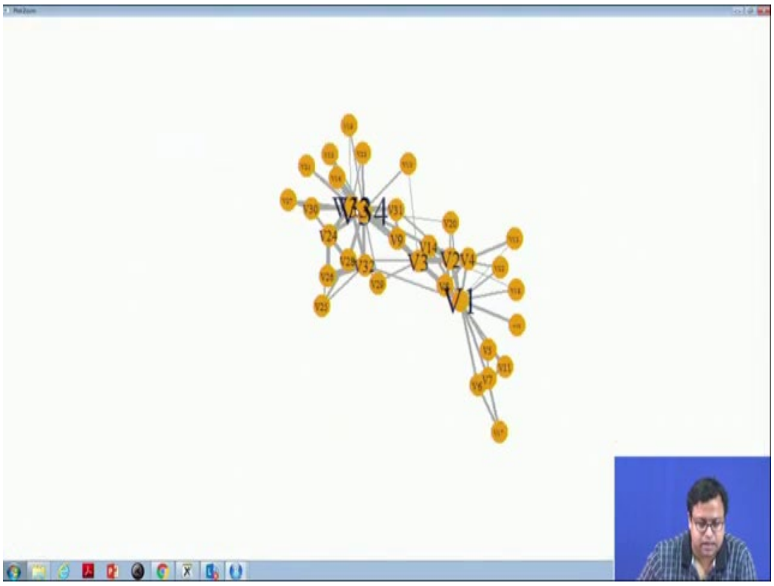
So if I by chance cut V3 or V9, this particular place I can create a cut. This guy if I vanish then there are lots of less connections will be there from this group and that group. So this is something which is an interesting insight for these marketing managers.

(Refer Slide Time: 14:27)



```
V(g)$label.color <- rgb(0, 0, .2, .8)
V(g)$frame.color <- NA
E(g)$width <- E(g)$weight

# plot the graph in layout1
plot(g, layout=layout1)
tkplot(g, layout=layout.kamada.kawai)
```

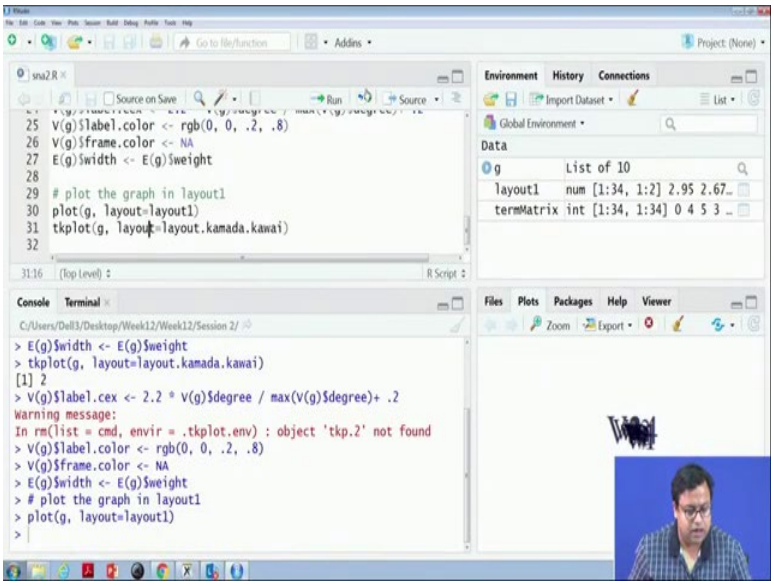
```
> V(g)$label.color <- rgb(0, 0, .2, .8)
> V(g)$frame.color <- NA
> E(g)$width <- E(g)$weight
> tkplot(g, layout=layout.kamada.kawai)
[1] 2
> V(g)$label.cex <- 2.2 * V(g)$degree / max(V(g)$degree)+ .2
Warning message:
In rm(list = cmd, envir = .tkplot.env) : object 'tkp.2' not found
> V(g)$label.color <- rgb(0, 0, .2, .8)
> V(g)$frame.color <- NA
> E(g)$width <- E(g)$weight
>
```



Now what they later do? I want to make it a little bit better looking. So to make it a little bit better looking I can say that the label will be based on the degrees, the label color can be changed and the width can also be changed. So this is something that I am putting up. Just 1 minute.

Okay so this is something that I have to run and then I can run this, fair enough and then if I now try to plot you will see that not only I am getting the V34, and 33 and V1, the size are bigger because as I was telling that I can specifically see that which one has higher things and which one is lower.

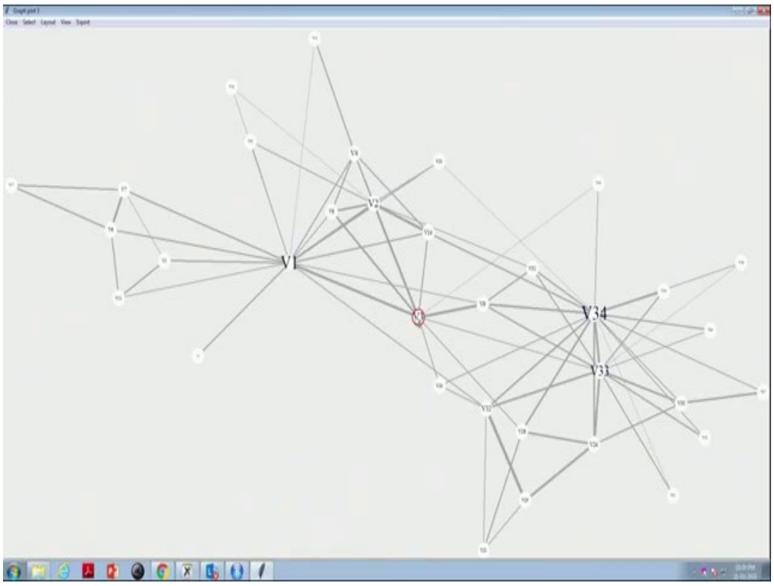
(Refer Slide Time: 15:21)



```
V(g)$label.color <- rgb(0, 0, .2, .8)
V(g)$frame.color <- NA
E(g)$width <- E(g)$weight

# plot the graph in layout1
plot(g, layout=layout1)
tkplot(g, layout=layout.kamada.kawai)
```

```
Data
g            List of 10
layout1      num [1:34, 1:2] 2.95 2.67…
termMatrix   int [1:34, 1:34] 0 4 5 3 …
```

```
C:/Users/Dell3/Desktop/Week12/Week12/Session 2/
> E(g)$width <- E(g)$weight
> tkplot(g, layout=layout.kamada.kawai)
[1] 2
> V(g)$label.cex <- 2.2 * V(g)$degree / max(V(g)$degree)+ .2
Warning message:
In rm(list = cmd, envir = .tkplot.env) : object 'tkp.2' not found
> V(g)$label.color <- rgb(0, 0, .2, .8)
> V(g)$frame.color <- NA
> E(g)$width <- E(g)$weight
> # plot the graph in layout1
> plot(g, layout=layout1)
> 
```



I can also plot in this interactive plot. The interactive plot will also try to, so if I just select all vertices and make them, make the color white and then if I just layout with this one, I can still say that see, 33, 34 is coming very big and V1 is coming so I think these two guys are most important guys. 33, 34 and V1 and V3 is a person who is in between, who is also trying to connect between these two people and I have to find out that which group is leaving.

If this group is leaving then 33 and 34 are more important, if this group is leaving then V1 is more important. If these guys are quarreling then V3 is the most important. So this is the social structure of the club participants and if there is lots of churn happening we as marketing manager of this particular club has to identify whom I can incentivize to reduces the chances of churn.

In this particular condition it is V33 with whom whatever be this person, V33, V34 and V1 these are the persons who can create a difference in this churn management kind of a scenario. So that is how I can say that we can apply social network analysis in a situation.

This is more of a exploratory analysis technique. This is not majorly, I would say very hardcore, machine learning kind of technique but from here I can find out the centrality, I would say cohesiveness and various other measures that are there about structure, social structure and from there we can find out that how that impacts the centrality or social structure or cohesiveness impacts the chances of reneging or staying in the club or not.

So that is all for this particular case study. We will be doing another case study in the next video and I will meet you in the next video. Thank you very much for being with me. See you in the next video.