The Fundamentals of Network Analysis

Dr. Suat ATAN

19.01.2021

Section 1: Warm-up

Introduction

Dr. Suat ATAN

www.suatatan.com

suatatan.wordpress.com

Dr. Suat ATAN
www.suatatan.com
suatatan.wordpress.com

I am delighted to be here today to tell you about Network Analysis. I'm Suat ATAN. I want to introduce myself briefly and then I want to get to know you. I have a PhD in Business Administration at Ankara University. I studied text mining. Before that, I studied engineering. I have three books in total on the R language, Python language and Network analysis. I am currently doing postdoctoral research at the University of Turin in Italy on a scholarship. I work as a software developer in a public institution. Now let's get to know you before moving on to the course contents.

What are networks?

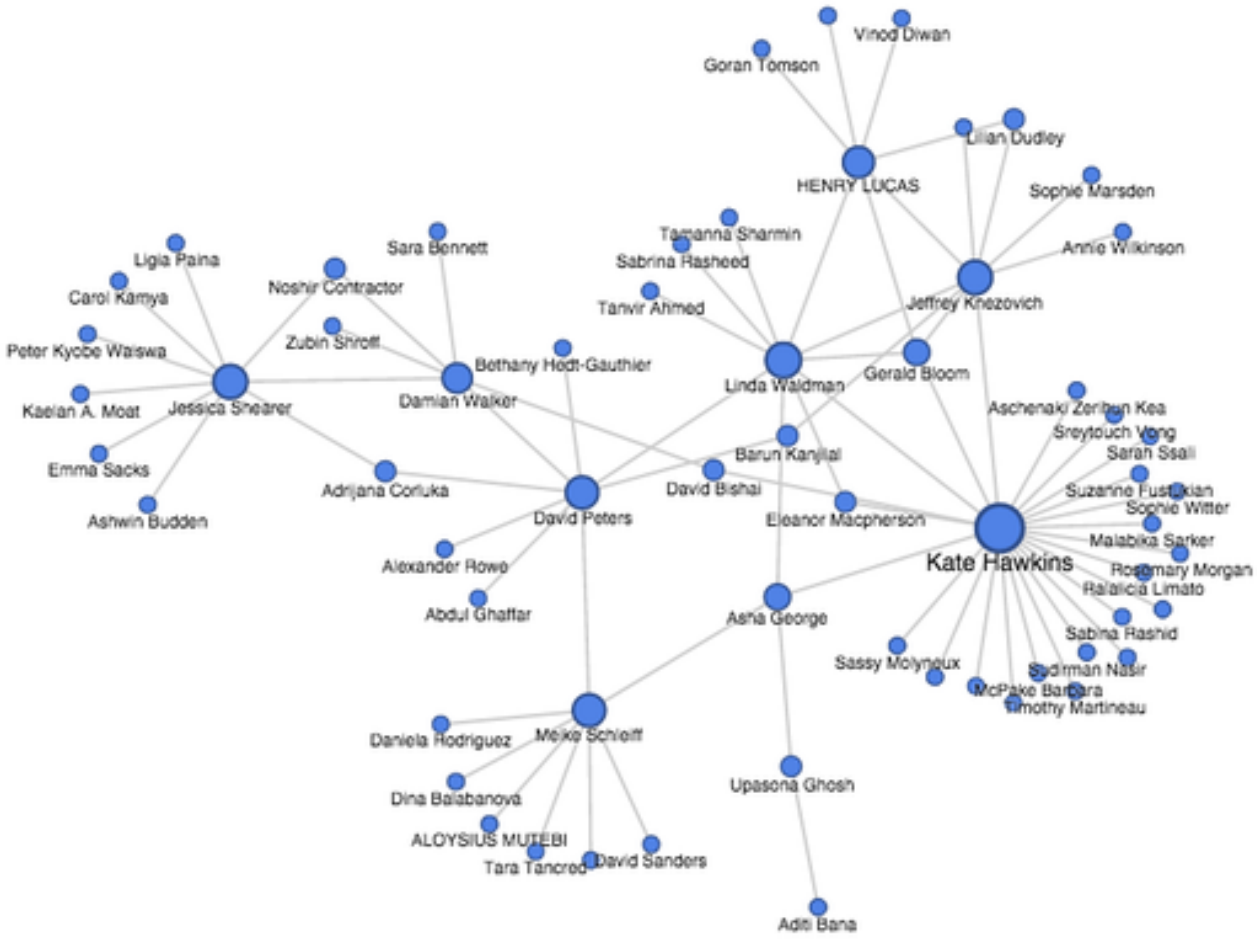


Figure 1: A network

Where are networks?

All around us are networks. Naturally, individuals organize networks themselves. Our families and friends are forming our social groups around each of us. We are living in the networks like:

- Twitter, Facebook and other social media
- Friendships
- Telephone communications

Where are networks?

All around us are networks. Naturally, individuals organize networks themselves. Our families and friends are forming our social groups around each of us. We are living in the networks like:

- Twitter, Facebook and other social media
- Friendships
- Telephone communications

What is Social Network Analysis? You've heard of social networks before from the online versions like Facebook and Twitter. Ans also more traditional social networks of people in a fraternity or in a small town. Clubs, groups and etc. You've probably also heard about some of the interesting things you can learn about these networks when you analyze them, from silly things like hearing how many degrees of any actor one actor is from another -odd but slightly creepy- like how Facebook algorithms try to predict products you wanna buy or who to add as a friend.

When you start to think about the world in terms of different overlaying networks that connect and transfer friendships, information, money and power, you start to see how analyzing things through the analysis of social networks can lead to new realization about culture, politics, history and lots of other interesting topics.

We have intuitive sense that the connections of the people around us are huge factor and what we know how we think of what we do. But researchers using standard statistical methods don't have a really good

How big are networks?

Figure 2: Bigger networks

Figure 2: Bigger networks

We can find out how one person connected or disconnected from people, groups and trends in a population and all those people seemed to be friends with everyone. Well, social network analysis can generate graphical representations that reveal individuals in populations that bridge social groups.

With social network analysis, you can study how individuals divide their energies between different social groups over time. We can study what makes a group strangers, start to form statement groups what networks are firm, we can see how things like power, beliefs or even an outbreak of disease flows through the individual connections.

It is true not only for the individuals in network analysis we can refer points.

They may be countries too. The international trade and other transfer among countries are consist network.

Who or what is in the networks?

Figure 3: Human Network

Figure 3: Human Network

As you can tell, there are practical questions that you can find quantitative answers and new insights with social network analysis that just weren't possible before. This opens up an exciting range of new options. One example is in colonial terrorist organizations. How they connect? Who is the top banana? Are there any isolated persons? So any other researchers can try to find out is there any correlation with friendships and race or age or economical status?

What happens inside networks?

Figure 4: n3

Figure 4: n3

Social network analysis is a very open field and there are lots of technical options to try out. Like adding geographic mapping data to understand how physical environments change network dynamics. For instance in road design what happens if any place has new road connection to another place. And for math geeks, it's still a very new field with lots of room for creating new analytical algorithms so we can create new forms of mapping network connections. See the power of social network analysis in your own life right now by looking at the tools in this lecture and may be you'll wanna to try it or on your Facebook, Twitter, and LinkedIn networks.

Wrap-up: More examples about networks?

- Money transfers
- Drug sellers
- Stakeholders - Partnerships
- Terrorist organizations

You probably won't expect everything that you see which is one of the reasons that social network analysis is so exciting. If you're a faculty member, researcher, or graduate student interested in learning more, or applying this to your own research It's a useful pathway to see how this approach can stimulate new ideas for you. If you're loving math learn social networks, network theory or graph theory in their descriptions for deeper understanding. Research on Google Scholars to see more examples about network analysis including my studies. Networks are everywhere, what will you discover with them

What makes a thing 'network'?

Network is consist two fundamental element:

- Points = Asset = Actor = Vertex = **Node** = Person = Organization
- Line = Connection = **Edge** = Transfer = Relation

Let's ask again

How do they connect? What happens when they connect? In this series of lectures, we will focus on questions like.

- What is the connection?
- Who or what can connect?
- What happens when everything becomes connected

What makes a thing 'network'?

Network is consist two fundamental element:

- Points = Asset = Actor = Vertex = **Node** = Person = Organization
- Line = Connection = **Edge** = Transfer = Relation

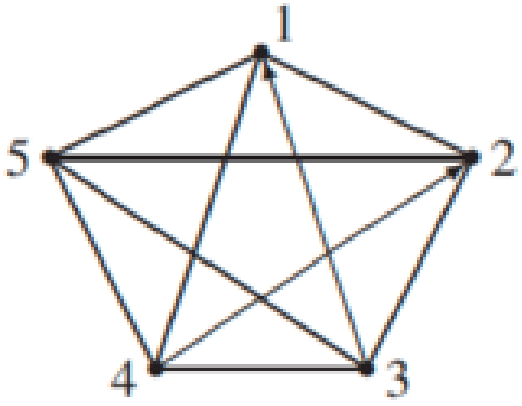
Let's ask again

How do they connect? What happens when they connect? In this series of lectures, we will focus on questions like.

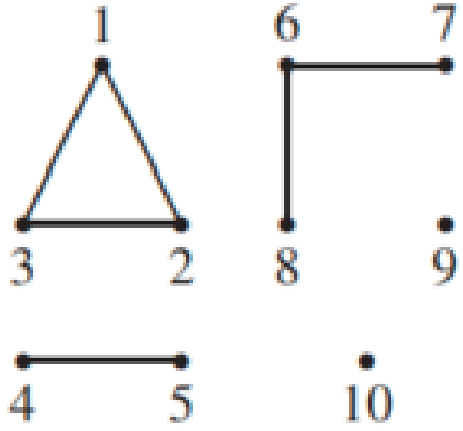
- What is the connection?
- Who or what can connect?
- What happens when everything becomes connected

We will use the bolder fonts in our lessons. We should focus the two elements that combines the network: points and connections or nodes and edges. IMPROMPTU

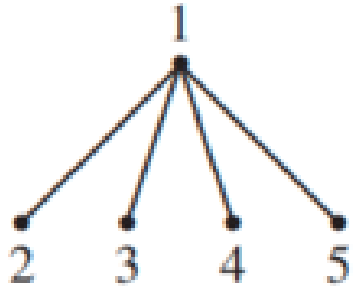
Network Imaginations



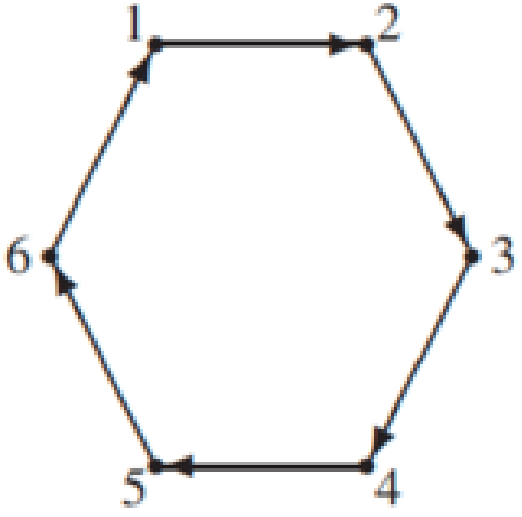
First network



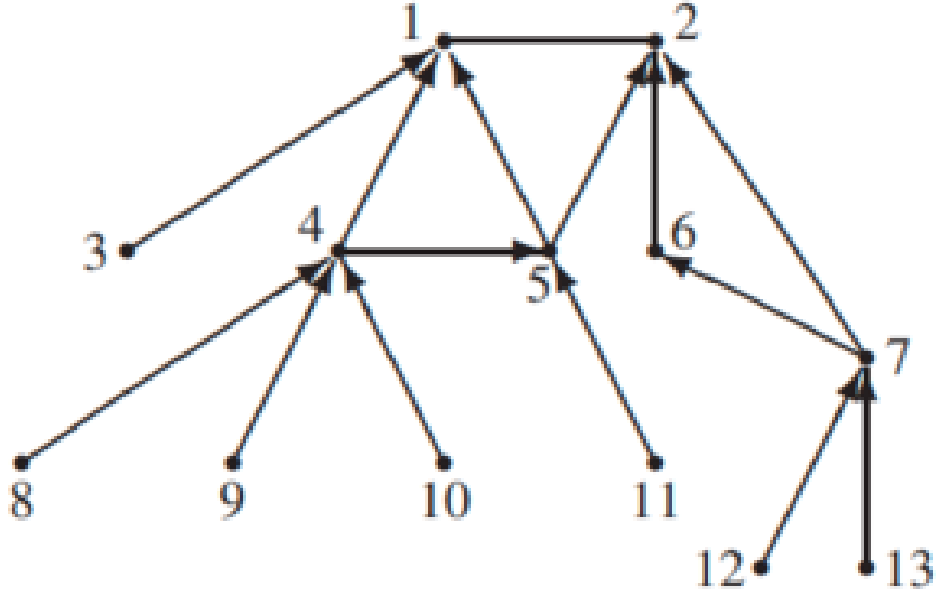
Second network



Third network

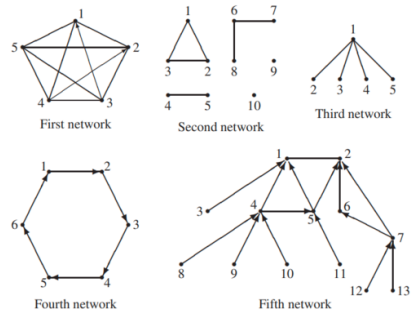


Fourth network



Fifth network

Network Imaginations



Let's get some sort of network structures. Let's ask ourselves where is the nodes and where is the edges. First are fully connected EXPLAIN Second includes no connection networks and three points becomes network as a triangular and just a form like F, EXPLAIN Third are an organizational network between a boss and subordinates ...NOTE there is directions some networks what are means? ASK I want to ask for you, what is meaning of arrow between points or persons can anyone give an example

Network Imaginations 2

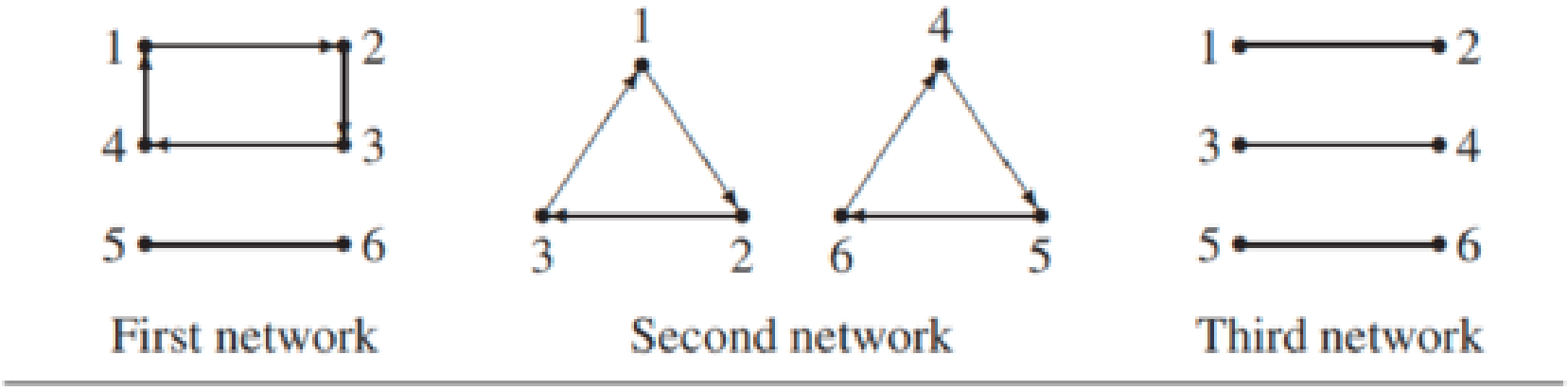


Figure 6: Types

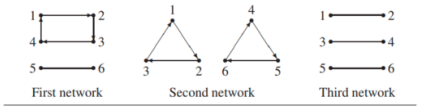


Figure 6: Types

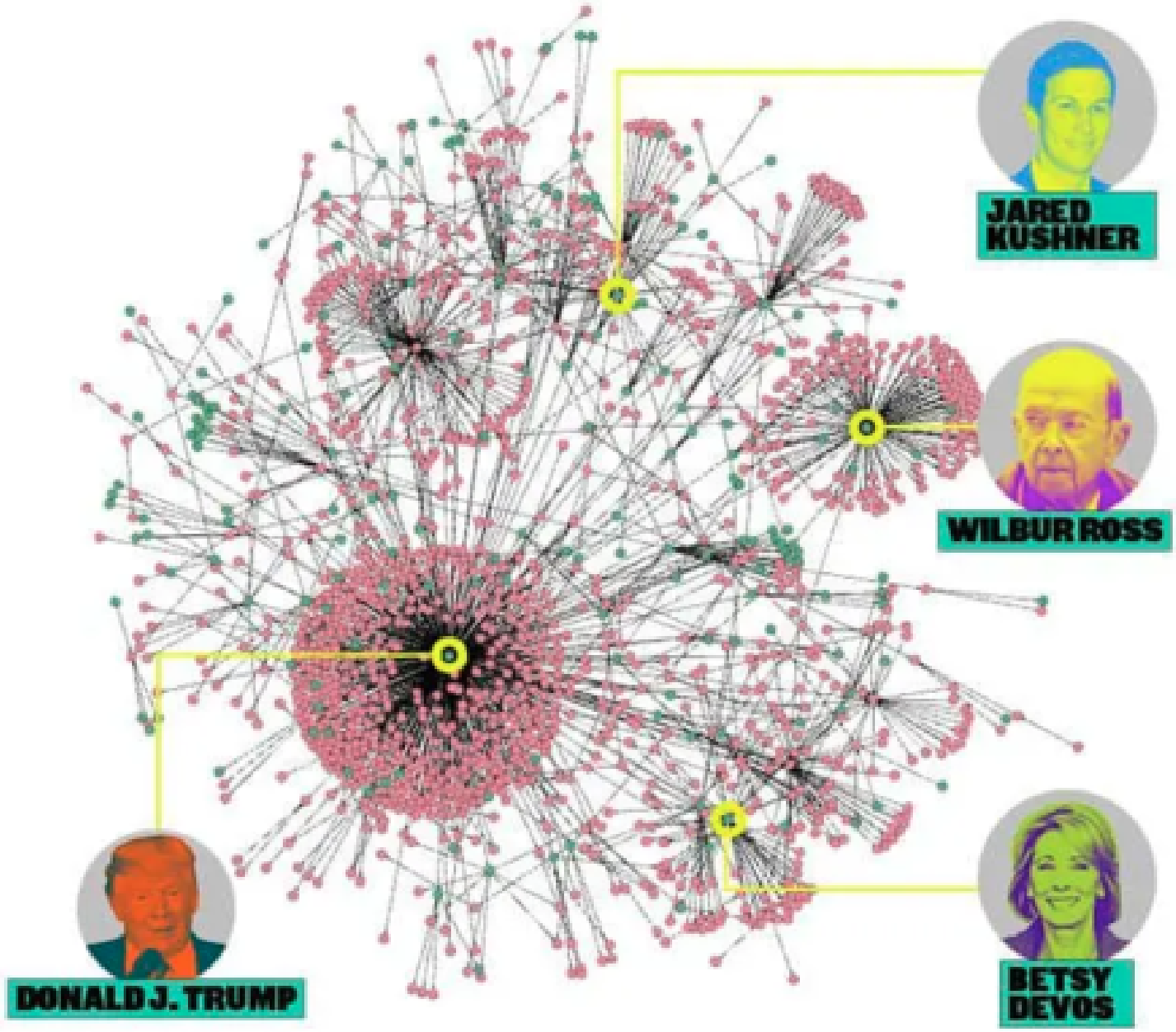
First is sequential network with direction

Show me an example!

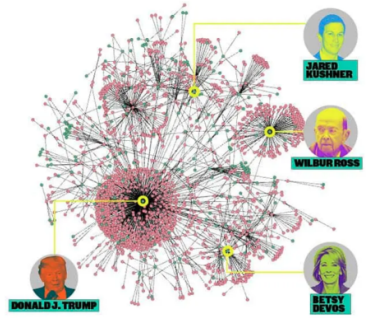
John Templon, Data Reporter for Buzzfeed, decided to use Social Network Analysis to create an interactive map to showcase Donald Trump's affiliations with various organizations.

https://digitaluncovered.com/use-social-network-analysis/

A ‘Very’ Real Network



A 'Very' Real Network



The team at Buzzfeed spent weeks curating this data from public documents and mapping over 2,400 people and organizations connected to Trump. In the article published Buzzfeed team asked viewers to suggest people and organizations connected to Trump which they would have probably missed out in the chart. Network analysis is increasingly used by brands and publishers to identify relations and the value they hold. What we see in this network: Every thing are connected and everyone can access to each other

Bird-eye-view

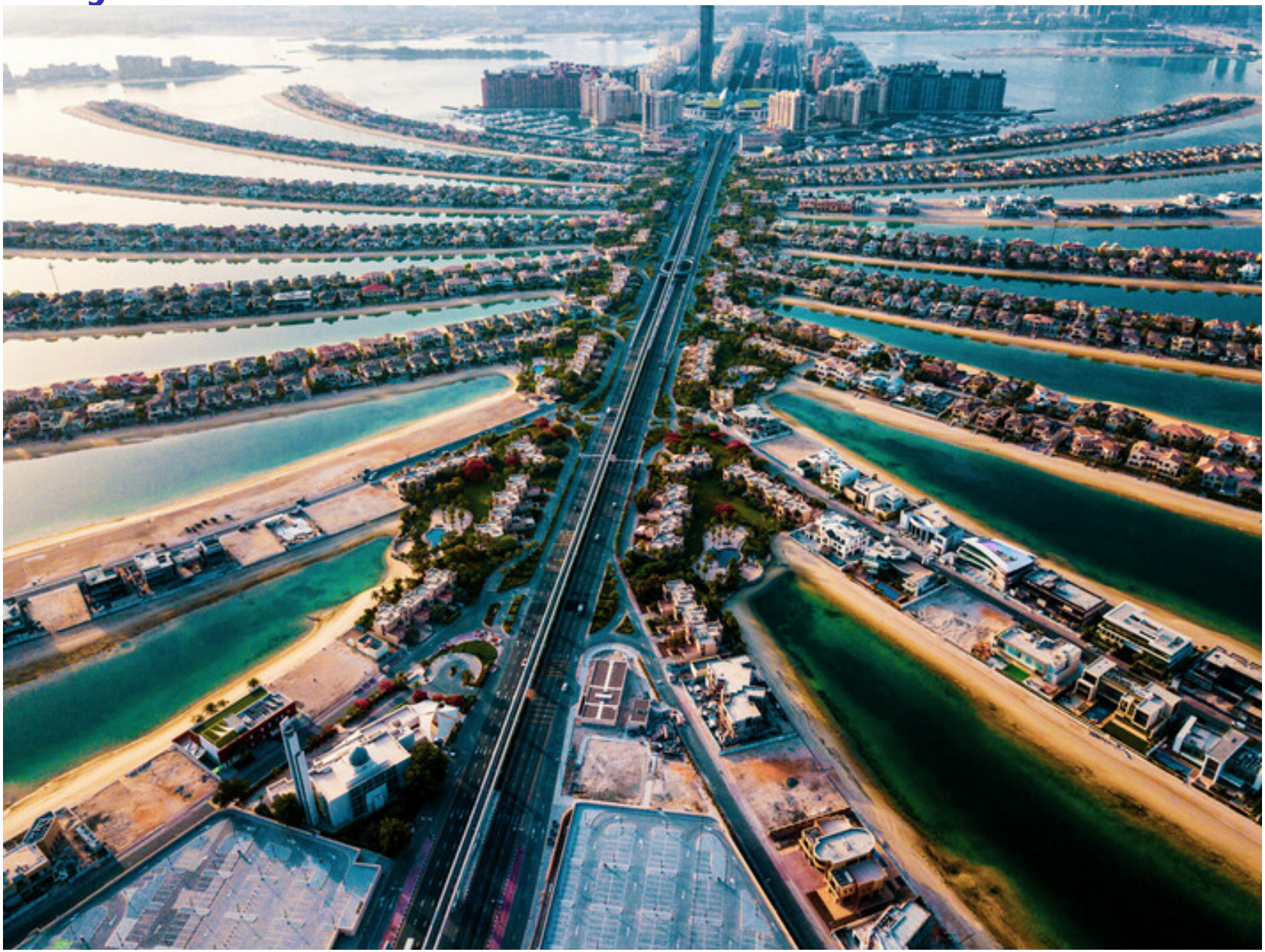


Figure 8: Dubai

Bird-eye-view

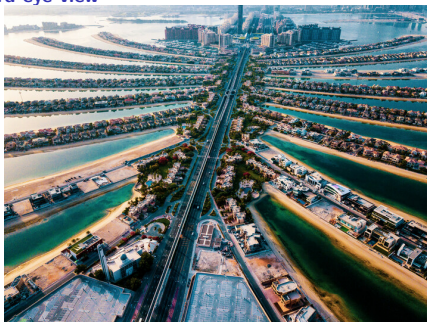


Figure 8: Dubai

RELAX, ASK FOR QUESTIONS WAIT 1 MIN

Non-Human Networks

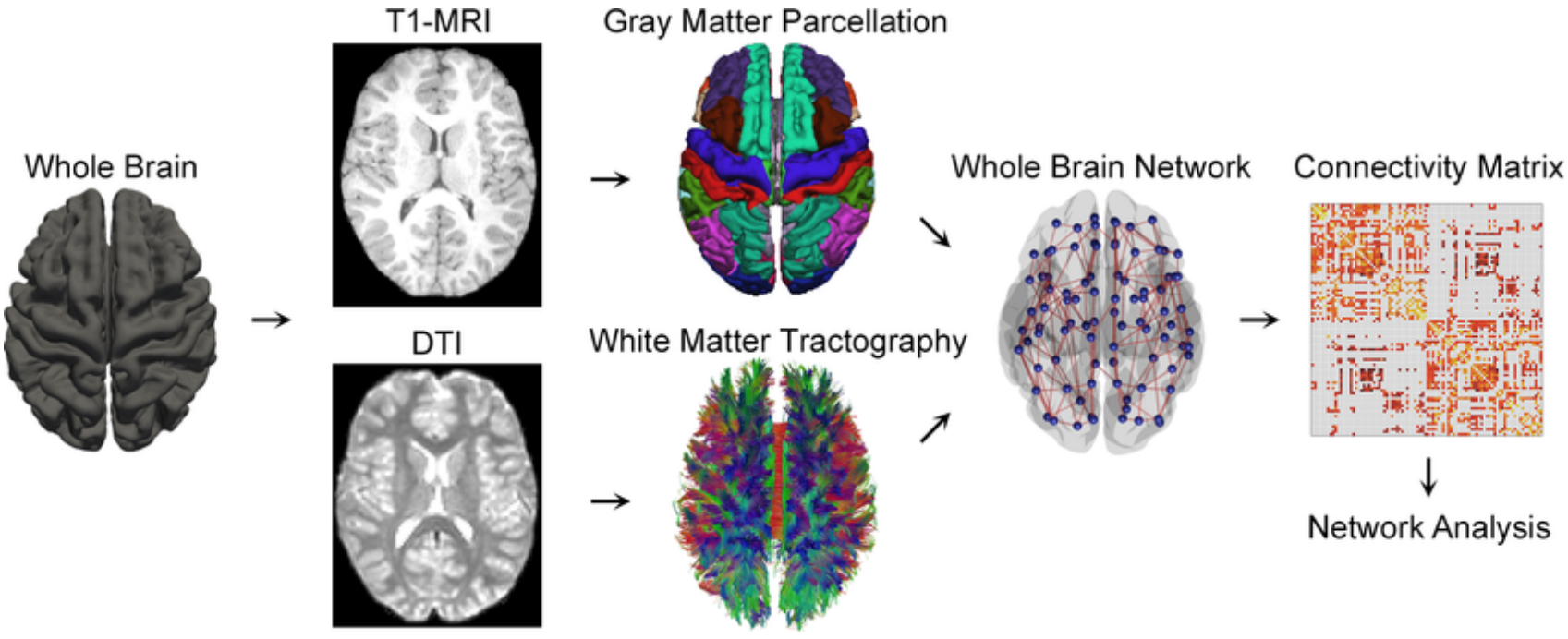


Figure 9: Brain Network

Besides, non-human network systems exist nearly anyplace you see. Our genes and proteins are connected through complex networks. Our brain is presently seen as a complex network of neurons.

Non-Human Networks

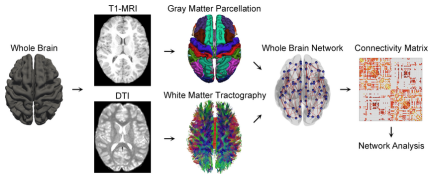


Figure 9: Brain Network

Besides, non-human network systems exist nearly anyplace you see. Our genes and proteins are connected through complex networks. Our brain is presently seen as a complex network of neurons.

For instance brain has its own connections between two hemisphere. This connection becomes network. I read a book named Dyslectic advantage. This book refers how the brain network of dyslexic persons are different and how they are powerful

Section 2: Theory

Get the ball rolling! Definition

> *A network consists of a set of **nodes** (also called vertices) and a set of **links** (also called edges) connecting some of them*

from Great Principles of Computing by Peter J. Denning, Craig H. Martell, Vint Cerf MIT Press, 2015 - Science

> *A network consists of a set of **nodes** (also called vertices) and a set of **links** (also called edges) connecting some of them*
>
> from Great Principles of Computing by Peter J. Denning, Craig H. Martell, Vint Cerf MIT Press, 2015 - Science

So far we tried to imagine the networks but let's go deeper and read the scholar definitions of network. They will be just repetition of what you learn so far. Do you remember the what we talked about? It is just refers them.

Another definition of network

> *A network is the connecting system (wireline or wireless) that allows shared resources among different computers, and usually among a wide range of users.*

from Contemporary Business by Louis E. Boone, David L. Kurtz, Susan Berston Wiley, 2019 Current events

> *A network is the connecting system (wireline or wireless) that allows shared resources among different computers, and usually among a wide range of users.*
>
> from Contemporary Business by Louis E. Boone, David L. Kurtz, Susan Berston Wiley, 2019 Current events

You may notice here, this definition is about computer science however the logic is same. Not only computers but also humans or other network assets interacts and transfers information. So, what happens when some nodes are more than others.

Why I should learn the networks

Assumption: You are like the statistic and data analytics. Imagine this scenario: Some companies have money transfer data like this. Let's name this scenario as **Antares Alliance**

From	To	Amount
A	B	100
A	C	120
A	D	200
B	E	20

Why I should learn the networks

Assumption: You are like the statistic and data analytics. Imagine this scenario: Some companies have money transfer data like this. Let's name this scenario as **Antares Alliance**

From	To	Amount
A	B	100
A	C	120
A	D	200
B	E	20

Think as if the company A sends many to B the amount is 100 dollar and also A sends to C 120 dollar two hundrden twenty dollar. How can we calculate most frequent sender. It may be easy but let's do things crazier. Which company is sends and gets more? Yes it is not picnic however not rocket science too

Why I should learn the networks

From	To	Amount
B	F	10
F	C	100
C	B	100
D	E	100

Now, try to answer these questions:

- Which company is *strong*
- Is there any similarity between company A and company B
- Which company is has not any money transfer (as sender or receiver)

Why I should learn the networks

From	To	Amount
B	F	10
F	C	100
C	B	100
D	E	100

Now, try to answer these questions:

- Which company is *strong*
- Is there any similarity between company A and company B
- Which company is has not any money transfer (as sender or receiver)

Network science starts around the these roundabouts.

Last and biggest questions:

- What if we cannot visualize this?
- What if we visualize but we just see the messy points and lines?

Last and biggest questions:

- What if we cannot visualize this?
- What if we visualize but we just see the messy points and lines?

Well, OK, Let's say you have calculated the answer of previous questions manually or you visualized the network and just understand the things checking visual inspection. What happens int thohse scenarios. IMPROMPTU

Network analysis can help you

- To understand *big-picture*
- To analyze *smallest picture* with its relations to the whole.
- Ang Sang Wahe Guru? (to be explained).
- https://www.youtube.com/watch?v=lPp2wtVquBU

Network analysis can help you

- To understand *big-picture*
- To analyze *smallest picture* with its relations to the whole.
- Ang Sang Wahe Guru? (to be explained).
- https://www.youtube.com/watch?v=IPp2wtVquBU

It is Hinduist mantra refers to connection of whole and part. So, Ang is 'a part.' Sang is 'in every,' or 'with every.' Wahe is 'the indescribable living ecstasy of Infinite Being.' Guru is 'the knowledge that transforms your mind, emotion and essence.' The whole phrase means, "The Infinite Being, God, is with me, and vibrates in every molecule and cell of my being.What to do there? The network analysis gives us to undersanda 'ang' and 'sang' or whole and part and their connections "

Network Analysis with Breaking Bad:

Gustavo's e-mail traffic:

Of course, we have not this kind of data however let's imagine:

From	To	Hour
Gustavo	El-Capo	5.00 am
Gustavo	Mr. White	8.00 am
Pinkman	Gustavo	9.00 am
Pinkman	El-Capo	7.00 am

Network Analysis with Breaking Bad:

Gustavo's e-mail traffic:

Of course, we have not this kind of data however let's imagine:

From	To	Hour
Gustavo	El-Capo	5.00 am
Gustavo	Mr. White	8.00 am
Pinkman	Gustavo	9.00 am
Pinkman	El-Capo	7.00 am

Have you watched the movie breaking bad? It tells about the drug production and trade that a professor started after he had lung cancer and which he grew a lot. The police just can't find them. Of course nobody sends e-mails to anyone in the movie, but let's say they do. Their email traffic would no doubt create an interesting network.In that case, we have from and to column again however we have another attribute it is an hour. So, what will happen when we analyze this? Simply, we can show which people interact with each other and even we can see it when. IMPROPTU

Why Should You Use Social Network Analysis?

- Visualization and Effectiveness and mapping
- Pattern detection
- Hypothesis test

Effectiveness and mapping Social Network Analysis is useful if you wish to understand the effectiveness of network both offline or online. Network Analysis also helps you to map the flow of information in a network.

Uncover trends and Pattern SNA is valuable if you are looking to discover new trends or patterns. Valdis E. Krebs, a network scientist in his widely praised paper, shared how network analysis can be used to identify terrorist networks.

Find value Test Hypothesis SNA can be used to test a hypothesis in online behavior and to determine causes for dysfunctional communities and networks, and to promote social cohesion and growth in the online community.

Section 3: Tool

Which tool we'll use? : R

- Integtrated
- Flexible
- Open Source

- Integtrated
- Flexible
- Open Source

Let's talk about which tool that we'll use. There is a not a lot of but a few tools for network analysis. Gephi, UCINET are some well known applications that are not requires to write code. However we like the code. The R statistical programming language and environment comprise a vast integrated system of thousands of packages and functions that allow it to handle innumerabledata management, analysis, or visualization tasks. As you know The R system includes a number of packages that are designed to accomplish specific network analytic tasks.However, by performing these network tasks within the R environment, the analyst can take advantage of any of the other capabilities of R. Most other network analysis programs (e.g., Pajek, UCINet, Gephi) are stand-alone packages, and thus donot have the advantages of working within an integrated statistical programming environment.

R is Open source

> *Open source software is software with source code that anyone can inspect, modify, and enhance.*
> *"Source code" is the part of software that most computer users don't ever see; it's the code computer programmers can manipulate to change how a piece of software—a "program" or "application"—works. Programmers who have access to a computer program's source code can improve that program by adding features to it or fixing parts that don't always work correctly.*

https://opensource.com/resources/what-open-source

R is Open source

> *Open source software is software with source code that anyone can inspect, modify, and enhance.*
> *"Source code" is the part of software that most computer users don't ever see; it's the code computer programmers can manipulate to change how a piece of software—a "program" or "application"—works. Programmers who have access to a computer program's source code can improve that program by adding features to it or fixing parts that don't always work correctly.*
>
> https://opensource.com/resources/what-open-source

I want to re-define or recall the Open source to emphasize the power of it. One of the important reasons for R's popularity and success is its free and open nature. This is formally ensured via the GNU General Public License (GPL) that R-code is released under. More informally, there is a vast R user and developer community which is continually working to enhance and improve R base code and the thousands of R packages that can be freely accessed.

R is compact

- Comparison of R with commerical tools like SPSS/SAS
- Network analysis capacity
- Reporting possibilities (shiny).
- e.g: https://suatatan.shinyapps.io/JournalAnalytics/

R is compact

- Comparison of R with commerical tools like SPSS/SAS
- Network analysis capacity
- Reporting possibilities (shiny).
- e.g: https://suatatan.shinyapps.io/JournalAnalytics/

Although there are many good network analysis programs available which canhandle a wide variety of network descriptive statistics and visualization tasks, no other network package has the same power to handle often complex data andproject management tasks for larger-scale network analyses compared to R. First,as suggested above, network analysis in R can take advantage of the powerful datamanagement, cleaning, import and export capabilities of base R.

Network analysis often starts by importing and transforming data from other sources into a form that can be analyzed by network tools. All network pack-ages have some data management capabilities, but no other program can match R'sbreadth and depth.Second, when conducting sophisticated scientific or commercial network analyses, it is important to have the right project management tools to facilitate codestorage and retrieval, managing analysis outputs such as statistical results and infor-mation graphics, and producing reports for internal and external audiences. Tradi-tional statistical

R package ecosystem

- Traditional statistic is just simple attribute of R
- R provides to import and export from various alternantives like Gephi or Ucinet.
- R package ecosystem are vibrant and well documented.

- Traditional statistic is just simple attribute of R
- R provides to import and export from various alternantives like Gephi or Ucinet.
- R package ecosystem are vibrant and well documented.

The primary reason R is ideal for network analysis is the breadth of packages that are currently available to manage network data and conduct network visualization, network description, and network modeling.

Recall R Fundamentals

```
myvar = 120
list_a = c(10,20,30)
list_b = c(11,21,33)
mydataframe = data.frame(list_a,list_b)
```

Recall R Fundamentals

```
myvar = 120
list_a = c(10,20,30)
list_b = c(11,21,33)
mydataframe = data.frame(list_a,list_b)
```

If you don't program with R every day, you may have forgotten some basic concepts and coding. Don't worry, we'll cover them very short now. EXPLAIN

Recall R Fundamentals

```
mydataframe
```

```
##   list_a list_b
## 1     10     11
## 2     20     21
## 3     30     33
```

```
mydataframe
```

```
##   list_a list_b
## 1     10     11
## 2     20     21
## 3     30     33
```

We can see our data frame with calling it directly

Filtering Data Frames by Dplyr

```
library(dplyr)
mydataframe %>% filter(list_a == 10)
```

```
##   list_a list_b
## 1     10     11
```

```
library(dplyr)
mydataframe %>% filter(list_a == 10)
```

```
##   list_a list_b
## 1     10     11
```

dplyr is primarily a set of functions designed to enable dataframe manipulation in an intuitive, user-friendly way.

Recall Dplyr Methods

select(), which is used to subset a dataframe by its columns;

arrange(), which is used to sort rows in a dataframe based on attributes held by particular columns;

select(), which is used to subset a dataframe by its columns;

arrange(), which is used to sort rows in a dataframe based on attributes held by particular columns;

EXPLAIN, select is select asterix from of SQL, arrange is for sorting

Recall Dplyr Methods

mutate(), which is used to create new variables, by altering and/or combining values from existing columns; and

summarize(), also spelled summarise(), which is used to collapse values from a dataframe into a single summary.

mutate(), which is used to create new variables, by altering and/or combining values from existing columns; and

summarize(), also spelled summarise(), which is used to collapse values from a dataframe into a single summary.

EXPLAIN mutate clonens, summarize is for counting or summing up values

Reading Data From CSV

```
df = read_csv("table.csv")
```

```
df = read_csv("table.csv")
```

Where do network diagrams come from, of course data. Remember the email traffic between the Breaking Bad movie heroes or the Antares Alliance. The data were in tabular form, right? Here's how we do the data reading in R.

Section: Quick-Start

Installation

```
#install.packages("statnet")
#install.packages("devtools")
#devtools::install_github("DougLuke/UserNetR")
```

Calling packages

```
library(statnet)
library(UserNetR)
```

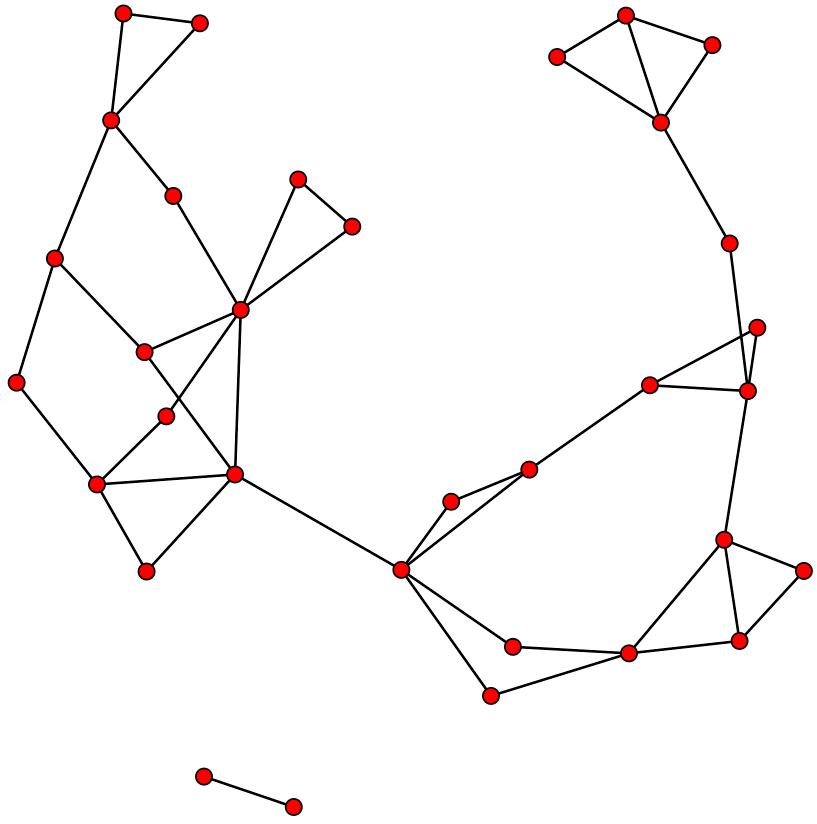
Inspecting Data

```
data(Moreno)
Moreno
```

```
##  Network attributes:
##   vertices = 33
##   directed = FALSE
##   hyper = FALSE
##   loops = FALSE
##   multiple = FALSE
##   bipartite = FALSE
##   total edges= 46
##     missing edges= 0
##     non-missing edges= 46
##
##  Vertex attribute names:
##     gender vertex.names
##
## No edge attributes
```

Introduction to Visualization

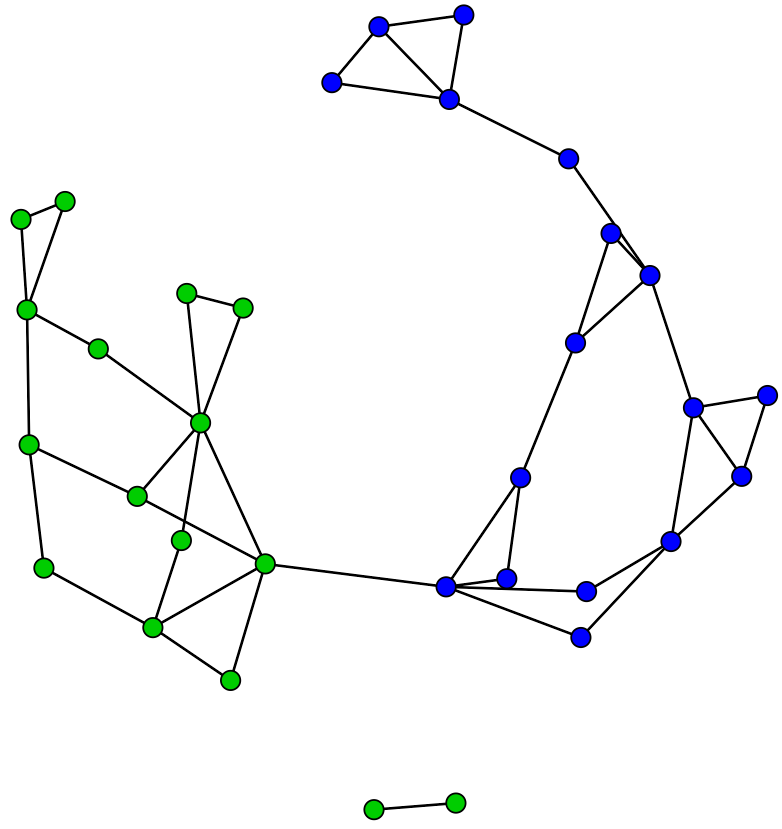
```
plot(Moreno)
```



Visualization by Gender

```
##  [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1
```

Visualization by Gender



Explaining Visualization

The resulting plot makes it clearly obvious how two subtly different subgroups, based on gender, make up the friendship network. The most significant structural characteristics found in the social network can often be revealed by a rapidly generated network graphic like this.

Size

```
network.size(Moreno)
```

```
## [1] 33
```

Size

```
network.size(Moreno)
```

```
## [1] 33
```

The most basic characteristic of a network is its size. The size is simply the number of members, usually called nodes, vertices or actors. The network.size() function is the easiest way to get this. The basic summary of a statnet network object also provides this information, among other things. The Moreno network has 33 members, based on the network.size and summary calls. (Setting the print.adj to false suppresses some detailed adjacency information that can take up a lot of room.)

Summarizing the descriptive characteristics of the network

```
summary(Moreno,print.adj=FALSE)
```

```
## Network attributes:
##   vertices = 33
##   directed = FALSE
##   hyper = FALSE
##   loops = FALSE
##   multiple = FALSE
##   bipartite = FALSE
##  total edges = 46
##    missing edges = 0
##    non-missing edges = 46
##  density = 0.08712121
##
## Vertex attributes:
##
##  gender:
##    numeric-valued attribute
```

Density (For Directed Networks)

- L number of observed connection in the network
- k maximum number of possible actors
- D is between 0 and 1:
- 0: no connection
- 1: full connection

$$D_d = \frac{L}{k \times (k-1)}$$

Density (For Directed Networks)

- L number of observed connection in the network
- k maximum number of possible actors
- D is between 0 and 1:
- 0: no connection
- 1: full connection

$$D_d = \frac{L}{k \times (k-1)}$$

Of all the basic characteristics of a social network, density is among the most important as well as being one of the easiest to understand. Density is the proportion of observed ties (also called edges, arcs, or relations) in a network to the maximum number of possible ties. Thus, density is a ratio that can range from 0 to 1. The closer to 1 the density is, the more interconnected is the network.

Density (For Undirected Network)

$$D_u = \frac{2L}{k \times (k-1)}$$

Density Function

```
gden(Moreno)
```

```
## [1] 0.08712121
```

Density (Zoom ++)

This network shows the interactions among the Jemaah Islamiyah terrorist group that carried out the bombings in Bali in 2002.

http://dougluke.github.io/UserNetR/reference/Bali.html

```
gden(UserNetR::Bali)
```

```
## [1] 0.4632353
```

This network object contains collaboration ties among 493 members of Washington University's Institute of Clinical and Translational Sciences (ICTS).

http://dougluke.github.io/UserNetR/reference/ICTS_G10.html

```
gden(UserNetR::ICTS_G10)
```

```
## [1] 0.01120566
```

Reviewing Network Fundamentals

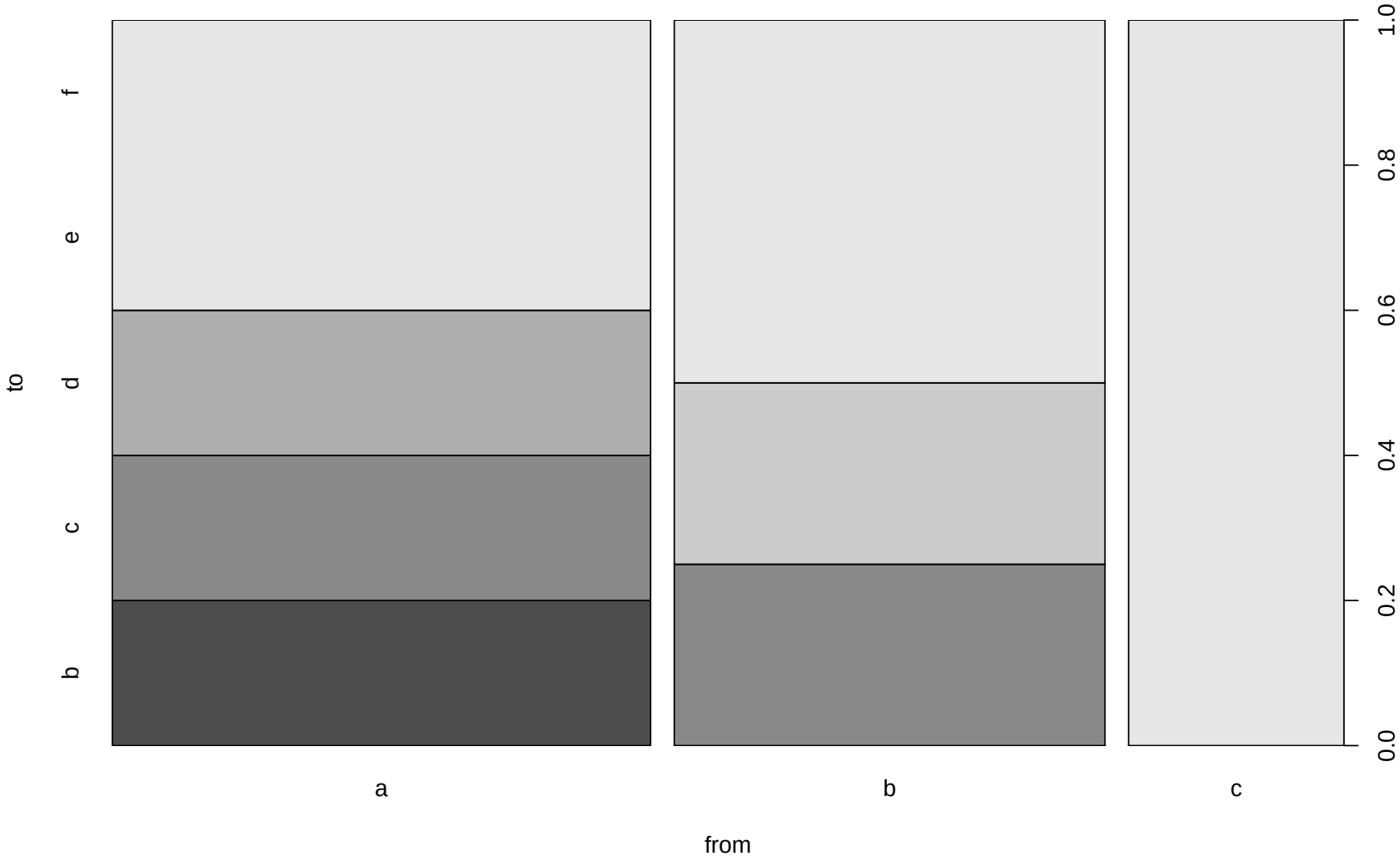
Let's read data from our local file system:

```
local <- read.csv("local.csv")
local
```

```
##    from to
## 1     a  b
## 2     a  c
## 3     a  d
## 4     b  c
## 5     b  e
## 6     b  f
## 7     a  f
## 8     c  f
## 9     a  f
## 10    b  f
## 11    c  f
```

Why?

```
library(statnet)
plot(local)
```



Nature of objects

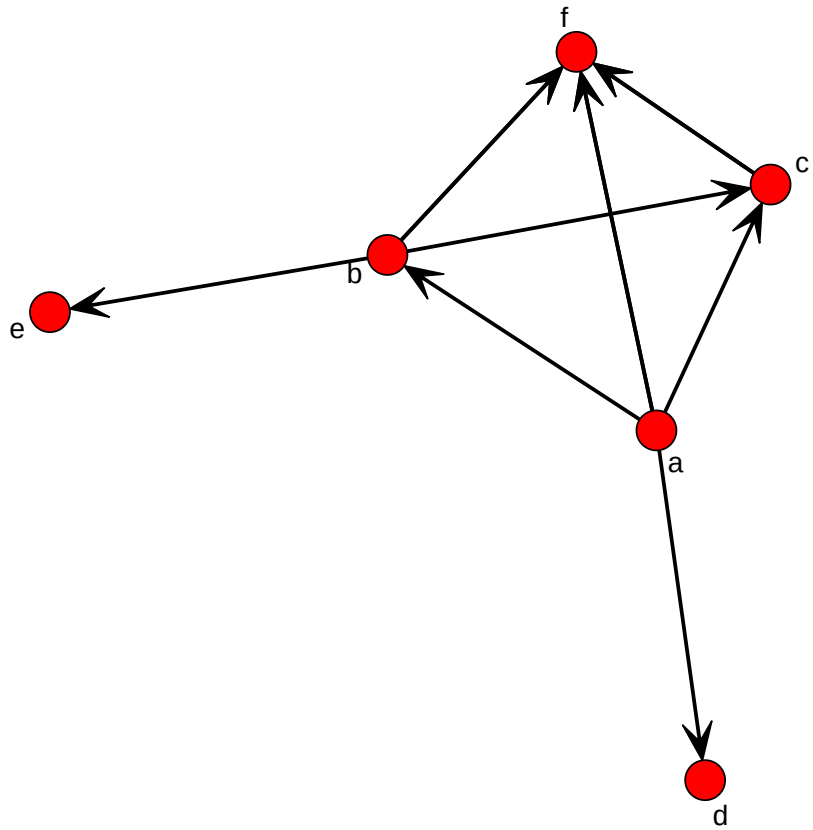
- Whereas we are working on data science we still need to be aware some concepts of OOP.
- The variable named by `local` is standard data frame object we need to convert it into `statnent` graphic data.
-

Type Conversion

```
network_obj = network(local)
```

Plotting from Data Frame

```
gplot(network_obj,displaylabels=TRUE)
```



Network Size

Network Density

```
gden(network_obj)
```

```
## [1] 0.3666667
```

What we learned?

- What is the network analysis
- Why it is important
- Recalling R data frames
- Installing `statmet`

What we learned?

- Size
- Density
- Interpreting density

End

Dr. Suat ATAN