

Introduction

–Seminar–

Yasemin Aslan

SPRU (Science Policy Research Unit)
Business School
University of Sussex



Week 1: 28 January 2022

Learning outcome	Assessment mode
1 Explain the concept of network and list the main network indicators	ESS
2 Describe and apply the major techniques for the collection of network data and their statistical analysis	ESS, GPN + GWS
3 Identify the main characteristics of networks by means of network measures	ESS, GPN + GWS
4 Employ network analysis techniques to produce network data-based infographics	GPN + GWS

Note: ESS: Essay; GPN: Group Presentation; GWS: Group Written Submission

- 1 The 'old school' exercise
- 2 R, RStudio, and igraph
- 3 Ready ...?
- 4 Some useful (and freely available) resources to learn R

The 'old school' exercise

The 'old school' exercise

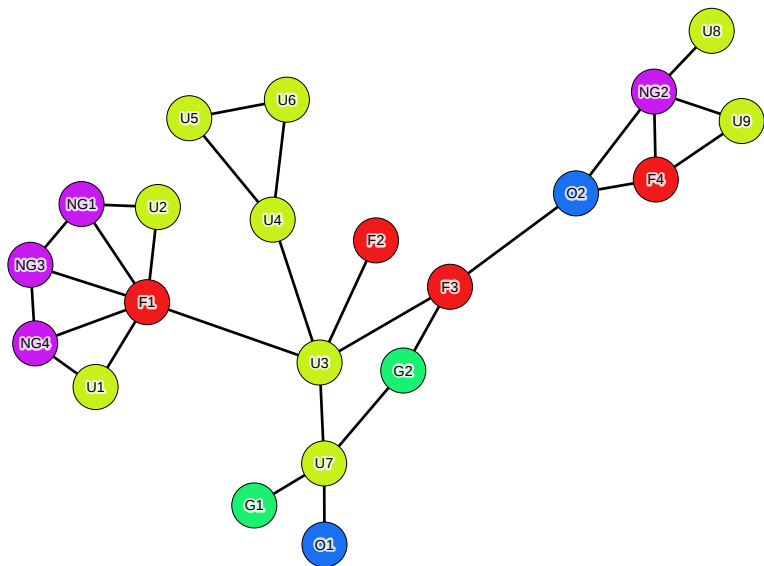
You are provided with a list 20 R&D projects. These projects involve different types of organisations. You are requested to:

- 1 Draw a network the depicts the collaboration activity of firms on projects (inter-organisational network)
- 2 Which are the most influential organisations in this network?
- 3 Which are the most critical organisations for the cohesion of the network?
- 4 Upload a picture of your network
<https://padlet.com/yaslan2/oysry42vouhcetvt>

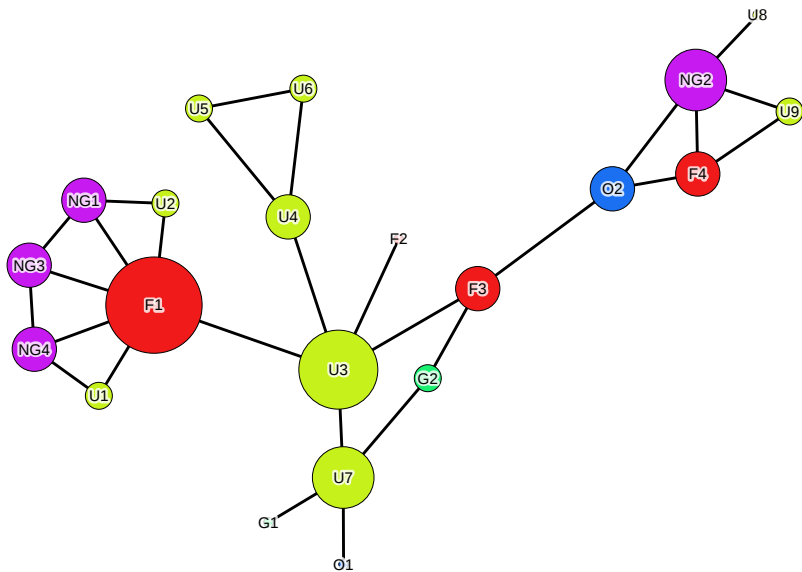
R&D project	List of partners
Proj01	U2, F1, NG1
Proj02	U1, NG4, F1
Proj03	NG3, NG1, F1
Proj04	NG3, NG4, F1
Proj05	U3, F1
Proj06	U3, F2
Proj07	U3, F3
Proj08	U3, U4
Proj09	F1
Proj10	U5
Proj11	U4, U5, U6
Proj12	U3, U7
Proj13	U7, G1
Proj14	U7, O1
Proj15	U7, G2
Proj16	G2, F3
Proj17	F3, O2
Proj18	O2, F4, NG2
Proj19	F4, U9, NG2
Proj20	NG2, U8

Firm (F); University (U);
Gov. (G); Non-Gov. (NG);
Other (O)

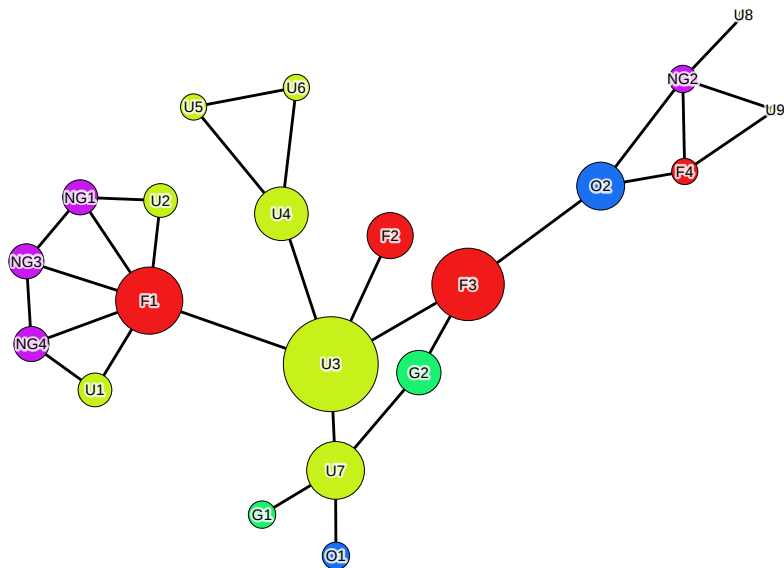
The 'old school' exercise



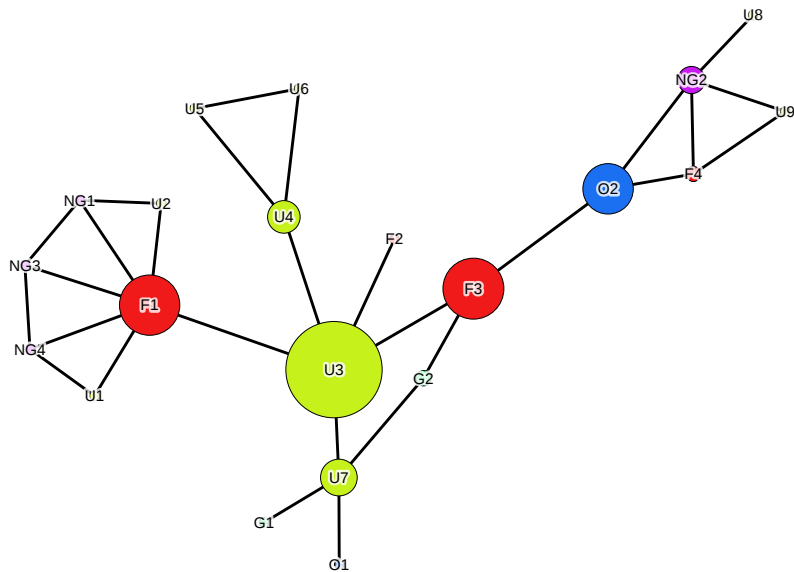
The 'old school' exercise



The 'old school' exercise

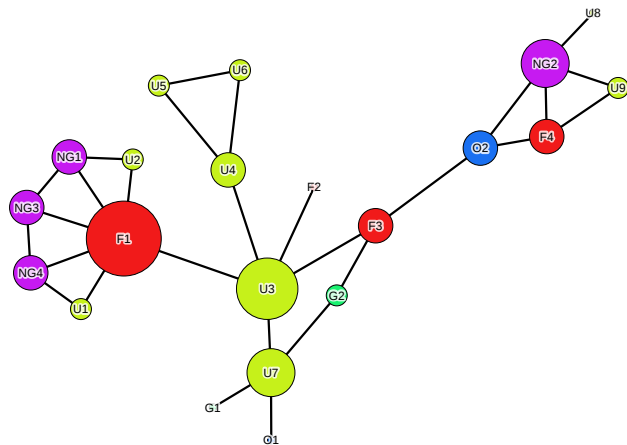


The 'old school' exercise



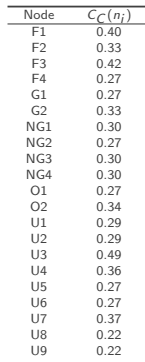
The 'old school' exercise

Degree centrality



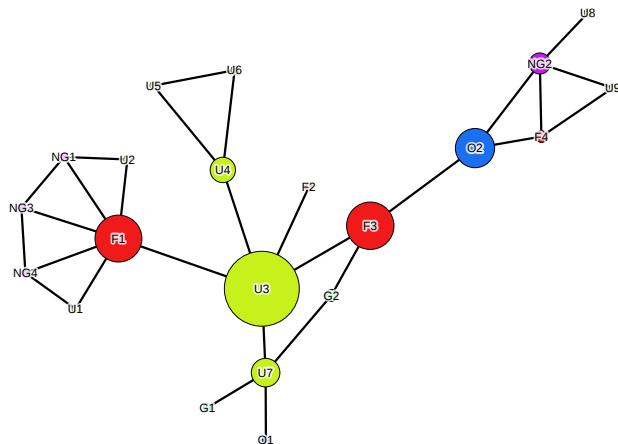
Node	$C_D(n_i)$
F1	6
F2	1
F3	3
F4	3
G1	1
G2	2
NG1	3
NG2	4
NG3	3
NG4	3
O1	1
O2	3
U1	2
U2	2
U3	5
U4	3
U5	2
U6	2
U7	4
U8	1
U9	2

Closeness centrality



The 'old school' exercise

Betweenness centrality



Node	$C_B(n_i)$
F1	0.42
F2	-
F3	0.42
F4	0.04
G1	-
G2	0.05
NG1	0.00
NG2	0.14
NG3	0.00
NG4	0.00
O1	-
O2	0.34
U1	-
U2	-
U3	0.72
U4	0.19
U5	-
U6	-
U7	0.22
U8	-
U9	-

R, RStudio, and igraph



www.r-project.org

- A **language** and **environment** for statistical computing and graphics



www.r-project.org

- A **language** and **environment** for statistical computing and graphics
- Runs on Windows, Linux, macOS



www.r-project.org

- A **language** and **environment** for statistical computing and graphics
- Runs on Windows, Linux, macOS
- Initially developed for **statistical analysis** by Robert Gentleman and Ross Ihaka (University of Auckland) – that's why "R".



www.r-project.org

- A **language** and **environment** for statistical computing and graphics
- Runs on Windows, Linux, macOS
- Initially developed for **statistical analysis** by Robert Gentleman and Ross Ihaka (University of Auckland) – that's why "R".
- **Core Group** with write access to the R source



www.r-project.org

- A [language](#) and [environment](#) for statistical computing and graphics
- Runs on Windows, Linux, macOS
- Initially developed for [statistical analysis](#) by Robert Gentleman and Ross Ihaka (University of Auckland) – that's why "R".
- [Core Group](#) with write access to the R source
- Constantly updated and with access to [19,000 packages](#) (January 2022, www.r-pkg.org)

R, RStudio, and igraph



R Console

Help Search

R version 3.3.0 (2016-05-03) -- "Supposedly Educational"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.68 (7202) x86_64-apple-darwin13.4.0]

[Workspace restored from /Users/danielerotolo/.RData]
[History restored from /Users/danielerotolo/.Rapp.history]

>

RMeSHqualifier.R

splitVec

Help search

1 #####
2 ##### ----- MeSH Qualifier Analysis ----- #####
3 ##### ----- Dr. Daniele Rotolo (www.danielerotolo.com) ----- #####
4 ##### ----- SPRU Science Policy Research Unit, University of Sussex ----- #####
5 #####
6
7 library(ggplot2)
8 library(plyr)
9
10 #####
11 # Defining parameters
12
13 # -- working directory -- #
14 setwd("/Users/danielerotolo/dropbox/r_routines/RMeSH_qualifier_analysis/")
15
16 # -- MeSH term used for the search -- #
17 mesh_term_searched<-"Diabetes Mellitus"
18
19 # -- Name of the file including MEDLINE/PubMed data -- #
20 file<-"Diabetes Mellitus.txt"
21
22 # -- Observation period
23 start<-1990
24 end<-2013
25
26 # -- Number of most occurring qualifier to visualise -- #
27 N<-10
28
29 #####
30
31
32 #####
33 # -- Data conversion starts
34
35 # 1) Read the data file and creating a new one file for the converted data
36 medline<-readLines(file)
37 new_file<-paste(mesh_term_searched, "_converted.txt", sep='')
38 write(paste("PMID", "PY", "MeSH", "Descriptor", "Descriptor_MT", "Qualifier", "Qualifier
_MT", sep='\t'),
39 new_file, sep='\t')



www.rstudio.com

- An [open-source editor](#) for R



www.rstudio.com

- An [open-source editor](#) for R
- Developed by a company called RStudio



www.rstudio.com

- An [open-source editor](#) for R
- Developed by a company called RStudio
- [Functionalities](#) that are otherwise not available in the standard R editor



www.rstudio.com

- An [open-source editor](#) for R
- Developed by a company called RStudio
- [Functionalities](#) that are otherwise not available in the standard R editor
- Support, tutorials, webinars

R, RStudio, and igraph



RStudio

```
igraph_session.R x
Source on Save Run Import Dataset Global Environment
99 h2 <- induced_subgraph(g, c(1,2,3,5))
100 plot(h2)
101
102
103
104 #---Change nodes's color and size / edges' colors and width
105 V(g)$color <- "lightblue"
106 V(g)$size <- 25
107 E(g)$color <- "red"
108 E(g)$width <- 2
109 plot(g, vertex.label = V(g)$id)
110
111 E(g)[5 %--% 7]$color <- "green" #Color specific nodes and edges
112 V(g)[5]$color <- "white"
113 plot(g)
114
115
116
117 #---Create new attributes for nodes and edges
118 V(g)$att <- c(1,0,1,1,0,1,0)
119 V(g)$color <- ifelse(V(g)$att == 1, "green", "yellow")
120 plot(g)
121
122 E(g)$value <- c(1.0,2.0,3.5,4.0,5.5,6.8,2.0,1.0)
123 plot(g, edge.label = E(g)$value)
124 plot(g, edge.width = E(g)$value)
117:1 R Script
```

Environment History
Global Environment

Data
data 5 obs. of 2 variables

Values

a	num [1:5] 1 2 3 4 5
b	num [1:3] 2 4 5
c	Factor w/ 5 levels "Aldo","Antonio",...: 1 4 2 3 4 5 2 2
d	List of 3
directory	"/Users/danielerotolo/Dropbox/SNA_IPTS_2015/Lab/igraph"
g	List of 10
h1	List of 10
h2	List of 10
x	5
v	4

Files Plots Packages Help Viewer
Zoom Export Publish

Console
~/Dropbox/SNA_IPTS_2015/Computer_session/igraph/ R
> #---Add nodes and vertices and select subgraphs
> h1 <- g + vertices(c(8,9))
> h1 <- h1 + edges(c(8,9), c(1,8))
> plot(h1)
> h2 <- induced_subgraph(g, c(1,2,3,5))
> plot(h2)
> #---Change nodes's color and size / edges' colors and width
> V(g)\$color <- "lightblue"
> V(g)\$size <- 25
> E(g)\$color <- "red"
> E(g)\$width <- 2
> plot(g, vertex.label = V(g)\$id)
> E(g)[5 %--% 7]\$color <- "green" #Color specific nodes and edges
> V(g)[5]\$color <- "white"
> V(g)[5]\$color <- "white"
> plot(g)
>

Elements of the interface:

- **Script:** the list of commands we want R to execute
- **Console:** the list of commands R executes and the outcomes of these
- **Environment, History:** variables, datasets, and executed commands
- **Files, Plot, Packages, Help, Viewer:** generated charts, loaded packages, etc.



<http://igraph.org/r/>

- Provides a relatively comprehensive [set of tools to perform network analysis](#)



<http://igraph.org/r/>

- Provides a relatively comprehensive [set of tools to perform network analysis](#)
- Developed by Gábor Csárdi (Harvard University) and others



<http://igraph.org/r/>

- Provides a relatively comprehensive [set of tools to perform network analysis](#)
- Developed by Gábor Csárdi (Harvard University) and others
- Excellent textbook [Kolaczyk and Csárdi, 2014]

Ready ...?

Ready ...?

R and RStudio

On campus*

- 1 Go to <http://rstudio.uscs.susx.ac.uk/>
- 2 Access the website by using your University of Sussex account

On your personal computer

- 1 Install R
- 2 Install RStudio

RStudio cloud

- 1 Go to <https://rstudio.cloud>
- 2 Register and sign in

*You can access the server off campus by using remote desktop
<http://www.sussex.ac.uk/its/services/software/windowsremote>

Ready ...?

Installing igraph

Through the **RStudio interface**:

- 1 Tools
- 2 Install Packages
- 3 Search for igraph
- 4 Tick the box "Install Dependencies"
- 5 Install

or

Through the **RStudio console**

```
1| install.packages("igraph")  
2| library("igraph")
```

Some useful (and freely available) resources to learn R

- **Books**

- ▶ [Harley Wickham and Garrett Golemund - R for Data Science](#): generic introduction to R and to the data analysis workflow
- ▶ [Claus O. Wilke - Fundamentals of Data Visualisation](#): mainly based on ggplot2, an R package for data visualisation

- **Online courses:** a host of possible choices!

- ▶ [RStudio](#) official learning pathways
- ▶ [Linkedin Learning](#)
- ▶ Coursera, EdX...

- **Online communities:** practical solutions to specific issues

- ▶ [StackOverFlow](#) - Q&A about coding problems
- ▶ [GitHub](#) - a bit more advanced, repository of routines

- **Twitter:** #rstats and @rstatstweet

Next time ...

- **Lecture: Network definition**

- ▶ Definition of network and different types of networks
- ▶ Overview of the historical and disciplinary origins of (social) network analysis
- ▶ Network visualisation standards

- **Seminar: Network definition**

- ▶ Short intro to R
- ▶ Basic commands



Kolaczyk, E. and Csárdi, G. (2014).

Statistical analysis of network data with R, volume 65.
Springer-Verlag, New York, NY, USA.