

PROGRAMA DE DOCTORADO EN INGENIERÍA INFORMÁTICA

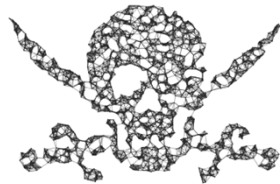
20 de noviembre de 2023





El problema de la ubicación de los comercios desde la perspectiva de las redes complejas

José Manuel Galán Ordax



- Ahedo, V., Santos, J.I., Galán, J.M. (submitted). Network-based quality index aggregation in the retail location problem. A supervised learning approach
- Ahedo, V., Santos, J.I., Galán, J.M. (2023). Combining Quality Indexes in the Retail Location Problem Using Generalized Linear Models. In: García Márquez, F.P., Segovia Ramírez, I., Bernalte Sánchez, P.J., Muñoz del Río, A. (eds) IoT and Data Science in Engineering Management. CIO 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 160, pp. 47-52. Springer, Cham. https://doi.org/10.1007/978-3-031-27915-7_9 (Best paper award)
- Sánchez-Saiz, R.M., Ahedo, V., Santos, J.I., Gómez, S., Galán, J.M. (2022) Identification of robust retailing location patterns with complex network approaches. Complex & Intelligent Systems, vol. 8, pp. 83-106, doi: <https://doi.org/10.1007/s40747-021-00335-8>
- Ahedo, V., Santos, J.I., Galán, J.M. (2021) Knowledge Transfer in Commercial Feature Extraction for the Retail Store Location Problem IEEE Access, vol. 9, pp. 132967-132979, 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3115712>.



VIRGINIA AHEDO



JOSÉ IGNACIO SANTOS



ROSA Mª SÁNCHEZ



SERGIO GÓMEZ





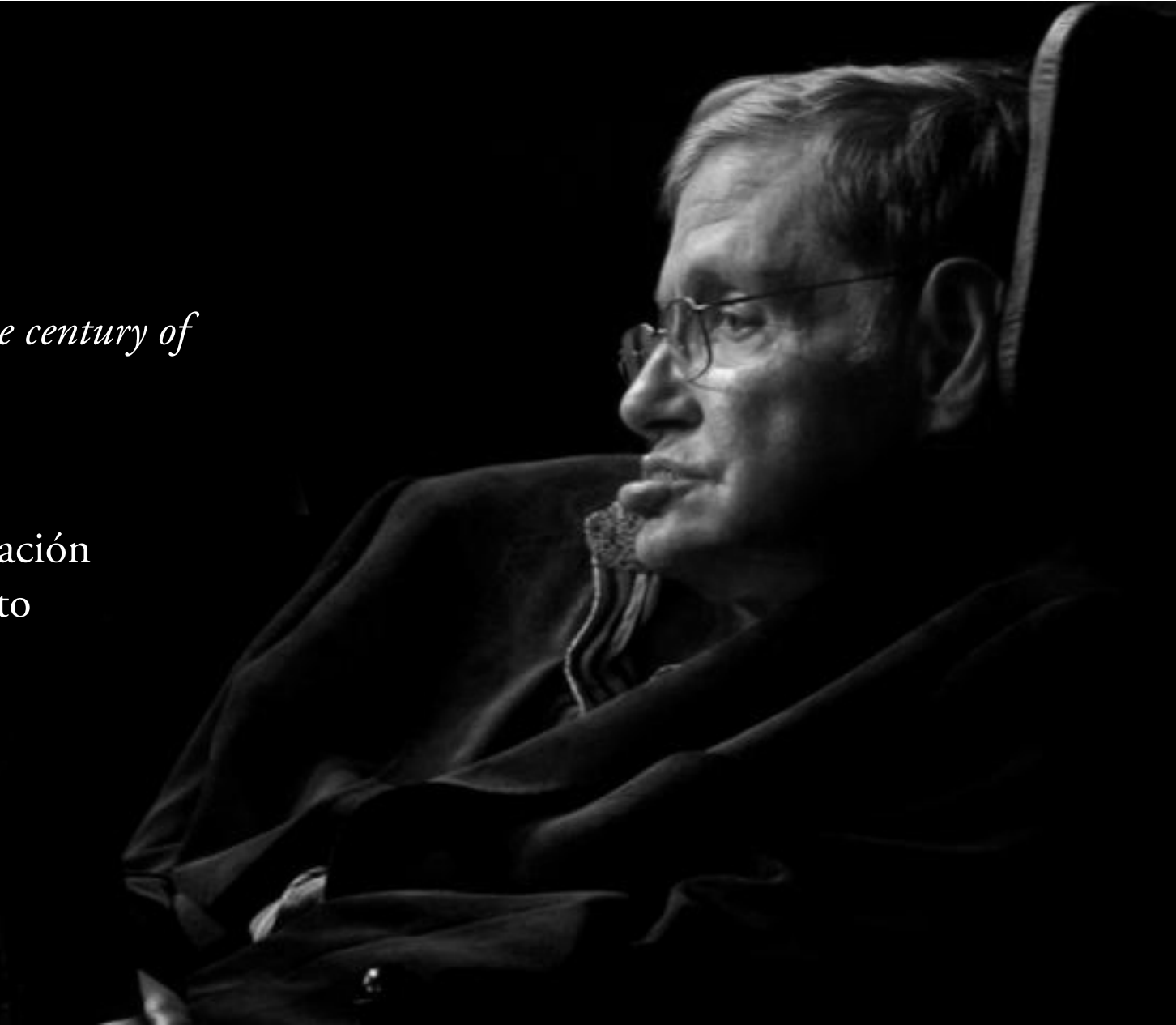


*“Step back and ask, what is the most striking
feature of the geography of economic activity?
The short answer is surely concentration”*

Paul Krugman. Premio Nobel de Economía

“I think the next century will be the century of complexity”

Stephen Hawking . Premio Fundación
BBVA Fronteras del Conocimiento





Objetivo

El objetivo general es proporcionar una visión más completa de las relaciones espaciales y las dinámicas y mecanismos de influencia entre los negocios minoristas en entornos urbanos utilizando el marco de los sistemas complejos



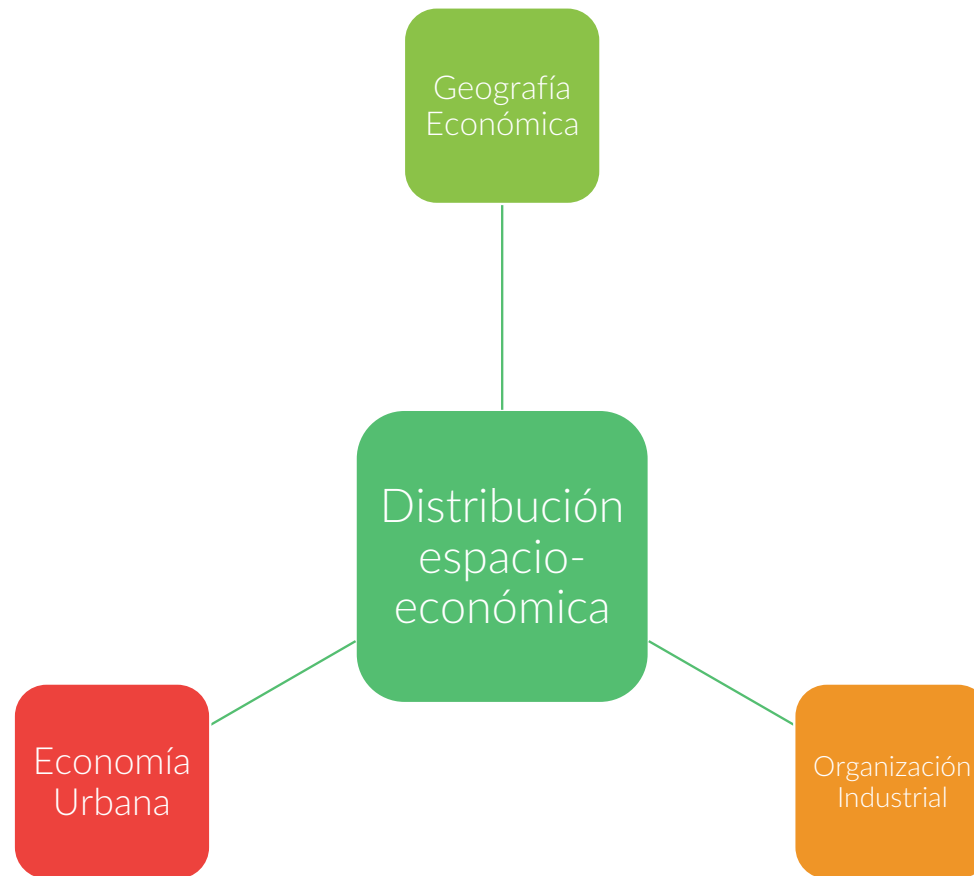
Estructura

- Antecedentes y estado actual
- Metodología
- Caso de estudio I
 - Experimentos
 - Resultados
- Caso de estudio II
 - Experimentos
 - Resultados
- Algunas conclusiones
- Limitaciones



Antecedentes y estado actual

Distribución geográfica de la actividad económica



Antecedentes y estado actual

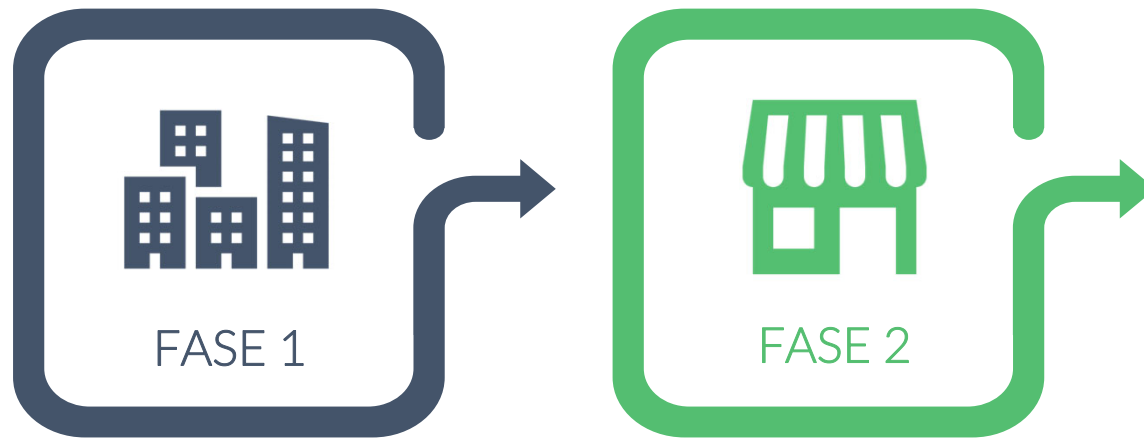
Importancia en la localización minorista

- Aproximadamente el 90% de los ingresos en venta minorista a día de hoy se produce en tiendas (Berman et al, 2018)



Antecedentes y estado actual

Factores en la decisión



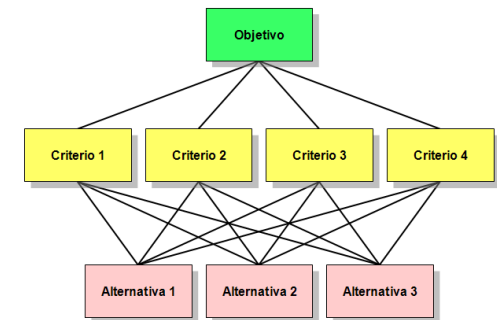
Análisis de área

Localización específica

"The checklists and analogues described in the literature do occur but the more complicated mathematical techniques do not"

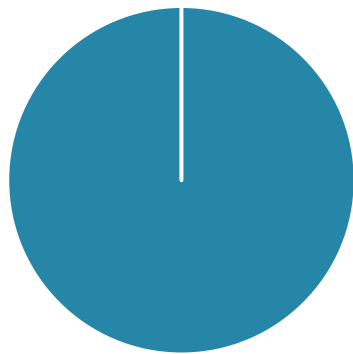
"Note that we used to use a neural network based modelling system but have moved away from it - too 'black boxy' and we never knew why it had come up with the numbers that it had."

(Reynolds & Wood, 2010)

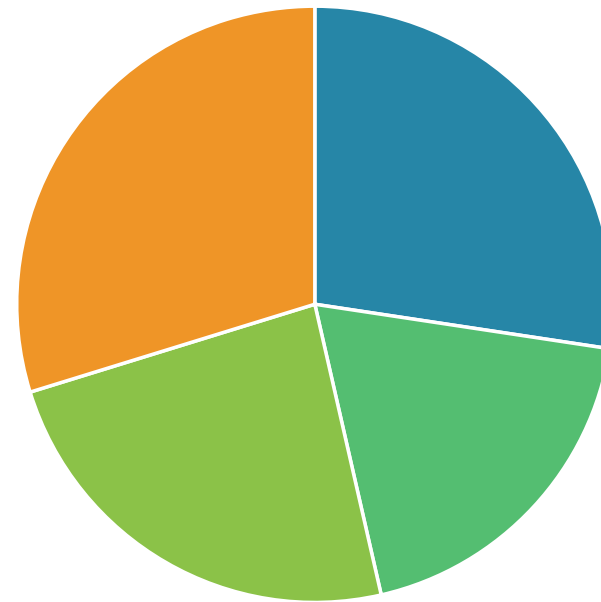


Antecedentes y estado actual

Comportamiento del sector



■ Empresa 1



■ Empresa 1 ■ Empresa 2 ■ Empresa 3 ■ Empresa 4

Aumento de mercado debido la incertidumbre sobre las preferencias y gustos y sobre las expectativas de precios más bajos (Konishi, 2005)

Antecedentes y estado actual

Balanced tenancy

“la estructura y balance de otras tiendas que puedan complementar el ecosistema comercial propio- cada vez se considera un factor más relevante en el proceso de localización” (Berman et al, 2018)

The Principle of Relatedness La probabilidad de que un nuevo comercio se instale en una zona determinada es función del número de actividades comerciales relacionadas presentes en esa zona (Hidalgo et al. 2018).

Proporcionar una visión más completa de las relaciones espaciales y las dinámicas y mecanismos de influencia entre los negocios minoristas en entornos urbanos, teniendo en cuenta para ello los aspectos adaptativos y estratégicos y como las interacciones, tanto positivas como negativas, entre los diferentes negocios conforman los entornos económicos en las ciudades.



Antecedentes y estado actual

Sistemas complejos

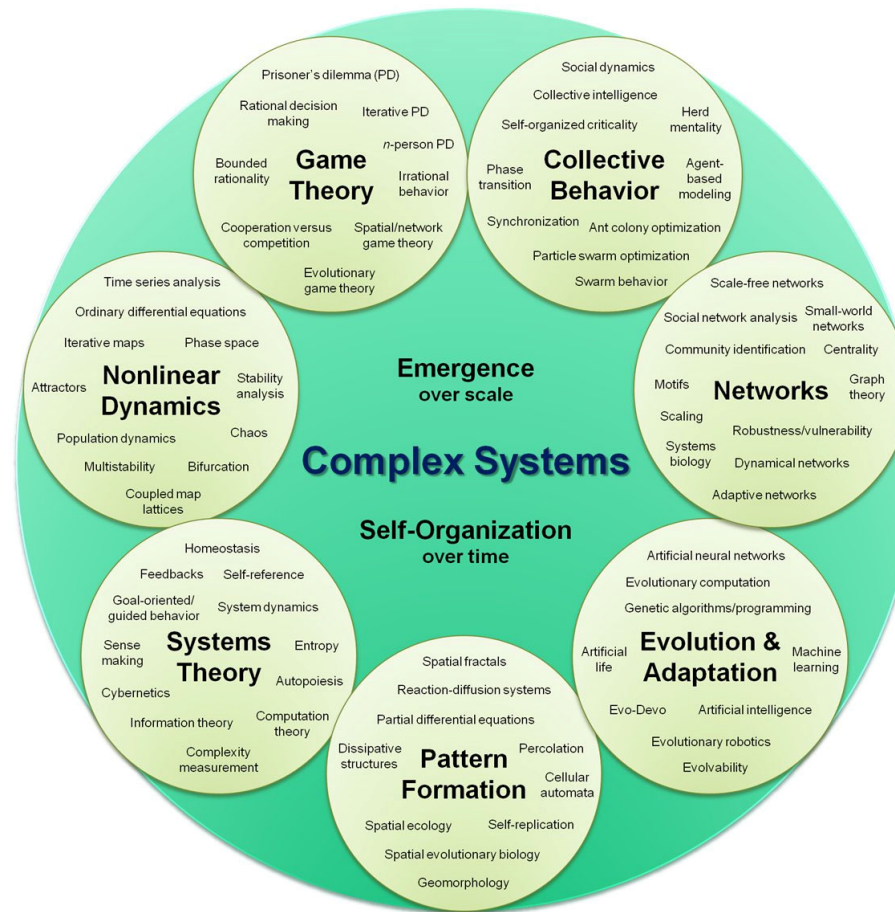


Imagen: https://en.wikipedia.org/wiki/Complex_system

Antecedentes y estado actual

Sistemas complejos



Antecedentes y estado actual

Sistemas complejos



A network-based prediction of retail stores commercial categories and optimal locations

Pablo Jensen

Laboratoire de Physique, CNRS UMR 5672, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

Laboratoire d'Economie des Transports, CNRS UMR 5593, ISH-Université Lyon-2, 14, Av. Berthelot, 69007 Lyon, France

Analyzing the Localization of Retail Stores with Complex Systems Tools

Pablo Jensen

Institut des Systèmes Complexes Rhône-Alpes, IXXI-CNRS,
Laboratoire de Physique, Ecole Normale Supérieure de Lyon and LET-CNRS,
Université Lyon-2, 69007 Lyon, France

Analysis of community structure in networks of correlated data

Sergio Gómez,¹ Pablo Jensen,² and Alex Arenas¹

¹*Departament d'Enginyeria Informàtica i Matemàtiques,
Universitat Rovira i Virgili, 43007 Tarragona, Spain*

²*IXXI - Institut des Systèmes Complexes, 5 rue du Vercors, 69007 Lyon, France*

(Dated: June 18, 2009)

No existe ningún estudio sistemático e integrado
que establezca las relaciones y patrones
existentes entre múltiples sectores y a lo largo de
múltiples ciudades en el problema de localización
en retailing

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

La metodología asume un área urbana con las tiendas minoristas situadas en N_T lugares diferentes

$$T = \{T_i; i = 1, \dots, N_T\}$$

Y divididas en diferentes categorías (clasificación del NAICS)

$$A = \{A_i; i = 1, \dots, N_A\}$$



Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

$$A = \begin{bmatrix} 0 & d_{12} & d_{13} & \dots & d_{1n} \\ d_{21} & 0 & d_{23} & \dots & d_{2n} \\ d_{31} & d_{32} & 0 & \dots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \dots & 0 \end{bmatrix}$$



Redes Jensen

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas



Intra-attraction coefficient

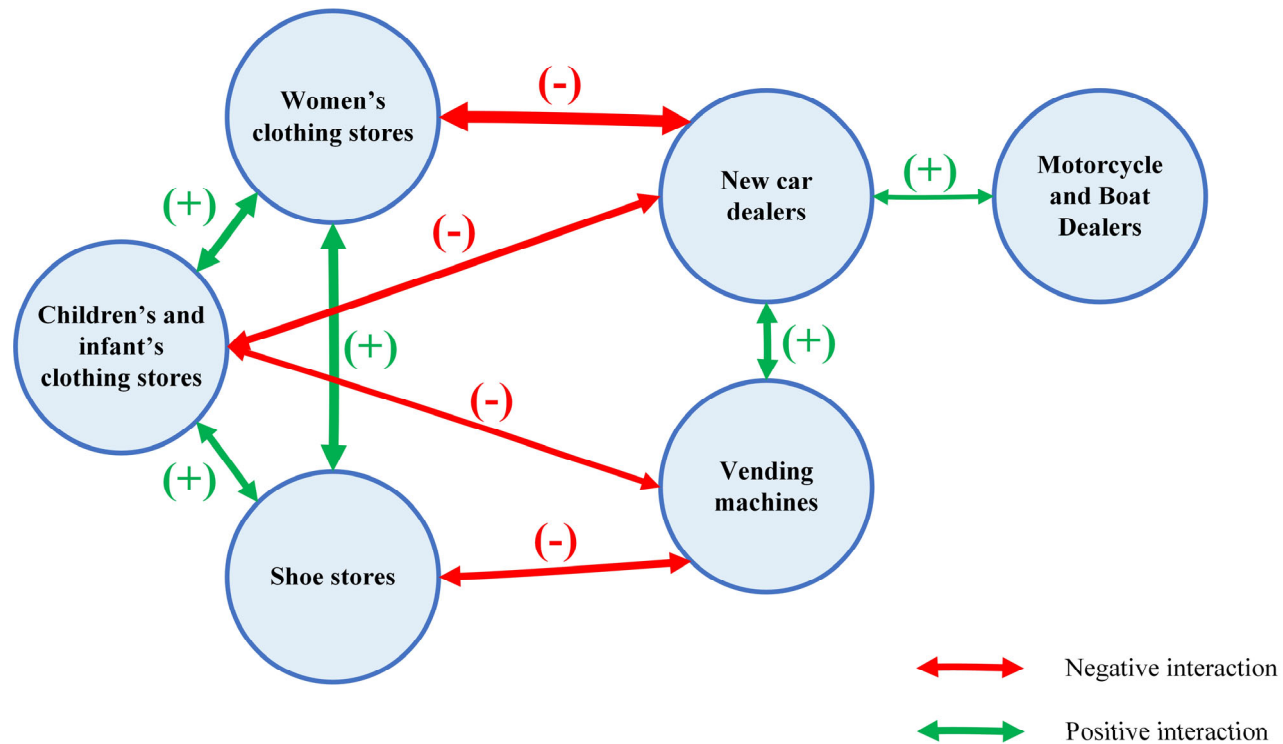
$$a_{AA}(r) = \log_{10} \frac{N_t - 1}{N_A(N_A - 1)} \sum_{i=1}^{N_A} \frac{N_A(A_i)}{N_t(A_i)}$$

Inter-attraction coefficient

$$a_{AB}(r) = \log_{10} \frac{N_t - N_A}{N_A N_B} \sum_{i=1}^{N_A} \frac{N_B(A_i)}{N_t(A_i) - N_A(A_i)}$$

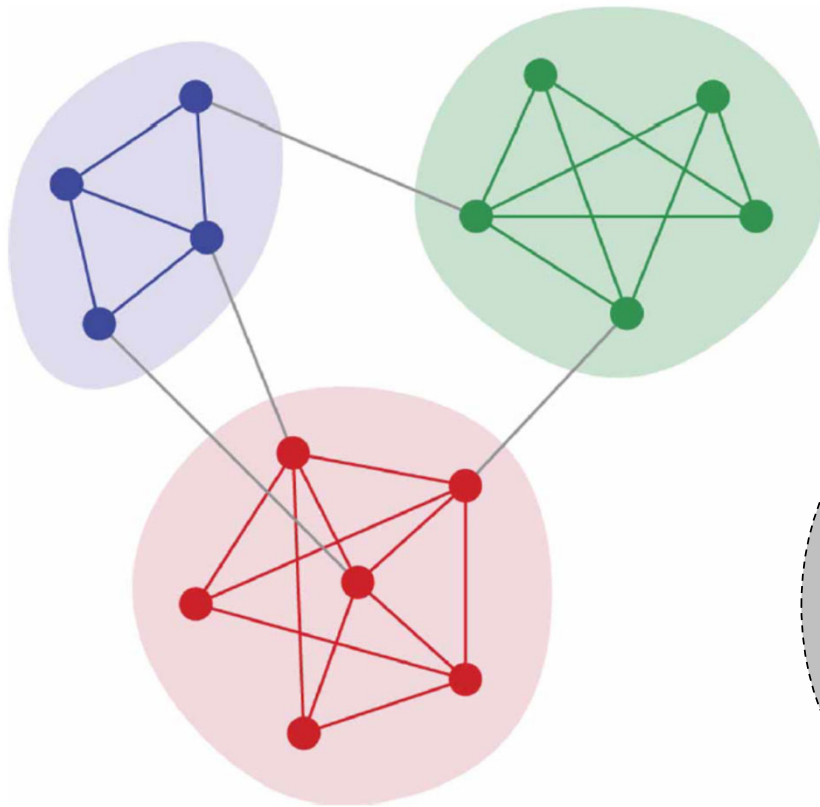
Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas



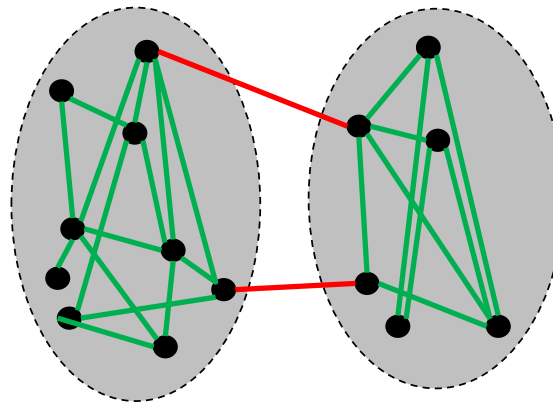
Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

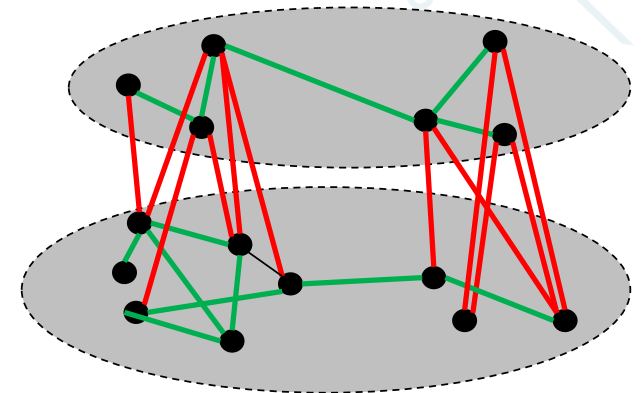


Redes Jensen

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$



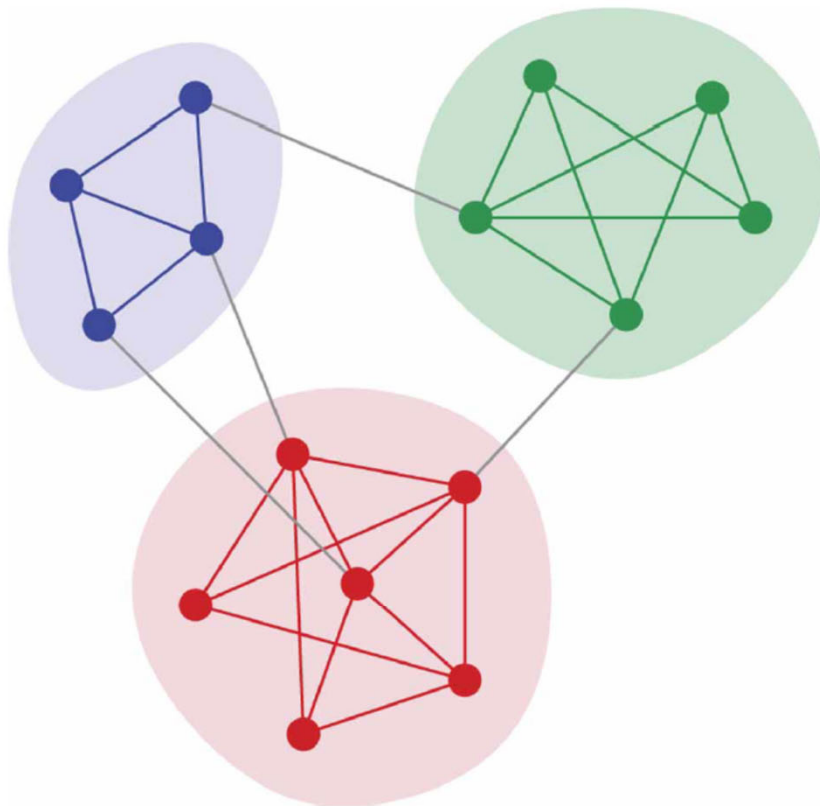
Muy Buena partición
(alta modularidad)



Una partición no tan buena
(modularidad menor)

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas



Redes Jensen

Redes dirigidas, pesadas, con signo y con autoenlaces

$$Q^+ = \frac{1}{2w^+} \sum_i \sum_j \left(w_{ij}^+ - \frac{w_i^+ w_j^+}{2w^+} \right) \delta(C_i, C_j)$$

$$Q^- = \frac{1}{2w^-} \sum_i \sum_j \left(w_{ij}^- - \frac{w_i^- w_j^-}{2w^-} \right) \delta(C_i, C_j)$$

$$Q = \frac{2w^+}{2w^+ + 2w^-} Q^+ - \frac{2w^-}{2w^+ + 2w^-} Q^-$$

Gómez S, Jensen P, Arenas A (2009) Analysis of community structure in networks of correlated data. Phys Rev E Stat Nonlinear Soft Matter Phys 80:16114. <https://doi.org/10.1103/PhysRevE.80.016114>

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

- Problemas de las métricas de Jensen

- Indeterminaciones

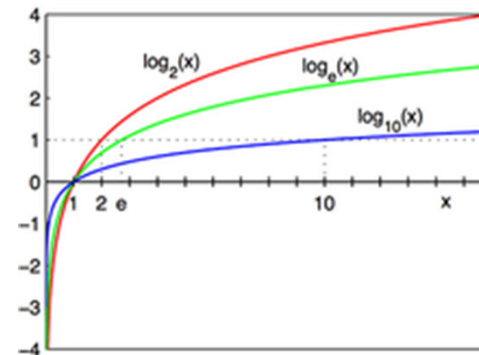
$$a_{AA}(r) = \log_{10} \frac{N_t - 1}{N_A(N_A - 1)} \sum_{i=1}^{N_A} \frac{N_A(A_i)}{N_t(A_i)}$$

- Puntos aislados

$$a_{AB}(r) = \log_{10} \frac{N_t - N_A}{N_A N_B} \sum_{i=1}^{N_A} \frac{N_B(A_i)}{N_t(A_i) - N_A(A_i)}$$

- El problema de las cotas al tomar logaritmos y la dificultad de interpretar las fuerzas

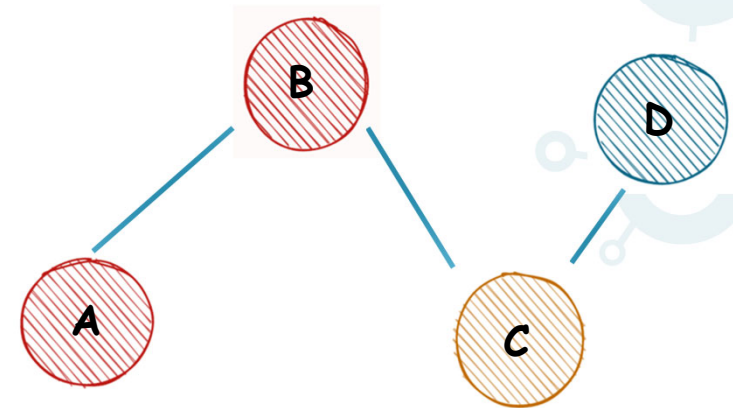
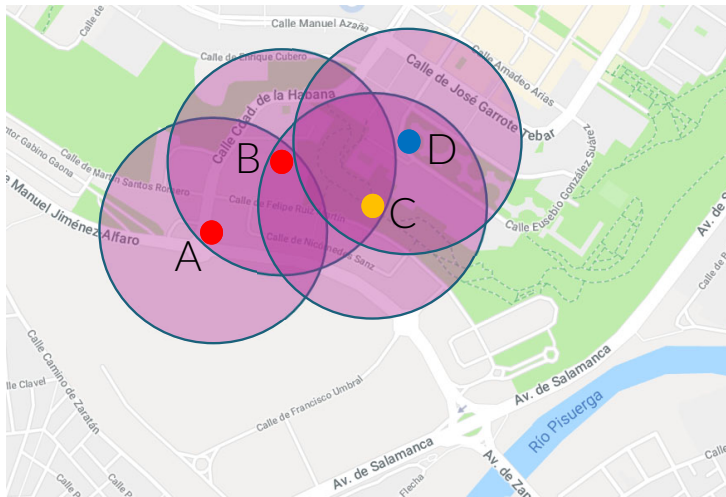
- - Inf



Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

La red inducida



Redes Jensen

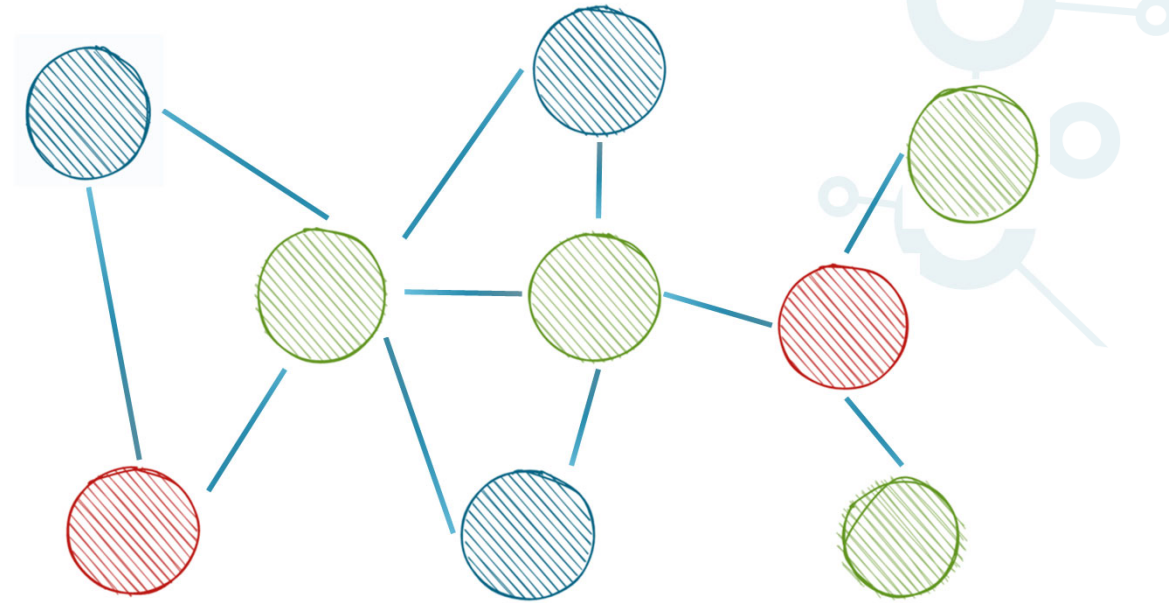
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 1. Dejar la estructura de la red y aleatorizar las categorías de negocios

- La idea implícita es que los negocios y la infraestructura comercial está dada, y los negocios la ocupan
- Esperanza y límites de confianza por Montecarlo
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

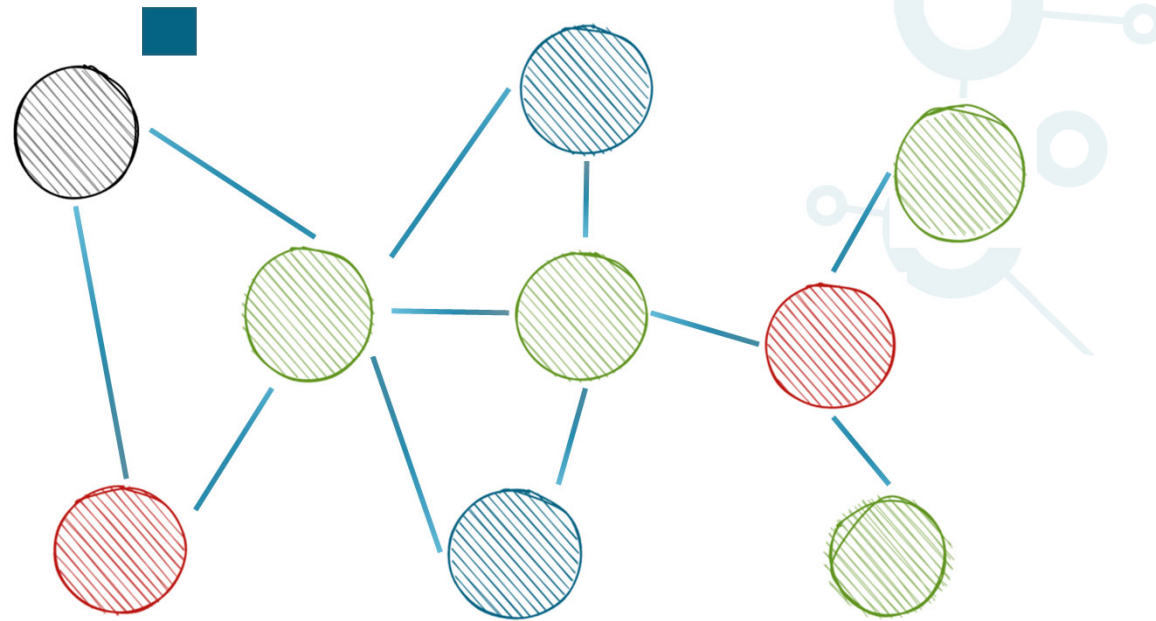
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 1. Dejar la estructura de la red y aleatorizar las categorías de negocios

- La idea implícita es que los negocios y la infraestructura comercial está dada, y los negocios la ocupan
- Esperanza y límites de confianza por Montecarlo
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

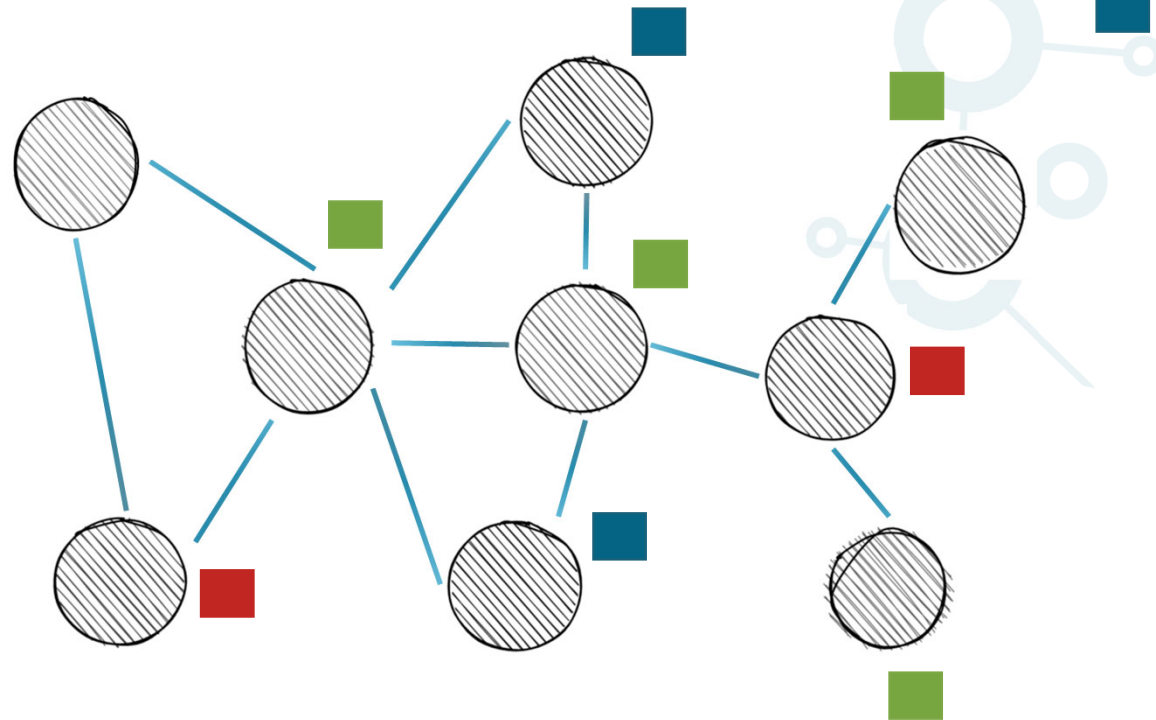
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 1. Dejar la estructura de la red y aleatorizar las categorías de negocios

- La idea implícita es que los negocios y la infraestructura comercial está dada, y los negocios la ocupan
- Esperanza y límites de confianza por Montecarlo
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

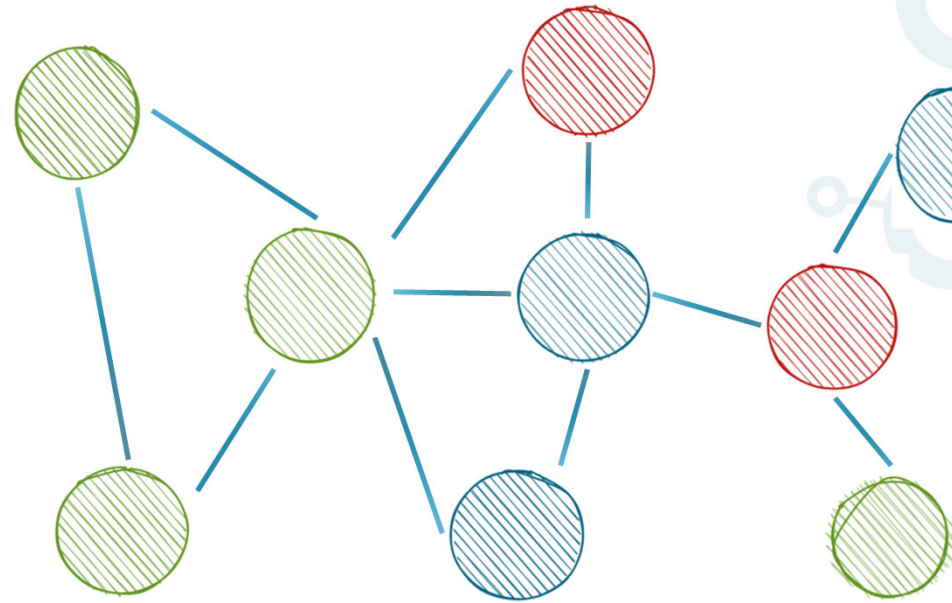
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 1. Dejar la estructura de la red y aleatorizar las categorías de negocios

- La idea implícita es que los negocios y la infraestructura comercial está dada, y los negocios la ocupan
- Esperanza y límites de confianza por Montecarlo
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

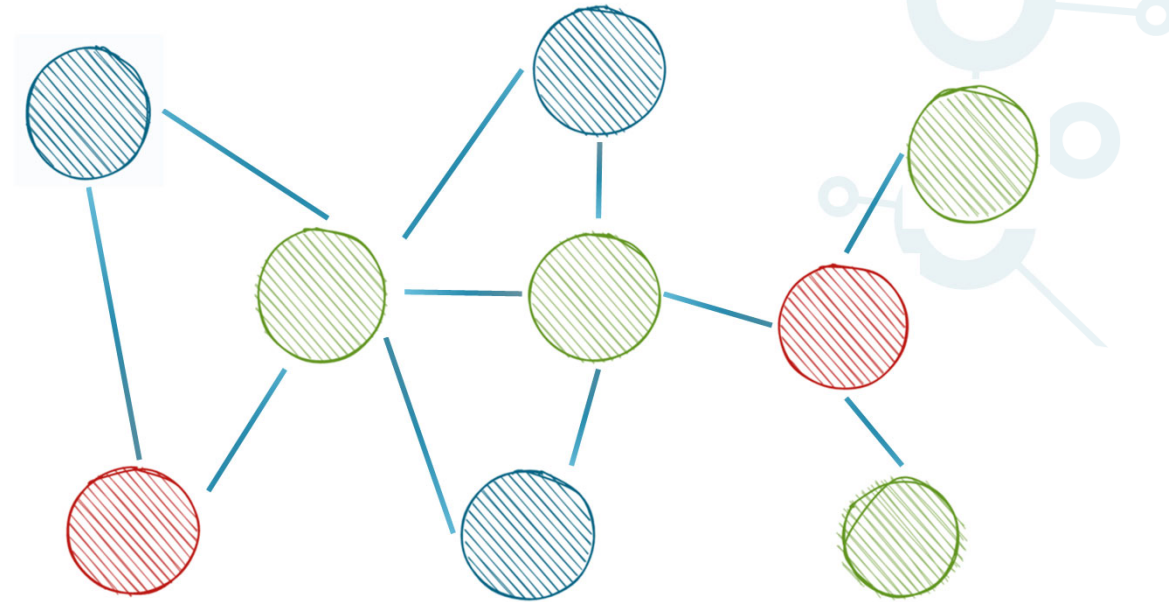
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 2. Mantener el grado de cada nodo y hacer rewiring

- Implícitamente está asumiendo que un negocio dado induce una estructura comercial en su entorno, que es lo que se mantiene en el modelo
- La esperanza se puede deducir analíticamente y probablemente los límites de confianza
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

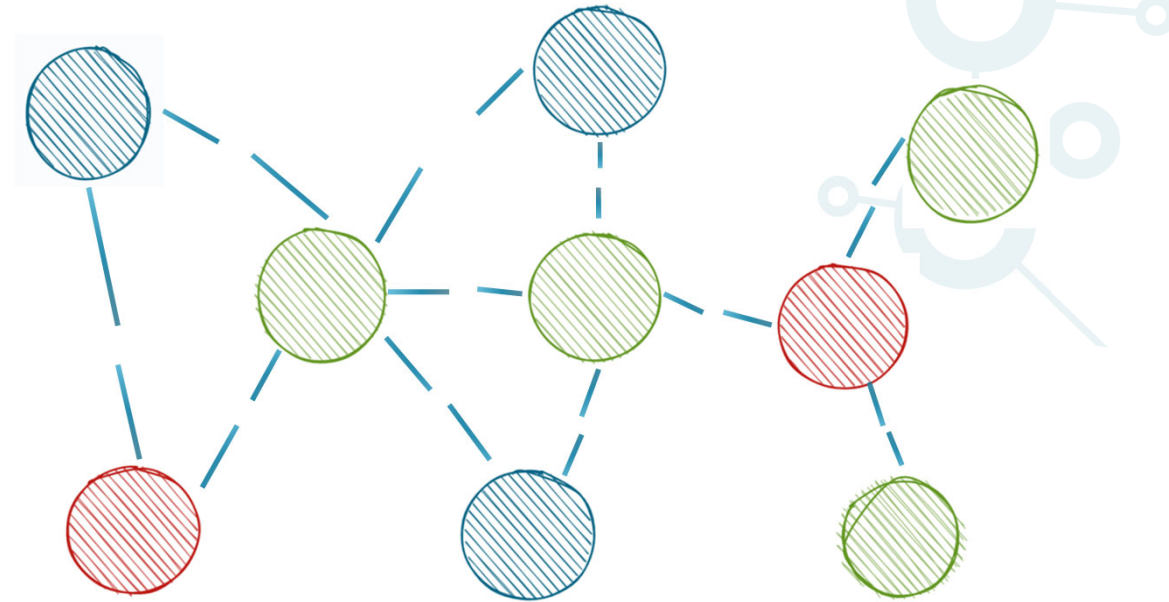
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 2. Mantener el grado de cada nodo y hacer rewiring

- Implícitamente está asumiendo que un negocio dado induce una estructura comercial en su entorno, que es lo que se mantiene en el modelo
- La esperanza se puede deducir analíticamente y probablemente los límites de confianza
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

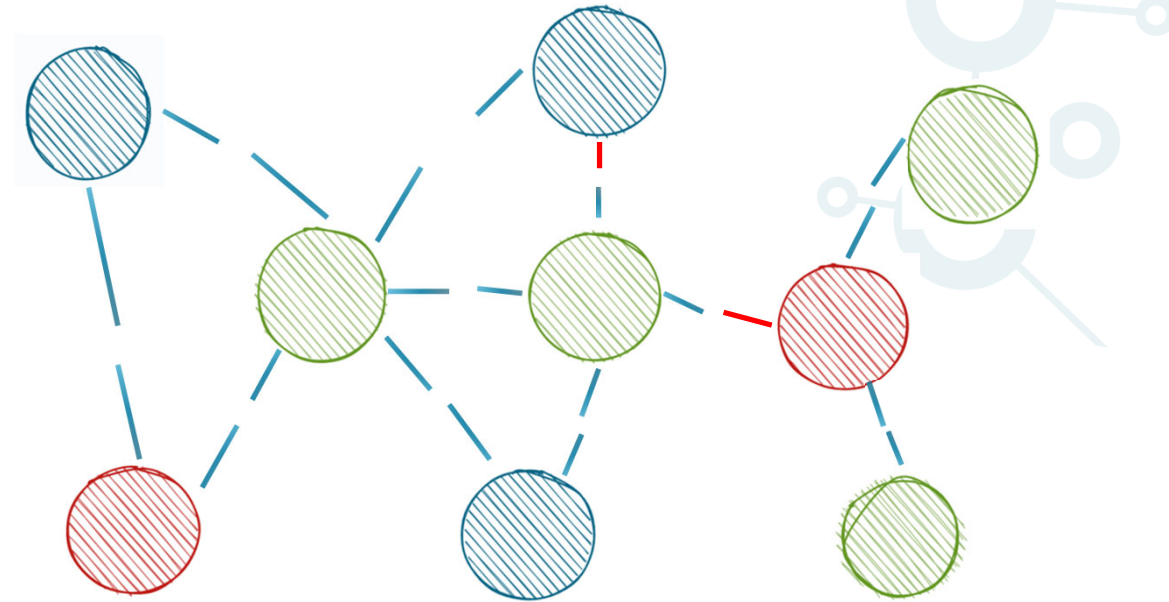
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 2. Mantener el grado de cada nodo y hacer rewiring

- Implícitamente está asumiendo que un negocio dado induce una estructura comercial en su entorno, que es lo que se mantiene en el modelo
- La esperanza se puede deducir analíticamente y probablemente los límites de confianza
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

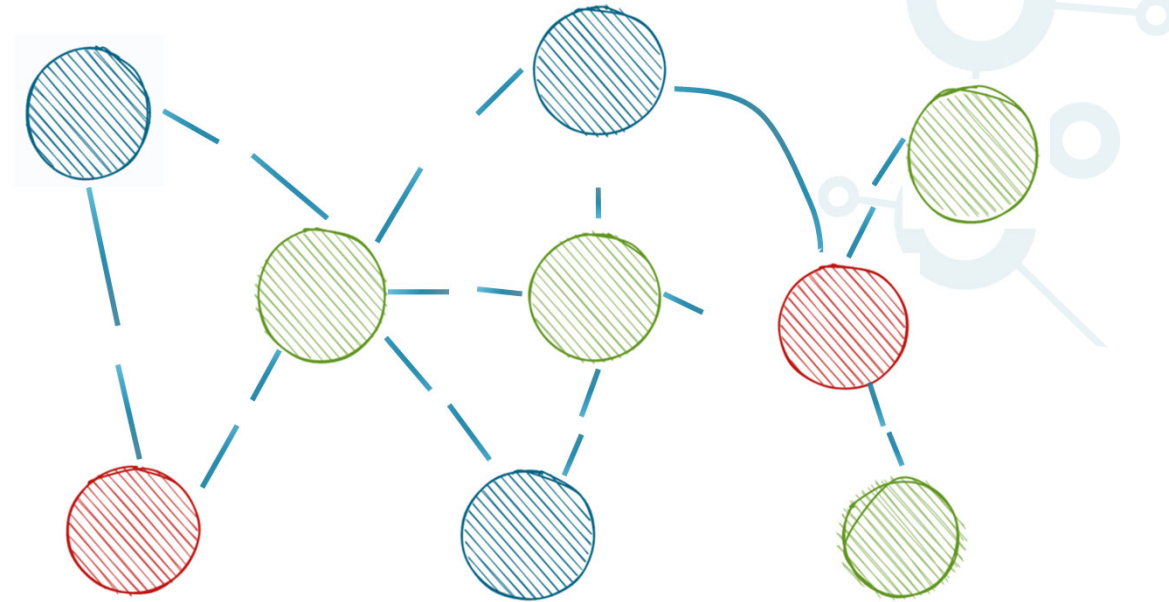
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 2. Mantener el grado de cada nodo y hacer rewiring

- Implícitamente está asumiendo que un negocio dado induce una estructura comercial en su entorno, que es lo que se mantiene en el modelo
- La esperanza se puede deducir analíticamente y probablemente los límites de confianza
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

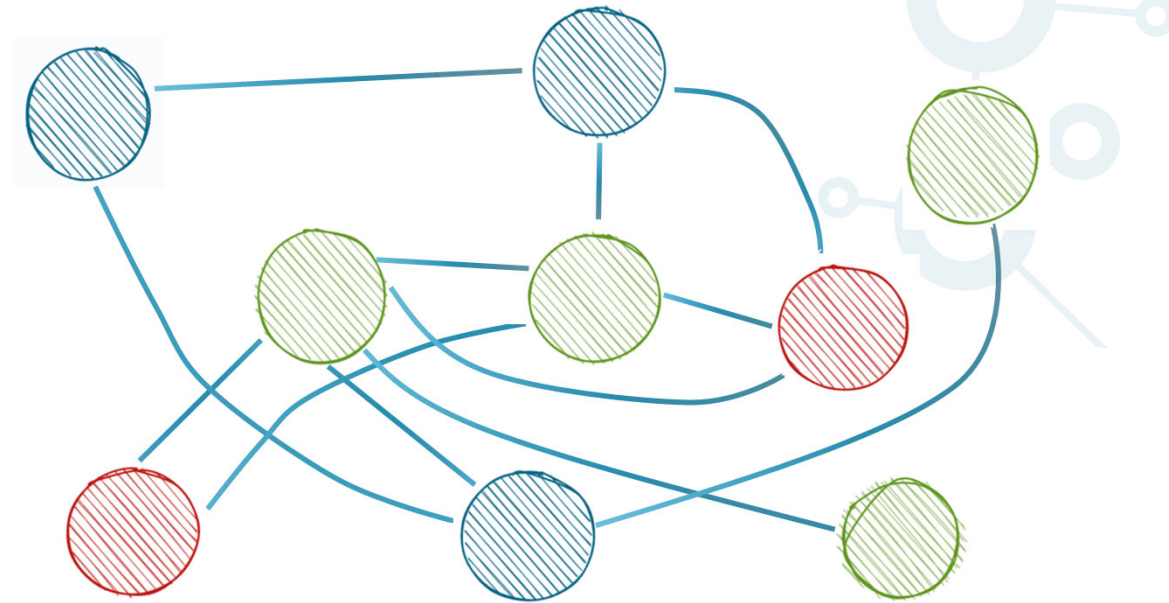
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 2. Mantener el grado de cada nodo y hacer rewiring

- Implícitamente está asumiendo que un negocio dado induce una estructura comercial en su entorno, que es lo que se mantiene en el modelo
- La esperanza se puede deducir analíticamente y probablemente los límites de confianza
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

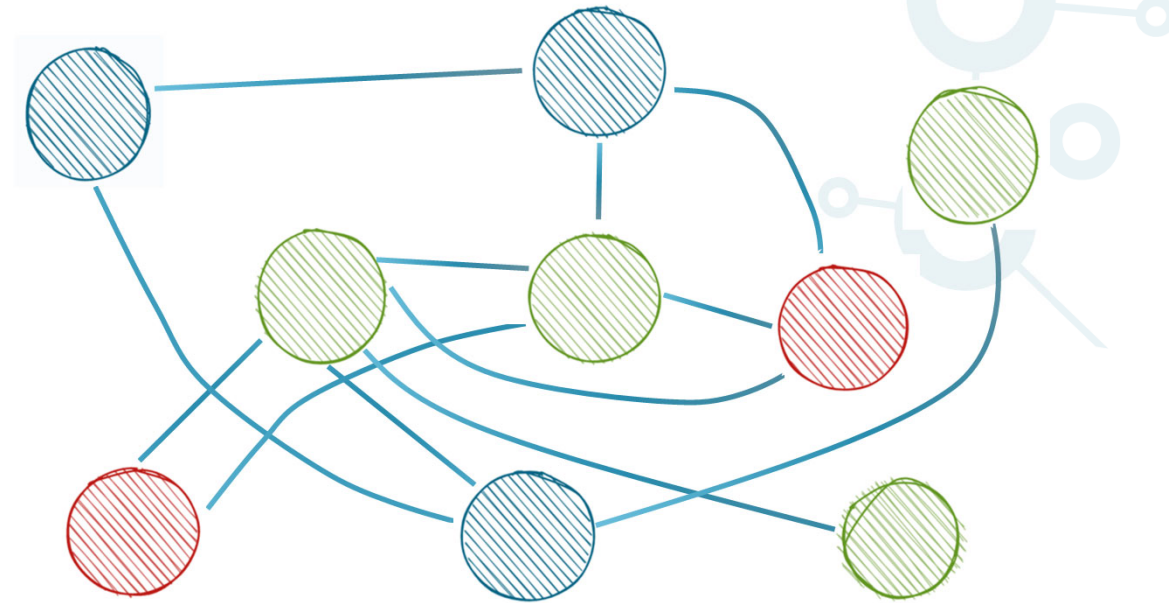
Índices y
significación

Metodología

Caracterización empírica de los sistemas comerciales minoristas en zonas urbanas

Modelo nulo 2. Mantener el grado de cada nodo y hacer rewiring

- Implícitamente está asumiendo que un negocio dado induce una estructura comercial en su entorno, que es lo que se mantiene en el modelo
- La esperanza se puede deducir analíticamente y probablemente los límites de confianza
- Cálculo de la z-score para la fuerza de atracción y repulsión



Redes Jensen

Índices y
significación

Caso de estudio I

Fuentes de datos



Datos

Complex & Intelligent Systems
<https://doi.org/10.1007/s40747-021-00335-8>

ORIGINAL ARTICLE



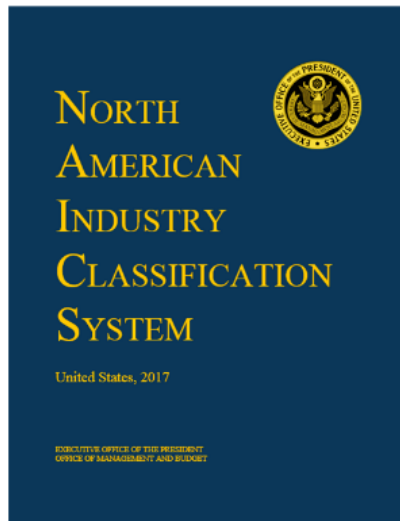
Identification of robust retailing location patterns with complex network approaches

Rosa María Sánchez-Saiz¹ · Virginia Ahedo¹ · José Ignacio Santos¹ · Sergio Gómez² · José Manuel Galán¹

Received: 30 September 2020 / Accepted: 9 March 2021
© The Author(s) 2021

Caso de estudio

Clasificación



44111 New Car Dealers	44112 Used Car Dealers
44121 Recreational Vehicle Dealers	44122 Motorcycle and Boat Dealers
44131 Auto Parts and Accessories	44132 Tire Dealers
44211 Furniture Stores	44221 Floor Covering Stores
44229 Other Home Furnishing Stores	44311 Appliances & Electronics Stores
44312 Computer Stores	44313 Camera and Photography Stores
44411 Home Centers	44412 Paint and Wallpaper Stores
44413 Hardware Stores	44419 Other Building Materials Stores
44421 Outdoor Power Equipment Stores	44422 Nursery and Garden Stores
44511 Grocery Stores	44512 Convenience Stores
44521 Meat Markets	44522 Fish and Seafood Markets
44523 Fruit and Vegetable Markets	44529 Other Specialty Food Markets
44531 Liquor Stores	44611 Pharmacy and Drug Stores
44612 Cosmetics and Beauty Stores	44613 Optical Goods Stores
44619 Other Health and Personal Care Stores	44711 Gasoline Stations with Convenience Stores
44719 Gasoline Stations w/o Conv. Stores	44811 Men's Clothing Stores
44812 Women's Clothing Stores	44813 Children's and Infant's Clothing Stores
44814 Family Clothing Stores	44815 Clothing Accessory Stores
44819 Other Apparel Stores	44821 Shoe Stores
44831 Jewelry Stores	44832 Luggage Stores
45111 Sporting Goods Stores	45112 Hobby, Toy, and Game Stores
45113 Sewing and Needlecraft Stores	45114 Musical Instrument Stores
45121 Book Stores	45122 Record, Tape, and CD Stores
45211 Department Stores	45291 Warehouse Superstores
45299 Other General Merchandise Stores	45311 Florists
45321 Office and Stationary Stores	45322 Gift and Souvenir Stores
45331 Used Merchandise Stores	45391 Pet and Pet Supply Stores
45392 Art Dealers	45393 Mobile Home Dealers
45399 Other Miscellaneous Retail Stores	45411 Mail Order and Catalog Stores
45421 Vending Machines	45431 Fuel Dealers
45439 Other Direct Selling Establishments	7211 Hotels & Other Travel Accommodations
7212 RV Parks	7213 Rooming and Boarding Houses
7221 Full Service Restaurants	7222 Limited Service Restaurants
7223 Special Food Services and Catering	7224 Drinking Places

Datos

Clasificación

Caso de estudio

Geo-referenciación



Google Maps Platform



OpenStreetMap

Datos

Clasificación

Geo-referenciación

Configuración de los experimentos



COMPARATIVA DE LOS TRES MÉTODOS

Para cada ciudad:

Obtención de la red empírica de interacciones entre categorías comerciales.

Tres métodos → Tres redes → Tres matrices de interacción empíricas

Simulaciones modelos nulos:

1.000 iteraciones Jensen

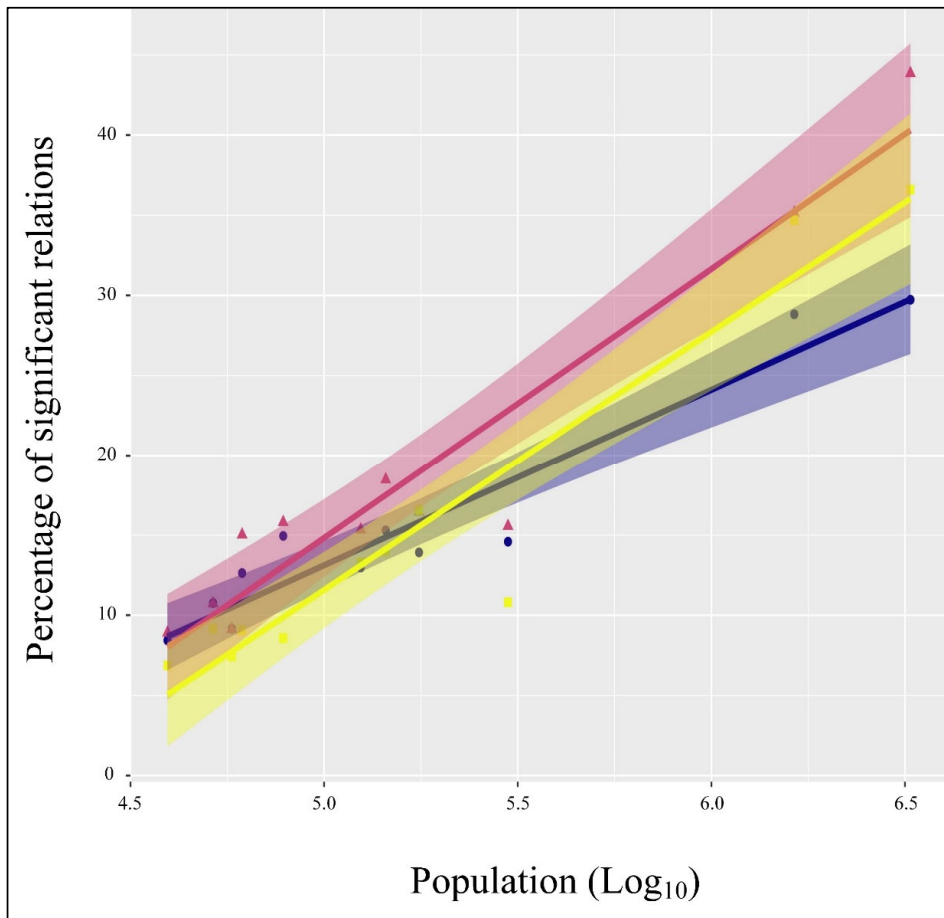
10.000 iteraciones Permutation

1.000 iteraciones Rewiring

Relaciones significativas → Nivel de significación del 0,05 (chequeadas contra el rango definido por los percentiles 2,5 y 97,5).

Resultados

¿Existe relación entre el tamaño de la ciudad y el número de relaciones significativas encontradas?

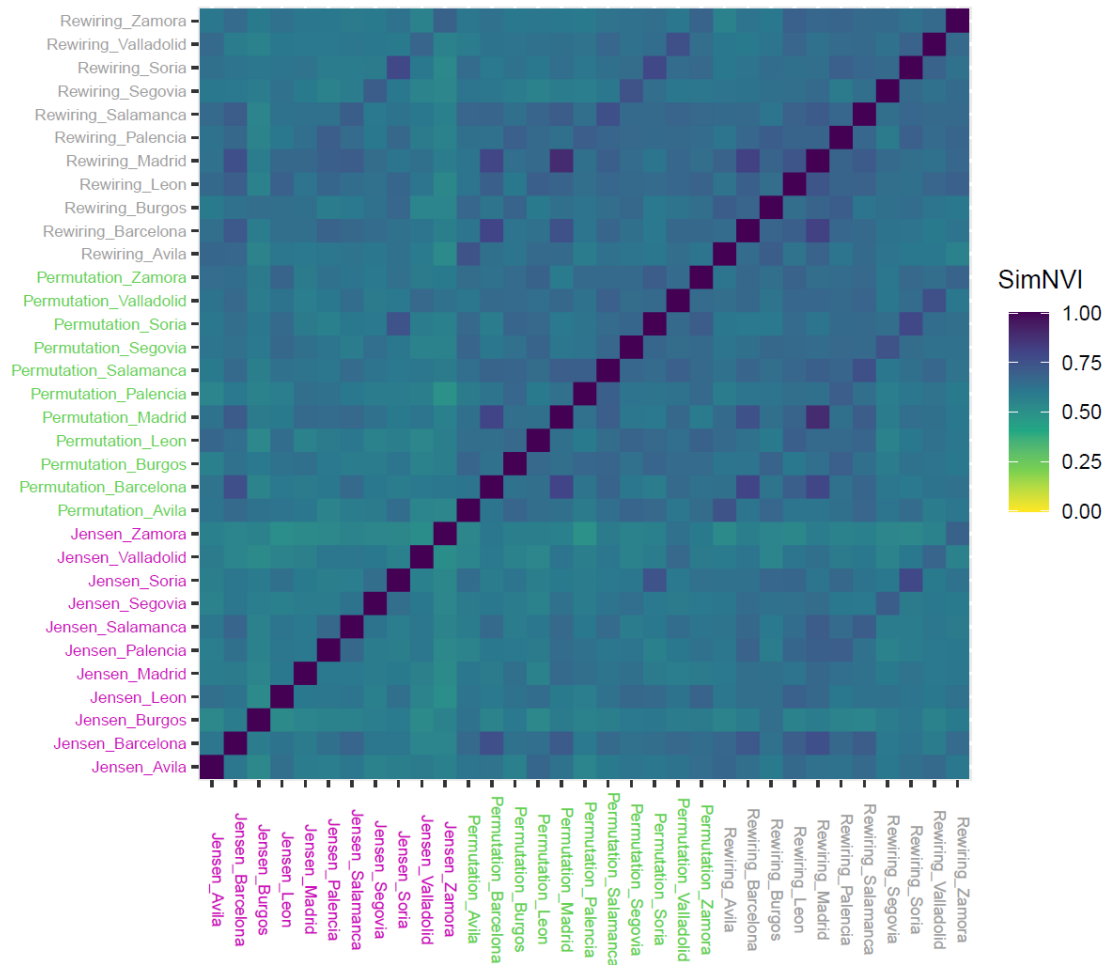


	Adjusted R ²	Spearman's rank correlation
Jensen	0.914 (2.7 10 ⁻⁶)	0.891 (4.3 10 ⁻⁴)
Z-score permutation	0.908 (3.6 10 ⁻⁶)	0.909 (5.6 10 ⁻⁵)
Z-score rewiring	0.904 (4.3 10 ⁻⁶)	0.882 (6.7 10 ⁻⁴)

En todos los casos el coeficiente de correlación de Spearman encuentra una relación monótonica significativa entre las dos variables → A mayor tamaño de la ciudad, mayor especialización comercial.

Resultados

Comparación entre métodos y ciudades



OJO: SimNVI está concebida para comparar dos particiones del mismo dataset. En nuestro caso, como tenemos 11 ciudades no siempre tenemos las mismas categorías comerciales → Trabajamos sobre la intersección de categorías común a todas las ciudades.



Resultados

Evaluación de la existencia de relaciones comerciales sólidas entre las ciudades

RED CONSENSO DE RELACIONES:

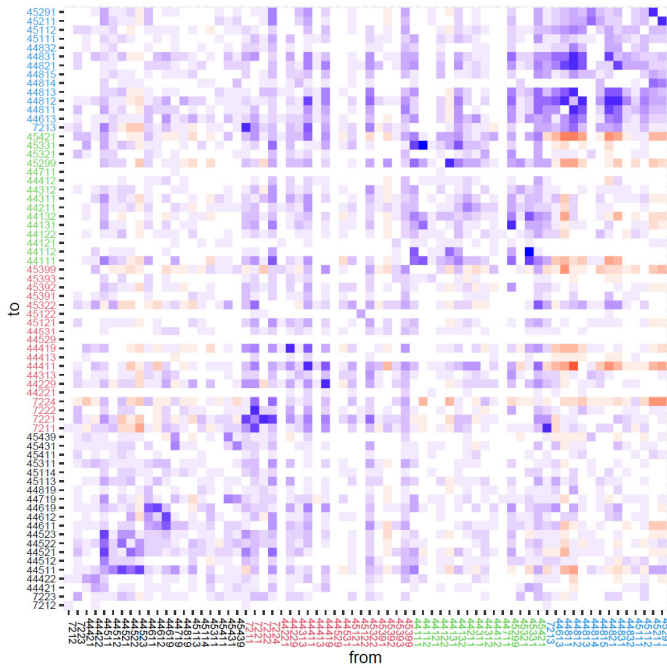
- Pesada y con signo
- Nodos → categorías comerciales
- Peso enlace → Para cada par de categorías, chequeamos a lo largo de las 11 ciudades si existe una relación significativa entre ellas: +1 si la relación es positiva, -1 si es negativa. Máx. valor absoluto = 11.
- Consideramos todas las categorías (no la intersección).
- Algoritmo detección comunidades → Gómez et al. 2009.

RED CONSENSO DE PARTICIONES:

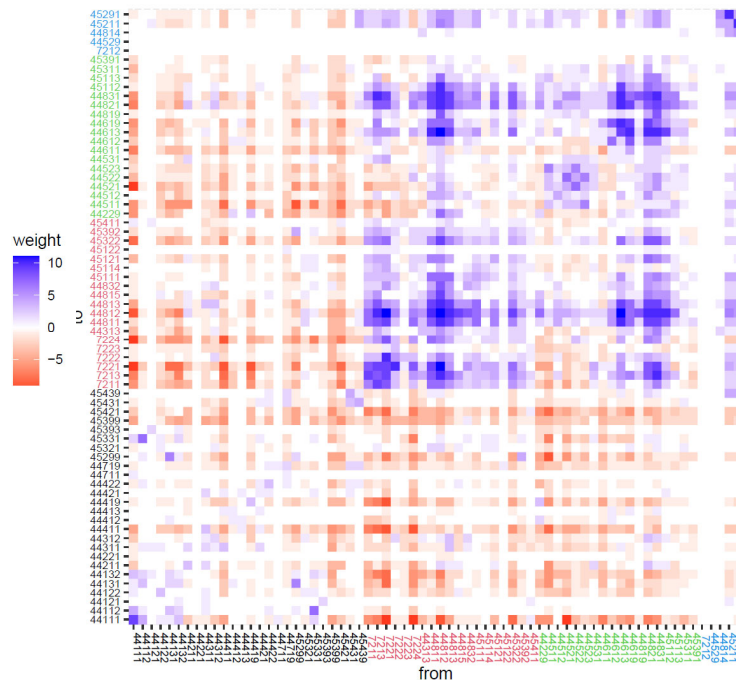
- Pesada y con enlaces positivos.
- Nodos → categorías comerciales
- Peso enlace → Número de ciudades en que las dos categorías pertenecen a la misma comunidad para algoritmo de Gómez et al. 2009.
- Algoritmo detección comunidades → OSLOM



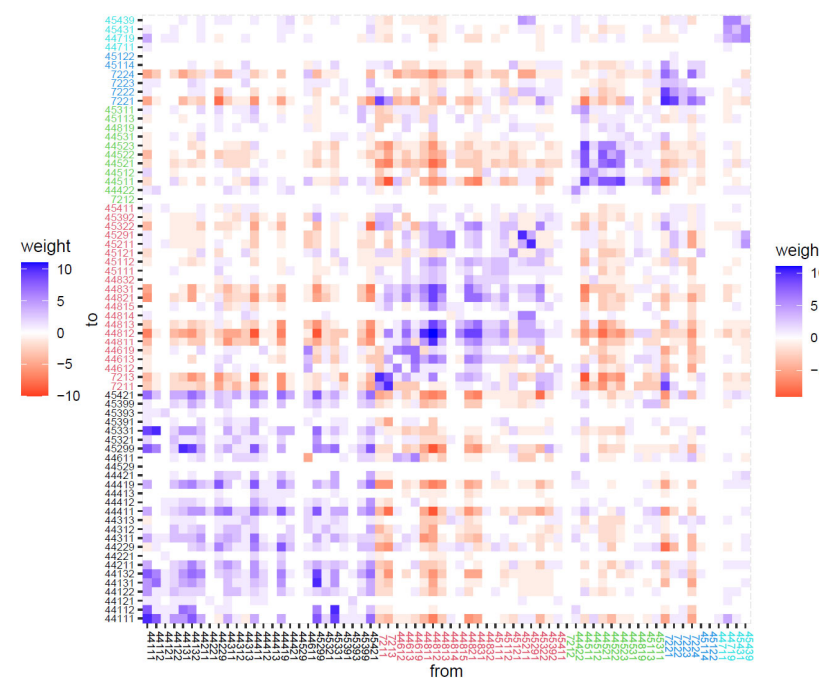
RED CONSENSO DE RELACIONES JENSEN. *THRESHOLD* = 0



RED CONSENSO DE RELACIONES PERMUTATION. THRESHOLD = 0



RED CONSENSO DE RELACIONES REWIRING. THRESHOLD = 0



Resultados



Tablas resumen redes consenso relaciones

From	To	Jensen	Permutation	Rewiring
Women's clothing stores	Men's clothing stores	10	10	10
Men's clothing stores	Women's clothing stores	10	10	10
Warehouse superstores	Department stores	10	10	10
Department stores	Warehouse superstores	10	10	10
Rooming and boarding houses	Hotels and other travel accommodations	10	9	10
Hotels and other travel accommodations	Rooming and boarding houses	10	9	10
Limited service restaurants	Full service restaurants	10	10	9
Full service restaurants	Limited service restaurants	10	10	9
Children's and infant's clothing stores	Women's clothing stores	10	10	9
Shoe stores	Women's clothing stores	10	10	9

Resultados



Tablas resumen redes consenso relaciones

From	To	Jensen	Permutation	Rewiring
Women's clothing stores	Home centers	- 3	- 8	- 9
Women's clothing stores	Vending machines	- 5	- 3	- 7
Women's clothing stores	New car dealers	- 9	- 9	- 7
Shoe stores	New car dealers	- 7	- 7	- 6
Rooming and boarding houses	Home centers	- 8	- 9	- 8
Women's clothing stores	Other general merchandise stores	- 3	- 5	- 9
Men's clothing stores	Vending machines	- 4	- 8	- 6
Shoe stores	Vending machines	- 5	- 10	- 6
New car dealers	Full service restaurants	- 10	- 10	- 5
Jewelry stores	Vending machines	- 4	- 6	- 7

REDES CONSENSO DE PARTICIONES, *Threshold* = 1

Significance:
0.00557547

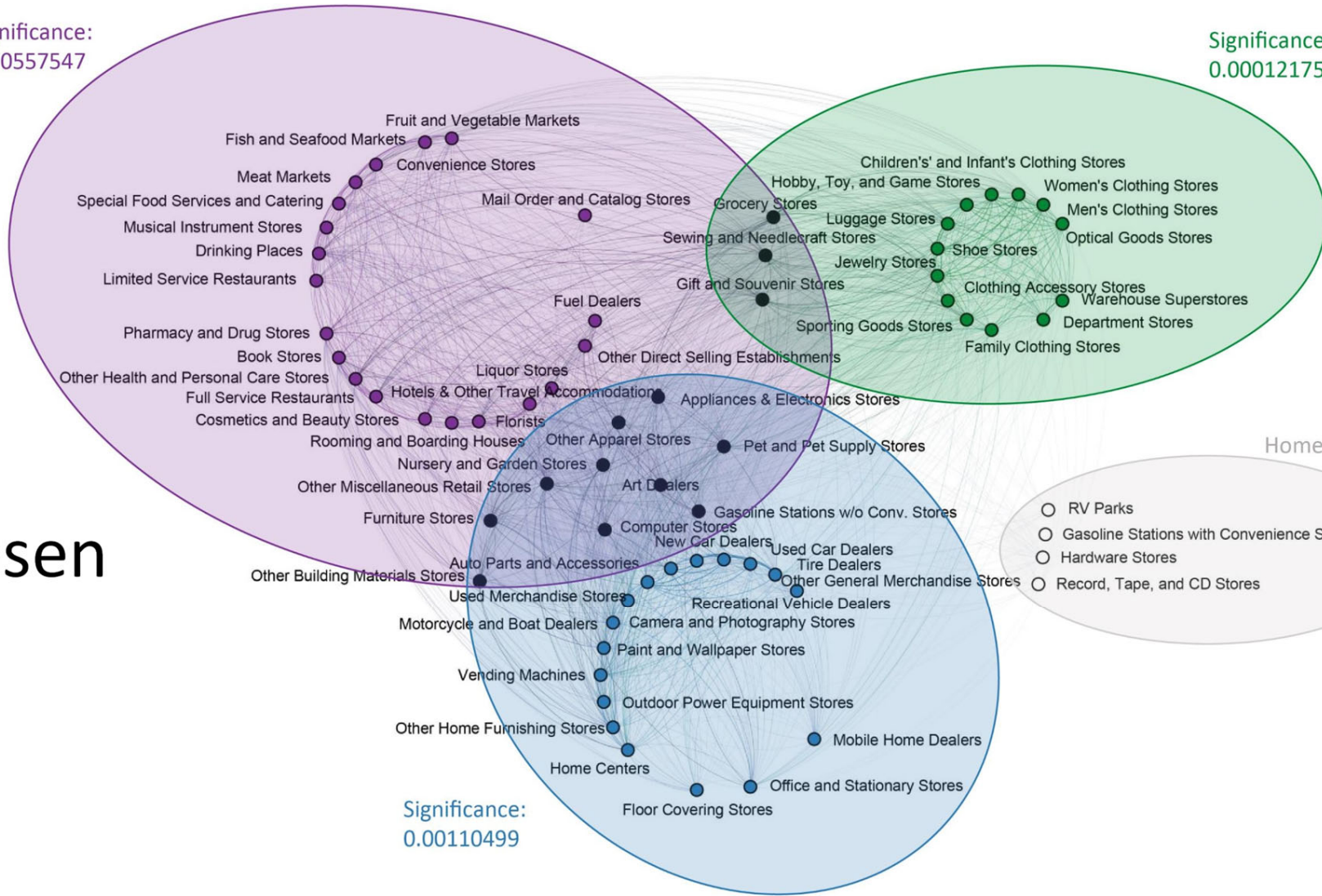
Significance:
0.000121758

Jensen

Significance:
0.00110499

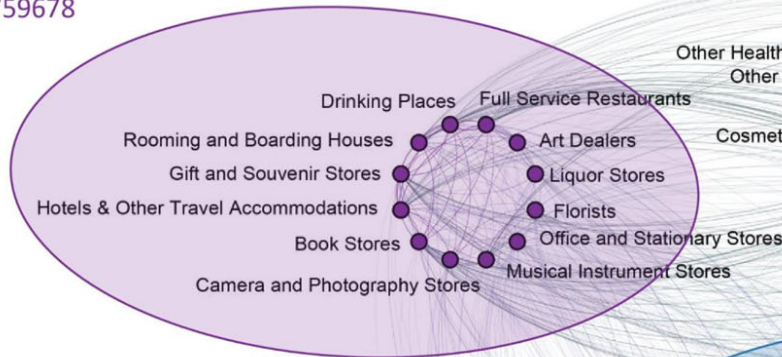
Homeless nodes

- RV Parks
- Gasoline Stations with Convenience Stores
- Hardware Stores
- Record, Tape, and CD Stores



REDES CONSENSO DE PARTICIONES, *Threshold* = 1

Significance:
0.00759678

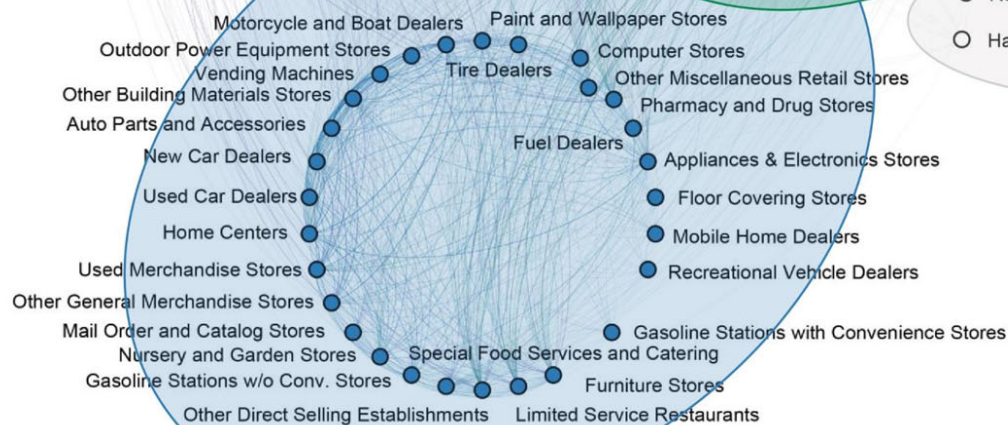


Significance:
7.82051e-05



Homeless nodes

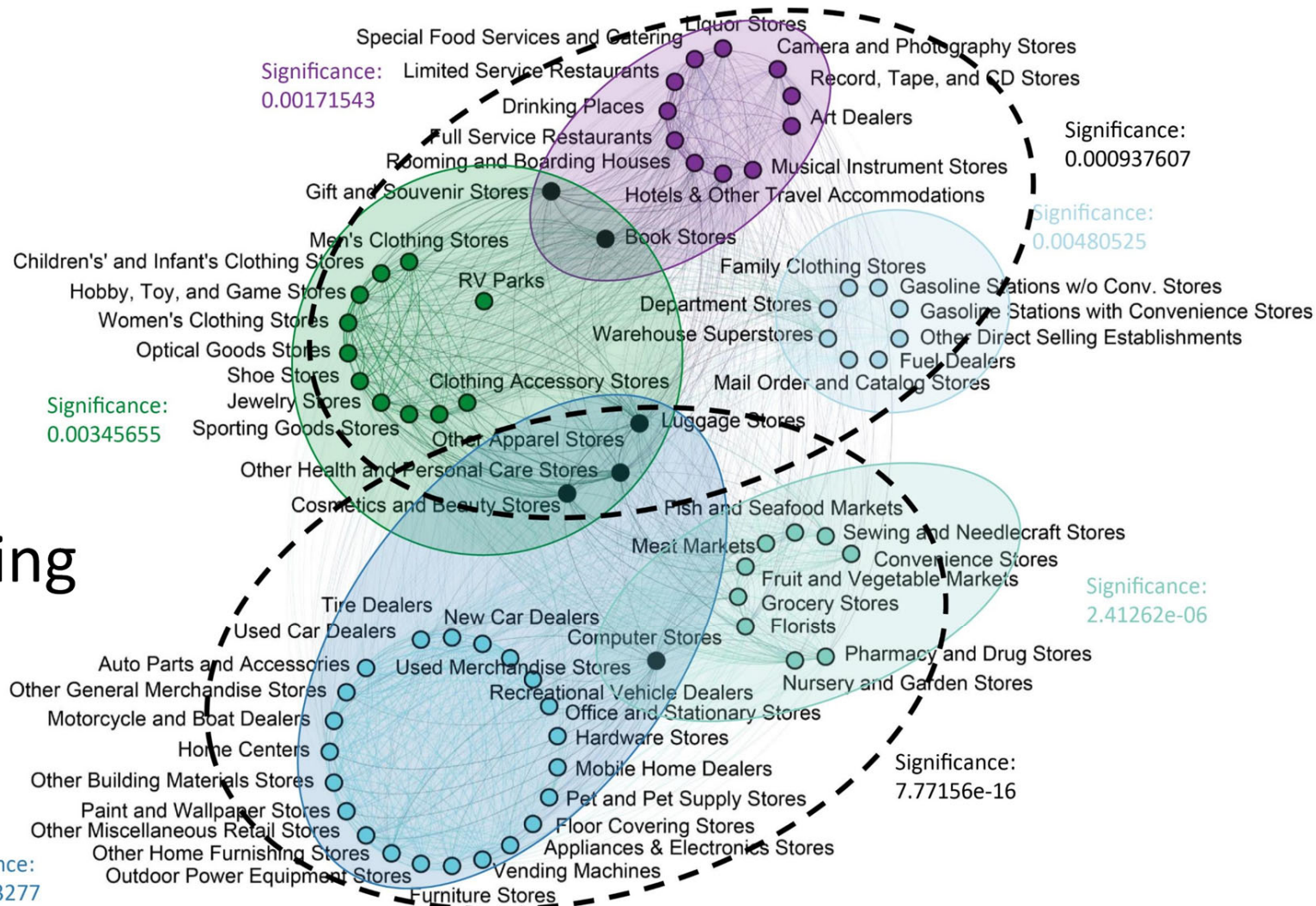
- Record, Tape, and CD Stores
- Hardware Stores



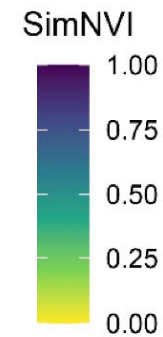
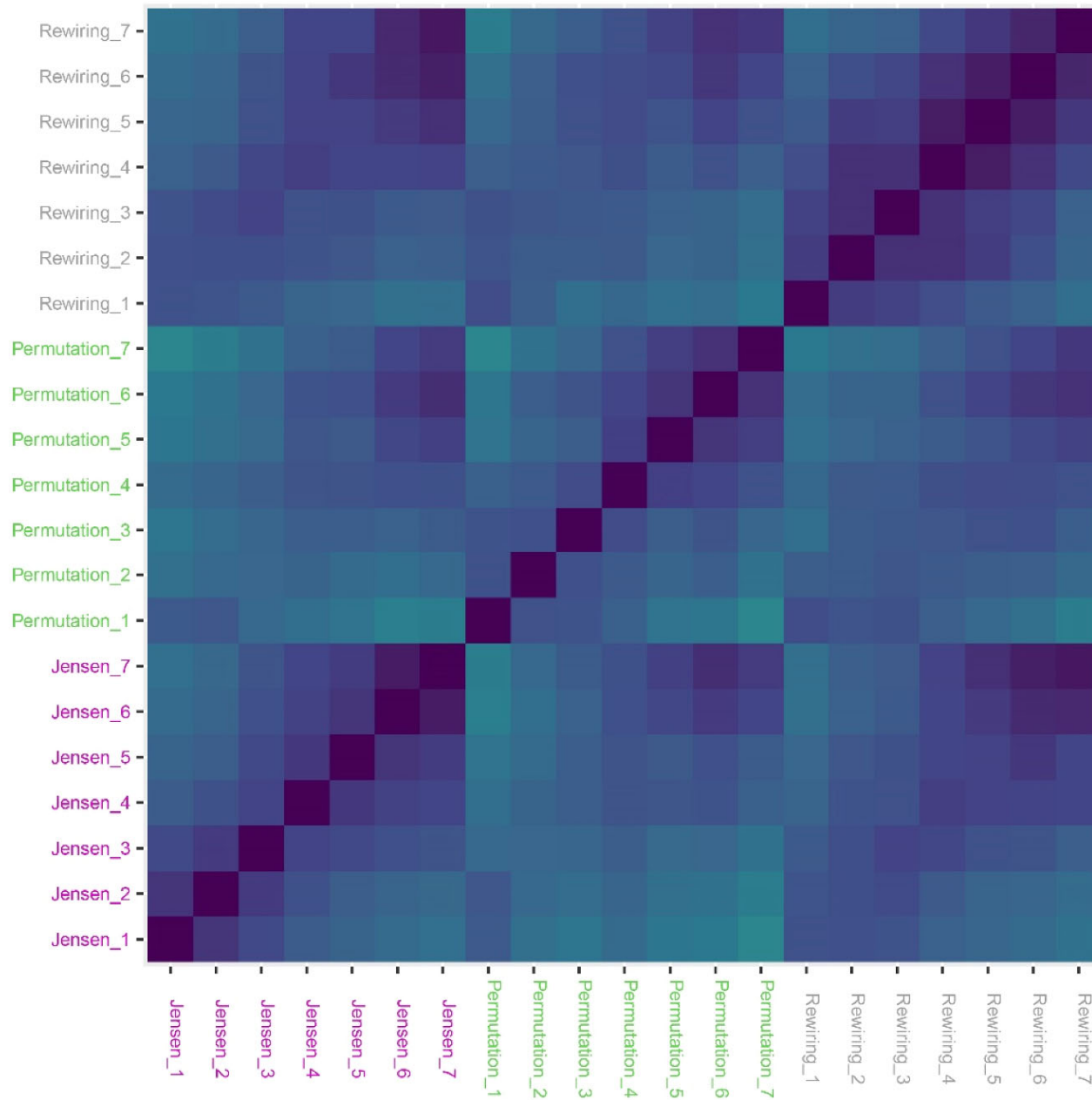
Significance:
0.00143953

Permutation

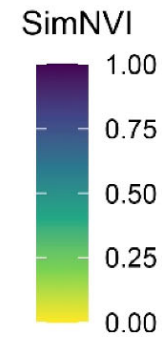
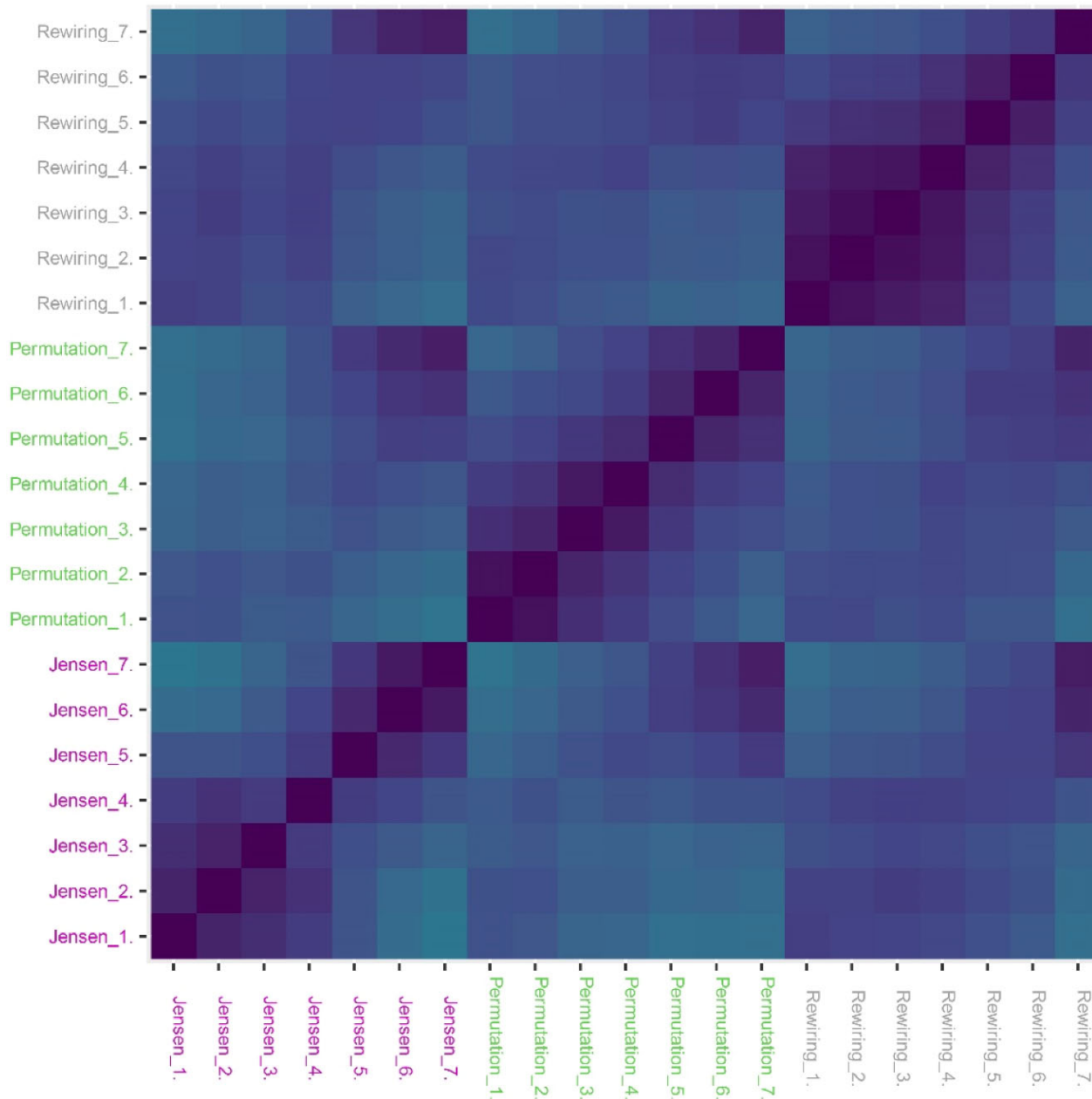
REDES CONSENSO DE PARTICIONES, *Threshold* = 1



Rewiring

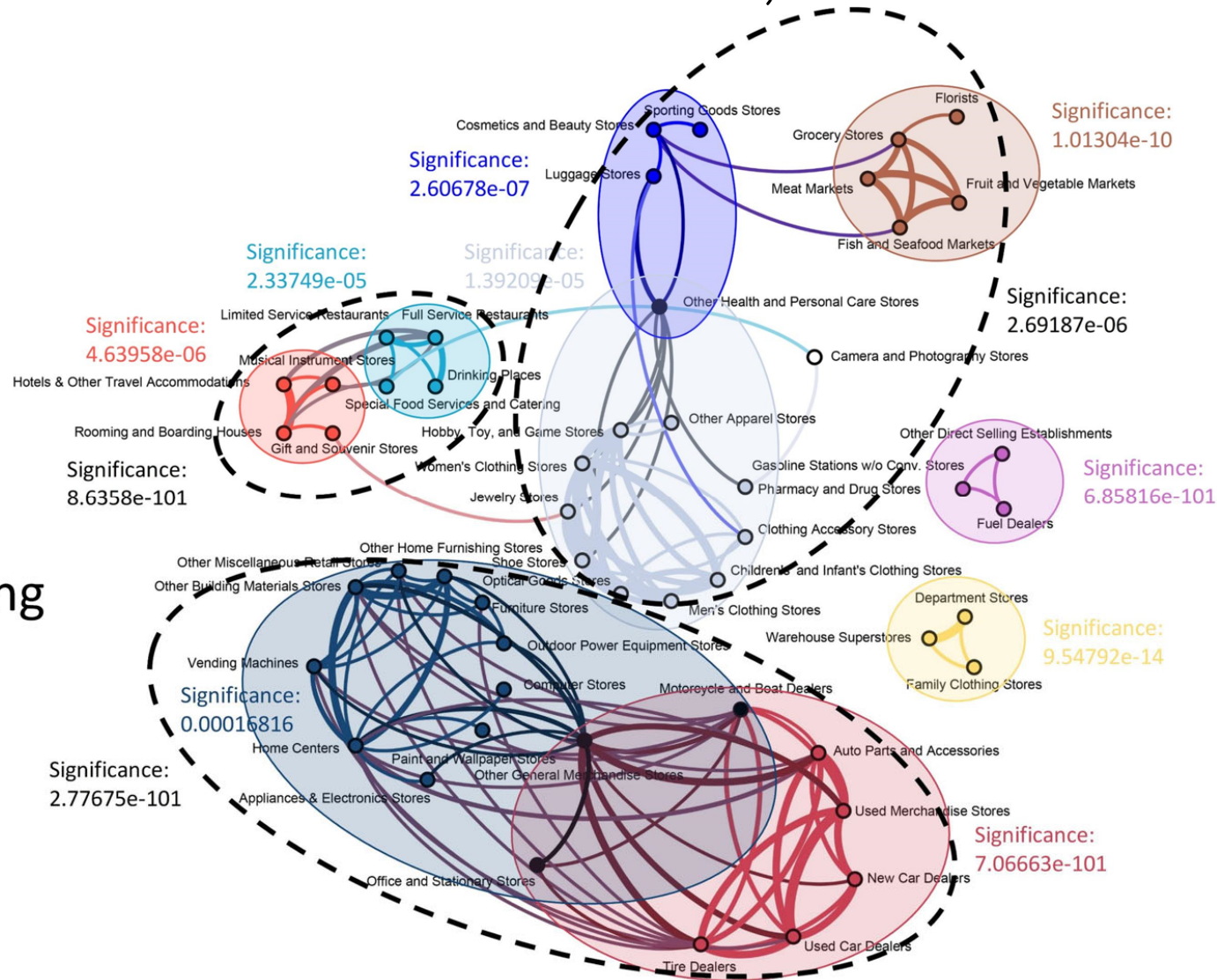


Similitud entre
particiones redes
consenso de relaciones
variando el *threshold*



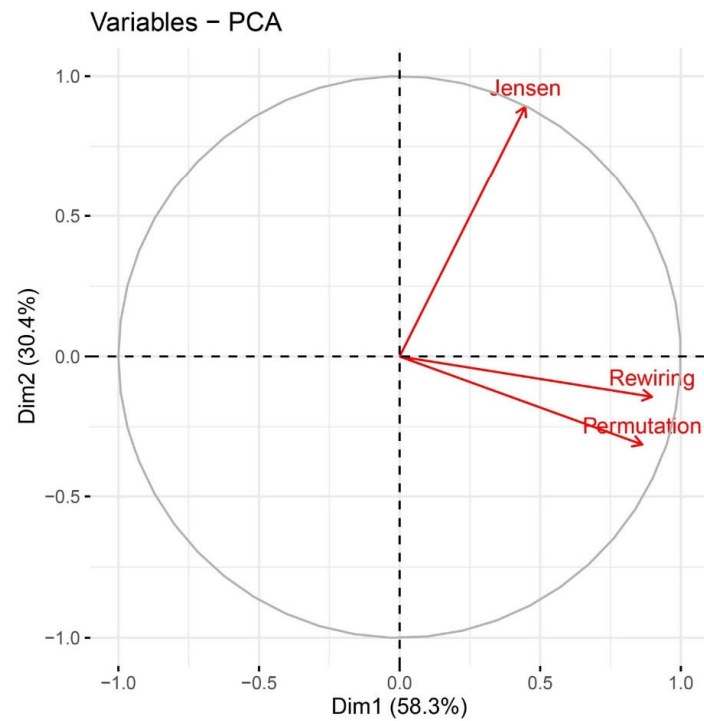
Similitud entre
particiones redes
consenso de particiones
variando el *threshold*

REDES CONSENSO DE PARTICIONES, *Threshold* = 6

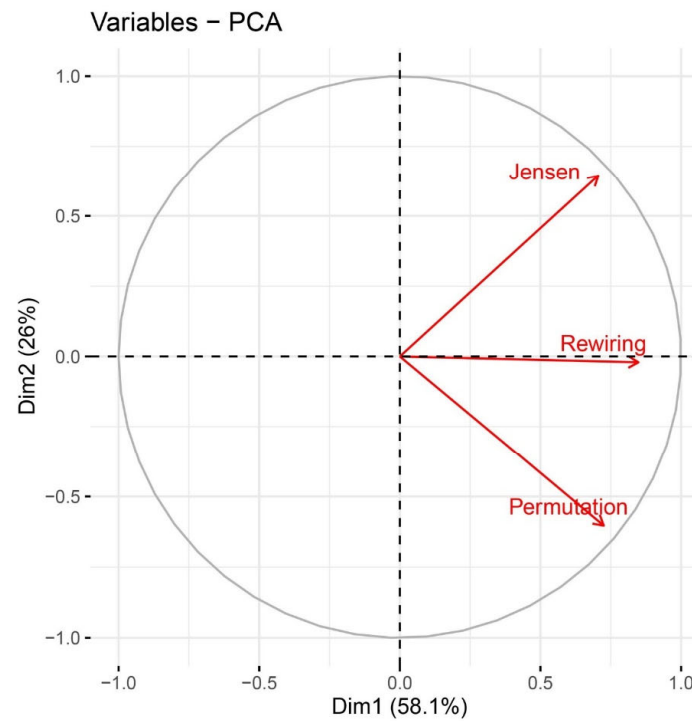


COMPARATIVA ENTRE MÉTODOS

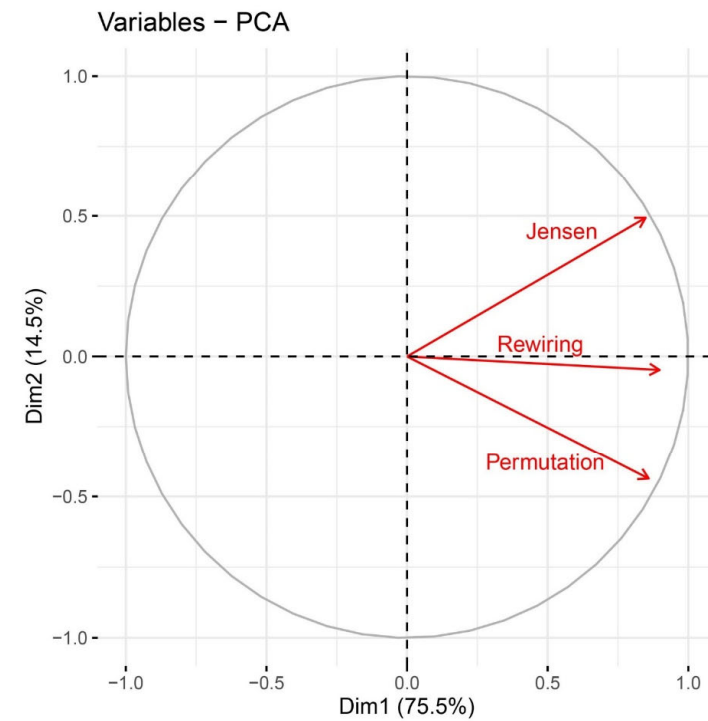
Disaggregated data



Consensus by significance



Consensus by partition



Caso de estudio II

Fuentes de datos



IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 7, 2021, accepted September 23, 2021, date of publication September 24, 2021,
date of current version October 4, 2021.
Digital Object Identifier 10.1109/ACCESS.2021.3015712

Knowledge Transfer in Commercial Feature Extraction for the Retail Store Location Problem

VIRGINIA AHEDO¹, JOSÉ IGNACIO SANTOS², AND JOSÉ MANUEL GALÁN³

¹Departamento de Ingeniería de Organización, Universidad de Burgos, 09006 Burgos, Spain

Corresponding author: José Manuel Galán (jmgalan@ubu.es)

This work was supported in part by the Spanish Ministry of Science, Innovation and Universities through Excellence Network under Grant RED2018-102518-T, in part by the Spanish State Research Agency under Grant PID2020-118906GB-I00/AEI/10.13039/501100011033, and in part by the Junta de Castilla y León—Consejería de Educación under Grant BDNS 425389.

DIFERENTES ÍNDICES DE CALIDAD

ÍNDICES DE CALIDAD RAW → Considerando el valor medio

1. Jensen's Quality Index

$$Q_{JENSEN_i}(x, y) = \sum_{j=1}^N a_{ij} (nei_{ij}(x, y) - \overline{nei_{ij}})$$

2. Permutation Quality Index

$$Q_{PERMUTATION_i}(x, y) = \sum_{j=1}^N a_{ij} (nei_{ij}(x, y) - \overline{nei_{ij}})$$

3. Rewiring Quality Index

$$Q_{REWIRING_i}(x, y) = \sum_{j=1}^N a_{ij} (nei_{ij}(x, y) - \overline{nei_{ij}})$$



DIFERENTES ÍNDICES DE CALIDAD

ÍNDICES DE CALIDAD RAW → Sin considerar el valor medio

4. Jensen's Quality Index Raw

$$Q_{JENSEN-RAW_i}(x, y) = \sum_{j=1}^N a_{ij} \left(nei_{ij}(x, y) \right) \text{ con } a_{ij} = \log M_{ij}$$

5. Permutation Quality Index Raw

$$Q_{PERMUTATION-RAW_i}(x, y) = \sum_{j=1}^N a_{ij} \left(nei_{ij}(x, y) \right) \text{ con } a_{ij} = \frac{x_{ij} - \overline{x_{ij}^{null_model_PERM}}}{s_{ij}^{null_model_PERM}}$$

6. Rewiring Quality Index Raw

$$Q_{REWIRING-RAW_i}(x, y) = \sum_{j=1}^N a_{ij} \left(nei_{ij}(x, y) \right) \text{ con } a_{ij} = \frac{x_{ij} - \overline{x_{ij}^{null_model_REW}}}{s_{ij}^{null_model_REW}}$$



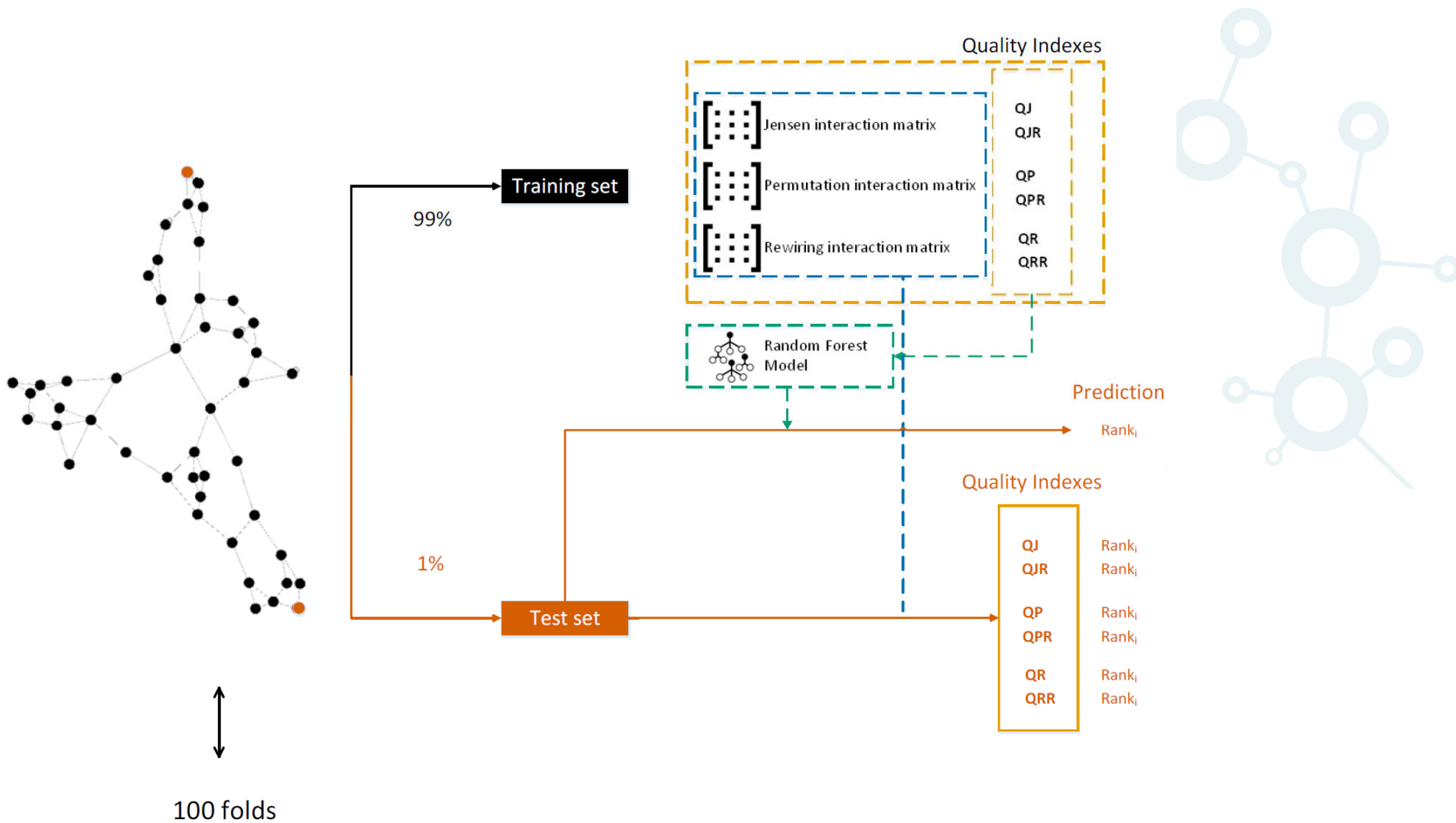
Diseño de experimentos

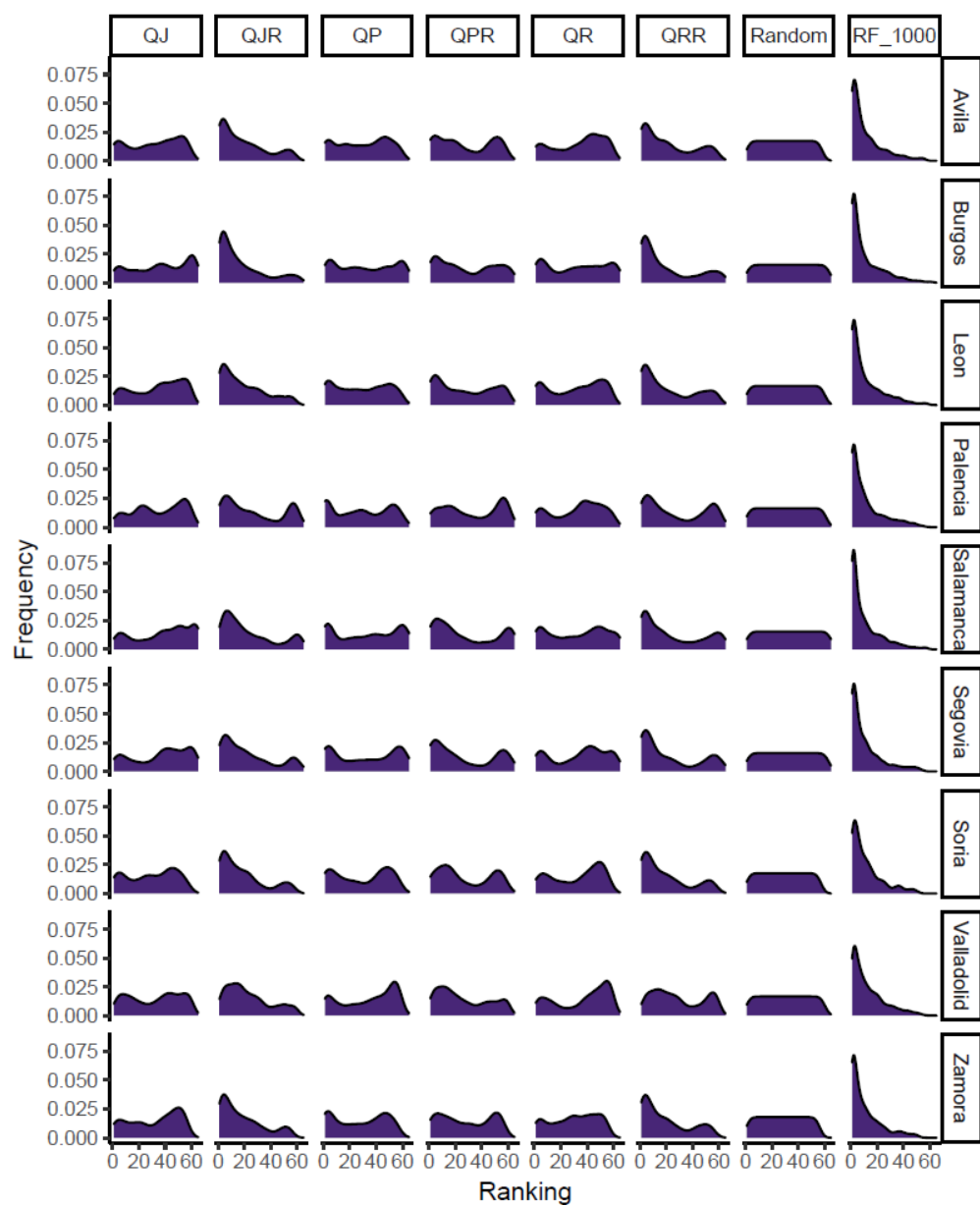
Caso de estudio II

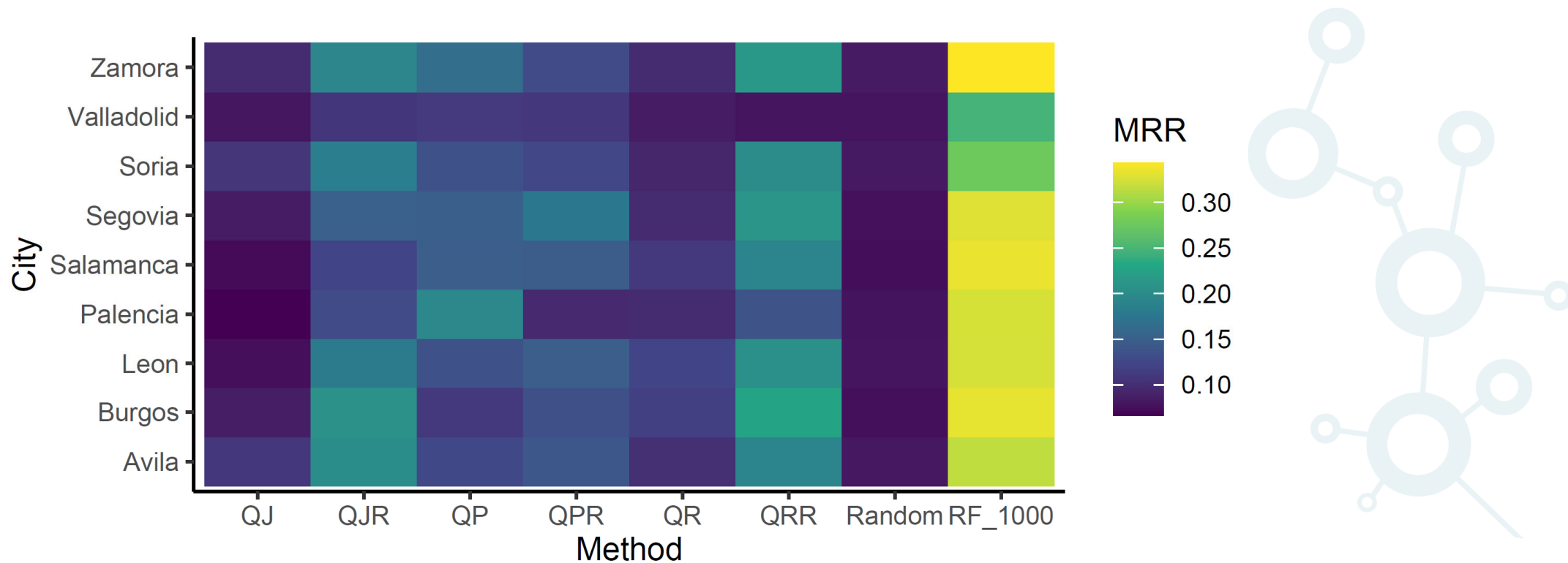


Para cada ciudad:

1. *Obtención de la red de interacciones entre comercios minoristas.*
2. *100-fold cross-validation:*
 - *En cada 99-fold training set obtener:*
 - *Las matrices de interacción entre categorías comerciales para Jensen, Permutation y Rewiring.*
 - *Los vectores de frecuencia de las distintas categorías comerciales.*
3. *Determinar la significación de las relaciones empíricas identificadas contra 250 iteraciones del modelo nulo correspondiente.*

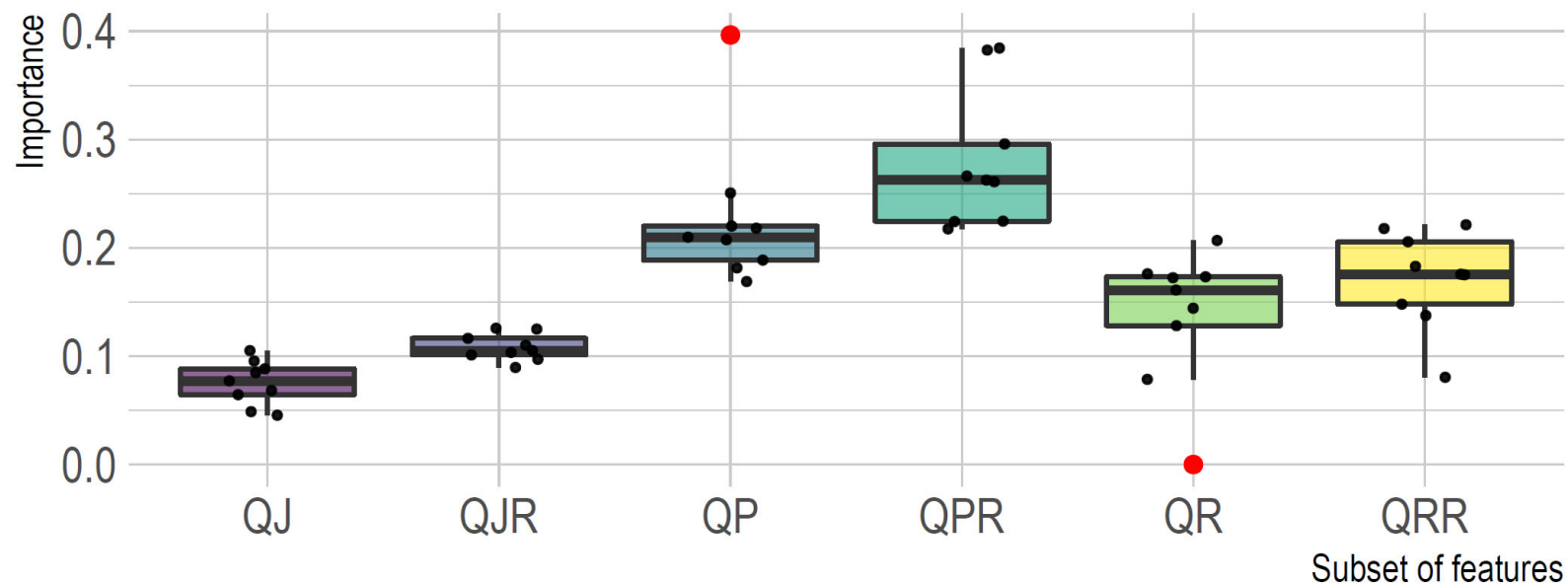






Mean Reciprocal Rank (MRR) → En Sistemas de Recomendación que devuelven un ranking de las categorías comerciales más adecuadas para una determinada localización, lo habitual es evaluarlos en base a la posición de la categoría real en el ranking.

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$



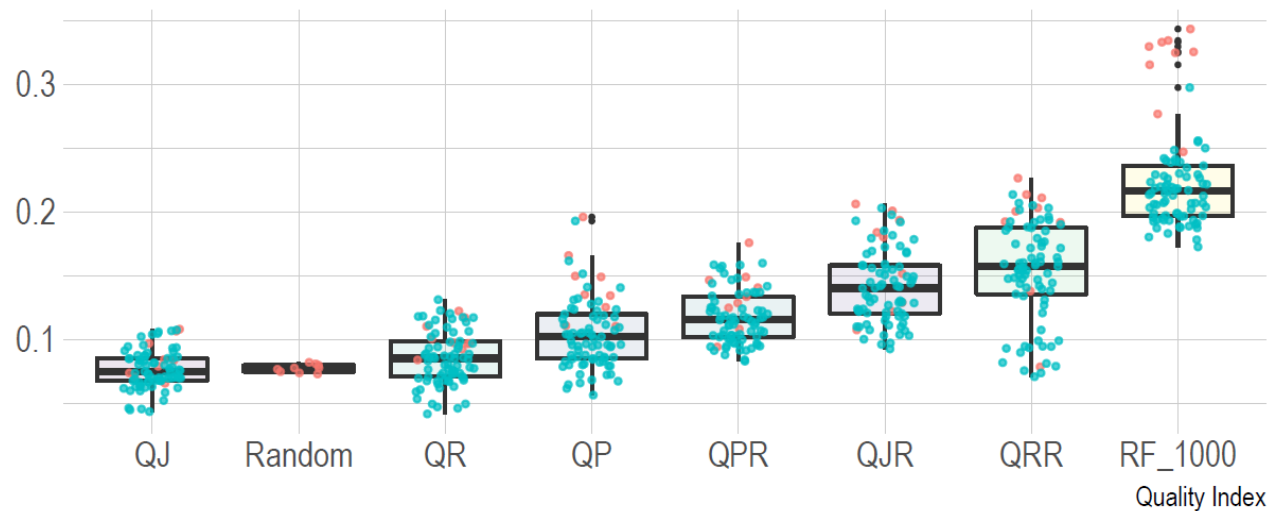
Análisis de importancia de las variables de grupo con Gini insesgado:

Los seis índices tienen una contribución significativa, lo que apoya la idea de complementariedad entre ellos.

TRANSFERENCIA CIUDAD A CIUDAD:

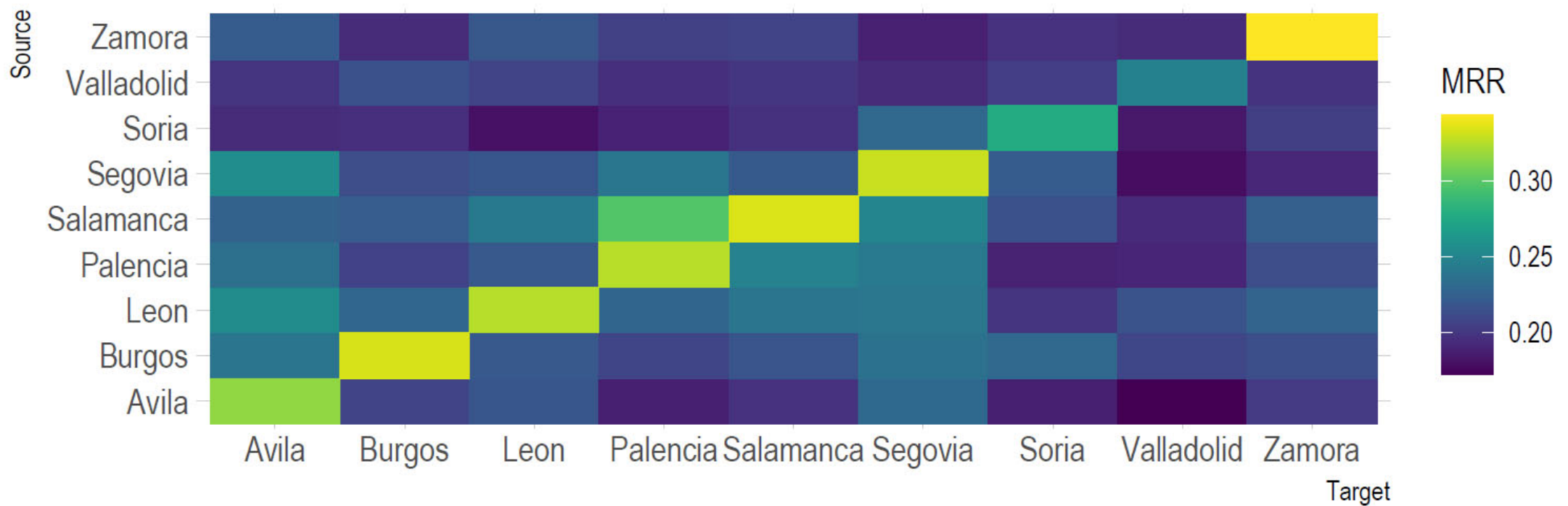
1. Potencial predictivo individual de cada índice de la ciudad origen con respecto a la ciudad objetivo.
2. Potencial predictivo conjunto de todos los índices de la ciudad origen (combinados en un *random forest*) con respecto a la ciudad objetivo.

MRR

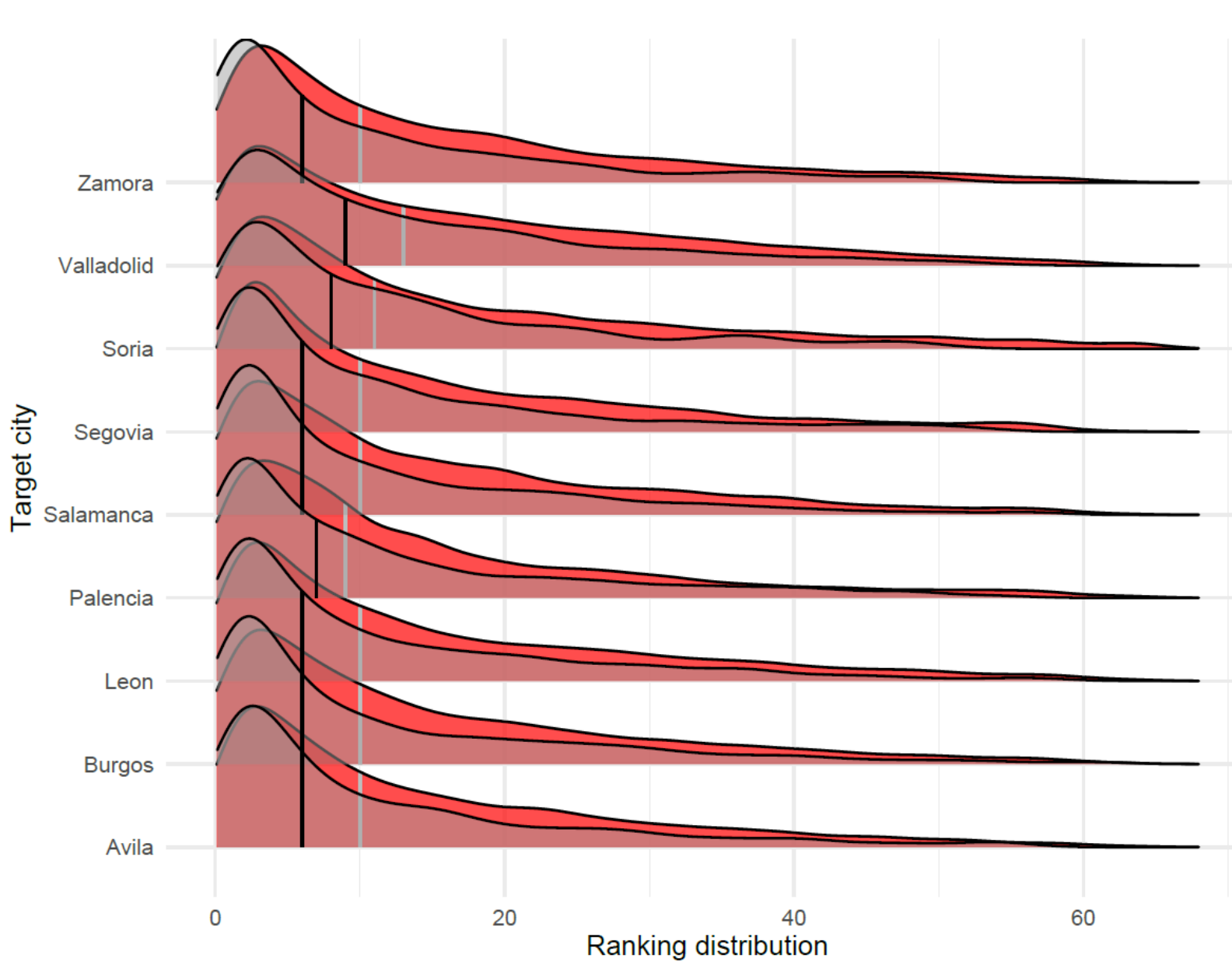


MRR obtenido para cada índice y para el RF que los combina al hacer transferencia (puntos azules) y cuando se utilizan los datos de la propia ciudad (puntos rojos).



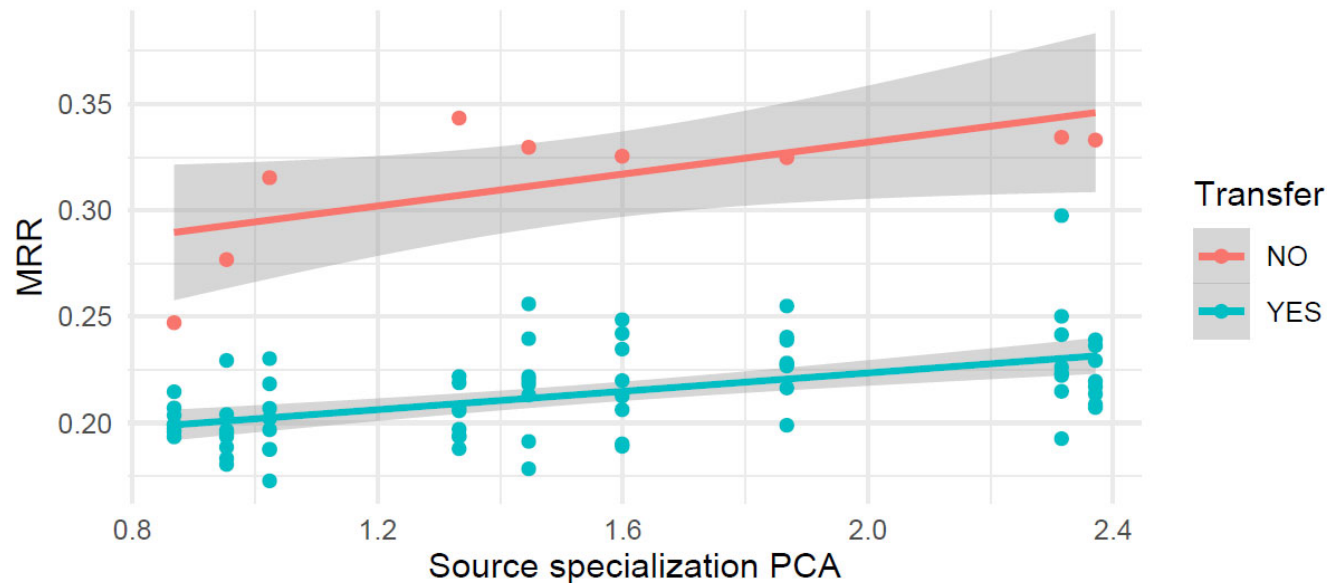


- MRR obtenido al combinar todos los índices de calidad de la ciudad origen mediante un *random forest* de 1000 árboles para hacer predicciones en otra ciudad.
- Los resultados obtenidos son sistemáticamente peores que cuando se usan los datos de la propia ciudad.



Distribuciones de ranking obtenidas cuando el *random forest* es entrenado sobre los datos de la propia ciudad (gris) y sobre datos transferidos (ciudad origen y objetivo no coincidentes).

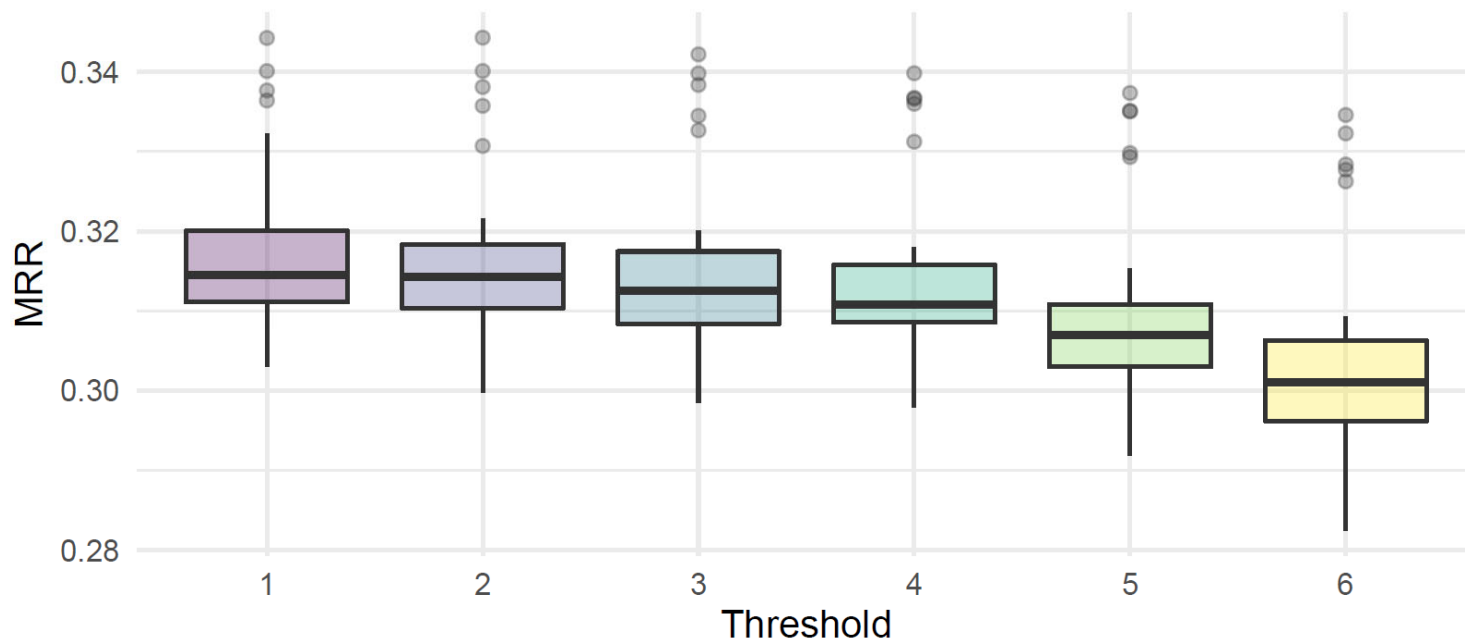
Se observa una pérdida sistemática de *performance* al utilizar datos de otra ciudad.



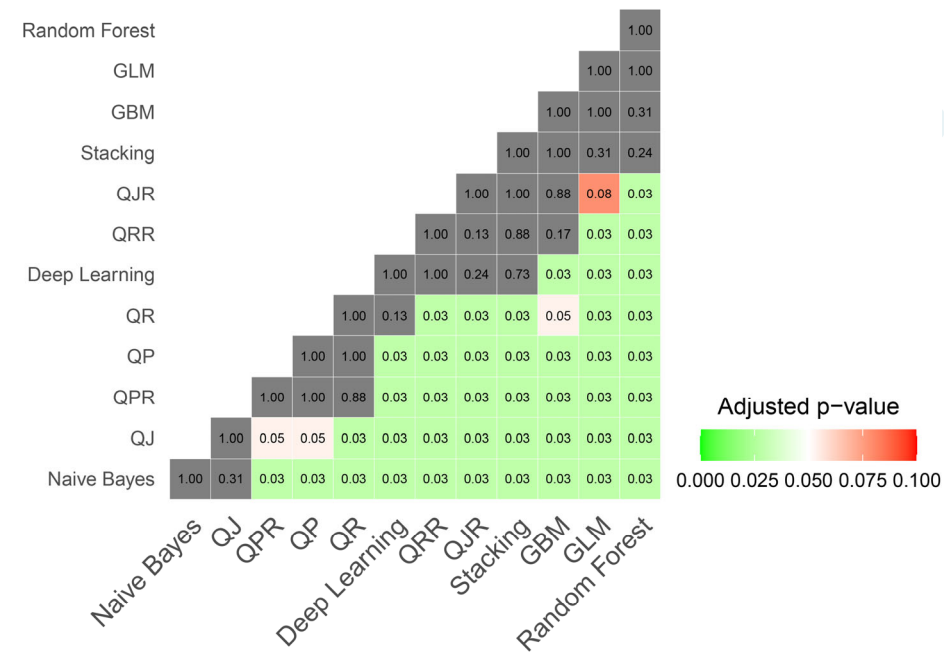
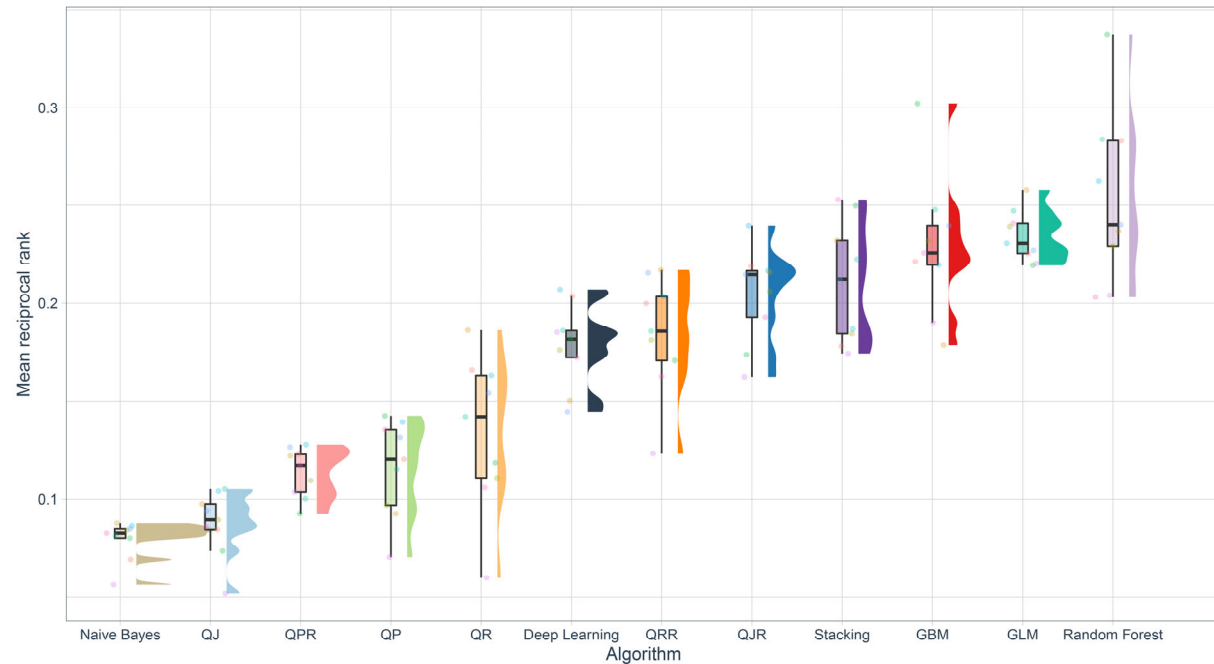
- Sólo hemos encontrado significativo el modelo aditivo.
- El nivel de especialización de la ciudad origen es significativo.
- El nivel de especialización de la ciudad objetivo sólo es significativo si el nivel de significación es 0,1.
- El efecto de la transferencia es significativo y se cuantifica en una disminución de 0,1 en el MRR.

Coefficients	Estimate	Std. Error	t value Pr(> t)
Intercept	0.266310	0.010914	24.402 < 2e-16 ***
Source	0.023346	0.004066	5.741 1.77e-07 ***
Target	0.008087	0.004066	1.989 0.0503 .
Transferred	-0.101069	0.006841	-14.773 < 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

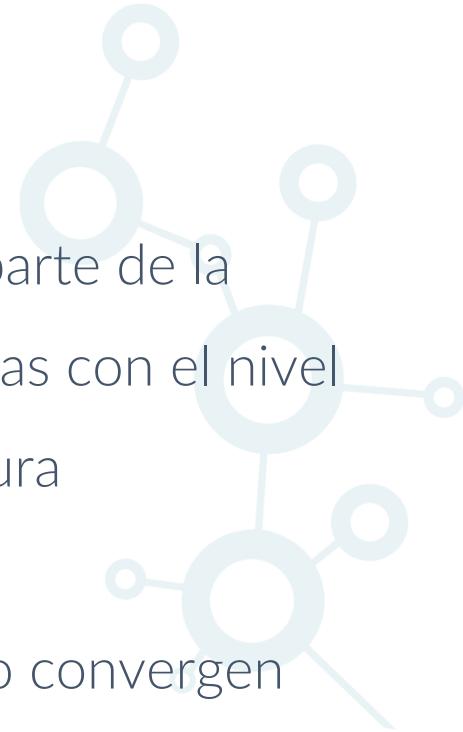


- En general, la aplicación de thresholds NO es útil en términos predictivos. Utilizar toda la información disponible es una mejor aproximación.
- A su vez, los enfoques consenso que incluyen también la información de la propia ciudad (como en esta figura) obtienen resultados sistemáticamente mejores.



Algunas conclusiones

- Los modelos nulos de Jensen, Permutation y Rewiring comparten parte de la información, pero presentan sus **propias particularidades** relacionadas con el nivel al que operan (**nodo/enlace**) y el grado en que preservan la estructura local/global.
- A medida que aumentamos el threshold, ambos enfoques consenso convergen hacia un **backbone (core robusto) de relaciones** entre categorías comerciales
- Existencia de una relación positiva entre el **tamaño de la ciudad y su nivel de especialización comercial**.
- Los índices **Quality Rewiring Raw** y **Quality Jensen Raw** usados individualmente capturan la estructura comercial de manera satisfactoria.



Algunas conclusiones

- No obstante, la capacidad predictiva aumenta sustancialmente cuando **combinamos todos los índices de calidad** en un modelo como el RF (o Stacking, GLM ó GBM)
- La transferencia de información comercial de una ciudad a otra se traduce en una **pérdida** de capacidad predictiva cuantificable en torno a un 0,1 en el MRR.
- El uso como **ciudades origen** de ciudades con un alto nivel de especialización comercial es **significativamente favorable** para la predicción.
- Las aproximaciones de transferencia mediante **redes consenso** de relaciones proporcionan resultados intermedios entre usar los datos de la propia ciudad y la mejor transferencia encontrada. Su rendimiento mejora si incluimos también la información de la propia ciudad.

Limitaciones



- No se han tenido en cuenta otros factores como preferencias culturales, precios de locales, densidad de población, etc.
- No todos los comercios aparecen en las Páginas Amarillas.
- No somos capaces de distinguir entre joining location y co-location
- Cierta homogeneidad cultural
- Efecto de otros elementos geográficos y POI
- Las distancias geodésicas igual no son las adecuadas en algunos casos
- Interpretabilidad de los modelos



Thank you



El problema de la ubicación de los comercios desde la perspectiva de las redes complejas

José Manuel Galán Ordax

