

Escuela de Verano de Inteligencia Artificial (EVIA 2016)



Modelado de Procesos de Difusión en Redes Sociales Aplicación a Procesos de Marketing Viral

Oscar Cordón García

Dpto. Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada

ocordon@decsai.ugr.es

COMPLEJIDAD, SISTEMAS COMPLEJOS Y REDES COMPLEJAS

Hay Sistemas Complejos en Todo Nuestro Entorno...

que son muy difíciles de comprender y analizar como:

- Las **sociedades biológicas** (p.e., colonias de hormigas, abejas, manadas, etc.) y la propia **sociedad humana**, cuyo funcionamiento requiere cooperación entre billones de individuos
- El **cerebro humano**, que requiere una actividad coherente de billones de neuronas
- Los **genomas** de los individuos, basados en la interacción constante entre miles de genes
- Los **sistemas financieros**, como por ejemplo el mercado de valores (la bolsa), en los que interaccionan millones de componentes a nivel mundial
- Los **sistemas de transmisiones**, como las redes de telecomunicaciones (ej: Internet) o las **redes de distribución eléctrica**
- etc.

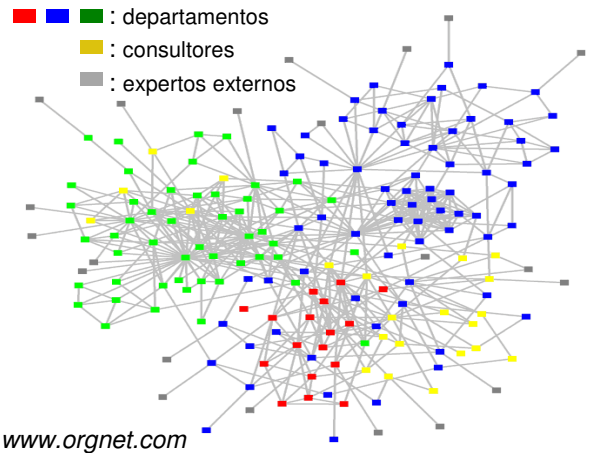
CARACTERÍSTICAS DE LOS SISTEMAS COMPLEJOS

- Compuesto de agentes simples (**auto-organizado**)
- **Descentralizado:** No existe un supervisor, no hay un control central
- **Comportamiento emergente:** No hay un plan global, se producen acciones complejas e inesperadas a partir de reglas simples
- **Robusto:** Las actuaciones se completan aunque un individuo falle

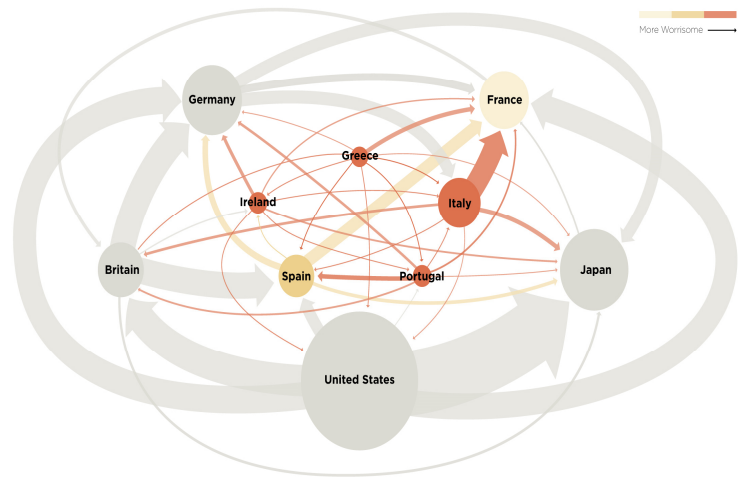


EJEMPLOS DE SISTEMAS COMPLEJOS

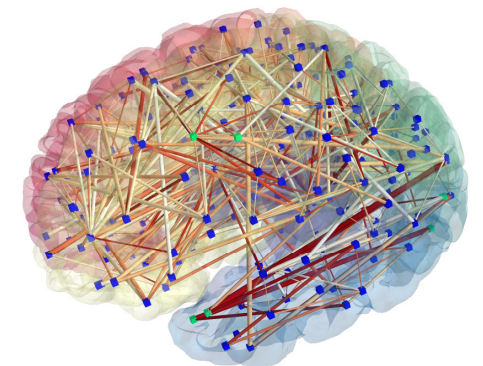
Organizaciones



Sistemas financieros

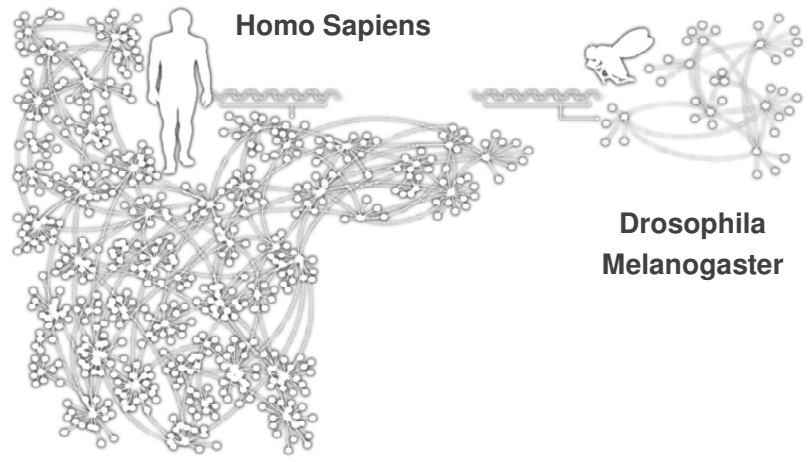


El cerebro

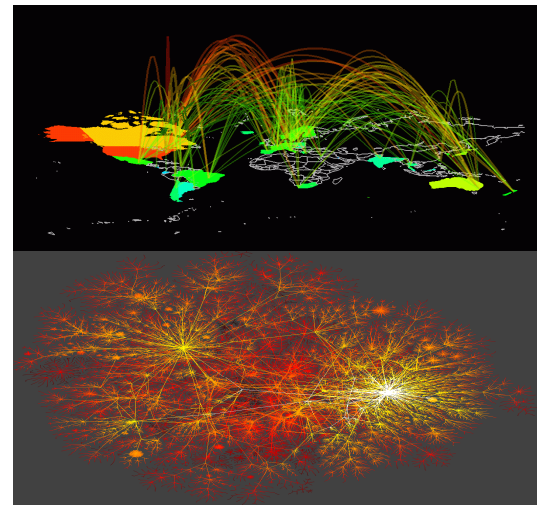


<http://www.humanconnectomeproject.org/>

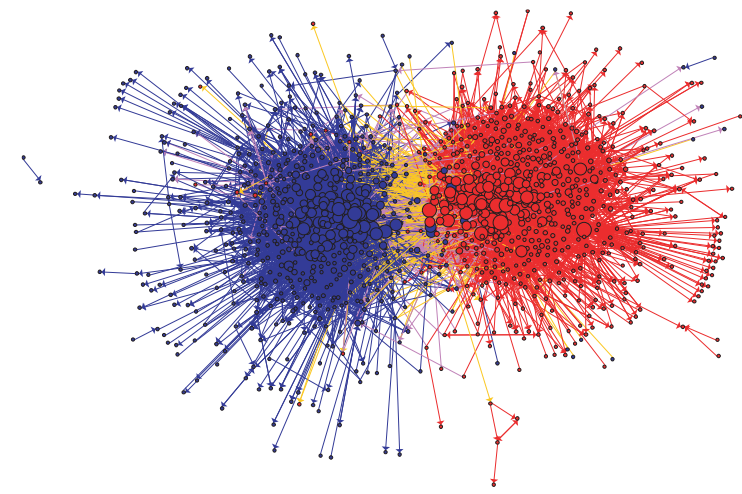
<http://www-personal.umich.edu/~mejn/netdata/>



Genomas



Red física de Internet



Sistemas políticos

COMPLEJIDAD

La comprensión, la predicción e incluso el control de sistemas complejos ha adquirido una importancia capital en la actualidad

Complex

[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]

—adjective

1.

composed of many interconnected parts; compound; composite: a complex highway system.

2.

characterized by a very complicated or involved arrangement of parts, units, etc.: complex machinery.

3.

so complicated or intricate as to be hard to understand or deal with: a complex problem.

Source: Dictionary.com

Complexity, a **scientific theory** which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behaviour, seem to occur in many complex systems involving living organisms, such as a stock market or the human brain.

Source: John L. Casti, Encyclopædia Britannica

Complexity

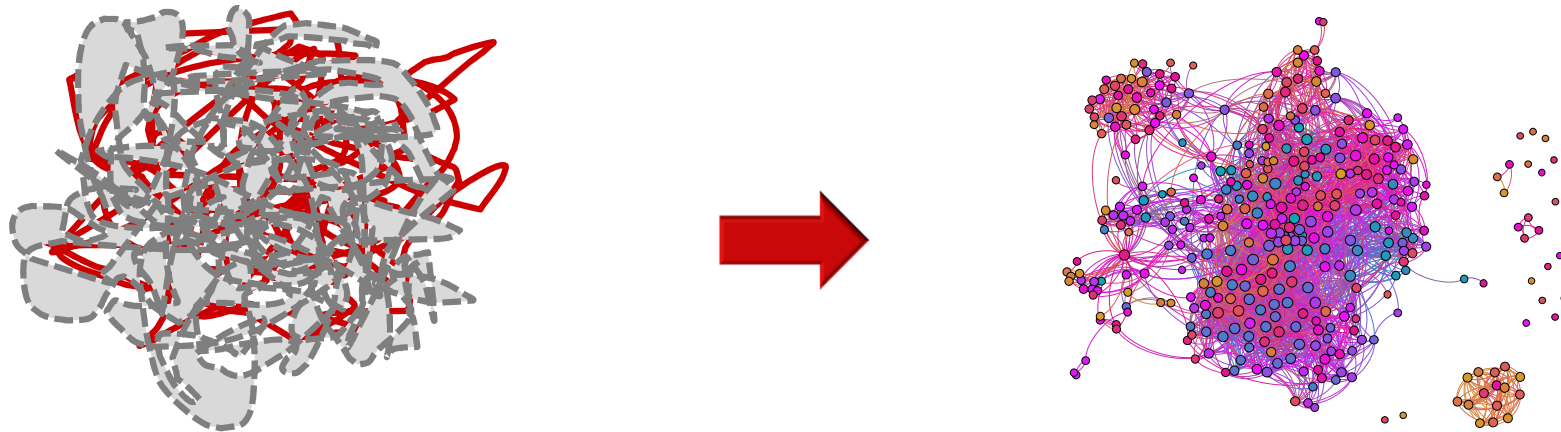


*“I think the next century
will be the century
of complexity.”*

Stephen Hawking
January 23, 2000

EL ROL DE LAS REDES

Detrás de cada sistema complejo estudiado siempre hay un diagrama de conexiones, una **red**, que define las interacciones entre sus componentes



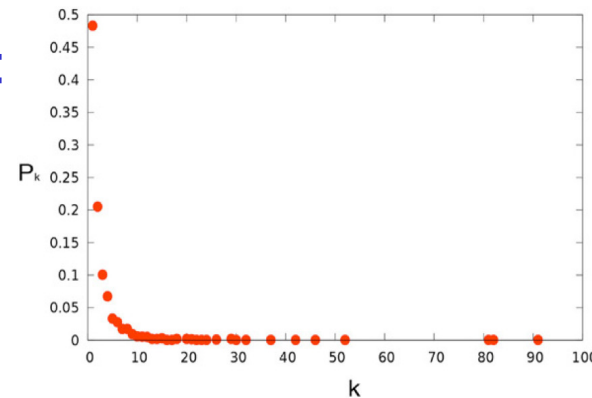
No seremos capaces de entender los sistemas complejos a menos que podamos mapear y comprender las redes que los soportan

A pesar de las diferencias aparentes en componentes e interacciones, **las redes que regulan los distintos sistemas complejos existentes en nuestro mundo son similares, siguen unas leyes comunes y presentan mecanismos reproducibles**

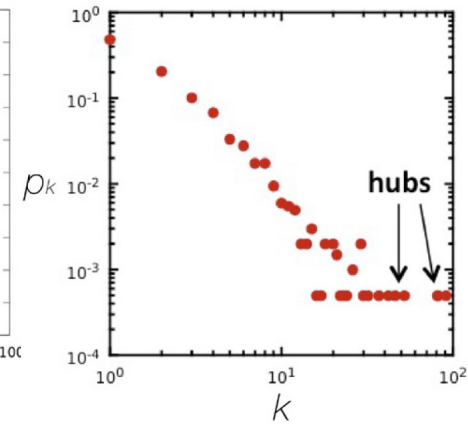
ANÁLISIS Y MODELADO DE REDES COMPLEJAS

La **Ciencia de Redes** se dedica al análisis de redes reales. Se usan medidas como:

Distribución (de probabilidad) de grados p_k :
Probabilidad de encontrar nodos con k enlaces
(grado medio, **presencia de hubs**, etc.)



Distancia media (de los caminos entre todos los nodos de la red) $\langle d \rangle$



Coefficiente de Clustering C_i : Mide la densidad local de la red, ¿qué proporción de los vecinos de cada nodo están conectados?

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

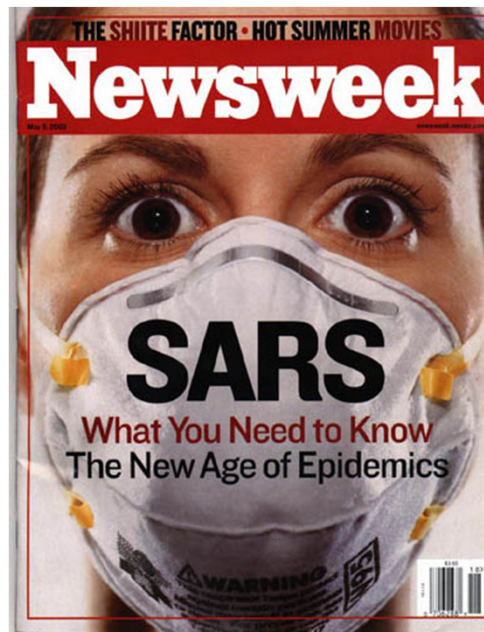
$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$

Los valores de estas medidas en una red concreta determinan el comportamiento dinámico del sistema complejo

Existen distintos tipos de redes en función de estas características:
regulares, aleatorias, mundos pequeños, libres de escala, ...

PROCESOS EPIDÉMICOS Y DE DIFUSIÓN

¿POR QUÉ ES TAN IMPORTANTE EL PROCESO DE PROPAGACIÓN?



Epi + demos *sobre pueblo*



<http://es.wikipedia.org/wiki/Epidemia>

Biología:

- Enfermedades transmitidas por el aire (gripe, gripe aviar, ...)
- Enfermedades venéreas (VIH, ...)
- Otras enfermedades infecciosas, incluidos algunos cánceres (VPH (virus del papiloma humano, ...))
- Parásitos (chinchas, malaria, ...)

TIC:

- Virus de ordenador, gusanos
- Virus de teléfonos móviles

Conceptos/Aspectos intelectuales:

- Difusión de innovaciones
- Rumores, memes, ...
- Prácticas empresariales

MODELOS CLÁSICOS DE PROPAGACIÓN DE EPIDEMIAS

[http://es.wikipedia.org/wiki/Modelaje matemático de epidemias](http://es.wikipedia.org/wiki/Modelaje_matemático_de_epidemias)

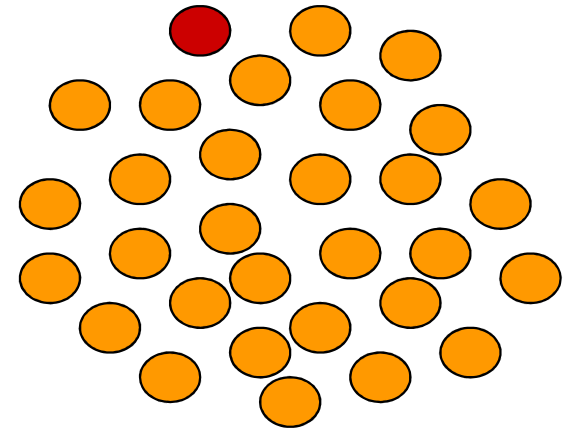
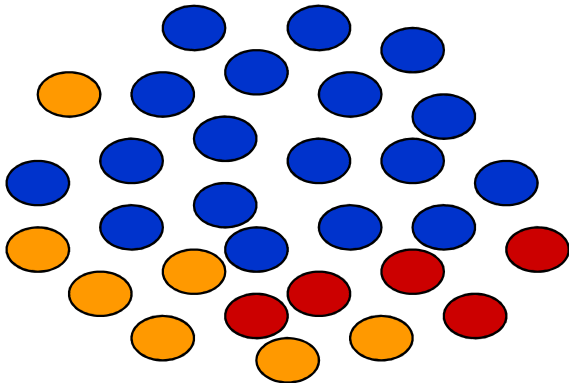
[http://www.uni-tuebingen.de/modeling/Mod Pub Software SIR en.html](http://www.uni-tuebingen.de/modeling/Mod_Pub_Software_SIR_en.html)

HIPÓTESIS FUNDAMENTALES DEL MODELADO EPIDÉMICO

1. Compartimentación:

Los individuos se clasifican en varios estados según la etapa de desarrollo de la enfermedad. La clasificación más sencilla incluye tres estados:

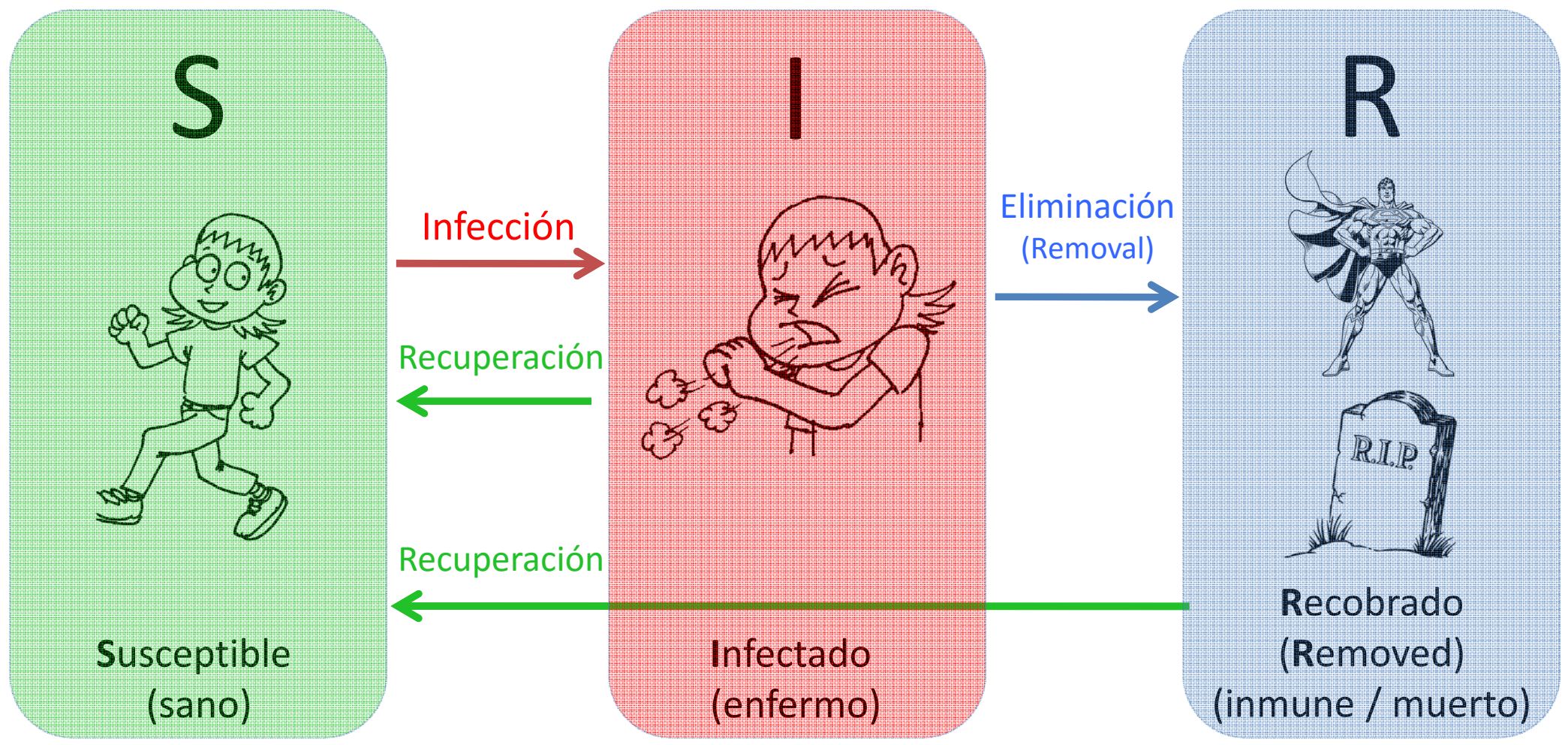
- *Susceptible (S)*: Individuos sanos que no han contraído la enfermedad
- *Infectado (I)*: Individuos contagiados que la han contraído y pueden infectar a otros
- *Recobrado (R)*: Individuos previamente infectados que se han recuperado de la enfermedad y ya no son infecciosos



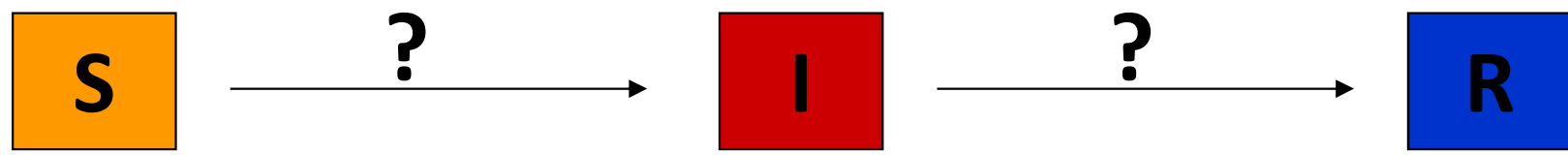
2. Mezclado Homogéneo:

Esta hipótesis asume que todos los individuos tienen la misma probabilidad de entrar en contacto con un individuo infectado. **Elimina la necesidad de conocer la red de contactos específica** sobre la que se difunde la enfermedad, reemplazándola con la asunción de que cualquiera puede infectar al resto.

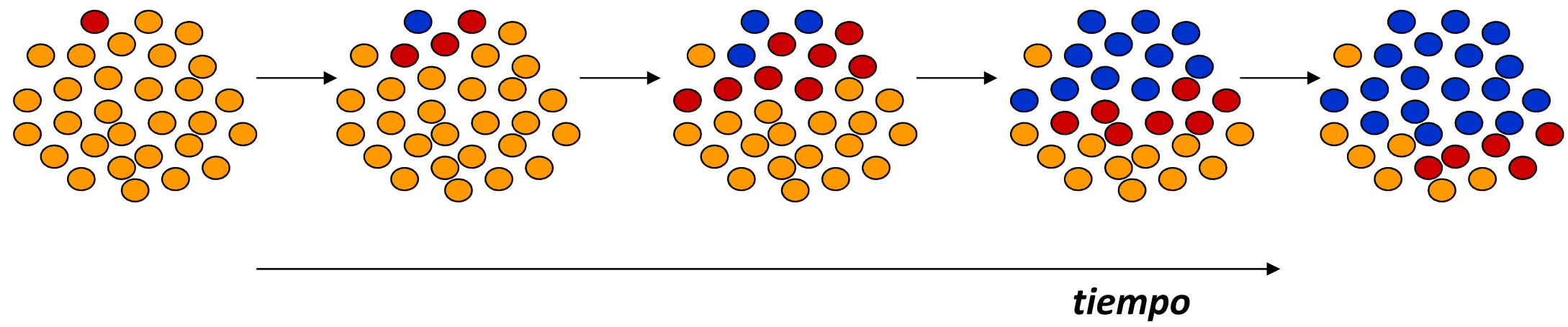
ESTADOS BÁSICOS Y TRANSICIONES DEL MODELO CLÁSICO SIR



DINÁMICA DE LAS VARIANTES DEL MODELO SIR



¿Dinámica del proceso epidémico?



COMPORTAMIENTOS IMPORTANTES DE LOS MODELOS EPIDÉMICOS

Comportamiento temprano: Patrón de comportamiento de la epidemia en las fases iniciales. Es importante porque:

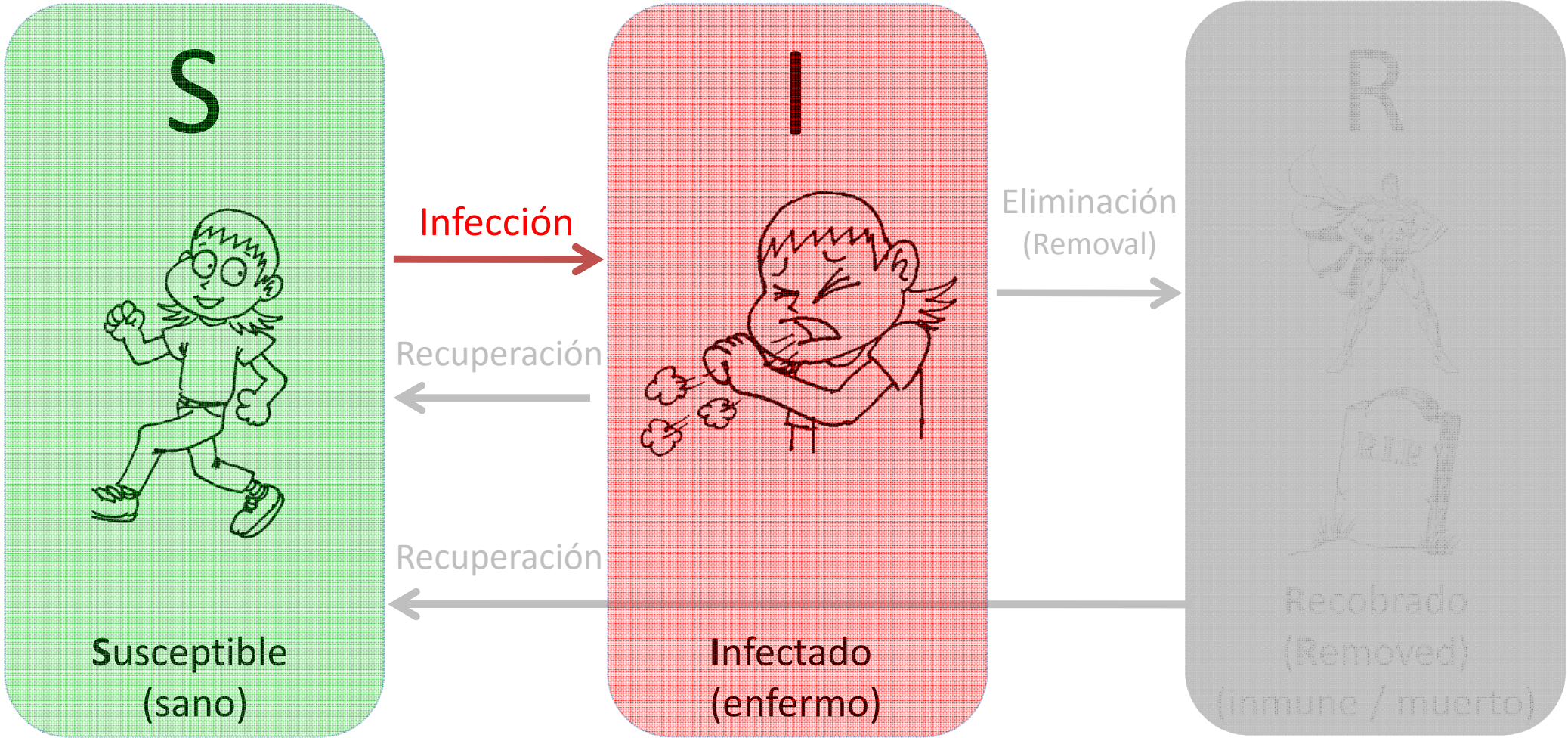
- Se necesita tiempo para desarrollar las vacunas y las intervenciones médicas
- La mejor forma de detener o contener la epidemia es la **cuarentena y/o la vacunación tempranas**

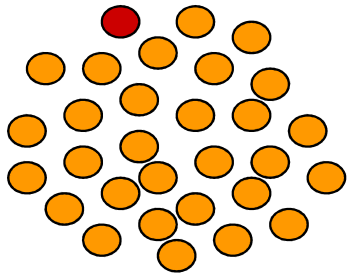
Comportamiento tardío: Patrón de comportamiento de la epidemia en las fases finales (cuando $t \rightarrow \infty$). Es importante porque:

- Permite medir (y por tanto predecir) el alcance de la epidemia, el número de individuos afectados (su pico), etc.

MODELO SI

Modelo teórico más simple





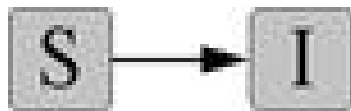
Mezclado Homogéneo: En cada unidad de tiempo, cada individuo tiene $\langle k \rangle$ contactos con otros individuos de la población escogidos aleatoriamente

La probabilidad que un individuo quede infectado por esos contactos es $\beta \in [0,1]$. El ratio de transmisión de la enfermedad es $\beta \cdot \langle k \rangle$

Si en la población de tamaño N hay I individuos infectados y S individuos susceptibles en el instante t , la probabilidad de encontrar un individuo infectado es $i=I(t)/N$ y susceptible es $s=S(t)/N$; $i+s=N$

La enfermedad solo se produce cuando hay contacto entre un individuo infectado y otro susceptible. En cada instante de tiempo, un individuo infectado tiene contacto con un promedio de $\langle k \rangle \cdot s$ individuos sanos, a los que transmite la enfermedad con probabilidad β (**CONTAGIO SIMPLE**)

Como hay i individuos infectados, el **ratio medio de nuevas infecciones** en cada instante de tiempo t es $\beta \cdot \langle k \rangle \cdot s \cdot i$ y las ecuaciones diferenciales del sistema son:



$$\frac{di}{dt} = \beta \cdot \langle k \rangle \cdot s \cdot i \quad ; \quad \frac{ds}{dt} = -\beta \cdot \langle k \rangle \cdot s \cdot i$$

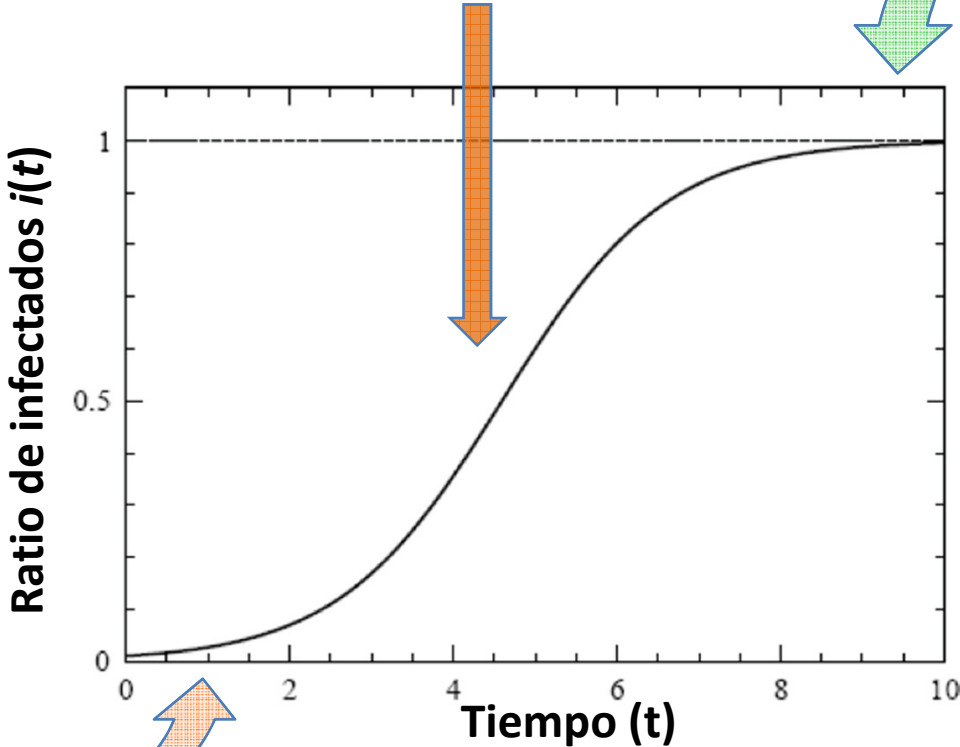
MODELO SI

Comportamiento

$$i(t) = \frac{i_0 \exp(\beta \cdot \langle k \rangle \cdot t)}{1 - i_0 + i_0 \exp(\beta \cdot \langle k \rangle \cdot t)}$$

La pendiente de la curva depende del ratio de transmisión $\beta \cdot \langle k \rangle$

Al principio, sólo hay uno o unos pocos individuos infectados
brote exponencial

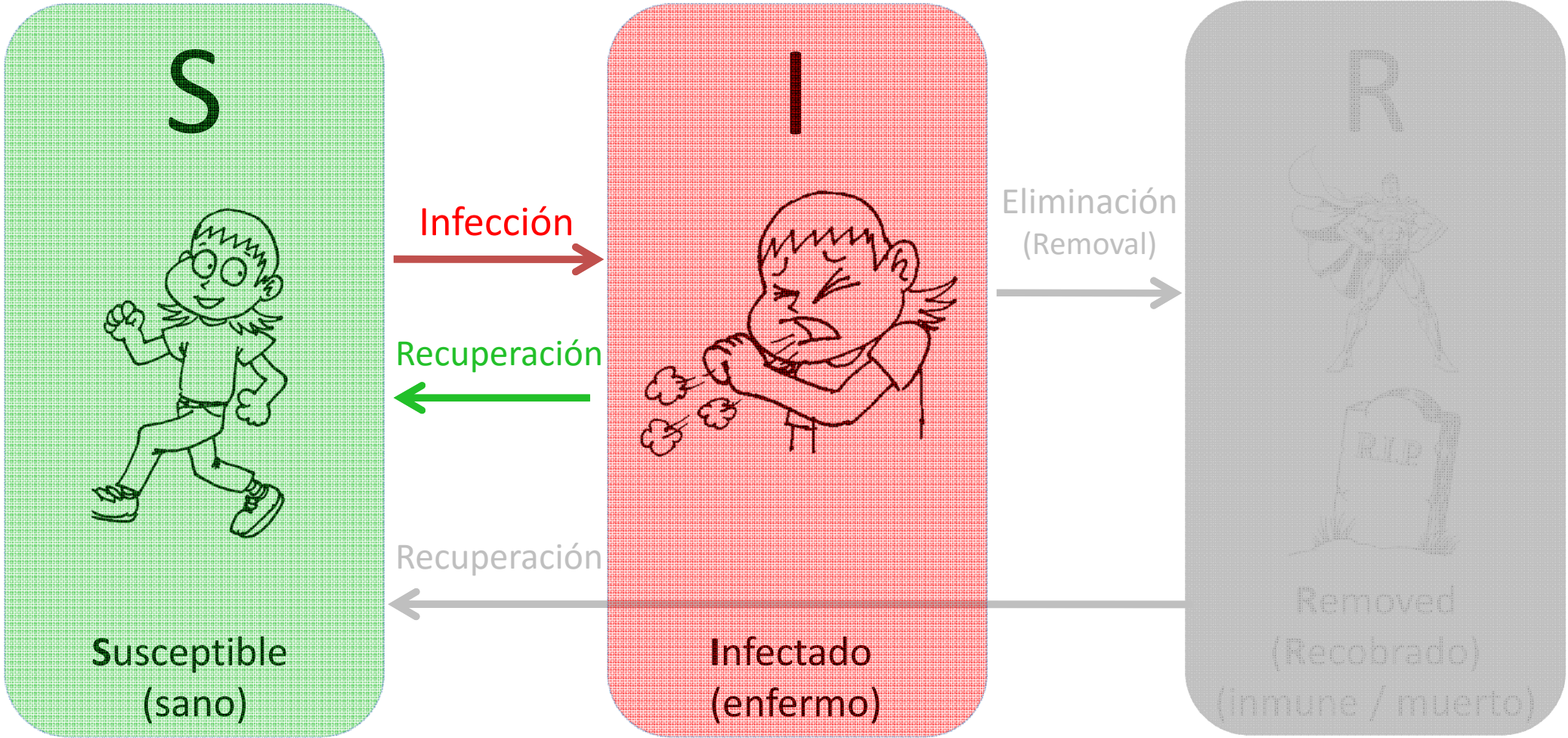


Al final, toda la población se infecta, aunque en cada paso hay menos nuevos infectados
saturación

Modelo SI: el ratio de infectados aumenta hasta que todo el mundo queda infectado

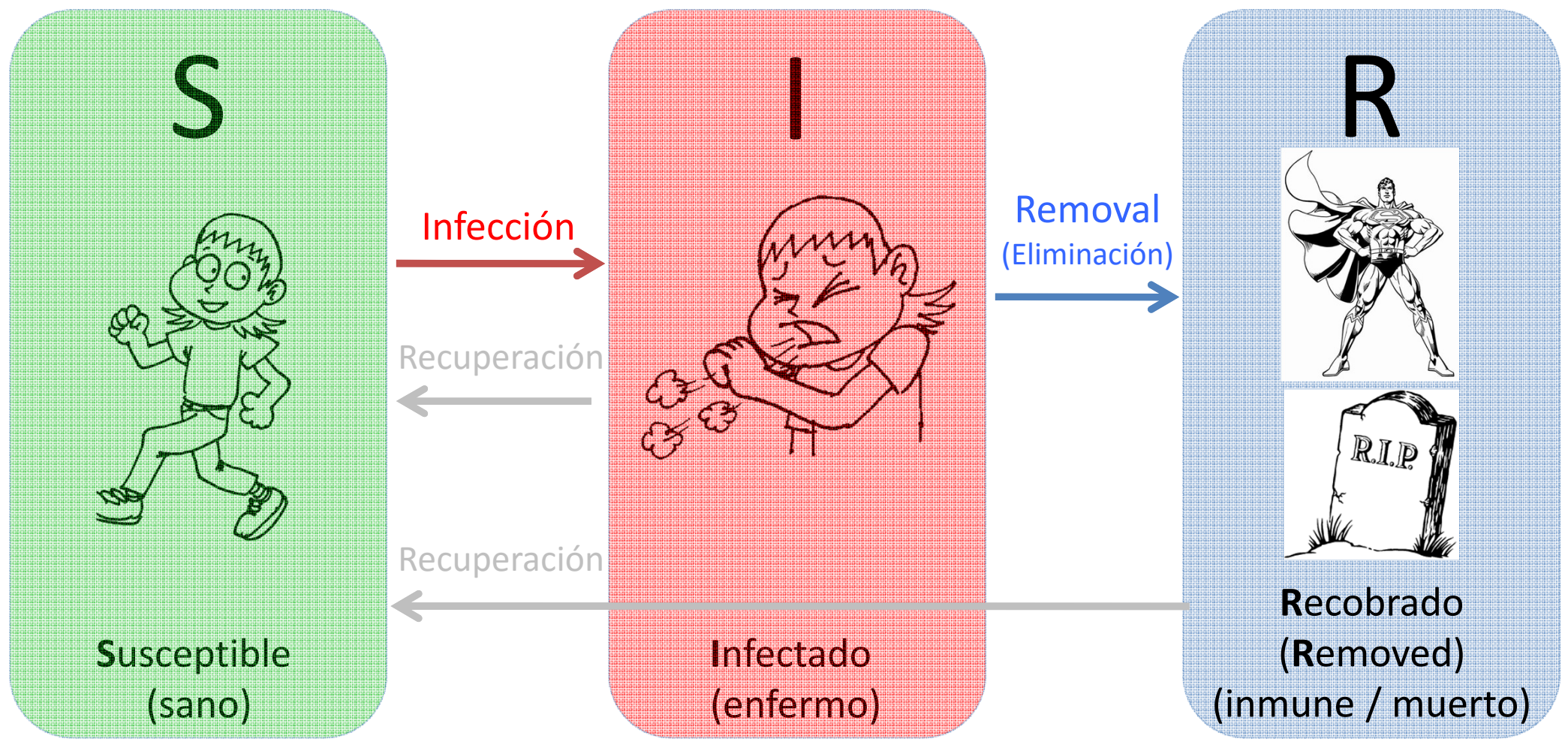
MODELO SIS

Resfriado común



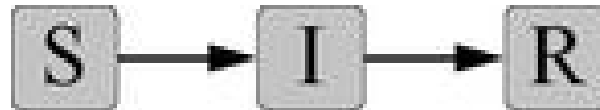
MODELO SIR

Gripe (aviar), plagas, virus informáticos, difusión de información, ...



El modelo SIR extiende al SI considerando la posibilidad de **recuperación/eliminación** de los individuos pero sin **reinfección**

Una vez que un individuo se infecta, puede quedar **permanente e inmune** a la enfermedad o “**morir**”



Aplica el mezclado homogéneo con tres estados: S , I y R . Los individuos infectados pasan periódicamente al estado R (dos transiciones posibles: $S \rightarrow I$ y $I \rightarrow R$)

A efectos del modelo, el estado R representa tanto la inmunidad como el fallecimiento. En ambos casos, el individuo deja de ser susceptible

La dinámica del proceso tiene dos fases. En la primera, los individuos infectados infectan a los susceptibles (**ratio de transmisión $\beta \cdot \langle k \rangle$**). En la segunda, los infectados se recuperan (o mueren) con una **probabilidad de recuperación μ**

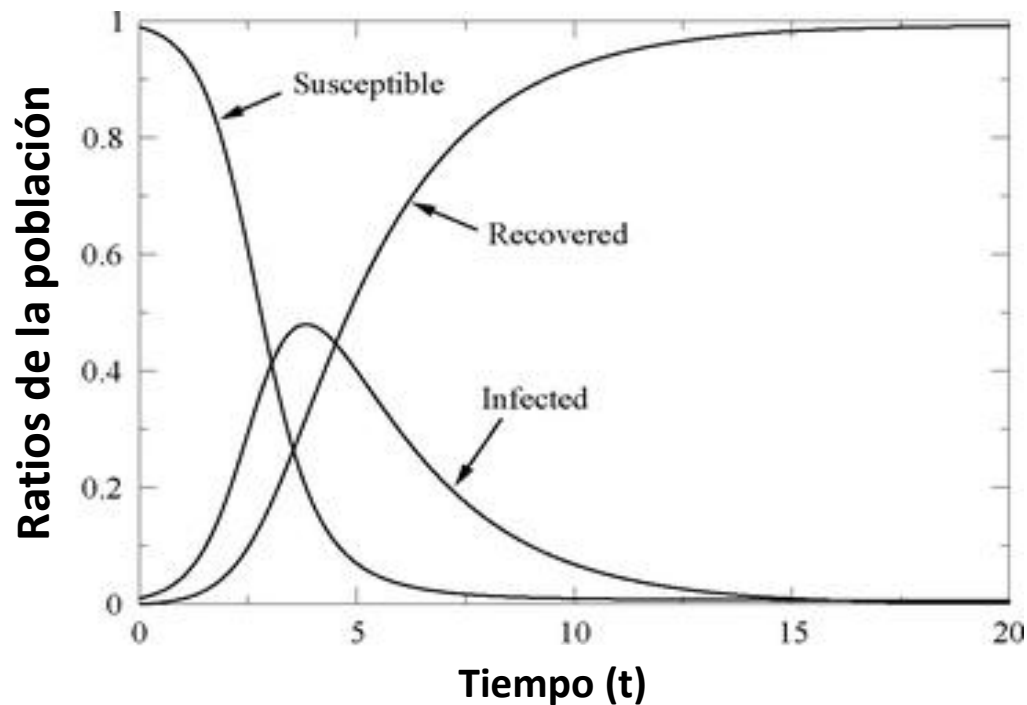
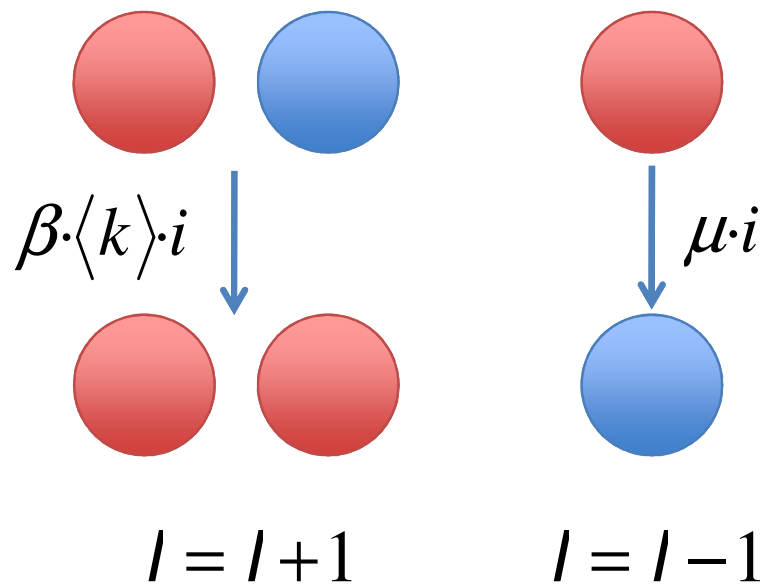
MODELO SIR

Umbral Epidemiológico y Número Reproductivo Básico

$$\frac{di}{dt} = \underbrace{\beta \cdot \langle k \rangle \cdot i}_{I} \cdot \underbrace{(1 - i)}_S - \underbrace{\mu \cdot i}_{I \rightarrow R}$$

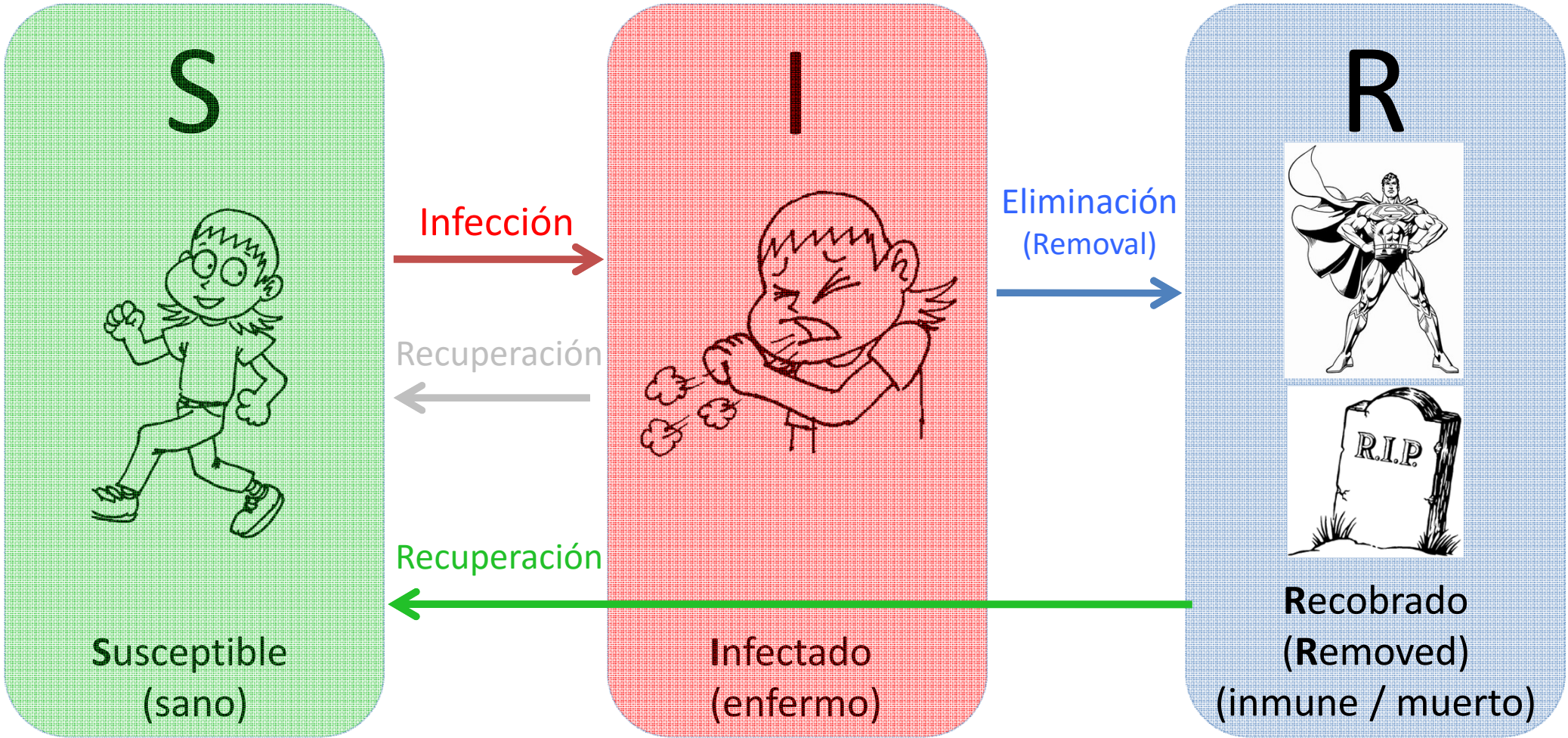
La relación entre los que se infectan ($\beta \cdot \langle k \rangle$) y los que se recuperan (μ) define un “**umbral epidemiológico**”

Si $\beta \cdot \langle k \rangle < \mu$: **Estado sano** (extinción del brote)
Si $\mu > \beta \cdot \langle k \rangle$: **Estado endémico** (epidemia)



VARIANTES

Modelo SIRS: virus informáticos, difusión de información, ...

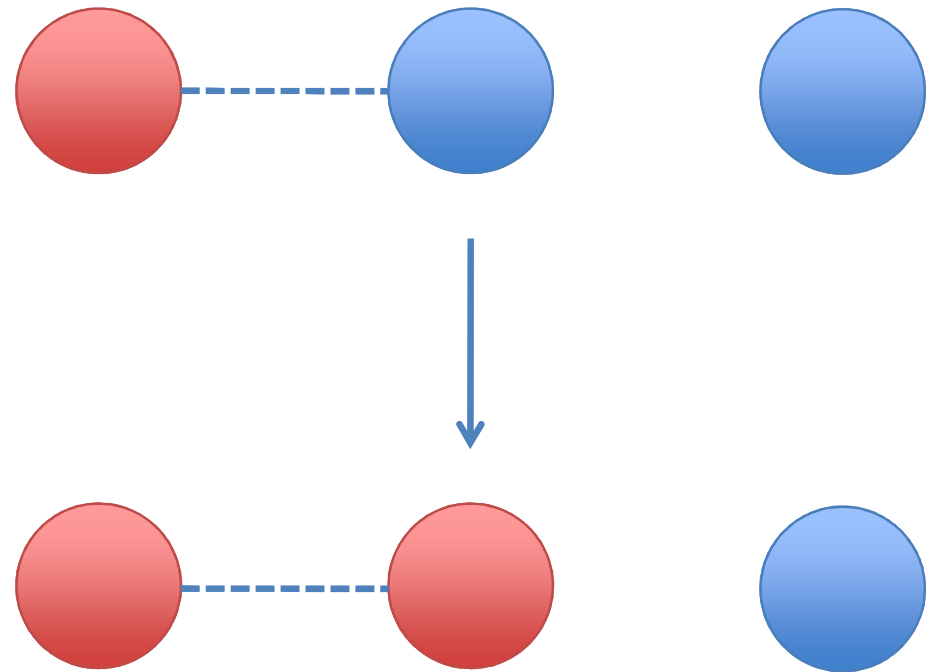


MODELOS DE PROPAGACIÓN DE EPIDEMIAS BASADOS EN REDES

¡La propagación de epidemias implica siempre una estructura de red subyacente!

La propagación solamente ocurre cuando los portadores de las enfermedades/virus/ideas están **conectados entre sí**

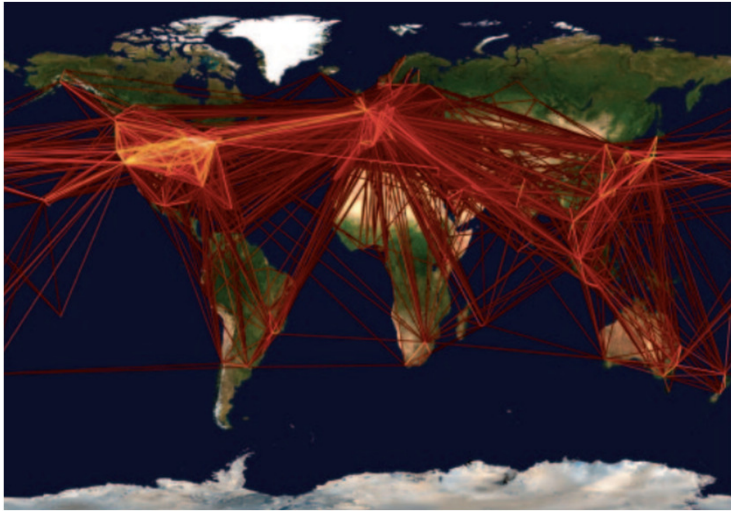
Las **características de la red de contactos condicionan totalmente el proceso de difusión**



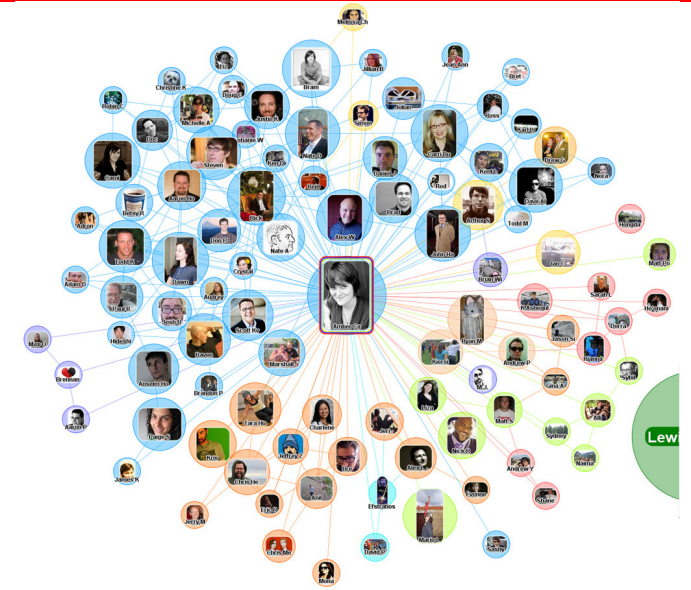
REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (2): Tipos

Fenómeno	Red compleja	Agente
Enfermedad venérea	Red sexual	Patógenos
Otras enfermedades infecciosas	Red de contactos, red de transporte	Patógenos
Propagación de rumores	Red de comunicaciones	Información, memes
Difusión de la innovación	Red de comunicaciones	Ideas
Gusanos de Internet	Internet	Malware (códigos binarios)
Virus de teléfonos móviles	Red social / Red de proximidad	Malware (códigos binarios)
Chinches	Red de hoteles – viajeros	Chinches
Malaria	Red de mosquitos – humanos	<i>Plasmodium</i>

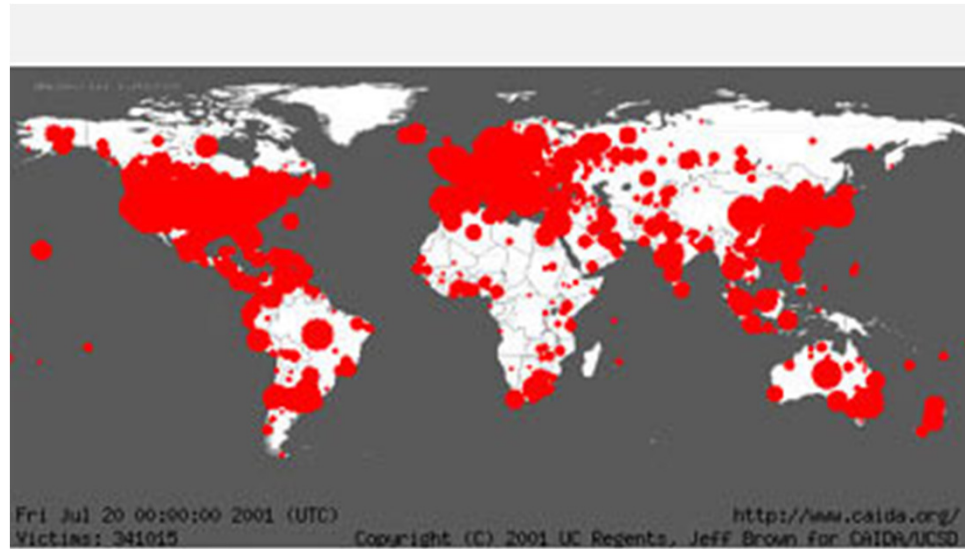
REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (3)



La red mundial de transporte



Redes sociales (electrónicas)



Internet: <http://www.caida.org/publications/visualizations/>

REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (4)

La estructura de la red, su evolución a lo largo del tiempo y su uso están mutuamente correlacionados y se deben estudiar conjuntamente:

La **topología de la red** (tipo de red, densidad, distribución de grados, distancia media, coef. de clustering, etc.) **influye en los procesos que ocurren en el sistema complejo:**

- ¿A qué estado convergen los nodos?
- ¿Cuánto se tarda en llegar a dicho estado?
- ¿Cómo se puede inmunizar (resistencia de la red ante caídas o ataques) o viralizar un sistema complejo con una topología de red concreta?

El mecanismo del proceso de difusión también influye en el proceso global:

- **contagio simple:** Cada nodo infectado conectado a ti te infecta con una cierta probabilidad
- **contagio complejo:** Para infectarte, es necesario el contacto con más de un individuo infectado

MODELO SIR EN REDES COMPLEJAS: FUNDAMENTOS (1)

Los modelos epidémicos basados en redes se comportan como los clásicos, **la diferencia es que consideran los contactos definidos por la red para la propagación**

Toda variante del modelo SIR puede generalizarse usando redes complejas pero no es fácil resolver analíticamente modelos de este tipo para redes complejas generales

La alternativa es simular el proceso en el ordenador mediante técnicas de modelado social (modelado basado en agentes, ABM)

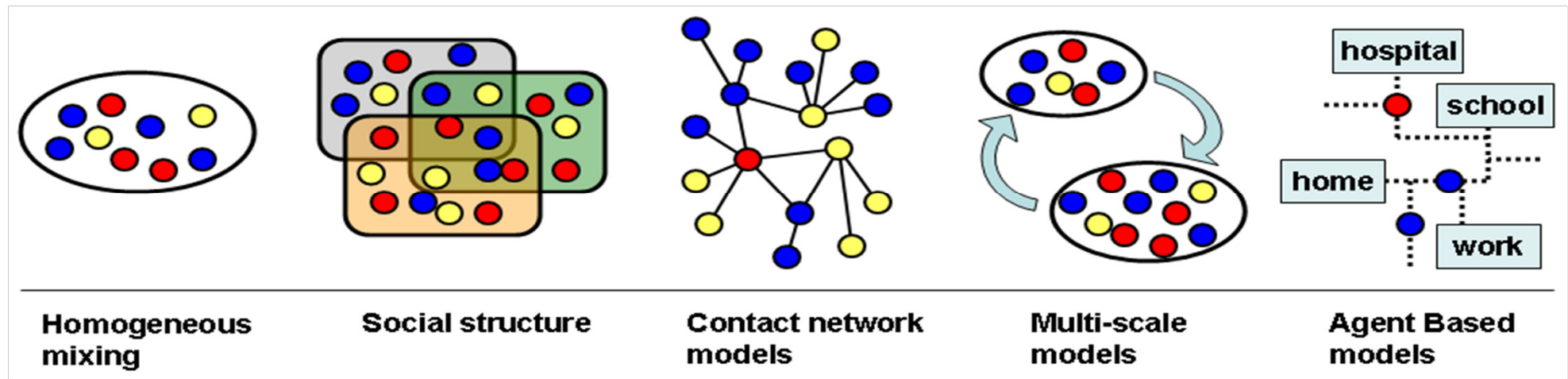
Esta metodología es mucho más potente. No se consideran globalmente los ratios de cada estado sino que se trabaja **a nivel de cada individuo (agente)**

Se conocen sus **propiedades individuales**, que pueden variar de unos a otros, y se consideran las **interacciones locales** (*comportamiento emergente*) y la aleatoriedad

P.ej. **se puede modelar el progreso del brote** en distintos escenarios con individuos de distintos tipos, distintas conexiones entre ellos, etc.

MODELO SIR EN REDES COMPLEJAS: FUNDAMENTOS (2)

En realidad, los distintos modelos tienen una complejidad creciente según se vayan considerando nuevas características más complejas de la realidad:



Simple



Realista

Habilidad para advertir/explicar tendencias a nivel de población

El realismo del modelo hace que se pierda transparencia
La validación es más compleja

<http://www.openabm.org/files/books/3443/ch13-SIRonnetwork.html>

The image shows a screenshot of an Agent-Based Modeling (ABM) simulation interface. The central window displays a network of nodes (colored green, blue, and grey) connected by edges, representing the spread of a virus. A large red arrow points to a button labeled "CLICK TO RUN".

On the left side, there are several sliders and buttons for configuring the simulation:

- number of nodes: 100
- average node degree: 6
- initial outbreak size: 3
- Buttons: setup, go
- virus spread chance: 10.0 %
- virus check frequency: 1
- recovery chance: 1
- gain resistance chance: 100 %

On the right side, there are more sliders and a graph:

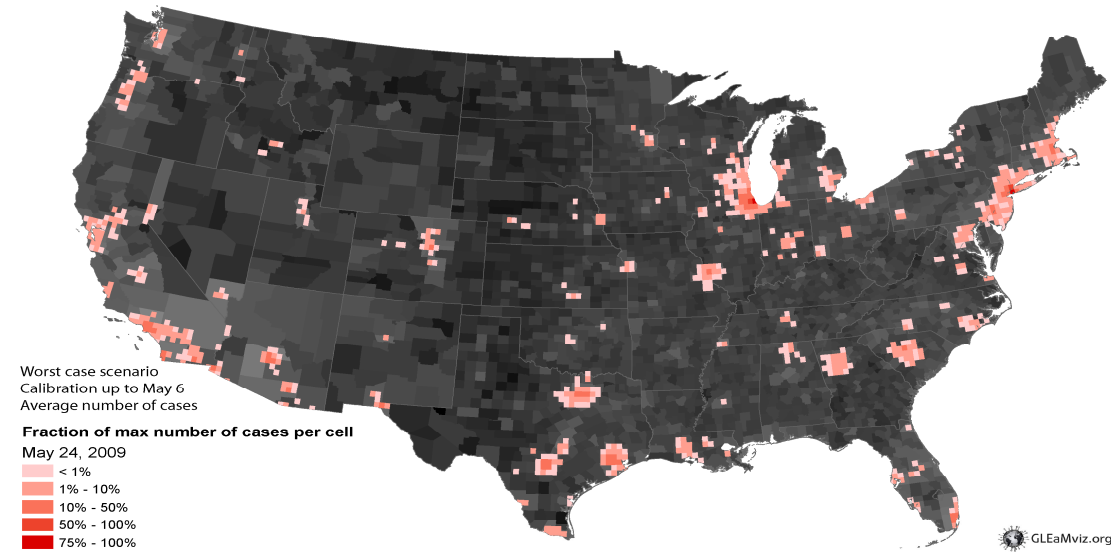
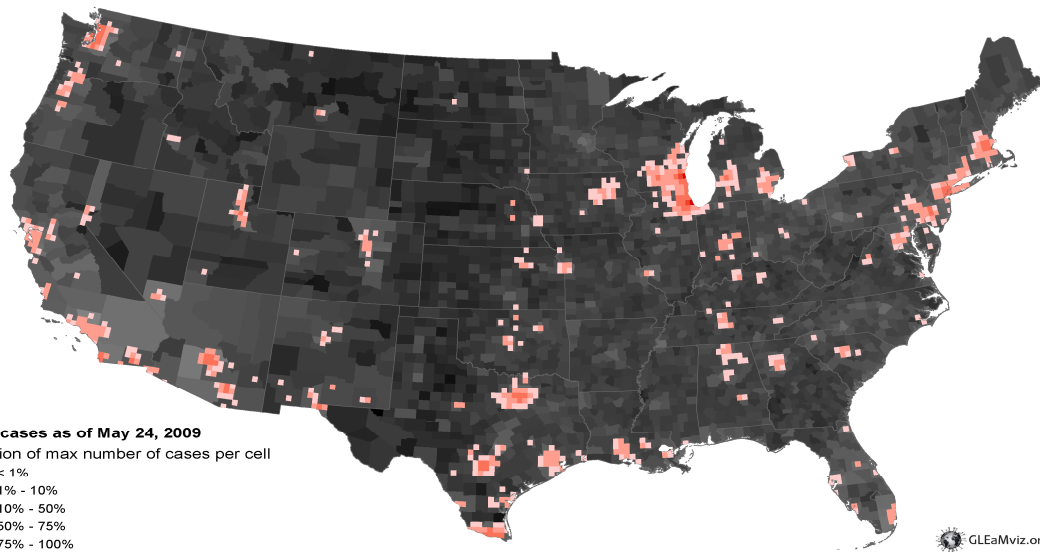
- numbers of isolated agents: 100
- wholistic degree: dropdown menu
- dist: 10
- Graph: A line graph showing the number of nodes over time, with a value of 46.3 at the end of the x-axis.

At the bottom left, there is a "Network Status" graph showing the percentage of nodes in different states over time:

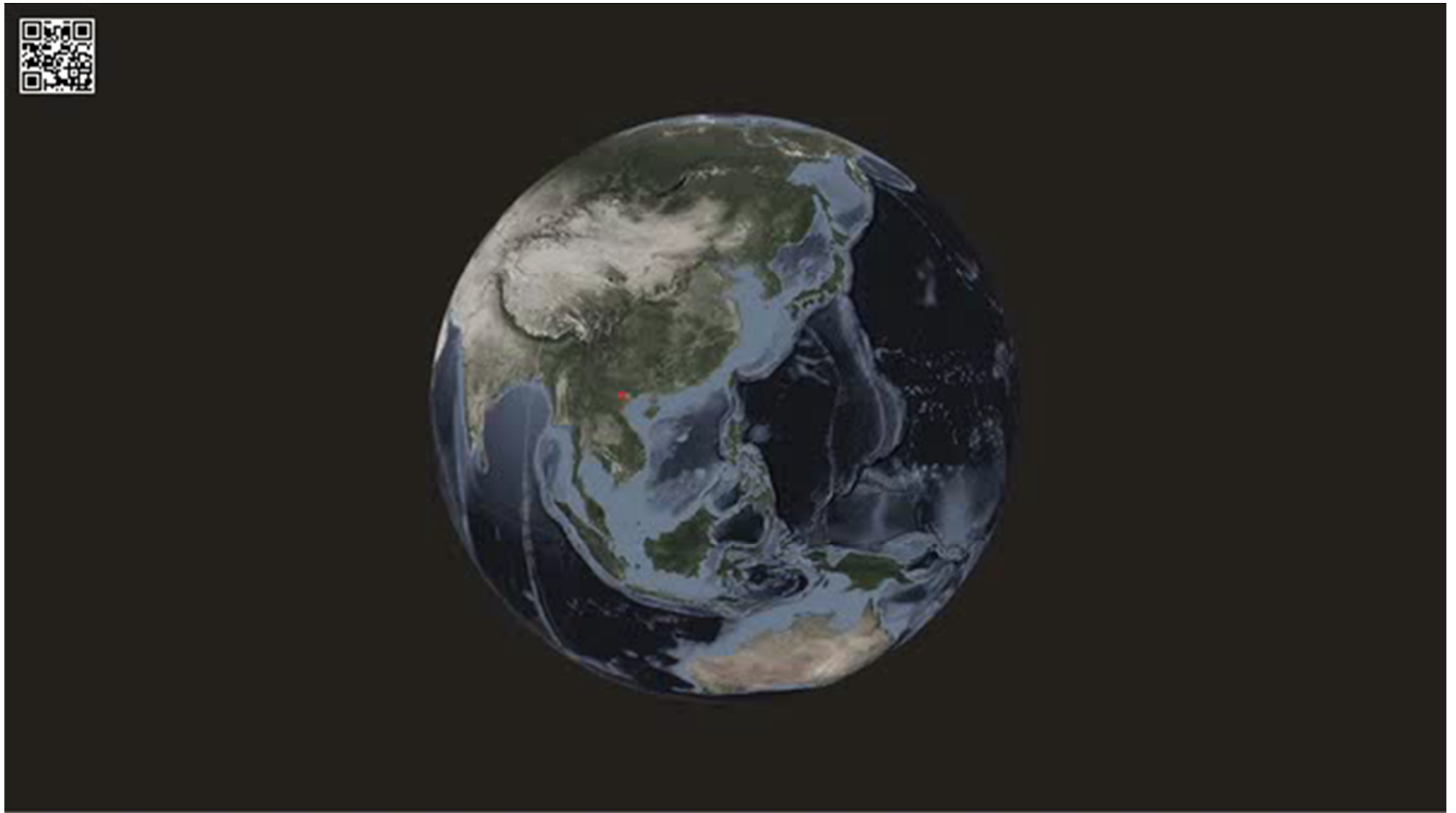
- Legend: susceptible (green), infected (red), resistant (blue)
- Y-axis: % of nodes (0 to 100)
- X-axis: time (0 to 12)

Real

Pronosticada



Primera pandemia analizada con redes complejas y ABM



<http://www.gleamviz.org/simulator/>

MODELOS DE DIFUSIÓN DE OPINIONES EN REDES

Easley y Kleinberg. Cascading Behavior in Networks. Networks, Crowds, and Markets: Reasoning about a Highly Connected World (Cap. 19). Cambridge University Press, 2010

Zafarani, Ali Abbasi y Liu. Information Diffusion in Social Media. Social Media Mining. An Introduction (Cap. 7). Cambridge University Press, 2014

Difusión de información: Proceso mediante el cual una unidad de información (conocimiento) se difunde **en una red** y alcanza a los individuos mediante interacciones

Se estudia en una gran cantidad de disciplinas como la Sociología, la Epidemiología, la Etnografía, la Minería de Medios Sociales o el Marketing

Existen distintas técnicas para modelar los procesos de difusión de información en redes

En Febrero de 2013, durante el tercer cuarto de la *Super Bowl*, un apagón interrumpió el partido durante 34 minutos

Durante ese parón, la compañía **Oreo** twiteó y publicó en su cuenta de *Facebook* el mensaje: “*Power out? No Problem, You can still dunk it in the dark*”

El tweet se propagó desmesuradamente, alcanzando **casi 15,000 retweets y 20,000 likes en menos de dos días**

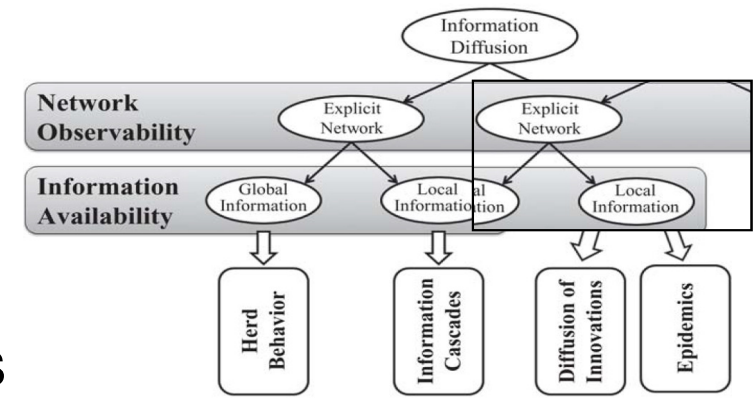
Este sencillo tweet se difundió rápidamente en una gran población de individuos

Ayudó a la compañía a conseguir reconocimiento con una inversión mínima (*earned media*) en un entorno en el que las empresas gastan 4 millones de dólares en anuncios de 30 segundos durante la *Super Bowl*

- Los modelos de difusión de información en redes existentes permiten modelar distintas situaciones:
 - **Cascadas de Información y Modelos Epidémicos en redes**: la difusión se produce sólo vía los amigos (**contagio/decisión con información local**)
 - **Difusión de innovaciones en redes**: Tres variantes con **información global y local**: sólo innovación (global), sólo imitación (local) y mixto (global y local)

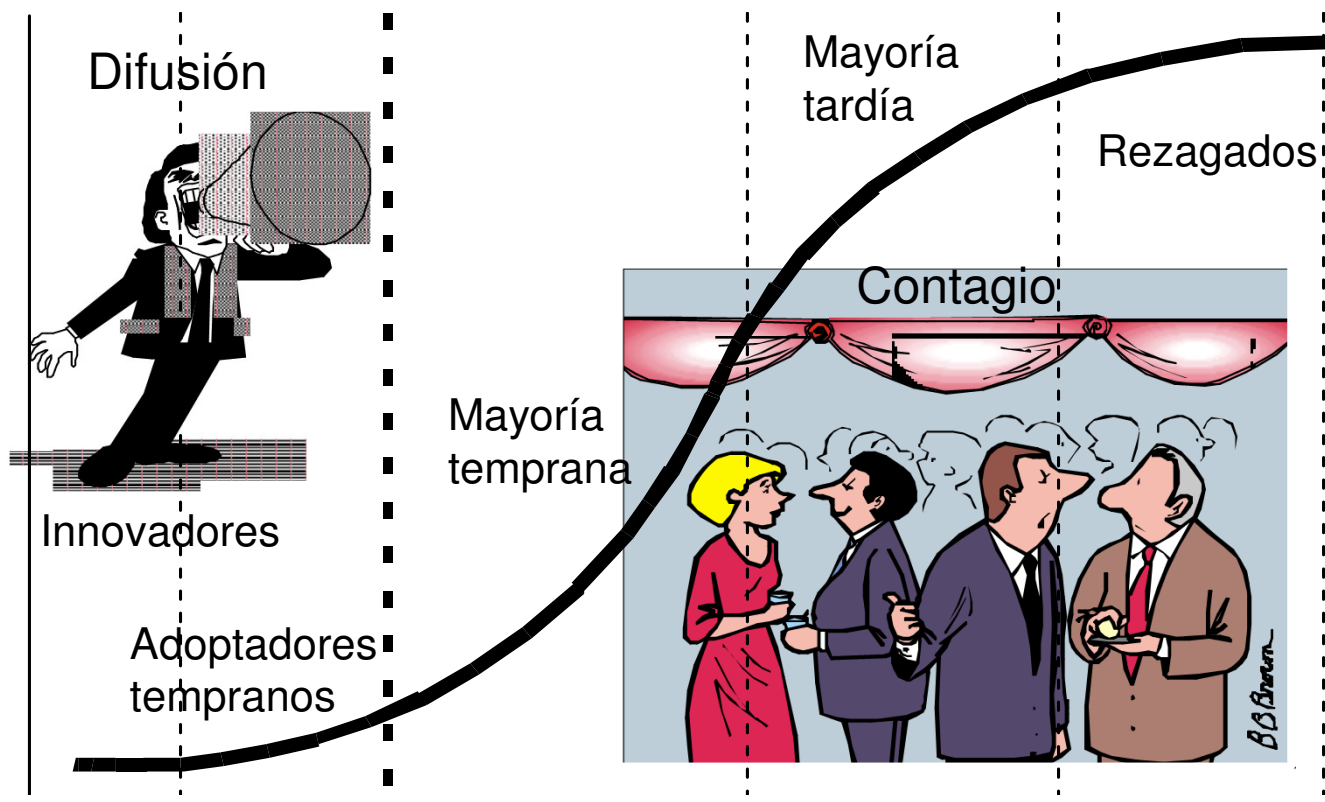
- Pueden ser *centrados en el emisor*, en los que los individuos no toman la decisión por si mismos, o en el *receptor*, donde si lo hacen

- Los **contagios** pueden ser **simples** (individuales por probabilidad) y **complejos** (múltiples)



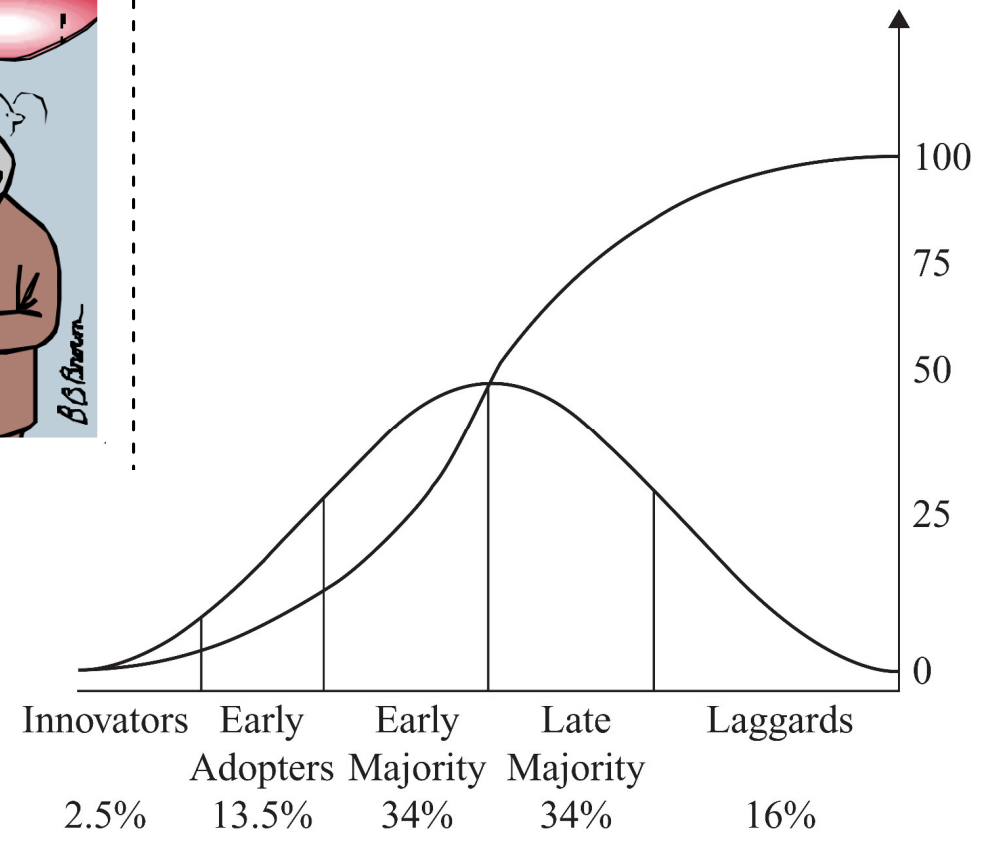
DIFUSIÓN DE LA INNOVACIÓN

Curvas de Adopción



Muy estudiada en **Investigación de Mercados**

Una **innovación** es una idea, práctica u objeto percibido como nuevo por un individuo y que puede ser adoptado o no



- La **Teoría de la Difusión de Innovaciones** estudia cómo y por qué se difunden y adoptan esas innovaciones
- Modelo clásico de Bass de 1969 (sin redes):

$$\frac{dA(t)}{dt} = i(t)[P - A(t)]. \quad i(t) = \alpha + \beta A(t)$$

α : factor de innovación (influencia externa, global)

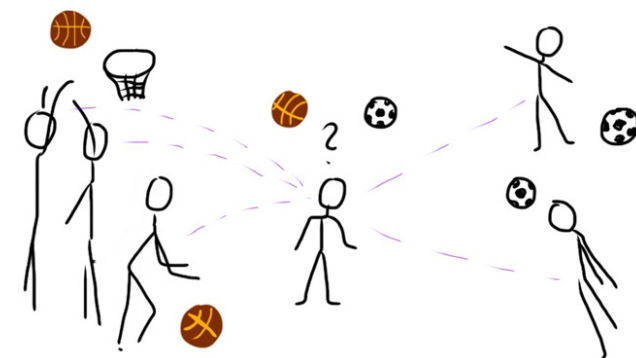
β : factor de imitación (influencia de los adoptadores, boca-a-oreja, local)

P: tamaño de la población

A(t): Ratio de adoptadores

- El modelo clásico es analítico: se asume mezclado homogéneo y sólo se consideran los ratios (obtención de las curvas de adopción)
- En los modelos basados en redes (ABMs) se considera el ratio de adoptadores entre los contactos del nodo (red social personal). Están centrados en el emisor

- En los **modelos de contagio complejo**, el contagio no se produce por un único contacto infectado sino requiere dos o más (Ej. credibilidad de una leyenda urbana)
- En los **modelos de umbral** la adopción requiere superar un umbral asociado a un porcentaje de contactos infectados (20%, 30%, ...). Los umbrales pueden ser comunes a toda la población o particulares a cada agente
- Pueden incluir información sobre las ventajas de la adopción (**recompensas**). Ej. dos opciones, fútbol o baloncesto:
 - d contactos. p : ratio que juega baloncesto (B), $1-p$: ratio que juega fútbol (F)
 - Cada amigo de B recompensa con r_b y de F con r_f
 - Si escoge B, recibe $p \cdot d \cdot r_b$. Si escoge F, recibe $(1-p) \cdot d \cdot r_f$
 - El agente escogerá B si $p \cdot d \cdot r_b \geq (1-p) \cdot d \cdot r_f$
 - **Umbral adopción B:** $p \geq r_f / (r_b + r_f)$



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 y hay interacciones que influyen en la compra

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

Modelo CONSUMAT 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



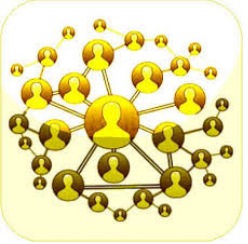
UTILIDAD DE PRODUCTO (1)

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)

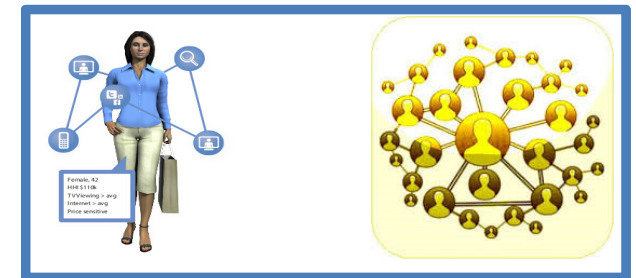
La parte de **efecto social** de la utilidad del producto aumenta con la cantidad de amigos que consumen el producto (**variante de contagio complejo**). Está relacionado con la necesidad del agente de parecerse al resto de agentes de su red social

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 β_i , el agente es más individualista (innovador)**

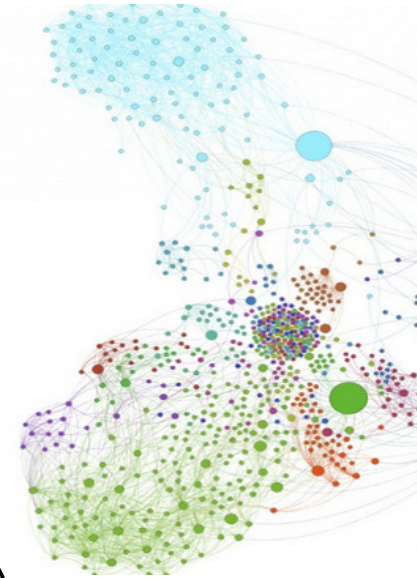


RED SOCIAL

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

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

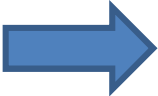
Cada agente que compra calcula su utilidad e incertidumbre globales y escoge una heurística de decisión entre cuatro posibles. **Las heurísticas determinan qué producto comprar de forma determinística (el que maximiza la utilidad) o probabilística**

Los umbrales y el parámetro de comportamiento social β_i determinan la toma de decisiones de cada agente consumidor (conservadora, arriesgada, imitadora, ...)

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	$E[Unc_{ij}] = (1 - \beta_i) \cdot (1 - x_j)$

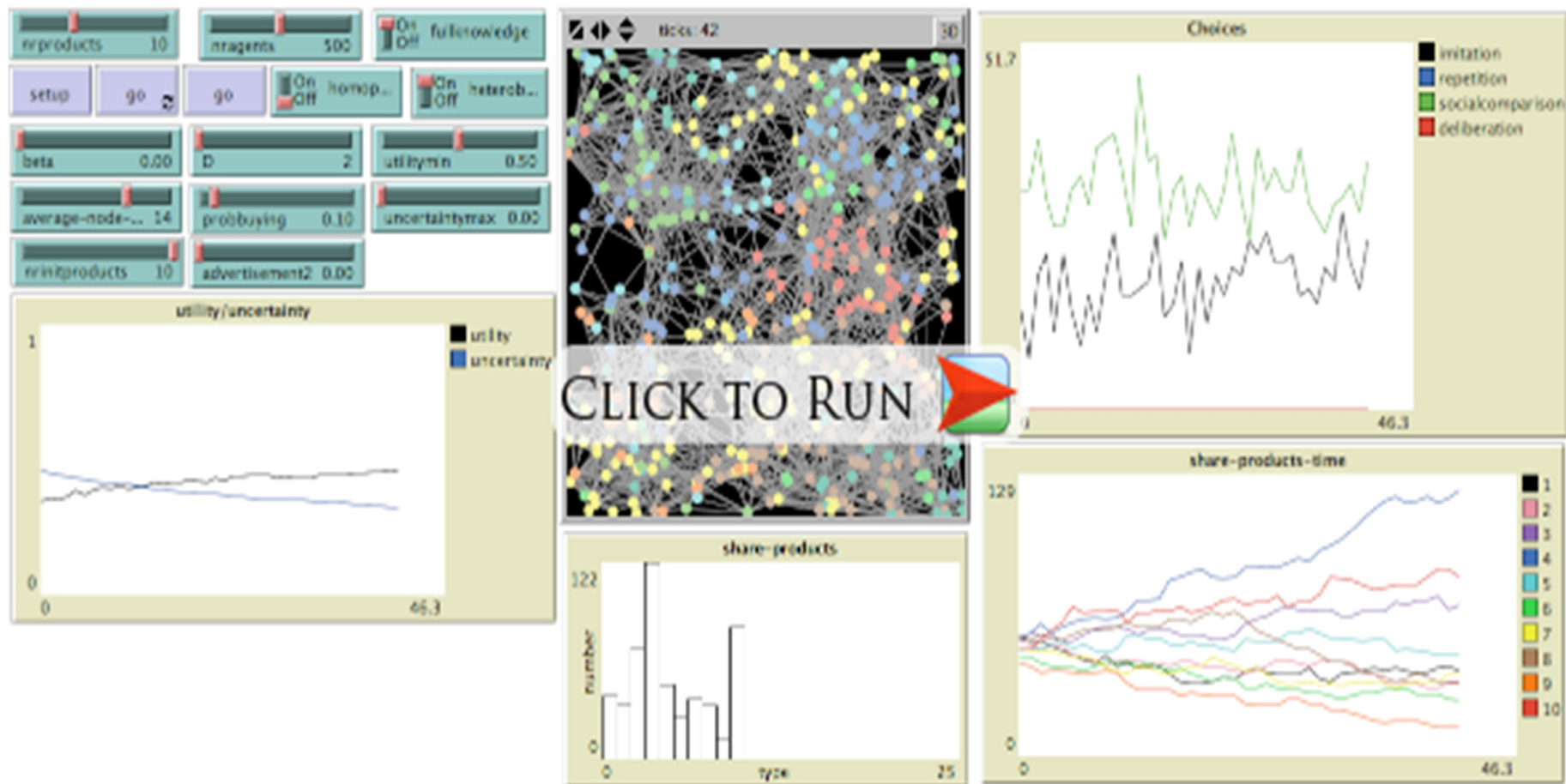
β_i : inclinación del consumidor i hacia adoptar sus propias decisiones o a dejarse sesgar por las de su entorno



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

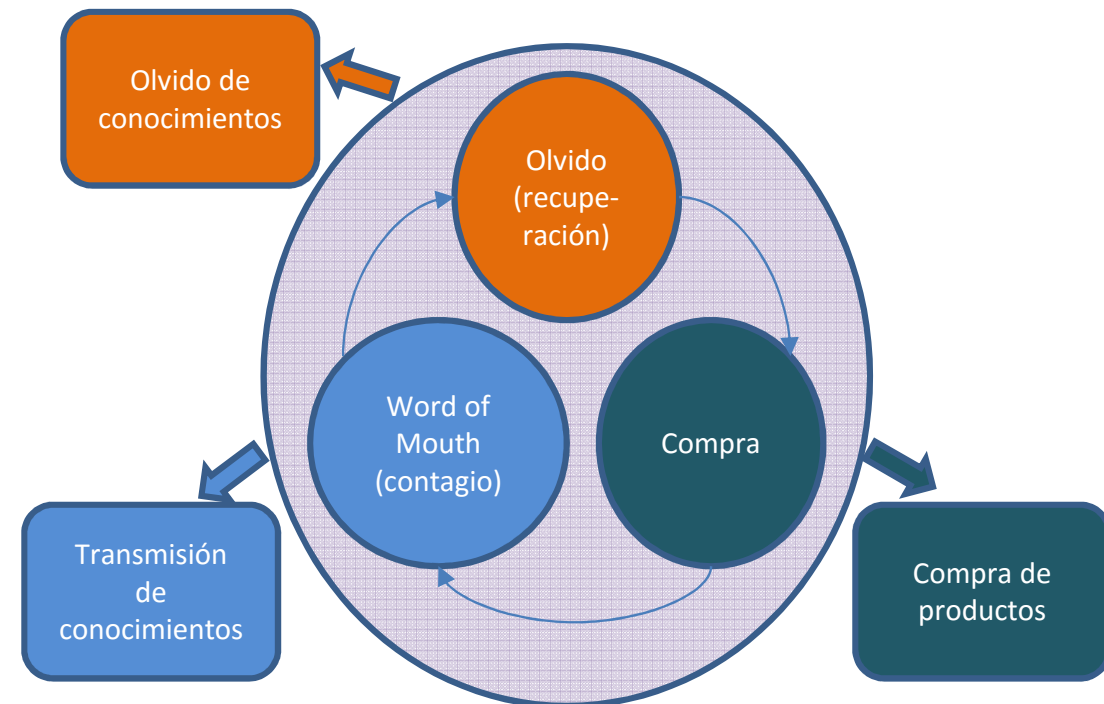
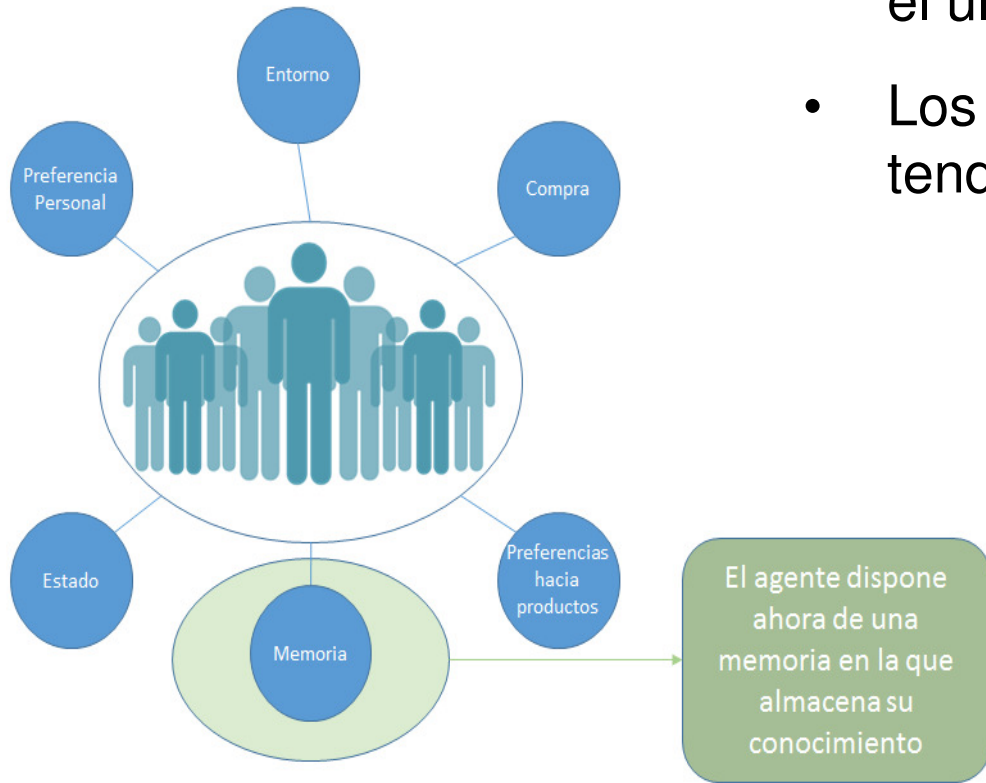
El agente ya no dispone de un conocimiento completo de los productos y **sólo puede adquirir aquellos que conoce (*racionalidad limitada*)**. El reconocimiento de cada producto se adquiere mediante el **WoM** con los contactos del agente

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)

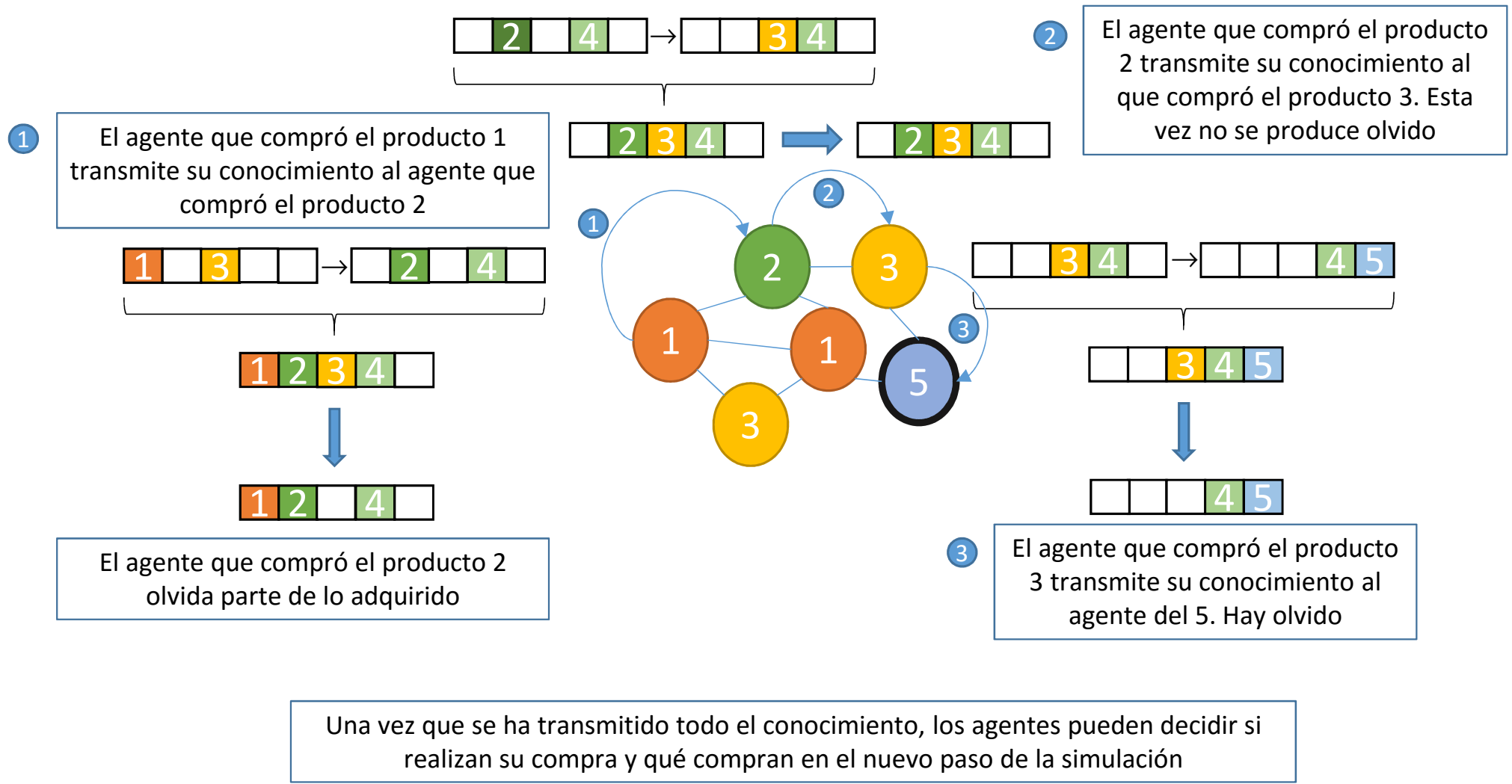
INCORPORACIÓN DEL RECONOCIMIENTO DE PRODUCTO (2)

- 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



INCORPORACIÓN DEL RECONOCIMIENTO DE PRODUCTO (3)

Ejemplo de funcionamiento



ALGUNOS EXPERIMENTOS (1)

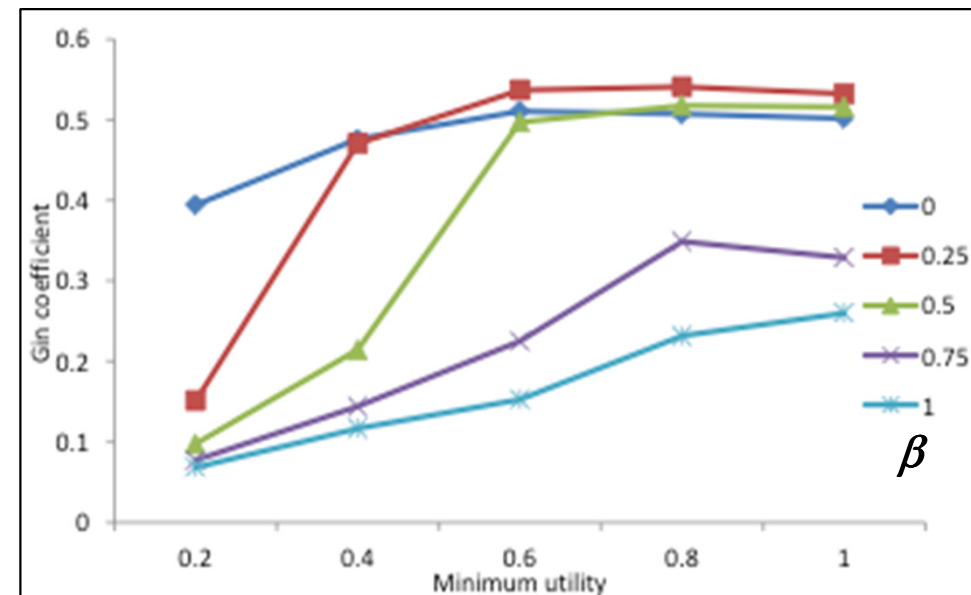
1. Modelo sin reconocimiento de productos:

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 β y umbral mínimo U_{\min} , comunes para toda la población; 100 simulaciones (**Montecarlo**) de 100 pasos

El índice Gini es bajo (poca dispersión) si β 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 U_{\min} crece por encima de 0.5 (agentes más sociales y más exigentes)



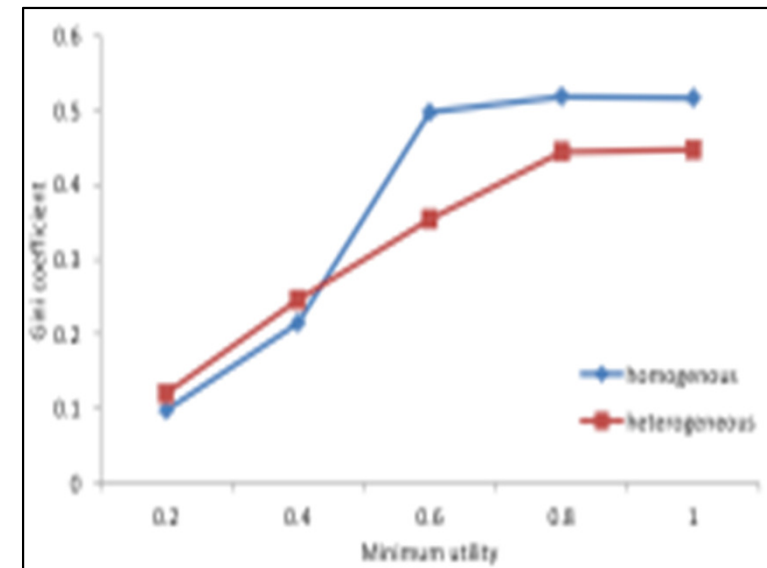
ALGUNOS EXPERIMENTOS (2)

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

P.ej., considerando un valor distinto de β_1 , 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 $\beta_1 \sim U[0,1]$ (**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.

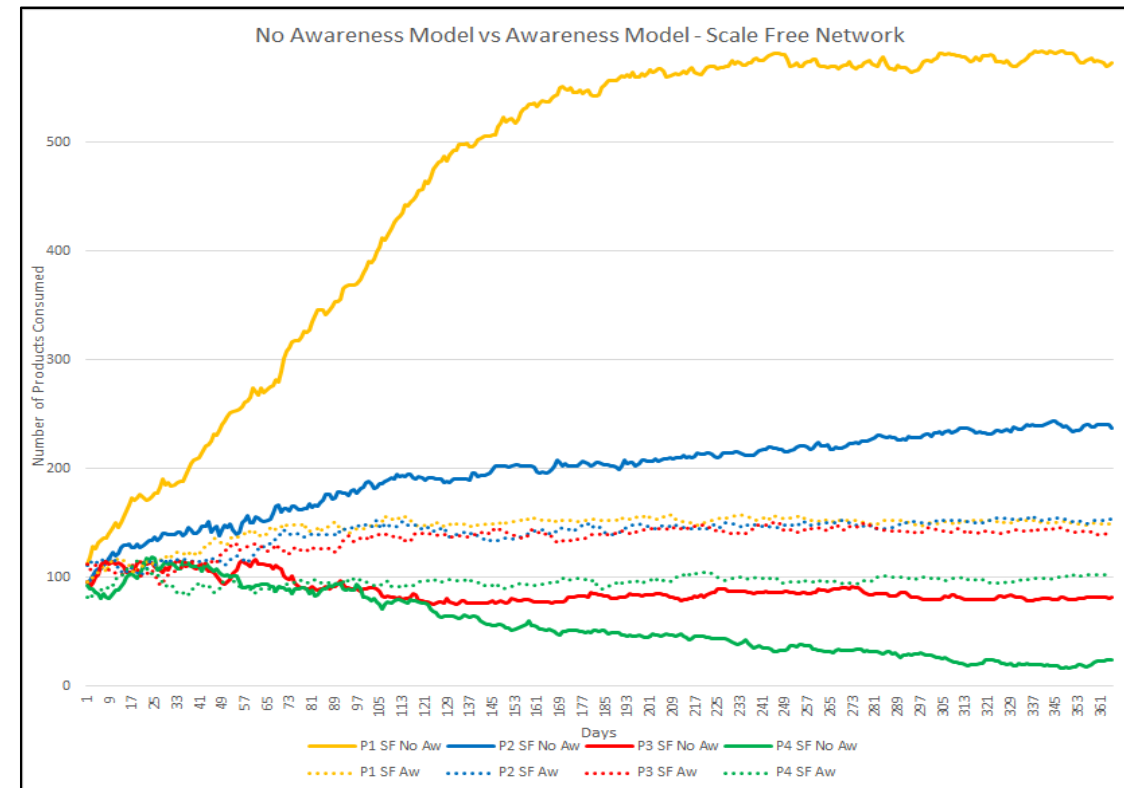
ALGUNOS EXPERIMENTOS (3)

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 $p_c=0.1$; 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**:

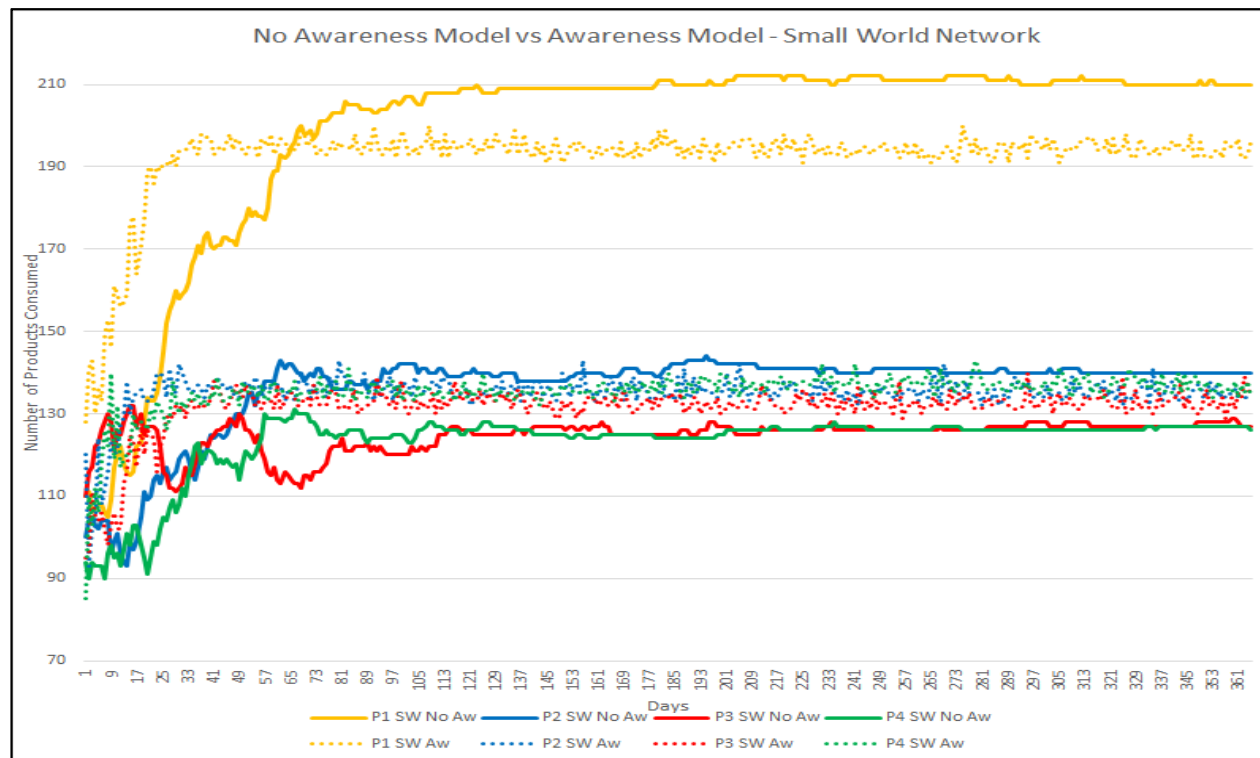
- Sin reconocimiento (línea continua): un producto satura el mercado
- Con reconocimiento (línea discontinua): mucha mayor competencia



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 **selección de las semillas para provocar cascadas en campañas 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 / Social Media Mining**:
 - **análisis de sentimientos** para incorporar información realista al modelo ABM,
 - **escalado de las redes** al tamaño de las redes sociales reales, etc.
- ...

APRENDIZAJE EVOLUTIVO DE ESTRATEGIAS DE MARKETING VIRAL (1)

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 (semillas) 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 (2)

Este 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 en la red social** como grado o coeficiente de clustering (CC)

Stonedahl et al. *Evolving viral marketing strategies*. GECCO 2010

APRENDIZAJE EVOLUTIVO DE ESTRATEGIAS DE MARKETING VIRAL (3)

J.F. Robles, M. Chica, O. Cerdón: Incorporating Awareness and Genetic-based Viral Marketing Strategies to a Consumer Behavior Model. IEEE CEC 2016

Hemos incorporado el método de Stonedahl a nuestro modelo CONSUMAT extendido con reconocimiento de productos

El algoritmo persigue obtener **la mejor selección posible de individuos influyentes para conseguir promover la compra de un producto**. Para la selección, se consideran combinaciones de los criterios locales

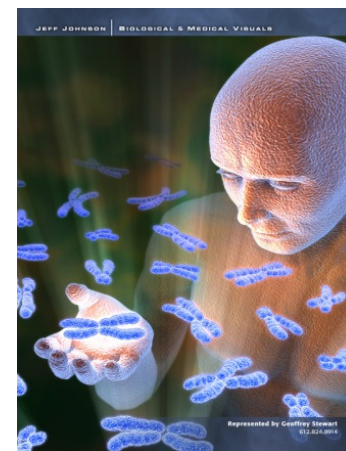
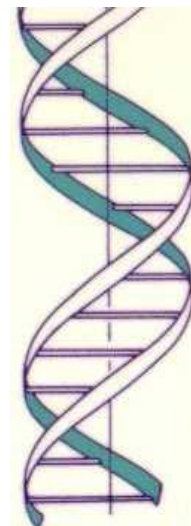
No sólo interesa que muchos consumidores compren el producto anunciado sino que lo hagan pronto para lograr más beneficios. Se usa un índice llamado *Net Present Value* (NPV) que mide ambos conceptos

Las estrategias diseñadas ordenan los individuos según su preferencia para ser escogidos. Así *se optimiza también* su número y *el presupuesto de la campaña*

El método se basa en el uso de *algoritmos genéticos*

ALGORITMOS GENÉTICOS

Métodos de optimización inspirados en los principios Darwinianos de Evolución Natural. Se basan en la evolución de poblaciones cuyos elementos representan soluciones a problemas



ERROR: stackunderflow
OFFENDING COMMAND: ~
STACK: