

COMPLEX SYSTEMS & NETWORK THINKING

Pierre-Alexandre Balland



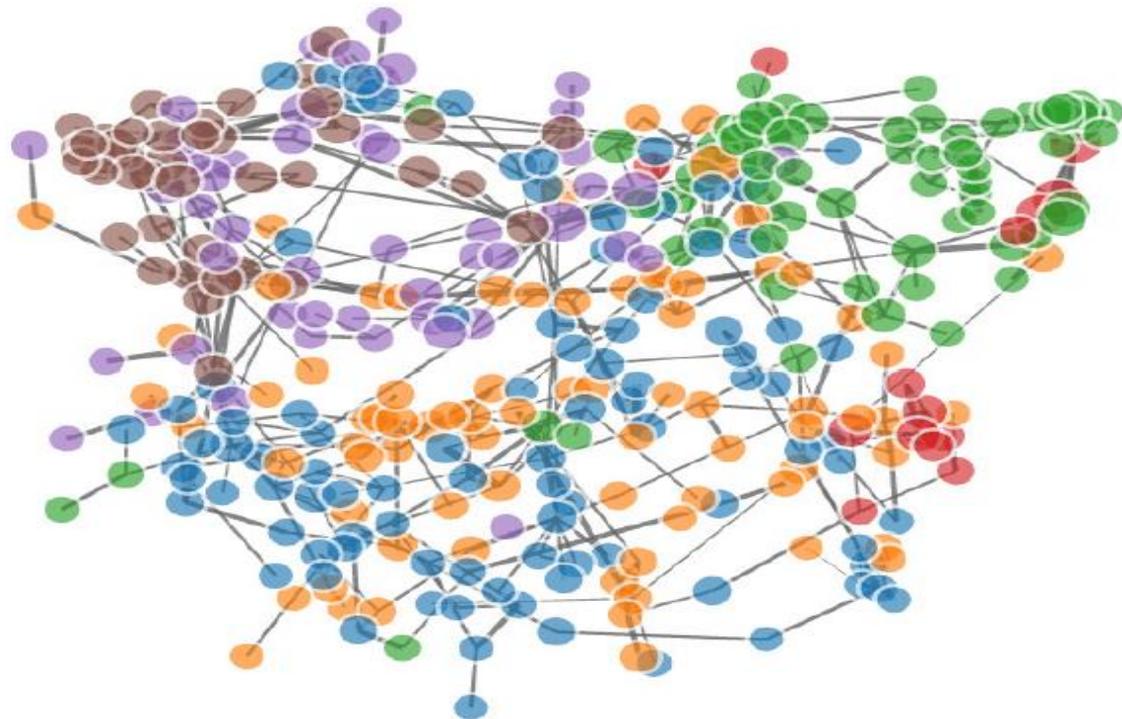
Universiteit Utrecht



Organizations and Networks



Economic and Organizational structures



ON is about applying network thinking (and complex system thinking) to solve economic and business problems

Today's objectives

- Real world networks in science and business

Today's objectives

- Real world networks in science and business
- Network thinking

Today's objectives

- Real world networks in science and business
- Network thinking
- Link organizations and networks

Today's objectives

- Real world networks in science and business
- Network thinking
- Link organizations and networks
- Structure of the class (topics, exam, project, ...)

On the side

- Discussion on big data
- Do we still need theory when we have big data?
- Data visualization techniques – art or science?

Lab #1

- Discuss project idea (& start forming groups)
- Introduction to R, RStudio and R packages
- First programming attempt

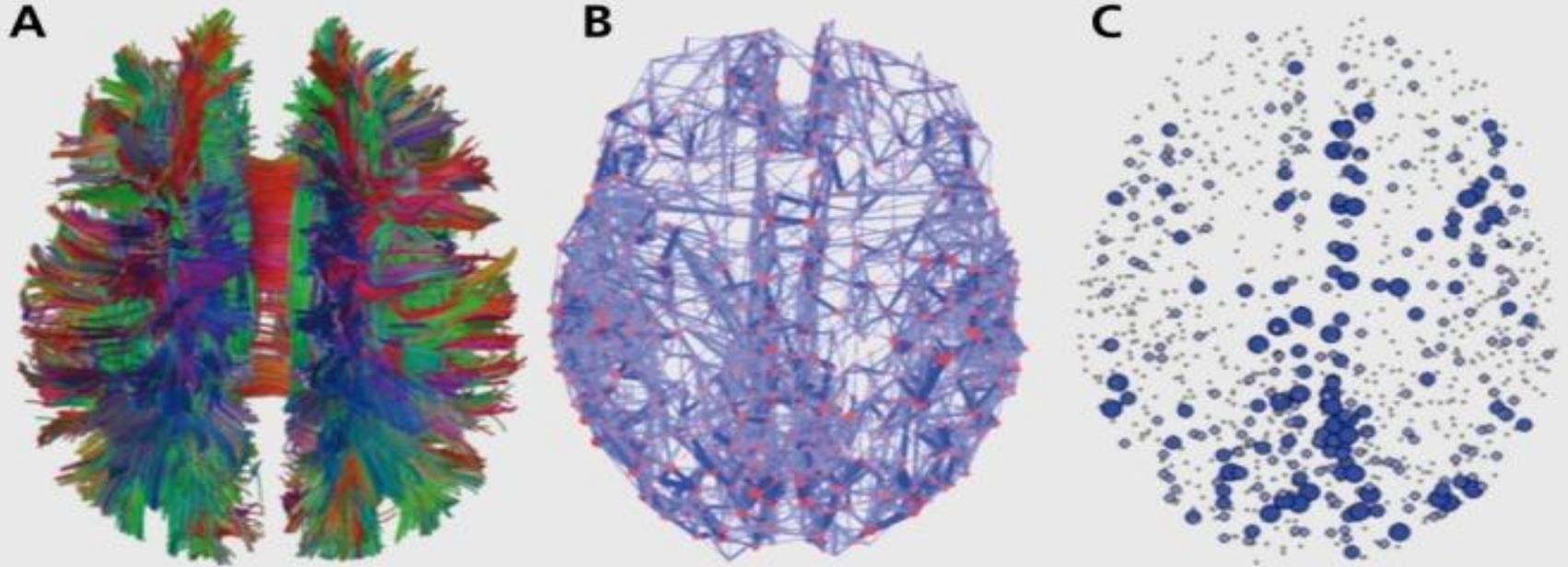
Class schedule & overview of the class

paballand.com/on

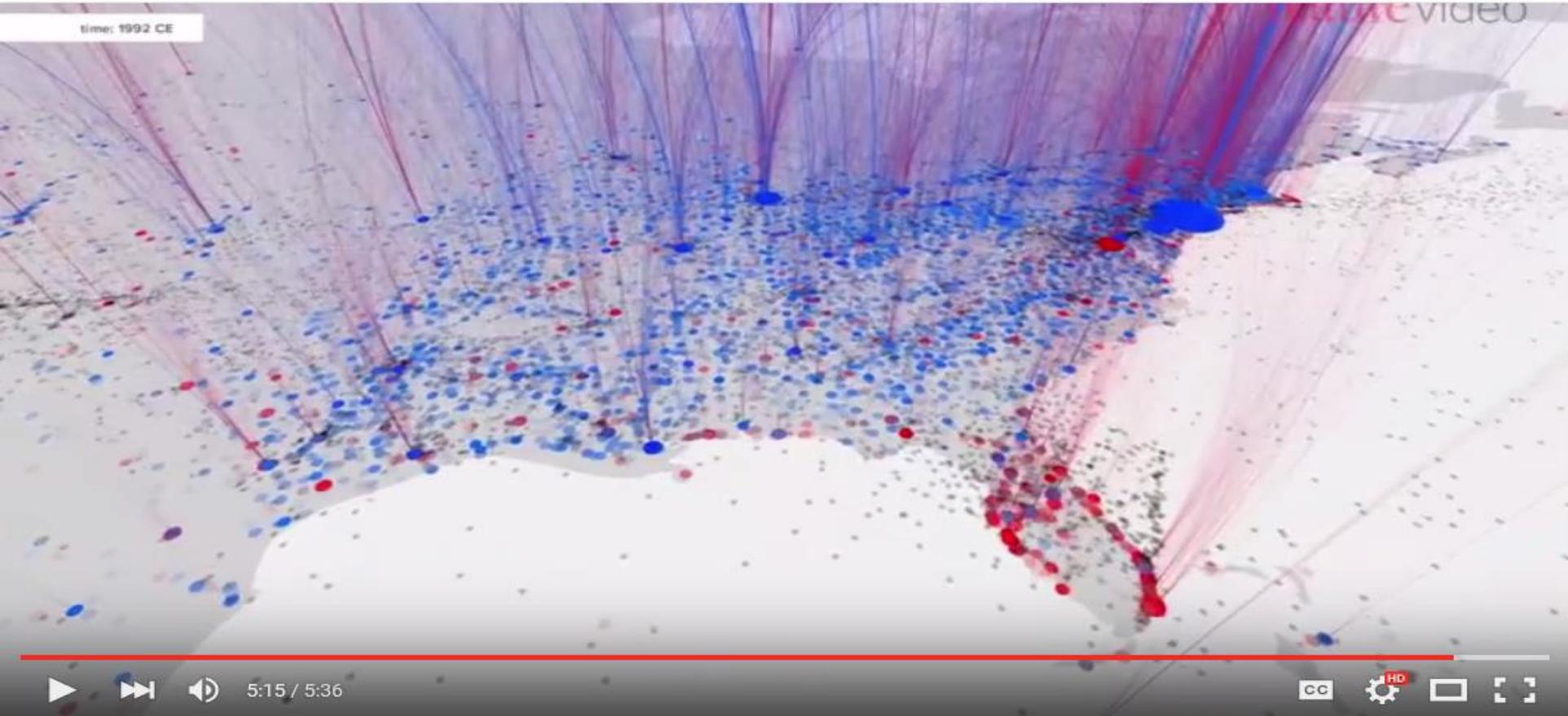
What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)

Network structure of the brain



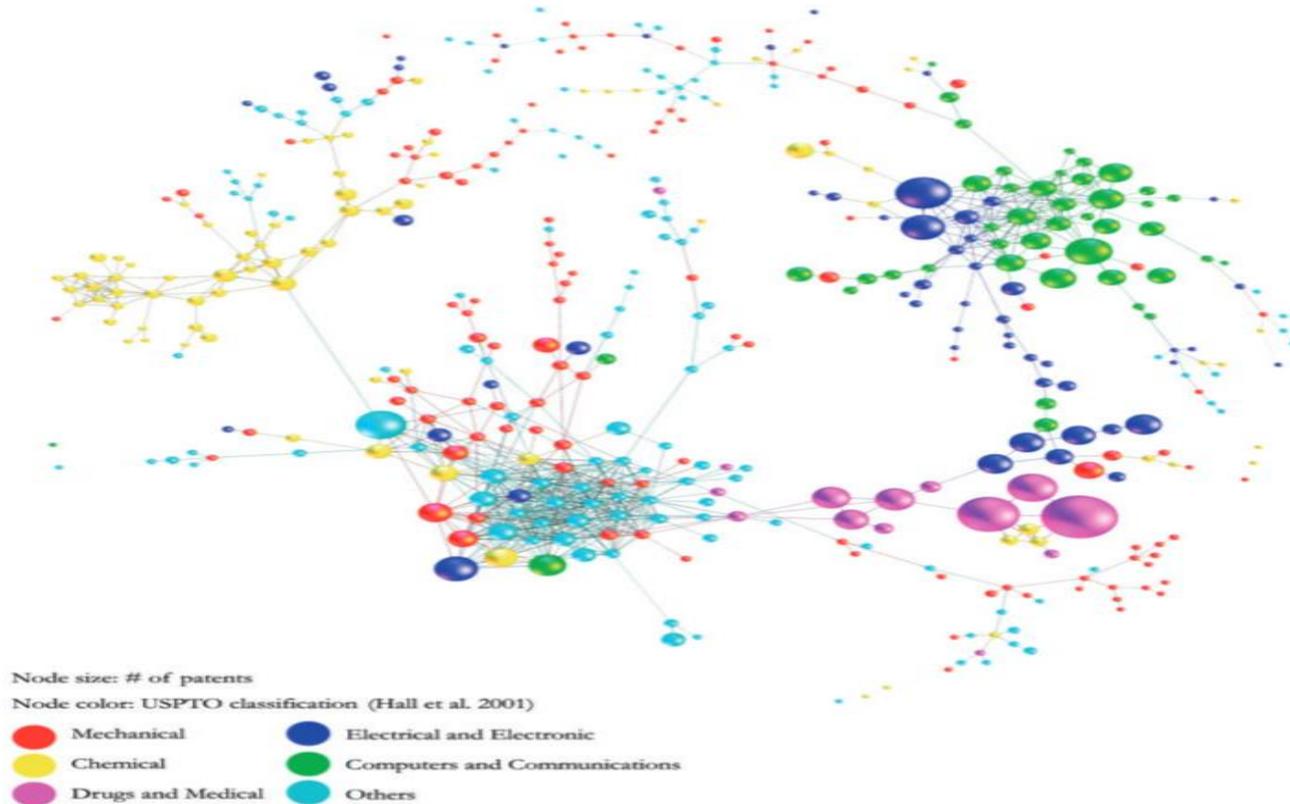
Migration flows



Knowledge flows



Knowledge relatedness



Boschma, Balland and Kogler (2013)

What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)...**but also business**

An interesting patent



US006285999B1

(12) **United States Patent**
Page

(10) **Patent No.:** **US 6,285,999 B1**
(45) **Date of Patent:** **Sep. 4, 2001**

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) **Inventor:** Lawrence Page, Stanford, CA (US)

(73) **Assignee:** The Board of Trustees of the Leland Stanford Junior University, Stanford, CA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) **Appl. No.:** 09/004,827

(22) **Filed:** Jan. 9, 1998

Related U.S. Application Data

(60) Provisional application No. 60035,205, filed on Jan. 10, 1997.

(51) **Int. Cl.**⁷ **G06F 17/30**

(52) **U.S. CL.** **707/5; 707/7; 707/501**

(58) **Field of Search** 707/100, 5, 7, 707/513, 1-3, 10, 104, 501; 345/440; 382/226, 229, 230, 231

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106	*	8/1990	Gausner et al.	345/440
5,400,535	+	9/1995	North	395/140
5,748,954	+	5/1998	Mauldin	395/610
5,752,241	+	5/1998	Cohen	707/3
5,832,494	+	11/1998	Egger et al.	707/102
5,848,407	+	12/1998	Ishikawa et al.	707/2
6,014,678	+	1/2000	Inoue et al.	707/501

OTHER PUBLICATIONS

- S. Jeromy Carriere et al., "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257-1267.
Wang et al "Prefetching in World Wide Web", IEEE 1996, pp. 28-32.
Ramer et al "Similarity, Probability and Database Organization: Extended Abstract", 1996, pp. 272-276.*

- Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221-231.*
L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometrika, vol. 18, pp. 39-43.
C.H. Hubbell, "An input-output approach to clique identification sociometry," 1965, pp. 377-399.
Mizruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26-48.
E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471-479.
Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297-312.
N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93-95.
P. Dorian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45-56.

(List continued on next page.)

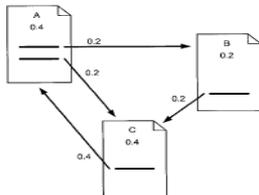
Primary Examiner—Thomas Black
Assistant Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder L.L.P.

(57)

ABSTRACT

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

29 Claims, 3 Drawing Sheets



The Google PageRank algorithm



US006285999B1

(12) **United States Patent**
Page

(10) **Patent No.:** US 6,285,999 B1
(45) **Date of Patent:** Sep. 4, 2001

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) **Inventor:** Lawrence Page, Stanford, CA (US)

(73) **Assignee:** The Board of Trustees of the Leland Stanford Junior University, Stanford, CA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) **Appl. No.:** 09/004,827

(22) **Filed:** Jan. 9, 1998

Related U.S. Application Data

(60) Provisional application No. 60/035,205, filed on Jan. 10, 1997.

(51) **Int. Cl.⁷** G06F 17/30

(52) **U.S. Cl.** 707/5; 707/7; 707/501

(58) **Field of Search** 707/100, 5, 7, 707/513, 1-3, 10, 104, 501; 345/440; 382/226, 229, 230, 231

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,953,106 *	8/1990	Gausner et al.	345/440
5,400,535 *	9/1995	North	395/140
5,748,954 *	5/1998	Mauldin	395/610
5,752,241 *	5/1998	Cohen	707/3
5,832,494 *	11/1998	Egger et al.	707/102
5,848,407 *	12/1998	Ishikawa et al.	707/2
6,014,678 *	1/2000	Inoue et al.	707/501

OTHER PUBLICATIONS

S. Jeromy Carriere et al., "Web Query: Searching and Visualizing the Web through Connectivity", Computer Networks and ISDN Systems 29 (1997), pp. 1257-1267.
Wang et al "Prefetching in World Wide Web", IEEE 1996, pp. 28-32.
Ramer et al "Similarity, Probability and Database Organization: Extended Abstract", 1996, pp. 272-276.*

Craig Boyle "To link or not to link: An empirical comparison of Hypertext linking strategies". ACM 1992, pp. 221-231.*
L. Katz, "A new status index derived from sociometric analysis," 1953, Psychometrika, vol. 18, pp. 39-43.
C.H. Hubbell, "An input-output approach to clique identification sociometry," 1965, pp. 377-399.
Mizuruchi et al., "Techniques for disaggregating centrality scores in social networks," 1996, Sociological Methodology, pp. 26-48.
E. Garfield, "Citation analysis as a tool in journal evaluation," 1972, Science, vol. 178, pp. 471-479.
Pinski et al., "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," 1976, Inf. Proc. And Management, vol. 12, pp. 297-312.
N. Geller, "On the citation influence methodology of Pinski and Narin," 1978, Inf. Proc. And Management, vol. 14, pp. 93-95.
P. Dorian, "Measuring the relative standing of disciplinary journals," 1988, Inf. Proc. And Management, vol. 24, pp. 45-56.

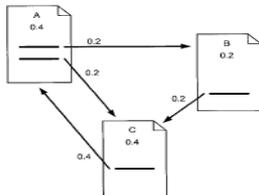
(List continued on next page.)

Primary Examiner—Thomas Black
Assistant Examiner—Uyen Le
(74) *Attorney, Agent, or Firm*—Harrity & Snyder L.L.P.

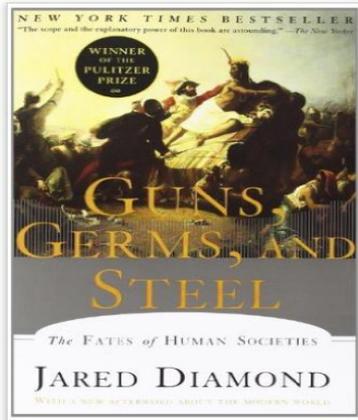
(57) **ABSTRACT**

A method assigns importance ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database. The rank assigned to a document is calculated from the ranks of documents citing it. In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

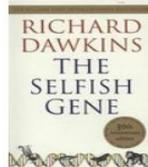
29 Claims, 3 Drawing Sheets



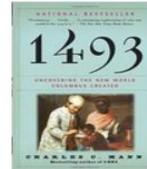
Amazon



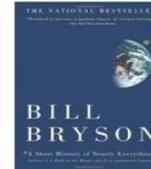
Customers Who Bought This Item Also Bought



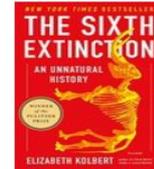
The Selfish Gene: 30th Anniversary Edition—with a new Introduction by the...
> Richard Dawkins
★★★★☆ 707
#1 Best Seller in Genetics
Paperback



1493: Uncovering the New World Columbus Created
> Charles C. Mann
★★★★☆ 513
Paperback
\$10.08 **Prime**



A Short History of Nearly Everything
> Bill Bryson
★★★★☆ 2,179
Paperback
\$9.60 **Prime**



The Sixth Extinction: An Unnatural History
> Elizabeth Kolbert
★★★★☆ 1,174
#1 Best Seller in Natural History
Paperback

Facebook recommendation



People You May Know

Add Friend

Remove

Social networks and population mapping



facebook

December 2010

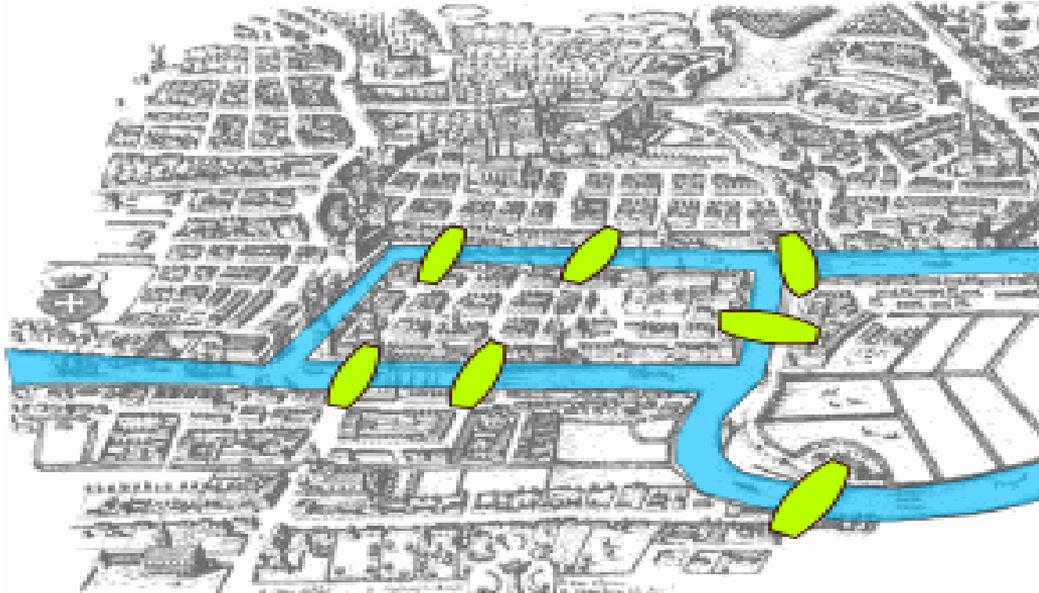
Paul Butler (Facebook - 2010)

What is network thinking?

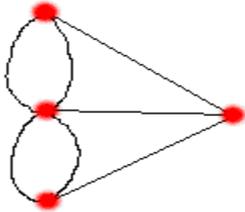
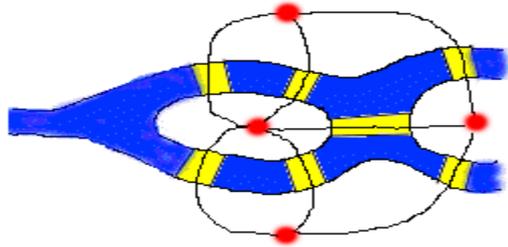
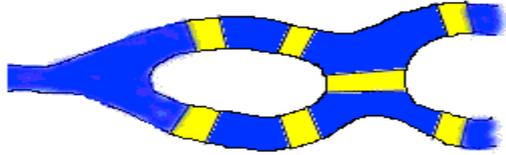
- A network-based paradigm is taking science by storm (Barabási, 2012)...but also business
- Network analysis is a broad intellectual approach instead of a narrow set of methods (Wellman, 1983)

The seven bridges of Königsberg

Fundamental problem in the history of mathematics : find a walk through the city that would cross each bridge once and only once



Leonhard Euler solution (1735)



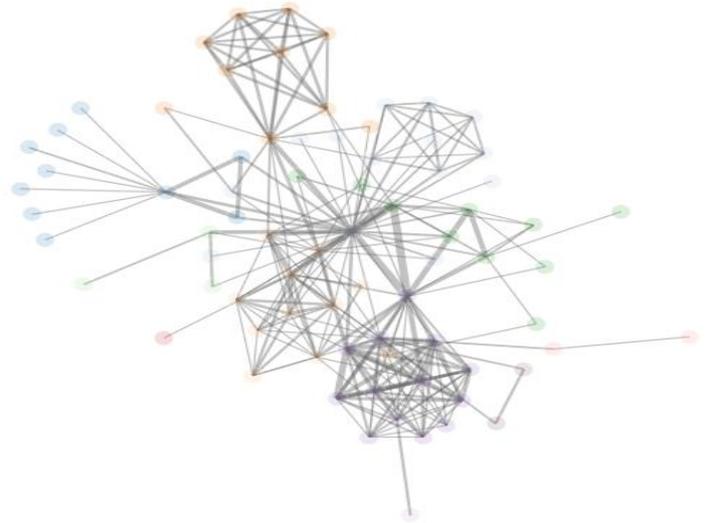
- The route inside each land mass is irrelevant, only the patterns of connection are important
- Abstract reformulation: collapse areas of land separated by the river into points (nodes) connected by the 7 bridges (edges)
- Euler used a network (graph-based) approach to prove that there is no path that would cross each bridge once and only once
- Foundation of graph theory and mathematical topology

What is network thinking?

- A network-based paradigm is taking science by storm (Barabási, 2012)...but also business
- Network analysis is a broad intellectual approach instead of a narrow set of methods (Wellman, 1983)
- A network-based paradigm shifts the unit of analysis from **individuals** and their attributes to (the structure of) their **relationships**

Network metrics & visualization

- Network centrality
- Brokerage
- Network density
- Core-periphery structure
- Average path length
- Clustering coefficient
- Communities
- Degree distribution
- Statistical model of network dyna
- ...



Economic complexity



CALL FOR PAPERS in [Research Policy](#) Special Issue on [Economic Complexity](#)

Guest editors

Pierre-Alexandre Balland ([Utrecht University](#) & [Collective Learning Group, MIT Media Lab](#)), Tom Broekel ([Utrecht University](#)), Dario Diodato ([CID Harvard](#)), Ricardo Hausmann ([CID Harvard](#)), Neave O'Clery ([Oxford](#)), and David Rigby ([University of California, Los Angeles](#))

Lead editor

Elisa Giuliani ([University of Pisa](#))



CLUe Training 16: EconGeo: Economic Geography in R

Time: 11:40-13:00 (with lunch), Friday 14 September

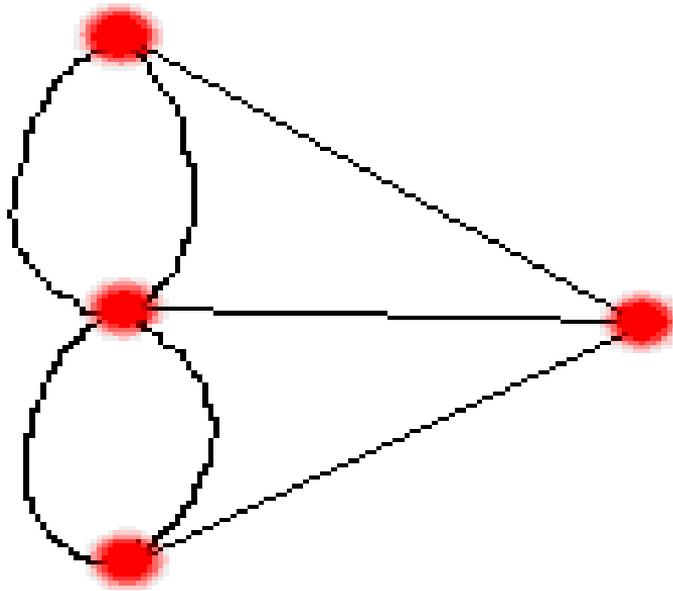
Content: Dr. Pierre-Alexandre Balland will introduce his R package EconGeo for understanding the spatial dimension of economic activities.

[Info & Register >>](#)

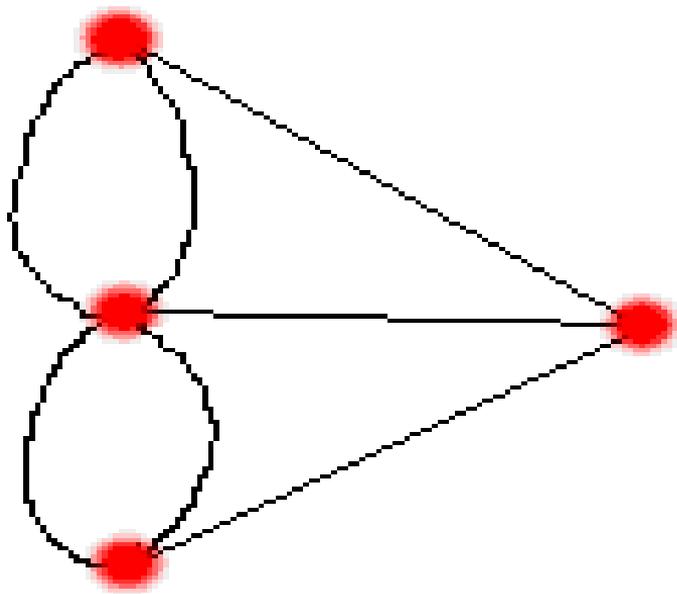
Network thinking in music



Network thinking reduces complexity



Hidden properties of network structures



Thanks!

paballand.com

github.com/PABalland/EconGeo

Computer lab: R & RStudio

- In this course we will perform structural network analysis with packages implemented in the R statistical software
- R is the software – but we will use Rstudio as an interface
- R is an open-source project lifted by a virtual community of thousands of developers and million of users worldwide

Why R?

- Reproducibility – R scripts
- Today R offers the most elegant and comprehensive language for the structural and dynamic analysis of networks
- It's free and contains state-of-the-art statistical and graphical routines not yet available in other software
- You can do all your analysis in R, but also data scrapping, create a webpage, or write your research paper

Getting started with R

- Using R is easier than it looks like. And once you master it, you save a ridiculous amount of time
- Afraid of R? It is just a big calculator (a very smart one)
- R is case sensitive
- The `#` character at the beginning of a line signifies a comment, it is ignored by R



The `#` character at the beginning of a line signifies a comment, it is ignored by R

RStudio

The image shows the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. Below the menu is a toolbar with icons for file operations and a search bar. The main editor window is titled 'Untitled1' and contains a blue-bordered box with the text 'EDITOR WINDOW: Your script (code)'. The console window at the bottom displays the R version information and help text. The environment and history panes are visible on the right side of the interface.

1

**EDITOR WINDOW:
Your script (code)**

```
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"  
Copyright (C) 2016 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (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.  
  
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.  
  
[Workspace loaded from ~/.RData]  
> |
```

RStudio

The image shows the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. Below the menu bar is a toolbar with icons for file operations and a search bar. The main workspace is divided into three panes: a source editor on the left, an environment/history pane on the top right, and a console/plots/packages/help/viewer pane on the bottom right. The source editor contains a single line of code, '1', and is annotated with a blue box containing the text 'EDITOR WINDOW: Your script (code)'. The console pane displays the R startup message and is annotated with a blue box containing the text 'CONSOLE: Run your code'.

EDITOR WINDOW:
Your script (code)

CONSOLE:
Run your code

```
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"  
Copyright (C) 2016 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)  
  
R is free software; you are welcome to redistribute it under certain  
conditions. Type 'license()' to see the full text of the license.  
  
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.  
  
[Workspace loaded from ~/.RData]  
  
> |
```

RStudio

The image shows the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. Below the menu is a toolbar with icons for file operations and a search bar. The main workspace is divided into several panes:

- Editor Window:** The top-left pane, containing a script file named 'Untitled1'. It is annotated with a blue box containing the text: **EDITOR WINDOW: Your script (code)**.
- Environment:** The top-right pane, currently empty. It is annotated with a blue box containing the text: **WORKSPACE: Your variables, functions, etc...**.
- Console:** The bottom-left pane, showing the R startup message and a prompt. It is annotated with a blue box containing the text: **CONSOLE: Run your code**. The console output includes:

```
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"  
Copyright (C) 2016 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)  
  
R is free software; you are welcome to redistribute it under certain conditions.  
You are welcome to receive free copies of the R source code.  
Type 'license()' to see the full license.  
  
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.  
  
[Workspace loaded from ~/.RData]  
  
> |
```
- Environment, History, and Presentation:** The top-right pane, currently empty.
- Files, Plots, Packages, Help, and Viewer:** The bottom-right pane, currently empty.

RStudio

The image shows the RStudio desktop application. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. Below the menu is a toolbar with icons for file operations and running code. The main workspace is divided into several panes:

- Editor Window:** The top-left pane, containing a script file named 'Untitled1'. A blue-bordered box with the text 'EDITOR WINDOW: Your script (code)' is overlaid on this pane.
- Console:** The bottom-left pane, showing the R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes" and copyright information. A blue-bordered box with the text 'CONSOLE: Run your code' is overlaid on this pane.
- Environment/History/Presentation:** The top-right pane, currently empty. A blue-bordered box with the text 'WORKSPACE: Your variables, functions, etc...' is overlaid on this pane.
- Files/Plots/Packages/Help/Viewer:** The bottom-right pane, currently empty. A blue-bordered box with the text 'EXTERNAL: Files, plots, packages,...' is overlaid on this pane.

```
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software; you are welcome to redistribute it under certain
conditions. Type 'license()' to see the full text of the license.

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.

[Workspace loaded from ~/.RData]

> |
```

Let's get started

The image shows the RStudio interface. The top-left pane is the script editor, containing the code `1 10+5|`. The top-right pane is the Environment, History, and Presentation view. The bottom-left pane is the Console, displaying the R startup message and the output of the command `> 10+5`, which is `[1] 15`. The bottom-right pane is the Files, Plots, Packages, Help, and Viewer view. A blue arrow points from the 'Run' button in the toolbar to the code in the script editor. Another blue arrow points from the console output back to the code in the script editor.

```
1 10+5|
```

Environment History Presentation x

Files Plots Packages Help Viewer

Zoom Export

Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (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.

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.

[workspace loaded from ~/.RData]

```
> 10+5  
[1] 15  
> |
```

Create a variable "x"

The screenshot displays the RStudio interface. In the top-left pane, the source editor shows the following R code:

```
1 10+5  
2 x <- 10+5
```

A blue arrow points from the second line of code to the Environment pane on the right. The Environment pane shows the Global Environment with a table of values:

values	
x	15

The bottom pane shows the R console output:

```
PLATFORM: x86_64-w64-mingw32/x64 (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.  
  
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.  
  
[Workspace loaded from ~/.RData]  
  
> 10+5  
[1] 15  
> x <- 10+5  
> |
```

Create a variable y

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains three lines of R code:

```
1 10+5
2 x <- 10+5
3 y <- "geography"
```
- Environment Pane:** Shows the 'Global Environment' with a 'Values' table:

Variable	Value
x	15
y	"geography"
- Console:** Shows the output of the R script:

```
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.

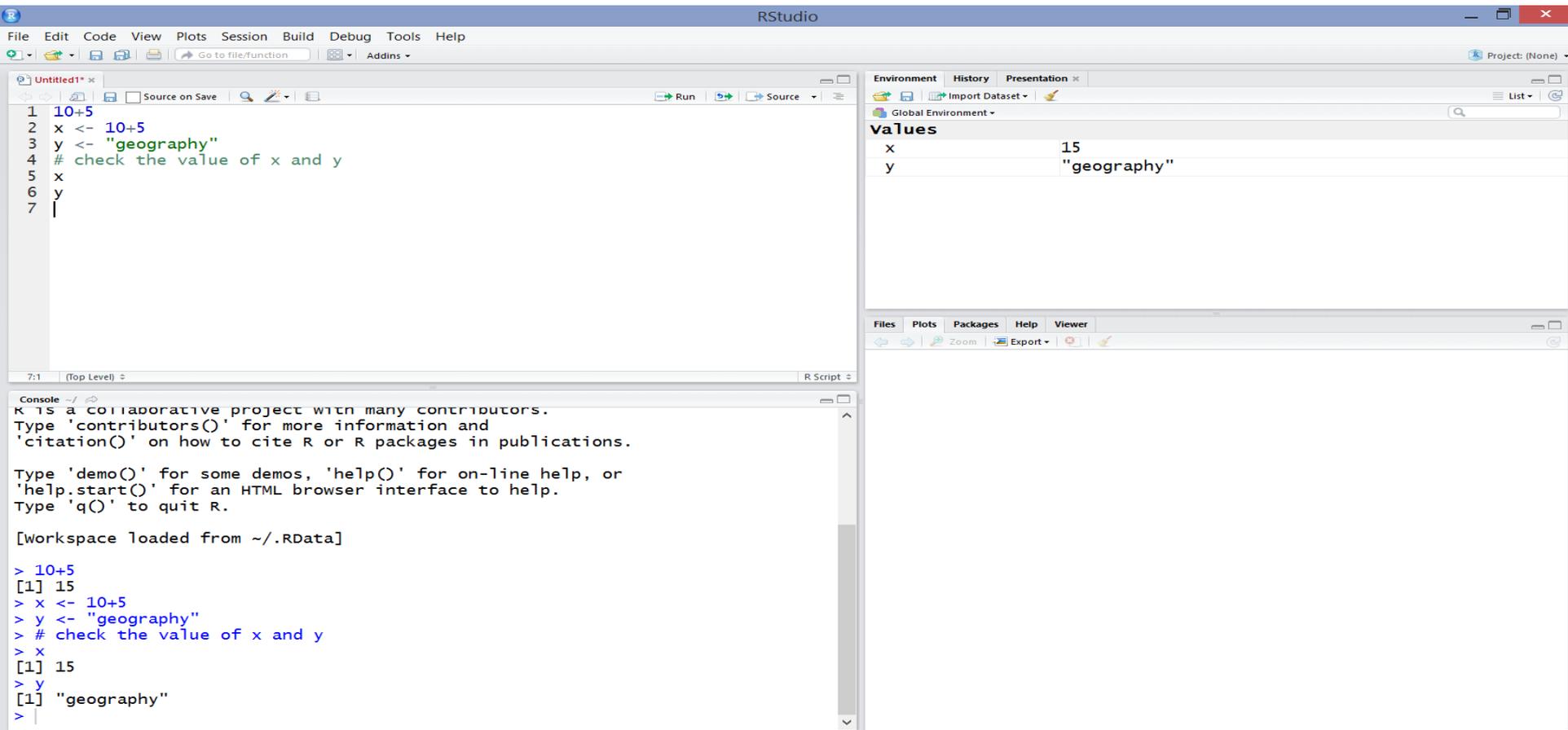
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

[workspace loaded from ~/.RData]

> 10+5
[1] 15
> x <- 10+5
> y <- "geography"
> |
```

Check the value of x and y



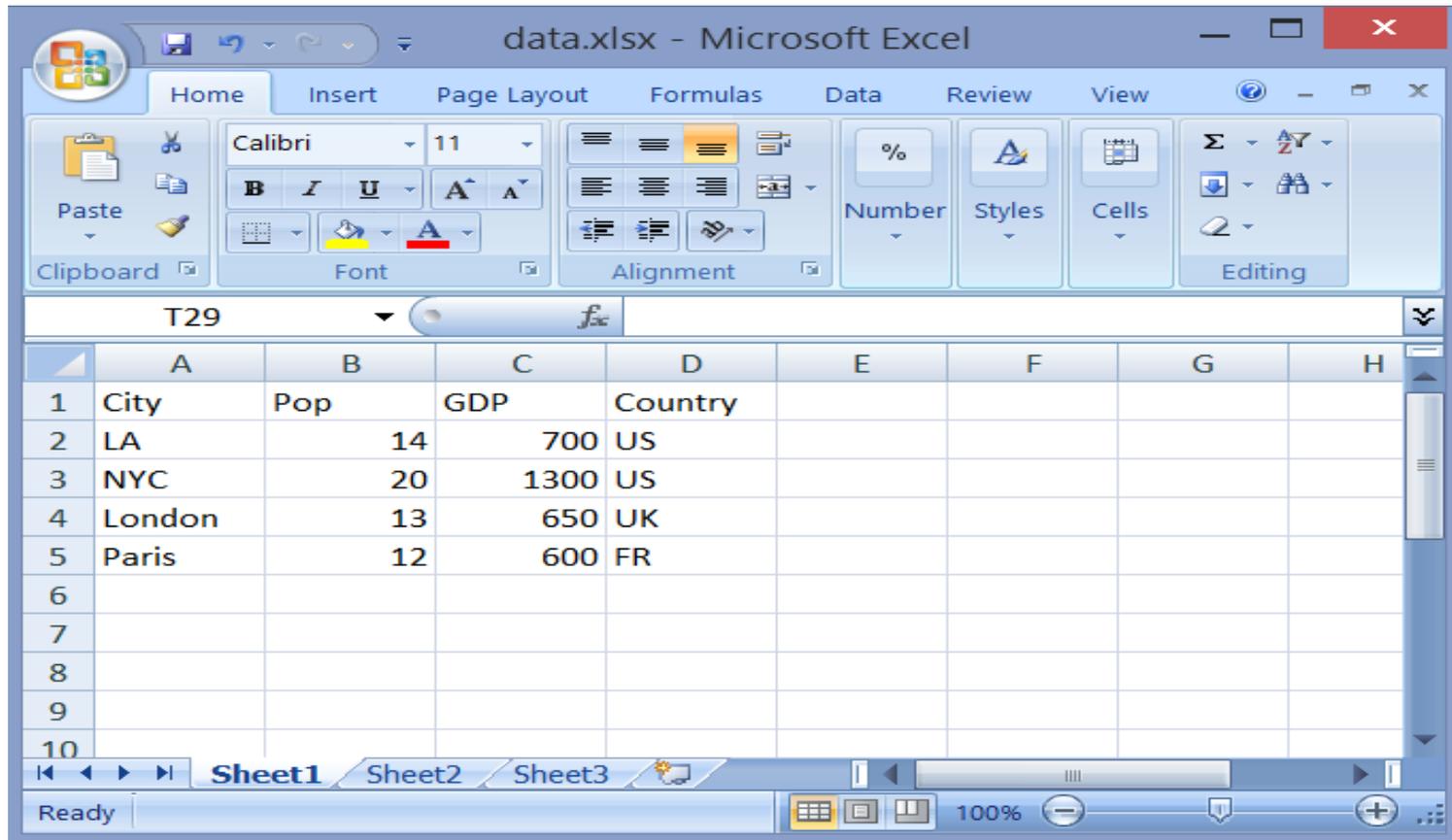
The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for calculating $10+5$, assigning the result to `x`, and assigning the string "geography" to `y`. A comment indicates the goal is to check the values of `x` and `y`.
- Environment Panel:** Shows the current environment with two variables:

Variable	Value
x	15
y	"geography"
- Console:** Shows the execution output, including the workspace loading message and the results of the R commands:

```
[Workspace loaded from ~/.RData]
> 10+5
[1] 15
> x <- 10+5
> y <- "geography"
> # check the value of x and y
> x
[1] 15
> y
[1] "geography"
```

Let's create a toy dataset

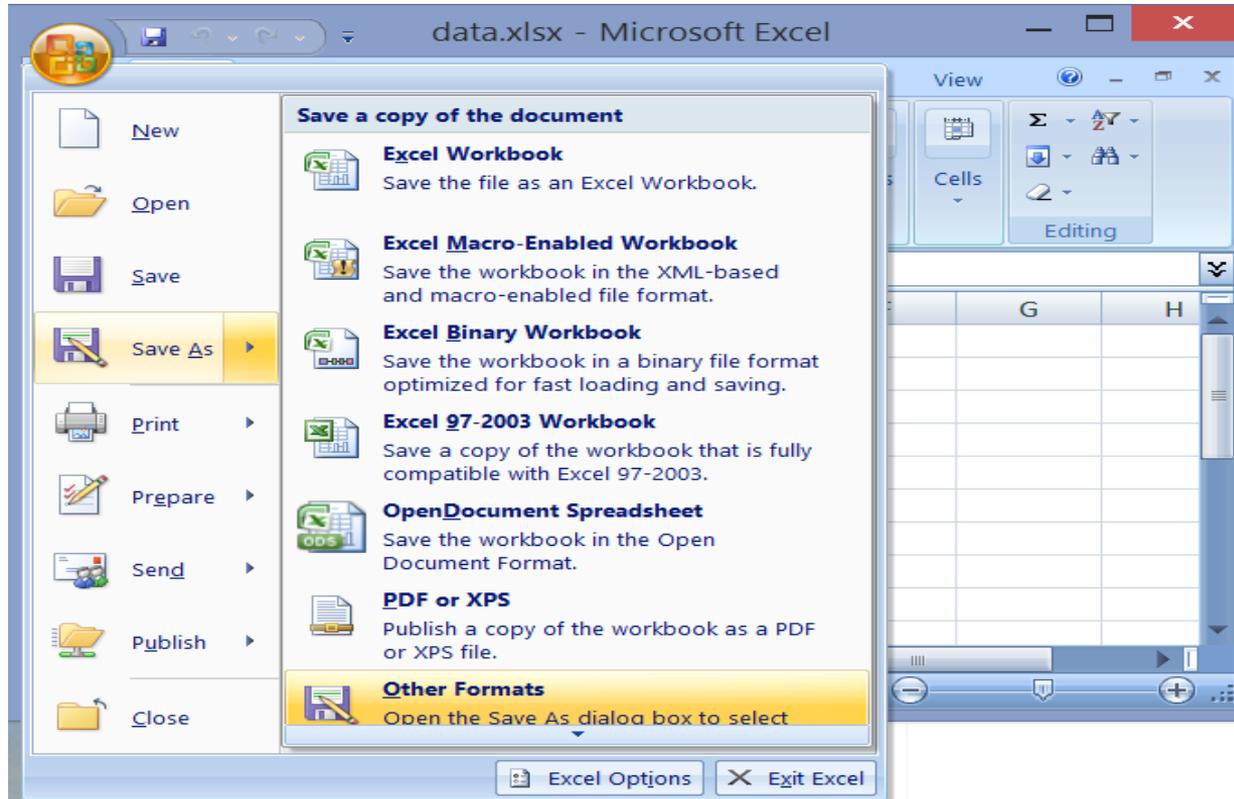


The screenshot shows the Microsoft Excel interface with a spreadsheet titled "data.xlsx". The ribbon is set to "Home", and the "Font" group is active. The spreadsheet contains a table with the following data:

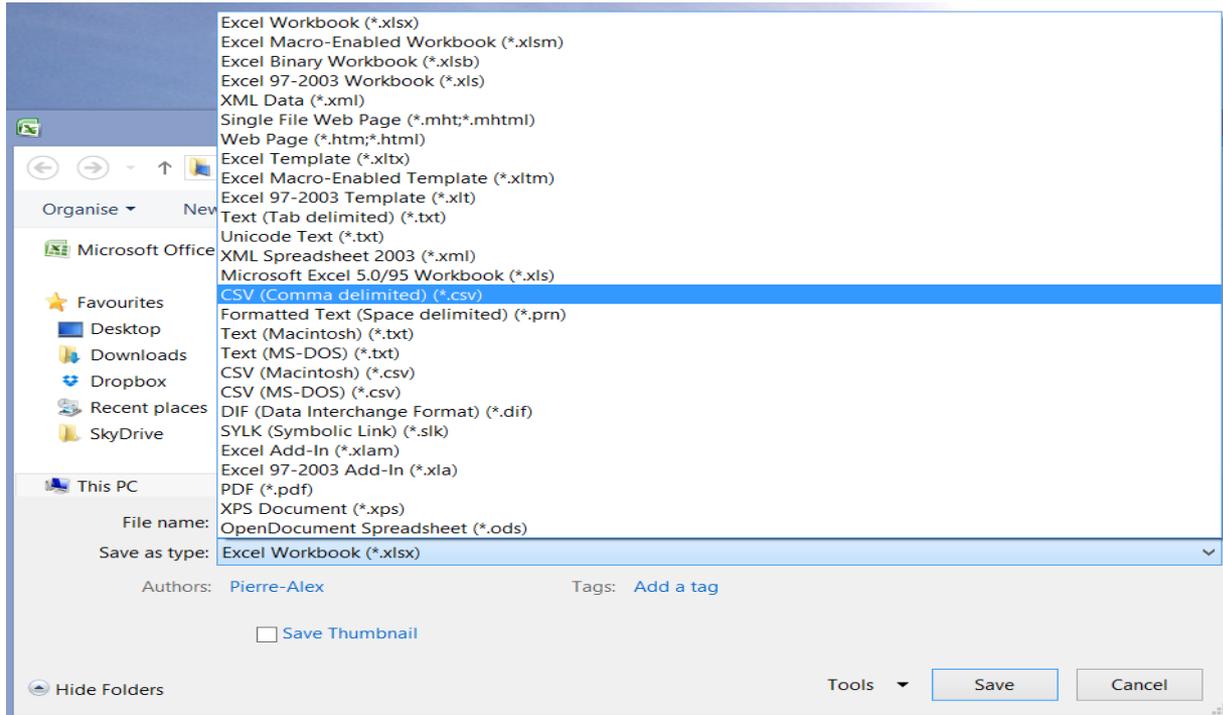
	A	B	C	D	E	F	G	H
1	City	Pop	GDP	Country				
2	LA	14	700	US				
3	NYC	20	1300	US				
4	London	13	650	UK				
5	Paris	12	600	FR				
6								
7								
8								
9								
10								

The status bar at the bottom indicates "Ready" and "100%". The sheet tabs at the bottom show "Sheet1", "Sheet2", and "Sheet3".

Save as a .csv file



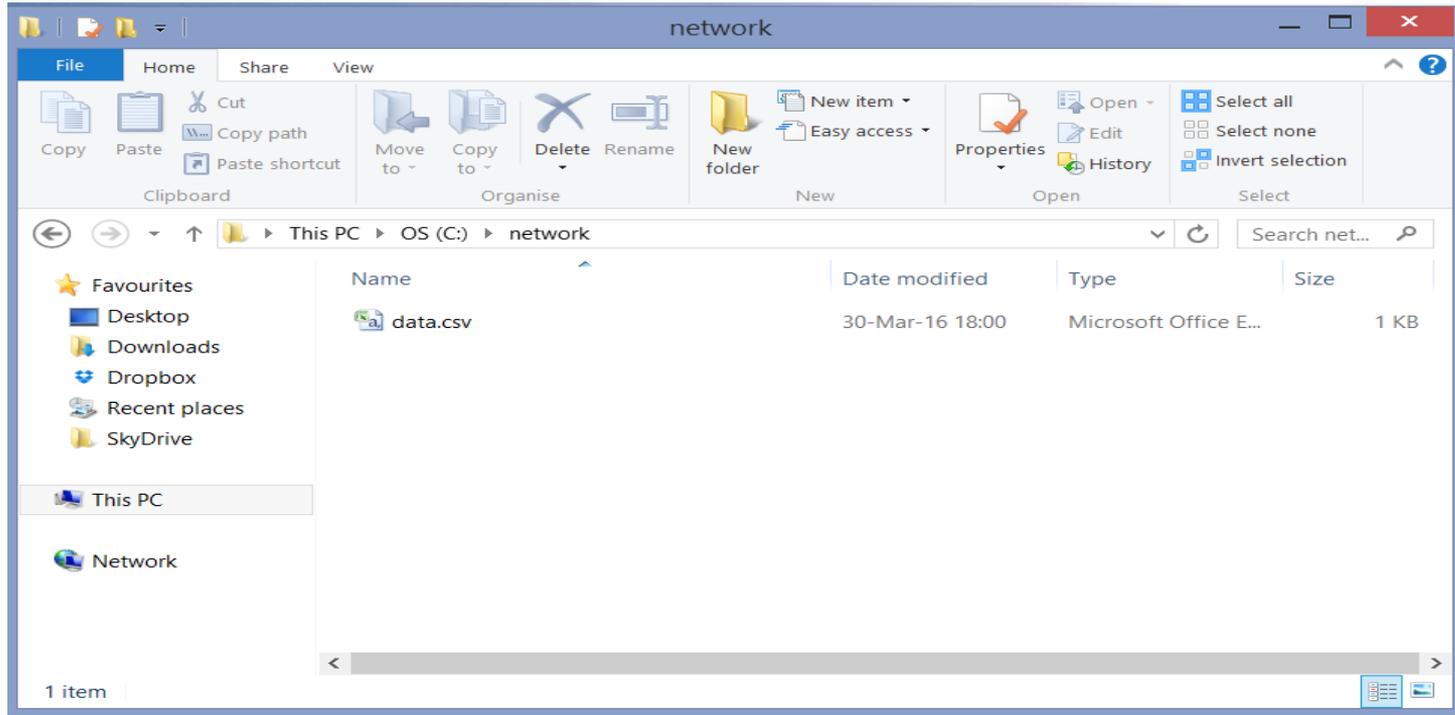
Save as a .csv file



Two warnings: ok



Create a new folder and move the .csv



This is your file path

