Statistical Computing with R: Masters in Data Science 503 (S15) Fourth Batch, SMS, TU, 2025

Shital Bhandary
Associate Professor
Statistics/Bio-statistics, Demography and Public Health Informatics
Patan Academy of Health Sciences, Lalitpur, Nepal
Faculty, Masters in Medical Research, NHRC/Kathmandu University
Faculty, FAIMER Fellowship in Health Professions Education, India/USA

Review Preview

- Social Networks:
  - Nodes/Vertices
  - Edges/Connection
  - Degree
  - Edge density
  - Closeness (centrality)
  - Betweenness (centrality)
  - Edge_betweenness etc.
- Social Network Analysis:
  - Hubs
  - Authorities
  - Community detection

Social Networks:

https://study.com/academy/lesson/what-are-social-networks-types-examples-quiz.html

- **Social networks** are simply networks of social interactions and personal relationships. Think about your group of friends and how you got to know them.
- Maybe you met them in elementary school, or maybe you met them through a hobby or through your community.
- Either way, you were exposed to social networks: **meeting other individuals in a social situation, while developing strong personal bonds over time**.

- If you're on Facebook, keep in mind that so are 1.15 billion? other people throughout the world.
- In fact, 72% of all Internet users are active on social media today, indulging in social interactions and developing personal relationships.
- **But you don't always have to go online to be exposed to social networks, as they come in a multitude of formats.**

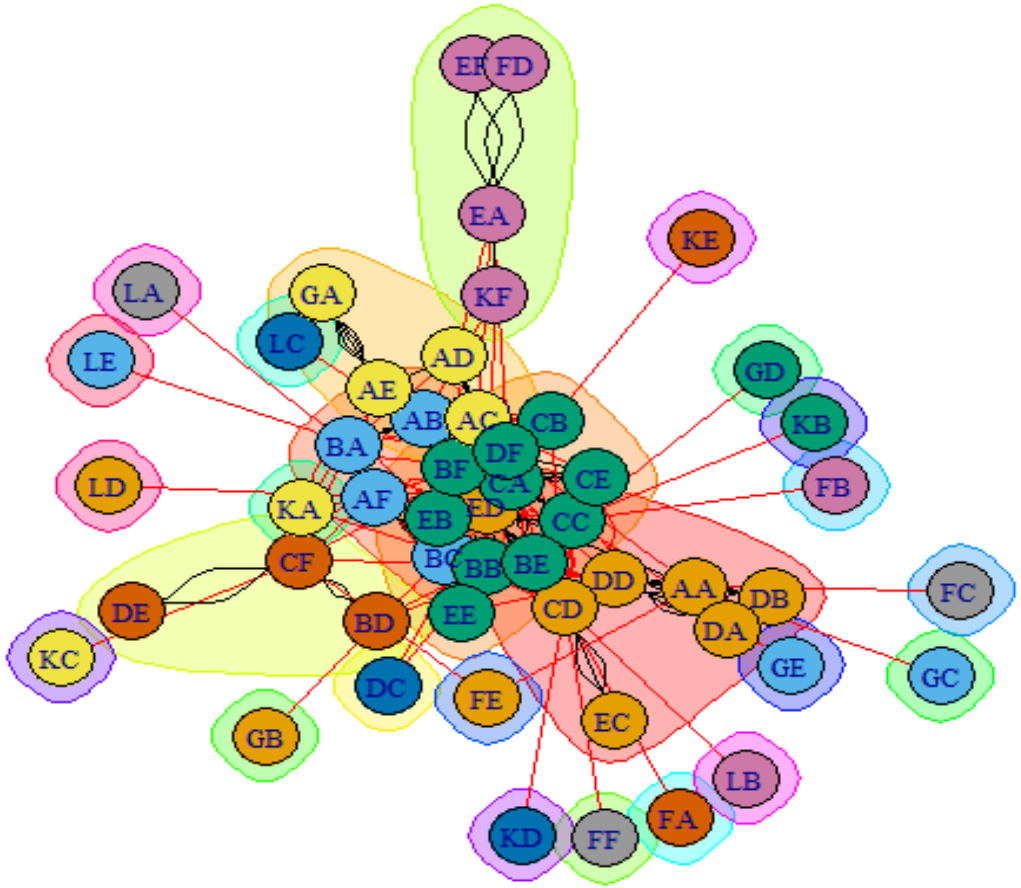
Why Should I Care About Social Network Analysis?

https://towardsdatascience.com/how-to-get-started-with-social-network-analysis-6d527685d374

- Social network analysis (SNA), also known as network science, is a field of data analytics that uses networks and graph theory to understand social structures.
- SNA techniques can also be applied to networks outside of the societal realm.
- Networks are all around us — such as road networks, internet networks, and online social networks like Facebook, Twitter …
- Learning SNA and its techniques will give you valuable tools to provide insight on a variety of data sources.
- In order to build SNA graphs, two key components are required: **actors and relationships**.

SNA graph:

- A social network graph contains both points and lines connecting those dots — similar to a connect-the-dot puzzle.
- The points represent the actors and the **lines** represent the **relationships**.
- The shaded area is “community”



SNA: Networks and Graph theory

https://en.wikipedia.org/wiki/Social_network_analysis

- **Social network analysis** (**SNA**) is the process of investigating social structures through the use of networks and graph theory.
- It characterizes networked structures in terms of *nodes* (individual actors, people, or things within the network) and the *ties*, *edges*, or *links* (relationships or interactions) that connect them.
- The advantages of SNA are twofold. Firstly, it can process a large amount of relational data and describe the overall relational network structure.
- It can also select term and parameter to confirm the influential nodes in the network, such as in-degree and out-degree centrality.
- Through analyzing nodes, clusters and relations, the communication structure and position of individuals can be clearly described

Discussion on “How to do SNA Guide?”

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491572/socnet_howto.pdf

- The aim of social network analysis is to understand a community by mapping the relationships that connect them as a network, **and then trying to draw out** key individuals, groups within the network (‘components’), and/or associations between the individuals.

- A network is simply a number of points (or ‘nodes’) that are connected by links.
- Generally in social network analysis, the nodes are people and the links are any social connection between them – for example, friendship, marital/family ties, or financial ties.
- **SNA for detecting network of gangs (of criminals)**

How SNA is used to analyze “gang” network?

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491572/socnet_howto.pdf

- Social network analysis can provide information about the **reach of gangs, the impact of gangs, and gang activity**.
- The approach may also allow you to identify those who may be at risk of gang-association and/or being exploited by gangs.
- The technique will generate diagrams that will show the relationships between individuals that are contained in your data, this could include: criminal links, social links, potential feuds, etc.
- SNA diagrams can include names, pictures and further details of individuals as required.

Key network statistics 1:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491572/socnet_howto.pdf

Statistic	Shows	Explanation
Size		
Number of **nodes** - the people in the network	Size of the network	Number of individuals in the network
Number of **links** - social connections/relationships between nodes (e.g. friendship, family ties)	How 'busy' the network in total	Number of relationships between individuals in the network (in total)
Number of unique links	How 'busy' the network is, taking out relationships that are duplicated	Number of relationships between individuals in the network, with duplicates removed
Cohesiveness		
Number of **components** – distinct groups in the network	Whether there may be sub-groups in the network	Number of discrete groups in the network
Density	The extent to which nodes are interconnected – lower density networks have fewer links between nodes	The proportion of all links that are actually present

Key network statistics 2:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491572/socnet_howto.pdf

Diameter	Size of the network	Greatest number of steps between any pair of nodes
Mean average distance between nodes	How 'close' (in network terms) the nodes are to each other	Average number of steps needed to go from one node to any other
Centrality		
Mean degree	How central (on average) nodes in the network are	Average number of links that pass through the nodes
Mean betweeness	How central (on average) nodes in the network are	Average number of unique paths that pass through the nodes

SNA Basics:

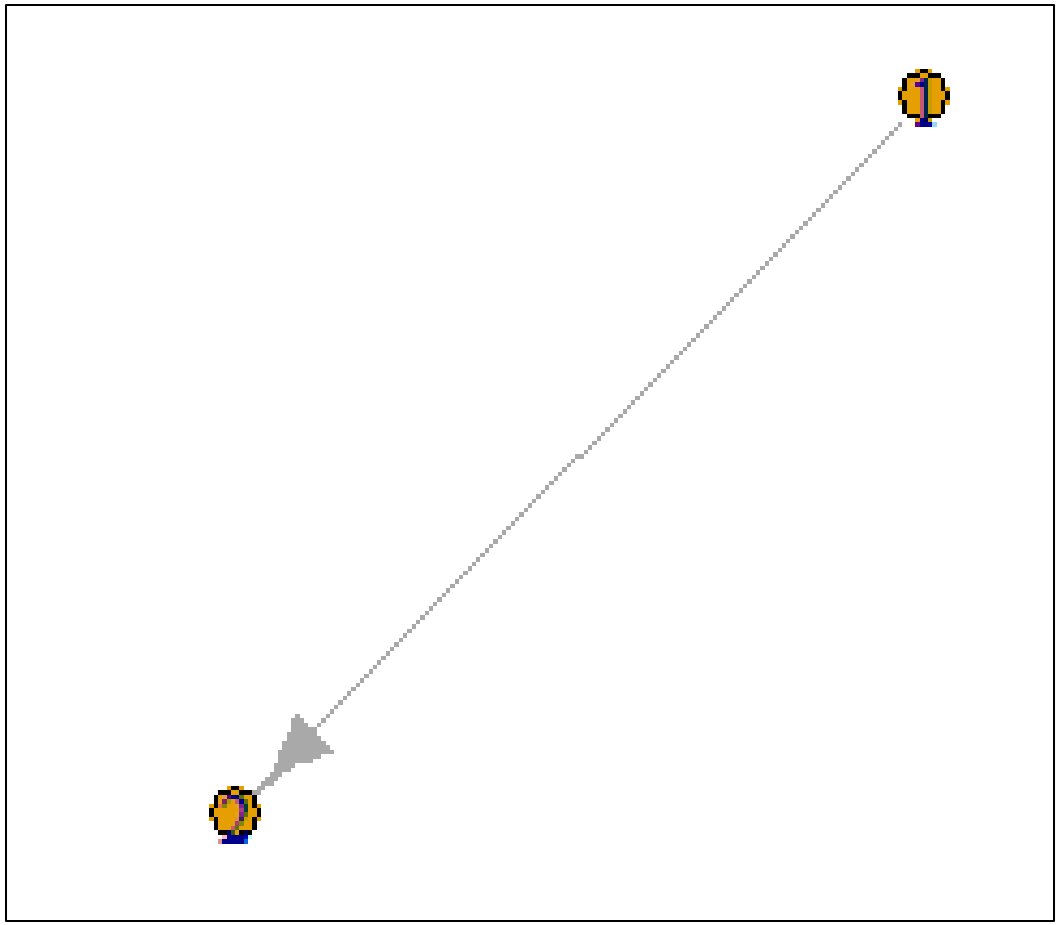
https://www.youtube.com/watch?v=0xsM0MbRPGE

library(igraph)

```
g <- graph(c(1,2))
plot(g)
```

- First **node** contains 1
- Second **node** contains 2

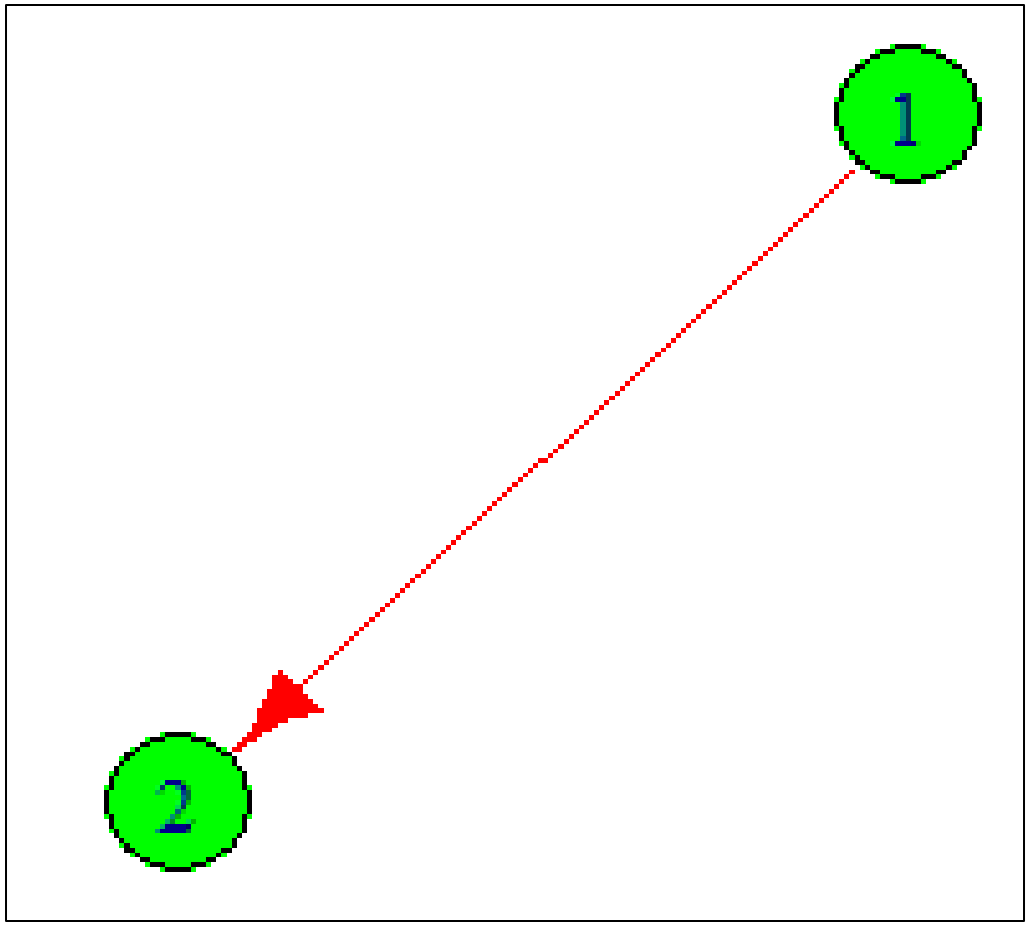
- The arrow (**edge**) goes from 1 to 2 as we defined that way in g!



SNA Basics: Changing size and color of node (vertex) and edge

```
plot(g,
        vertex.color = “green”,
        vertex.size = 40,
        edge.color = “red”,
        edge.size = 20)
```

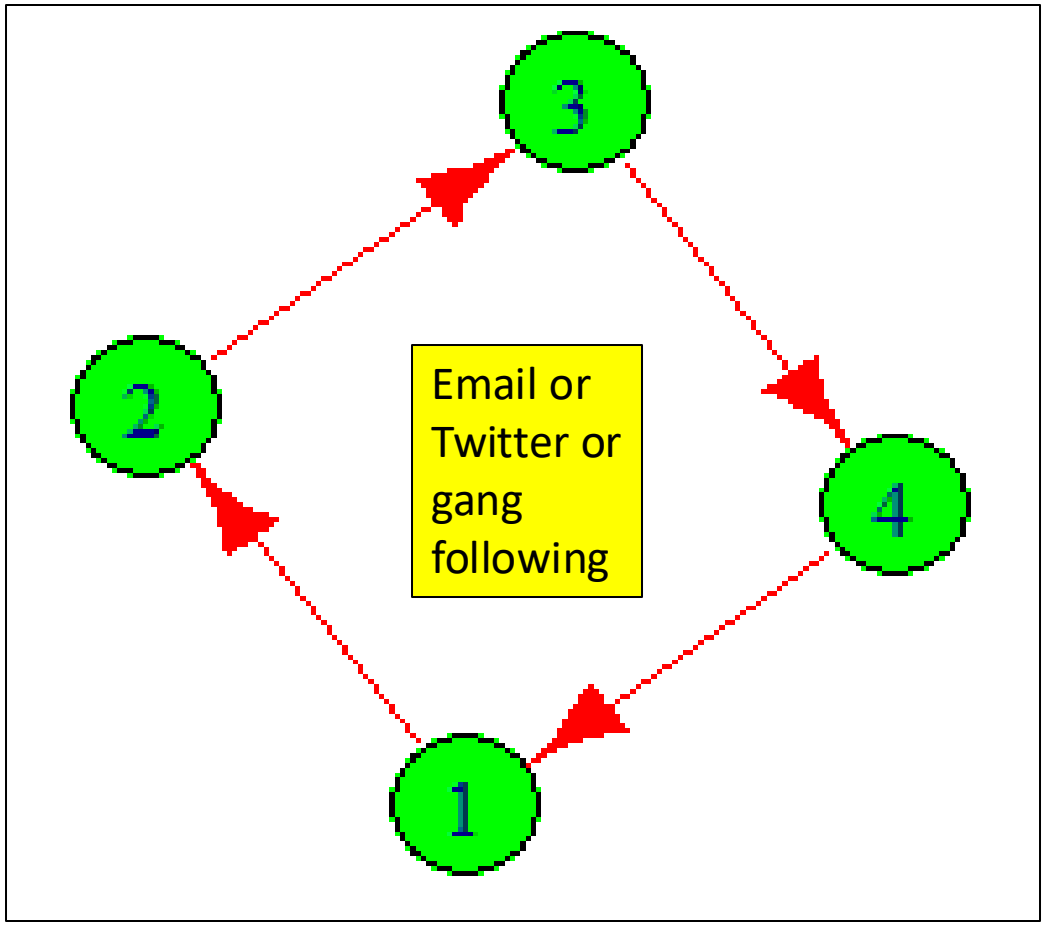
Note: Here information (email, twitter following, gang following) is flowing from 1 to 2!



SNA Basics: Adding more data points

```
g <- graph(c(1,2,2,3,3,4,4,1)
plot(g,
        vertex.color = “green”,
        vertex.size = 40,
        edge.color = “red”,
        edge.size = 20)
```

Note: This is a directed graph as we can see “arrow” here.

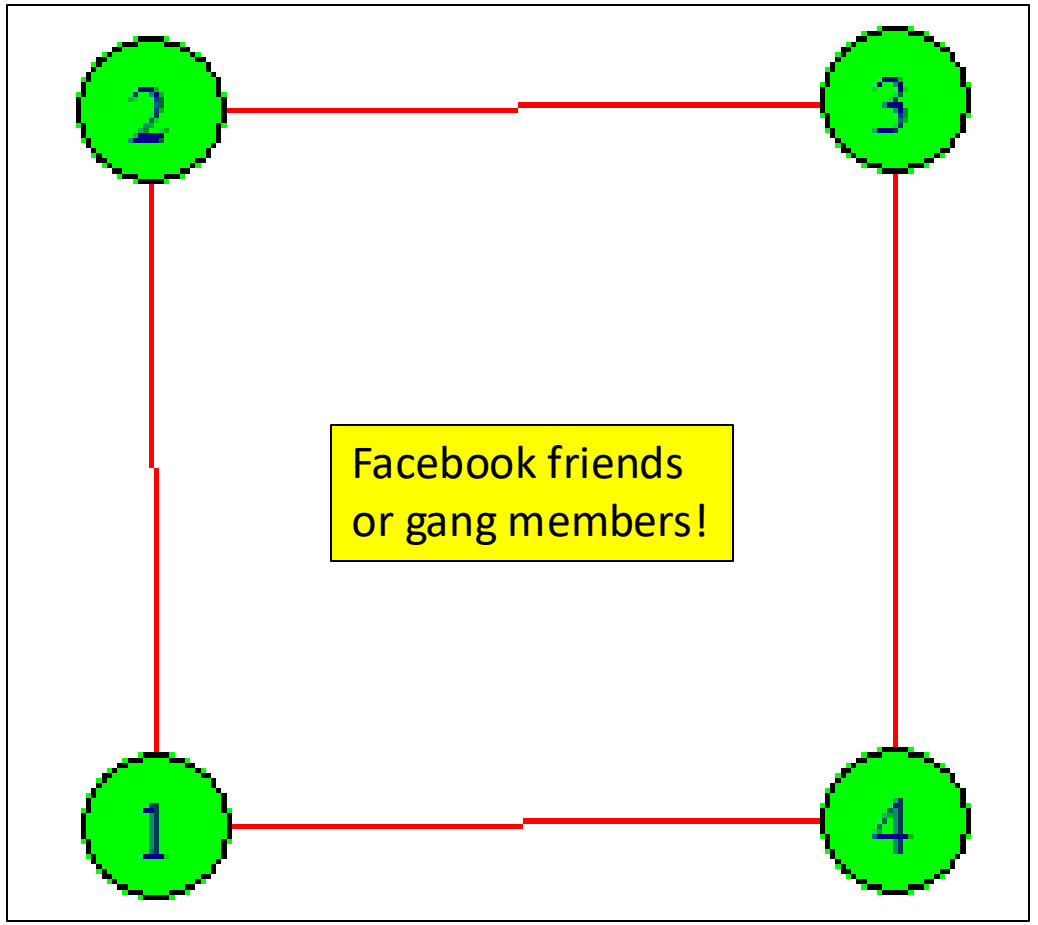


SNA Basics: Undirected data points

```
g <- graph(c(1,2,2,3,3,4,4,1),
directed = F)

plot(g,
        vertex.color = “green”,
        vertex.size = 40,
        edge.color = “red”,
        edge.size = 20)
```

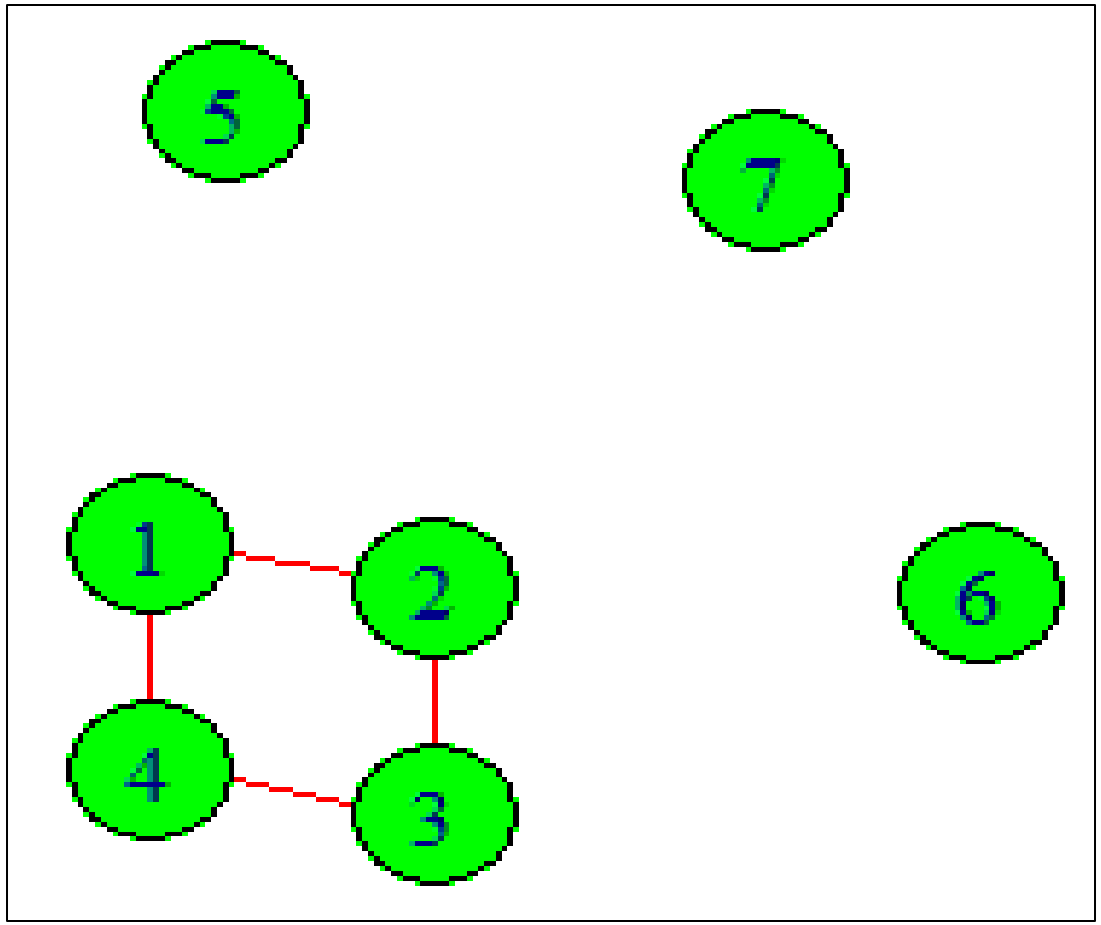
Note: This is not a directed graph as we cannot see “arrow” here.



SNA Basics: Adding related & unrelated nodes

```
g <- graph(c(1,2,2,3,3,4,4,1),
directed = F, n=7)
plot(g,
        vertex.color = “green”,
        vertex.size = 40,
        edge.color = “red”,
        edge.size = 20)
```

Note: Three unrelated nodes are shown without links.



SNA Basics: Adding related & unrelated nodes

g[]

This will give us the matrix used to produce the earlier graph

The dimension of this matrix is 7x7

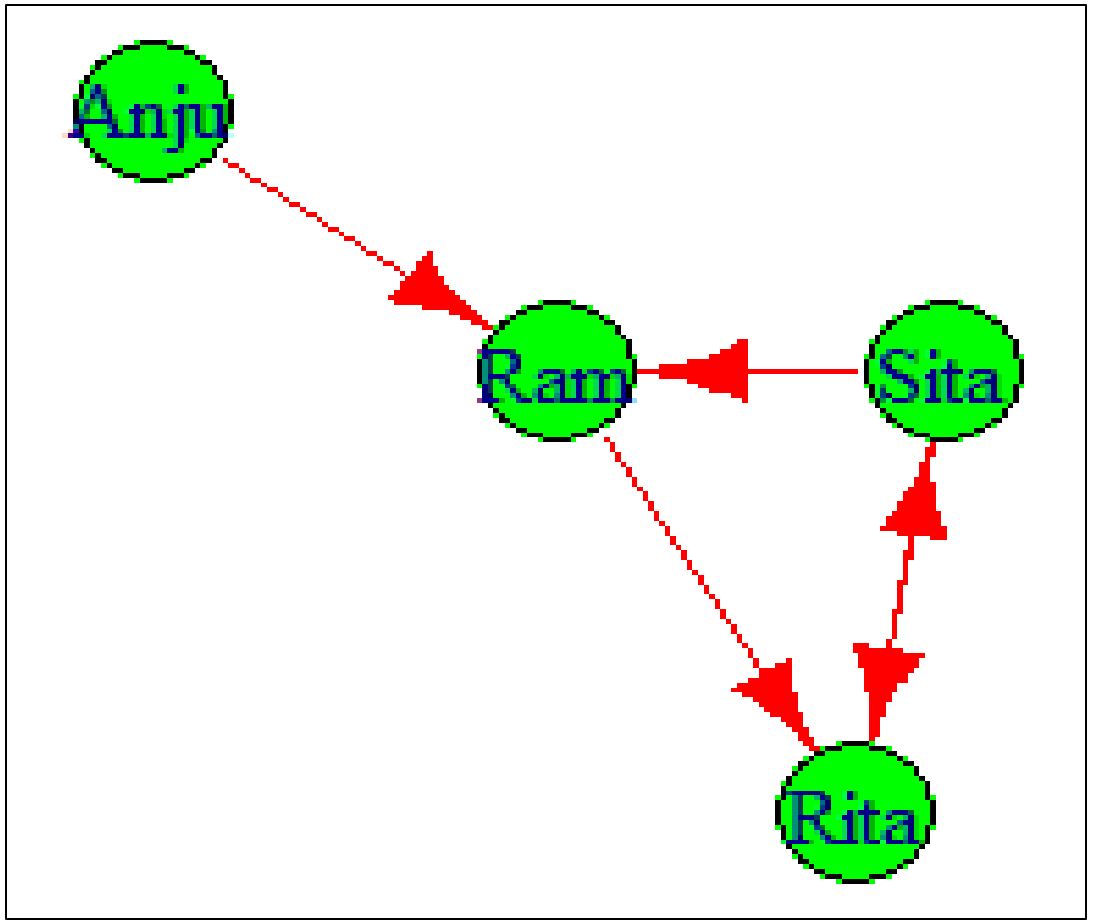
The dot(.) means no relation (connection) and 1 mean the connection with the nodes e.g. 1 has connection with 2 and 4

7 x 7 sparse Matrix of class "dgCMatrix"

[1,] . 1 . 1 . . .
[2,] 1 . 1 . . . .
[3,] . 1 . 1 . . .
[4,] 1 . 1 . . . .
[5,] . . . . . .
[6,] . . . . . .
[7,] . . . . . .

SNA Basics: Defining nodes with text data

```
g1 <-
graph(c("Sita","Ram","Ram","Rita"
,"Rita","Sita","Sita","Rita", "Anju",
"Ram"))
plot(g1,
    vertex.color = "green",
    vertex.size = 40,
    edge.color = "red",
    edge.size = 5)
```



SNA Basics: Getting info of “g1”

g1

D=Directed, N=Names
4 = Four vertices (nodes)
5 = Five edges (lines)

Pairs: Sita->Ram
Ram->Rita
Rita->Sita
Sita->Rita
Anju->Ram

Output in R:

IGRAPH 0adac86 DN-- 4 5 --
+ attr: name (v/c)
+ edges from 0adac86 (vertex names):
[1] Sita->Ram Ram ->Rita Rita->Sita Sita->Rita Anju->Ram

SNA Basics: Getting degrees of “g1”

degree(g1) or degree(g1, mode=“all”)

Sita Ram Rita Anju

3 3 3 1

degree(g1, mode=“in”)

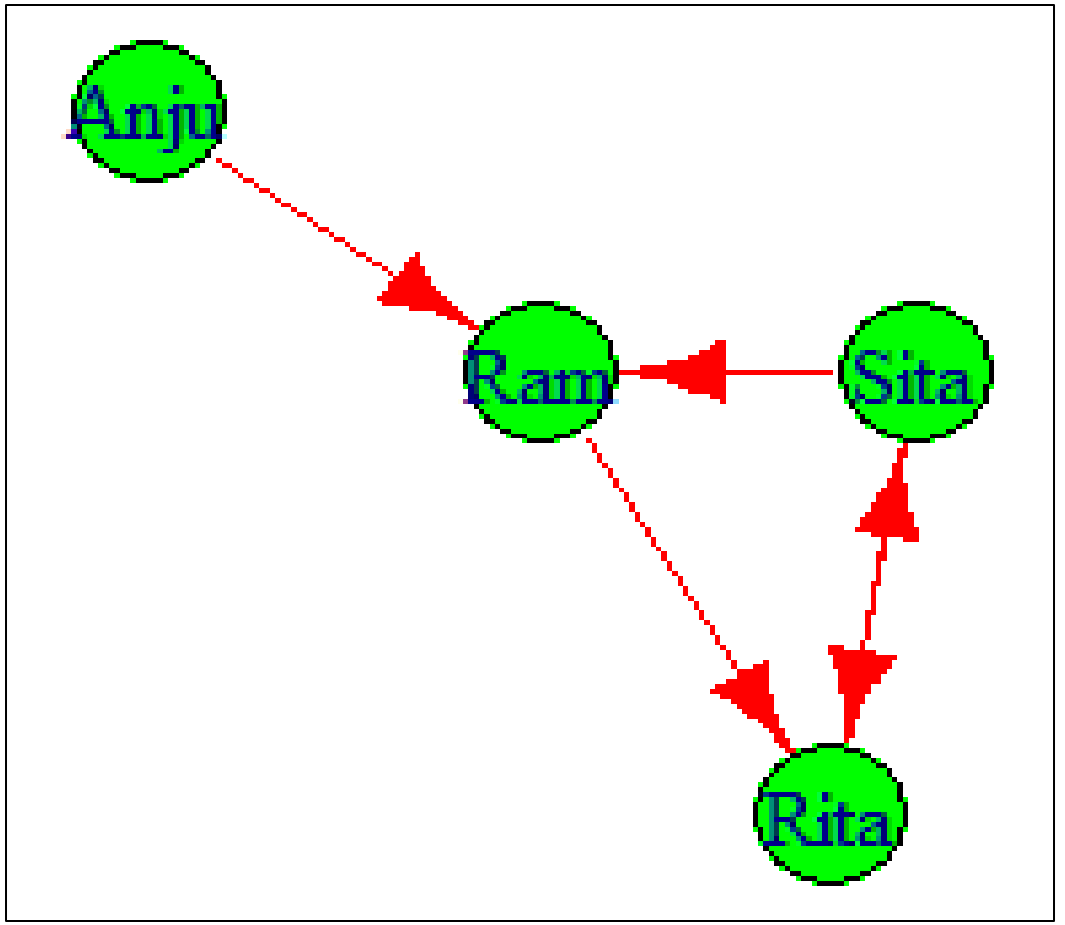
Sita Ram Rita Anju

1 2 2 0

degree(g1, mode=“out”)

2 1 1 1

“degree” means = Number of connections for each node



SNA Basics: Getting diameter of “g1”

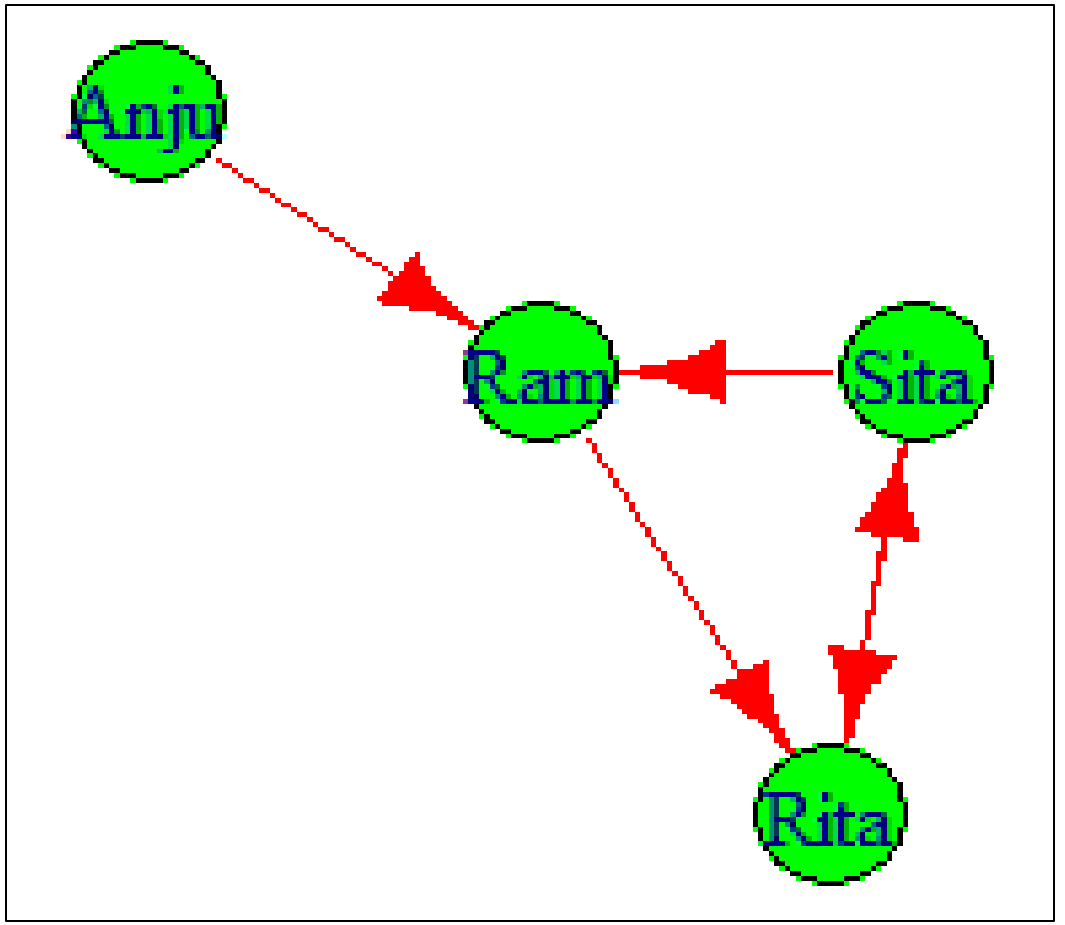
#Diameter

diameter(g1, directed = F, weights = NA)

[1] 2

“diameter” means = number of edged inside and outside of SND
i.e. Anju -> Ram and Ram -> Rita

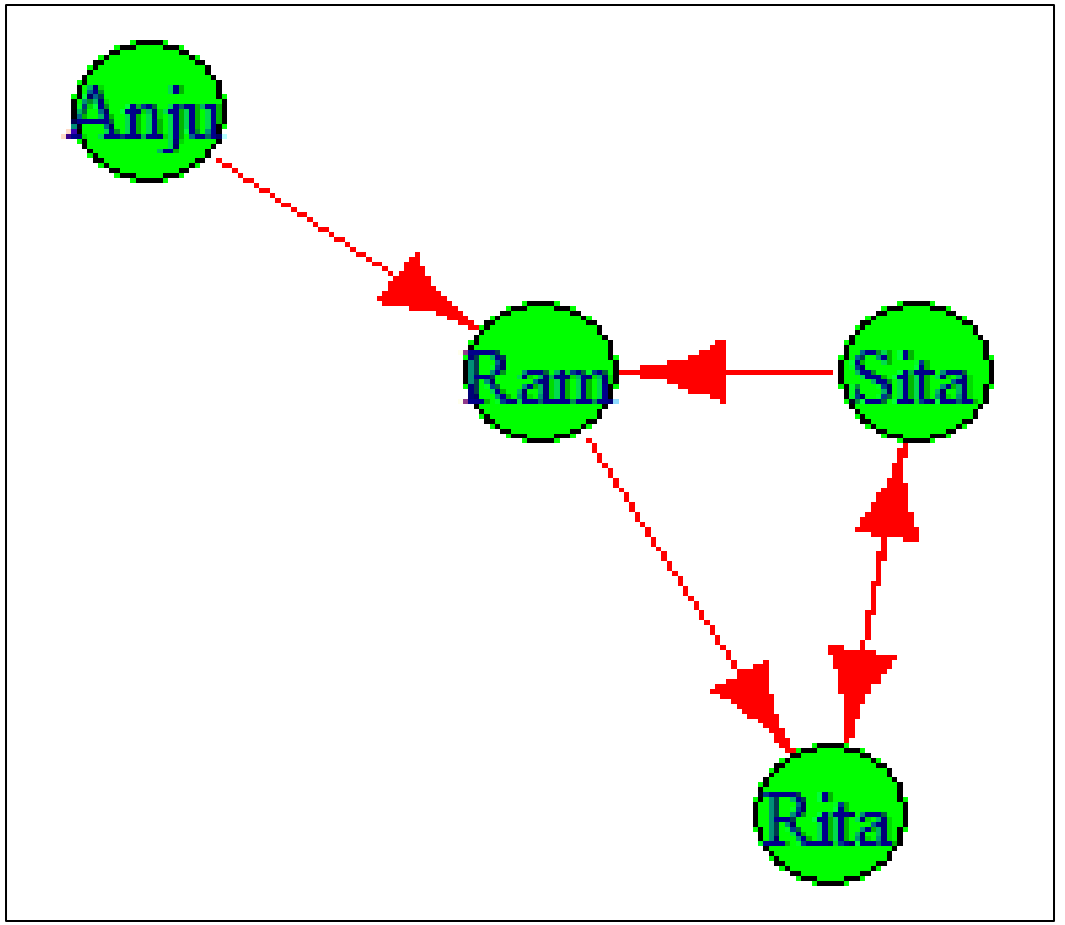
Or Anju -> Ram and Ram -> Sita



SNA Basics: Getting edge density of “g1”

```
#Edge density
edge_density(g1, loops = F)
[1] 0.4166667


#Edge density
ecount(g1)/(vcount(g1)*(vcount(g1)
-1))
5/4*(4-1)
[1] 0.4166667
```



SNA Basics: Getting reciprocity of “g1”

#Reciprocity of directed graph

#Percentage reciprocated ties

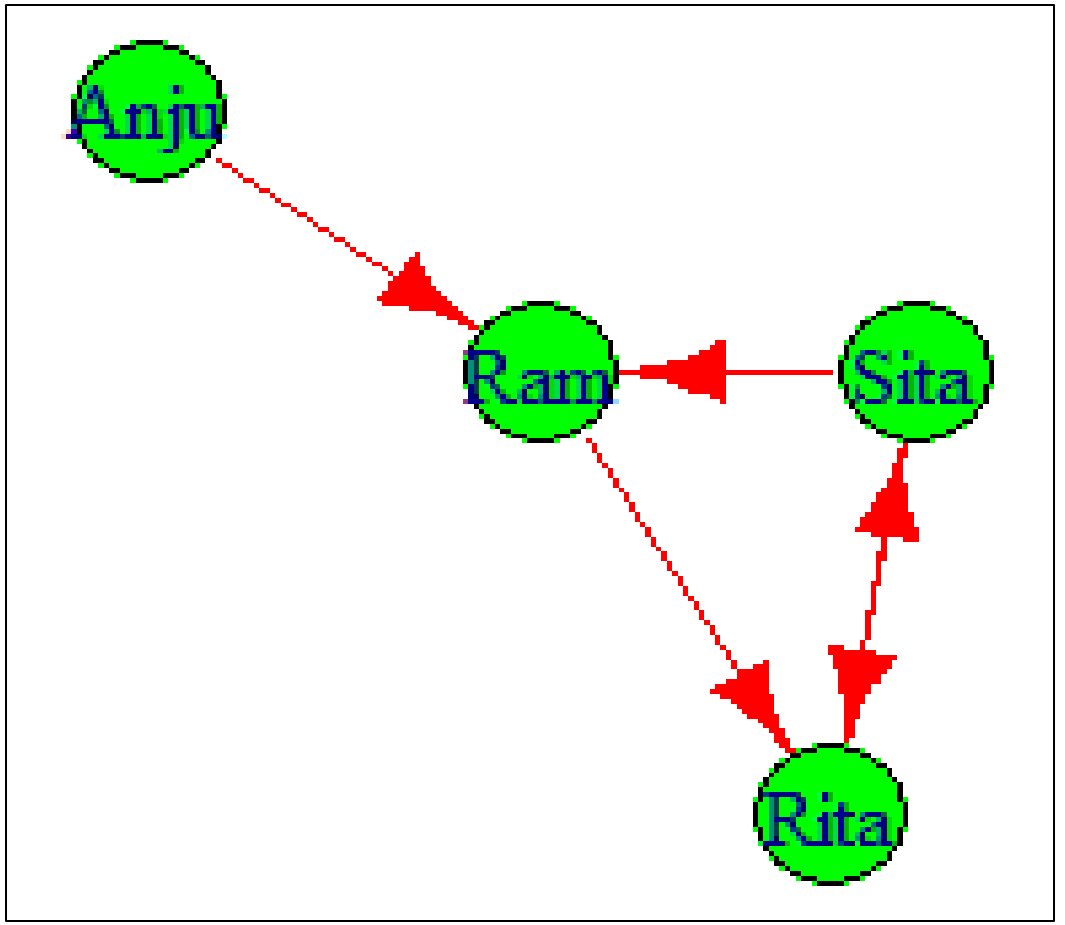
reciprocity(g1)

[1] 0.4

Total edges = 5

Tied edges = 2

Reciprocity = 2/5 = 0.4



SNA Basics: Getting closeness of “g1”

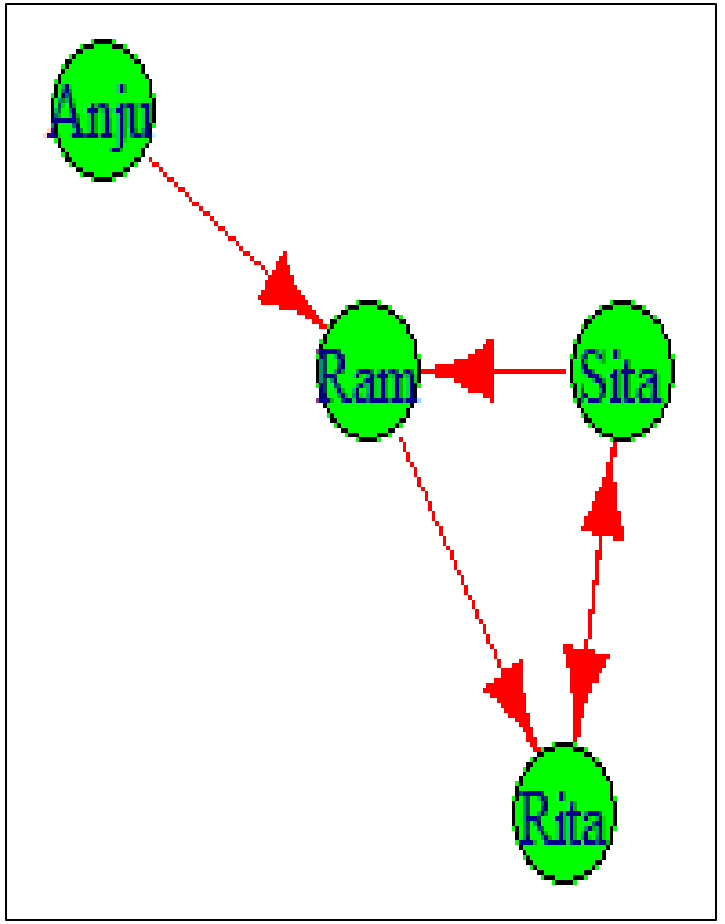
#Closeness

closeness(g1, mode = "all", weights = NA)

Sita	Ram	Rita	Anju
0.2500000	0.3333333	0.2500000	0.2000000

Ram is closest to other three persons

Anju is farthest to other three persons



SNA Basics: Getting betweenness of “g1”

#Betweenness

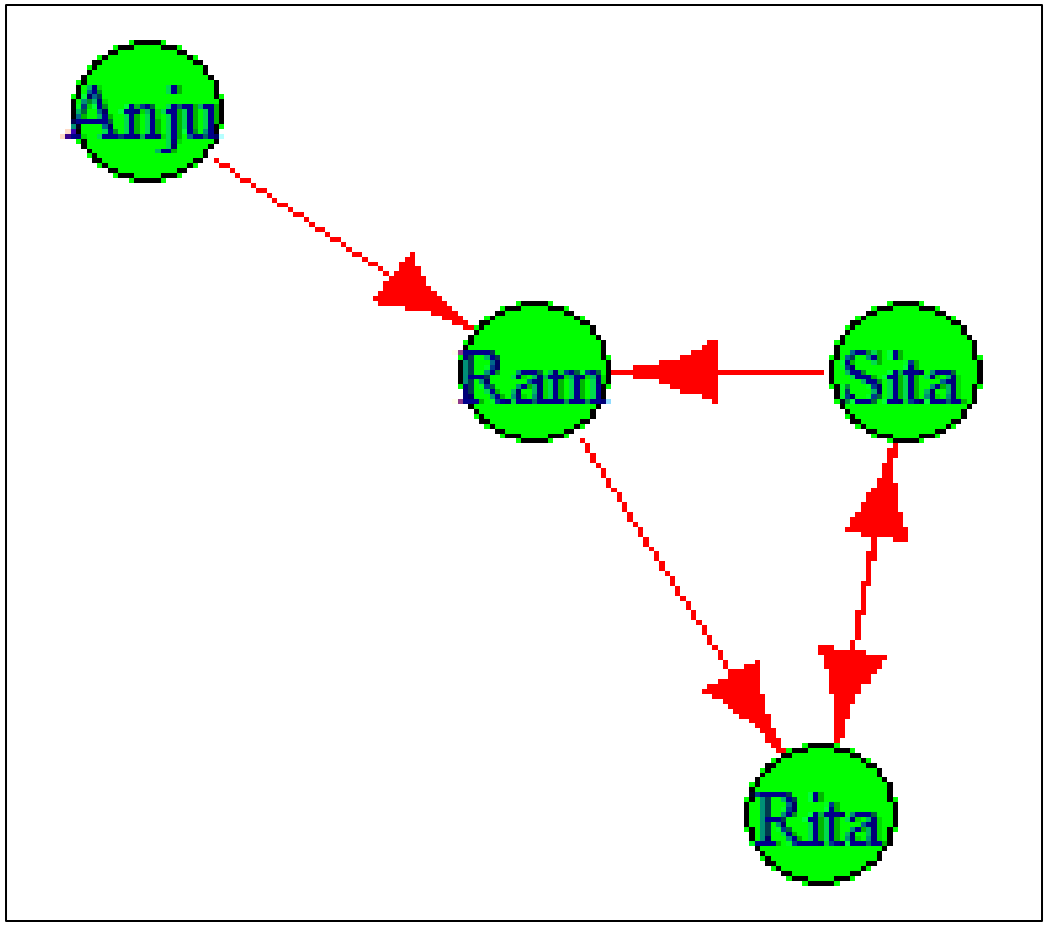
betweenness(g1, directed = T, weights = NA)

Sita	Ram	Rita	Anju
1	2	2	0

Ram and Rita have two “inner” edges, Sita has 1 and Anju has 0!

edge_betweenness(g1, directed = T, weights = NA)

2 4 4 1 3 **#Learn on your own!**



Question/queries so far?

- More are here: https://igraph.org/r/html/latest/

Self-Practice: SNA with “SNA_School.csv” data file

Follow this video: https://www.youtube.com/watch?v=0xsM0MbRPGE

```
#Read the data in R

data <- read.csv(file.choose(), header=T)

#Save the first two columns as y

y <- data.frame(data$first, data$second)

#Save it as network graph data

net <- graph.data.frame(y, directed=T)
```

first	second	grade	spec
AA	DD	6	Y
AB	DD	6	R
AF	BA	6	Q
DD	DA	6	Q
CD	EC	6	X
DD	CE	6	Y
CD	FA	6	X
CD	CC	6	W
BA	AF	6	R
CB	CA	6	T
CC	CA	6	U
CD	CA	6	Q
BC	CA	6	U
DD	DA	6	Y
ED	AD	6	R
AE	AC	6	Z
AB	BA	6	Y
CD	EC	6	X
CA	CC	6	U

SNA with a data file: networkdata.csv

#Vertices – 52 unique vertices

V(net)

#Edges – 290 edges

E(net)

#Names as labels

V(net)$label #Result = NULL

#Define the labels

V(net)$label <- V(net)$name

V(net)$label # 52 vertices as labels

- + 52/52 vertices, named, from 58abab2:
- [1] AA AB AF DD CD BA CB CC BC ED AE CA EB BF BB AC DC BD DB CF DF BE EA CE EE EF
- [27] FF FD GB GC GD AD KA KF LC DA EC FA FB DE FC FE GA GE KB KC KD KE LB LA LD LE

SNA with a data file: networkdata.csv

#Define degree

V(net)$degree #Result = NULL

V(net)$degree <- degree(net)

V(net)$degree

What does it means here?

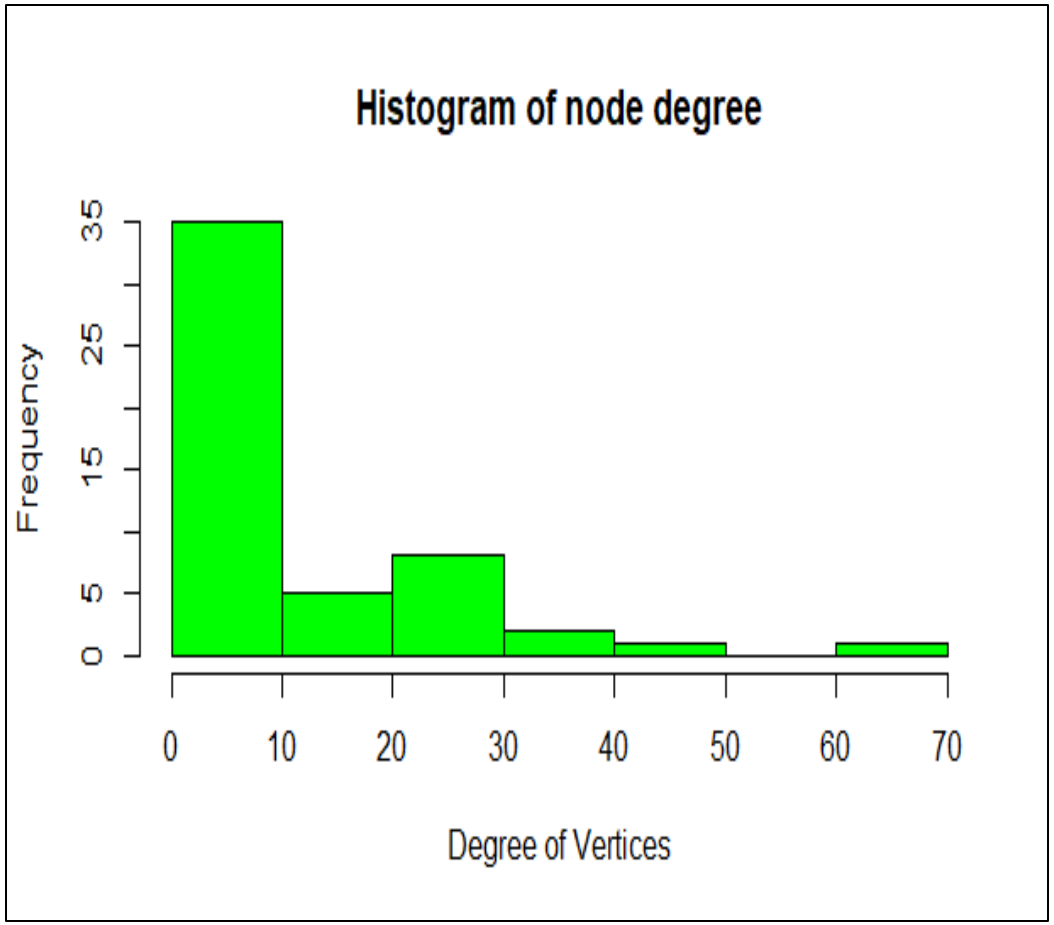
Number of connections for each nodes (vertices)

table(degree(net)) ???

- [1] 18 9 23 36 40 26 24 50 21 27 15 62 7 12 23 27 2 4 8 12 23 20 8 10 6 8
- [27] 1 8 1 1 1 9 3 3 1 7 3 1 1 2 1 2 5 1 1 1 1 1 1 1 1 1
- [1] "AA" "AB" "AF" "DD" "CD" "BA" "CB" "CC" "BC" "ED" "AE" "CA" "EB" "BF" "BB" "AC" "DC" "BD" "DB" "CF" "DF" "BE" "EA" "CE" "EE" "EF"
- [27] "FF" "FD" "GB" "GC" "GD" "AD" "KA" "KF" "LC" "DA" "EC" "FA" "FB" "DE" "FC" "FE" "GA" "GE" "KB" "KC" "KD" "KE" "LB" "LA" "LD" "LE"

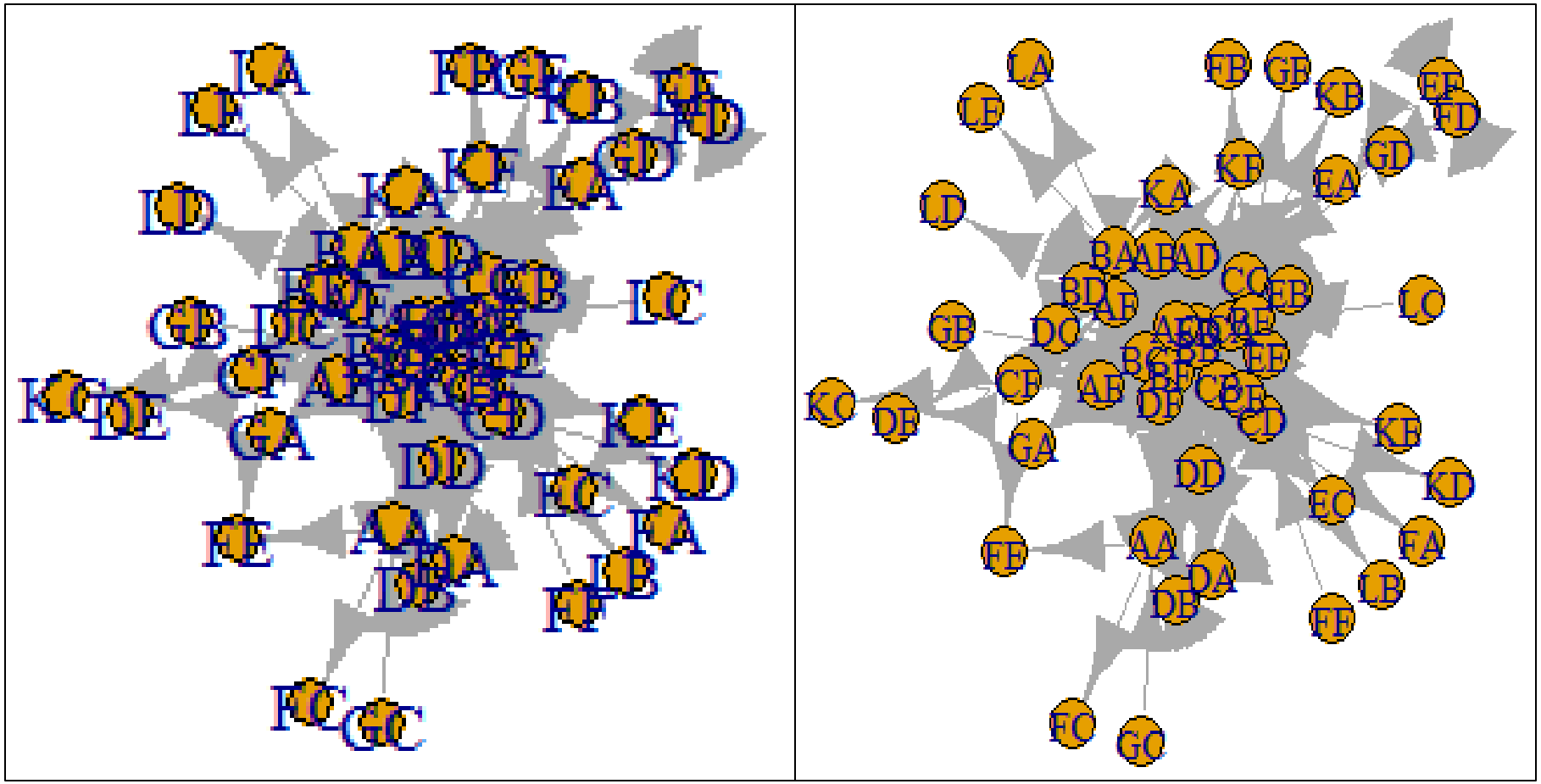
Histogram of node degree i.e. connections

- #Histogram of node degree
- hist(V(net)$degree,
  col = "green",
  main = "Histogram of node degree",
  ylab = "Frequency",
  xlab = "Degree of Vertices")



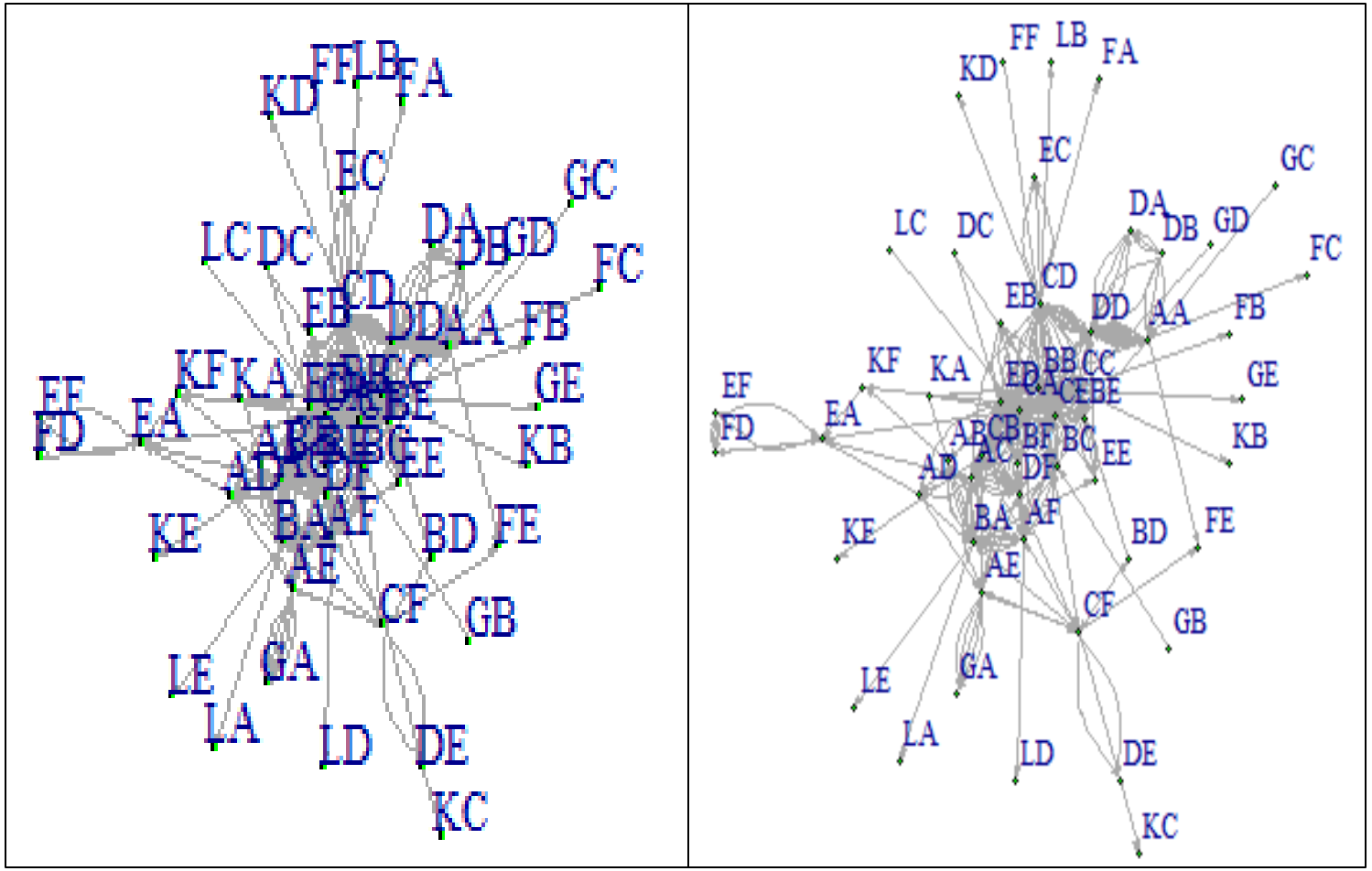
Network diagram:

- set.seed(222)
- plot(net)



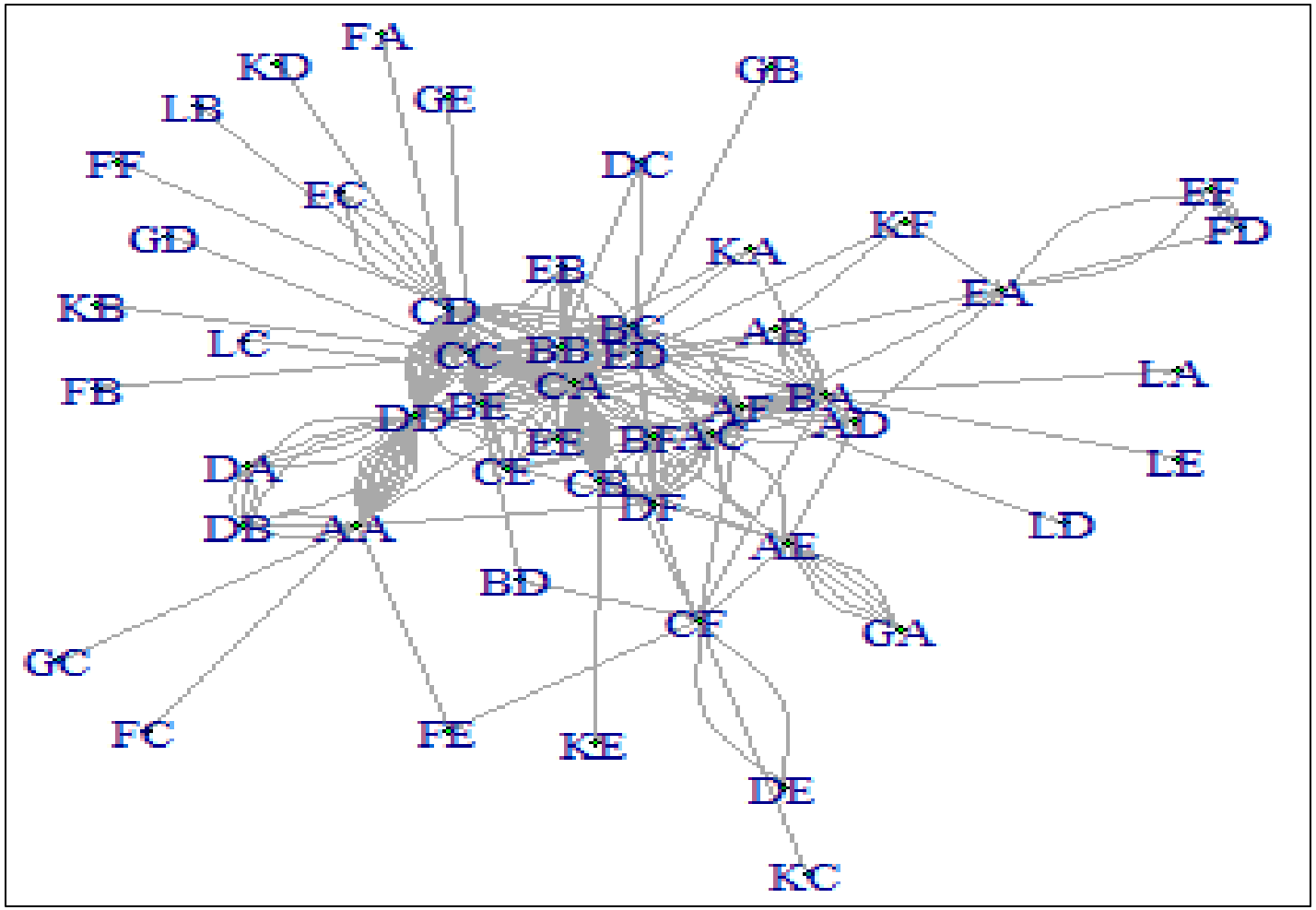
Network diagram: A bit of tweaking!

- plot(net,
- vertex.color = "green",
- vertex.size = 2,
- vertex.label.dist = 1.5,
- edge.arrow.size = 0.1,
- vertex.label.cex = 0.8)



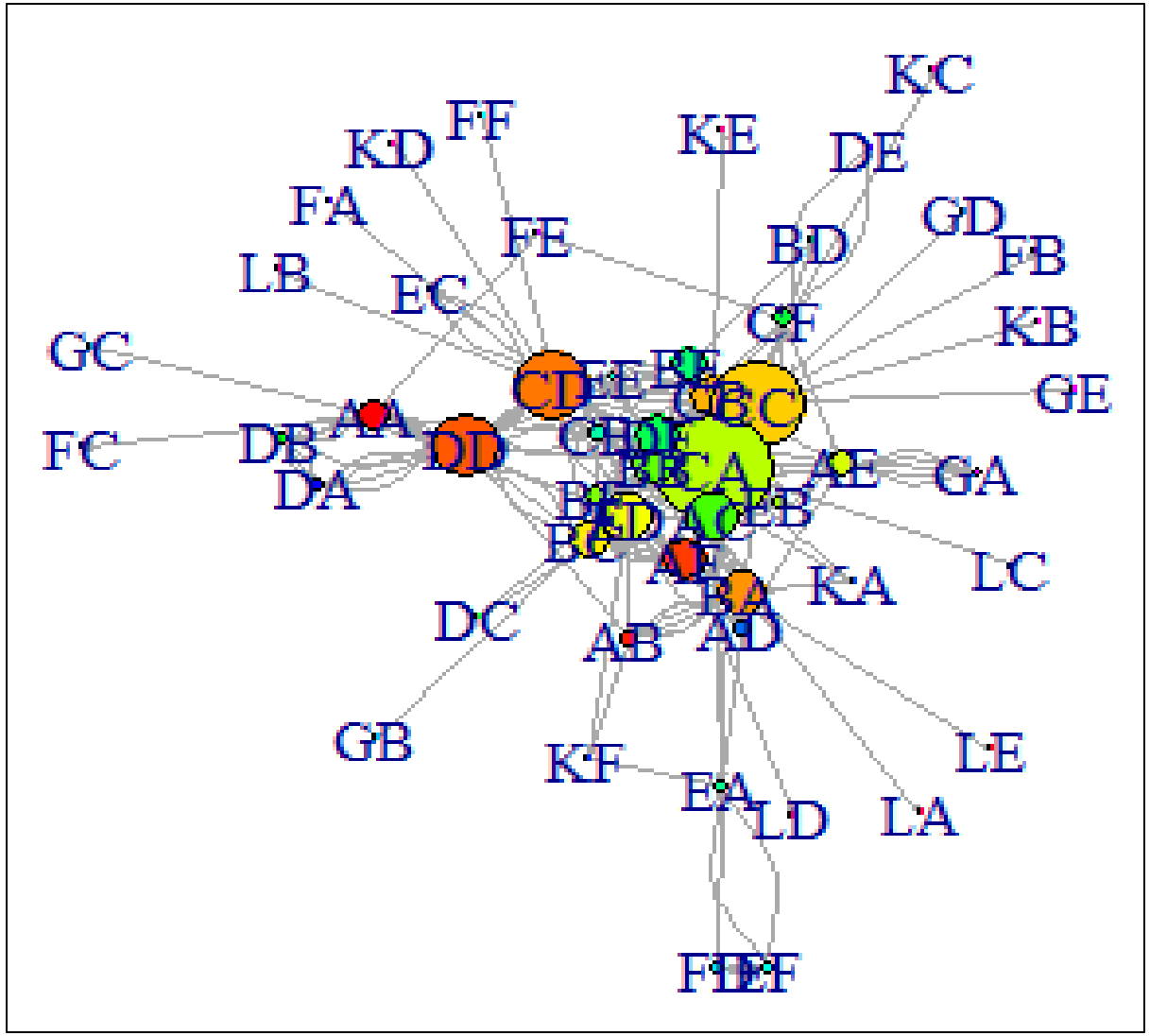
Network diagram: A little bit of tweaking!

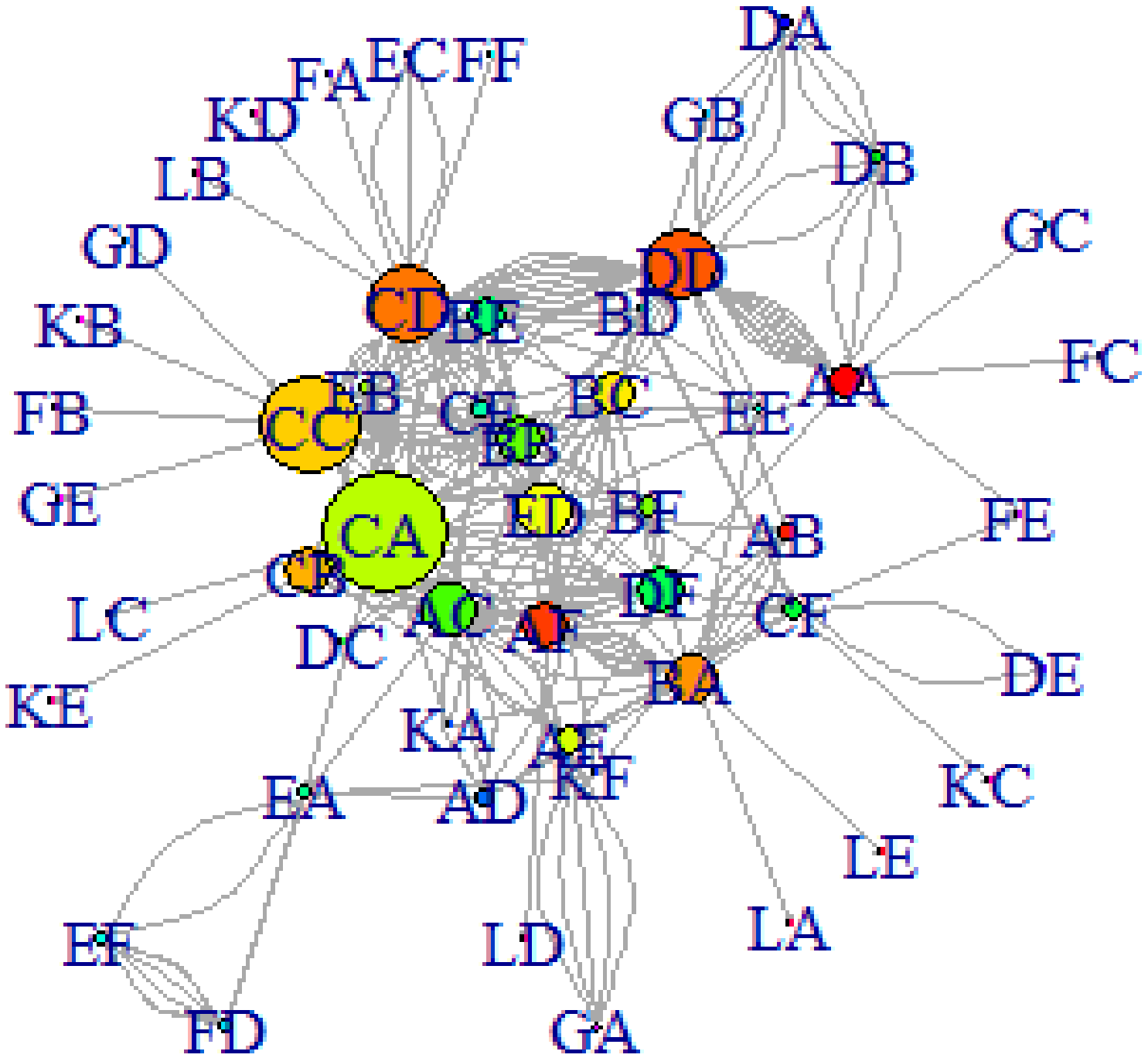
- plot(net,
- vertex.color = "green",
- vertex.size = 2,
- edge.arrow.size = 0.1,
- vertex.label.cex = 0.8)



Network diagram: layout 1!

```
plot(net,
vertex.color = rainbow(52),
vertex.size = V(net)$degree*0.4,
edge.arrow.size = 0.1,
layout=layout.fruchterman.reingold)
```






```
#Next layout i.e. layout 2
plot(net,
     vertex.color = rainbow(52),
     vertex.size = V(net)$degree*0.4,
     edge.arrow.size = 0.1,
     layout=layout.kamada.kawai)
```

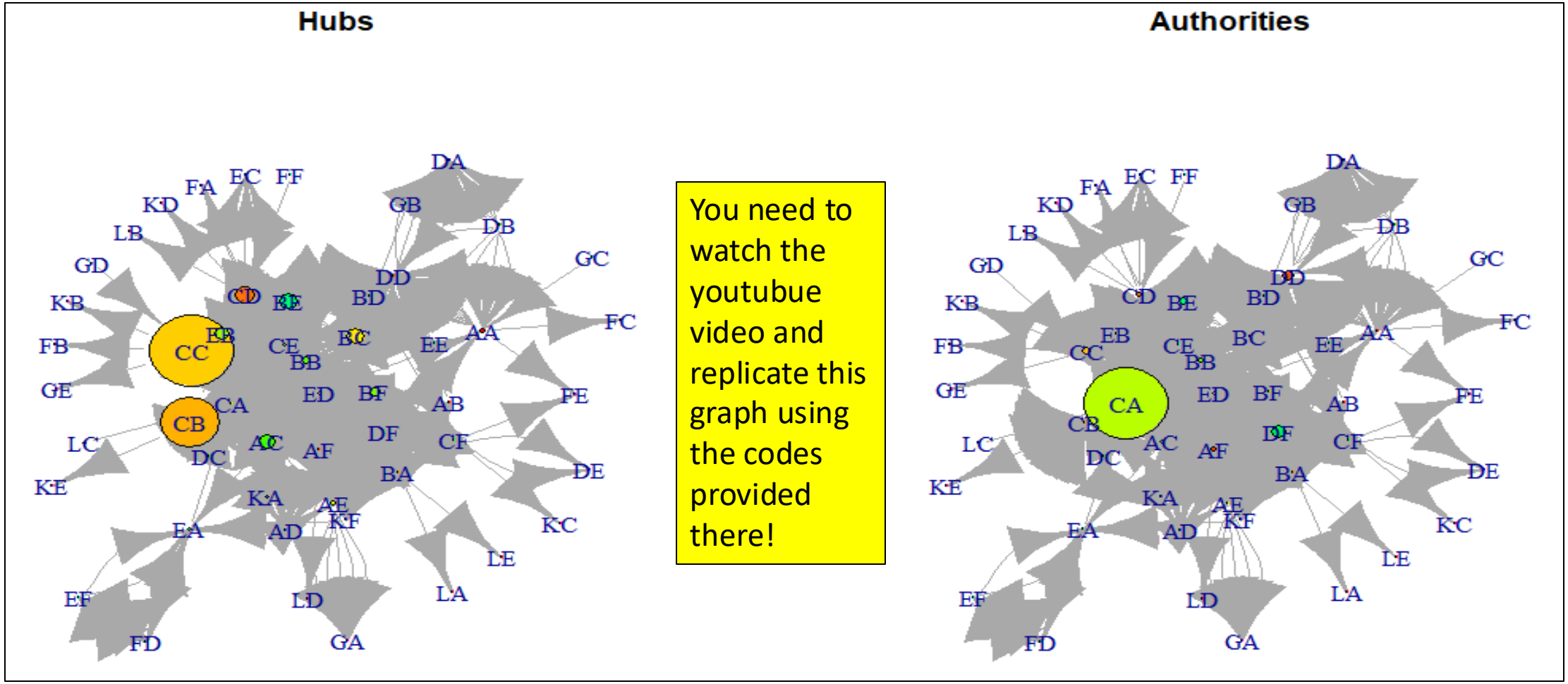
Which nodes are "hubs"?

- **Nodes with most outer edges**
- **We need "hub score"**

Which nodes are "authorities"?

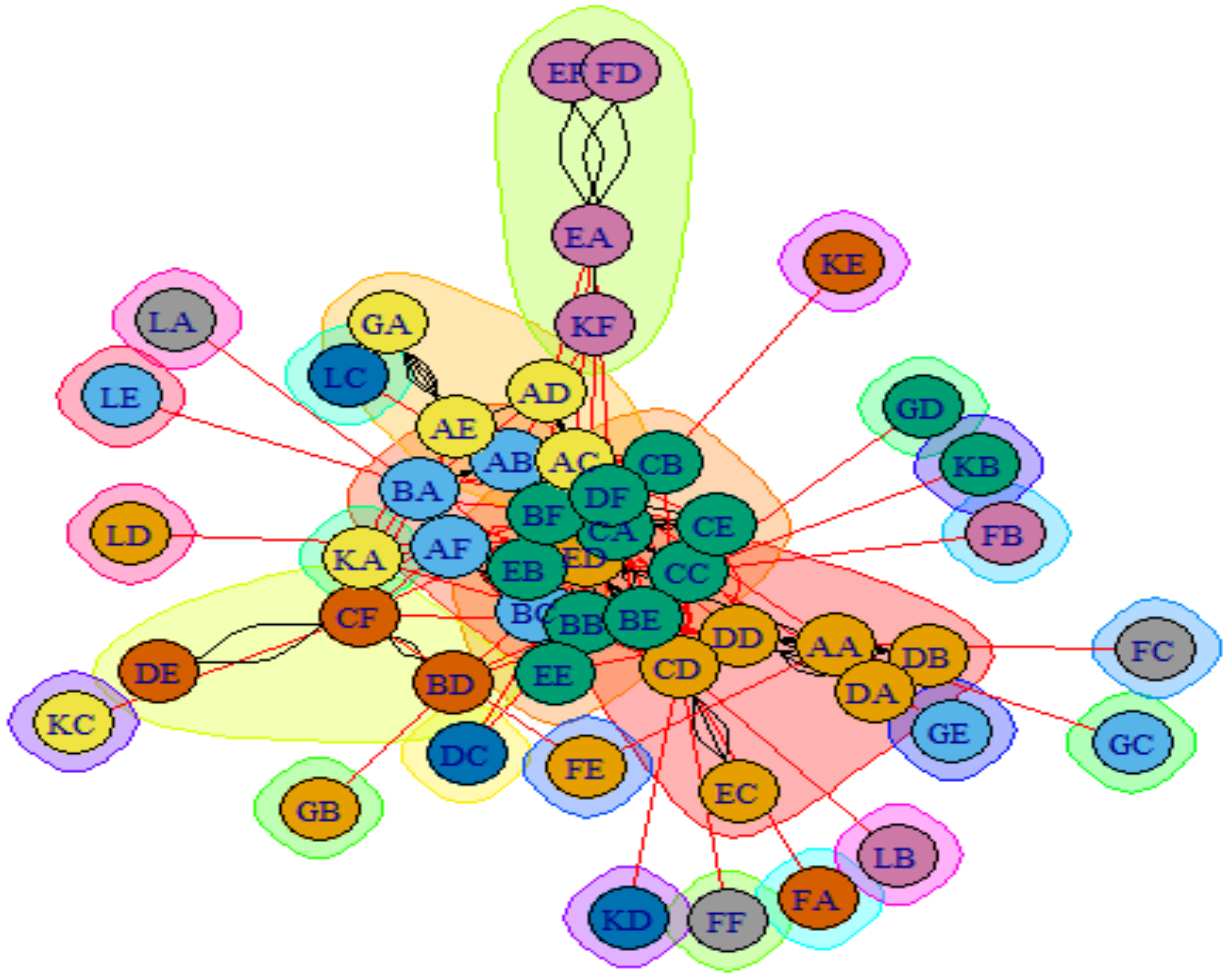
- **Nodes with most inner edges**
- **We need "authority score"**

Hubs and authorities: With hub score & authority scores of the network data



https://www.youtube.com/watch?v=0xsM0MbRPGE

Community (cluster) detection:




```
#Community detection
net <- graph.data.frame(y, directed = F)
cnet <- cluster_edge_betweenness(net)
plot(cnet,
    net,
    vetex.size = 10,
    vertex.label.cex = 0.8)
```

cluster_edge_betweenness is the function available in the igraph package to fit the clustering algorithm!

More are available here:
https://igraph.org/r/html/latest/

Resources:

- **Read and learn about “sna” package on your own! Also check the reading on facebook social network analysis and try to do the same with your own facebook data!**

- **SNA statistics: https://www.latentview.com/blog/a-guide-to-social-network-analysis-and-its-use-cases/**

- **Application of SNA:**
  - **https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-019-1599-6**
  - **https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0264-6**

Question/Queries?

Next class:

- **Grammar of graphics**
- ggplot2 packages and its use in R
- **Read Chapter 3: Data Visualization with ggplot2 of your course text book carefully before coming to the next class**
- Link: https://r4ds.had.co.nz/data-visualisation.html

Thank you!

@shitalbhandary