# Análisis computacional

# y de redes en Ciencias Sociales

Ernesto Calvo

[ecalvo@umd.edu](mailto:ecalvo@umd.edu)

Universidad de Maryland

Tiago Ventura

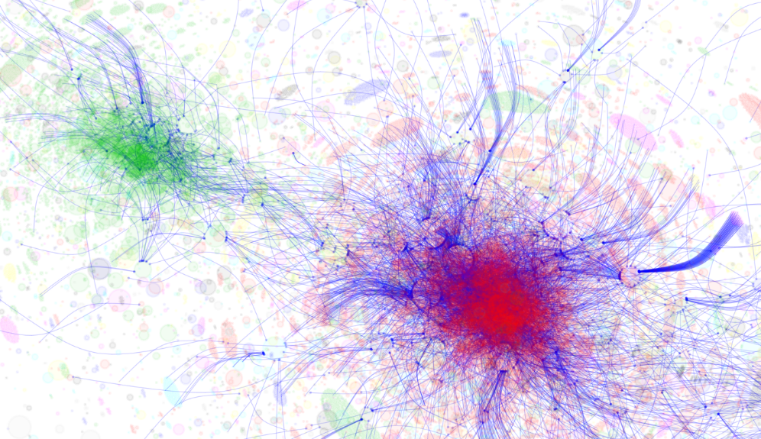
[venturat@umd.edu](mailto:venturat@umd.edu)

Universidad de Maryland

Flacso/Mexico

Agosto, 2021

Este curso está orientado a profesionales y estudiantes de Ciencias Sociales, interesados en el uso de métodos cuantitativos para el análisis de grandes datos, con énfasis en el análisis datos no estructurados y de redes sociales. El objetivo es ofrecer una exposición sistemática de los conceptos fundamentales para quienes quieran participar del vínculo cada vez más estrecho entre el Análisis de Datos y las Ciencias Sociales.

El seminario está estructurado en 10 clases de 3 horas. Se complementarán las clases con un laboratorio/práctico durante la tarde. Será dictado en R/R-Studio.

**Programa:**

* Introducción a R: Interfaz de R-Studio, operaciones básicas, vectores. Carga y manipulación de datos, *Data Frames*.
* Análisis de texto: Bolsa-de-palabras vs. Análisis semántico. Codificación de Diccionario.
* Análisis semántico: extracción de espacio latente de un corpus.
* Extracción de conocimiento: tópicos de aprendizaje automático.
* Análisis de redes sociales: conceptos de redes, detección de comunidades y visualización de redes.
* Análisis de corpus de textos en redes sociales: hashtags, análisis de clusters, redes semánticas, visualización de contenidos.
* **Requisitos:**Concurrir con una computadora con R y RStudio instalado. (<https://cran.r-project.org/> y [**https://www.rstudio.com/**](https://www.rstudio.com/)).
* Tener un conocimiento básico de R, como el disponible en el [siguiente tutorial.](http://www.fundacionsadosky.org.ar/mailings/pcd/intro.html)

**Para instalar los paquetes requeridos para la clase correr en R o R-Studio el código:**

sapply(c("rvest","RCurl","XML","httr","xml2","RCurl","tm","igraph","foreign","twitteR", "httpuv","lmer","streamR","quanteda","readtext"),install.packages)

**Repositorio**

<https://github.com/TiagoVentura/ventura_calvo_flacso_workshop/blob/main/install_fest.md>

**Selección de Textos de Apoyo**

<https://www.dropbox.com/sh/38x2cj4rwouv092/AAArq9fg8RffvG2Rnkp2zZUpa?dl=0>

# Programa

**Semana 1: Tiago Ventura**

**Readings**

* [Salganik, M. 2017. Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press.](https://www.bitbybitbook.com/en/1st-ed/preface/)
* [Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. O’Reilly Media, Inc.](https://r4ds.had.co.nz/)
* [Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis, 21(3), 267-297.)](https://web.stanford.edu/~jgrimmer/tad2.pdf)
* [Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. O’Reilly Media, Inc.](https://www.tidytextmining.com/)

**Sitio:**

Todos los materiales de la primera semana del curso se pueden encontrar en el siguiente enlace:

<https://github.com/TiagoVentura/ventura_calvo_flacso_workshop>

**Pre llegada:**

Antes de comenzar nuestra clase el lunes, debe trabajar con este material de incorporación que preparamos para usted. Sitio abajo:

<https://github.com/TiagoVentura/ventura_calvo_flacso_workshop/blob/main/install_fest.md>

**Día 1**

**Modulo 1: Teórico - Introducción as Ciencias Sociales Computacionales**

* 9 a 10hs: Introducción a las ciencias sociales computacionales*.*

*Leer: Salganik, M. 2017. Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press, Capítulo 1 y 2*

**Módulo 2, Introducción a R**

* 10:00 a 12:00hs: Introducción a R. Flujo de trabajo: Scripts. Tipos de datos. Importación de datos

Leer: R for Data Science, Hadley Wickham, capítulos 4, 6, 11, 20.

**Día 2**

**Módulos 3 y 4, Tidyverse: Data Wrangling.**

* Tópicos: Pipes, Dados en formato tidy, dplyr, joins.
* Leer: R for Data Science, capítulos 5, 12, 13.

**Día 3**

**Módulo 5: Extra en Tidyverse**

* Topicos: Strings en R.
* Leer: R for Data Science, capítulo 14

**Módulo 6: Programación de Funciones y Iteración**

* Topicos: Funciones, Purr, Loops
* Leer: R for Data Science, capítulo 19 y 21

**Día 4**

**Módulo 7: Webscrapping: Datos en Sitios on-line**

* **Tópicos: Rvest, Extraccion de Htmls y CSS**

**Módulo 8: Webscrapping: Datos de APIs**

* **Tópicos: APIs, httr, Jsons.**

**Día 5**

**Módulo 9: Introducción teórica al Análisis de Texto:**

* **Leer:** Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis, 21(3), 267-297.

**Módulo 10: Análisis de Texto con datos tidy**

* **Tópicos:** Datos de Text em R, Datos de Texto en Formato Tidy, tidytext, Análisis de Sentimiento y diccionarios
* **Leer:** Text Mining with R

**Semana 2: Ernesto Calvo**

**Módulo 11: Big Data con Redes.**

* Datos relacionales
* Nodos, Aristas, tipos de redes.

**Módulo 12: Filas JSON, Vertices y Aristas, Manipulacion de Redes en R.**

* Introduccion a iGraph.
* V(net) vs. E(net)
* Centralidad, betweennes, paths

**Módulo 13: Layouts.**

* Force-Directed, MDS, Random.
* Core, Periphery.

**Módulo 14: Community Detection.**

* Greedy algorithms, Random.Walk.
* Mapping attributes

**Módulo 15: Activation.**

* Dependency and the Friendship paradox.

**Módulo 16: Análisis de texto en Twitter I.**

* Quanteda: An introduction.
* From text to nodes.

**Módulo 17: Bag of words and maximum discrimination.**

* Grams, nGrams, Frequency Matrix.

**Módulo 18: Propagación de Informacion en Redes II.**

* Ejercicios con redes.

**Módulo 19: A donde va el estudio de “Big Data”**

* Más redes, más nodos, más problemas.
* Deep learning, NPL 2.0.

**Módulo 20: Presentacion de Resultados y Cierre.**

* Sobre los trabajos practicos.
* Donde continuar nuestra formación profesional.