

Práctico 1 - Manipular grafos de redes

Análisis de Redes

Pablo Rodríguez-Bocca

1) Crear ambiente para el lenguaje R.

Se instalará el ambiente de trabajo (IDE) y los paquetes básicos necesarios para todos los prácticos. Este proceso puede demorar y se recomienda hacerlo previo al práctico.

A) Instalar Rstudio. Descargar desde:

<https://www.rstudio.com/products/rstudio/>

B) Instalar Paquetes básicos. En una consola de R:

```
install.packages("sand")
library(sand)
install_sand_packages()
```

Si el procedimiento anterior falla (porque el sistema o los paquetes se encuentran desactualizados) entonces recomiendo agregar los siguientes comandos a la instalación:

```
libs = c("igraph", "igraphdata", "GO.db", "GOstats", "ROCR", "ape", "car",
        "eigenmodel", "ergm", "fdrtool", "ggplot2", "huge", "kernlab",
        "lattice", "mixer", "network", "networkDynamic",
        "networkTomography", "ngspatial", "org.Sc.sgd.db", "sna", "vioplot")
new.packages <- libs[!(libs %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
```

2) Tutorial del lenguaje R.

Es necesario conocer el lenguaje R para ganar fluidez en el práctico. Quien no conoce el lenguaje debe realizar algun tutorial rápido, y al menos el siguiente:

```
library("igraph")
igraph_demo("crashR")
```

3) Tutorial de manipular grafos de redes.

Si aun no lo tiene, descargar el libro de práctico [SANDR]: Kolaczyk, E.D. and Csárdi, G. “Statistical Analysis of Network Data with R”. Use R!, Springer New York, 2014. ISBN 9781493909834. Disponible en <http://timbo.org.uy/>.

Seguir el texto del Capítulo 2 del libro [SANDR], ejecutando el código fuente incluido. Utilizar el manual de referencia del paquete `igraph` para extender el conocimiento:

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

4) Explorar una red de comunicaciones: emails de la empresa Enron.

La red de emails de la empresa Enron se publicó durante una investigación federal de EE.UU¹. En éste ejercicio haremos cálculos sencillos sobre el grafo de esta red usando R.

¹Si usted esta interesado en leer más sobre el escandalo de Enron, ver http://en.wikipedia.org/wiki/Enron_scandal

La red completa está disponible online en <http://www.cs.cmu.edu/~enron/> y contiene 517.431 correos extraídos de los directorios de correo de 150 usuarios. Aquí utilizaremos un grafo más pequeño, preparado por C. E. Priebe et al., que consiste en 34.427 correos enviados entre 184 correos de empleados de Enron (ver <http://cis.jhu.edu/~parky/Enron/enron.html> por más detalles). Los correos son del periodo 13/11/1998 al 21/06/2002 (44 meses).

A) Descargar el subconjunto de datos de `Enron.zip` de la página del curso, o directamente de:

<https://eva.fing.edu.uy/pluginfile.php/132941/course/section/14262/Enron.zip>

El `.zip` contiene dos archivos creados en Matlab. Para cargar archivos de Matlab en R, utilizar el paquete `R.matlab`. El arreglo de tres dimensiones $\mathbf{Y} \in \mathbb{R}^{184 \times 184 \times 44}$ contiene la secuencia de matrices de adyacencia que representa la red de quién le envió correo a quién para cada uno de los 44 meses. Notar que el grafo de la red es dirigido y ponderado, dado que el elemento $[\mathbf{Y}]_{ijt}$ indica la cantidad de correos enviados por el empleado i al empleado j durante el mes t . Construir la matriz binaria de adyacencia $\mathbf{A} \in \{0, 1\}^{184 \times 184}$ que agrega todos los correos a lo largo del tiempo (es decir, sumar \mathbf{Y} en la tercera dimension, y definir $[\mathbf{A}]_{ij} = \mathbb{I} \left\{ \sum_{t=1}^{44} [\mathbf{Y}]_{ijt} > 0 \right\}$). Además, también cargar el arreglo `employees` que contiene el nombre, la cuenta y cargo en la empresa para cada uno de los 184 empleados (vértices) en G .

B) Construir el grafo (con *igraph*) a partir de la matriz de adyacencia \mathbf{A} y agregar los atributos de vértices del arreglo `employees`.

Para cada uno de las siguientes partes, realizar los cálculos sobre el grafo (usando métodos de *igraph*), y usando la matriz de adyacencia \mathbf{A} (usando álgebra de matrices).

C) Calcular la cantidad de aristas (arcos) en la red (es decir, la cantidad de parejas únicas ordenadas $(u, v) \in E$ donde $u, v \in V$).

D) Calcular la cantidad de aristas no dirigidas en la red (es decir, la cantidad de parejas únicas no ordenadas $(u, v) \in E$ donde $u, v \in V$. (Esto significa que si $(u, v) \in E$ o $(v, u) \in E$, entonces se debe contar el par como una única arista.)

E) Calcular la cantidad de arcos mutuos en la red (es decir, la cantidad de parejas (u, v) donde $\{(u, v), (v, u)\} \subseteq E$ y $u, v \in V$). (Esto significa que si existen ambas $(u, v) \in E$ y $(v, u) \in E$, debe contarse el par como mutual.)

F) Calcular la cantidad de vértices con $d_v^{\text{in}} = 0$, y listar los nombres de los empleados correspondientes.

G) Calcular la cantidad de vértices con $d_v^{\text{out}} = 0$, y listar los nombres de los empleados correspondientes.

H) Calcular la cantidad de empleados que fueron contactados por 30 o más empleados.

I) Calcular la cantidad de empleados que contactaron a 30 o más empleados.

J) Calcular la cantidad de triángulos dirigidos en G . (Recordar que G es dirigido y la orientación de los triángulos importa, es decir, $i \rightarrow j \rightarrow k \rightarrow i$ es lo mismo que $k \rightarrow i \rightarrow j \rightarrow k$, pero diferente a $i \rightarrow k \rightarrow j \rightarrow i$.)

K) ¿Es G débilmente conexo?