

# Minería de Medios Sociales

Máster en Ciencia de Datos e Ingeniería de Computadores

Bloque I: Redes Sociales y Ciencia de Datos en Redes

Sesión I.6: Modelado Basado en Agentes. Aplicaciones en Marketing

Oscar Cordón García

Dpto. Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada  
[ocordon@decsai.ugr.es](mailto:ocordon@decsai.ugr.es)

# **MODELOS DE COMPORTAMIENTO DE CONSUMIDOR CON ABMs Y REDES SOCIALES**

# EL MARKETING ES UN SISTEMA COMPLEJO

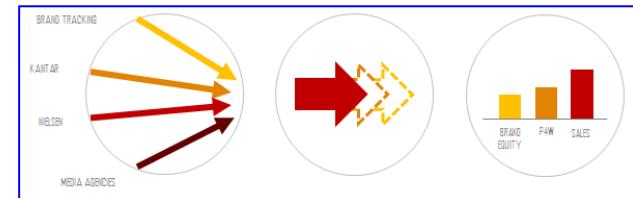
Los cambios en los mercados han alterado la forma en que éstos reciben y procesan la información:

- fragmentación continua de los medios publicitarios
- disponibilidad de mucha información descentralizada en las empresas
- nuevas dinámicas entre consumidores (Web 2.0, medios sociales)
- imposibilidad de modelar procesos complejos: WoM, marketing viral, etc.



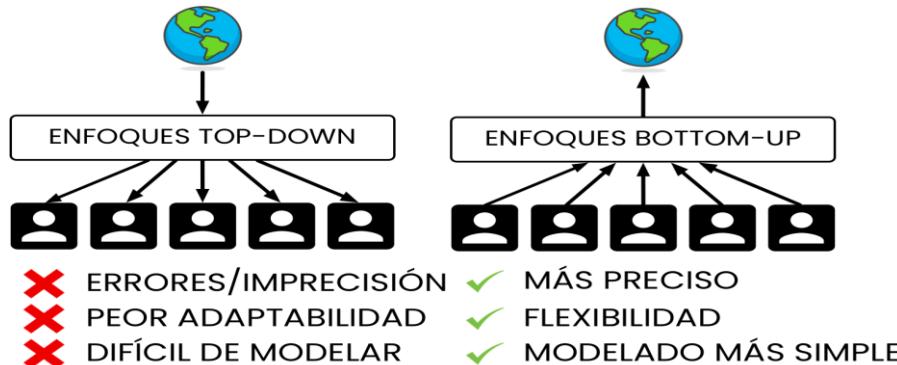
**Objetivo:** diseño de un **simulador** basado en ABMs y redes sociales que integre toda la información disponible (**tablero de mandos global**) y permita al director de marketing **definir y validar sus estrategias en un mercado virtual**, con un control absoluto

**MODELAR → SIMULAR → OPTIMIZAR**



# CARACTERÍSTICAS PRINCIPALES DE LOS MODELOS ABM

- Una técnica de **modelado descriptivo bottom-up**
- Un conjunto de agentes con comportamientos simples
- Pueden ser **heterogéneos** (consumidores, empresas, organismos, ...)
- Un sistema donde los agentes actúan e interactúan, habitualmente conectados a través de una **red social** que modela sus interacciones
- El comportamiento global del sistema se infiere de la agregación de todos los comportamientos individuales de los agentes



**PUEDEN SER ENFOQUES COMPLEMENTARIOS:**

Modelos Econométricos  
+  
ABMs

## EL MODELADO DEL COMPORTAMIENTO DEL CONSUMIDOR ES UNA TAREA CLAVE

<https://www.gartner.com/en/newsroom/press-releases/2022-10-10-gartner-identifies-seven-technology-disruptions-that-will-impact-sales-through-2027>

STAMFORD, Conn., October 10, 2022

### Gartner Identifies 7 Technology Disruptions That Will Impact Sales Through 2027

Generative AI, Machine Customers and AR/VR are Expected to Transform Sales in the Next Five Years

#### Digital Twin of the Customer

The **digital twin** of the customer is a dynamic virtual representation of a customer. It ingests data from personas and runs tests for sales messaging, sales processes and marketing campaigns to predict what does and doesn't work for that customer. The market for digital twin enabling software and services is expected to reach \$150 billion worldwide by 2030, up from \$9 billion in 2022.

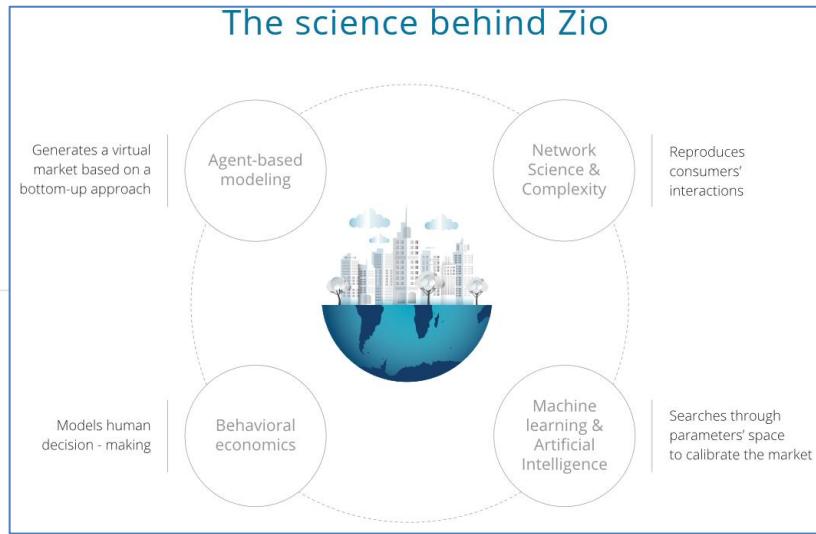
#### Digital Twin

A digital twin is a digital representation of a real-world entity or system. The implementation of a digital twin is an encapsulated software object or model that mirrors a unique physical object, process, organization, person or other abstraction. Data from multiple digital twins can be aggregated for a composite view across a number of real-world entities, such as a power plant or a city, and their related processes.

# ZIO: MODELADO Y SIMULACIÓN DE MERCADOS VIRTUALES



<https://www.zio-analytics.com/>



Desarrollado con reputados  
expertos de marketing y  
validada en empresas  
nacionales y multinacionales

# UN GEMELO DIGITAL DE COMPORTAMIENTO DE CONSUMIDOR...

Cada **agente consumidor** es un **elemento individual autónomo** con propiedades y comportamientos específicos o dependientes de su grupo, que modela a una serie de individuos reales **prototípicos**

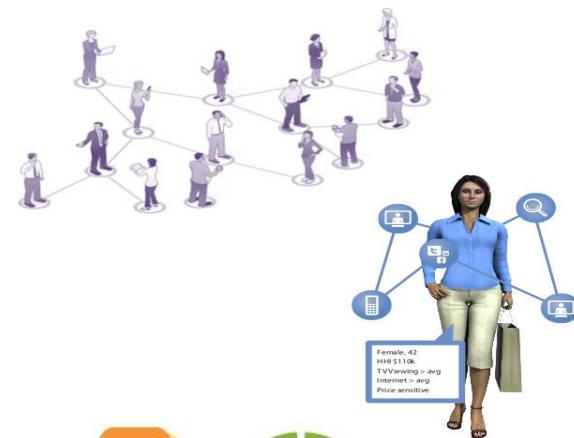
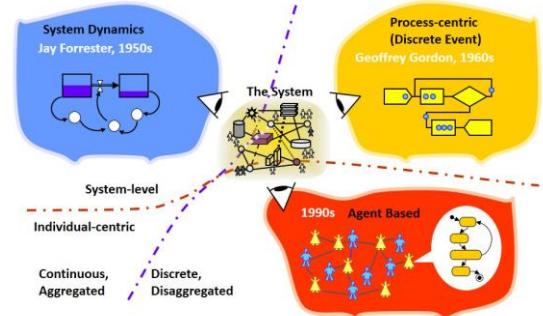
Los agentes consumidores deciden qué producto comprar de entre aquellos que conocen (**reconocimiento de marca**) en función de sus necesidades (**drivers de compra**) y de los **atributos** de éstos

La población está formada por miles de agentes con **comportamientos heterogéneos** modelando el mercado real

Se relacionan entre ellos (**red social de conversaciones**) y con el entorno (ej. con las **marcas** vía los **touchpoints**) y cambian su estado mental (reconocimiento, percepciones, ...)

Es un modelo **bottom-up** en el que no se diseña una dinámica global, sino micro, y la salida global (KPIs) se obtiene de la agregación de las acciones individuales (ej. **volumen de compras** o **Word of Mouth (WoM)**)

Generan patrones complejos que resultan de reglas simples (**comportamiento emergente**) y que no se pueden describir con un modelado clásico del sistema



# ZIO: MODELADO Y SIMULACIÓN DE MERCADOS VIRTUALES

## FLUJO DE TRABAJO EN ZIO



Delimitación del mercado | Configuración del mercado | Validación del mercado | Escenarios "what-if" | Optimización de estrategias

# ZIO: ABMs para marketing



## Comportamientos

Estacionalidad consumo  
Frecuencia de compra  
Frecuencia de uso



## Perfiles

Drivers  
Conocimiento  
Percepciones



## Redes

Conexiones  
Relaciones  
Influencia



## Touchpoints

Campañas  
Eficacia de los medios  
Alcance y frecuencia



## Decisión de compra

Racionales  
Emocionales  
Sociales



Sistema basado en agentes  
(ABMs)

Calibración y validación del sistema

Optimización de estrategias

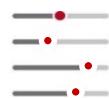
**Análisis automático de datos de simulación**

## ZIO basics: Consumidores

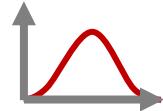
Se inicializan los consumidores en el mercado virtual especificando sus propiedades y comportamientos:



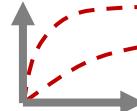
Segment 3



- Size
- Buying share



- Drivers
- Brand awareness
- Brand image



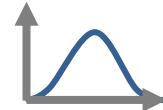
- Consumption frequency
- WOM activity
- Posting online



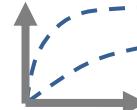
Segment 2



- Size
- Buying share



- Drivers
- Brand awareness
- Brand image



- Consumption frequency
- WOM activity
- Posting online



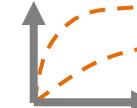
Segment 1



- Size
- Buying share



- Drivers
- Brand awareness
- Brand image



- Consumption frequency
- WOM activity
- Posting online

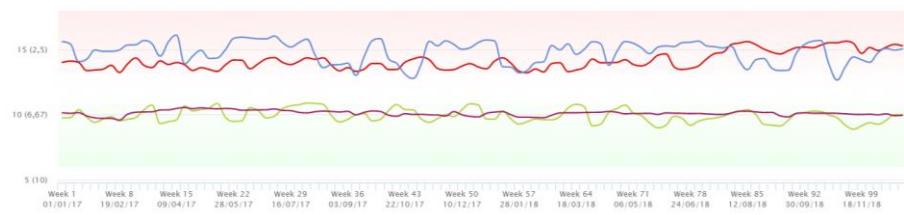
## ZIO basics: Marcas/productos

Para inicializar el mercado virtual se incorporan también las **propiedades y el comportamiento de las principales marcas** del sector en el periodo estudiado

- **Penetración inicial**



- **Experiencia del consumidor**

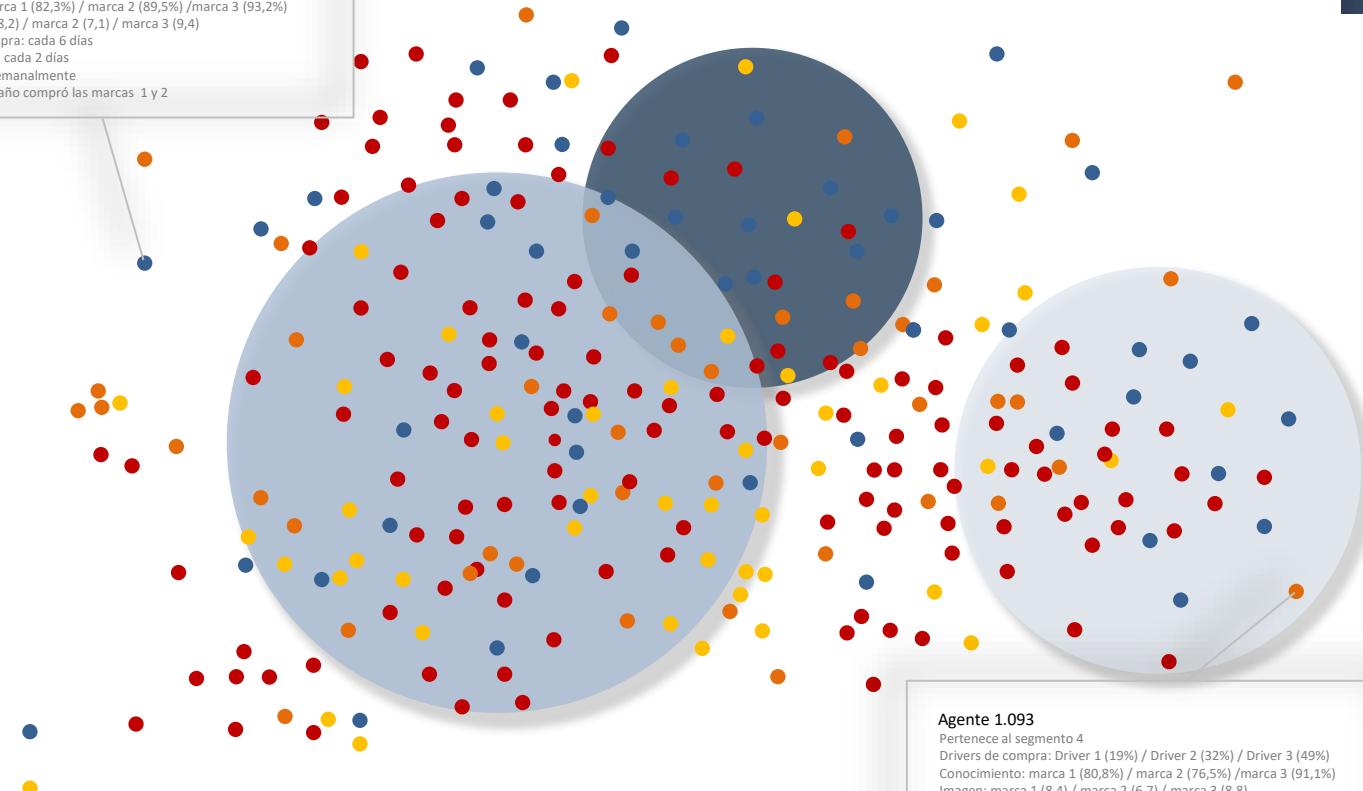


Product / Service availability

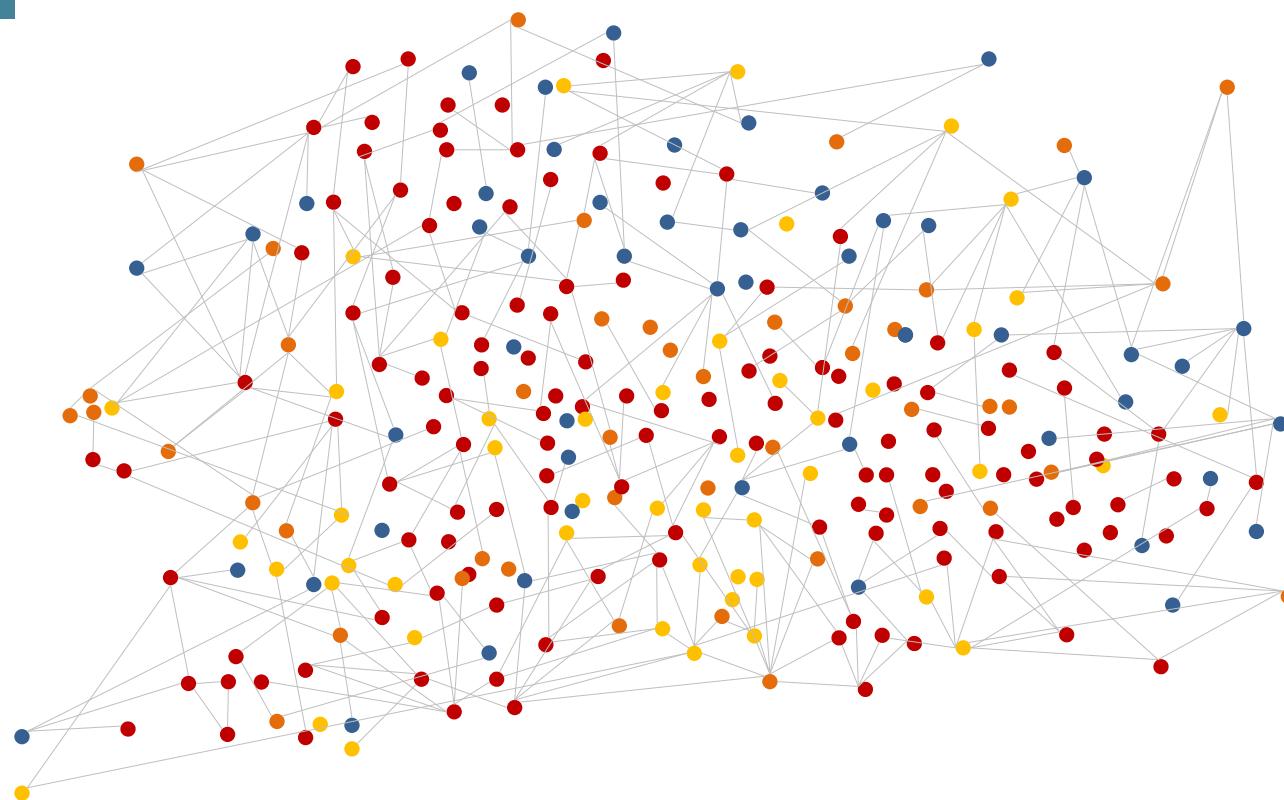


**Agente 6.744**  
 Pertenece al segmento 1  
 Drivers de compra: Driver 1 (22%) / Driver 2 (38%) / Driver 3 (40%)  
 Conocimiento: marca 1 (82,3%) / marca 2 (89,5%) / marca 3 (93,2%)  
 Imagen: marca 1 (8,2) / marca 2 (7,1) / marca 3 (9,4)  
 Frecuencia de compra: cada 6 días  
 Frecuencia de uso: cada 2 días  
 Lee blogs online semanalmente  
 Durante el último año compró las marcas 1 y 2

- Segmento 1
- Segmento 2
- Segmento 3
- Segmento 4
  
- Brand A
- Brand B
- Brand C



**Agente 1.093**  
 Pertenece al segmento 4  
 Drivers de compra: Driver 1 (19%) / Driver 2 (32%) / Driver 3 (49%)  
 Conocimiento: marca 1 (80,8%) / marca 2 (76,5%) / marca 3 (91,1%)  
 Imagen: marca 1 (8,4) / marca 2 (6,7) / marca 3 (8,8)  
 Frecuencia de compra: cada 3 días  
 Frecuencia de uso: cada 2 días  
 Lee blogs online cada 2 semanas  
 Durante el último año compró la marca 2



## ZIO basics: Touchpoints y Consumer journey

Las propiedades (**reconocimiento e imagen de marca**) y comportamientos (**probabilidad de hablar**) de los consumidores virtuales van cambiando por la acción de las marcas (**touchpoints de pago y ganados**) y por las interacciones entre los propios consumidores

El modelado de cada touchpoint de pago se realiza indicando la inversión semanal de cada marca (datos obtenidos de la empresa: GRPs, impactos o impresiones). Se puede especificar la creatividad de la campaña, que afecta según los drivers de los consumidores

Marketing plan Nidina | Total investment: 3.749.222€ | Share of voice: 35,13% | Initial penetration: 29%

Touchpoints: 12 | Flights: 33

Touchpoint Paid	Investment (€)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
TV (NZ)	677.836																								
TV (Brand)	399.600																								
Magazines (NZ)	9.258																								
Magazines (Brand)	114.426																								
Digital VOL (NZ)	303.626																								
Digital VOL (Brand)	425.722																								
Digital Display (NZ)	29.149																								
Digital Display (Brand)	760.290																								
Digital SEM (NZ)	1.000																								
Digital SEM (Brand)	547.607																								
Social (NZ)	61.528																								
Social (Brand)	428.182																								

Download creativity file | Load creativity file | Summary creativity

Creativity

Average creativity values for the whole calendar

Touchpoint	Taste	Price	Trust in Brand	Brand Expertise	Recommendation (Pediatric)
TV (NZ)	Emphasis: 20	20	20	20	20
	Quality: 50	50	50	50	50
TV (Brand)	Emphasis: 20	20	20	20	20
	Quality: 50	50	50	50	50
Magazines (NZ)	Emphasis: 20	20	20	20	20
	Quality: 50	50	50	50	50
Magazines (Brand)	Emphasis: 20	20	20	20	20
	Quality: 50	50	50	50	50
Digital VOL (NZ)	Emphasis: 20	20	20	20	20
	Quality: 50	50	50	50	50
Digital VOL (Brand)	Emphasis: 20	20	20	20	20
	Quality: 50	50	50	50	50

Customer Journey

1st Jan OOH Magazine Buy Consumer experience WOM 31 Dec

Compras del agente

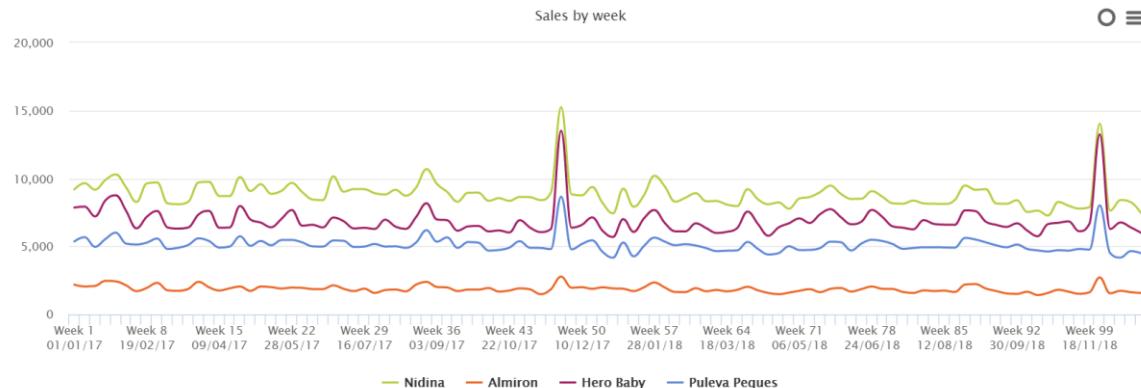
Customer Journey

1st Jan OOH Magazine Buy Consumer experience WOM 31 Dec

Imagen de marca del agente

## ZIO basics: proceso de compra

Se incorporan al modelo el histórico de ventas de cada marca durante el periodo simulado, incluyendo la estacionalidad correspondiente:



En la simulación se realizan exactamente ese nº de ventas en el momento concreto por parte de los consumidores virtuales

Los eventos de compra se asignan a los consumidores pero cada uno decide qué producto concreto comprar según sus propiedades (reconocimiento e imagen de marca, peso de sus drivers de compra) y una serie de [heurísticas realistas de compra](#)

[El ABM es un modelo descriptivo de la dinámica del mercado no de predicción de ventas.](#) Una vez validado, se puede utilizar como herramienta predictiva con los escenarios what-if y la optimización de estrategias



Social Network

Listen  
WoM

Talk  
WoM

Ad  
Impact

- Investment
- Creativity (emphasis / quality)
- Total reach
- Reach speed

Paid touchpoints for each brand

TV	Radio
DOOH	Cinema



Read  
Post  
Online

Posts  
repository

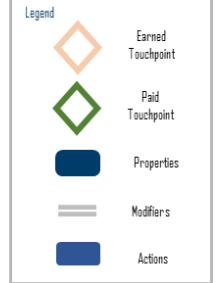
Write  
Post  
Online

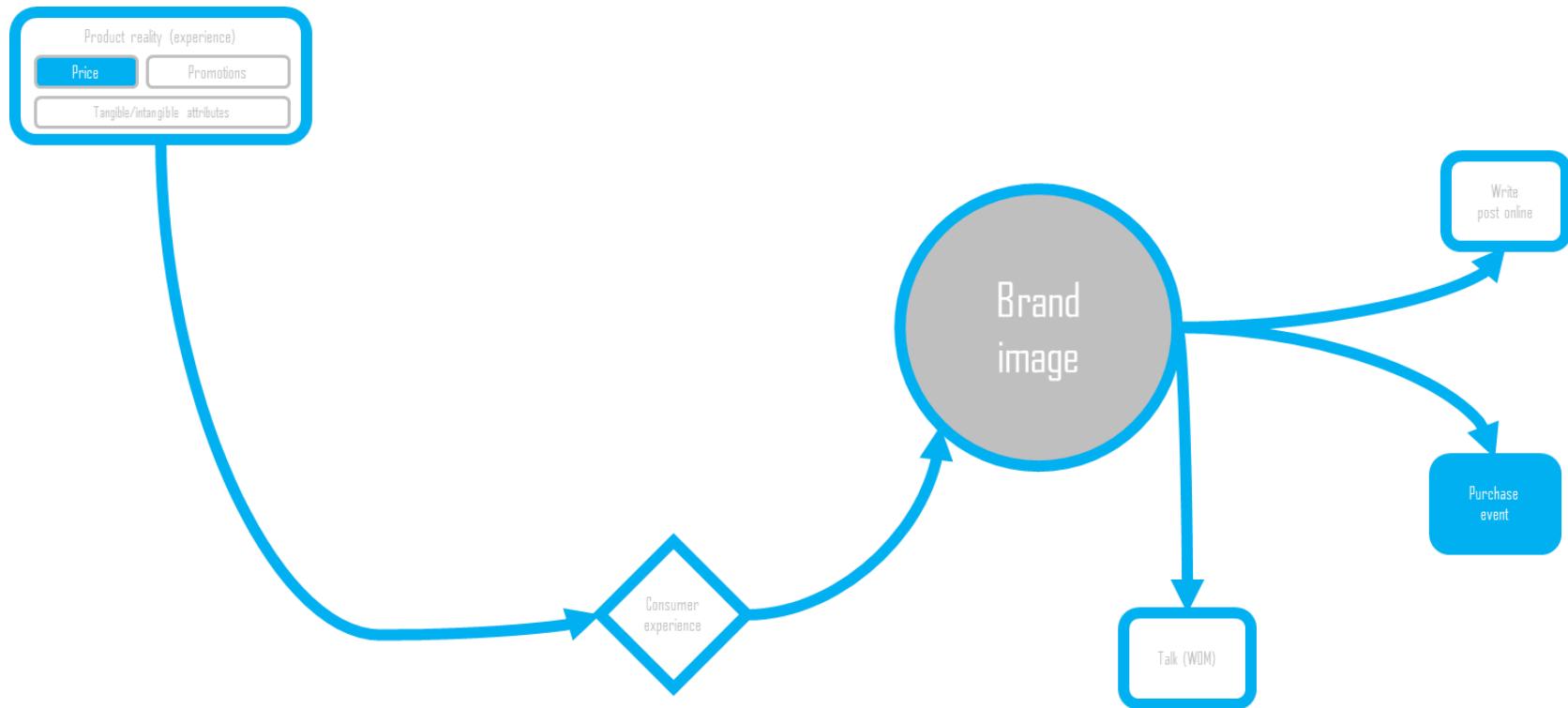
Drivers

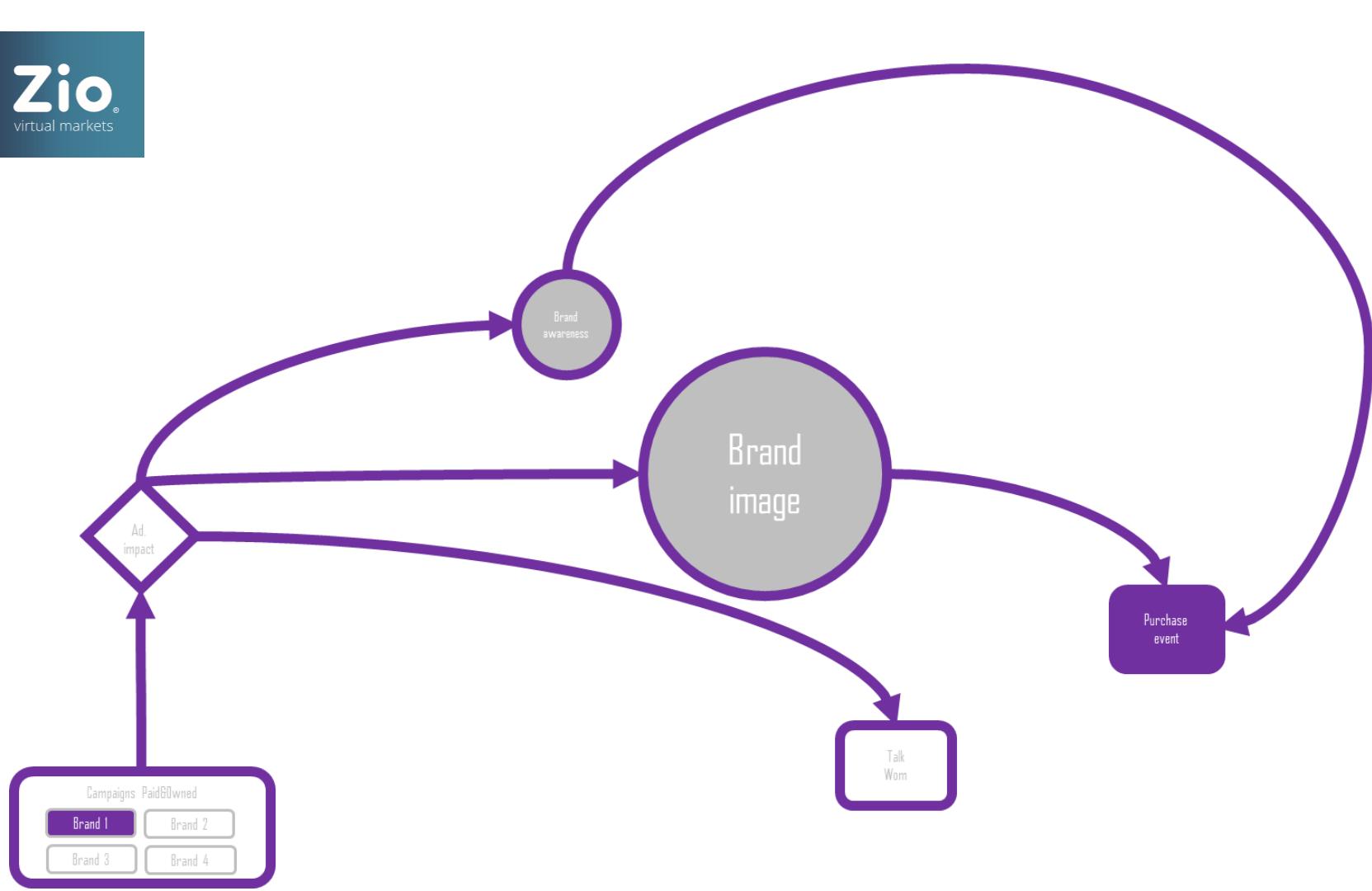
Market Seasonality

Purchase  
events

Consumer experience (promotions,  
product or service reality)

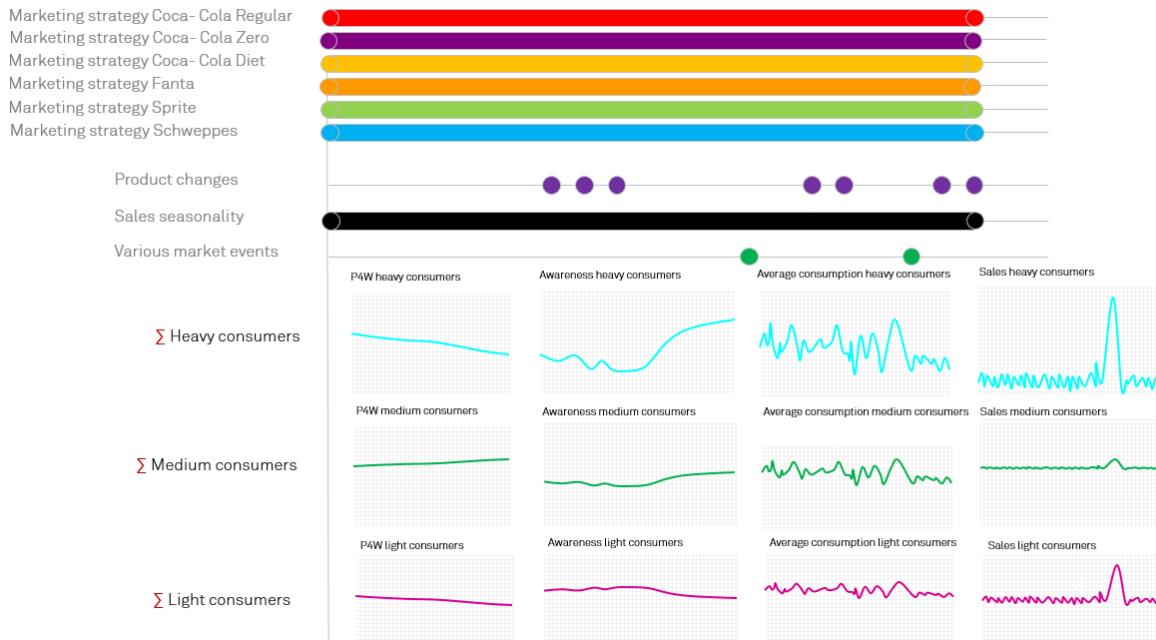




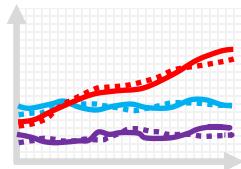


## ZIO basics: Cálculo de los KPIs

Para obtener los valores globales de los KPIs en el mercado virtual, se agregan los valores de la población de consumidores



# CALIBRACIÓN Y VALIDACIÓN DEL MODELO



Ventas

- Marca 1
- Marca 2
- Marca 3

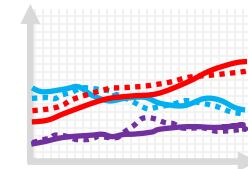
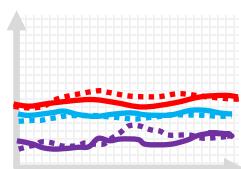
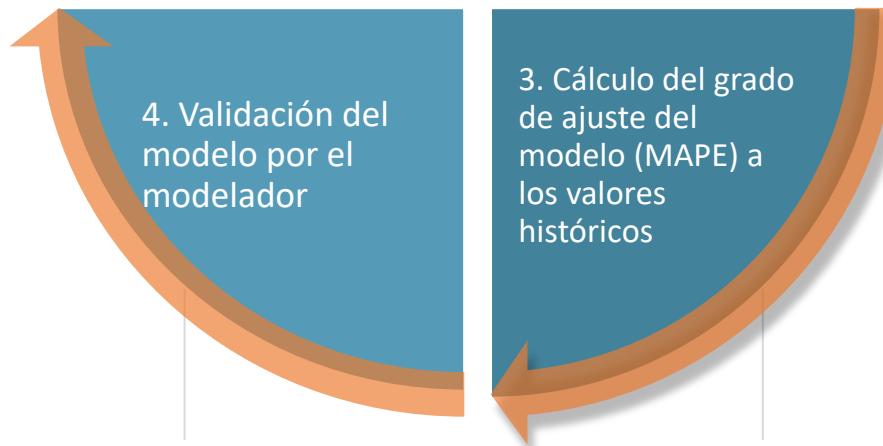


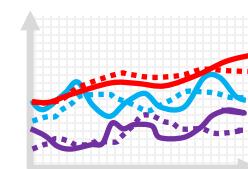
Imagen de marca

- Marca 1
- Marca 2
- Marca 3



Reconocimiento de marca

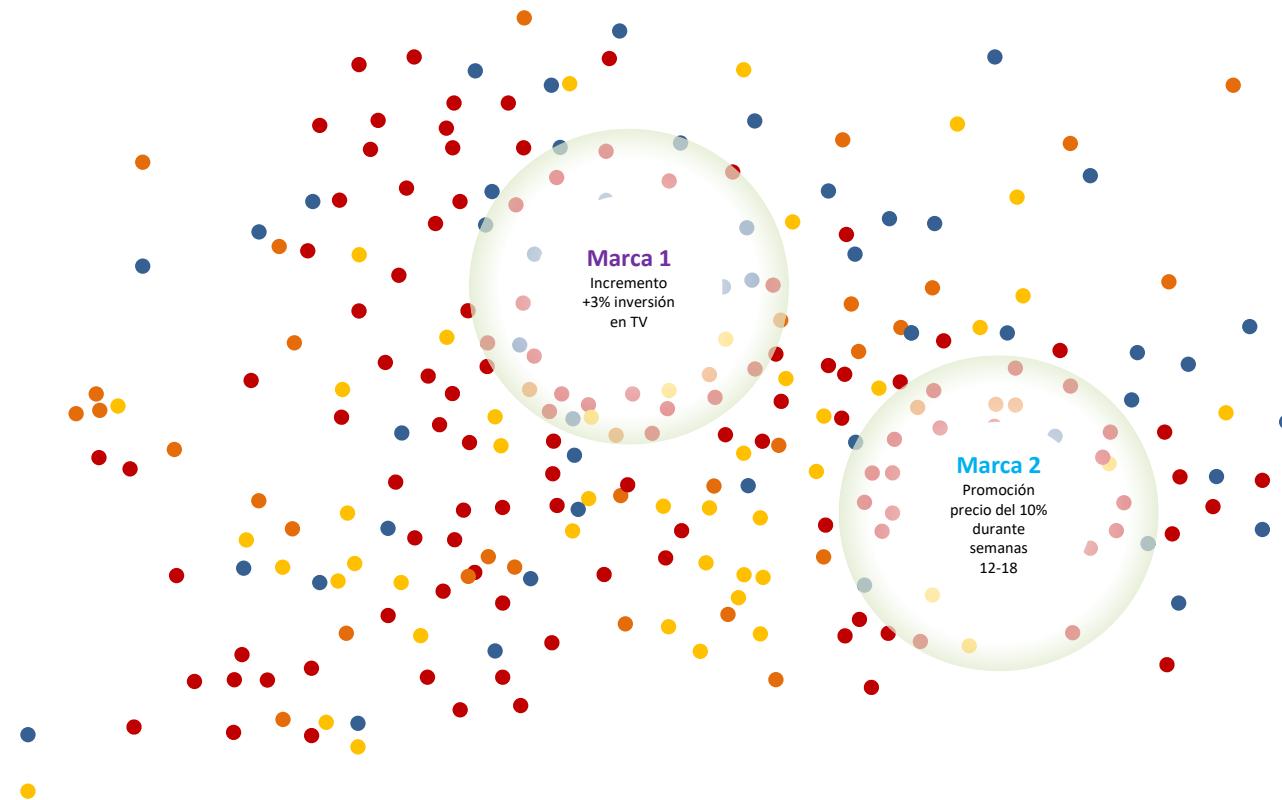
- Marca 1
- Marca 2
- Marca 3



Word of Mouth

- Marca 1
- Marca 2
- Marca 3

## ESCENARIOS WHAT-IF



# PRODUCTO COMERCIAL: ZIO ANALYTICS (2)



# CASO DE ESTUDIO ZIO: Optimización del Mix de una Marca de Perfumes de Lujo

Mercado **muy competitivo** y **altamente estacional** en el que la publicidad juega un rol fundamental

**OBJETIVO:** Entender la dinámica del mercado con respecto a la competencia y a las marcas propias. **Optimizar el mix de medios de las 5 marcas propias para aumentar las ventas**

## Market delimitation

Target: Spanish women 25-64 yo.

Period modelled: 2017 – 2018.

## Brands (12)

### Brands Client's

- Brand 1
- Brand 2
- Brand 3
- Brand 4
- Brand 5

### Competitors'

- Armani
- Carolina Herrera
- Chanel
- Gucci
- Lancome
- Paco Rabanne
- YSL

## Purchase drivers (6)

- Scent and Permanence
- Brand image
- Quality / Price
- Promotions / Gifts
- Recommendations
- Packaging

## Customer segments (2)

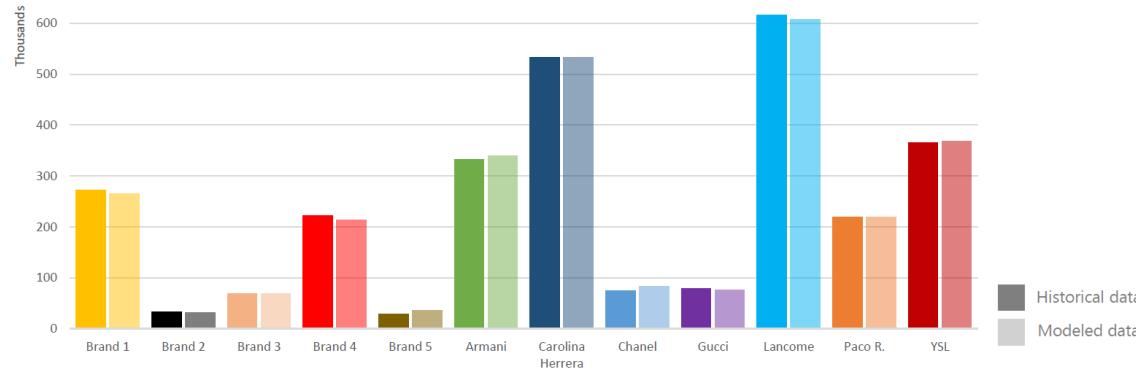
- Women 25 – 44
- Women 45 – 64

## Touchpoints (10)

- TV <= 20" – Night
- TV <= 20" – Other
- TV > 20" – Night
- TV > 20" – Other
- Cinema
- Outdoors
- Print
- Radio
- Video Online
- Display

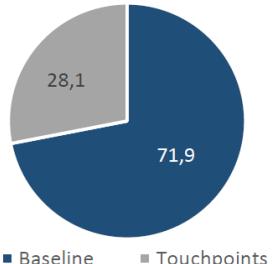
# CASO DE ESTUDIO ZIO: Optimización del Mix de una Marca de Perfumes de Lujo

**CALIBRACIÓN DEL MODELO:** Ajuste con un **98,44%** de precisión en las **Ventas Globales**, obtenido usando el KPI Items vendidos:

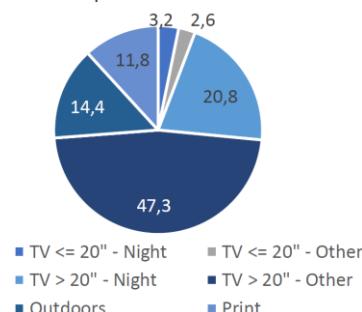


## DIAGNÓSTICO DE LA MARCA 1: Estudio de la contribución de los touchpoints:

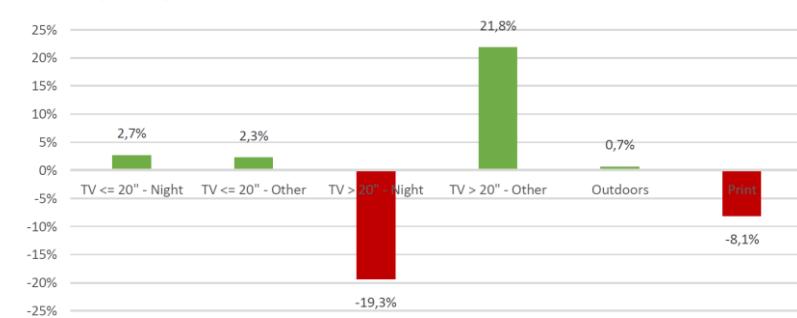
Baseline sales vs Media contrib.



Touchpoint contribution

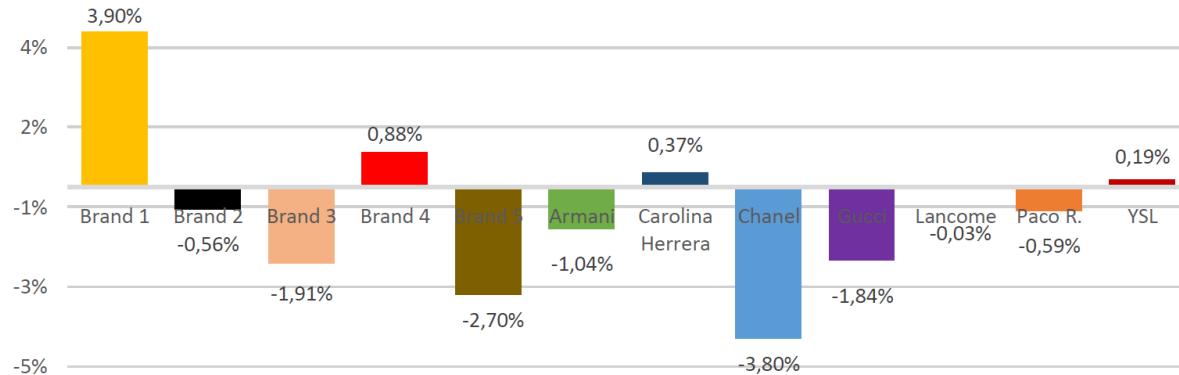


Efficiency analysis



# CASO DE ESTUDIO ZIO: Optimización del Mix de una Marca de Perfumes de Lujo

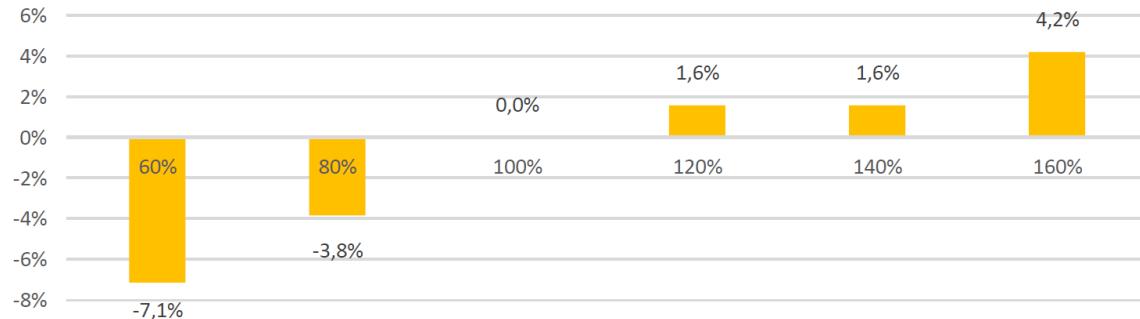
**ESCENARIOS WHAT-IF Marca 1:** Análisis de ganadores y perdedores con un nuevo mix de medios:



## RESULTADO FINAL:

Comprensión de la dinámica del mercado para todas las marcas. **Potencial incremento de ventas del 7% cambiando el mix de medios de las 5 marcas propias**

**Análisis de sensibilidad de la inversión en medios**  
del escenario para saber la **elasticidad de la inversión**:



# ANÁLISIS DE INFLUENCIA DE LA COMUNICACIÓN EN PROCESOS ELECTORALES USANDO ABMs Y REDES SOCIALES

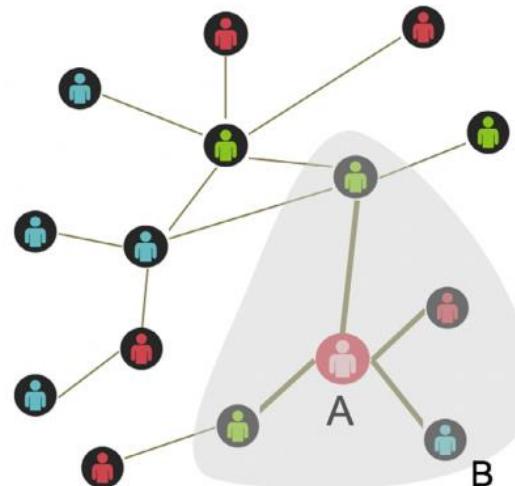
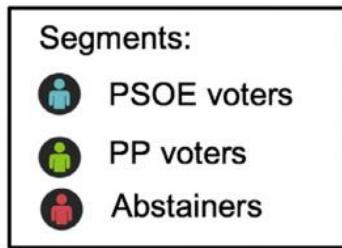
## OBJETIVOS DEL MODELO

- Analizar el impacto en los votantes de la información del atentado (AlQaeda vs ETA) proporcionada por los medios entre el 11 y el 14 de Marzo de 2004 considerando la **Teoría del Marco**
- Modelar las principales opciones políticas, la red social de votantes y la cobertura de medios (usando *touchpoints*) con un ABM
- Calibrar los parámetros del modelo a partir de los informes del CIS y de los resultados reales de las elecciones generales del 14M
- Evaluar distintos escenarios *what-if* para la cobertura de los medios y los efectos del “boca a oído” (*Word of Mouth*, WoM) sobre los votantes

# ESTRUCTURA GENERAL DEL MODELO



(1) Generates message about attacks  
with a given polarization



*Scale-free social network of voting  
agents with a specific density*

(2) Message is received by agent  
'A' and modifies its resilience state

(3) 'A' talks to its neighbor 'B'  
about the message and  
modifies its resilience state

## COMPONENTES DEL MODELO (1)

24,109 agentes (ratio 1 votante virtual  $\leftrightarrow$  1.212 votantes reales) para un periodo de ejecución de 72 pasos (1 paso = 1 hora) entre el 11 de Marzo (atentado de Madrid) y el 14 de Marzo (elecciones generales)

**Opciones de voto:** PP, PSOE, Abstención

**Medios (= touchpoints):** modelados mediante la polarización del mensaje  $m_c \in [-2, 2]$  (-2 si apunta a AlQaeda, 2 a ETA y 0 si es totalmente neutral)

El etiquetado se realiza de forma manual por un grupo de expertos en comunicación y polítólogos

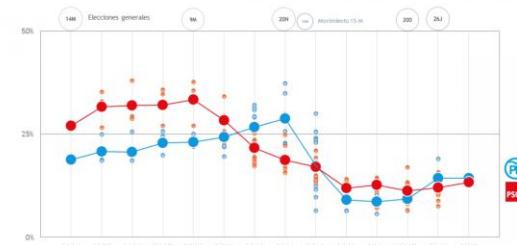
Cada *touchpoint* tiene un **alcance (reach)** (datos reales) y unos parámetros de **influencia en los votantes** y **buzz** fijados por un calibrador automático

Selected mass media channels	
Channel name	Type
El País	Press
El Mundo	Press
ABC	Press
Cadena Ser	Radio
TVE	Television
Antena 3	Television
Telecinco	Television

## COMPONENTES DEL MODELO (2)

**Votantes virtuales (agentes):** Agrupados en segmentos para definir sus parámetros/estado mental iniciales: (S1) votantes iniciales del PP, (S2) votantes iniciales del PSOE, (S3) abstencionistas iniciales

El declive del bipartidismo PP-PSOE entre la gente joven\*  
En porcentajes de intención de voto a PP y PSOE en promedio



El estado mental de los votantes se basa en **su resistencia ( $\mu_i$ )**, una variable real de la Teoría del Marco definida en  $[0,10]$ , que cambia durante la simulación vía los *touchpoints* o el WoM

En el paso final de la simulación los agentes votan de forma determinística según el valor de su variable resistencia ( $0 \rightarrow$ PSOE,  $1 \rightarrow$ ABS,  $2 \rightarrow$ PP):

$$v_A(\mu_A(t)) = \begin{cases} 0, & \text{if } \mu_A(t) \in [0, 3.3), \\ 1, & \text{if } \mu_A(t) \in [3.3, 6.7), \\ 2, & \text{if } \mu_A(t) \in [6.7, 10]. \end{cases}$$



## COMPONENTES DEL MODELO (3)

Los agentes están conectados por una red social (topología **libre de escala**) para difundir sus opiniones políticas a sus vecinos directos (**WoM**) con una cierta probabilidad



Cuando un agente  $A$  habla con otro agente  $B$ , el valor de resistencia de  $B$  cambia en función del estado de  $A$  y de la influencia de su segmento ( $\Delta_A$ ):

$$\mu_B(t+1) = \mu_B(t) + |\mu_B(t) - \mu_A(t)|\Delta_A.$$

Se aplica un mecanismo para olvidar la influencia de los demás a lo largo del tiempo (*decay*) :

$$\delta_A(t) = \sum_{i=1}^{i=t} (\mu_{A_i} - \mu_{A_{i-1}}). \quad \mu_A(t+1) = \mu_A(t) - (\delta_A(t)d\Delta_A).$$

# PROCESO DE CALIBRACIÓN Y VALIDACIÓN (1)

Todo modelo necesita ser validado para ser útil. Usamos un mecanismo de calibración automático para **ajustar los 44 parámetros desconocidos del modelo de acuerdo a los resultados finales de las elecciones (datos reales)**:

List of parameters to be automatically calibrated by the memetic algorithm.

---

Parameters to be calibrated (44 parameters)

---

Mass media (35 in total)				
	El País	El Mundo	ABC	Cadena Ser
Buzz	$(\tau, d\tau)$	$(\tau, d\tau)$	$(\tau, d\tau)$	$(\tau, d\tau)$
Influence	$(\Delta_{max}, \Delta, d\Delta)$	$(\Delta_{max}, \Delta, d\Delta)$	$(\Delta_{max}, \Delta, d\Delta)$	$(\Delta_{max}, \Delta, d\Delta)$

---

	TVE	Antena 3	Telecinco
Buzz	$(\tau, d\tau)$	$(\tau, d\tau)$	$(\tau, d\tau)$
Influence	$(\delta_{max}, \Delta, d\Delta)$	$(\delta_{max}, \Delta, d\Delta)$	$(\delta_{max}, \Delta, d\Delta)$

---

**WOM process (3 per segment, 9 in total)**

---

Initial talking prob.	$(p_{PSOE}(0), p_{ABST}(0), p_{PP}(0))$
Influence change	$(\Delta_{PSOE}, \Delta_{ABST}, \Delta_{PP})$
Influence decay	$(d\Delta_{PSOE}, d\Delta_{ABST}, d\Delta_{PP})$

---

## PROCESO DE CALIBRACIÓN Y VALIDACIÓN (2)

La calibración se realiza con una técnica de IA denominada algoritmo evolutivo. La función de ajuste mide la desviación entre la salida del modelo y los valores históricos (SMAPE)

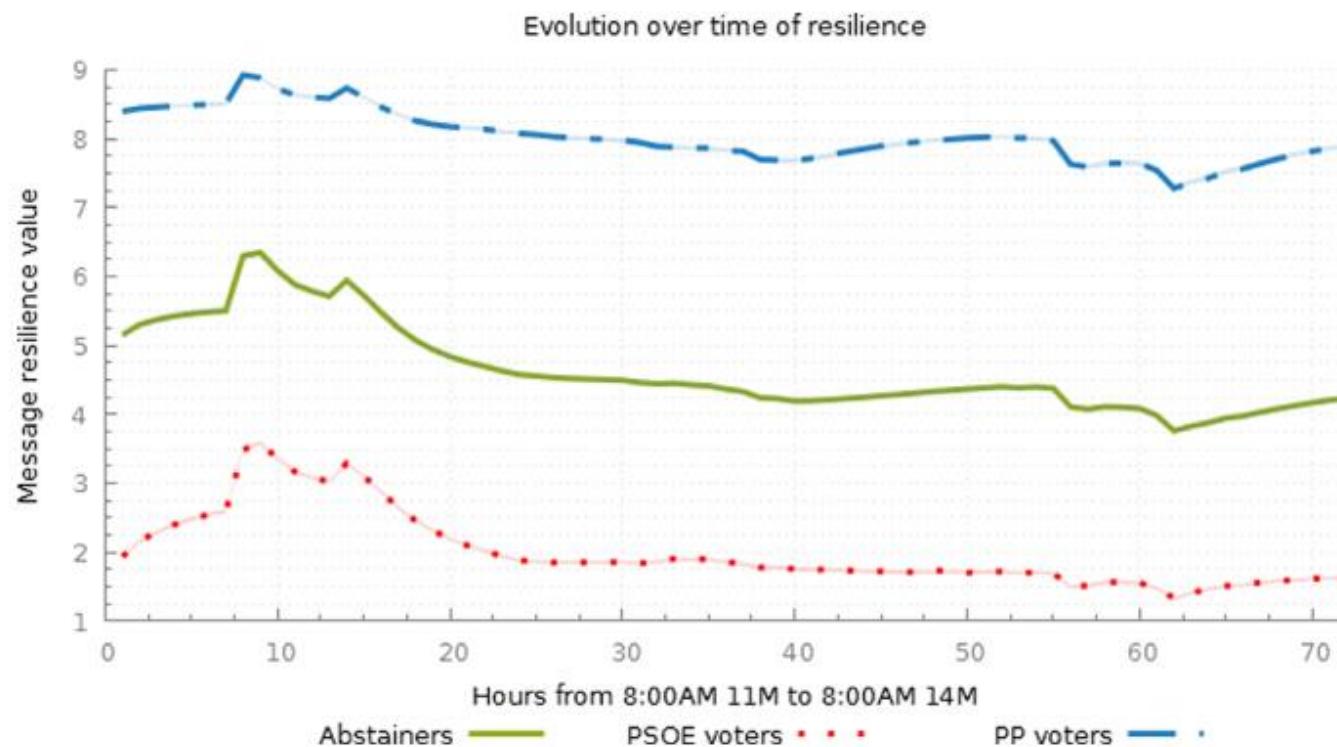
Fitting values of the calibrated model and three additional model variation scenarios.

Party Votes	Real data Election results	Models			
		Complete	No media	No diffusion	No influence
PSOE	11,026,163	11,020,144	10,329,618	10,259,170	9,941,145
PP	9,763,144	9,766,804	10,577,439	10,948,887	11,403,078
Abstention	8,449,355	8,451,711	8,331,602	8,030,602	7,894,438
% Total votes					
PSOE	37.71%	37.69%	35.33%	35.09%	34%
PP	33.39%	33.4%	36.18%	37.45%	39%
Abstention	28.9%	28.91%	28.5%	27.47%	27%
Global fitting		99.13%	84.49%	77.06%	68.39%

Comparamos distintas variantes del modelo para mostrar la importancia de cada componente

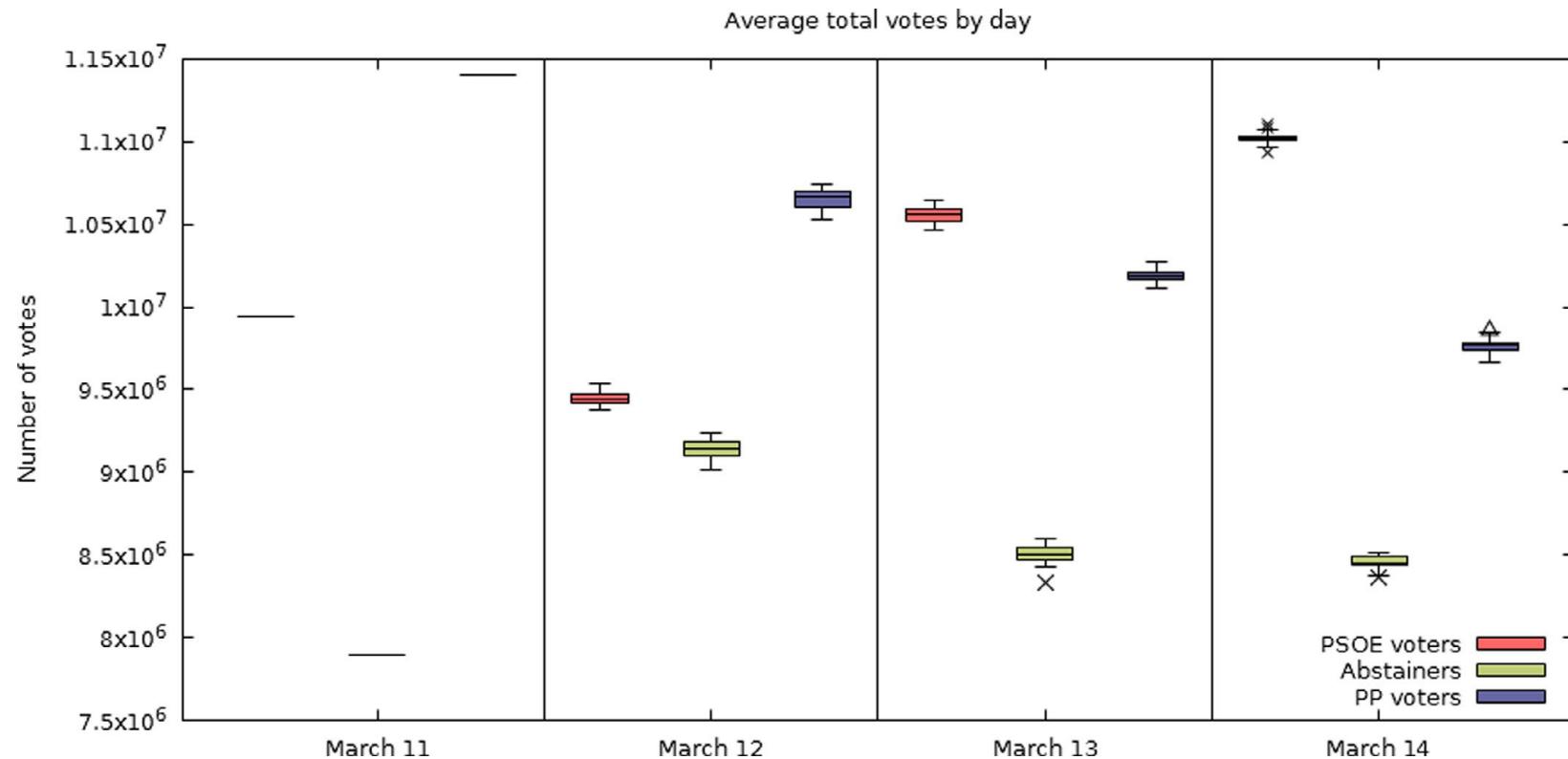
## RESULTADOS (1)

Una vez validado el modelo, podemos analizar la **variación del valor de resistencia de cada segmento** (PP, PSOE, Abstención) durante el periodo simulado:



## RESULTADOS (2)

Esa variación de la resistencia se refleja directamente en la **variación de la intención de voto** y en los resultados finales de las elecciones (box-plots de 30 ejecuciones de Montecarlo):



## RESULTADOS (3)

También se puede explorar el volumen del WoM (**volumen de conversaciones**) dependiente del momento de la simulación y de los eventos ocurridos:

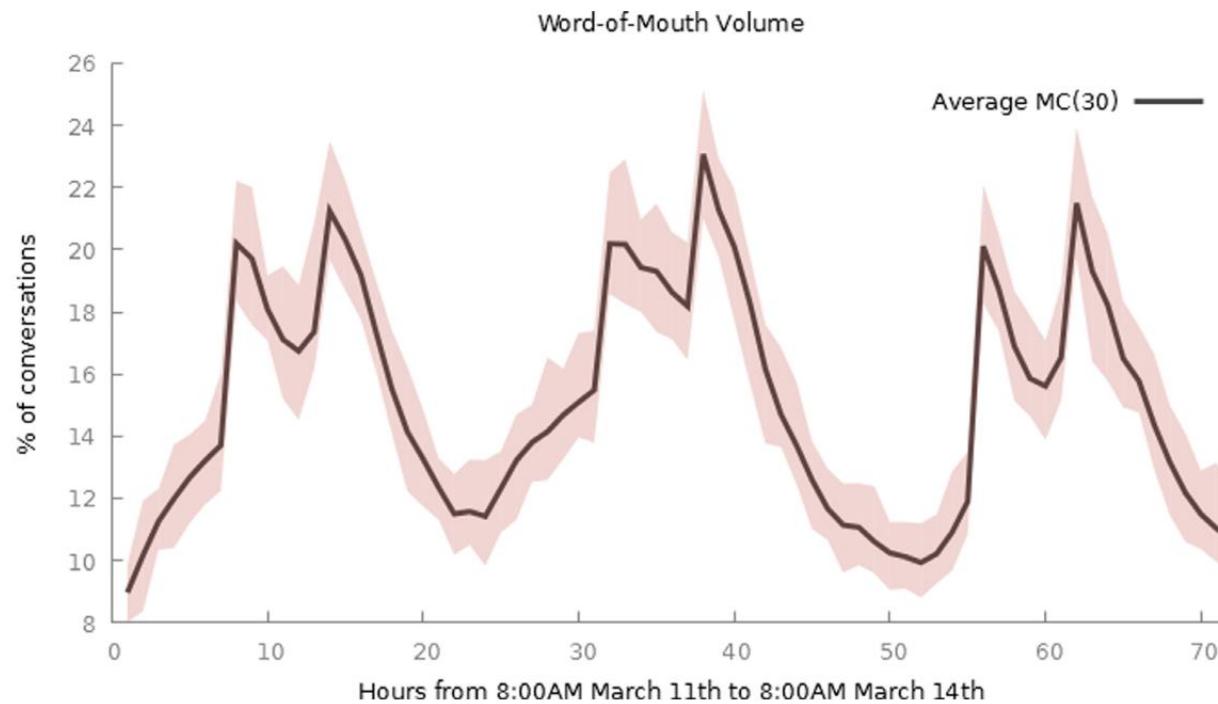
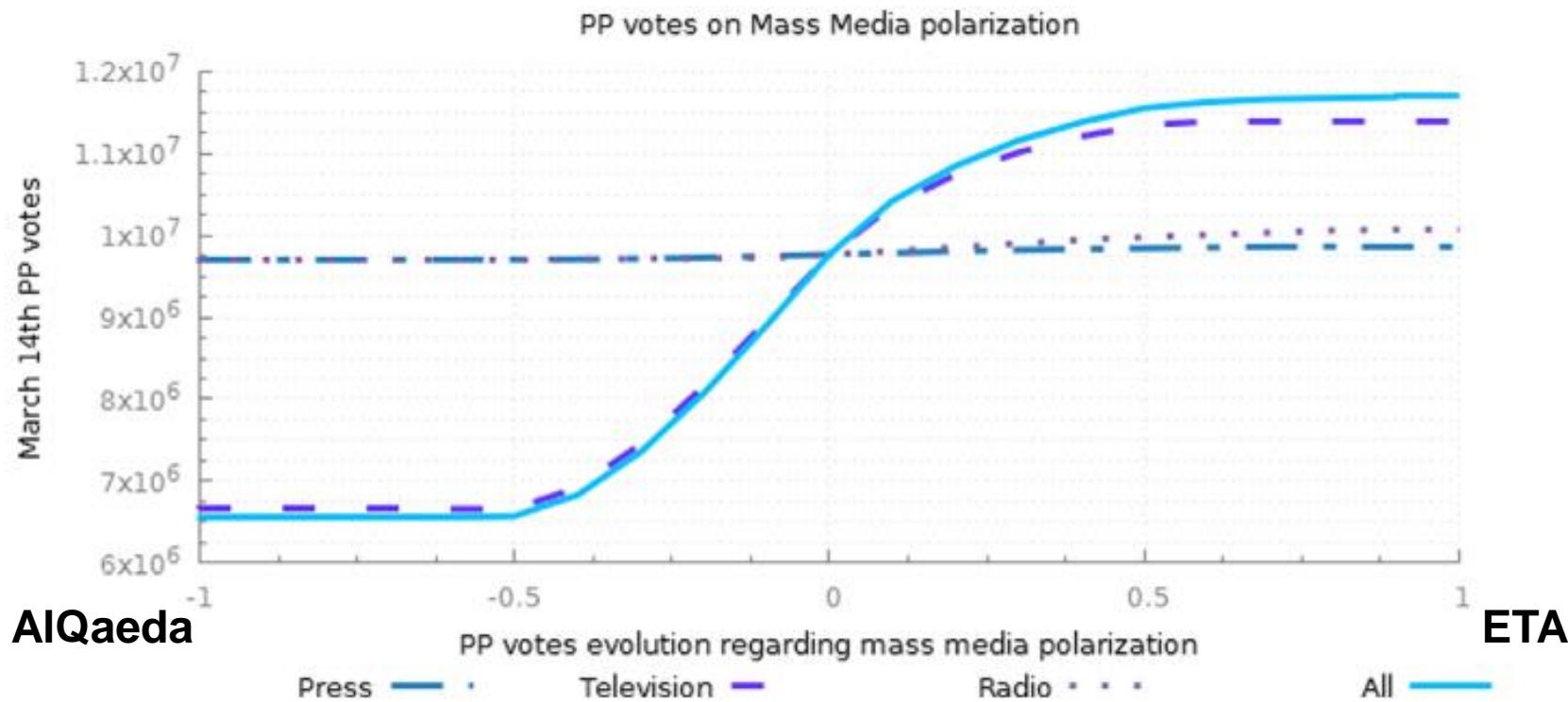


Fig. 6. Percentage of conversations made by agents within their social network.

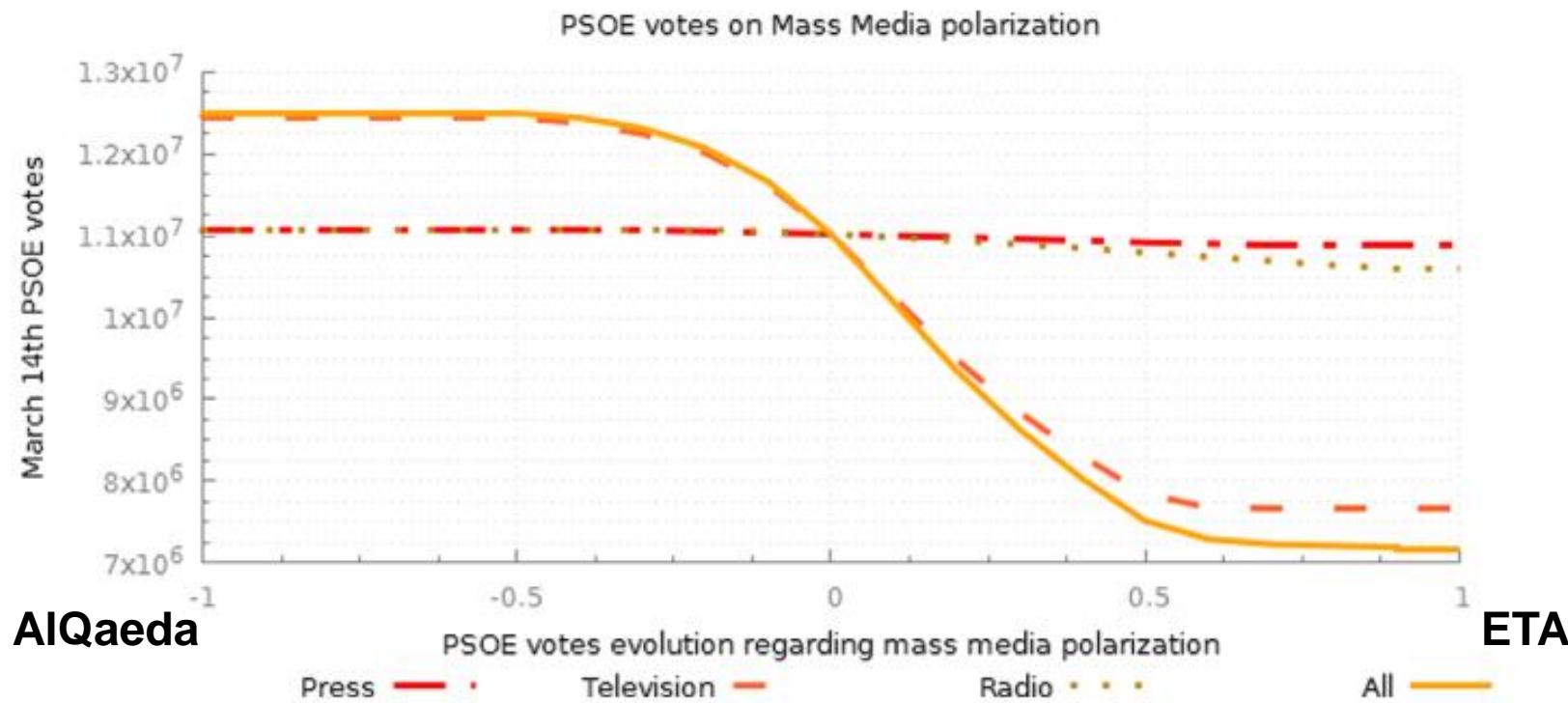
## RESULTADOS (4)

Análisis de sensibilidad del mensaje de los medios de comunicación (polarización AlQaeda – ETA) en el número final de votantes del PP:



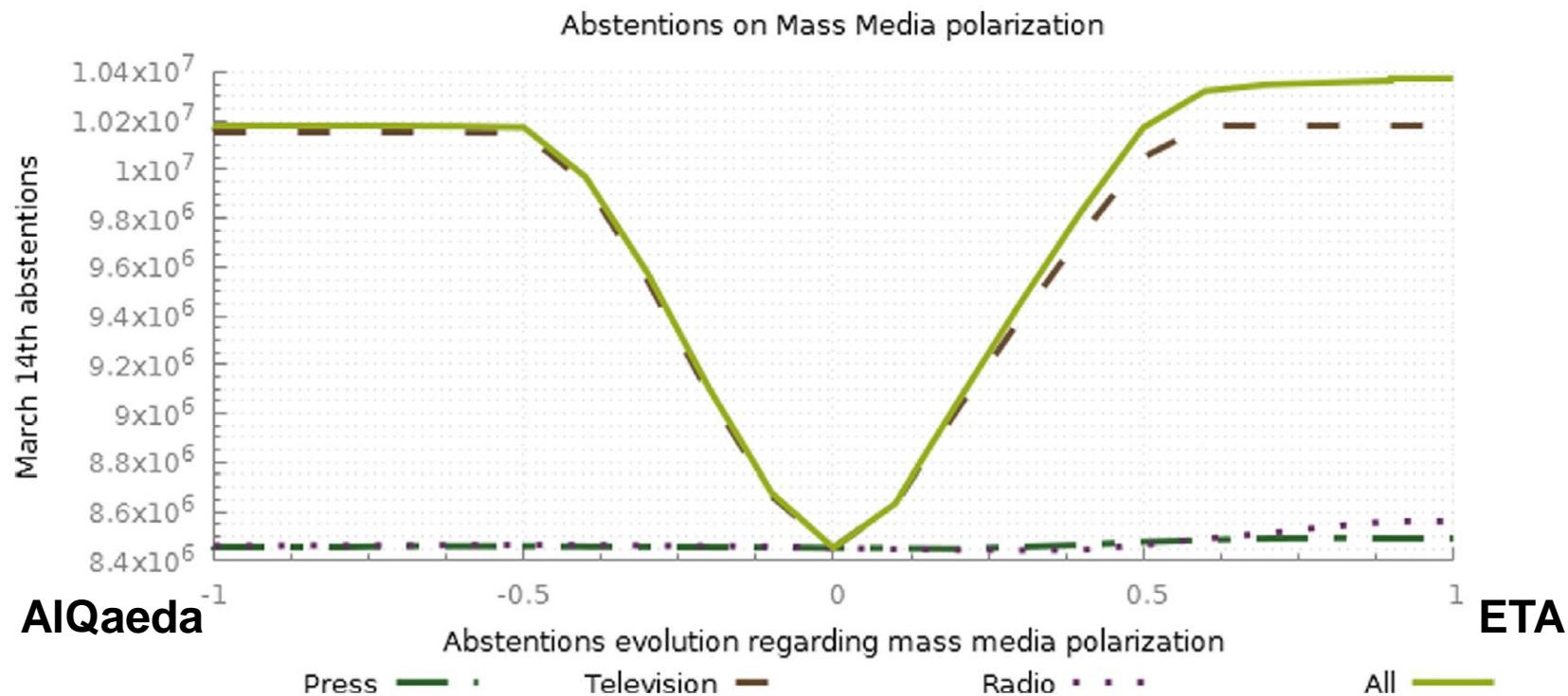
## **RESULTADOS (5)**

## **Análisis de sensibilidad del mensaje de los medios de comunicación (polarización AlQaeda – ETA) en el número final de votantes del PSOE:**



## RESULTADOS (6)

### Análisis de sensibilidad del mensaje de los medios de comunicación (polarización AlQaeda – ETA) en el número final de abstenciones:



# CONCLUSIONES

Primer modelo sobre el impacto del tratamiento informativo del atentado del 11M sobre el resultado electoral del 14M (efecto marco). Se ha usado un ABM con una red social de votantes virtuales, calibrado con datos reales (ajuste del 99%)

Observamos que:

- Hay un cambio significativo en el voto cuando se modifica los parámetros del mensaje y del WoM. La red social de conversaciones entre votantes tiene una gran influencia
- El PSOE es la opción política más influenciada
- El número de votos cambia suavemente cuando los mensajes se polarizan hacia ETA (i.e., hay una mayor sensibilidad del modelo cuando se polarizan hacia AlQaeda)
- La TV fue el medio (touchpoint) más influyente. Aún así, la agregación del efecto de todos los medios tuvo más influencia que la suma de cada uno por separado

# GRAN IMPACTO EN LOS MEDIOS DE COMUNICACIÓN

LAVANGUARDIA | Política

Al Minuto Internacional Política Opinión Vida Deportes Economía Local Gente Cultura Sucesos Temas

Descárgate la  
app Club Vanguardia y  
aprovecha todas sus ventajas

Descárgatela



AND-ANIVERSARIO  
11M

## Una investigación muestra que información sobre 11M influyó en las elecciones



Comparte en Facebook

Comparte en Twitter

Comparte por Whatsapp

Granada, 11 mar (EFE).- Un estudio de la Universidad de Granada ha demostrado científicamente con simulaciones basadas en inteligencia artificial que la información de los medios, los SMS y el boca a boca de los votantes tras el atentado del 11M influyeron en el resultado de las elecciones celebradas tres días después.

EcoDiario.es

Sábado, 11 de Marzo de 2017 Actualizado a las 11:27

ESPAÑA

Portada Global España Deportes Ciencia Cultura Programación TV Viralplus Cartelera El tiempo elEconomista.es

DESTACAMOS

El Barcelona golea a la Roma y deja casi sentenciado su pase a semifinales de Champions

DEPORTES

El Liverpool manda a la lona al Manchester City y se asoma a las semifinales de Champions

## Una investigación muestra que información sobre 11M influyó en las elecciones



11/03/2017 - 11:27

f t g+ in e-mail print

Más noticias sobre: 11-M DEMOCRACIA AUSTRALIA NEWCASTLE PSOE AL QAEDA PP



Granada, 11 mar (EFE).- Un estudio de la Universidad de Granada ha demostrado científicamente con simulaciones basadas en inteligencia artificial que la información de los medios, los SMS y el boca a boca de los votantes tras el atentado del 11M influyeron en el resultado de las elecciones celebradas tres días después.

Fondos de inversión  
**BESTINVER**

INFORMATE

30 AÑOS  
**BESTINVER**  
acciona

30 años creando valor

# ANÁLISIS DE LA TIPOLOGÍA DE ALOJAMIENTOS DE AIRBNB: UN ENFOQUE BASADO EN ANÁLISIS DE REDES SOCIALES

# INTRODUCCIÓN (1)

## Contexto:

La **economía colaborativa (P2P)** es una parte fundamental en el **sector de alojamientos turísticos**

Plataformas digitales como Airbnb o Vrbo ofrecen alternativas a los turistas diferentes a los hoteles

Si quiero invertir en este mercado...

- ¿En qué zona?
- ¿Características del alojamiento?
- ¿Tipo de gestión?
- ¿La diferenciación importa?

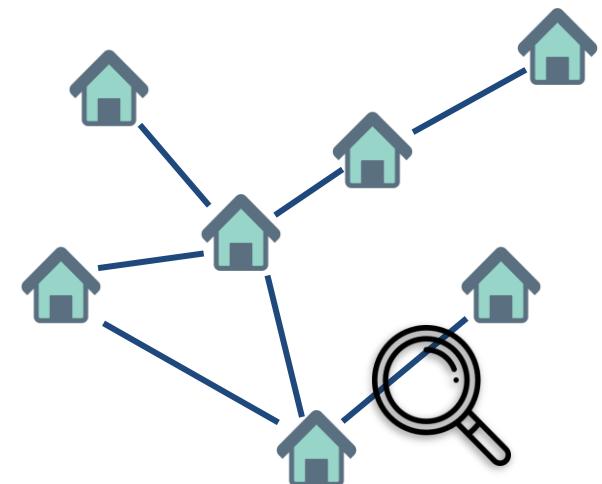
## INTRODUCCIÓN (2)

### Propuesta de metodología:

Construcción de un sistema de apoyo a la toma de decisiones basado en la representación visual de las relaciones de similitud entre los alojamientos

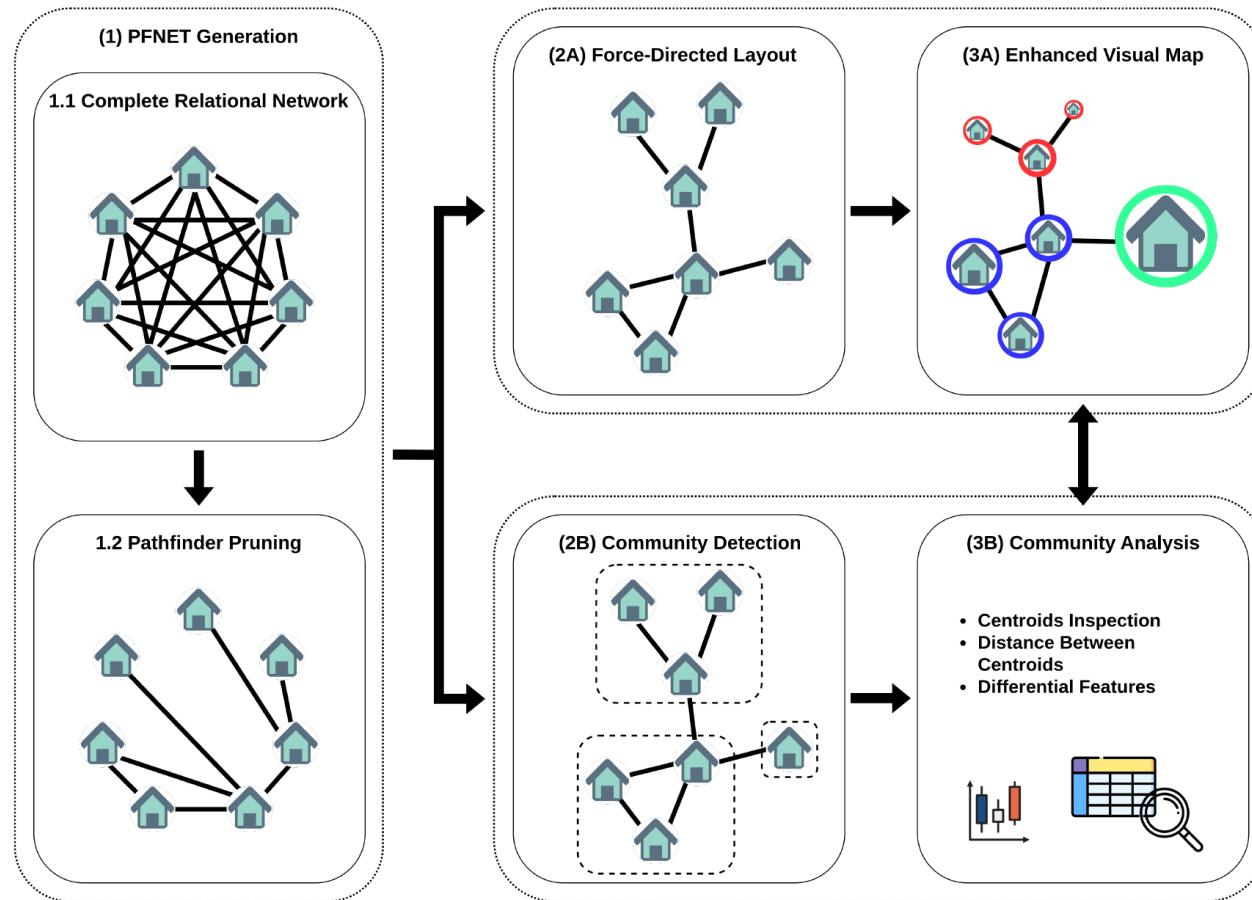
### Claves:

- Enfoque basado en el análisis de redes sociales (SNA)
- La detección automática de comunidades en la red relacional puede ayudar a identificar nichos de mercado
- Permite al decisor ver de un vistazo la situación del mercado y la posición relativa de su alojamiento respecto al resto
- Metodología genérica aplicable a otros dominios



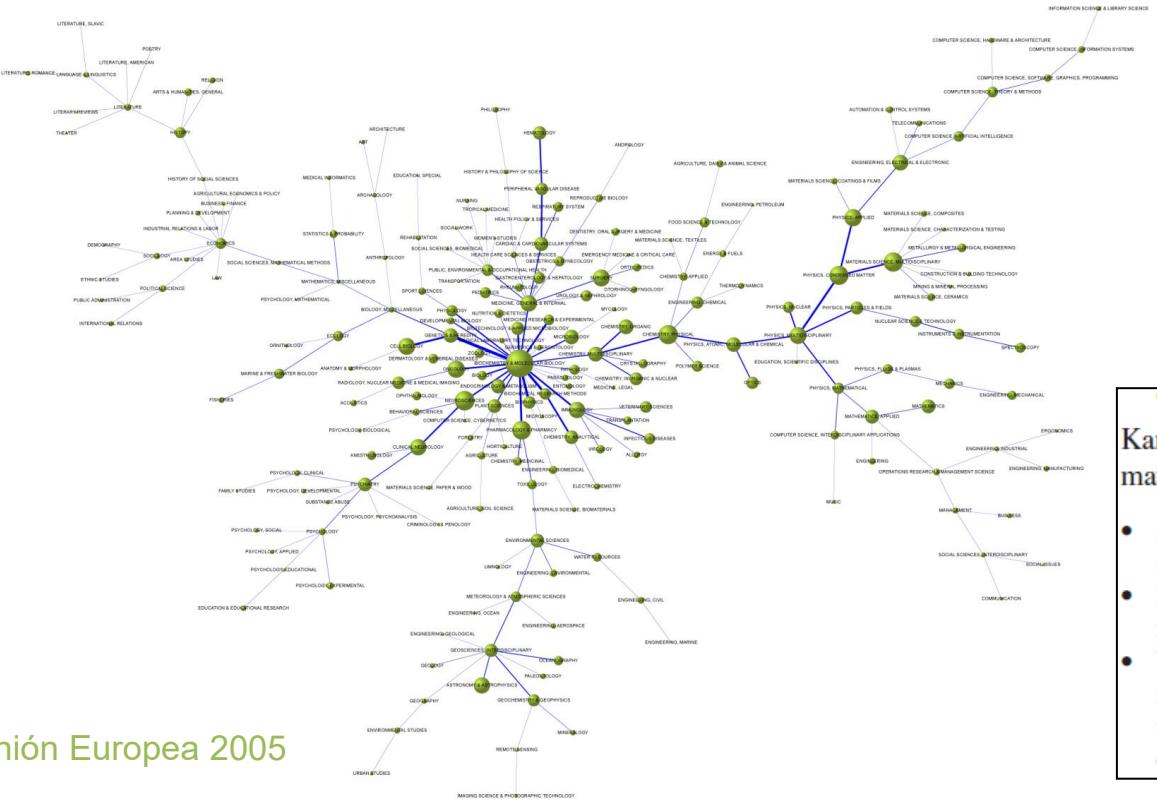
# INTRODUCCIÓN (3)

## Propuesta de metodología:



# INTRODUCCIÓN (4)

## Inspiración de la metodología: Cienciogramas (mapas visuales de la ciencia)



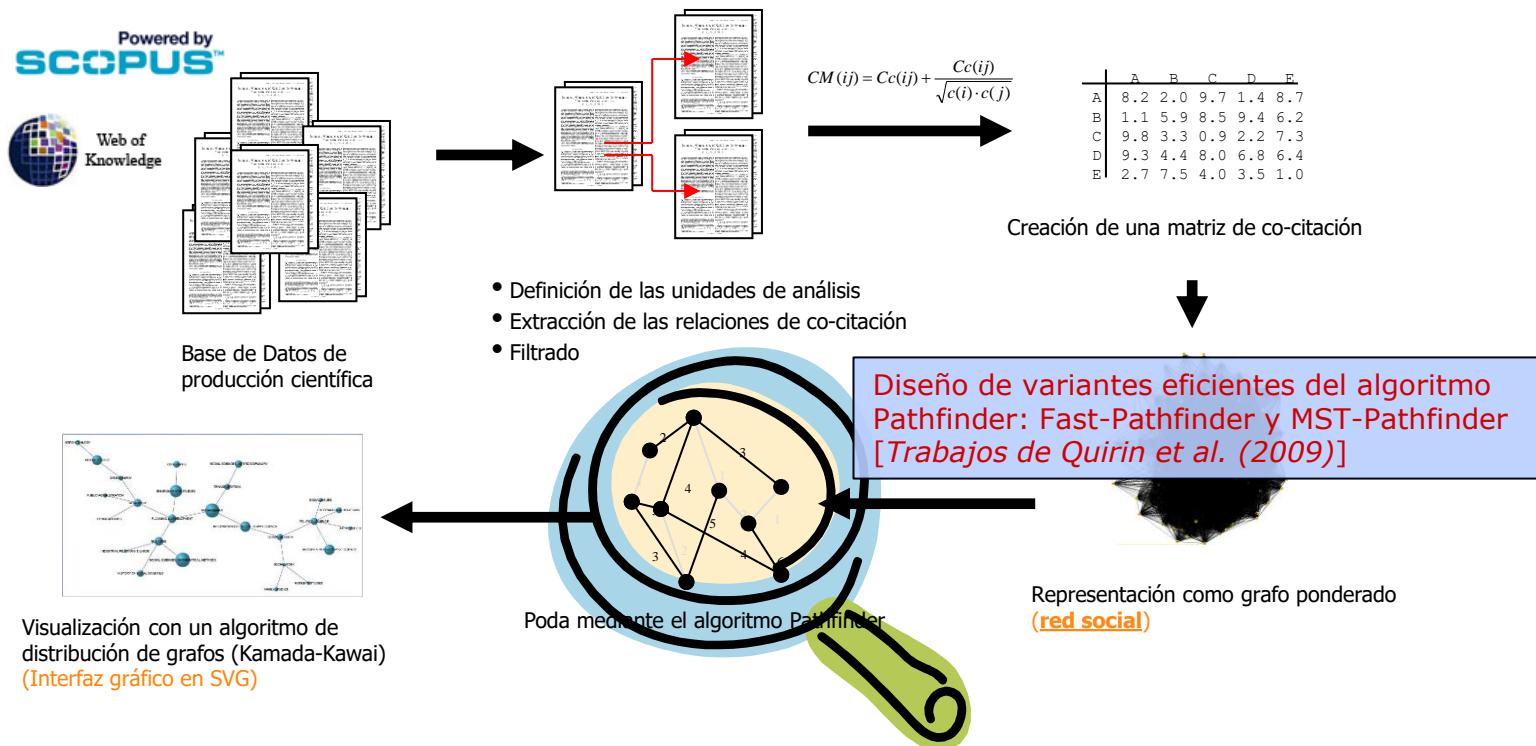
# Representación Gráfica de la Información

The result obtained by combining PFNET with the Kamada–Kawai algorithm is as spectacular and visually informative as the map of an underground metro or railroad system:

- At one glance, the center and the outer limits of the system (i.e., domain) can be seen.
  - It is easy to get from one station (i.e., category) to another, following the trails or links.
  - We can effortlessly see which are the most important nodes in terms of the number of their connections and, in turn, which points act as intermediaries with other lines, as hubs or forking points.

# INTRODUCCIÓN (5)

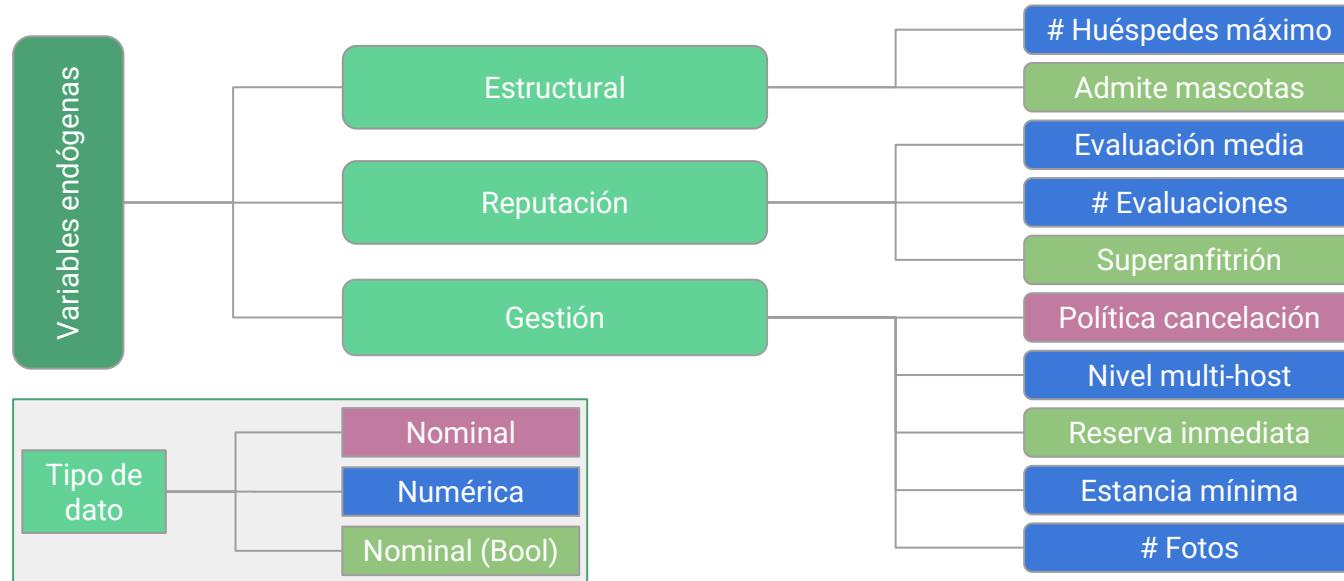
## Inspiración de la metodología: Cienciogramas (mapas visuales de la ciencia)



# CASO DE ESTUDIO

## Industria de alojamientos P2P en las Islas Canarias:

- Dataset con los alojamientos de Airbnb en las Islas Canarias a octubre de 2021
- Alojamientos tras limpieza de los datos: 9412
- Variables económicas / exógenas: beneficios anuales (Revenue) y rendimiento económico diario (ADR)
- 10 variables endógenas



# METODOLOGÍA: Creación red relacional PFNET (1)

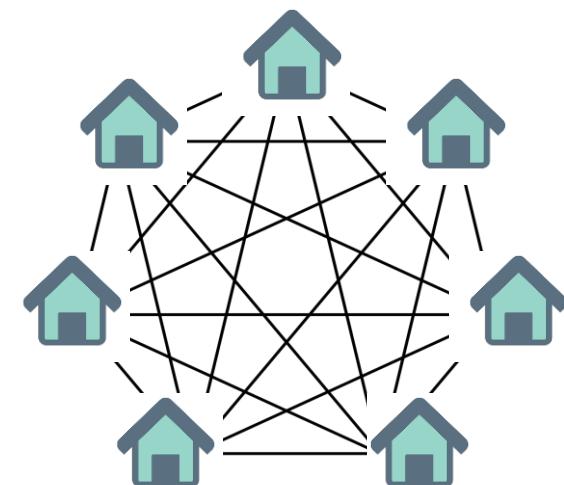
## Red relacional completa:

1. Matriz D: distancias entre todos los pares de alojamientos.
2. Normalización de D en rango [0, 1].
3. Creación matriz de similitudes:  $S = 1 - D$

S es una matriz simétrica que define la red relacional ponderada no dirigida

Distancia para manejar tanto variables nominales como numéricas:

$$HEOM(x_i, x_j) = \sqrt{\sum_{a=1}^m d_a(x_i^a, x_j^a)^2},$$
$$d_a(x_i, x_j) = \begin{cases} 0 & \text{si } a \text{ es nominal y } x_i^a = x_j^a, \\ 1 & \text{si } a \text{ es nominal y } x_i^a \neq x_j^a, \\ \frac{|x_i^a - x_j^a|}{\max_a - \min_a} & \text{if } a \text{ es numérica} \end{cases}$$



## METODOLOGÍA: Creación red relacional PFNET (2)

### Poda Pathfinder:

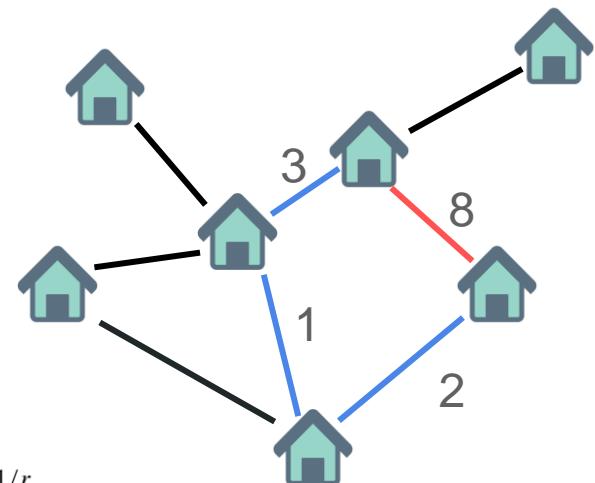
El uso de la red relacional completa sobrecargaría al decisor de información, dificultando la interpretación del mapa visual



Es necesario eliminar enlaces poco relevantes

### Método Pathfinder:

- Poda que preserva las distancias geodésicas de la red
- Elimina de forma iterativa aquellos enlaces que no satisfacen la desigualdad triangular
- Mantiene las relaciones más significativas

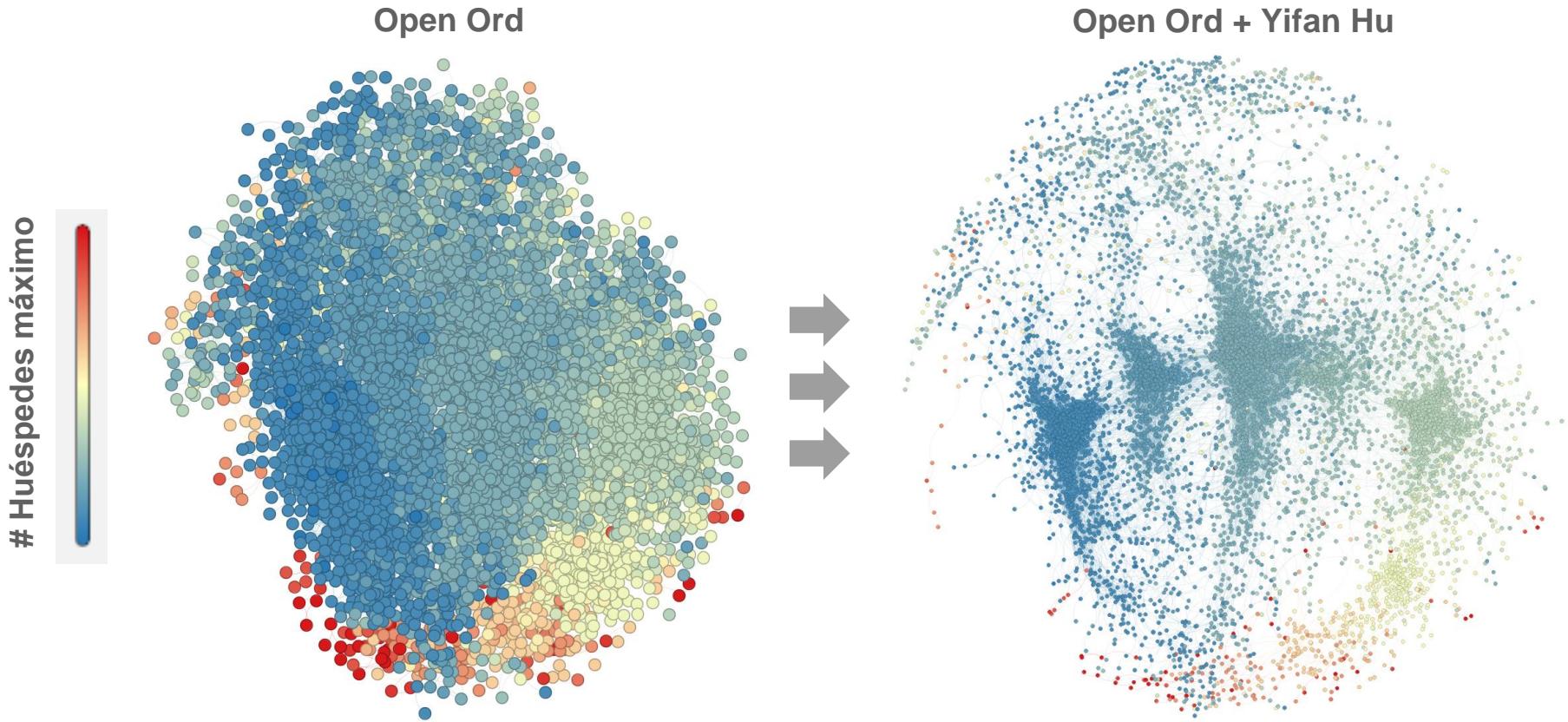


### Parámetros:

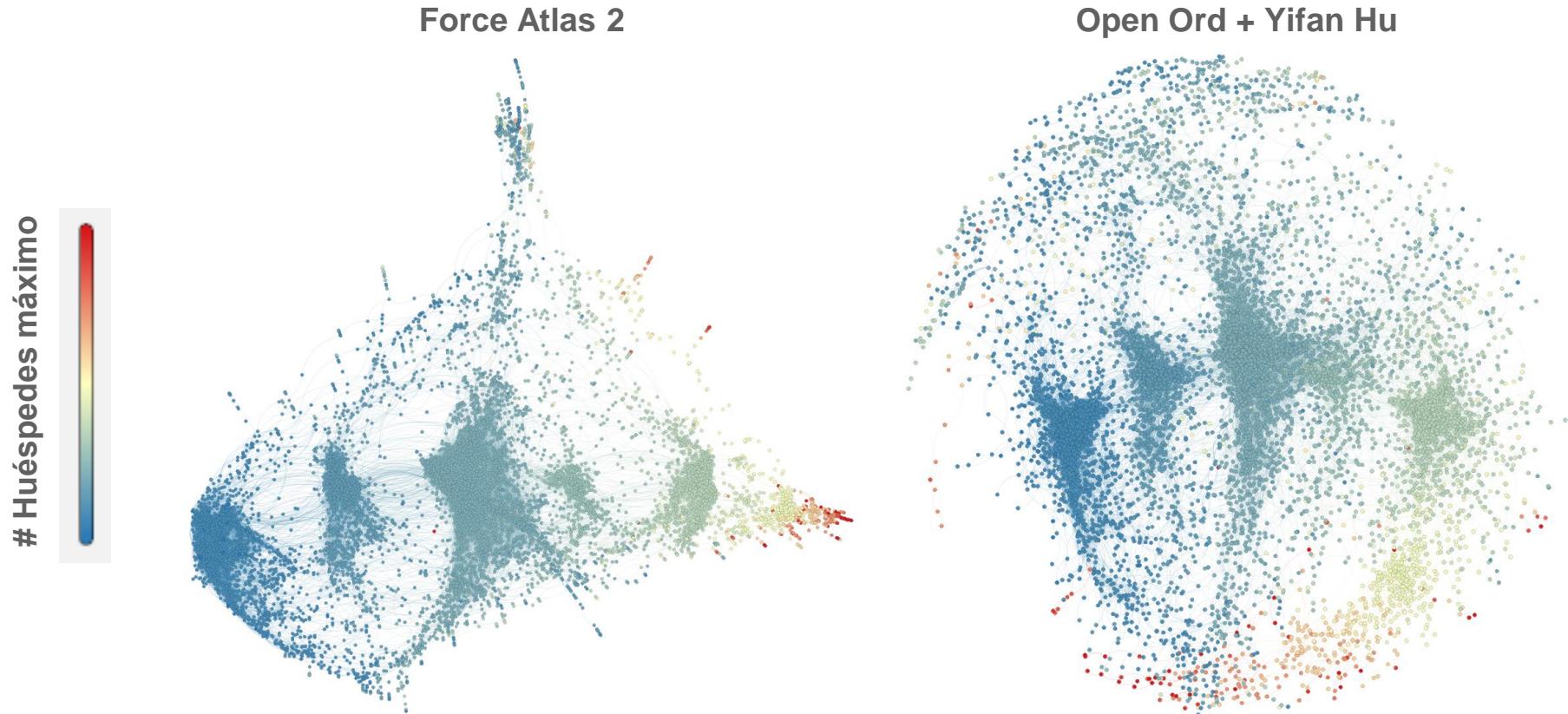
- $q$ : longitud máxima de los caminos considerados
- $r$ : configuración distancia Minkowski
- Configuración PFNET ( $q=n-1$ ,  $r=\infty$ ) → Poda máxima
  - Equivalente a la unión de todos los MSTs del grafo

$$D = \left( \sum_i d_i^r \right)^{1/r}$$

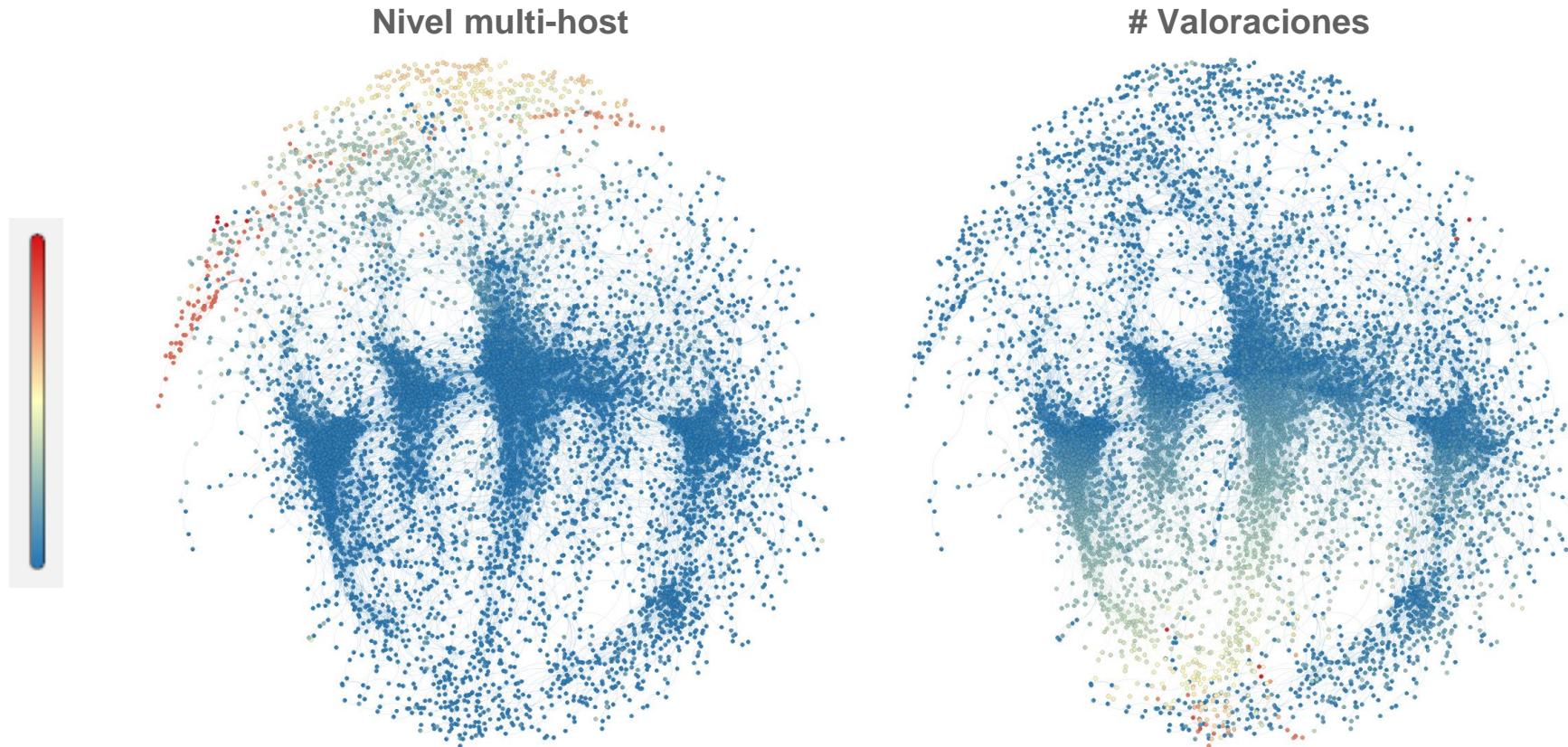
# METODOLOGÍA: Visualización red relacional (1)



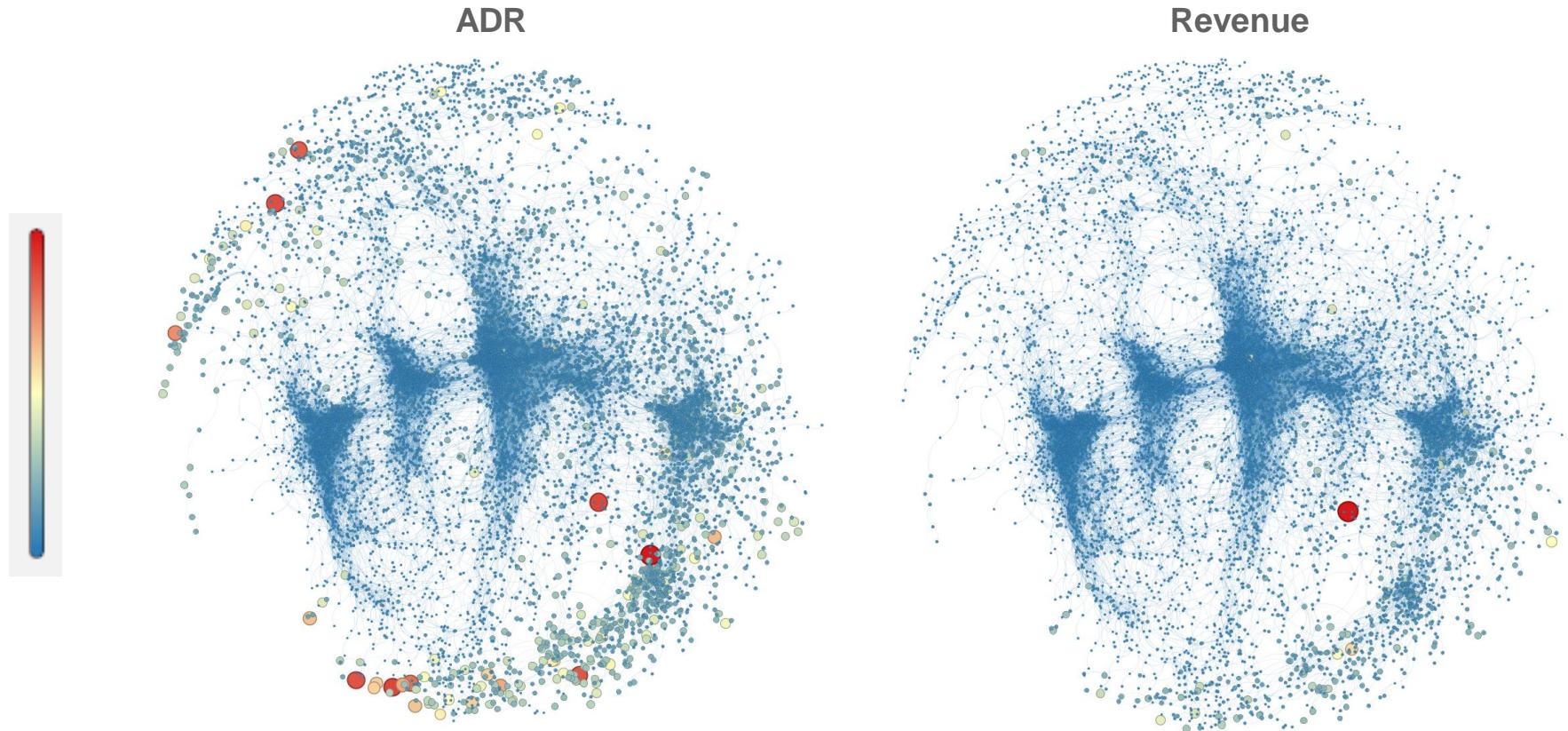
## METODOLOGÍA: Visualización red relacional (2)



## METODOLOGÍA: Visualización red relacional (3)



## METODOLOGÍA: Visualización red relacional (4)



# METODOLOGÍA: Detección de comunidades

**Comunidad:** grupo de nodos densamente conectados internamente y con pocas conexiones externas

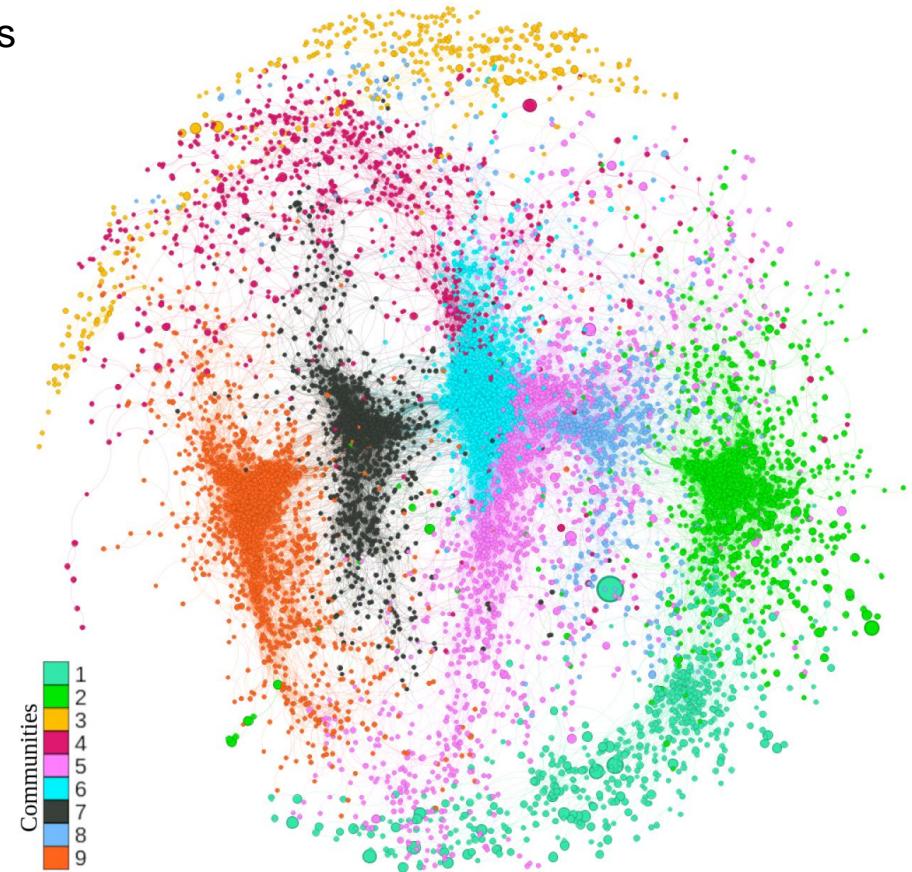
## Método de Lovaina:

Método preciso y muy eficiente. Iteración de dos procesos:

1. Optimización greedy de la modularidad ( $Q$ )
2. Agregación de comunidades en supernodos

## Resultados:

- 9 comunidades, todas con un tamaño significativo
  - La más pequeña tiene 383 alojamientos (4%)
- Partición de muy buena calidad:  $Q = 0.715$ 
  - Umbral empírico de calidad mínima:  $Q \geq 0.3$

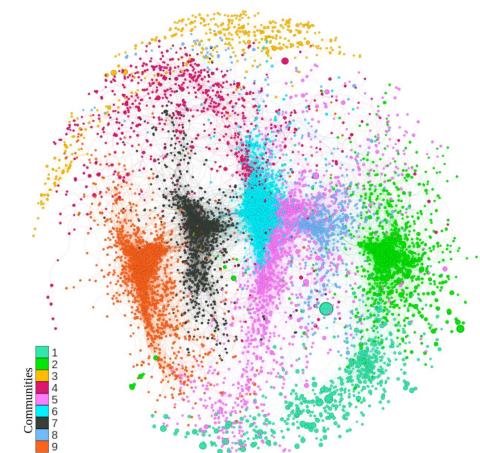


# ANÁLISIS DE RESULTADOS (1)

	Size	Size (%)	Revenue (\$)	ADR (\$)	Superanfitrión	Política Cancelación	Reserva Inmediata	Mascotas	# Huéspedes	Estancia Mínima	Nivel Multihost	# Fotos	# Valoraciones	Valoración Media
1	618	7%	30.311,95	264,11	False (80.4%)	Strict (48.2%)	True (57.3%)	False (72.2%)	9,34	4,55	9,01	43,59	24,99	95,52
2	1.241	13%	17.650,52	151,81	False (78.8%)	Strict (40.1%)	True (64.1%)	False (77.0%)	6,23	4,11	10,39	35,95	26,33	94,66
3	383	4%	12.729,90	157,68	False (100.0%)	Strict (62.9%)	True (99.5%)	False (83.3%)	4,53	1,91	231,61	31,13	14,97	88,43
4	764	8%	13.737,92	126,71	False (96.7%)	Strict (37.8%)	True (81.8%)	False (88.9%)	4,46	3,48	84,71	27,34	16,83	90,73
5	1.432	15%	13.794,16	90,94	False (69.5%)	Moderate (33.5%)	True (64.0%)	False (78.8%)	3,98	3,68	8,07	46,19	73,60	95,66
6	1.638	17%	10.009,89	91,14	False (79.0%)	Strict (37.5%)	True (64.3%)	False (81.1%)	4,02	3,78	6,22	23,29	20,53	94,09
7	1.115	12%	9.386,20	73,81	False (76.3%)	Moderate (34.7%)	True (68.8%)	False (85.0%)	2,99	3,54	8,25	26,35	30,78	94,10
8	564	6%	12.427,48	105,66	False (83.3%)	Strict (34.8%)	True (70.4%)	False (83.2%)	4,97	4,08	7,74	30,00	28,58	91,23
9	1.657	18%	9.468,61	70,22	False (73.0%)	Flexible (34.1%)	True (66.2%)	False (85.3%)	2,05	3,54	7,24	27,03	41,75	94,86
NO CONSIDERADAS PARA LAS DISTANCIAS				CONSIDERADAS PARA LAS DISTANCIAS										
INFLUENCIA MEDIA DISTANCIAS				0,11	0,18	0,10	0,04	0,30	0,00	0,17	0,04	0,06	0,01	

## Análisis de los centroides:

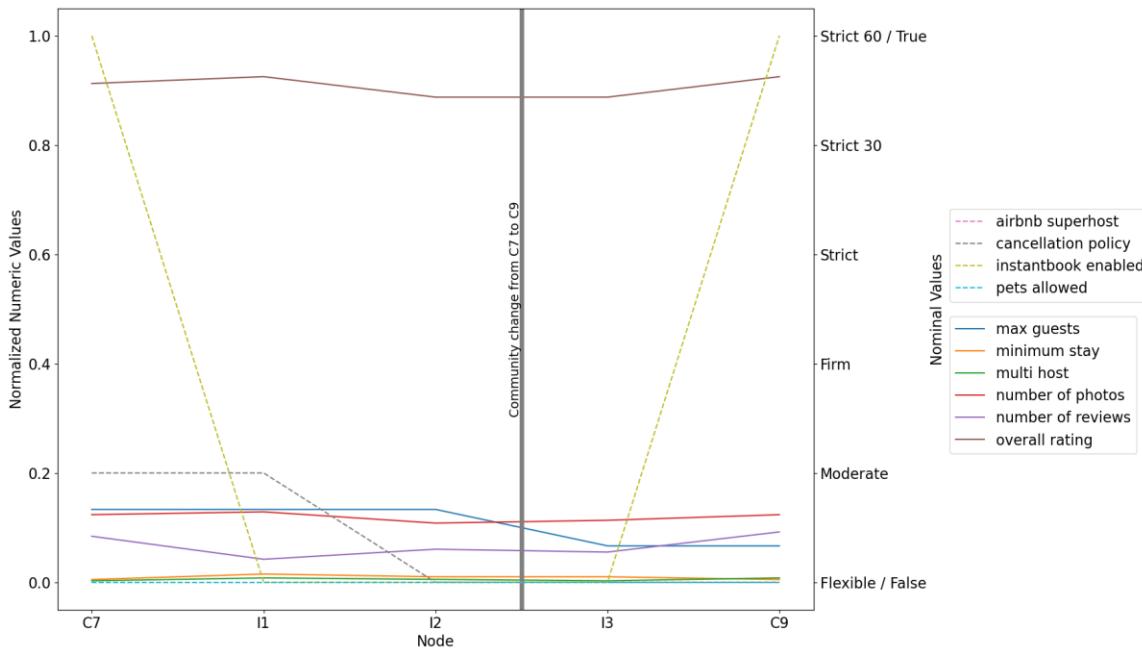
- C1 tiene el mayor Revenue y ADR. Alojamientos amplios con buena valoración
- C2 es una variante de C1 con valores más intermedios
- C3 se corresponde con grandes propietarios / empresas
  - Alto multihost, reserva automática cercana al 100% (poco trato)
  - Tiene precios altos, pero su Revenue es bajo: no se reserva demasiado
- C4 es una variante suavizada de C3, es el puente con el resto
- C5-9 tienen los beneficios más bajos, a la vez que guardan una posición más central
- Factor más discriminativo a nivel global: # Huéspedes**



## ANÁLISIS DE RESULTADOS (2)

### Análisis de la variabilidad de las variables entre comunidades: caminos geodésicos:

Variación de los valores de las variables en las distintas propiedades situadas en el camino geodésico que conecta las dos propiedades más cercanas al centroide de dos comunidades:



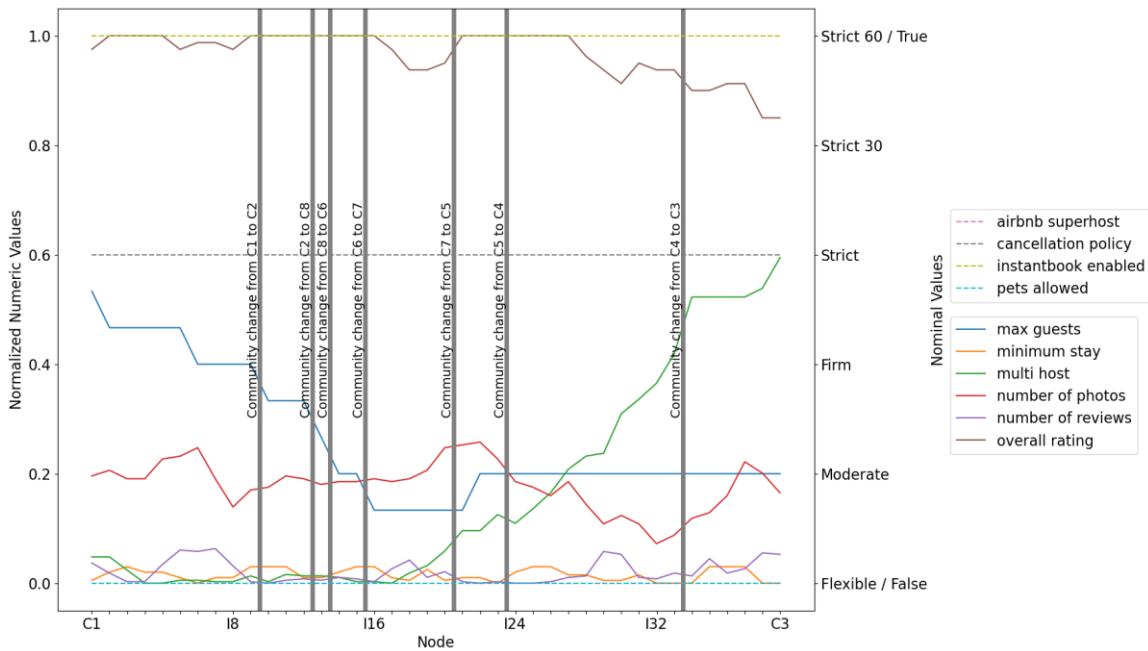
### 2 comunidades “cercanas”

- Líneas sólidas: variables numéricas
- Líneas punteadas: variables nominales
- Línea vertical: cambio de comunidad

# ANÁLISIS DE RESULTADOS (3)

## Análisis de la variabilidad de las variables entre comunidades: caminos geodésicos:

Variación de los valores de las variables en las distintas propiedades situadas en el camino geodésico que conecta las dos propiedades más cercanas al centroide de dos comunidades:



### 2 comunidades “lejanas”

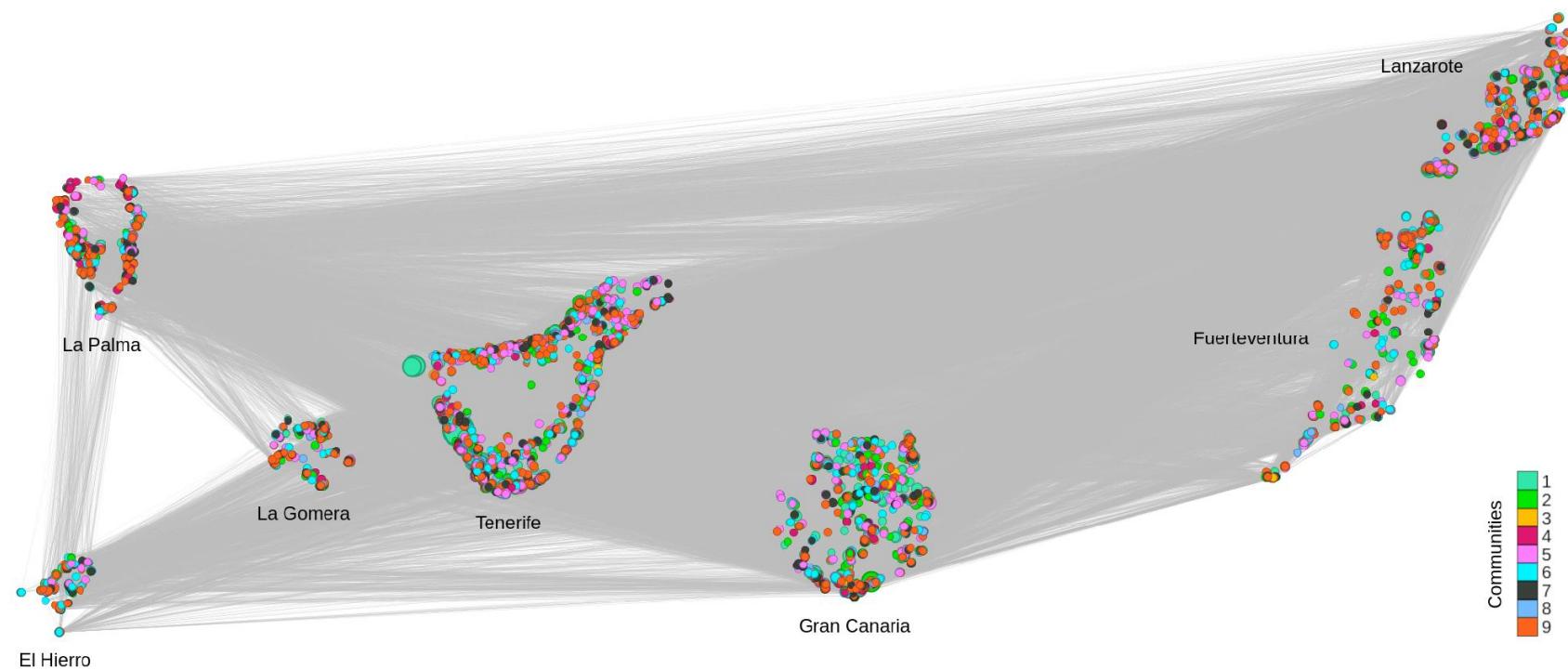
- Líneas sólidas: variables numéricas
- Líneas punteadas: variables nominales
- Línea vertical: cambio de comunidad

# ANÁLISIS DE RESULTADOS (4)

## Geovisualización:

Las comunidades están distribuidas por todas las islas:

- Configuraciones concretas de las variables endógenas usadas no son intrínsecas a ninguna isla



# CONCLUSIONES

## Metodología:

- La visualización de la red relacional PFNET posiciona adecuadamente los alojamientos: las distancias entre ellos representan claramente sus similitudes y diferencias
- Esto permite realizar un análisis centro-periferia de los alojamientos para determinar cuáles son los más estándar (centro) y cuáles más peculiares (periferia)
- El mapa visual permite ver de un vistazo la situación relativa de cada alojamiento respecto al mercado completo
- Los caminos geodésicos muestran la variación de las variables entre los alojamientos de los distintos grupos
- Metodología genérica aplicable a cualquier dominio en el que pueda definirse una medida de distancia entre los datos

## Caso de estudio:

- ¿La diferenciación importa? Claramente sí
  - Los alojamientos con mayor Revenue se concentran en una parte de la periferia
  - La diferenciación no es siempre positiva: la comunidad más peculiar no tiene un Revenue destacable
- Los alojamientos con mayores beneficios se caracterizan por ser amplios
- No hay islas con un “alojamiento tipo”: las configuraciones/comunidades encontradas se reparten entre todas las islas

# RETOS Y LÍNEAS DE INVESTIGACIÓN ABIERTAS

## RETOS Y LÍNEAS DE INVESTIGACIÓN ABIERTAS

- Influencia de la **estructura de la red** en procesos de difusión y cascada
- Influencia de la **estructura de comunidades** sociales
- Estrategias de **marketing viral**. Determinación de los actores influyentes (**influentials**)
- Definición de **heurísticas** de toma de decisión realistas: compra, posteo, etc.
- **Calibración** de los modelos ABM
- **Big Data:**
  - **análisis de sentimientos** para incorporar información realista al modelo ABM,
  - **escalado de las redes** al tamaño de las redes sociales reales, etc.
- ...

# MARKETING VIRAL: SELECCIÓN DE LAS SEMILLAS

El **marketing viral** (o marketing WoM) se basa en que las conversaciones de consumidores sobre un producto son una herramienta más poderosa que la publicidad tradicional

Una estrategia habitual es “recompensar” a unos pocos individuos para promocionar el producto entre sus amigos buscando una adopción grande y rápida (cascada)

El problema consiste en decidir qué individuos escoger para maximizar cantidad y ratio de adopción del producto. Cuantos más se recompensen en la campaña promocional, mayor será la velocidad de adopción pero también mayor el coste

Dada una red social (on-line), es un problema NP-duro. Hay que decidir bien los individuos y el presupuesto de la campaña



## APRENDIZAJE EVOLUTIVO DE ESTRATEGIAS DE MARKETING VIRAL (1)

Esta área de investigación está directamente relacionada con la de la **dinámica de los individuos influyentes y la estructura de las redes en los procesos de adopción**

Otra dificultad es que la topología de la red social no se conoce al completo por las restricciones de privacidad

Existen varias versiones del problema dependiendo de la información disponible (red social, presupuesto máximo, etc.)

Las versiones clásicas (globales: GVMP) son **poco realistas** al asumir un conocimiento completo de la red

Stonedahl et al. proponen una versión local más realista (LVMP) en la que la información disponible son **medidas locales de los individuos** como grado o coeficiente de clustering (CC)

## APRENDIZAJE EVOLUTIVO DE ESTRATEGIAS DE MARKETING VIRAL (2)

Se dispone de una red social de consumidores (desconocida a nivel global) y de una función de adopción que indica la probabilidad de que un individuo adopte el producto en tiempo  $t$  ([procesos de contagio simple/complejo o de umbral](#))

Usan [algoritmos genéticos para obtener los individuos \(semillas\) que maximizan tanto la cantidad como el ratio de adopción del producto](#)

No sólo interesa que lo adopten muchos consumidores sino que lo hagan pronto para lograr más beneficios. Se usa un índice llamado [Net Present Value](#) (NPV) que mide ambos conceptos

Las estrategias, simples y combinadas, ordenan los individuos según su preferencia para ser escogidos. Así, [se puede optimizar también el presupuesto de la campaña](#)

Se consideran varios tipos de redes (regular, ER, mundos pequeños, libres de escala y Twitter) para estudiar su influencia ([distinta dinámica del proceso de contagio](#))

## EXPERIMENTOS (1)

Estudian escenarios con **dos niveles de viralidad en la función de adopción**, media y alta (en la realidad, distintos mercados de productos), para analizar la robustez de las estrategias

Implementación en Netlogo disponible:  
<http://behaviorsearch.org/>. Resultados medios 30 ejecuciones del AG con 1000 agentes

Se obtienen estrategias que mejoran el NPV para todas las redes

**Las redes libres de escala facilitan la cascada. Sus estrategias requieren menos inversión**

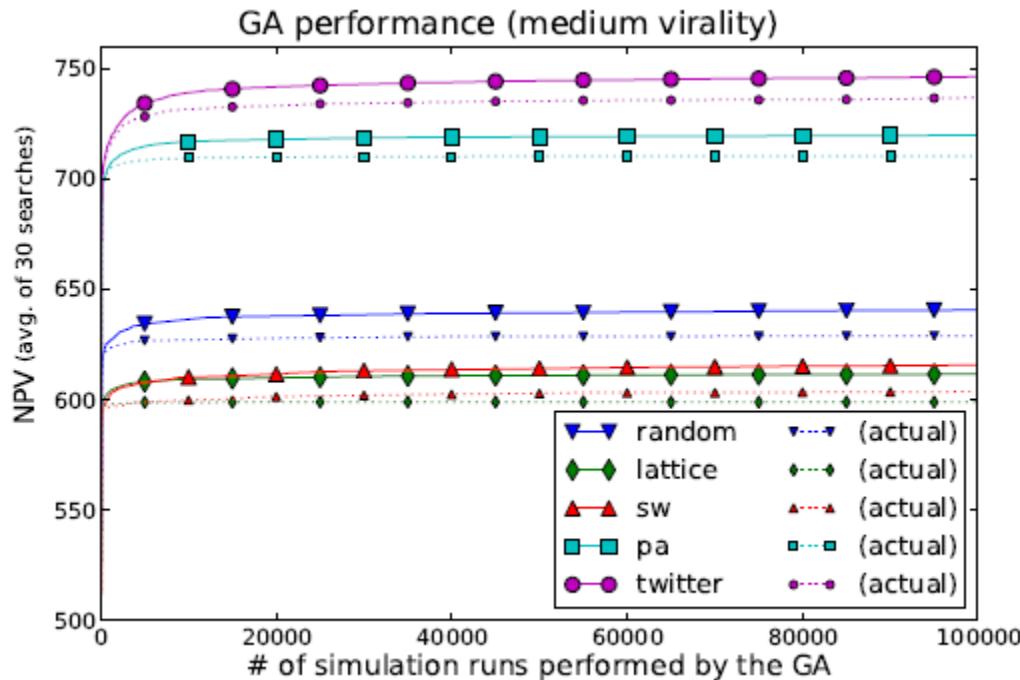
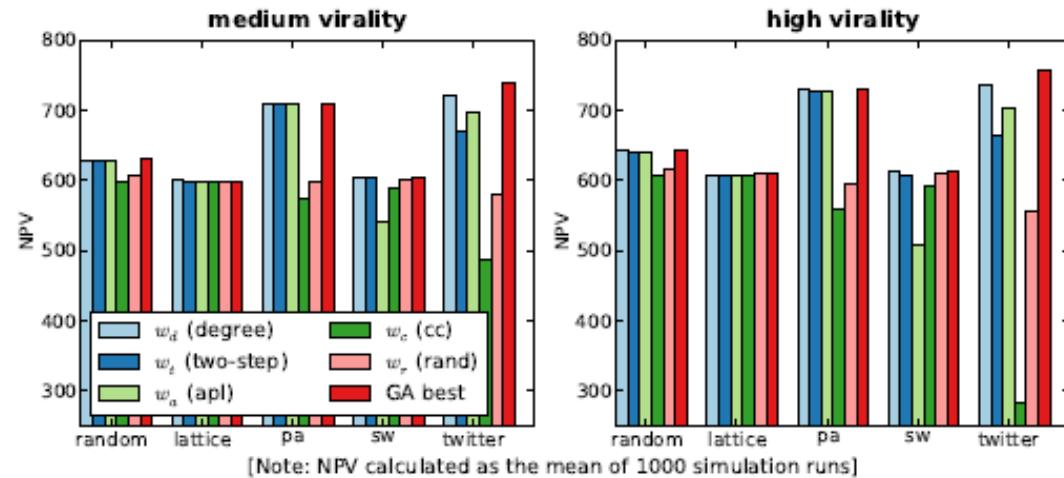


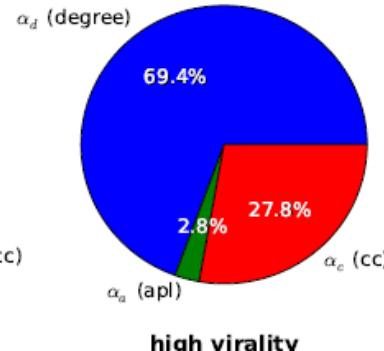
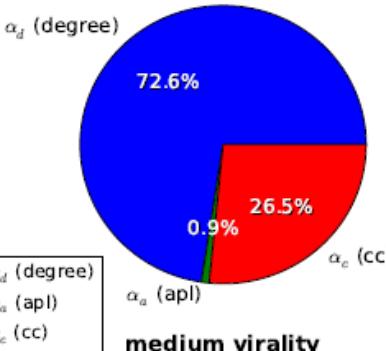
Figure 3: GA progress (averaged across 30 searches) by network topology, for the ‘medium virality’ scenario. GA’s reported best-of-run fitness (solid lines) are compared with the actual NPV values (dotted lines), estimated by 1000 simulation runs, showing the effect of noise. (Error bars too small to show.)

## EXPERIMENTOS (2)

La composición de las estrategias combinadas es diferente según el tipo de red, aunque sólo se obtienen mejoras significativas sobre las simples en la red de Twitter



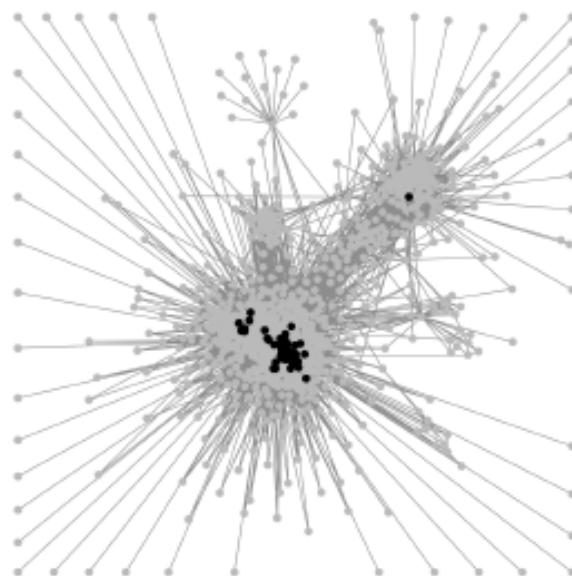
Best primary strategies for twitter network



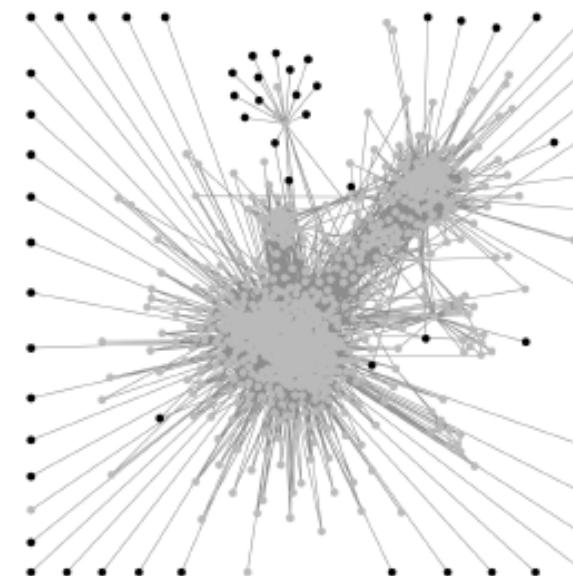
En la red de Twitter, el CC sólo no funciona bien pero la estrategia combinada de grado y CC mejora el NPV en un 2.5% en ambos escenarios de viralidad

## EXPERIMENTOS (3)

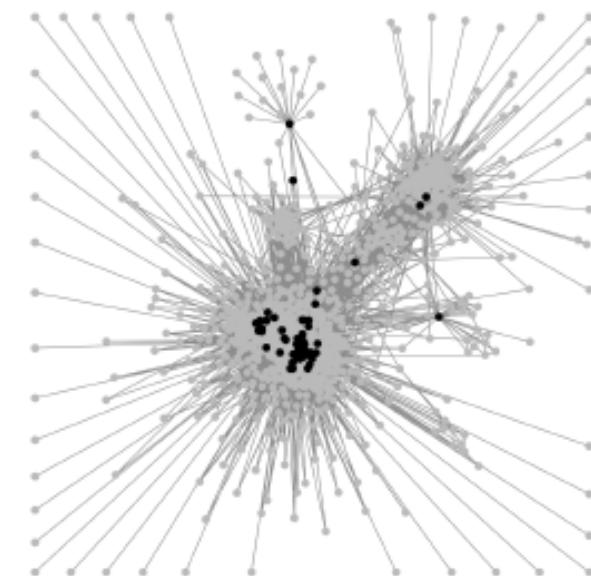
La explicación parece ser que **hay algunos individuos importantes alejados de los hubs centrales que hacen de “*brokers*” a individuos/grupos no conectados directamente a esos hubs**



a) pure degree



b) pure cc



c) GA's best strategy

Figure 8: Visualization of three seeding strategies on the *twitter* network.