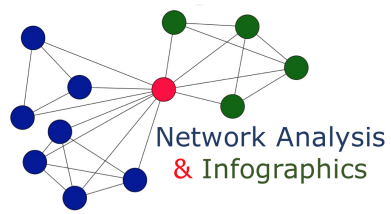
Network Data Collection
–Seminar–

Yasemin Aslan

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



Week 3: 11 February 2022

Learning Outcomes

	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

Overview

1. Introduction to R [recap]
2. Importing/exporting data in R
3. Network file formats
4. Introduction to igraph

Introduction to R [recap]

Introduction to R [recap]

- R creates and manipulate objects
- Different types or classes of objects
  - Data objects
    - ★ vector: an ordered collection of values
    - ★ matrix: a 2-dimensional vector
      (a vector with > 2 dimensions is called array)
    - ★ data frame: variables and observations
    - ★ list: an ordered sequences of objects
    - ★ factor: categorical data (e.g. "male", "female")
  - Function objects
    - ★ Functions can read, manipulate and analyse data
    - ★ Packages provide users with additional functions (e.g. igraph)
- Objects have attributes
- R can read and export a variety of data formats

Importing/exporting data in R

Importing/exporting data in R

- R can deal with many different data formats

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...
- Text editors: Notepad++ (Windows), TextWrangler (Mac)

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...
- Text editors: Notepad++ (Windows), TextWrangler (Mac)
- R packages to read and transform data:

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...
- Text editors: Notepad++ (Windows), TextWrangler (Mac)
- R packages to read and transform data:
  - "readr" (csv, txt, delimited)

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...
- Text editors: Notepad++ (Windows), TextWrangler (Mac)
- R packages to read and transform data:
  - "readr" (csv, txt, delimited)
  - "readxl" (Excel)

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...
- Text editors: Notepad++ (Windows), TextWrangler (Mac)
- R packages to read and transform data:
  - "readr" (csv, txt, delimited)
  - "readxl" (Excel)
  - "haven" (STATA, SPSS, SAS)

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...
- Text editors: Notepad++ (Windows), TextWrangler (Mac)
- R packages to read and transform data:
  - "readr" (csv, txt, delimited)
  - "readxl" (Excel)
  - "haven" (STATA, SPSS, SAS)
  - "tidyverse"

Importing/exporting data in R

- R can deal with many different data formats
- Which data formats do you know?
- Most commonly used data formats
  - txt: fields are usually separated by *comma*, *tab*, *colon*
  - csv: fields are separated by *comma*
  - dta: format used by STATA to store data
  - xls/xlsx: format used by Excel to store data
  - ...
- Text editors: Notepad++ (Windows), TextWrangler (Mac)
- R packages to read and transform data:
  - "readr" (csv, txt, delimited)
  - "readxl" (Excel)
  - "haven" (STATA, SPSS, SAS)
  - "tidyverse"
- Install these packages in R (you should now know how to do it)

Importing/exporting data in R

Let's import/export data in RStudio

Network file formats

Network file formats

Composition variables (attributes)

Case	Affiliation	Variable 1	$\cdots$	Variable K
1				
2				
$\vdots$				
N				

Structural variables (adjacency matrix)

		Case			
		1	2	$\cdots$	N
Case	1				
	2				
	$\vdots$				
	N				

Network file formats

Case-affiliation adjacency matrix

		Affiliation			
		1	2	$\cdots$	K
Case	1				
	2				
	$\vdots$				
	N				

Network file formats

Case-affiliation adjacency matrix

		Affiliation			
		1	2	$\cdots$	K
Case	1				
	2				
	$\vdots$				
	N				

(1)

Case-case adjacency matrix

		Case			
		1	2	$\cdots$	N
Case	1				
	2				
	$\vdots$				
	N				

Network file formats

Case-affiliation adjacency matrix

		Affiliation			
		1	2	$\cdots$	K
Case	1				
	2				
	$\vdots$				
	N				

(1)

Case-case adjacency matrix

		Case			
		1	2	$\cdots$	N
Case	1				
	2				
	$\vdots$				
	N				

(2)

Affiliation-Affiliation adjacency matrix

		Affiliation			
		1	2	$\cdots$	K
Affiliation	1				
	2				
	$\vdots$				
	K				

Network file formats

- The variety of software packages an program languages has led to a variety of network file formats
  - GML
  - GEXF
  - GDF
  - GraphML
  - Pajek NET
  - GraphViz DOT
  - UCINET DL
  - ...

- You do not need to remember all these file formats
  - Network file formats are often created from simple tables of nodes and edges
  - Most of these formats can be imported in R
  - New formats are frequently created

Introduction to igraph

Introduction to igraph

We focus on:

- How to create a network
- How to visualize a network
- How to add attributes to nodes and edges

Introduction to igraph

Let's practice this in RStudio (open S3_R_Script/S3_Script.R)

Questions

Next time ...

Next time ...

- **Lecture: Descriptive network analysis A**
  - Network measures at the level of the whole network
- **Seminar: Descriptive network analysis A**
  - Import network data in igraph
  - Calculate network-level measures in igraph