

Unha xenealoxía das teses doutorais en economía e empresa en Galicia co paquete *ggenealogy*

Jose Blanco Álvarez ¹

¹Universidade de Santiago de Compostela, Dep. Ec. Aplicada, grupo ICEDE

Santiago de Compostela; 18 Outubro, 2023

Resumo rápido

Os estudos universitarios en economía e empresa en Galicia impleméntanse desde o ano 1967

Grande esforzo na formación dunha forza de traballo (cadros de persoal) de PDI

Empregamos *Python* para data scrapping de bases de datos de teses doutorais defendidas no SUG e *Stata* para o dataclean

Usamos en *R* o paquete *ggenealogy* para estudar as árbores xenealoxicas das teses doutorais en economía e empresa presentadas no SUG desde 1976

Simple, ampliable a outras áreas e con posibilidades na difusión de resultados, pero con eivas por mellorar

Contidos

- A creación dos estudos de economía en Galicia
- Obtención dos datos
- Dataclean
- Interpretación e Representación dos datos
- Difusión dos resultados
- Discusión e Conclusións

A creación dos estudos de economía en Galicia

En resumo, (Guisán Seijas, 2018; Suárez-Llanos Gómez, 2008)

- Primeira promoción da licenciatura en ciencias económicas no curso 1967/68 na Universidade de Santiago de Compostela (USC)
- Conversión das antigas Escolas de Comercio en Escolas Universitarias e posterior segregación da Universidade da Coruña (UDC) e da Universidade de Vigo (UdeVigo) desde os 90
- Incremento constante dos cadros de persoal, con importante papel da USC como formadora dos doutores que se integrarían na UDC e UVigo

Obtención dos datos (I)

- **TESEO** como repositorio que contén información sobre as teses doutorais defendidas en universidades españolas desde 1976
- interfaz de consulta moi limitada...
- ... necesidade de consultar múltiples rexistros
- *Python con BeautifulSoup* como alternativa
- Descarga masiva vía *id* de cada entrada desde 1 ata 2.200.000
- Precaución contra as restriccións temporais do IP

Obtención dos datos (II)

- Alternativa de Dialnet
- ... toma os datos do propio TESEO
- ... máis estricta á hora de prohibir consultas masivas
- traballos académicos como (Guisán Seijas, 2018) para complementar á información con teses non rexistradas en TESEO

Dataclean (I)

Entre os datos incluídos/descargados do TESEO:

- Autor
- Título tese
- Data defensa
- Universidad - Departamento - Programa doutoramento
- Resumo - *Keywords*/descriptores
- Director e codirectores
- Tribunal
- Mención internacional
- Hipervínculo ó repositorio que contén a tese / .pdf da tese

Dataclean (II)

Empregando *Stata*¹:

- Limpamos ás entradas baleiras
- Harmonizamos os datos
- Seleccionamos aquelas defendidas nas tres universidades do SUG
- Seleccionamos aquelas relacionadas coa economía e empresa:
 - Adscritas a departamentos de eco e empresa
 - Adscritas a programas de doutoramento en eco e empresa
 - Con descritores que conteñen referencias a eco e empresa

Ficamos con 806 entradas, correspondentes a 417 teses defendidas na USC, 210 na UdeVigo e 179 na UDC.

- Transformamos os datos ó formato necesario para *ggenealogy*.
Pasamos a ter 1038 entradas (fenómeno co-direccións)

¹Por familiaridade do autor co software. As mesmas tarefas poderían realizarse en *R*.

Interpretación e Representación dos datos (I)

Temos un dataframe onde cada fila/entrada rexistra a relación autor-director da tese, conxuntamente co ano de defensa e universidade

	child	parent	year	universidad	titulo
	Nombre y apellidos del autor de la tesis		Año	Universidad en la que se han cursado los estudios	Título de la tesis
1	ABEAL VAZQUEZ, JOSE PABLO	SANCHEZ SANTOS, JOSE MANUEL	2011	UNIVERSIDAD DE A CORUNA	LA PROPAGACION
2	ABOUSAMRA, ALAA ABDELAZIZ ABDELFAH	RUZO SANMARTIN, EMILIO	2017	UNIVERSIDAD DE SANTIAGO DE COMPOSTELA	INTEGRATED INFO
3	ABOUSAMRA, ALAA ABDELAZIZ ABDELFAH	VARELA GONZALEZ, JOSE ANTONIO	2017	UNIVERSIDAD DE SANTIAGO DE COMPOSTELA	INTEGRATED INFO
4	AGBOR TABOT NTOUNG, LIOUS	CIBRAN FERRAZ, PILAR	2015	UNIVERSIDAD DE VIGO	THE INTERACTION
5	AGBOR TABOT NTOUNG, LIOUS	PISON FERNANDEZ, IRENE	2015	UNIVERSIDAD DE VIGO	THE INTERACTION
6	AGUAYO LORENZO, EVA	GUISAN SEJAS, MARIA DEL CARMEN	2001	UNIVERSIDAD DE SANTIAGO DE COMPOSTELA	EMPLEO REGIONA
7	AGUIAR MARAGOTO FERNANDO J.	DOLDAN TIE, FELIX RAMON	1994	UNIVERSIDAD DE A CORUNA	ANALISIS TEORIC

Figura: Exemplo do dataset final empregado con *ggenealogy*

Alén de cálculos sobre a complexidade destas redes, centrémonos nas capacidades de visualización (Rutter et al., 2019)

Interpretación e Representación dos datos (II)

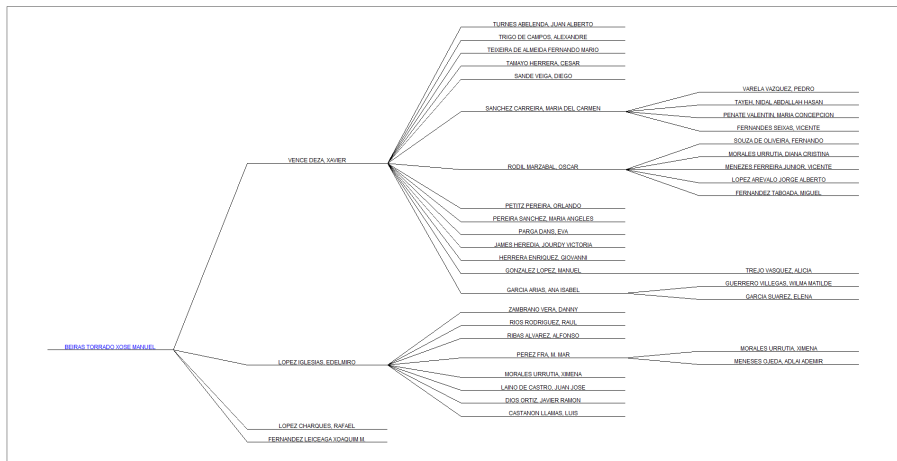


Figura: Doutorandos identificados coa dirección de Xosé Manuel Beiras Torrado

Interpretación e Representación dos datos (III)

Limitacións identificadas:

- As figuras producidas polo paquete *ggenealogy* son demasiado simples e non ofrecen opcións básicas de personalización:
 - Ano de defensa da tese
 - Universidade de defensa
 - etc.
- O incremento das co-direccións leva a que algúns directores-autores sexan considerados tamén siblings/irmáns na mesma xeración na medida en que comparten un proxenitor/director de tese

Interpretación e Representación dos datos (IV)

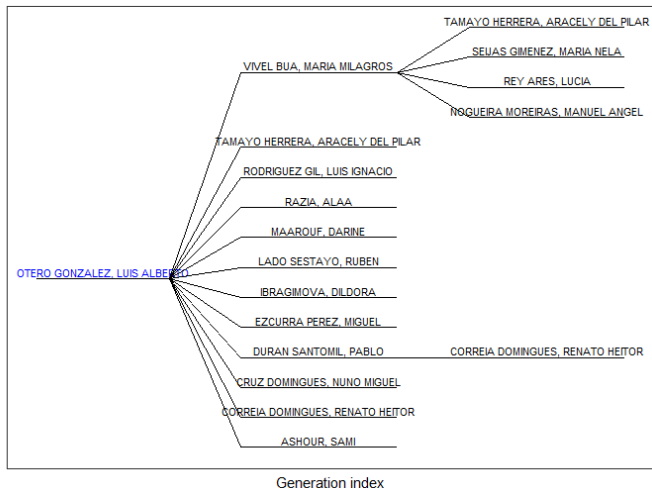


Figura: Problemática das co-direccións na representación gráfica

Difusión dos resultados

- O paquete *ggenealogy* inclúe integración con *Shiny*. Exemplo
- Podemos dispoñibilizar unha app online de consulta das relacións contidas na base do TESEO
 - Grande avance comparado coa limitada web actual do TESEO
 - Posibilidade de visualizar gráficamente as relacións sen ningún coñecemento técnico
 - Aínda por facer...

Discusión e Conclusións

- O estudo sobre a evolución da ciencia en Galicia está por desenvolver
- O exemplo coas teses no campo da economía e empresa e facilmente extensíbel a outras disciplinas científicas e ámbitos xeográficos, a partir dos datos xa dispoñibles
- *R* e o paquete *ggenealogy* ofrecen posibilidades para esta cuestión, especialmente no relativo a visualización e difusión deste estudo entre o público xeral
- Fronte a facilidade de uso, deben mellorarse e ampliarse as opcións de representación gráfica e reflexionar sobre o fenómeno das co-direccións e a súa repercusión na lóxica dunha árbore xenealóxica

Grazas pola atención

Contacto para dúbidas ou suxestións:

jose.blanco2@usc.es

Agradecementos

O traballo foi posible grazas a colaboración das seguintes persoas:

Na parte técnica (*programación en Python*):

- Felipe Vieira de Moraes Tavares

Na parte teórica (evolución dos estudos de economía en Galicia):

- Andrés González Rodríguez
- Fernando de la Torre Cuevas
- Helena Martínez Cabrera

Referencias I

- Rutter, L., VanderPlas, S., Cook, D., & Graham, M. A. (2019). Ggenealogy: An R Package for Visualizing Genealogical Data. *Journal of Statistical Software*, 89(13), 1–31. <https://doi.org/10.18637/jss.v089.i13>
- Guisán Seijas, M. d. C. (2018). 50 Anos de investigación económica en Galicia, 1967-2017. *Revista Galega de Economía*, 27(3), 143–166. <https://doi.org/10.15304/rge.27.3.5627>
- Suárez-Llanos Gómez, L. (2008). Testemuña duns comezos: actos conmemorativos do corenta aniversario da Facultade de Ciencias Económicas e Empresariais da USC. *Revista galega de economía: Publicación Interdisciplinar da Facultade de Ciencias Económicas e Empresariais*, 17(2), 7–16. Obtido el 6 de agosto de 2023, desde <https://dialnet.unirioja.es/servlet/articulo?codigo=2767597>