



# **Minería de Medios Sociales**

## **Máster en Ciencia de Datos e Ingeniería de Computadores**

### **Bloque I: Redes Sociales y Minería de Datos en Redes**

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

**Oscar Cordon García**

*Dpto. Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada*  
*ocordon@decsai.ugr.es*

# MODELOS DE COMPORTAMIENTO DE CONSUMIDOR CON ABMs Y REDES SOCIALES

# MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR

En los modelos de cascada, los agentes sólo adoptan un producto porque sus amigos lo hacen. En la realidad, las personas compran productos en función de sus necesidades y de los atributos de éstos

Los modelos más realistas integran reglas de decisión basadas en estas **necesidades individuales** con procesos de difusión (**Word of Mouth, WoM**) (**necesidades sociales**)

**Modelo básico de Janssen:** Los agentes consumidores deciden qué producto comprar de entre  $N$  posibles. Está conectados en una red social. Cada producto tiene una serie de atributos que definen sus características y su **utilidad** para el agente



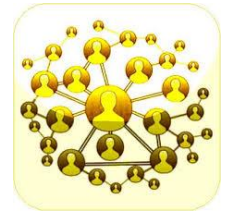
# UTILIDAD DE PRODUCTO (1)

Está compuesta por dos partes, **individual** (grado de ajuste entre sus características y las preferencias del consumidor) y **de efecto social** (influencia del consumo del producto por parte de los amigos del agente)

La **utilidad individual** se modela como la diferencia entre la preferencia del consumidor  $i$ , expresada como  $p_i \in [0,1]$ , y el valor de las características del producto, agrupadas en una única variable  $\delta_j \in [0,1]$

La parte individual de la utilidad de un producto  $j$  para el agente  $i$  es:

$$1 - |p_i - \delta_j|$$



## UTILIDAD DE PRODUCTO (2)

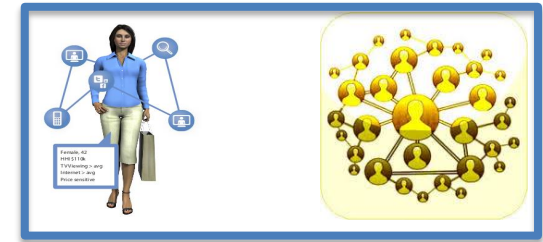
La parte de **efecto social** de la utilidad del producto aumenta con la cantidad de amigos que consumen el producto (**variante de contagio complejo**). Está relacionado con la necesidad del agente de parecerse al resto de agentes de su red social

Se modela con una variable  $x_j \in [0,1]$  que mide la distancia promedio entre la elección del agente y las de sus vecinos (**incertidumbre**). Para cada agente se calcula el ratio de cada producto consumido por los vecinos

La utilidad total esperada de consumo del producto  $j$  para el agente  $i$  es:

$$E[U_{ij}] = \beta_i \cdot (1 - |\delta_j - p_i|) + (1 - \beta_i) \cdot x_j$$

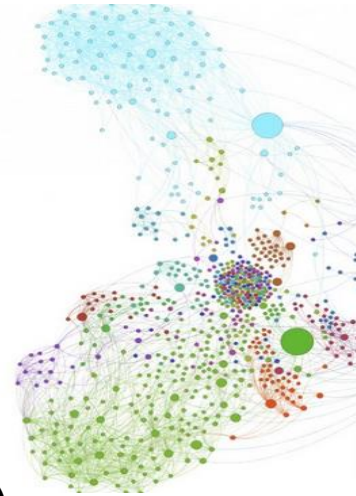
$\beta_i \in [0,1]$  es un peso que determina el **comportamiento social** de cada agente. A mayor valor de  $\beta_i$ , el agente es más individualista (innovador)



# RED SOCIAL (1)

## Representa las relaciones entre los consumidores del modelo:

- Comunicación y difusión de conocimientos y opiniones (WoM)
- Entorno de los agentes (“amigos” y conocidos)



La estructura de la red influye en los procesos de difusión y, por tanto, en el comportamiento de los consumidores (**micro-efectos**) y en los resultados finales del modelo de un mercado concreto (**macro-efectos**)

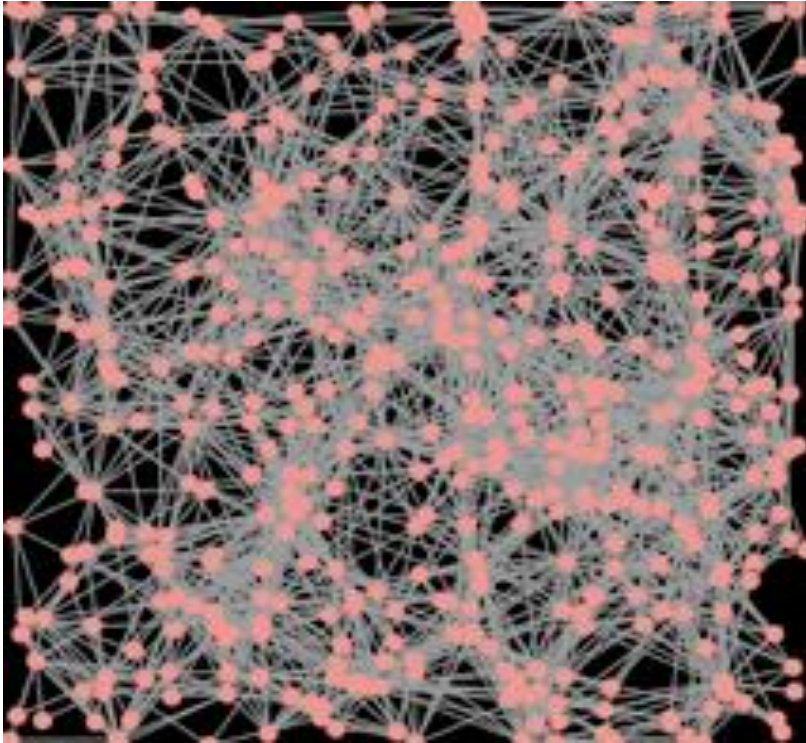
Se pueden considerar distintas topologías (redes aleatorias, de mundo pequeño, libres de escala, ...) y con distintos parámetros (densidad, grados, etc.):

- **La densidad de la red define el volumen de conversaciones sobre los productos** (características del mercado). Ej: papel higiénico vs. Móviles
- **La presencia de hubs determina la existencia de “influencers”**
- Las redes con mayor grado medio y la presencia de hubs aceleran la difusión

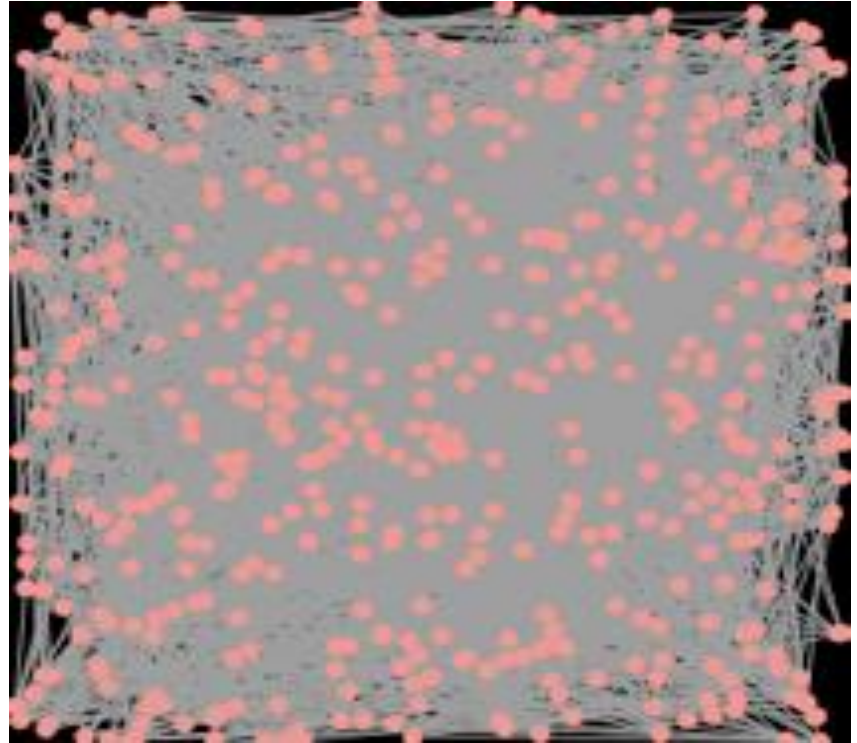


## RED SOCIAL (2): VOLUMEN DE CONVERSACIONES

(a) Red social con  $D=2$



(b) Red social con  $D=10$



# PROCESO DE COMPRA (1)

**Características del mercado:** Se asume un **conocimiento completo** de todos los productos del mercado y de las compras realizadas por los contactos. Se definen unos **umbrales** de utilidad mínima exigida  $U_{\min}$  e incertidumbre tolerada  $Unct$

Cada unidad de tiempo (p.ej. día), el consumidor decide si quiere comprar **probabilísticamente**. La probabilidad de comprar es un parámetro  $p_b$

Cada agente que compra calcula su utilidad e incertidumbre globales y escoge una heurística de decisión entre cuatro posibles. **Las heurísticas determinan que producto comprar de forma determinística (el que maximiza la utilidad) o probabilística**

**Los umbrales y el parámetro de comportamiento social  $\beta_i$  la toma de decisiones de cada agente consumidor (conservadora, arriesgada, imitadora, etc.)**



# PROCESO DE COMPRA (2)

|                             |  |
|-----------------------------|--|
| Cálculo de la utilidad      | $E [U_{ij}] = \beta_i \cdot (1 -  d_j - p_i ) + (1 - \beta_i) \cdot x_j$ |
| Cálculo de la incertidumbre | $E[Unc_{ij}] = (1 - \beta_i) \cdot (1 - x_j)$                            |

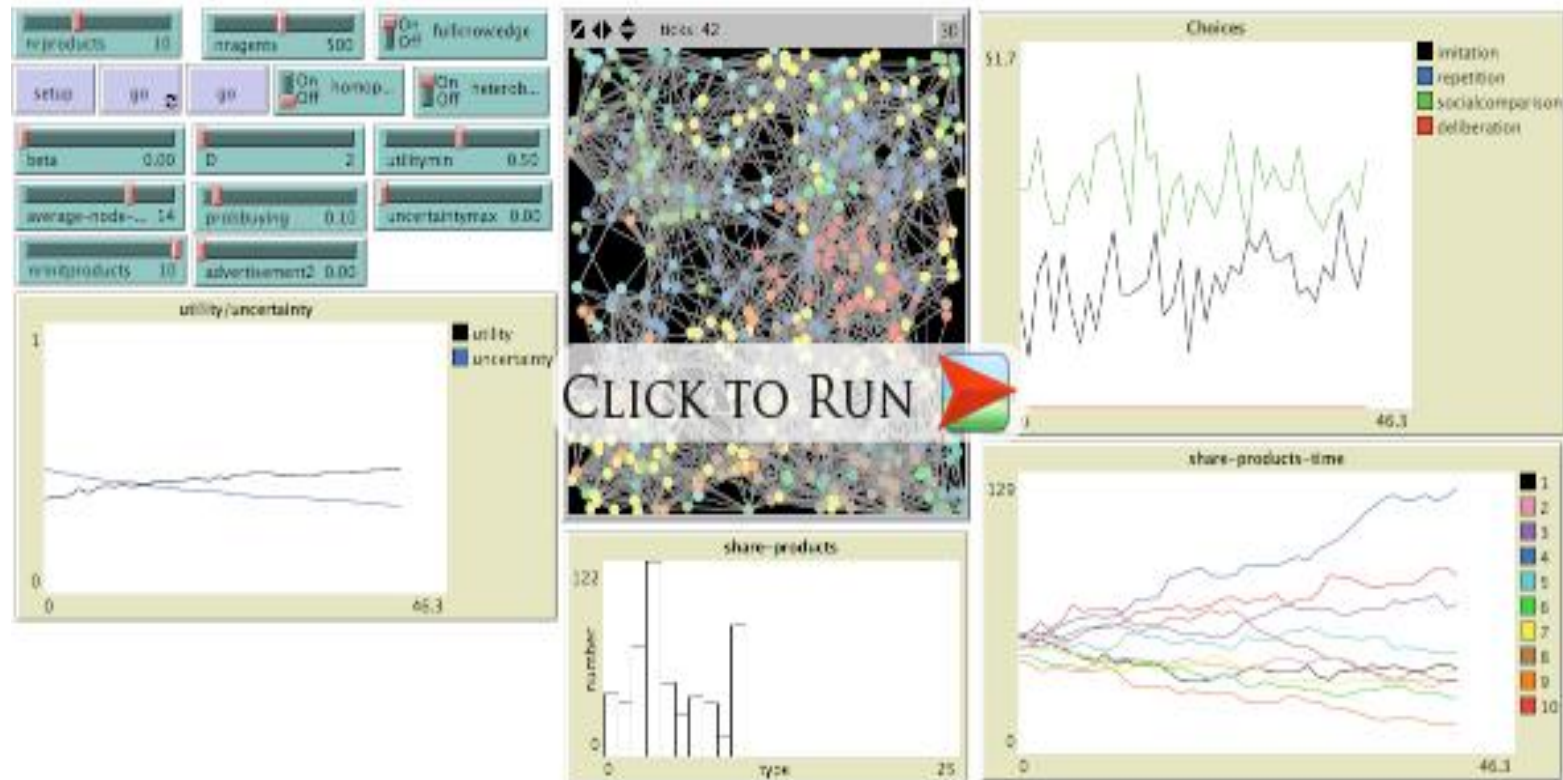
$\beta_i$ : inclinación del consumidor  $i$  hacia adoptar sus propias decisiones o a dejarse sesgar por las de su entorno



| Heurística   | Utilidad           | Incertidumbre      |
|--------------|--------------------|--------------------|
| Repetición   | $U_i \geq U_{min}$ | $Unc_i \leq Unc_t$ |
| Deliberación | $U_i < U_{min}$    | $Unc_i \leq Unc_t$ |
| Imitación    | $U_i \geq U_{min}$ | $Unc_i > Unc_t$    |
| Comp. Social | $U_i < U_{min}$    | $Unc_i > Unc_t$    |

# IMPLEMENTACIÓN DE CONSUMAT EN NETLOGO

<http://www.openabm.org/files/books/3443/ch14-consumats.html>



# INCORPORACIÓN DEL RECONOCIMIENTO DE PRODUCTO (1)

Hemos añadido un nuevo filtro en el comportamiento del consumidor que representa el **reconocimiento (*awareness*)** de los productos

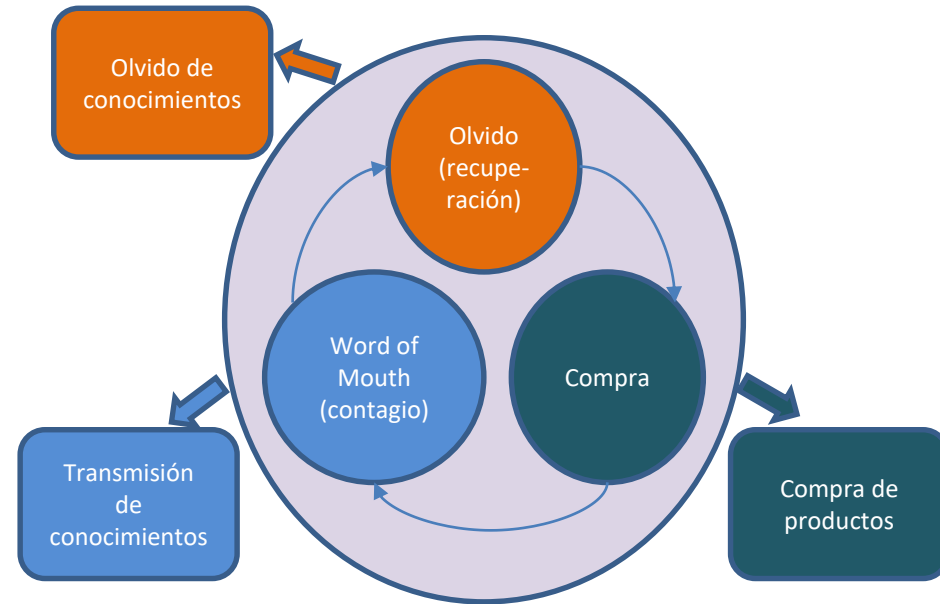
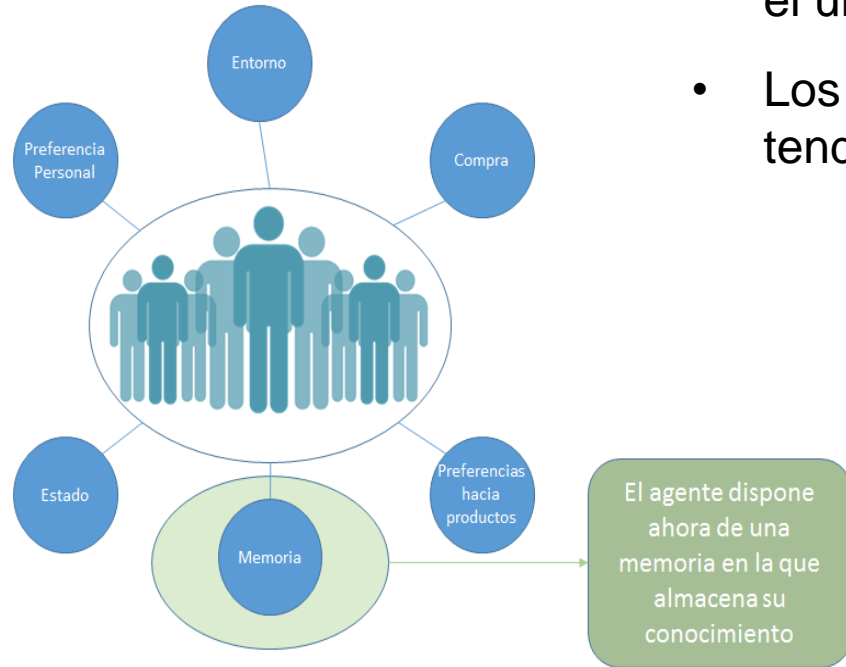
El agente ya no dispone de un conocimiento completo de los productos y **sólo puede adquirir aquellos que conoce** (*racionalidad limitada*). El reconocimiento de cada producto se adquiere mediante el **WoM** con los contactos del agente

Se modela así como un **proceso de contagio SIS** independiente basado en dos variables:

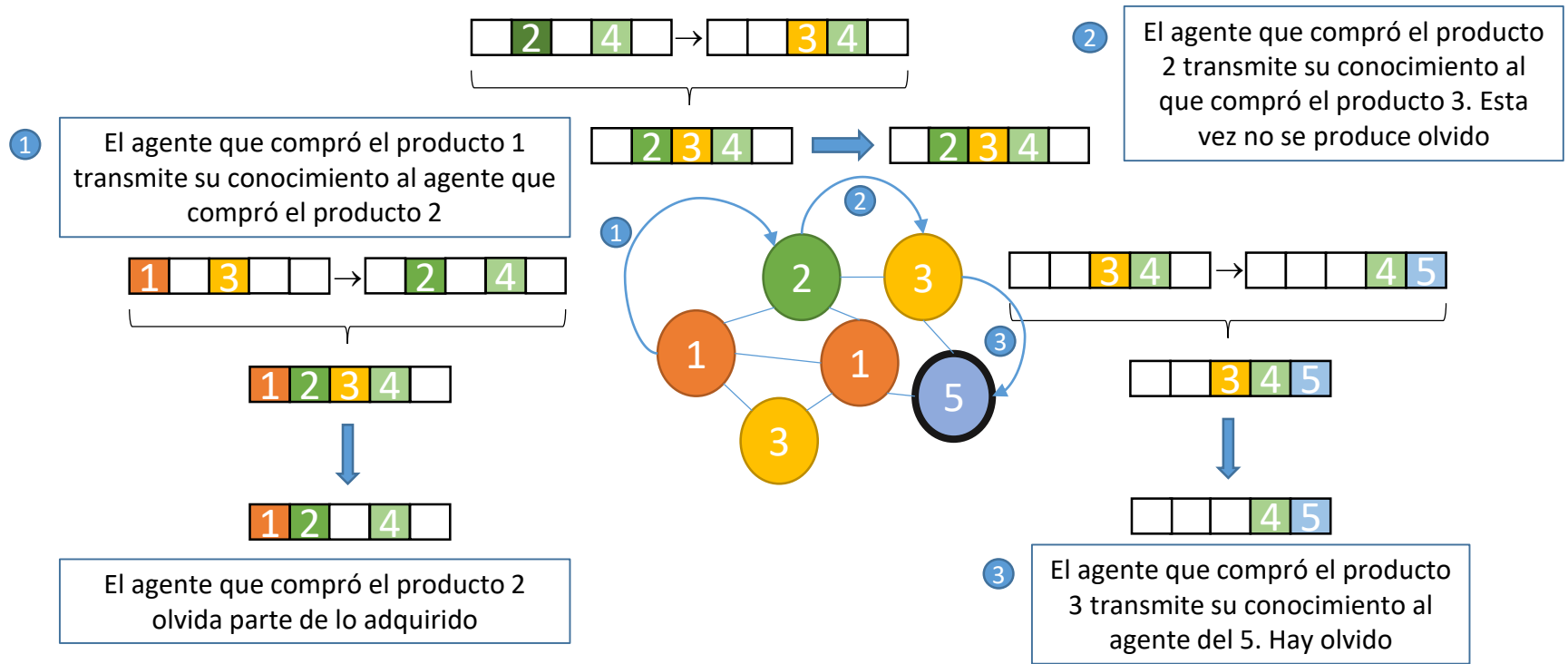
- una **probabilidad de contagio de reconocimiento**  $p_{cj}$ , que permite a los agentes que conocen un producto  $j$  compartirlo con sus contactos en la red
- una **probabilidad de olvido**  $p_{oj}$ , que permite a un agente olvidar el producto  $j$  (dejar de estar infectado)

# INCORPORACIÓN DEL RECONOCIMIENTO DE PRODUCTO (2)

- No se permite que el agente consumidor olvide el último producto que ha adquirido
- Los productos más conocidos (más adquiridos) tendrán mayor probabilidad de ser difundidos



# Ejemplo de funcionamiento



Una vez que se ha transmitido todo el conocimiento, los agentes pueden decidir si realizan su compra y qué compran en el nuevo paso de la simulación



# ALGUNOS EXPERIMENTOS (1)

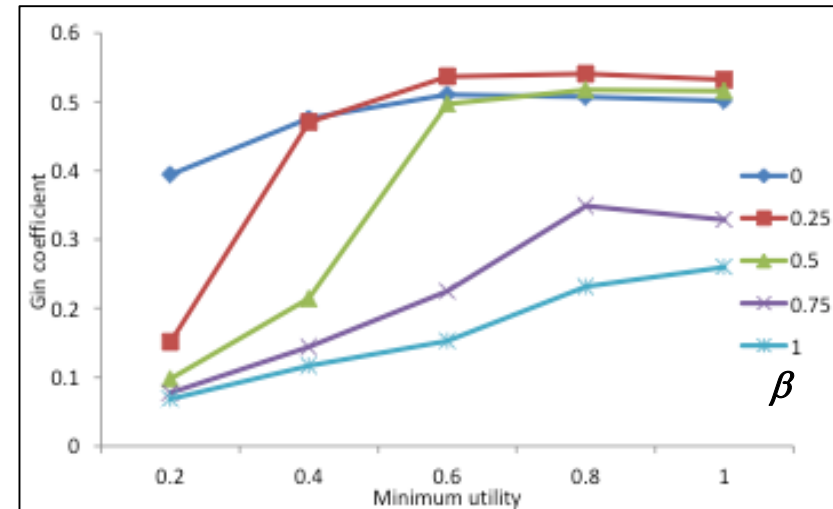
## 1. Modelo sin reconocimiento de productos:

**Indicador:** Dispersión de la distribución de productos comprados a lo largo de la simulación entre la población de agentes (**índice de Gini**)

$N=10$  productos; 500 agentes; probabilidad de compra  $p_b=0.1$ ; valores variables para peso social  $\beta$  y umbral mínimo  $U_{\min}$ , comunes para toda la población; 100 simulaciones (**Montecarlo**) de 100 pasos

El índice Gini es bajo (poca dispersión) si  $\beta$  es alto y  $U_{\min}$  bajo (agentes muy individualistas)

En cuanto  $\beta \leq 0.5$ , la dispersión aumenta y 1-2 productos dominan el mercado, sobre todo si el umbral crece por encima de 0.5 (agentes más sociales y más exigentes)



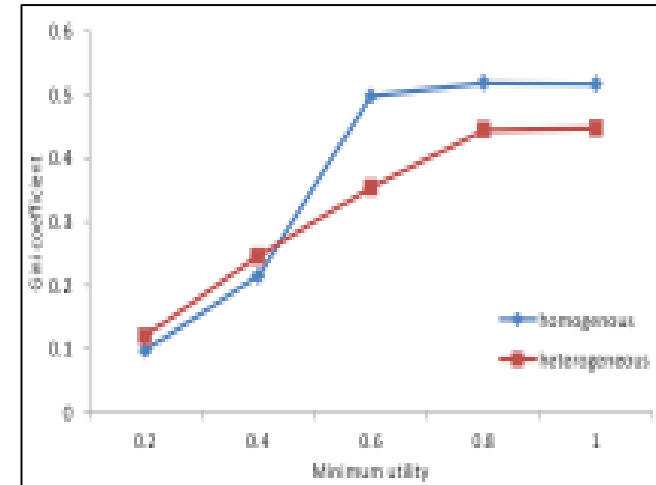
## ALGUNOS EXPERIMENTOS (2)

Los resultados varían si cambian los parámetros de la población de consumidores

P.ej., considerando un valor distinto de  $\beta_i$ , se tiene una situación más realista, en la que coexisten consumidores con distintos comportamientos sociales en la población

Habrán individuos innovadores, que adquieren productos que satisfacen sus preferencias, y otros más lentos para adoptarlos pero más influenciados por su entorno social

Si se compara  $\beta = 0.5$  (**homogéneo**) con  $\beta_i \sim U[0,1]$  (**heterogéneo**), la dispersión de productos es menor, especialmente con umbrales más altos



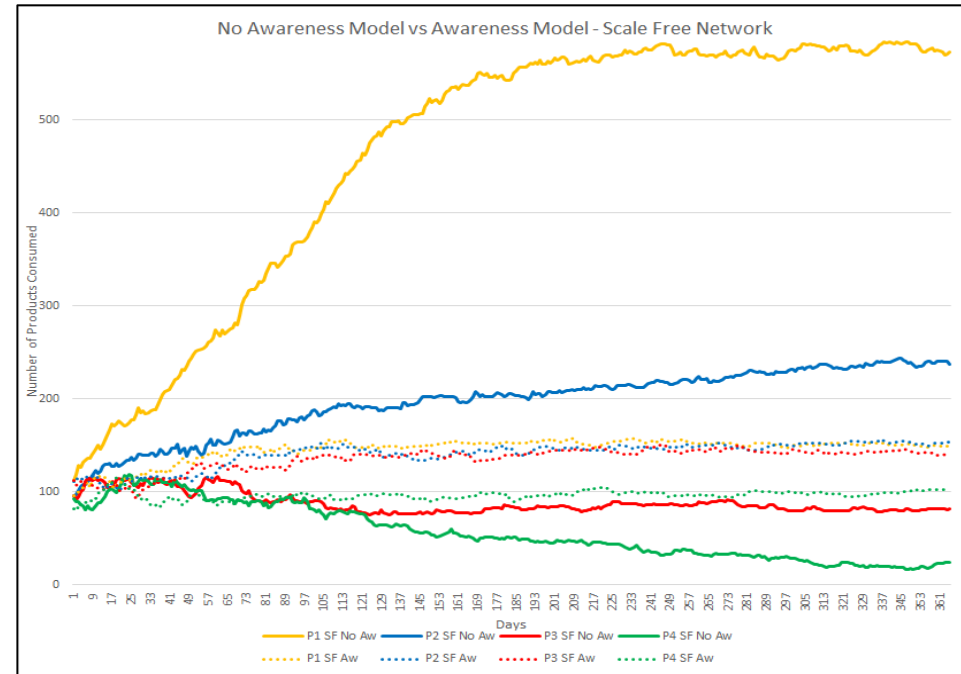
**Los resultados estarían también afectados por la estructura de la red social:** tipo de red, conexiones, agentes “influenciadores”, etc.

## 2. Comparativa modelos con/sin reconocimiento de productos:

$N=10$  productos; 1000 agentes;  $p_b=1$ ;  $\beta=0.6$ ;  $U_{\min}=0.4$ ;  $Unct=0.2$ ; probabilidad de contagio  $p_c=0.4$ ; probabilidad de olvido  $p_o=0.1$ ; comunes para toda la población; grado medio red = 10; 15 simulaciones de 365 pasos

Comparativa de volumen de ventas de los cuatro productos más vendidos, **red libre de escala**:

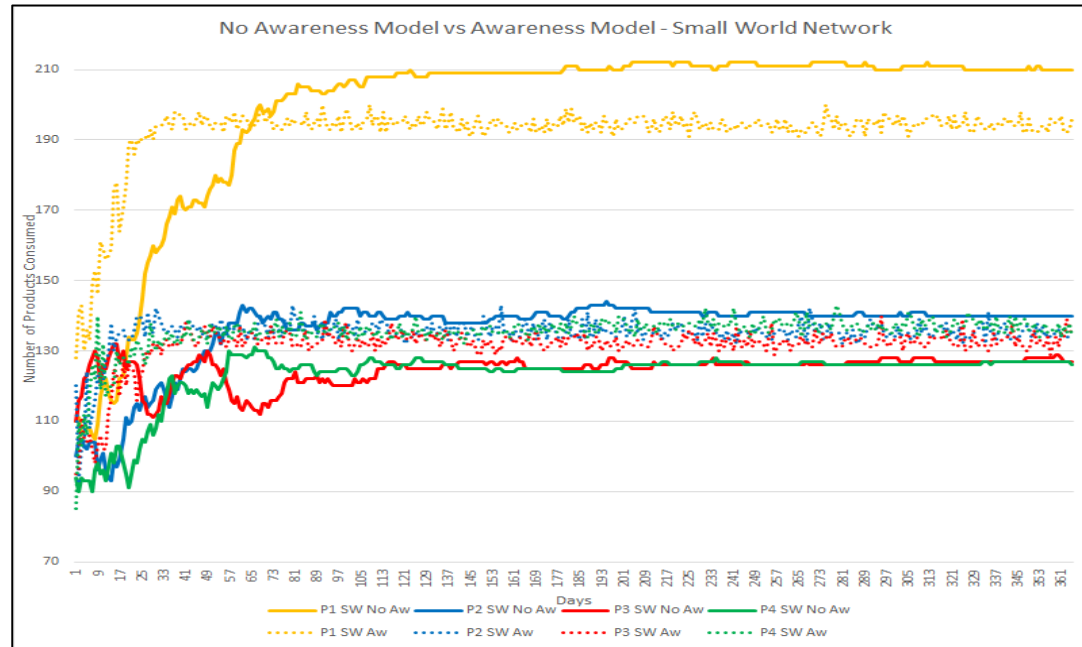
- Sin reconocimiento (línea continua): un producto satura el mercado
- Con reconocimiento (línea discontinua): mucha mayor competencia



## ALGUNOS EXPERIMENTOS (4)

Comparativa de volumen de ventas en la **red de mundo pequeño**:

- El comportamiento de ambos modelos (sin y con reconocimiento) es más similar
- La causa es que el proceso de difusión del reconocimiento es más lento en la red de mundo pequeño que en la libre de escala, que tiene *hubs*



# 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.
- ...

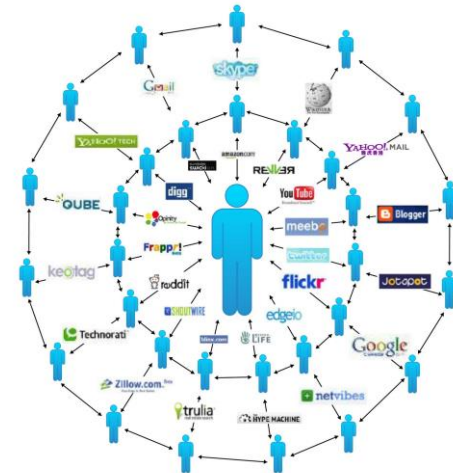
# MARKETING VIRAL: SELECCIÓN DE LAS SEMILLAS

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)

Stonedahl et al. Evolving viral marketing strategies. GECCO 2010

## 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)

En su aproximación, 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**

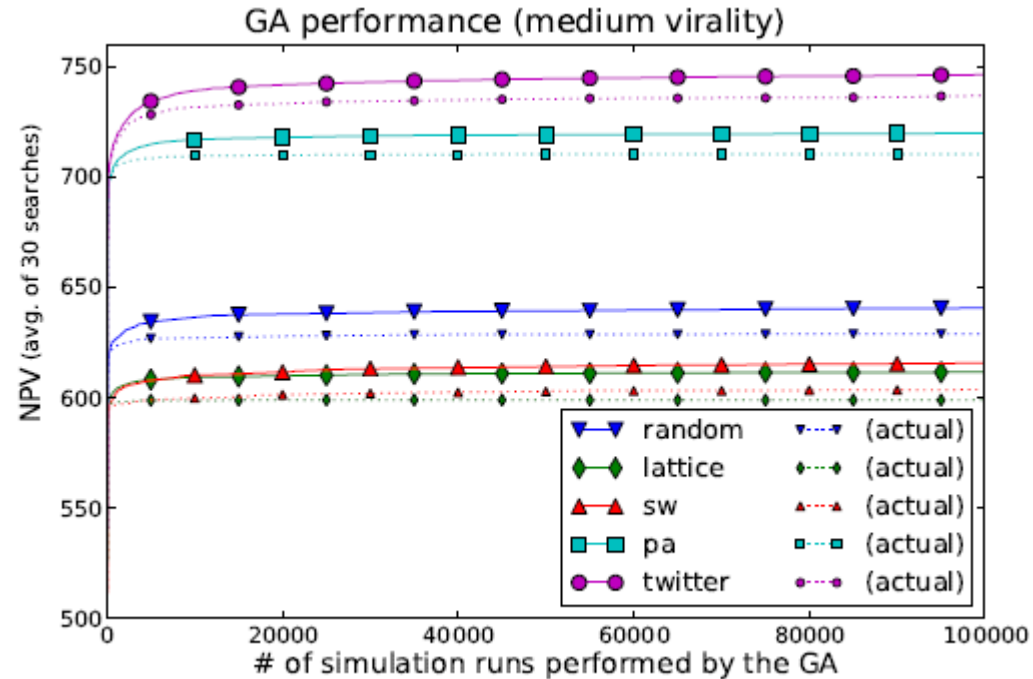
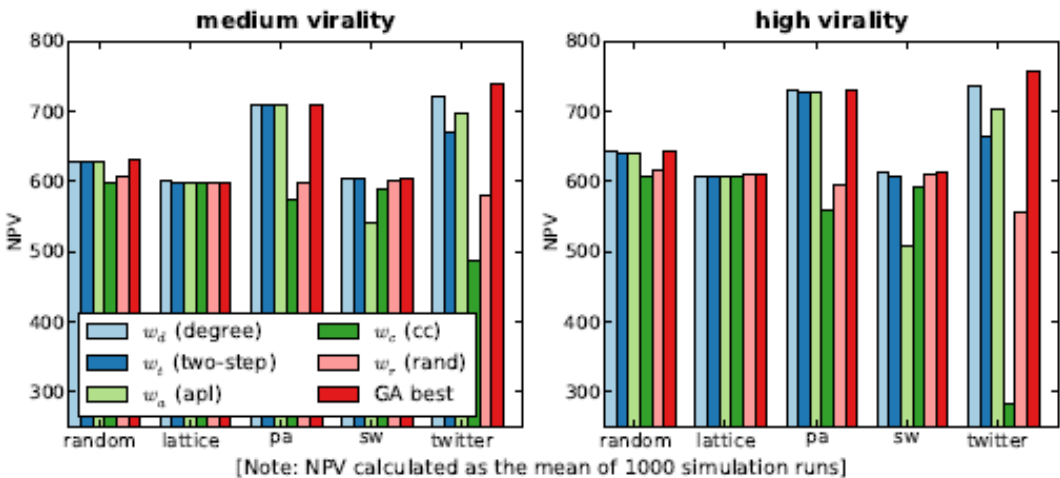


Figure 3: GA progress (averaged across 30 searches) by network topology, for the 'medium virality' scenario. GA's reported best-of-run fitness (solid lines) are compared with the actual NPV values (dotted lines), estimated by 1000 simulation runs, showing the effect of noise. (Error bars too small to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

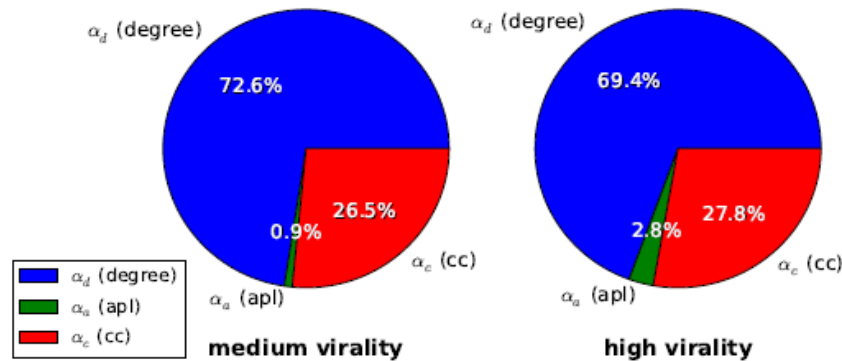


# EXPERIMENTOS (2)

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



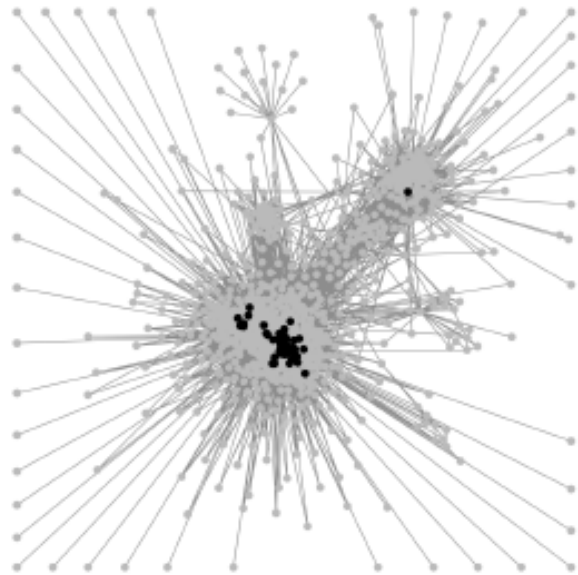
Best primary strategies for twitter network



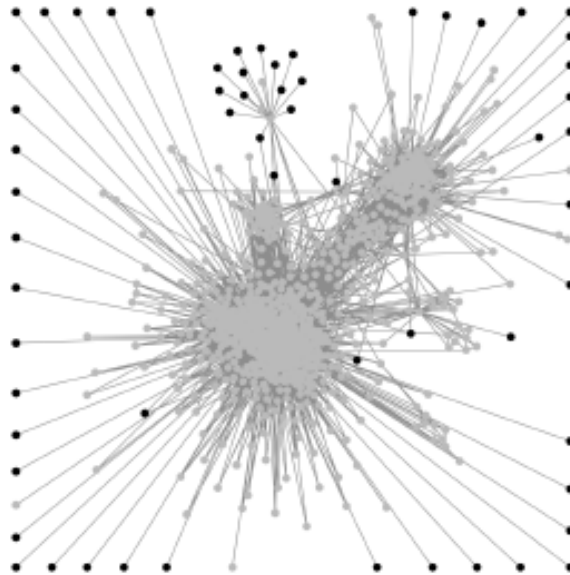
En la red de Twitter, el CC sólo no funciona bien pero la estrategia combinada de grado y CC mejora el NPV en un 2.5% en ambos escenarios de viralidad

## EXPERIMENTOS (3)

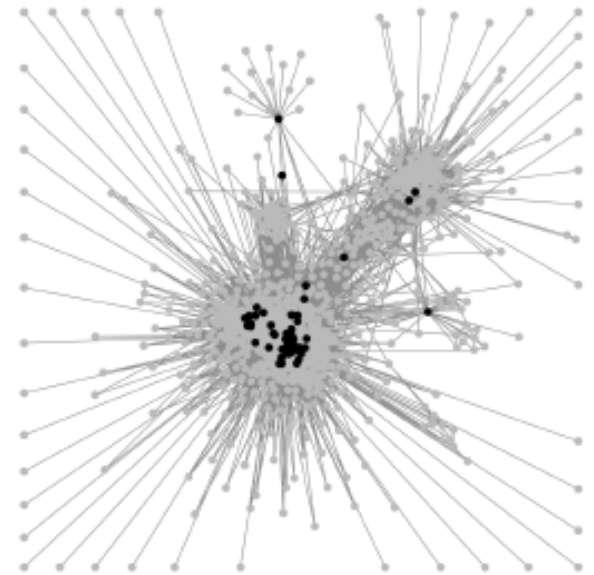
La explicación parece ser que **hay algunos individuos importantes alejados de los hubs centrales que hacen de “*brokers*” a individuos/grupos no conectados directamente a esos hubs**



a) pure degree



b) pure cc



c) GA's best strategy

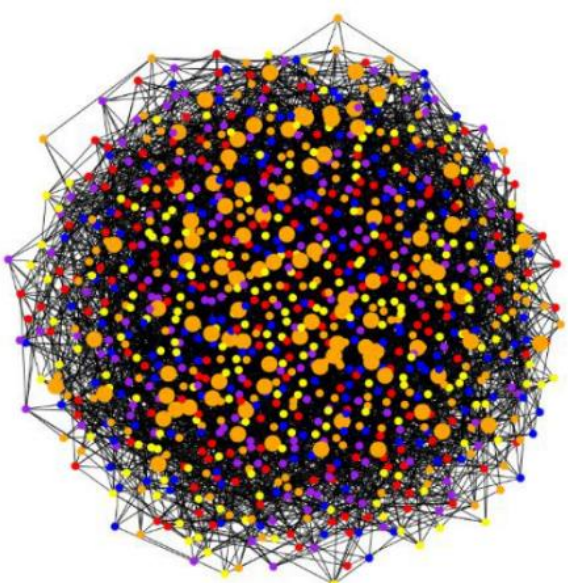
Figure 8: Visualization of three seeding strategies on the *twitter* network.

# EXPERIMENTOS (4)

Incorporación del método de Stonedahl a nuestro CONSUMAT con reconocimiento de productos

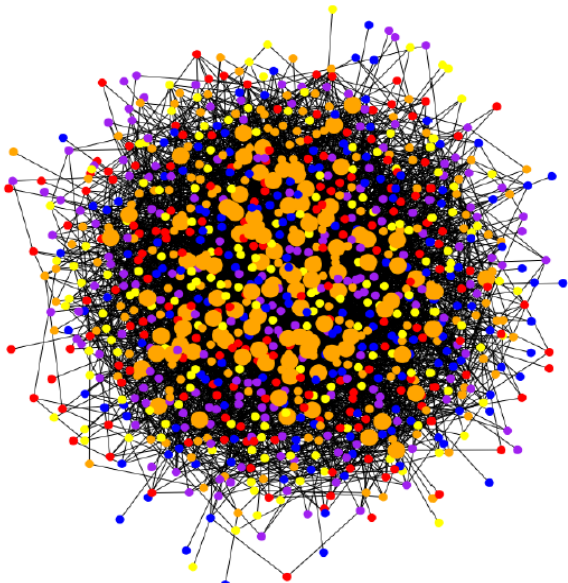
Simulaciones de 1000 agentes durante 365 pasos. Los nodos grandes naranjas identifican las semillas seleccionadas, el resto a consumidores normales indicando el producto adquirido:

| $W_d$  | $W_{2s}$ | $W_{cc}$ | $F_s$ | Benefits | Sales |
|--------|----------|----------|-------|----------|-------|
| 0.5292 | 0.6703   | 0.6168   | 82    | 2678.651 | 621   |



Red libre de escala

| $W_d$ | $W_{2s}$ | $W_{cc}$ | $F_s$ | Benefits | Sales |
|-------|----------|----------|-------|----------|-------|
| 0.355 | 0.629    | 0.122    | 99    | 718.4988 | 948   |



Red de mundo pequeño