

Gráficos Interactivos con Redes para el Análisis de Encuestas de Opinión.

M. Escobar y P. Cabrera (modesto@usal.es & pablocal@usal.es)

Universidad de Salamanca

XIV Congreso de la AECPA

Salamanca, 10 de julio de 2019



Contenido

- 1 **Introducción**
 - Grafos
 - Los grafos en la investigación social
- 2 CARG
 - Datos
 - Coincidencias
 - Grados
 - Regresiones
- 3 Programas
- 4 Propuestas



Grafo

Definición

- Un grafo \mathcal{G} es una colección de nodos o vértices x_1, x_2, \dots, x_n (expresados en el conjunto \mathcal{N}), y una colección de líneas o aristas l_1, l_2, \dots, l_m (expresadas en el conjunto \mathcal{L}) que vinculan a todos o a algunos de esos nodos. El grafo \mathcal{G} es, por tanto, se expresa simbólicamente mediante la pareja de conjuntos $(\mathcal{N}, \mathcal{L})$.
(Christofides, 1975)



Grafo

Operacionalización

- Esta pareja de conjuntos $(\mathcal{N}, \mathcal{L})$ se puede representar solo con una $n \times n$ matriz \mathbf{M} cuyos elementos m_{jk} representan la (fuerza de la) conexión de cada punto x_j con cualquier otro x_k .
- Hay dos otros dos modos de guardar los grafos:
 - Una lista de adyacencias en la que cada fila n es un nodo y las columnas son todos y cada uno de los correspondientes nodos vecinos
 - Una lista de vínculos en la que cada fila representa una de las m aristas de las que se compone el grafo (Mihura, 2011:7)



Representación de grafos

Ejemplos de disposición

Matriz de adyacencias

	x_1	x_2	x_3	x_4
	A	B	C	D
x_1 A	—			
x_2 B	0	—		
x_3 C	2	2	—	
x_4 D	1	1	1	—

Lista de adyacencias

Nodos	Vecinos
A	C C D
B	C C D
C	D

Lista de vínculos

Origen	Destino
A	C
C	A
B	C
C	B
C	D
A	D
B	D



Grafos en la investigación social

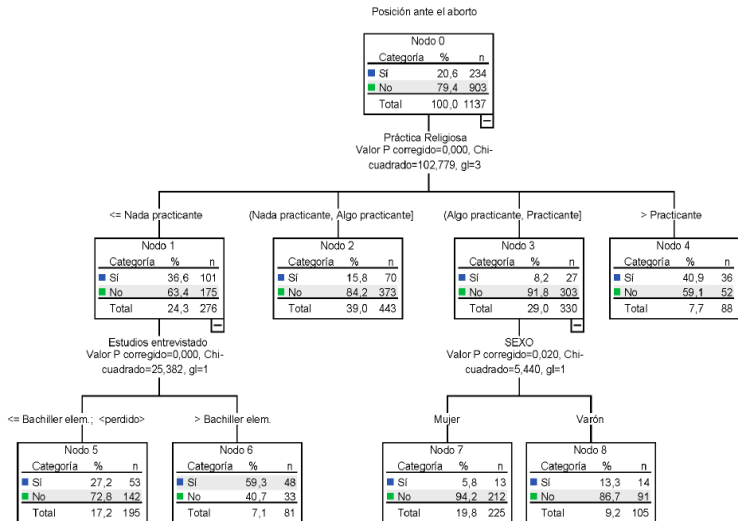
Clasificación

- Grafos (Análisis de redes sociales)
- Árboles (Árboles de clasificación y regresión)
- Senderos (Análisis de senderos)
- Modelos causales y de medida (Modelos de ecuaciones estructurales)
- Grafos reticulares de clasificación y regresión. (NCARG)



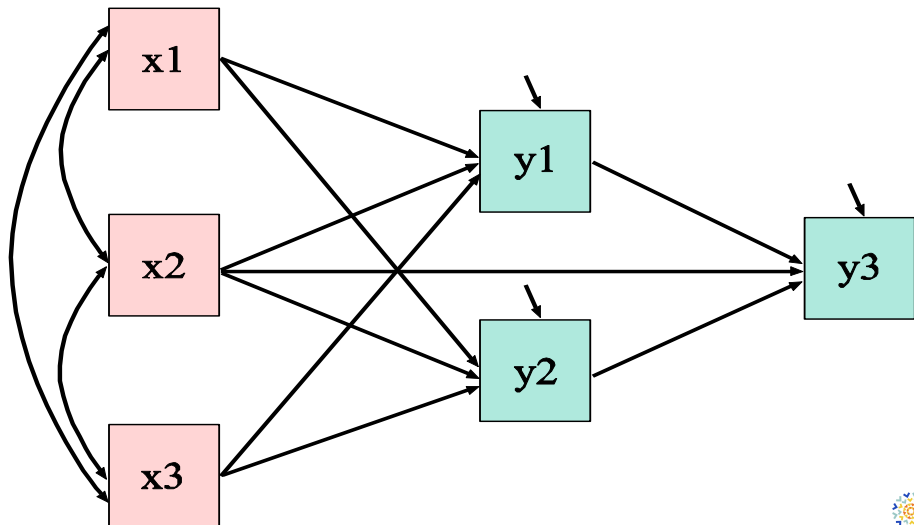
CART

Necesidad percibida de inmigrantes (Fuente: CIS-2511)



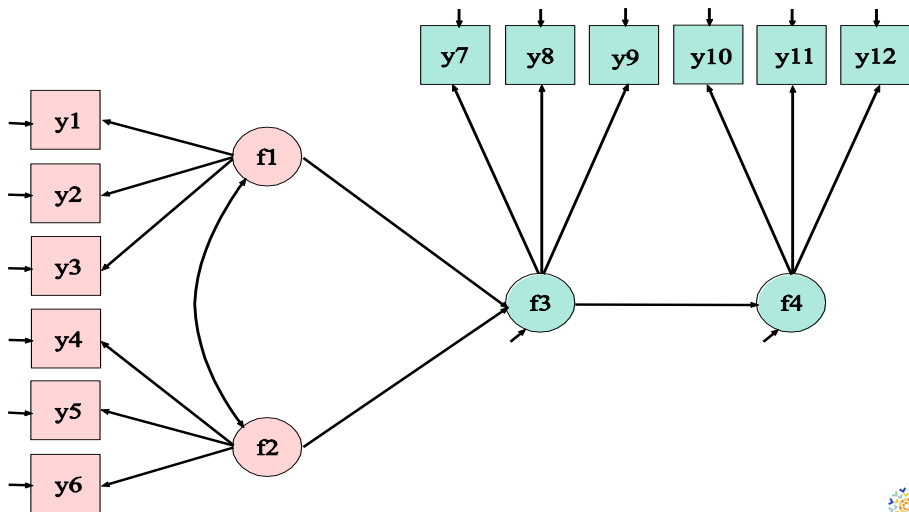
Análisis de senderos

Análisis de senderos con variables continuas (Muthén & Muthén 2017,32)



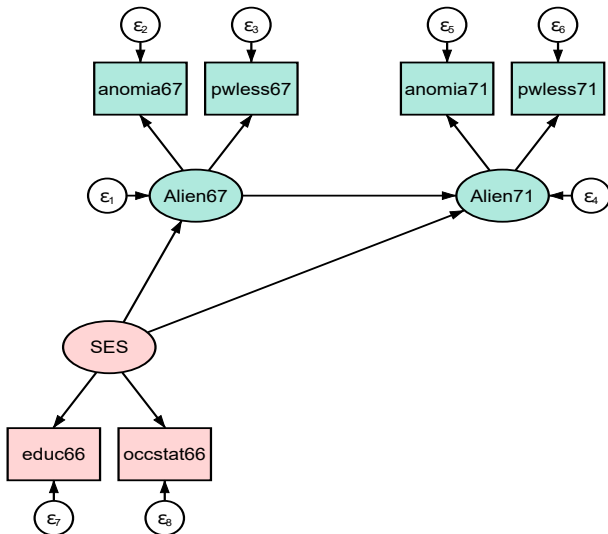
SEM

SEM causal con variables continuas observadas y latentes (Muthén & Muthén 2017,76)



SEM

Ejemplo clásico de la alienación (Wheaton et al. 1977,84-136)



Contenido

- 1 Introducción
 - Grafos
 - Los grafos en la investigación social
- 2 CARG
 - Datos
 - Coincidencias
 - Grados
 - Regresiones
- 3 Programas
- 4 Propuestas



Datos

Fuente: CIS-3.244 (2019)

- Vamos a ver los grafos reticulares de coincidencias y regresión con datos de un superbarómetro del CIS.
- Por simplicidad se usarán solo 3 variables: voto, ideología y religión.
- El voto se considerará una variable endógena, mientras que ideología y religión será tratadas como exógenas.



Distribución de frecuencias

Voto, ideología y religión (sin ponderar)

Tabla: Distribución de frecuencias

Categoría	%	Categoría	%
Voto:VOX	4.42	Ideología:1-3	26.09
Voto:PP	10.65	Ideología:4	14.05
Voto:Ciudadanos	7.84	Ideología:5	20.33
Voto:PSOE	21.20	Ideología:6	9.38
Voto:Podemos	7.08	Ideología:7	6.99
Voto:Otro	10.05	Ideología:8-10	9.46
Voto:Indeciso	29.58	Ideología:NC	4.35
Voto:No votará	9.18	Ideología:NS	9.35
		Religión:Católico/a no practicante	48.62
		Religión:Católico/a practicante	22.74
		Religión:No creyente	24.41



Análisis de coincidencias

Definición

El análisis de coincidencias es un conjunto de técnicas cuyo objeto consiste en detectar y representar qué sucesos, objetos o sujetos tienden a aparecer al mismo tiempo en unos espacios delimitados.

- Estos N espacios delimitados (i) se denominan escenarios y pueden considerarse unidades de análisis (registros).
- En cada uno de estos escenarios (campos) un conjunto de J sucesos (x_{ij}) pueden estar presentes (1) o ausentes (0).
- Un conjunto de escenarios forman una matriz binaria de incidencias (\mathbf{X}) con dimensiones ($N \times J$).
- Estos escenarios pueden agruparse en H subconjuntos para poderlos comparar.



7 grados de coincidencias

Clasificación

Las coincidencias entre sucesos pueden graduarse del modo siguiente:

- Sin coincidencia (sucesos mutuamente excluyentes)
- Mera coincidencia (al menos coinciden en un escenario)
- Probable ($p(\mathbf{x}_j|\mathbf{x}_k) > 0.5$)
- Estadísticamente probable ($P(p(\mathbf{x}_j|\mathbf{x}_k) \leq 0.5) < c$)
- Condicional ($p(\mathbf{x}_j) < p(\mathbf{x}_j|\mathbf{x}_k)$)
- Estadísticamente condicional ($P(p(\mathbf{x}_j) - p(\mathbf{x}_j|\mathbf{x}_k) \leq 0) < c$)
- Total (siempre ocurren en los mismos escenarios)



Características de los grados de coincidencias

Cuadro

Tabla: Grados de coincidencias

Grado	Definición	Asimétrica	Prueba
Nula	$f_{jk} = 0$	No	No
Simple	$f_{jk} > 0$	No	No
Probable	$f_{jk} / f_{kk} > 0,5$	Sí	Sí
Dependiente	$f_{jk} > f_{jk}^*$	No	Sí
Subtotal	$f_{jk} = f_{jj} < f_{kk}$	Sí	No
Total	$f_{jk} = f_{jj} = f_{kk}$	No	No

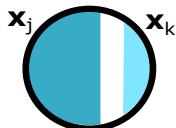


Representación de las coincidencias

Nula, mera, probable, subtotal, total

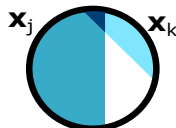
Coincidencias

Nula



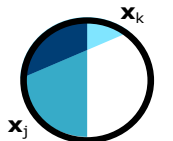
$$f_{jk}=0$$

Mera



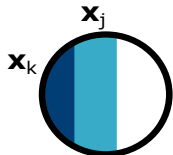
$$f_{jk}>0$$

Probable



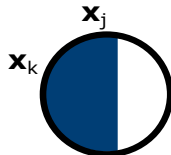
$$f_{jk}/f_j < 0.5 < f_{jk}/f_k$$

Subtotal



$$f_{jk}=f_k$$

Total



$$f_{jk}=f_j=f_k$$

Leyenda

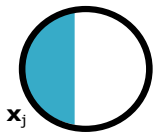


Representación de las coincidencias

Independencia y coincidencia condicional

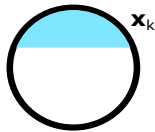
Coincidencias condicionales

Marginal de \mathbf{x}_j



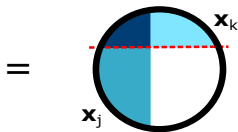
$$f_{jj} < 50\%$$

Marginal de \mathbf{x}_k



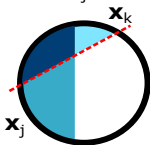
$$f_{kk} < 50\%$$

Independencia de \mathbf{x}_j y \mathbf{x}_k



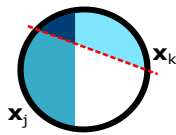
$$f_{jk} = f_{jj} f_{kk} / n$$

Coincidencia condicional
de \mathbf{x}_j y \mathbf{x}_k



$$f_{jk} > f_{jj} f_{kk} / n$$

No coincidencia condicional
de \mathbf{x}_j y \mathbf{x}_k



$$f_{jk} < f_{jj} f_{kk} / n$$

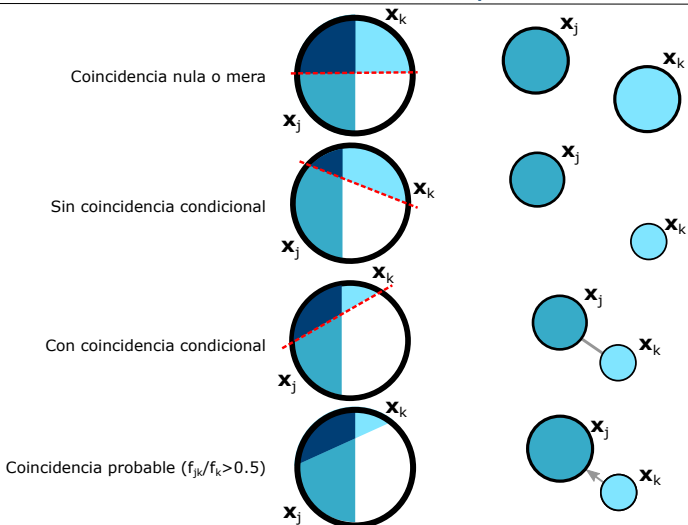
Leyenda



De las coincidencias a las adyacencias

Ejemplos de distintos grados de coincidencias

De las coincidencias a las adyacencias



Coincidencias entre voto e ideología

Residuos corregidos (r_{jk})

Tabla: Residuos corregidos entre voto e ideología (r_{jk})

Voto	Id:1-3	Id:4	Id:5	Id:6	Id:7	Id:8-10	Id:NC	Id:NS
VOX	-14.47	-8.99	-4.38	3.52	7.86	34.05	-3.21	-3.79
PP	-23.90	-15.10	-12.60	11.11	28.66	42.02	-6.02	-2.76
Ciudadanos	-17.43	-8.89	14.12	20.71	8.91	-1.06	-3.97	-7.30
PSOE	23.41	30.50	-5.48	-15.84	-15.61	-19.00	-9.54	-8.84
Podemos	37.55	0.16	-11.67	-9.26	-8.44	-10.33	-6.16	-9.35
Otro	15.65	-1.04	-0.72	-3.38	-6.99	-9.53	-0.72	-1.80
Indeciso	-12.27	-2.89	13.05	3.72	-0.63	-9.36	4.16	7.23
No votará	-11.35	-6.87	1.66	-5.37	-6.64	-7.36	25.56	23.72



Coincidencias entre voto y religión

Residuos corregidos (r_{jk})

Tabla: Residuos corregidos entre voto y religión (r_{jk})

Voto	R:Practicante	R:No practicante	R:No creyente
VOX	3.56	3.62	-6.71
PP	29.94	-6.96	-18.94
Ciudadanos	-2.60	7.24	-4.32
PSOE	-6.76	7.63	-2.83
Podemos	-13.39	-13.31	28.23
Otro	-8.59	-5.60	15.13
Indeciso	3.23	3.17	-7.00
No votará	-6.79	-0.05	3.29



Adjacencies among action ways to resolve social problems

($\phi \geq .25$)

Tabla: Coincidencias Voto con ideología y religión ($r_{ij} \geq 3$)

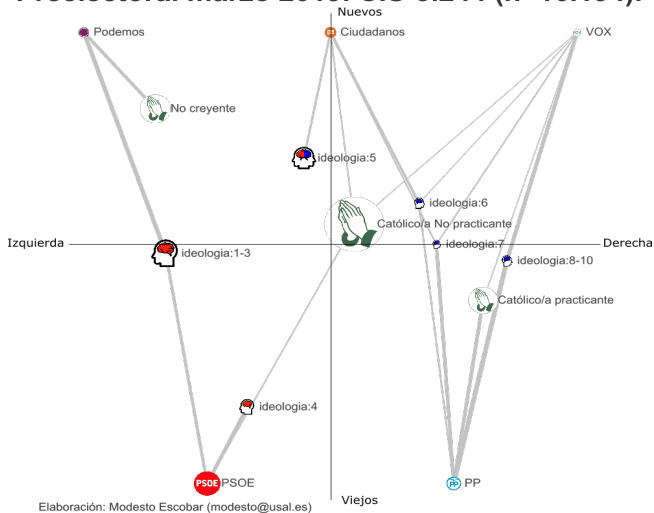
Voto	I:<4	I:4	I:5	I:6	I:7	I:>7	I:NC	I:NS	R:CP	R:NP	R:NC
VOX	0	0	0	1	1	1	0	0	1	1	0
PP	0	0	0	1	1	1	0	0	1	0	0
Ciudadanos	0	0	1	1	1	0	0	0	0	1	0
PSOE	1	1	0	0	0	0	0	0	0	1	0
Podemos	1	0	0	0	0	0	0	0	0	0	1
Otro	1	0	0	0	0	0	0	0	0	0	1
Indeciso	0	0	1	1	0	0	1	1	1	1	0
No votará	0	0	0	0	0	0	1	1	0	0	1



Ilustración de coincidencias

Coincidencias condicionales significativas ($p \leq 0.05$)

Preelectoral marzo 2019. CIS-3.244 (n=16.194).



Medidas de coincidencia

Haberman (residuo corregido), ϕ y probabilidad

Tabla: Medidas de coincidencia: r_{ij} , ϕ y significación

Origen	Destino	r_{ij}	ϕ	$p(Z)$
Ideo:8-10	PP	42.02	0.36	0.00
Ideo:1-3	Podemos	37.55	0.32	0.00
Ideo:1-3	<i>No creyente</i>	35.79	0.30	0.00
Ideo:8-10	VOX	34.05	0.29	0.00
Ideología:4	PSOE	30.50	0.26	0.00
Católico/a practicante	PP	29.94	0.25	0.00
Ideología:7	PP	28.66	0.24	0.00
No creyente	Podemos	28.23	0.24	0.00
Ideo:NC	No votará	25.56	0.22	0.00
Ideo:8-10	<i>Católico/a practicante</i>	24.09	0.20	0.00
Ideología:NS	No votará	23.72	0.20	0.00



De las coincidencias a las regresiones

Con el mismo ejemplo

- Ahora se distinguirá claramente entre variables dependientes e independientes
- Como, en este ejemplo, solo se está trabajando con variables cualitativas, las categorías endógenas se convierten variables dicotómicas para ser incluidas en regresiones logísticas como variables dependientes.
- Tantas como categorías dependientes consideremos de relevancia. En este caso, solo se emplean los cinco partidos más votados.
- Como variables independientes pueden introducirse variables cuantitativas o cualitativas (éstas como indicadores dicotómicos)
 - Religión
 - Ideología
- De este modo, se trabajaría con cinco regresiones logísticas, cada una de ellas con dos variables independientes de 4 y 8 categorías respectivamente.



Regresión logística del voto al PP

Sobre religión e ideología

Tabla: Regresión del voto al PP sobre religión e ideología

	Estimador	Error típ.	valor z	Pr(> z)
(Constante)	-5.4423	0.4123	-13.20	0.0000
Católico/a no practicante	-0.8156	0.0670	-12.17	0.0000
No creyente	-1.5612	0.1392	-11.21	0.0000
Otra religión	-1.2133	0.2091	-5.80	0.0000
Ideología: 4	1.5425	0.4728	3.26	0.0011
Ideología: 5	2.9734	0.4202	7.08	0.0000
Ideología: 6	4.6218	0.4157	11.12	0.0000
Ideología: 7	5.4992	0.4152	13.25	0.0000
Ideología: >7	5.6725	0.4138	13.71	0.0000
Ideología: NS	3.6774	0.4218	8.72	0.0000
Ideología: NC	2.7713	0.4688	5.91	0.0000



Regresión logística del voto al PSOE

Sobre religión e ideología

Tabla: Regresión del voto al PSOE sobre religión e ideología

	Estimador	Error típ.	valor z	Pr(> z)
(Constante)	-0.1759	0.0658	-2.67	0.0075
Católico/a no practicante	-0.0989	0.0634	-1.56	0.1186
No creyente	-0.9307	0.0739	-12.60	0.0000
Otra respuesta	-0.2015	0.1204	-1.67	0.0942
Ideología: 4	0.3751	0.0586	6.40	0.0000
Ideología: 5	-1.1726	0.0632	-18.55	0.0000
Ideología: 6	-2.8167	0.1440	-19.56	0.0000
Ideología: 7	-3.8717	0.2639	-14.67	0.0000
Ideología: > 7	-4.4367	0.2937	-15.11	0.0000
Ideología: NS	-1.6785	0.0952	-17.62	0.0000
Ideología: NC	-2.4032	0.1788	-13.44	0.0000



Problemas con las regresiones

- Los contrastes son una herramienta flexible para comprobar los efectos de los predictores nominales en los análisis de varianza y regresión.
- En las regresiones habituales los contrastes presentan estas características
 - Se contraste en relación con otra categoría.
 - Generalmente se realizan con la primera categoría (la que arbitrariamente presenta el valor menor).
- Sin embargo, hay otro tipo de contrastes
 - Contrastes adyacentes (invertidos o no).
 - Diferencias en relación con la media de niveles previos o posteriores.
 - **Diferencias en relación con la media global (ponderada o sin ponderar)**
 - Contrates polinómicos ortogonales.



Contraste de diferencias con la media global ponderada

Matriz de contrastes

- Cada nivel de una variable cualitativa se compara con la media ponderada de todos los niveles.
- En consecuencia, en el caso de una variable con cuatro categorías, la matriz de contrastes sería

$$\mathbf{C} = \begin{pmatrix} 1 - f_j/f. & -f_k/f. & -f_k/f. & -f_k/f. \\ -f_k/f. & 1 - f_j/f. & -f_k/f. & -f_k/f. \\ -f_k/f. & -f_k/f. & 1 - f_j/f. & -f_k/f. \\ -f_k/f. & -f_k/f. & -f_k/f. & 1 - f_j/f. \end{pmatrix}$$

- La función glm de regresión ha de recibir el siguiente argumento:

$$\mathbf{c} = \mathbf{C}^{-1}[\cdot, -1]$$



Regresión alternativa del voto al PP (1^{er} paso)

Cambiar la base de los predictores nominales: menor coeficiente

Tabla: Regresión alternativa del voto al PP -cambio de base-

	Estimador	Error típ.	valor z	Pr(> z)
(Constante)	-7.0036	0.4209	-16.64	0.0000
Católico/a practicante	1.5612	0.1392	11.21	0.0000
Católico/a no practicante	0.7456	0.1387	5.37	0.0000
Otra respuesta	0.3479	0.2417	1.44	0.1500
Ideología: 4	1.5425	0.4728	3.26	0.0011
Ideología: 5	2.9734	0.4202	7.08	0.0000
Ideología: 6	4.6218	0.4157	11.12	0.0000
Ideología: 7	5.4992	0.4152	13.25	0.0000
Ideología: 8-10	5.6725	0.4138	13.71	0.0000
Ideología: NS	3.6774	0.4218	8.72	0.0000
Ideología: NC	2.7713	0.4688	5.91	0.0000



Regresión alternativa del voto al PP (2° paso)

Aplicar el contraste de la media ponderada

Tabla: Regresión alternativa del voto al PP -cambio de contraste-

	Estimador	Error típ.	valor z	Pr(> z)
(Constante)	-2.9951	0.0857	-34.94	0.0000
Católico/a practicante	0.8290	0.0547	15.17	0.0000
Católico/a no practicante	0.0134	0.0423	0.32	0.7510
Otra respuesta	-0.3843	0.1991	-1.93	0.0537
Ideología: 4	-1.0975	0.2314	-4.74	0.0000
Ideología: 5	0.3334	0.1359	2.45	0.0142
Ideología: 6	1.9818	0.1315	15.07	0.0000
Ideología: 7	2.8592	0.1310	21.82	0.0000
Ideología: 8-10	3.0325	0.1262	24.02	0.0000
Ideología: NS	1.0374	0.1465	7.08	0.0000
Ideología: NC	0.1313	0.2467	0.53	0.5945



Regresión alternativa del voto al PP (3^{er} paso)

Eliminar las categorías con coeficiente negativo

Tabla: Regresión alternativa del voto al PP -sin estimadores no contribuyentes-

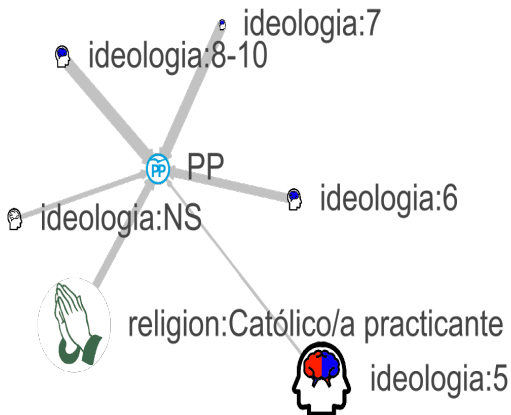
	Estimador	Error típ.	valor z	Pr(> z)
(Constante)	-2.9951	0.0857	-34.94	0.0000
Católico/a practicante	0.8290	0.0547	15.17	0.0000
Ideología: 5	0.3334	0.1359	2.45	0.0142
Ideología: 6	1.9818	0.1315	15.07	0.0000
Ideología: 7	2.8592	0.1310	21.82	0.0000
Ideología: 8-10	3.0325	0.1262	24.02	0.0000
Ideología: NS	1.0374	0.1465	7.08	0.0000



Grafo de regresión

Regresión del voto al PP

Intención de voto en Elecciones Generales España (Barómetro del CIS de enero-2019)



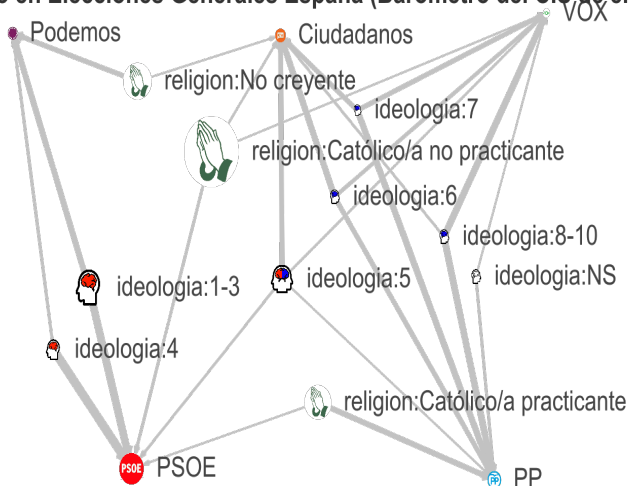
Fuente: CIS (3226) . Analisis: Modesto Escobar. Gráfico elaborado con netCoin. Escobar, M. et al. (2019). netCoin: Interactive analytical networks. Retrieved from CRAN.R packages.



De una a varias regresiones

PP, PSOE, Ciudadanos, U. Podemos, VOX

Intención de voto en Elecciones Generales España (Barómetro del CIS de enero-2019)



Fuente: CIS (3226) . Analisis: Modesto Escobar. Gráfico elaborado con netCoin. Escobar, M. et al. (2019). netCoin: Interactive analytical networks. Retrieved from CRAN.R packages.



Contenido

- 1 Introducción
 - Grafos
 - Los grafos en la investigación social
- 2 CARG
 - Datos
 - Coincidencias
 - Grados
 - Regresiones
- 3 Programas
- 4 Propuestas



¿Cómo hacer estos gráficos

netCoin

- netCoin es un paquete gratuito para R.
 - netCoin calcula
 - Una matriz de coincidencias a partir de una matriz de incidencias (coin).
 - Matrices de semejanzas (distancias) a partir de una matriz de coincidencias (sim).
 - netCoin representa gráficamente
 - Desde una lista de vínculos (netCoin)
 - Desde un conjunto de variables cuantitativas (netCorr)
 - Desde un conjunto de variables dicotómicas (allNet)
 - Desde un conjunto de datos en SPSS o Stata (surCoin and glmCoin)
 - Desde un objeto lavaan (pathCoin)
 - netCoin convierte
 - Una matriz de incidencias/coincidencias/semejanzas/distancias en una lista de vínculos.
 - Un gráfico netCoin en otro igraph o Pajek.



Contenido

- 1 Introducción
 - Grafos
 - Los grafos en la investigación social
- 2 CARG
 - Datos
 - Coincidencias
 - Grados
 - Regresiones
- 3 Programas
- 4 Propuestas



Algunas propuestas (A)

- En las ciencias sociales deberíamos asumir un papel importante en el mundo emergente del big data.
- Hemos de ser conscientes del enorme potencial de los grandes conjuntos de datos construidos mediante las interacciones sociales (datos transaccionales, tuits, ...)
- Podemos contribuir aportando métodos robustos y visuales en el análisis de datos e interpretaciones ilustrativas.



Algunas propuestas (B)

- Proponemos los grafos reticulares de coincidencias y regresión como un marco analítico visual.
 - Representar los nodos por su importancia (frecuencia) en lugar de por su grado.
 - Representar los enlaces en función de su significación en lugar de por sus frecuencias.
 - Expresar solo asociaciones positivas en relación con las medias ajustadas ponderadas.



Última diapositiva

Agradecimiento

Muchas gracias.
modesto@usal.es



Bibliografía

Análisis reticular de coincidencias

Bibliografía sobre análisis reticular de coincidencias

- Escobar, M. (2009). Redes semánticas en textos periodísticos: propuestas técnicas para su representación. *Empiria*, 17, 13-39.
- Escobar, M., y Gómez Isla, J. (2015). "La expresión de la identidad a través de la imagen: los archivos fotográficos de Miguel de Unamuno y Joaquín Turina". *Revista Española de Investigaciones Sociológicas*, 152, 23-46.
- Escobar, M. (2015). "Studying Coincidences with Network Analysis and Other Multivariate Tools". *The Stata Journal*, 15(4), 1118-1156.
- Escobar, M. (2016). "Ensayo sobre las coincidencias". En A. Almarcha, P. González, y L. Román (Eds.), *Donde la Sociología te lleve*. A Coruña: Universidad de A Coruña.
- Escobar, M., y C. Tejero (2018). "El análisis reticular de coincidencias". *Empiria*, 39, 129-148.
- Escobar, M. y L. Martínez (under review) "Network Coincidence Analysis: the netCoin R Package".



netCoin

Otros ejemplos interactivos

- Bases de datos
 - M. de Unamuno's album
 - British Library
- Encuestas
 - CIS
 - Series de TV
 - Otras iniciativas
- Coincidencias históricas
 - Historia de la ciencia
 - History de la cultura
- Análisis de contenido
 - TST
 - Blogs de mujeres
- Modelos de ecuaciones estructurales
 - Innovacion empresarial (CIE)

