



Escuela de Verano de Inteligencia Artificial (EVIA 2014)

Redes Sociales y 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

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.

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

SISTEMAS COMPLEJOS (2)

- 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

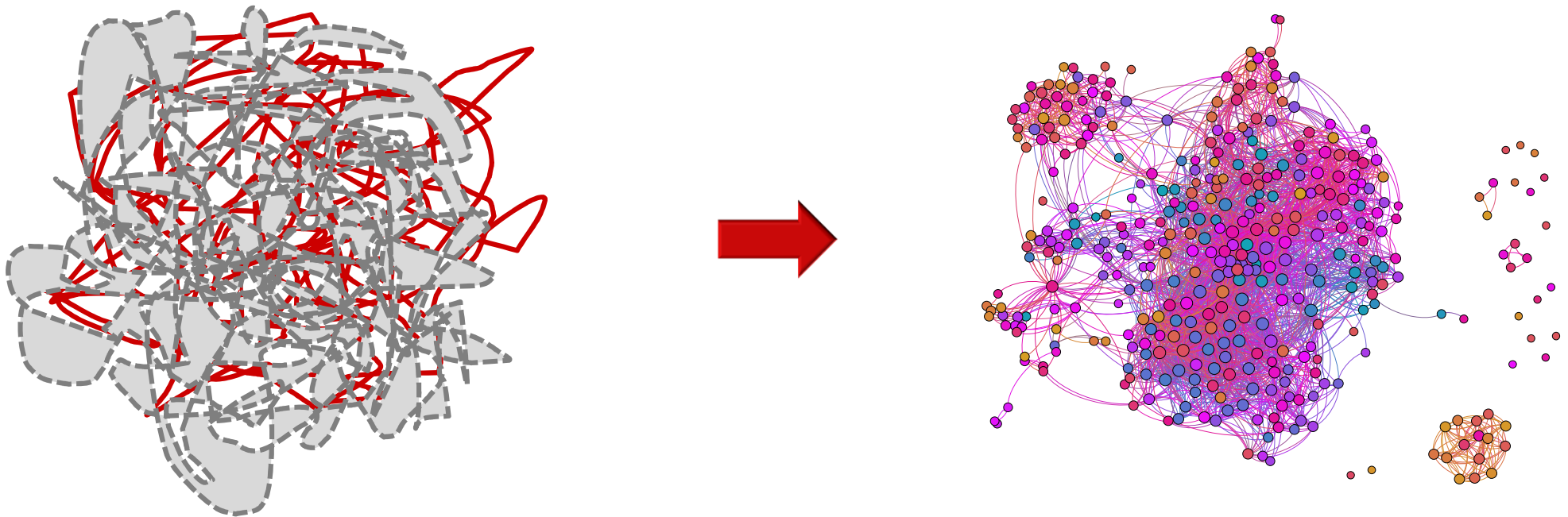




*“I think the next century
will be the century
of complexity.”*

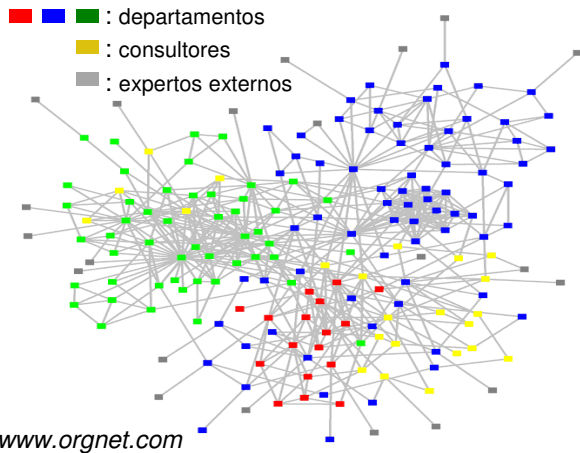
Stephen Hawking
January 23, 2000

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

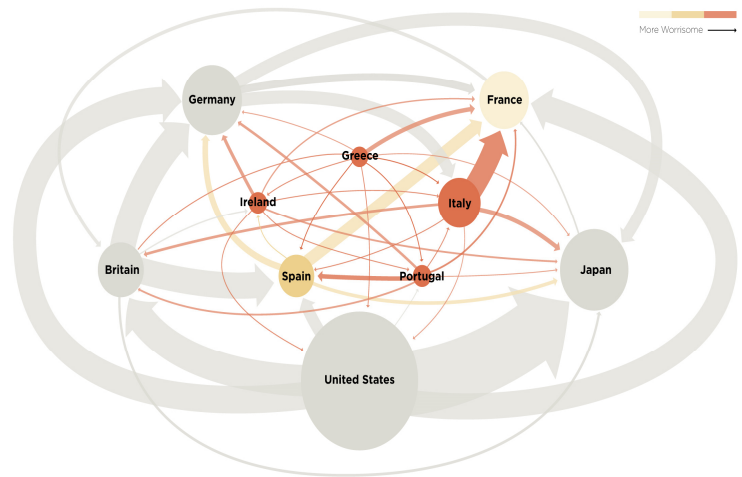


EJEMPLOS DE REDES COMPLEJAS REALES

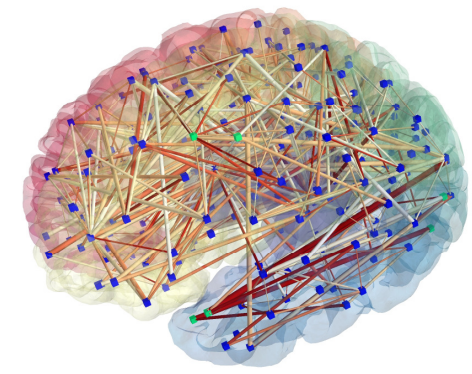
Organizaciones



Redes financieras

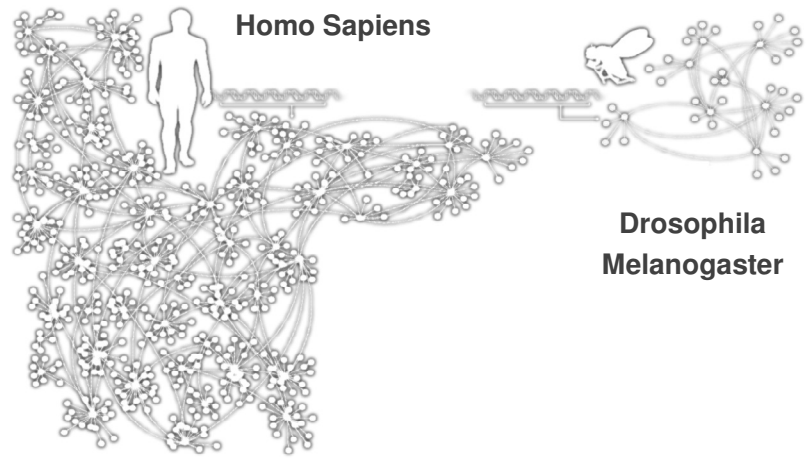


Redes cerebrales

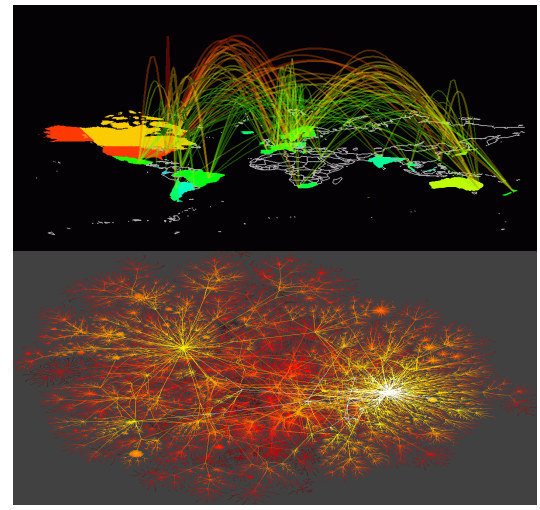


<http://www.humanconnectomeproject.org/>

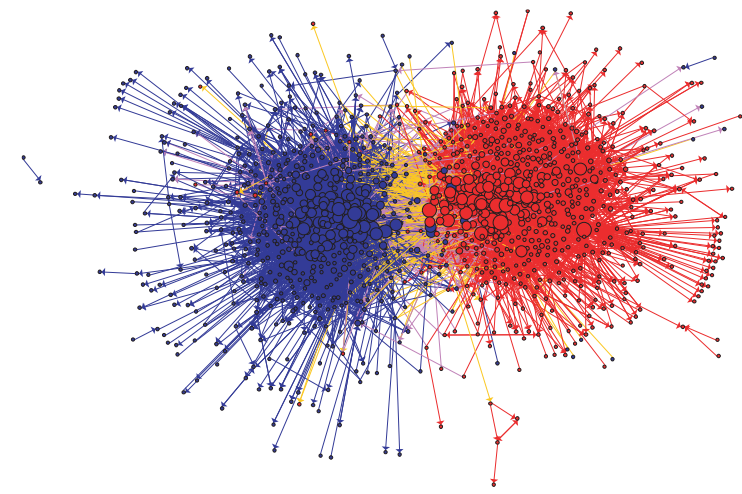
<http://www-personal.umich.edu/~mejn/netdata/>



Redes genéticas



Red física de Internet



Redes políticas

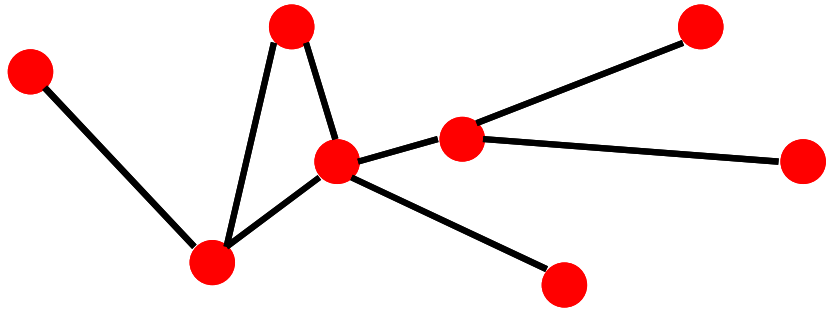
EL ROL DE LAS REDES

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**

COMPONENTES Y NOMENCLATURA DE UN SISTEMA COMPLEJO



▪ **componentes:** nodos, vértices **N**

▪ **interacciones:** enlaces, arcos **L**

▪ **sistema:** red, grafo **(N,L)**

red se suele referir a sistemas reales

- www: red de páginas conectadas por URLs,
- red social: grupo de individuos conectados por una relación familiar, de amistad o profesional,
- red metabólica: conjunto de las reacciones químicas que tienen lugar en una célula

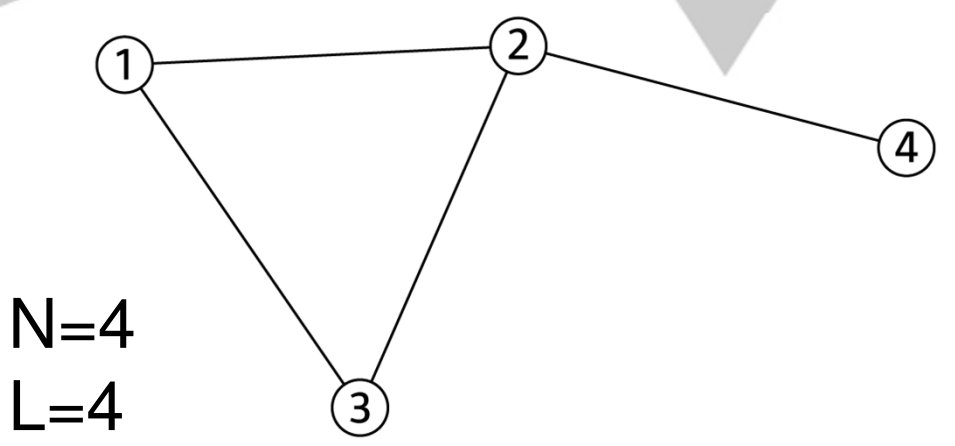
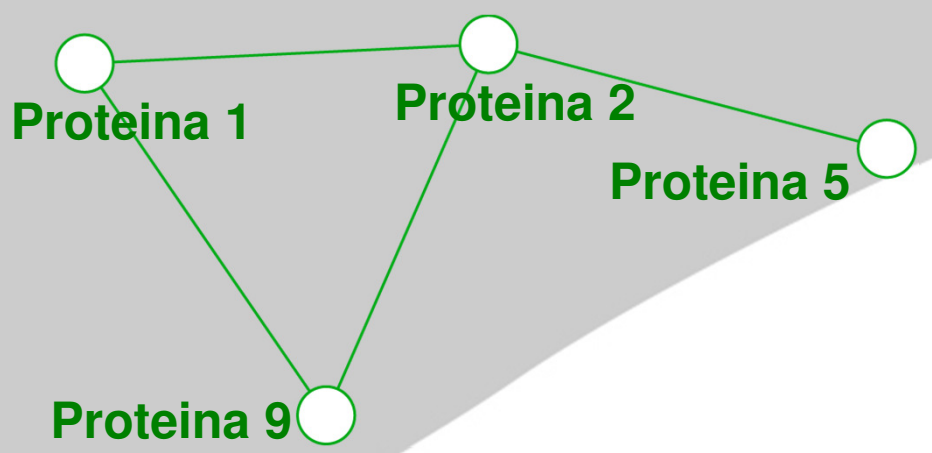
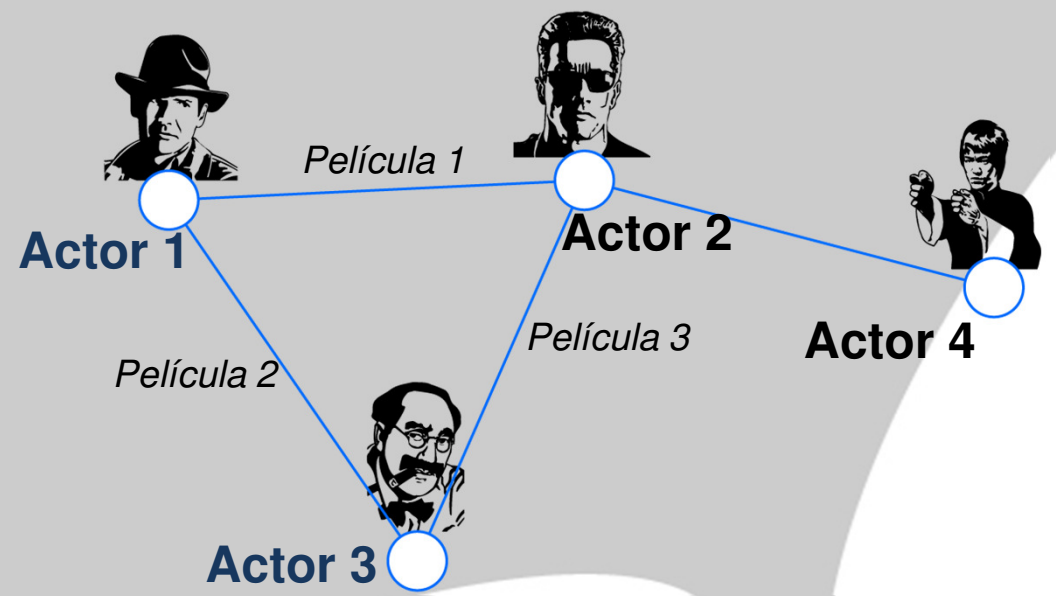
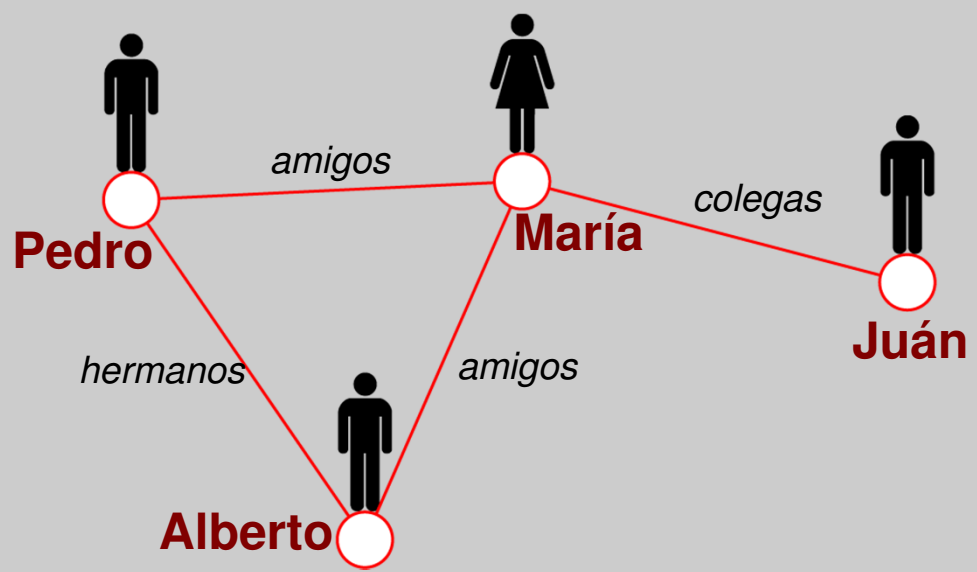
Nomenclatura: (Red, nodo, enlace)

grafo: representación matemática de una red

- grafo de la web,
- grafo social (término de Facebook)

Nomenclatura: (Grafo, vértice, arco)

LAS REDES OFRECEN UN LENGUAJE COMÚN PARA ESTUDIAR SISTEMAS



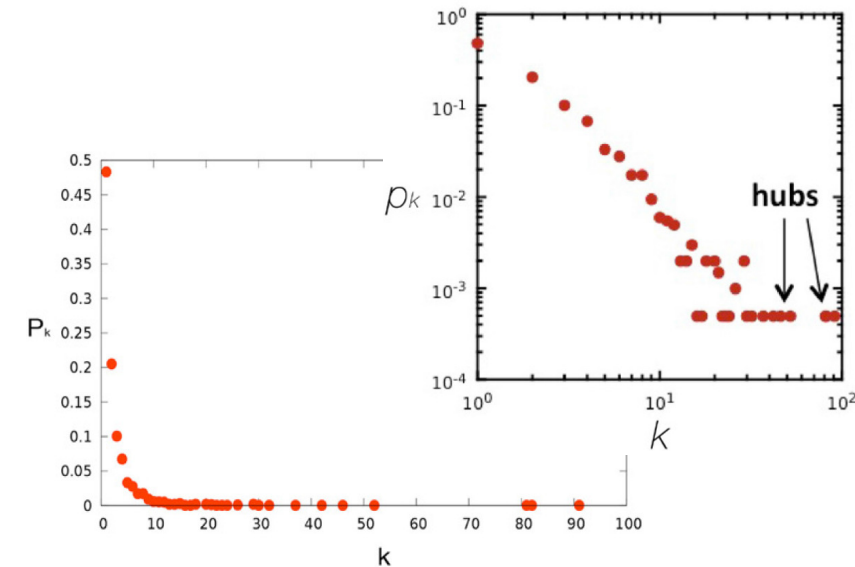
LAS TRES MEDIDAS PRINCIPALES EN TEORÍA DE REDES

El objetivo principal de la Teoría de Redes es el análisis de redes reales. Al ser de gran tamaño (desde cientos a millones de nodos), se usan **medidas** como:

Distribución (de probabilidad) de grados p_k :
Permite caracterizar distintos modelos de redes y determina muchos fenómenos de la red como su robustez y los procesos de difusión

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

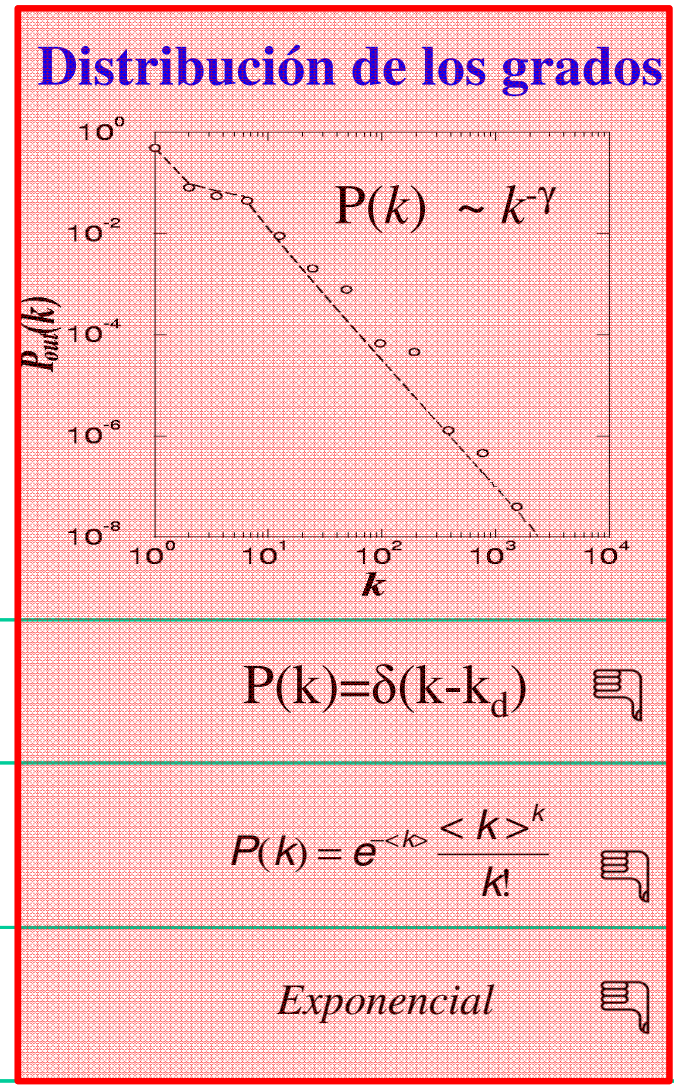
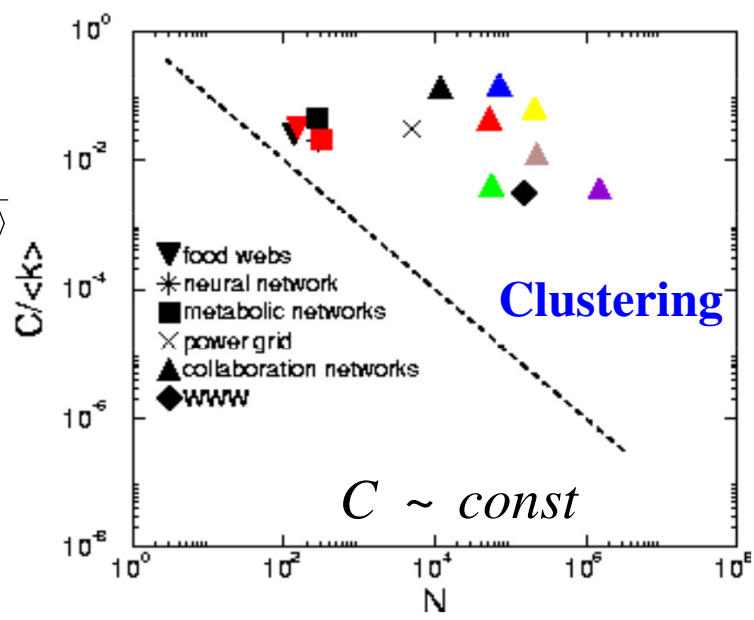
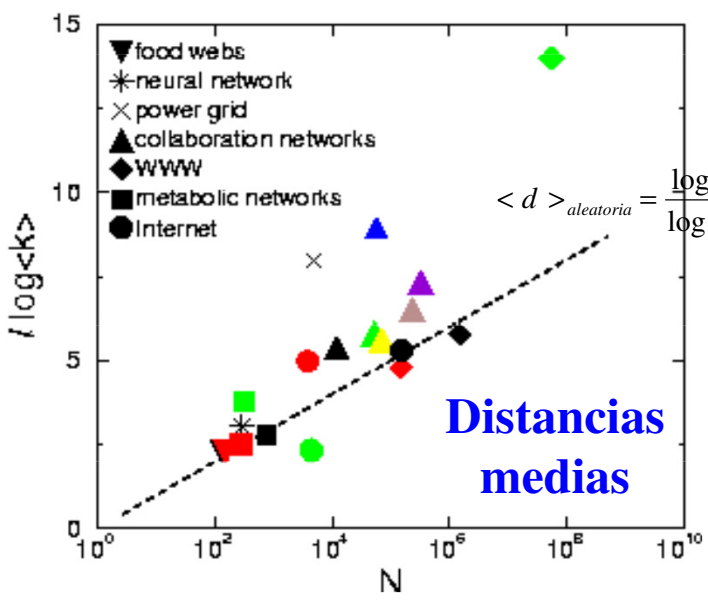
Coefficiente de Clustering C_i : Mide la densidad local de la red, ¿qué proporción de los vecinos de cada nodo están conectados?



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

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

REDES REALES Y MODELOS DE REDES COMPLEJAS



| | | | | | | |
|------------------------------|---|---|---|---|--|---|
| Redes regulares | $\langle d \rangle \approx N^{1/D}$ | 👎 | $C \sim const$ | 👍 | $P(k) = \delta(k - k_d)$ | 👎 |
| Redes ER (aleatorias) | $\langle d_{aleatoria} \rangle \approx \frac{\log N}{\log \langle k \rangle}$ | 👍 | $C_{aleatoria} = p = \frac{\langle k \rangle}{N}$ | 👎 | $P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$ | 👎 |
| Mundos pequeños | $\langle d_{aleatoria} \rangle \approx \frac{\log N}{\log \langle k \rangle}$ | 👍 | $C \sim const$ | 👍 | <i>Exponencial</i> | 👎 |
| Libres de escala | $\langle d_{aleatoria} \rangle \approx \frac{\log N}{\log \log N}$ | 👍 | $C \sim const$ | 👍 | $P(k) \sim k^{-\gamma}$ | 👍 |

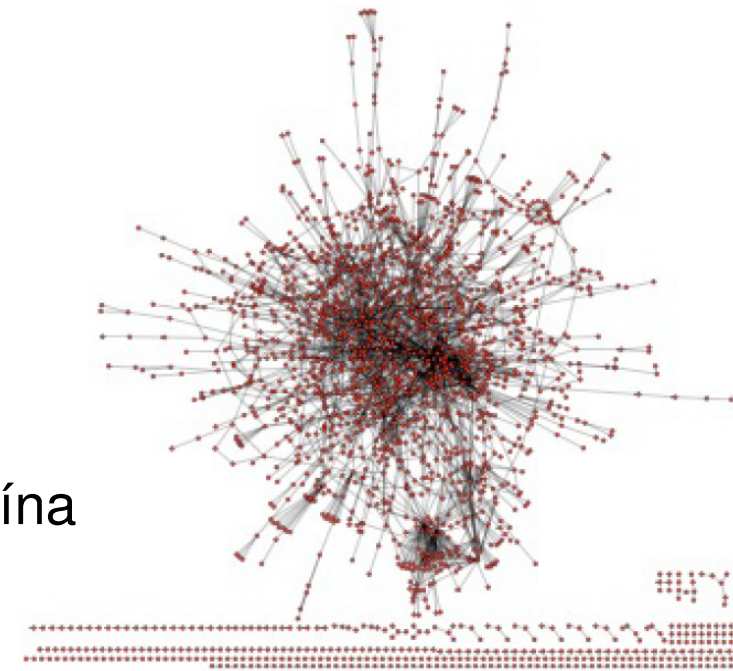
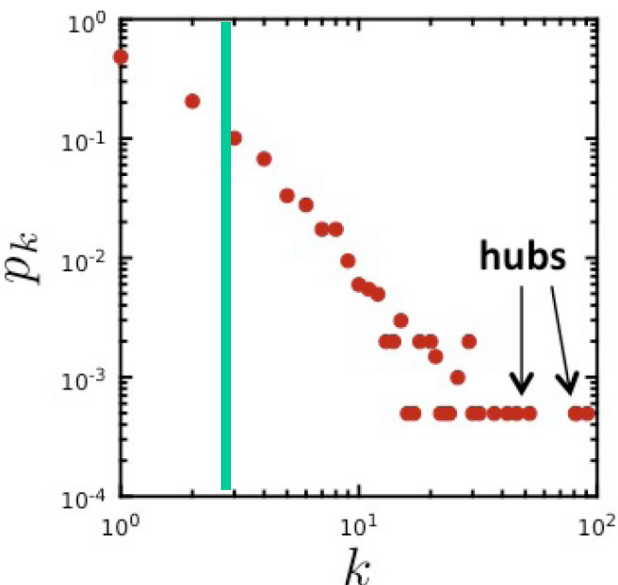
EJEMPLO: ANÁLISIS DE LA RED DE PROTEÍNAS DE LA LEVADURA

Red de interacción entre proteínas (PPI) de la levadura.
Los nodos son proteínas y están conectados si hay evidencia empírica de que interactúan

$N=2018$ nodos, $L=2930$ enlaces, $Densidad=0.00144$

Grado medio: $\langle k \rangle \equiv \frac{2L}{N} = 2.90$ interacciones por proteína

Red libre de escala



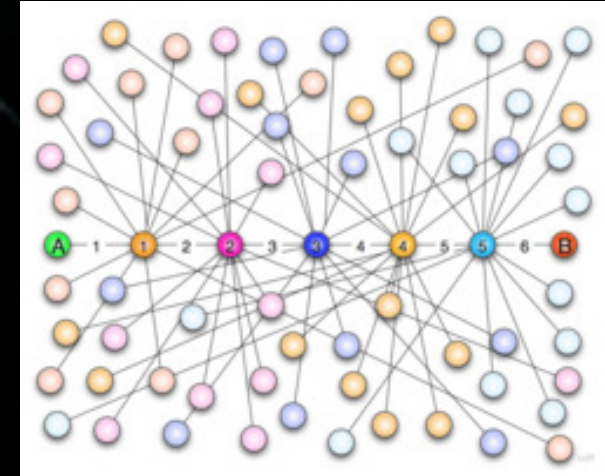
Distancia media: $\langle d \rangle = 5.61$ (**mundos pequeños**)

Coeficiente de **clustering medio:** $\langle C \rangle = 0.12$ (**bastante alto**)

Conectividad: 185 componentes conexas. La componente gigante agrupa el 81% de los nodos: 1647 de los 2018

DIFUSIÓN GENERAL: DOCUMENTAL

HOW KEVIN BACON CURED CANCER




Kevin Bacon y la
Teoría de los seis
grados de separación

<http://barabasilab.com/>
<http://youtu.be/zK1Cb9qj3qQ>

ANÁLISIS DE REDES Y MEDIOS SOCIALES

DEFINICION Y PARTICULARIDADES DE LAS REDES SOCIALES

Una **red social** es un **tipo concreto de red** que modela las relaciones o interacciones existentes entre un conjunto de **entidades sociales**, tales como personas, grupos u organizaciones

| | | |
|-----------------|------------------|---|
| Mathematics | Computer Science |  |
| Vertex/vertices | Node | |
| Edge | Link/connection | |

Según Newman, las redes reales se pueden categorizar en cuatro tipos principales:

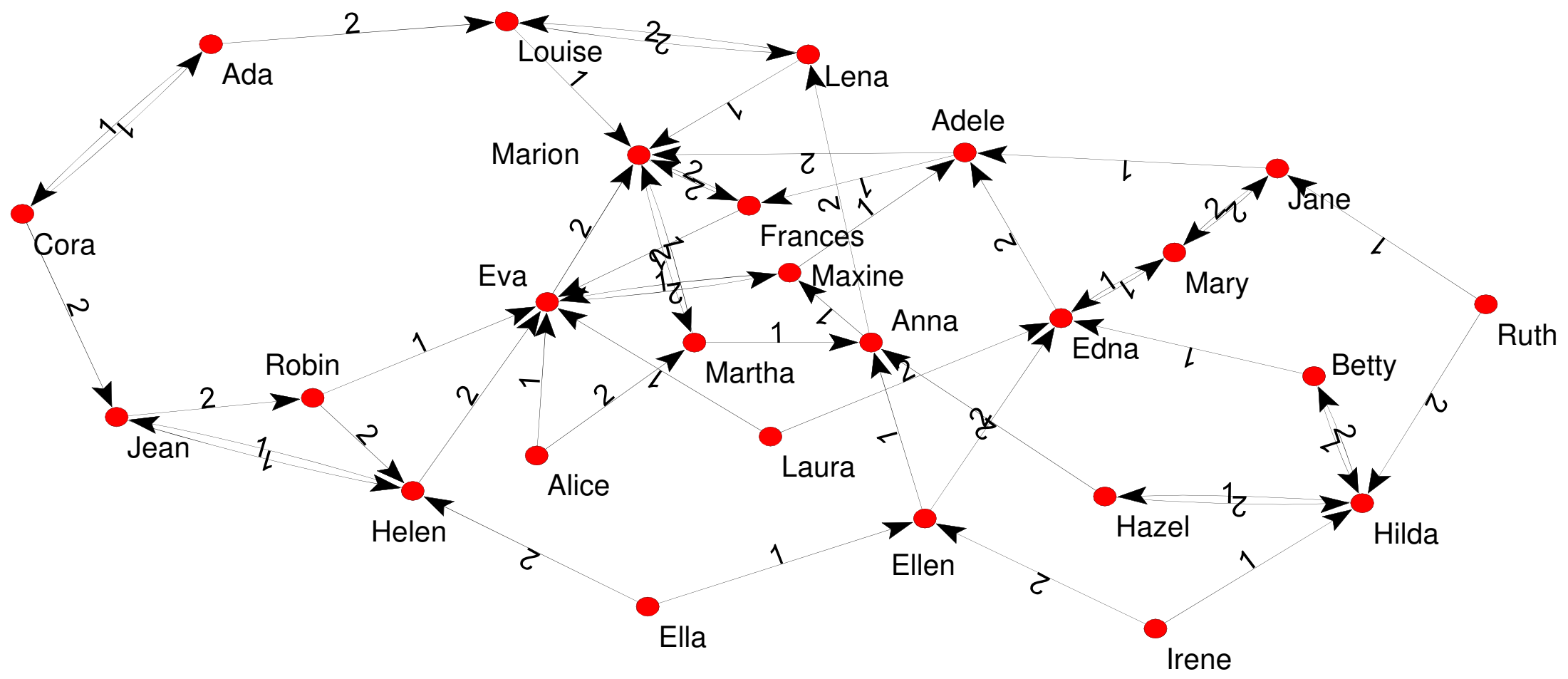
- redes sociales,
- redes de información o conocimiento,
- redes tecnológicas y
- redes biológicas

M.E.J Newman. The structure and function of complex networks. SIAM Reviews 2003, 45:167–228

Las redes sociales son las que resultan de interacciones humanas y sociales, como las redes de amistad, las redes informales de comunicación entre empresas y las redes de colaboración (p.ej., la red de actores de Hollywood y las redes de coautoría científica), entre otras

EJEMPLO CLÁSICO DE RED SOCIAL

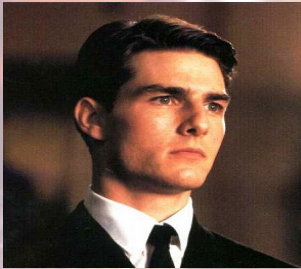
1ª y 2ª elección de compañeras de mesa en una residencia femenina
(Moreno, *The sociometry reader*, 1960)



RED DE ACTORES DE HOLLYWOOD

Nodos: actores

Enlaces: actuaciones conjuntas



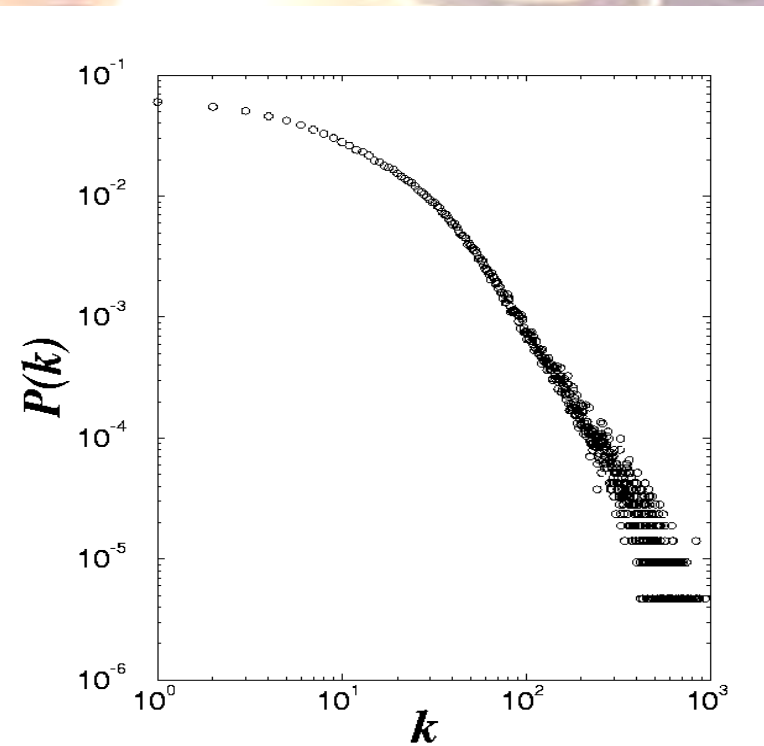
Days of Thunder (1990)
Far and Away (1992)
Eyes Wide Shut (1999)



$N = 212,250$ actores
 $\langle k \rangle = 28.78$

$$P(k) \sim k^{-\gamma}$$

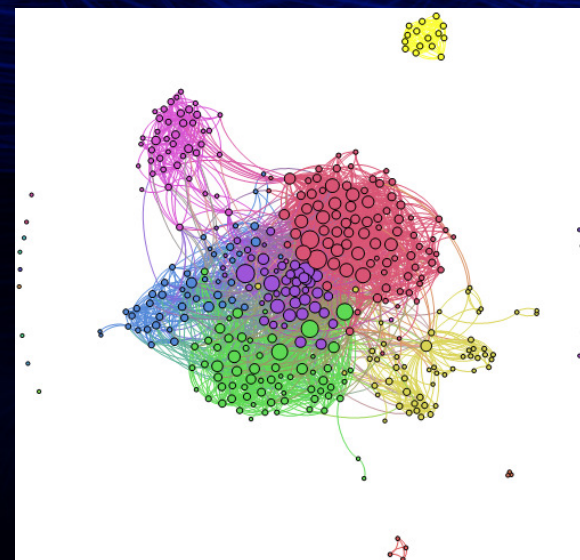
$$\gamma = 2.3$$



REDES SOCIALES ELECTRÓNICAS: FACEBOOK

El "Grafo Social" que hay detrás de Facebook

Keith Shepherd's "Sunday Best"
<http://baseballart.com/2010/07/shades-of-greatness-a-story-that-needed-to-be-told/>



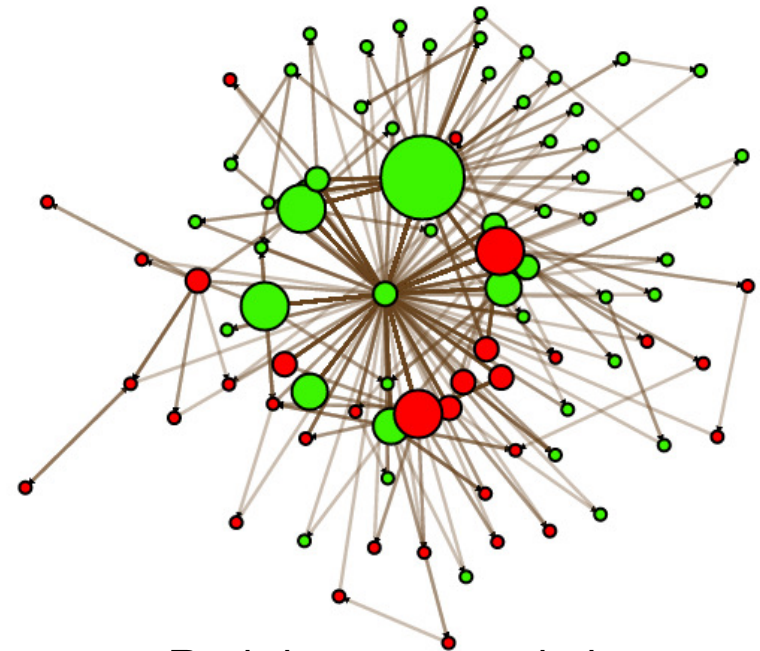
REDES SOCIALES ELECTRÓNICAS: TWITTER (1)

Principalmente, se consideran dos tipos de redes sociales en Twitter:

- **Redes de amistad** (*friend-follower*): Los nodos son usuarios y los enlaces (dirigidos) indican “quién sigue a quién”
- **Redes de flujo de información** (*information flow*): La visualización de los *retweets* refleja el flujo de información en las redes de Twitter



Ejemplo de red de amistad



Red de *retweets* de los tópicos “#zuccotti” y “#nypd”

REDES SOCIALES ELECTRÓNICAS: TWITTER (2)

Existen otras posibilidades como **redes de co-ocurrencia de *hashtags***

Los nodos son *hashtags* y los enlaces indican que dichos *hashtags* aparecen conjuntamente en el mismo *tweet*

Los enlaces están ponderados, indicando el número de co-ocurrencias



ANÁLISIS DE REDES SOCIALES (*SOCIAL NETWORK ANALYSIS*)

El **Análisis de Redes Sociales** (SNA) se centra en el descubrimiento de patrones de interacción entre **actores sociales** en redes sociales

Su objetivo principal es examinar tanto los contenidos como los **patrones de relación en redes sociales para entender las relaciones entre los actores** y las implicaciones de esas relaciones

Son tareas habituales del SNA:

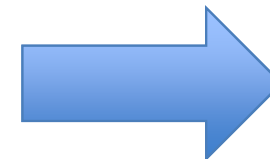
- identificar los actores más **influyentes**, **prestigiosos** o **centrales**, mediante medidas estadísticas,
- identificar **hubs** y **autoridades**, usando algoritmos de análisis de enlaces, y
- Descubrir grupos de actores cohesionados, empleando técnicas de detección de **comunidades**

Wasserman S., Faust, K. Social Network Analysis. Methods and Applications. Cambridge University Press; 2008

ANÁLISIS DE MEDIOS SOCIALES (SOCIAL MEDIA MINING)

Conocimiento útil:

| | | |
|---|---|--|
|  <p>Online Social Networks</p> |  <p>Blogging</p> |  <p>Microblogging</p> |
|  <p>Wikis</p> |  <p>Social News</p> |  <p>Social Bookmarking</p> |
|  <p>Media Sharing</p> |  <p>Opinions and Reviews</p> |  <p>Answers</p> |



| Category | Item | Value | Percentage |
|----------|-------------|-------|------------|
| Top 100 | Google | 1.2 | 1.2% |
| | Facebook | 1.1 | 1.1% |
| | Twitter | 1.0 | 1.0% |
| | LinkedIn | 0.9 | 0.9% |
| | YouTube | 0.8 | 0.8% |
| | Orkut | 0.7 | 0.7% |
| | Myspace | 0.6 | 0.6% |
| | WordPress | 0.5 | 0.5% |
| | Plurk | 0.4 | 0.4% |
| | Del.icio.us | 0.3 | 0.3% |



EJEMPLOS DE APLICACIÓN DEL ANÁLISIS DE REDES Y MEDIOS SOCIALES

- **Marketing viral: maximización de la difusión “boca a boca” de productos** de una compañía dirigiéndose a los clientes de mayor valor en la red (aquellos con una mayor influencia y soporte)
- Análisis de las redes de llamadas telefónicas en compañías de telecomunicaciones para **identificación de perfiles de los usuarios y recomendación de tarifas personalizadas** de acuerdo a dichos perfiles
- Uso para **predicción de la deserción de clientes** (*churn prediction*) identificando cambios en sus patrones de contactos telefónicos
- **Detección de fraudes**, por ejemplo, en comunicaciones organizacionales (conjunto de datos de Enron) para analizar la frecuencia y la dirección de los envíos de e-mails formales/informales que pueden revelar los patrones de comunicación entre empleados y jefes
Estos patrones pueden ayudar a identificar personas implicadas en actividades fraudulentas

TIPOS DE MEDIDAS DE ANÁLISIS DE REDES SOCIALES

De la misma forma que en el análisis de redes en general, el análisis de la estructura de las redes sociales pretende entender el comportamiento de los sistemas complejos (en este caso, *sistemas sociales*) que generan dichas redes

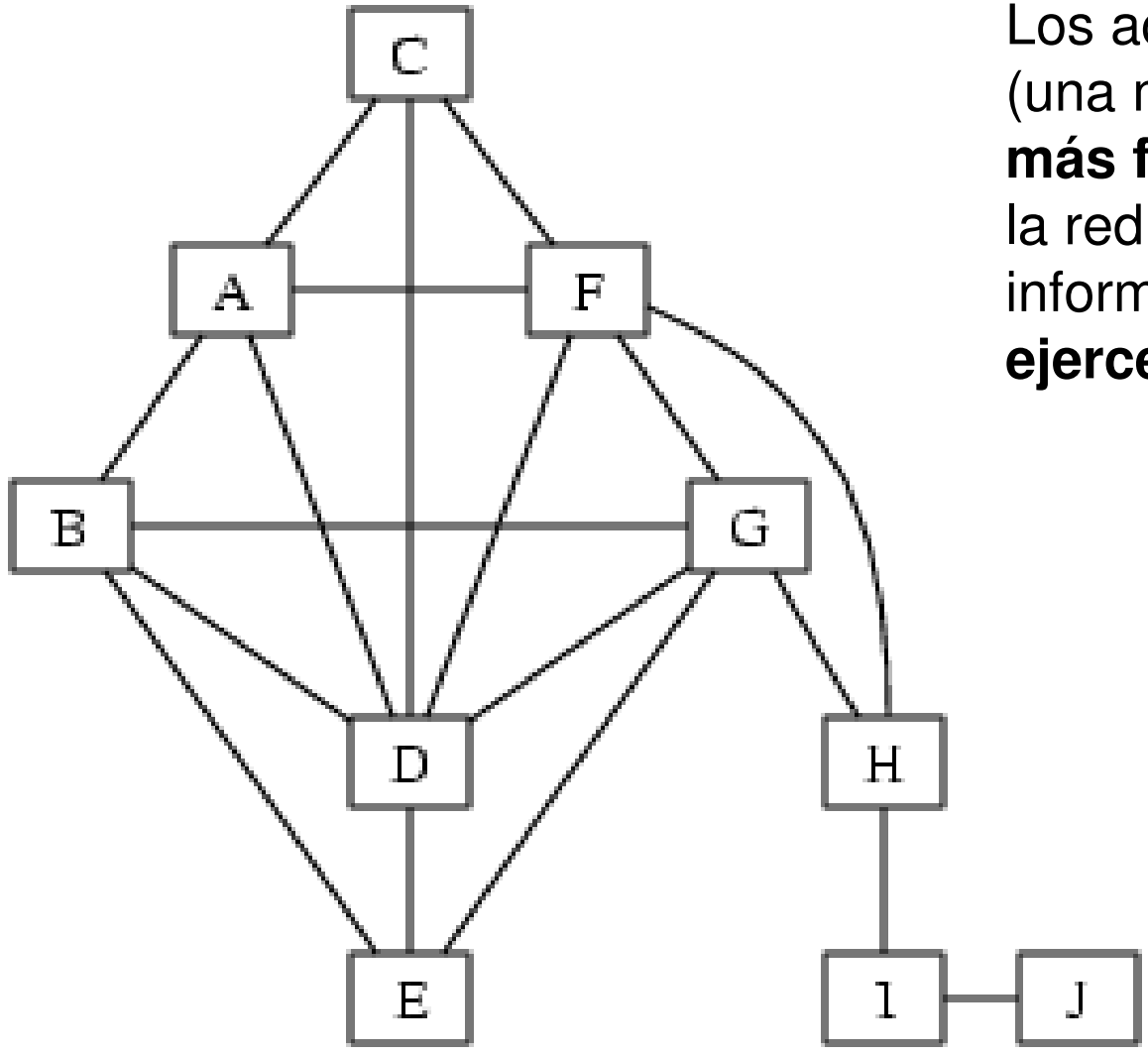
Existen dos tipos de medidas:

- **Medidas locales (a nivel de actores):** Todas ellas están basadas en el concepto general de **centralidad** (redes no dirigidas) o **prestigio** (redes dirigidas), una medida general de la posición de un actor en la estructura global de la red social

Se usan para identificar los **actores clave** de la red. Muestran como las relaciones se concentran en unos pocos individuos, dando una idea de su *poder social*

- **Medidas globales (a nivel de red):** Proporcionan información más compacta que permite evaluar la estructura global de la red, proporcionando información sobre propiedades importantes de los fenómenos sociales subyacentes

La mayoría coinciden con las empleadas para analizar cualquier otro tipo de red

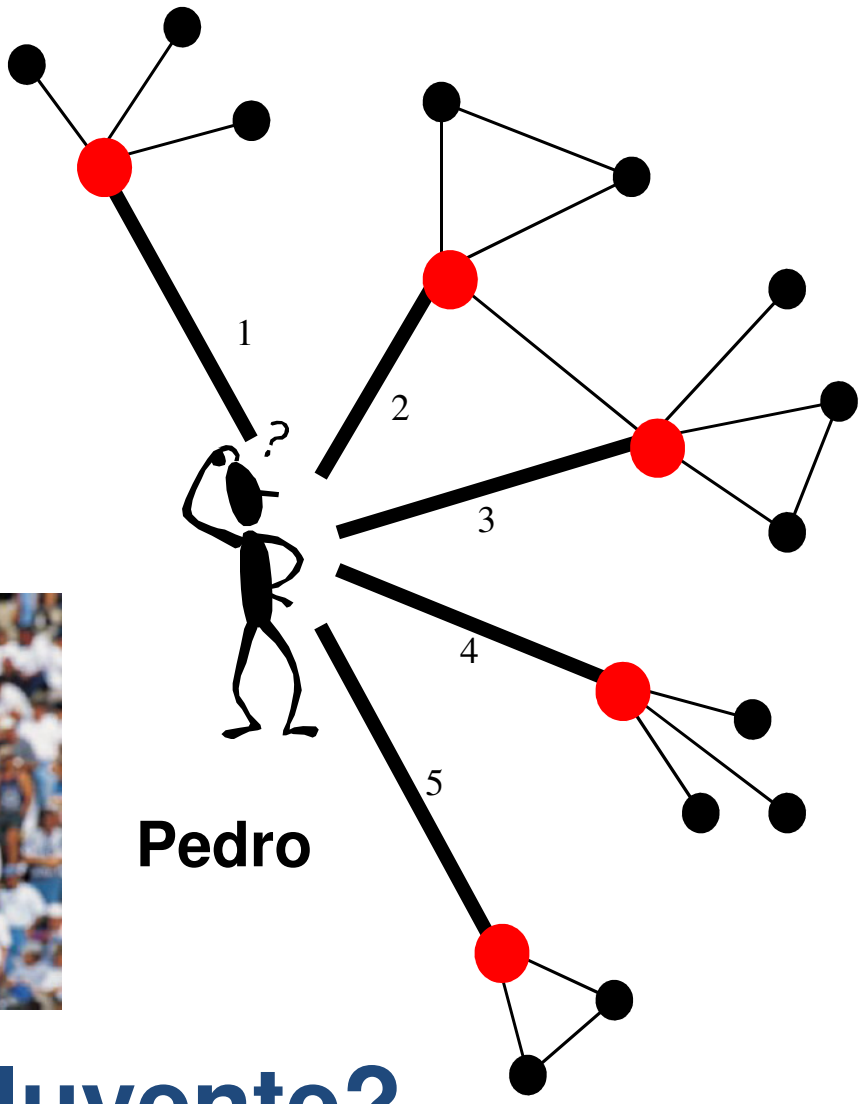
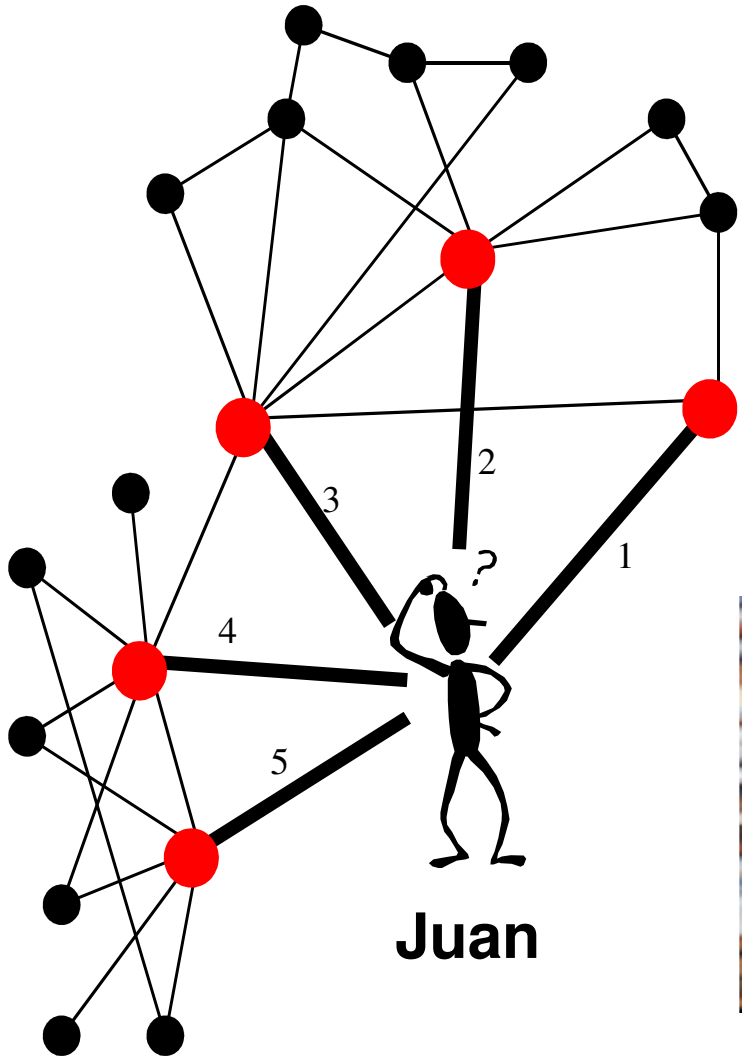


Los actores con una **“posición más central”** (una mayor centralidad) tienen un **acceso más fácil y rápido a los demás actores** de la red (útil para acceder a recursos como información) y una **mayor capacidad para ejercer un control del flujo** entre ellos

¿Quién es central en esta red?

MEDIDAS LOCALES DE CENTRALIDAD

Concepto de Centralidad

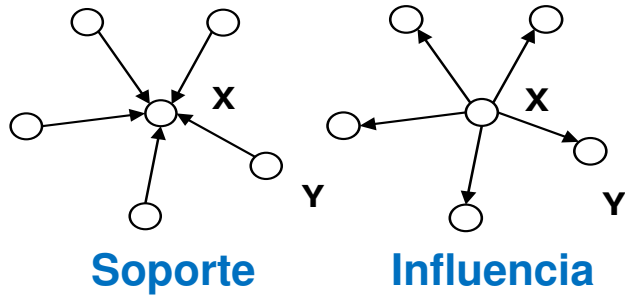


¿Quién es más influyente?

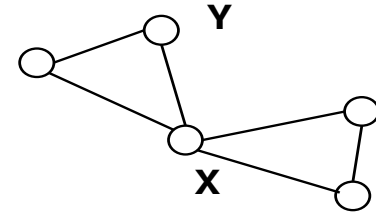
MEDIDAS LOCALES DE CENTRALIDAD

Existen varias medidas distintas de centralidad:

1. Grado:



2. Intermediación
(*betweenness*):



3. Cercanía (*closeness*):

x y

4. Excentricidad

5. Centralidad de vector propio

6. Coeficiente de clustering

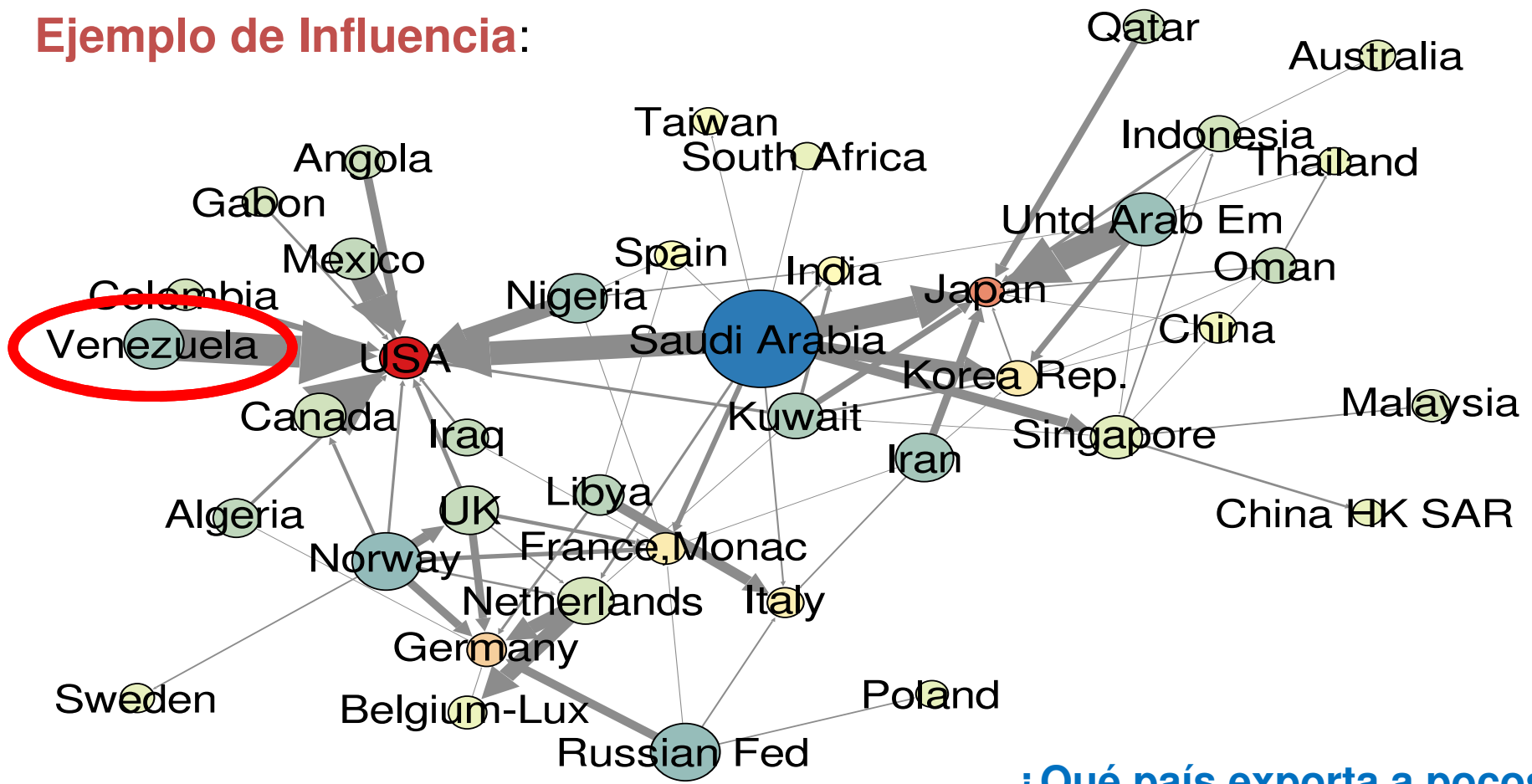
Es conveniente conocer bien las características de cada medida y usar varias

P.ej., las Centralidades de grado son medidas importantes pero no tienen en cuenta la estructura global de la red

MEDIDAS LOCALES DE CENTRALIDAD

Influencia

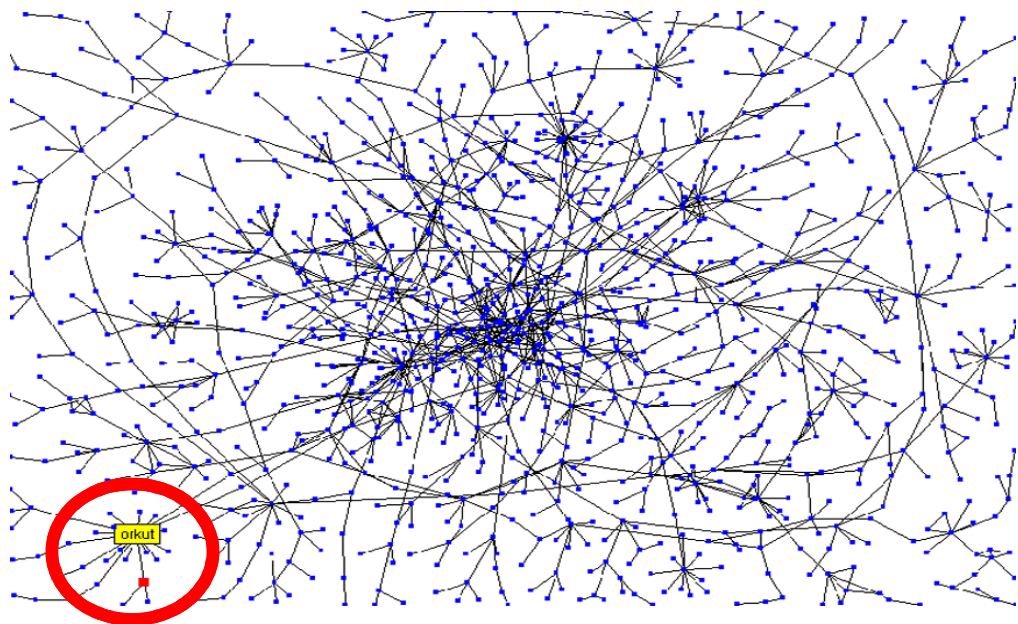
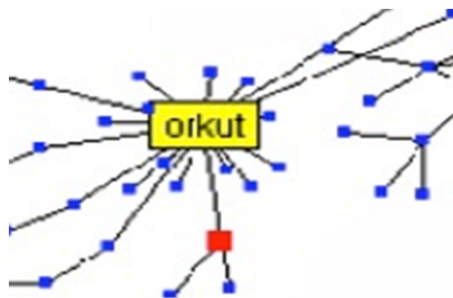
Ejemplo de Influencia:



¿Qué país exporta a pocos países (*out-degree* bajo) pero lo hace en gran cantidad (grosor de los arcos = volumen exportado)?
¿Arabia Saudí, Japón, Irak, USA, Venezuela?

MEDIDAS LOCALES DE CENTRALIDAD

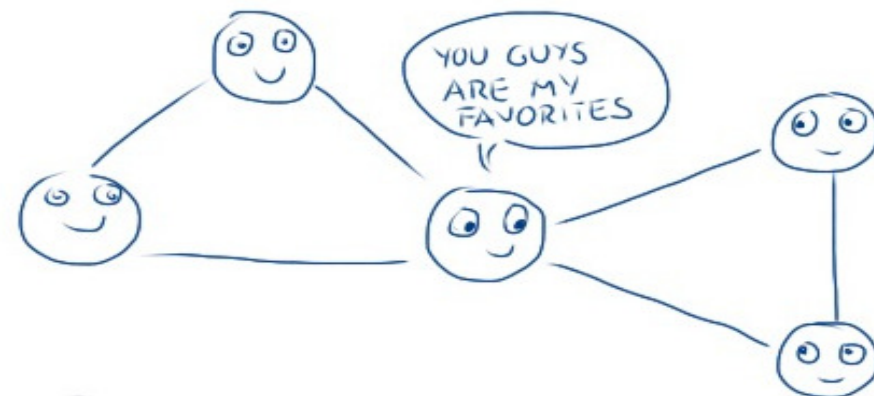
Stanford Social Web (ca. 1999)

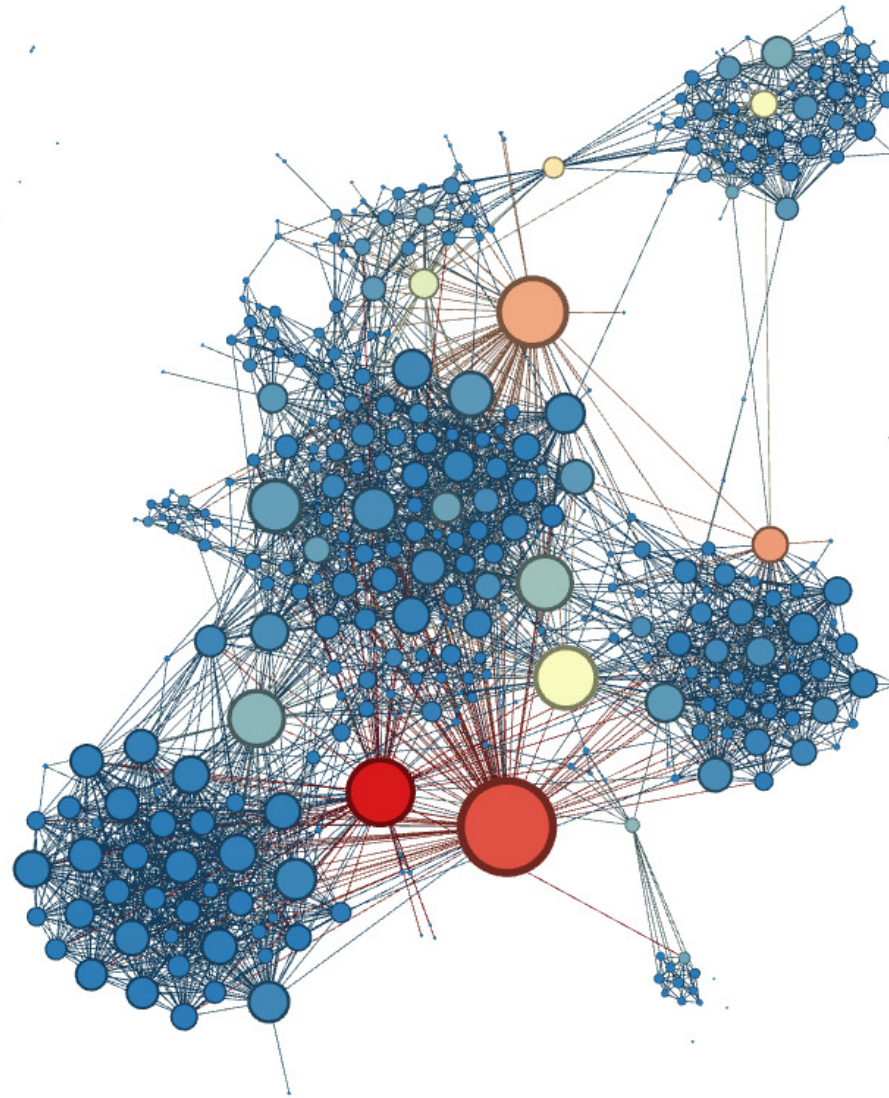


Red de páginas web personales en Stanford

Comportamiento

El grado no captura las “corredurías” (*brokerage*)



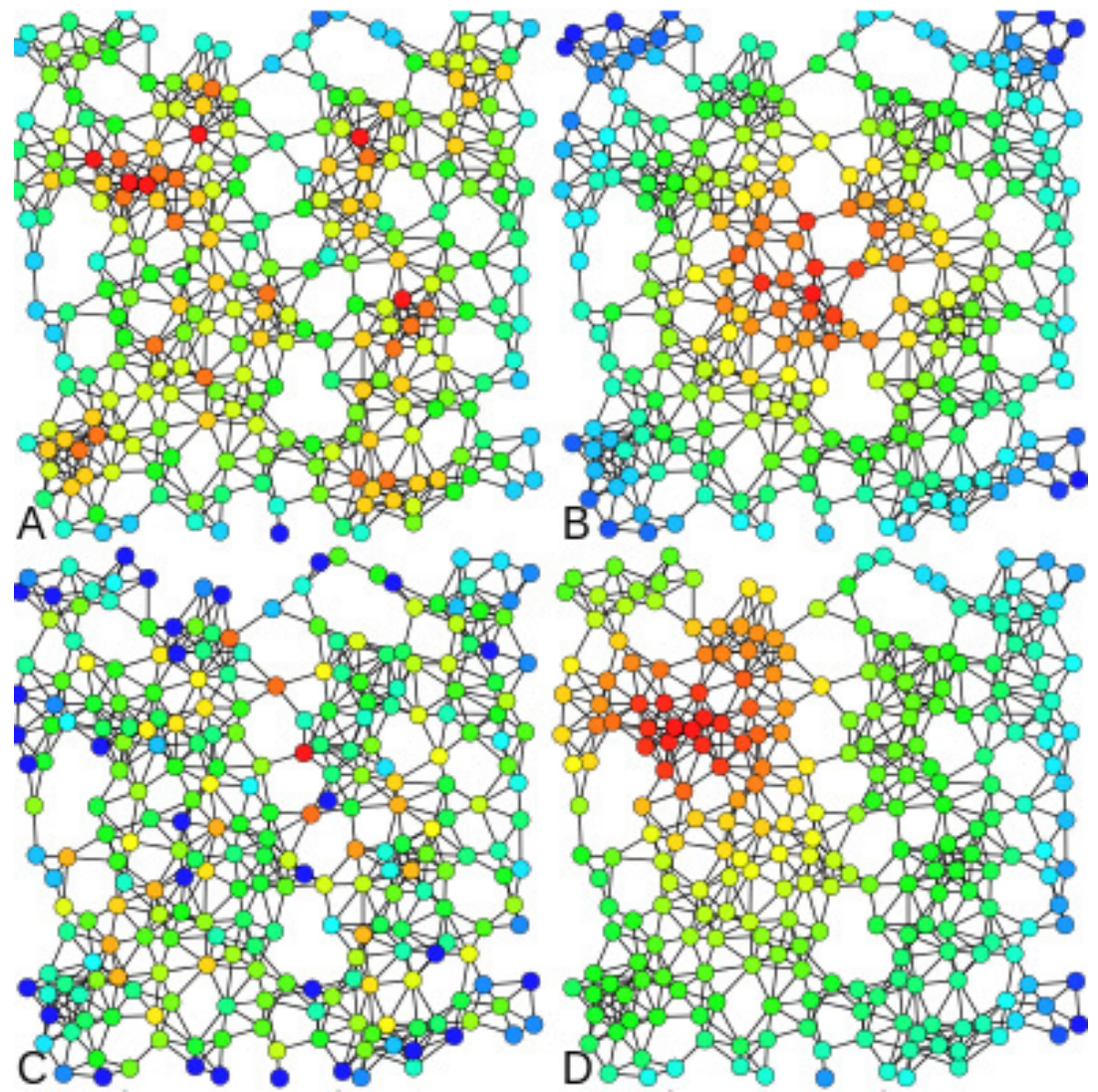


Red Personal de Contactos de Facebook de Oscar Cordón: el tamaño de los nodos indica el **grado** y el color la **intermediación** (más **azul**, menor valor; más **rojo**, mayor valor)

MEDIDAS LOCALES DE CENTRALIDAD

Comparativa

A) centralidad de grado; B) cercanía; C) intermediación; D) centralidad de vector propio

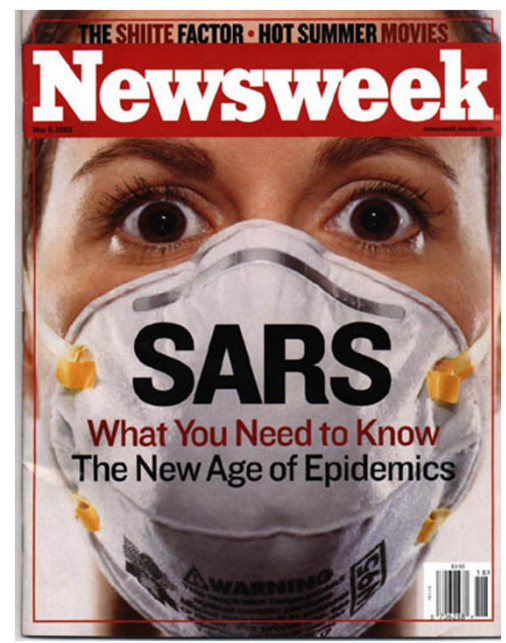


azul = menor valor

rojo = mayor valor

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:

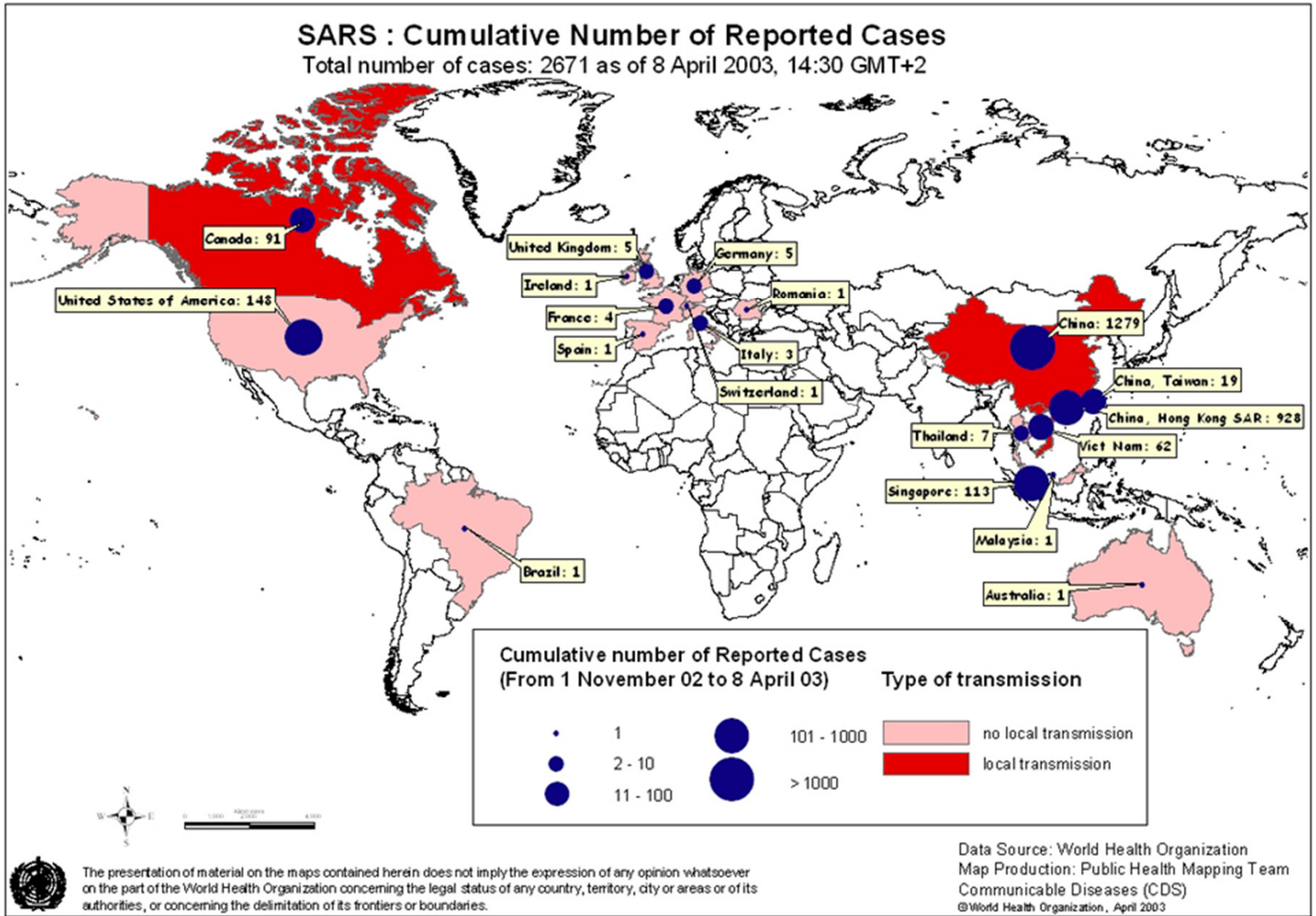
- Enfermedades transmitidas por el aire (gripe, gripe aviar, ...)
- Enfermedades venéreas (VIH, ...)
- Otras enfermedades infecciosas, incluidos algunos cánceres (VPH (virus del papiloma humano, ...))
- Parásitos (chinchas, malaria, ...)

TIC:

- Virus de ordenador, gusanos
- Virus de teléfonos móviles

Conceptos/Aspectos intelectuales:

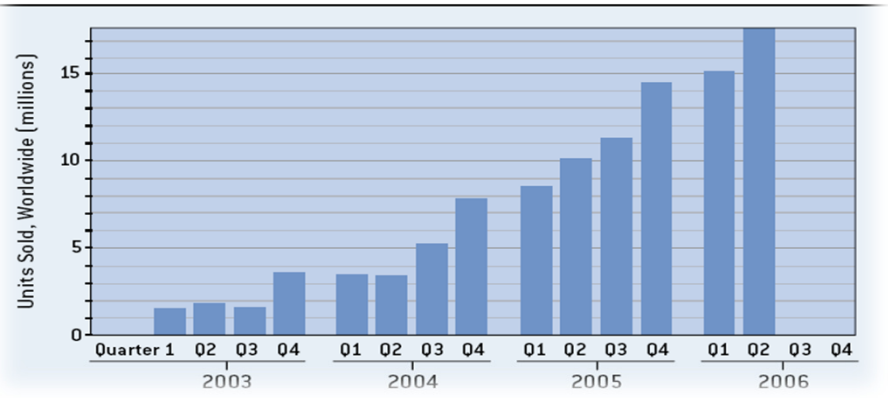
- Difusión de innovaciones
- Rumores
- Prácticas empresariales



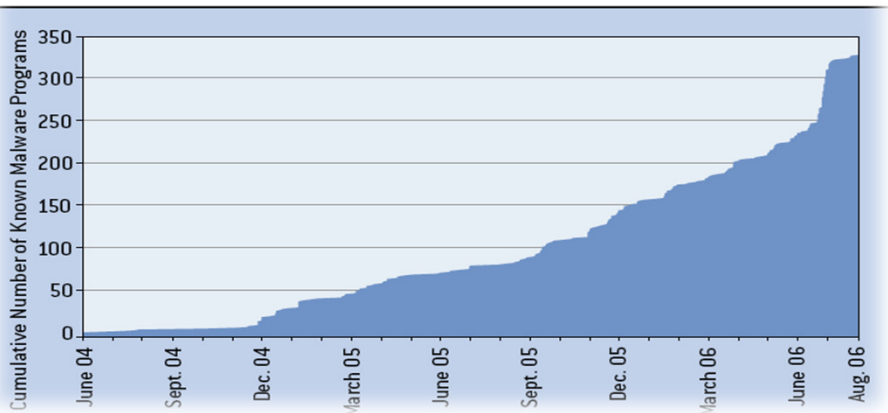
INTRODUCCIÓN

Epidemias Tecnológicas: Virus Informáticos y de Teléfonos Móviles

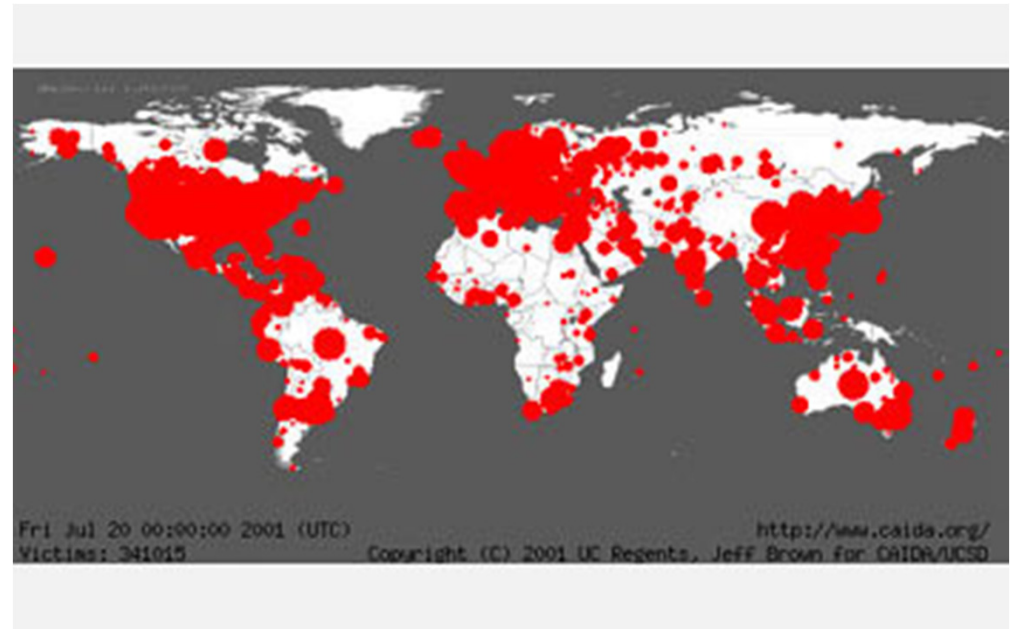
SMARTPHONES ON THE RISE



GROWTH IN MOBILE MALWARE



El código del *Red Worm* paralizó Internet en muchos países

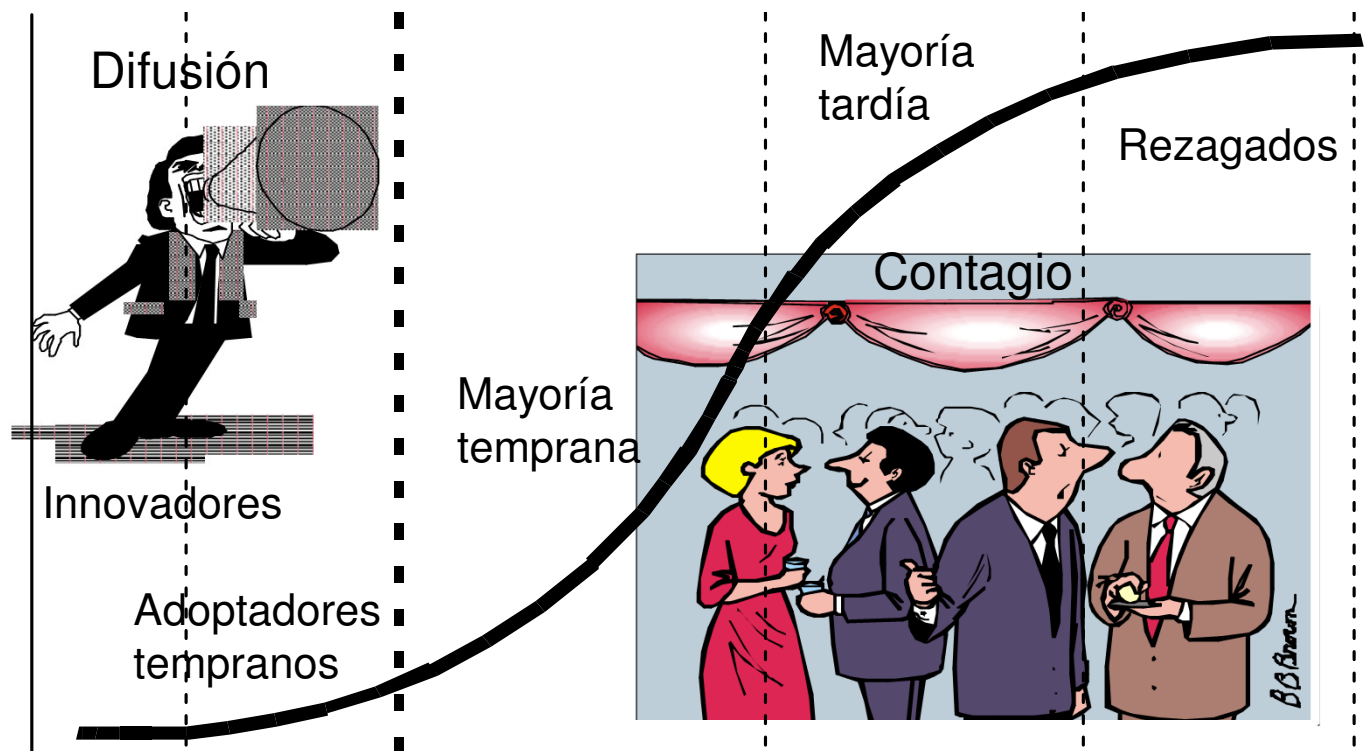


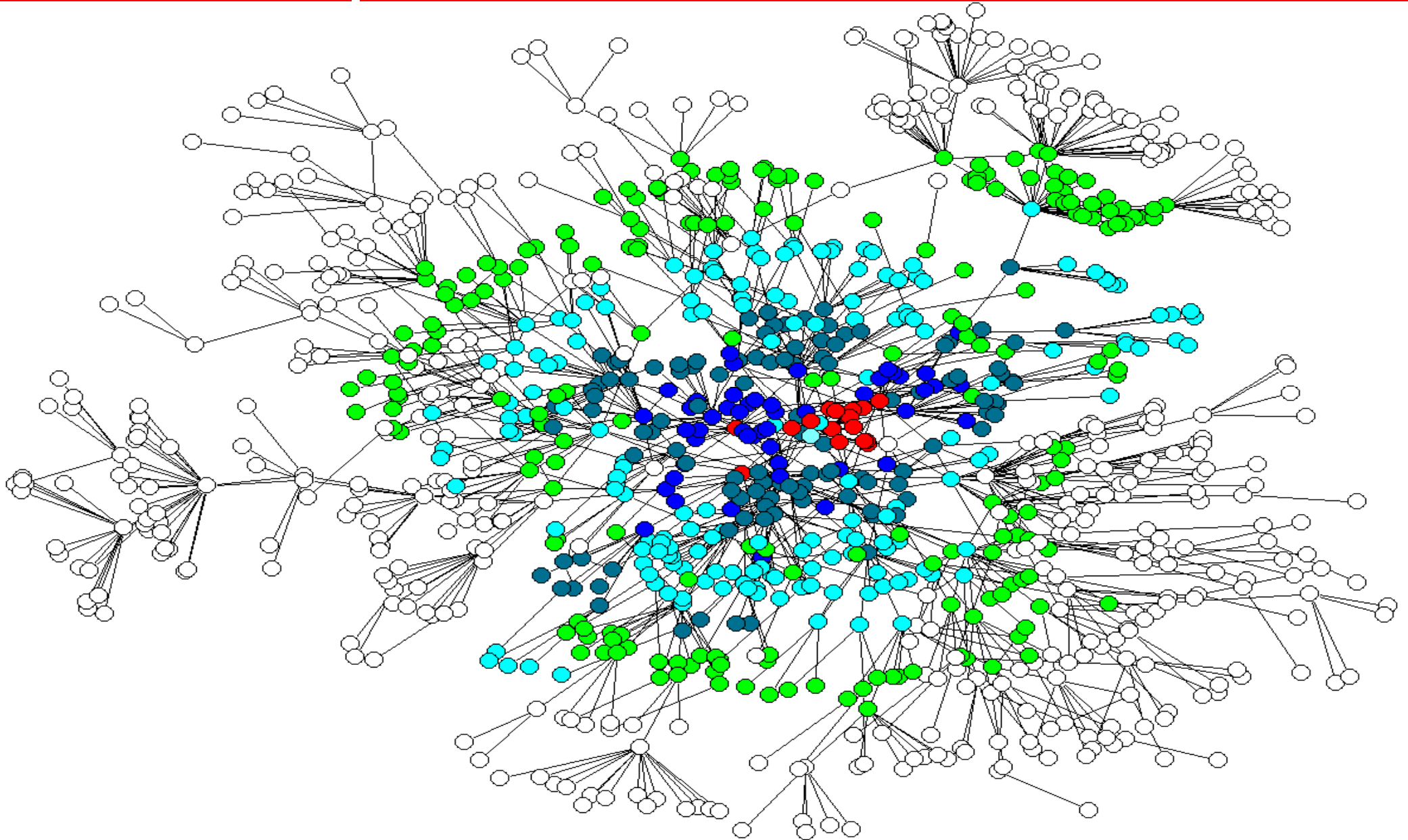
<http://www.caida.org/publications/visualizations/>

Hypponen M. Scientific American (Nov. 2006) 70-77

INTRODUCCIÓN

Difusión de la Innovación – Curva de Adopción



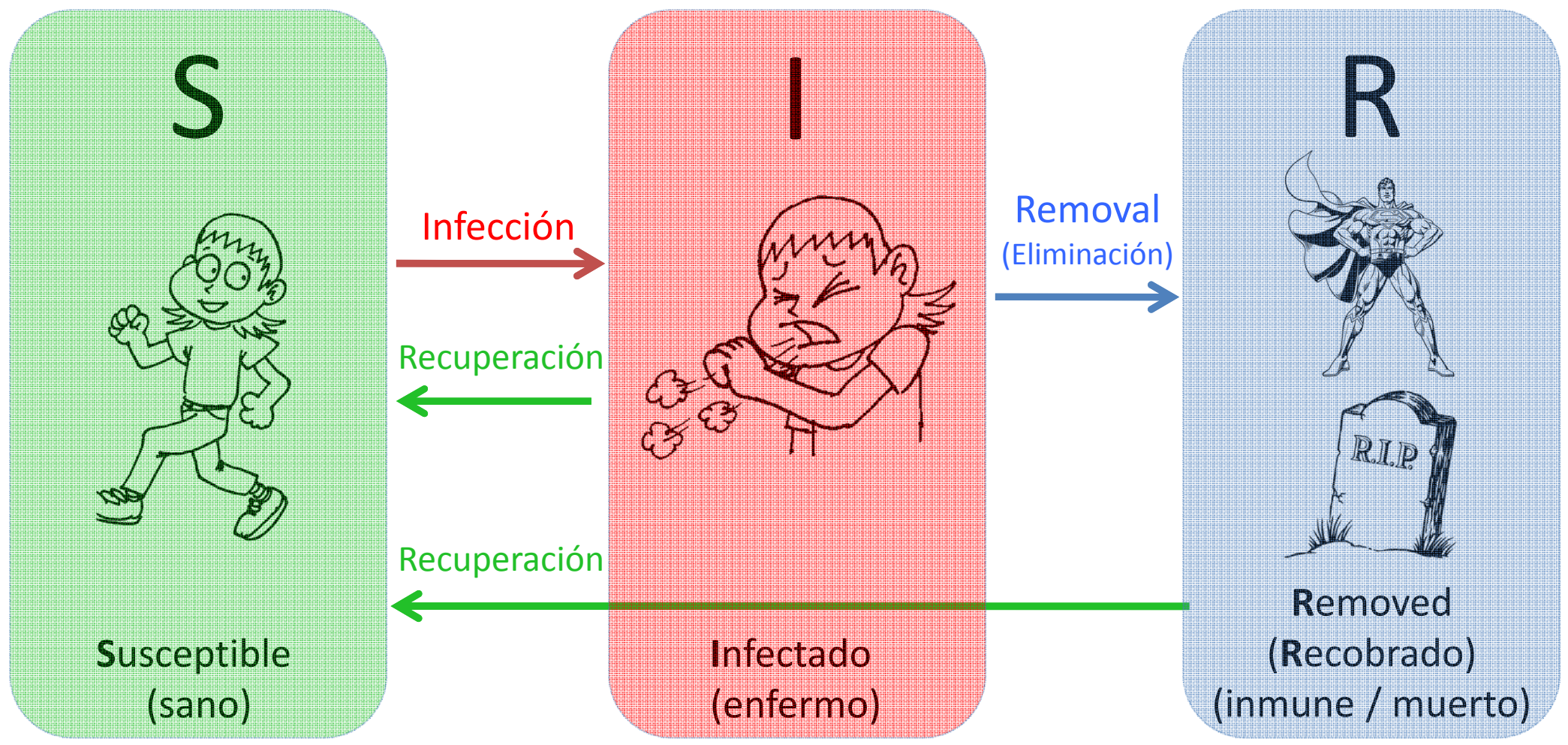


MODELOS CLÁSICOS DE PROPAGACIÓN DE EPIDEMIAS

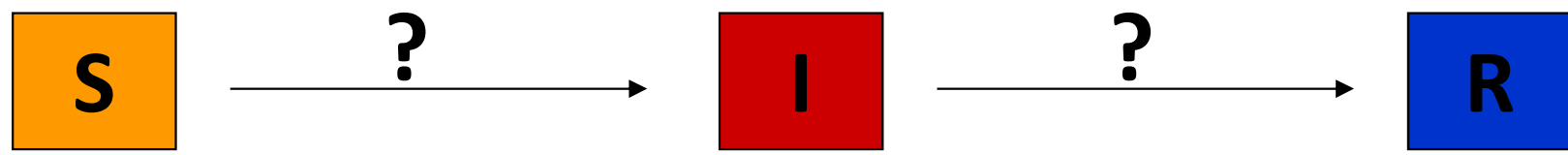
[http://es.wikipedia.org/wiki/Modelaje matemático de epidemias](http://es.wikipedia.org/wiki/Modelaje_matemático_de_epidemias)

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

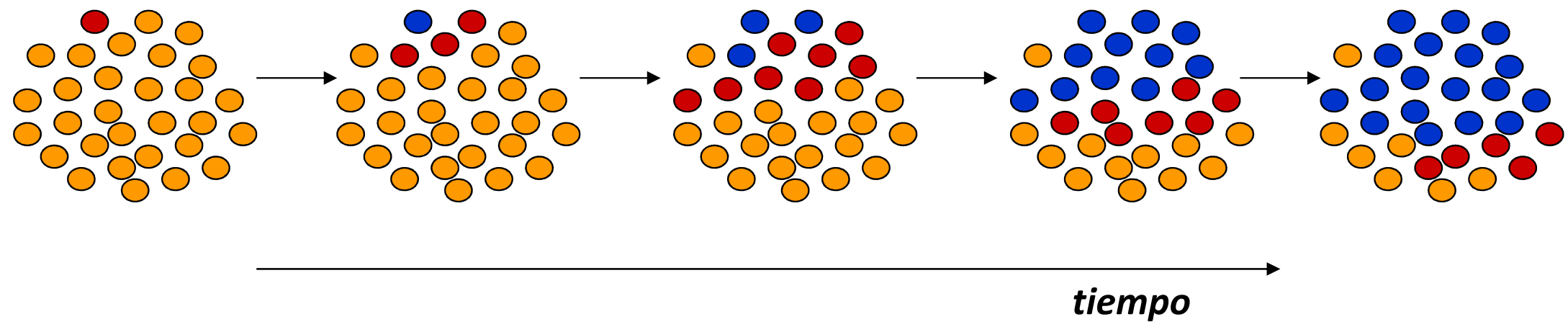
ESTADOS BÁSICOS Y TRANSICIONES DEL MODELO CLÁSICO SIR



DINÁMICA DE LAS VARIANTES DEL MODELO SIR

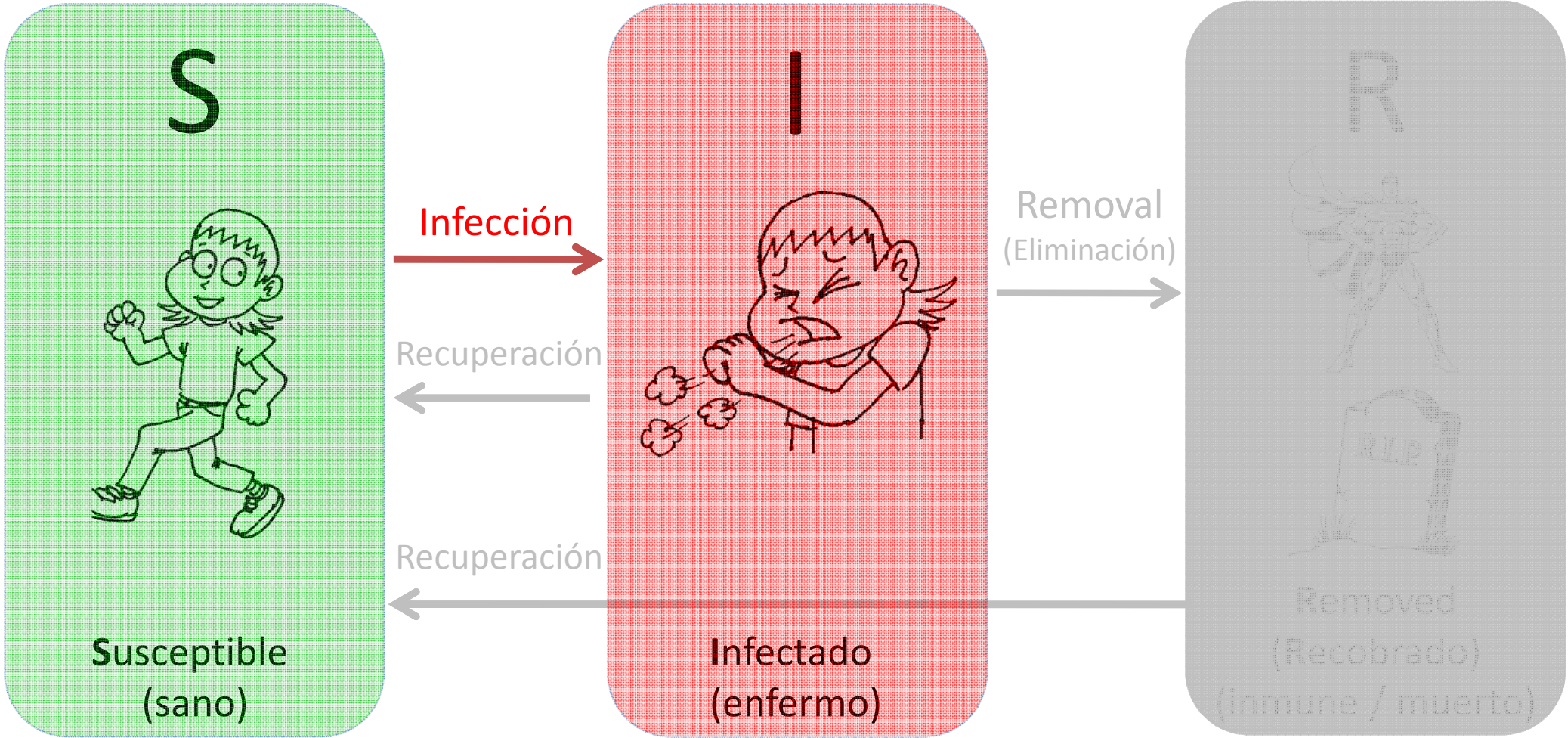


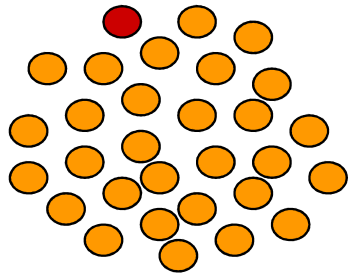
¿Dinámica del proceso epidémico?



MODELO SI

Modelo teórico más simple





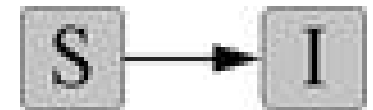
Mezclado Homogéneo (sin red compleja subyacente):

En cada unidad de tiempo, cada individuo tiene β contactos con otros individuos de la población escogidos aleatoriamente y puede quedar infectado como consecuencia de esos contactos

Si en la población de tamaño N hay I individuos infectados y S individuos susceptibles ($N=I+S$), la probabilidad de encontrar un individuo susceptible es 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 $\beta \cdot S/N$ individuos sanos

Por tanto, el **ratio medio de nuevas infecciones** es: $\frac{\beta \cdot S \cdot I}{N}$



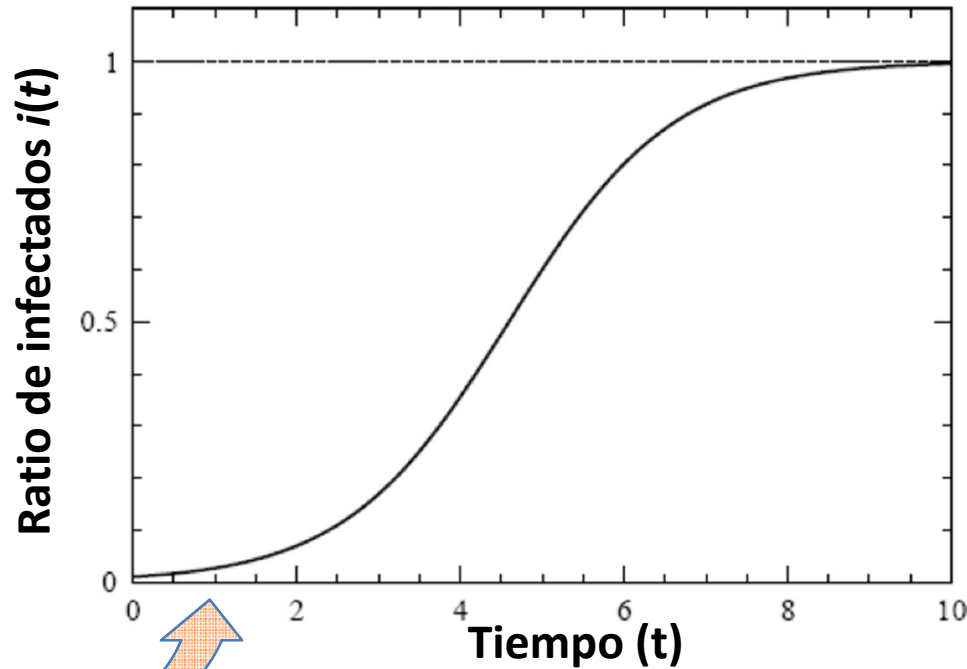
Y las ecuaciones diferenciales del sistema son: $\frac{dI}{dt} = \beta \cdot \frac{S \cdot I}{N}$; $\frac{dS}{dt} = -\beta \cdot \frac{S \cdot I}{N}$

Ecuación logística: modelo básico de crecimiento de poblaciones

$$\frac{di}{dt} = \beta \cdot i \cdot (1 - i)$$

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

Si $i(t)$ es pequeño,
 $\frac{di}{dt} \approx \beta \cdot i$
 $i \approx i_0 \exp(\beta \cdot t)$
brote exponencial



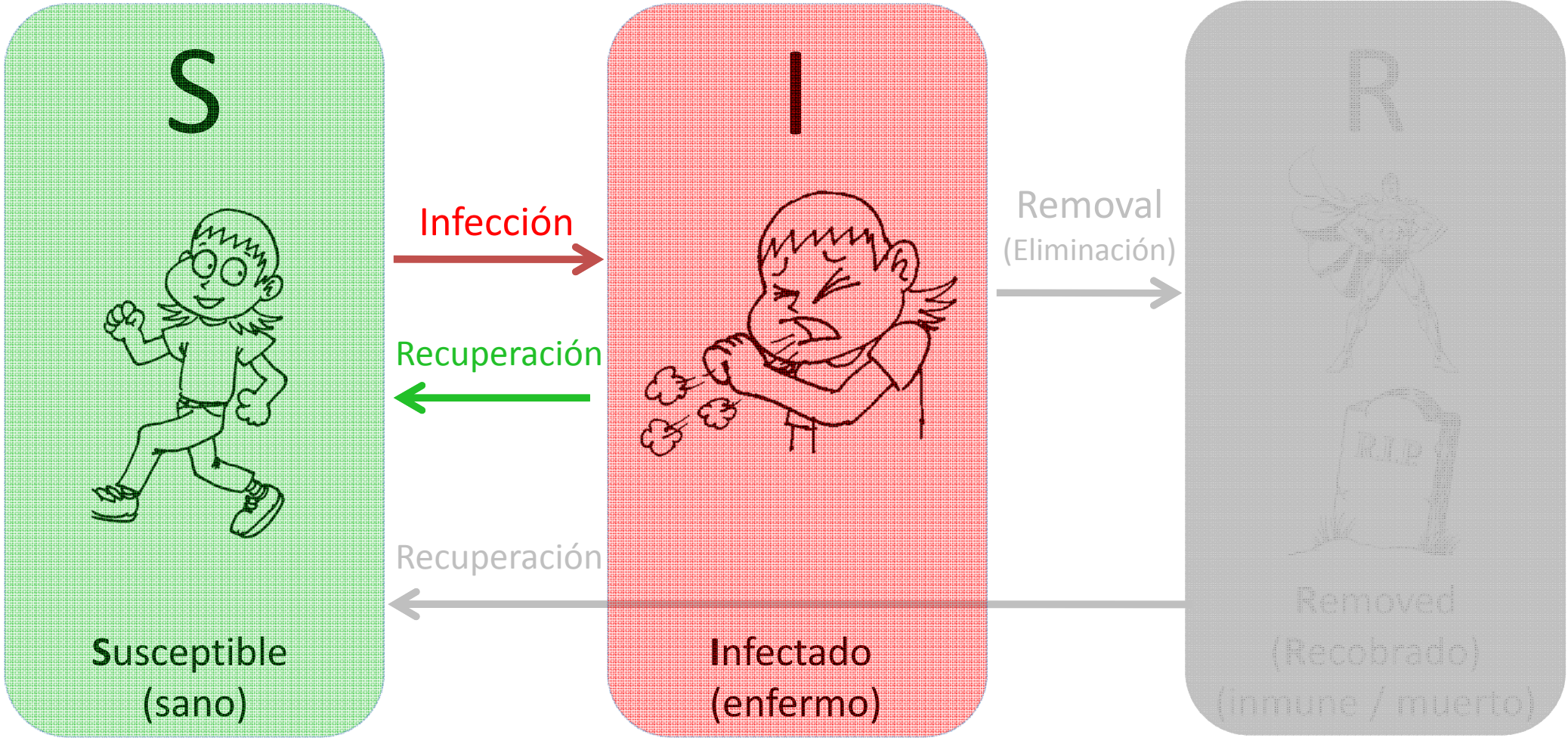
Cuando $i(t) \rightarrow 1$,
 $\frac{di}{dt} \rightarrow 0$
 saturación

http://es.wikipedia.org/wiki/Función_logística
<http://mathworld.wolfram.com/LogisticEquation.html>

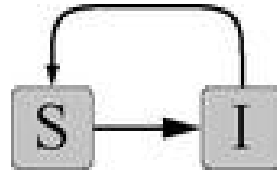
Modelo SI: el ratio de infectados aumenta hasta que toda la población queda infectada

MODELO SIS

Resfriado común



El modelo SIS extiende al SI considerando la posibilidad de **recuperación y reinfección** de los individuos:



Sigue manteniendo el mezclado homogéneo y la existencia de sólo dos estados: S e I . Los individuos infectados vuelven al estado susceptible después de su recuperación (dos transiciones posibles: $S \rightarrow I$ y $I \rightarrow S$)

Se mantiene el ratio de contactos β y **se incorpora una probabilidad de recuperación μ para los individuos infectados**

Las ecuaciones diferenciales del sistema son:

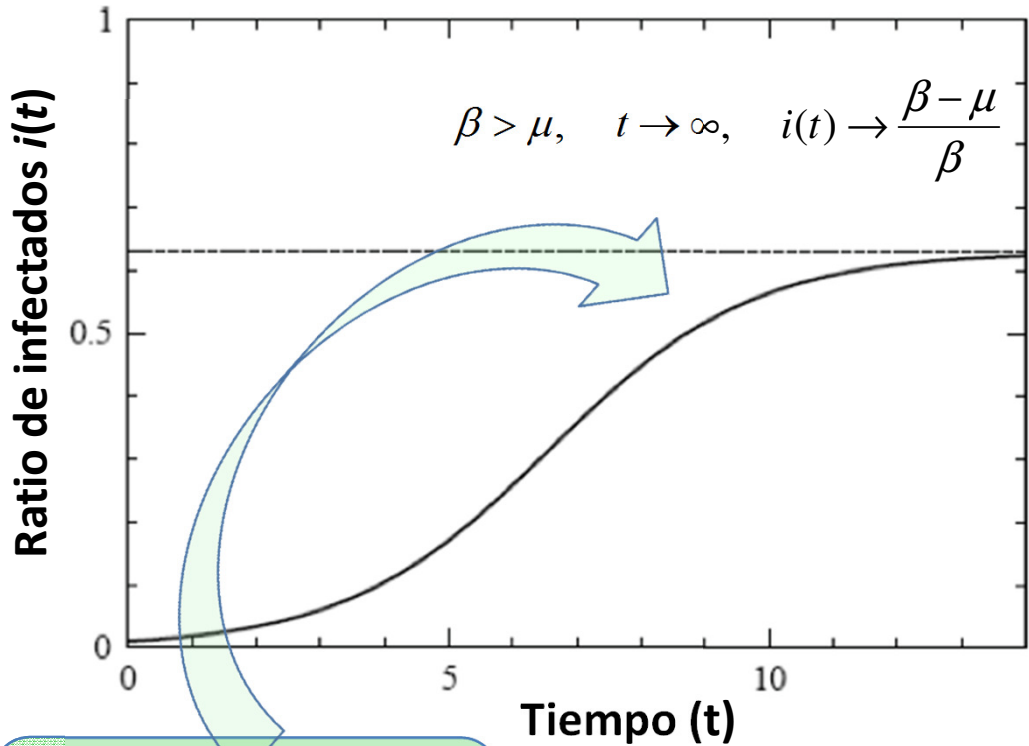
$$\frac{di}{dt} = \beta \cdot s \cdot i - \mu \cdot i \quad ; \quad \frac{ds}{dt} = \mu \cdot i - \beta \cdot s \cdot i \quad ; \quad s + i = 1$$

MODELO SIS

Comportamiento

$$\therefore i(t) = \left(1 - \frac{\mu}{\beta}\right) \frac{C \cdot e^{(\beta-\mu) \cdot t}}{1 + C \cdot e^{(\beta-\mu) \cdot t}}$$

$$C = \frac{\beta \cdot i_0}{\beta - \mu - \beta \cdot i_0}$$



Estado estacionario:
 $\frac{di}{dt} = \beta \cdot i \cdot (1 - i) - \mu \cdot i = 0$

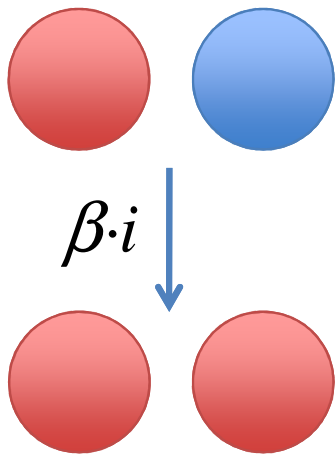
Modelo SIS: Cuando $\beta > \mu$, el ratio de infectados satura por debajo de 1, al contrario que en el modelo SI

Al final, siempre se tiene el mismo ratio de infectados, $(\beta - \mu) / \beta$, aunque no sean exactamente los mismos individuos en cada instante concreto

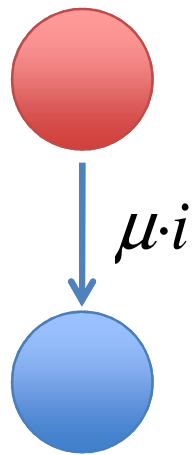
En Epidemiología, ese estado se denomina **estado endémico de la enfermedad**

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

Si $\mu \approx \beta$, $i \rightarrow 0$
 “Umbral epidemiológico”
 Transición de fase



$$I = I + 1$$



$$I = I - 1$$

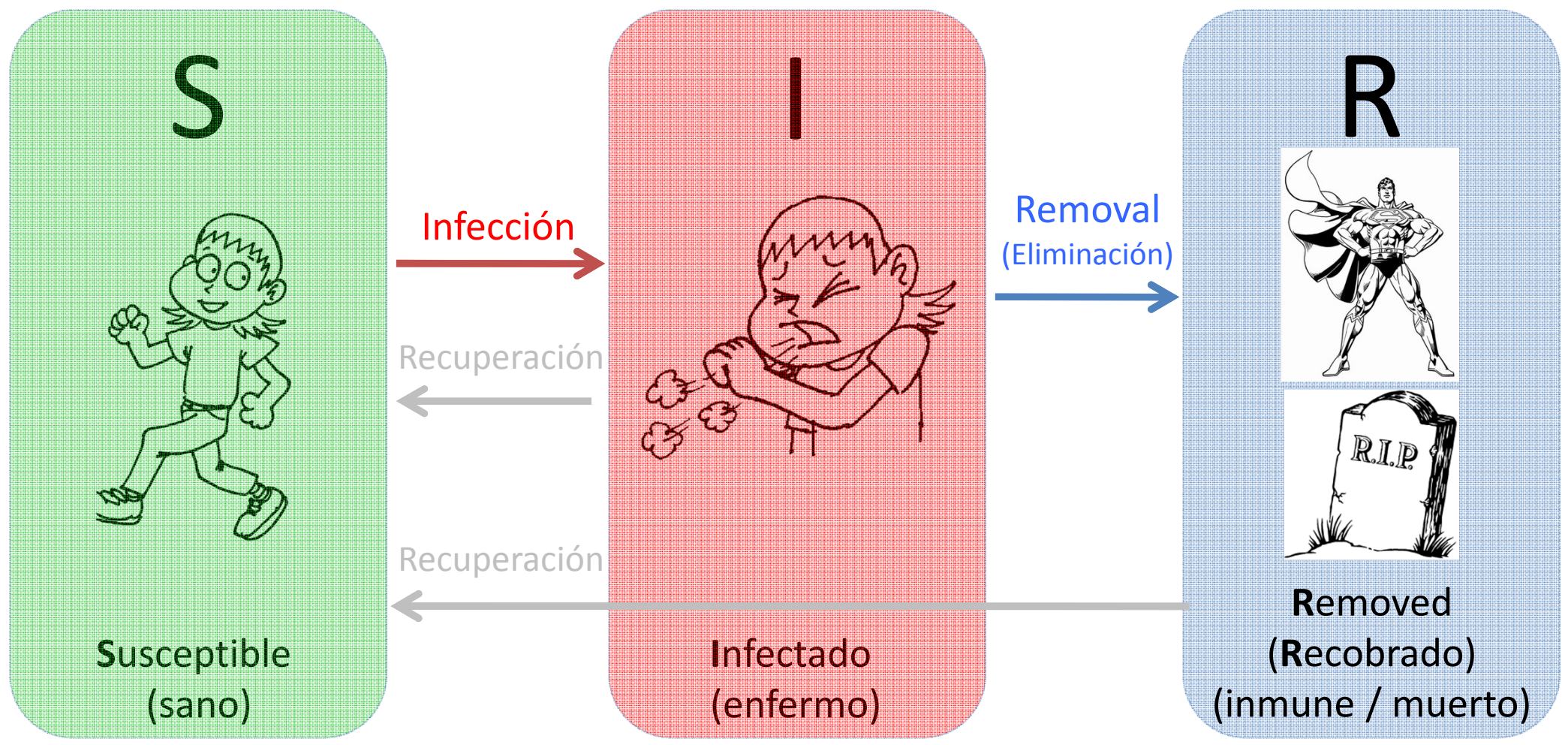
$\lambda(R_0) \equiv \frac{\beta}{\mu}$ “Ritmo reproductivo básico”
 En promedio, ¿cuántos individuos serán infectados por un individuo infectado concreto?

$\lambda > 1$: Epidemia

$\lambda < 1$: Extinción del brote epidémico

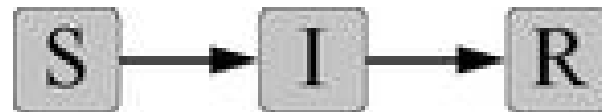
MODELO SIR

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



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

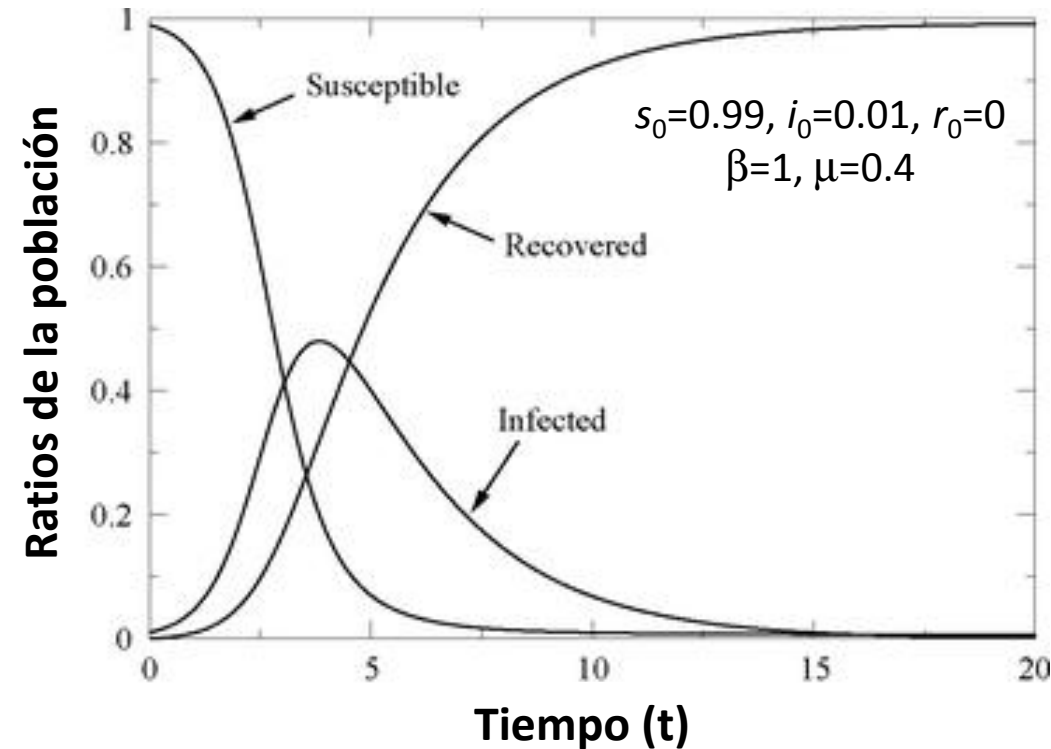
Una vez que un individuo se infecta, puede quedar **permanentemente inmune** a la enfermedad o “**morir**”



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 contactos β). En la segunda, los infectados se recuperan (o mueren) con una probabilidad de recuperación μ



Condiciones iniciales más habituales:
o bien $I=1$ o bien $I=c$, un número pequeño

$$i_0 = \frac{c}{N} ; s_0 = 1 - i_0 = 1 - \frac{c}{N} ; r_0 = 0$$

Modelo SIR: Cuando $\beta > \mu$, i crece hasta un pico máximo y luego decrece hasta valer 0

El ratio de susceptibles decrece de forma monótona y el de recobrados crece igual

s satura pero no llega nunca a cero porque con $i \rightarrow 0$, ya no hay individuos que puedan infectar

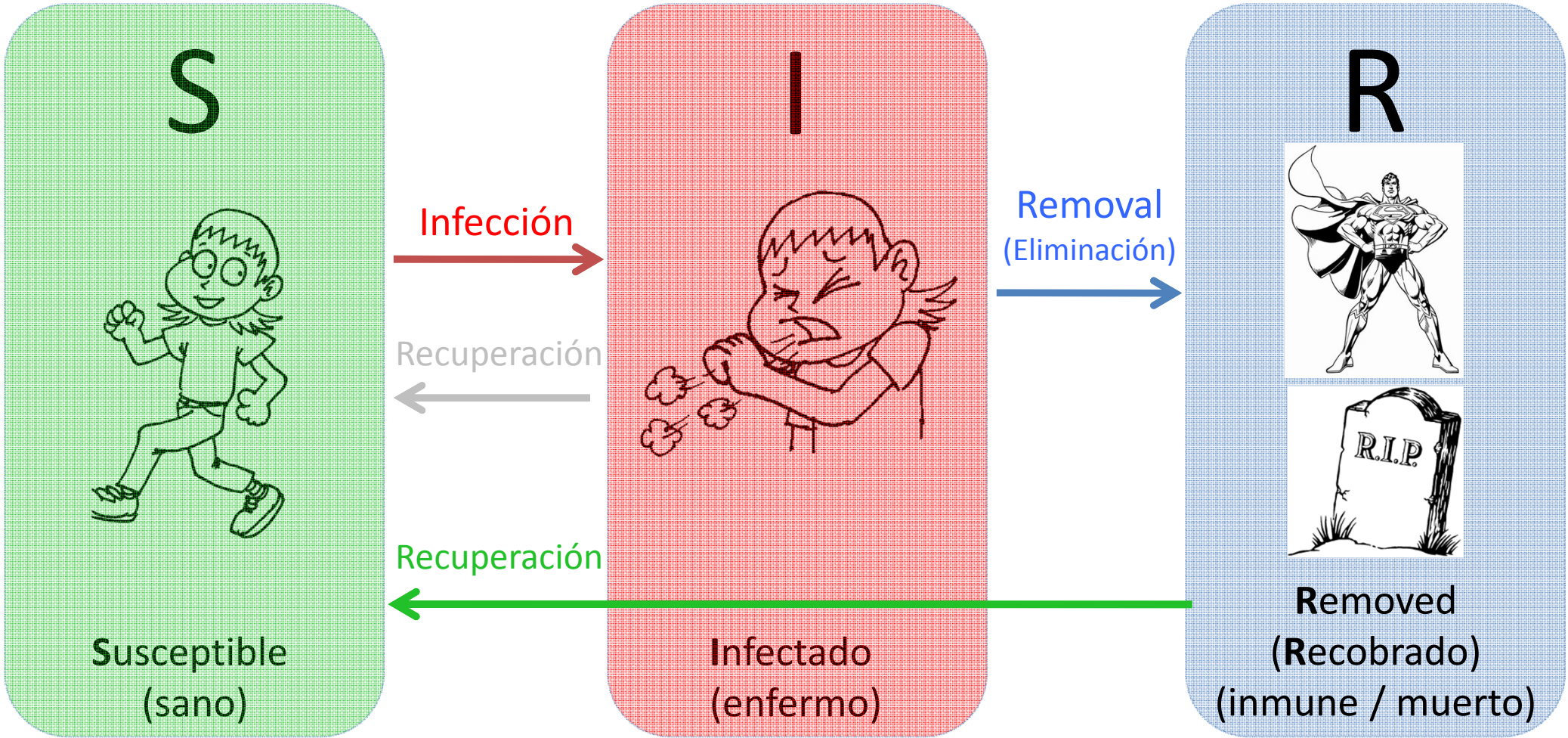
Los individuos que consiguen mantenerse sanos hasta fases avanzadas, pueden no infectarse nunca. Igualmente, r nunca llega a valer 1

El valor asintótico de r es el número de individuos afectados por la epidemia:

$$r = 1 - s_0 \cdot e^{-\beta \cdot r / \mu}$$

VARIANTES

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



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**

El modelo SI es el más relevante para este comportamiento

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.

Aplicación de los modelos epidemiológicos para predecir el ciclo de vida (la **dinámica de adopción y abandono de usuarios**) en redes sociales on-line (RSO) como **Facebook**

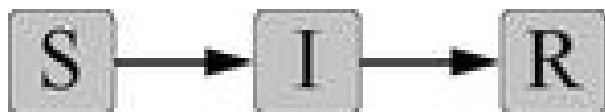
Se basa en la idea de que dicha dinámica se asemeja a los procesos epidémicos porque los usuarios se unen y abandonan la RSO porque sus amigos lo hacen



Se cambia ligeramente el modelo SIR tradicional y se usan datos de la única RSO conocida que “desapareció”, **MySpace**, para predecir el comportamiento futuro de Facebook

El nuevo modelo se denomina **infectious recovery SIR** (irSIR). Se modifica la recuperación porque los usuarios no piensan abandonar la RSO en principio. En teoría, los usuarios que se unen antes tienen más interés y se espera que permanezcan más tiempo

El proceso contagioso de abandono está controlado por los usuarios que dejan la RSO. Sus contactos consideran también abandonarla, como en el **churning** en compañías de móviles



S= individuos susceptibles de unirse a la RSO vía contagio. I= usuarios actuales de la RSO. R= individuos contrarios a unirse a la RSO (los que nunca se han unido y los que la han dejado)

Para incluir la dinámica de recuperación de la infección **se modifica su ratio para que sea proporcional a la población de recuperados**, los que provocan el abandono de la RSO:

Modelo SIR: $\frac{di}{dt} = \beta \cdot s \cdot i - \mu \cdot i$; $\frac{ds}{dt} = -\beta \cdot s \cdot i$; $\frac{dr}{dt} = \mu \cdot i$; $s + i + r = 1$

Modelo irSIR: $\frac{di}{dt} = \beta \cdot s \cdot i - \nu \cdot i \cdot r$; $\frac{ds}{dt} = -\beta \cdot s \cdot i$; $\frac{dr}{dt} = \nu \cdot i \cdot r$; $s + i + r = 1$

La probabilidad de recuperación se nota por ν para indicar el cambio (μ en el modelo SIR)

| Symbol | Units | Disease Model Parameter | Equivalent OSN Model Parameter |
|----------|--------------------|-------------------------|--|
| S | People | Susceptible | Potential OSN users |
| I | People | Infected | OSN users |
| R | People | Recovered/Immune | Population opposed to OSN use |
| β | Time ⁻¹ | Infection rate | Rate at which potential users join OSN |
| γ | Time ⁻¹ | recovery rate | - |
| ν | Time ⁻¹ | - | OSN abandonment rate |

El ratio inicial de recuperados r_0 (individuos que se resisten inicialmente a unirse a la RSO) es un parámetro importante en el modelo irSIR . **Si $r_0=0$, no hay abandonos**

El umbral epidemiológico depende de ν , β , s_0 y r_0 . Se considera al revés que en SIR, se produce infección de recuperación (el proceso de contagio de abandono provoca que se reduzca progresivamente el número de usuarios de la RSO) si:

$$\frac{\nu}{\beta} > \frac{s_0}{r_0}$$

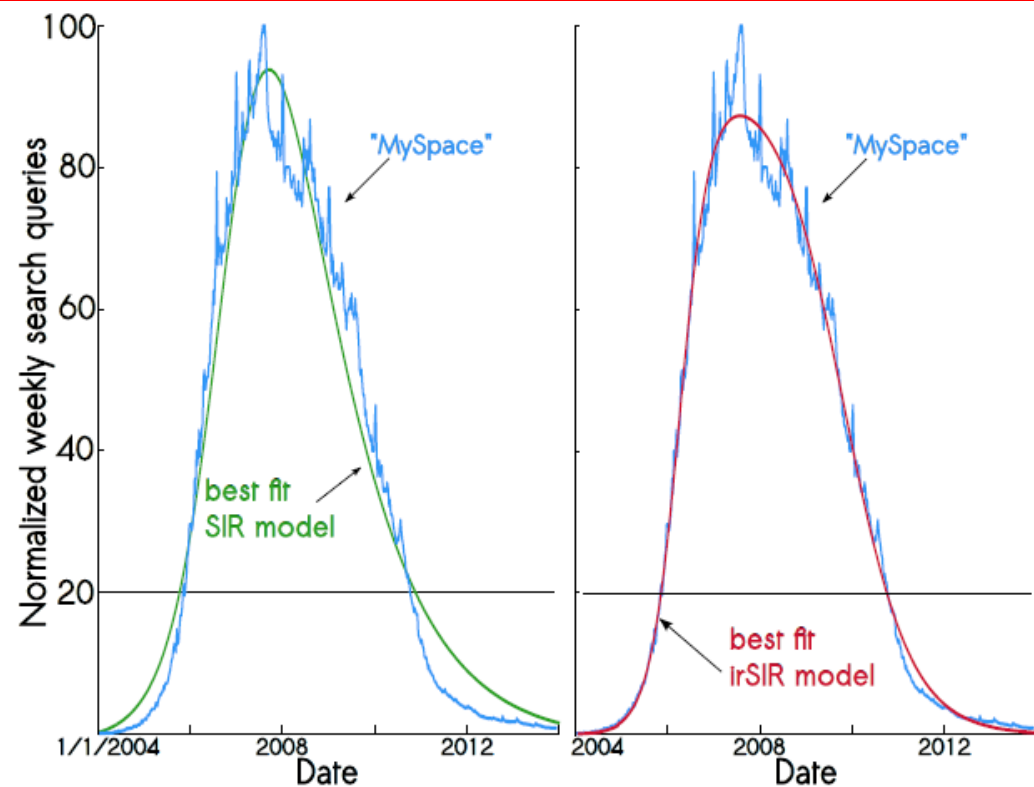
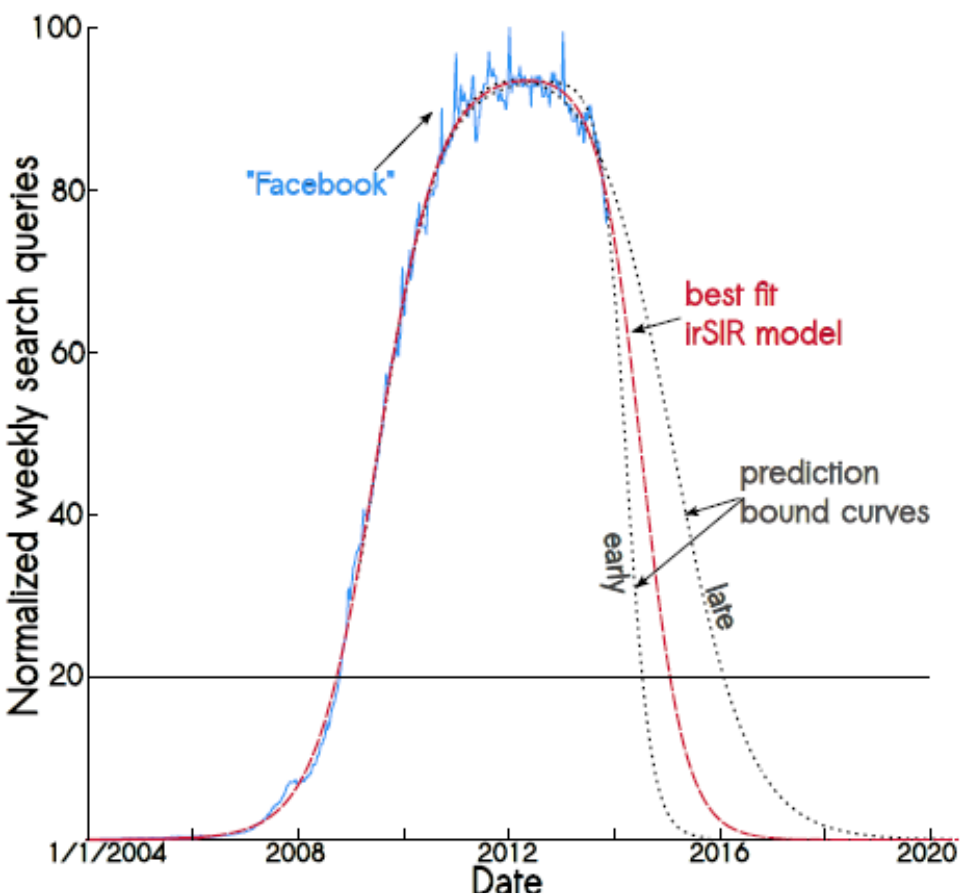
Salvo que $r_0=0$, esa es la dinámica habitual. Se espera que, **tarde o temprano, las RSOs siempre acaben perdiendo usuarios**

CASO DE ESTUDIO

Modelado de epidemias en redes sociales on-line (4)

En los experimentos realizados, se ajustan los parámetros de los modelos SIR e irSIR con datos de Google Trends conocidos para MySpace.

El ajuste del modelo irSIR es mucho mejor



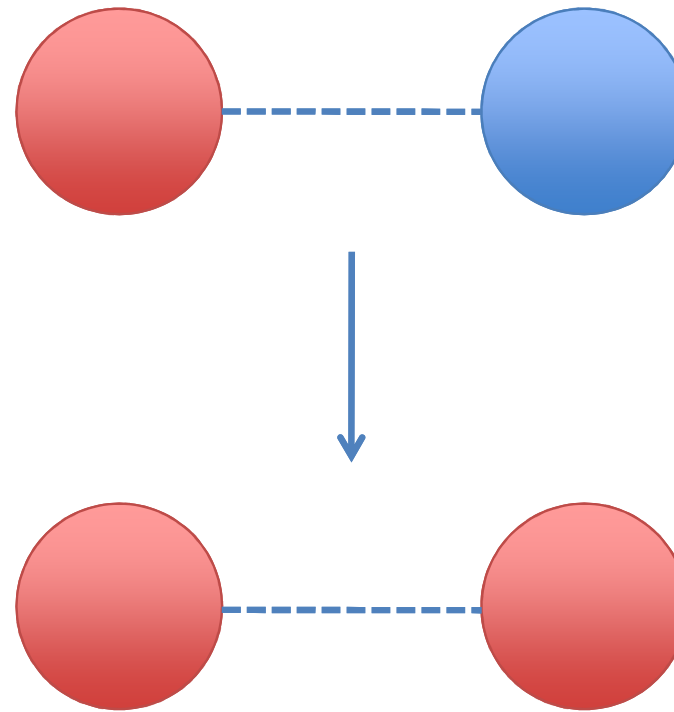
Para Facebook, se ajusta el modelo irSIR con datos 2004-2013 y se predice el comportamiento futuro

Se muestra también un intervalo de confianza con los tiempos de declive más cercano y más lejano

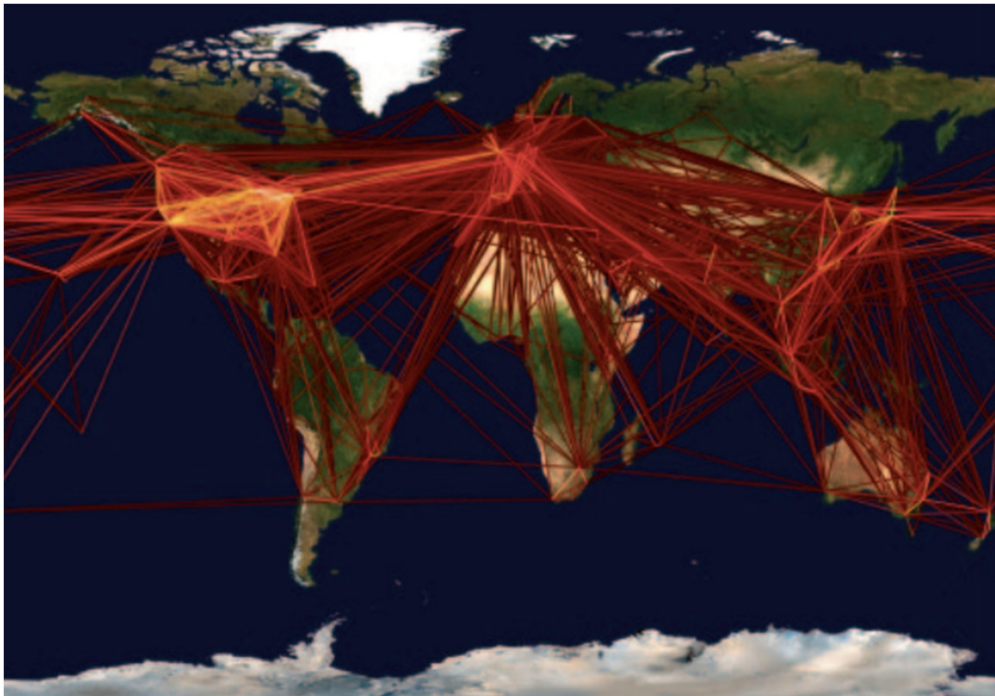
MODELOS DE PROPAGACIÓN DE EPIDEMIAS BASADOS EN REDES

¡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**

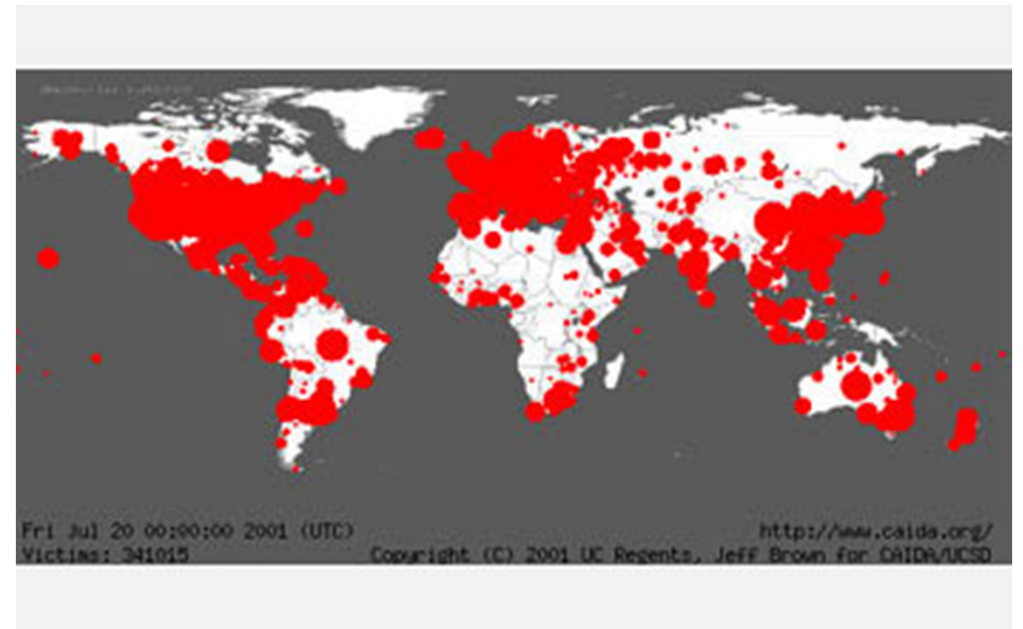


REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (2)



La red mundial de transporte

L. Hufnagel et al. Forecast and control of epidemics in a globalized world. PNAS 101 (2004) 15124- 15129



Internet

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

REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (3): 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 (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 influye 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 un sistema complejo con una topología de red concreta?

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

- Contagio simple vs. contagio complejo: En cada unidad de tiempo,
 - contagio simple: Cada “amigo” (nodo conectado a ti) infectado te infecta con una cierta probabilidad
 - contagio complejo: sólo realizas una acción si una cantidad/porcentaje de tus “amigos” lo hacen
- Coordinación y aprendizaje
- Resistencia de la red ante caídas o ataques (redes eléctricas, informáticas, etc.)

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 en lugar de asumir que los contactos son posibles entre todos los individuos

Dado un **ratio de infección β** (una probabilidad de contagio, dependiente de la enfermedad concreta) es posible definir modelos para su propagación sobre una red. **Todas las variantes del modelo SIR pueden generalizarse usando redes complejas**

En cada unidad de tiempo, los nodos infectados propagan la enfermedad a sus vecinos susceptibles con probabilidad β , difundiendo la enfermedad por la red

Al contrario que en los modelos clásicos, **no es fácil resolver analíticamente modelos de este tipo para redes complejas generales**

MODELO SIR EN REDES COMPLEJAS: FUNDAMENTOS (2)

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 que la basada en considerar las ecuaciones del sistema. No se consideran globalmente los ratios de cada estado sino que se trabaja **a nivel de cada individuo (agente)**

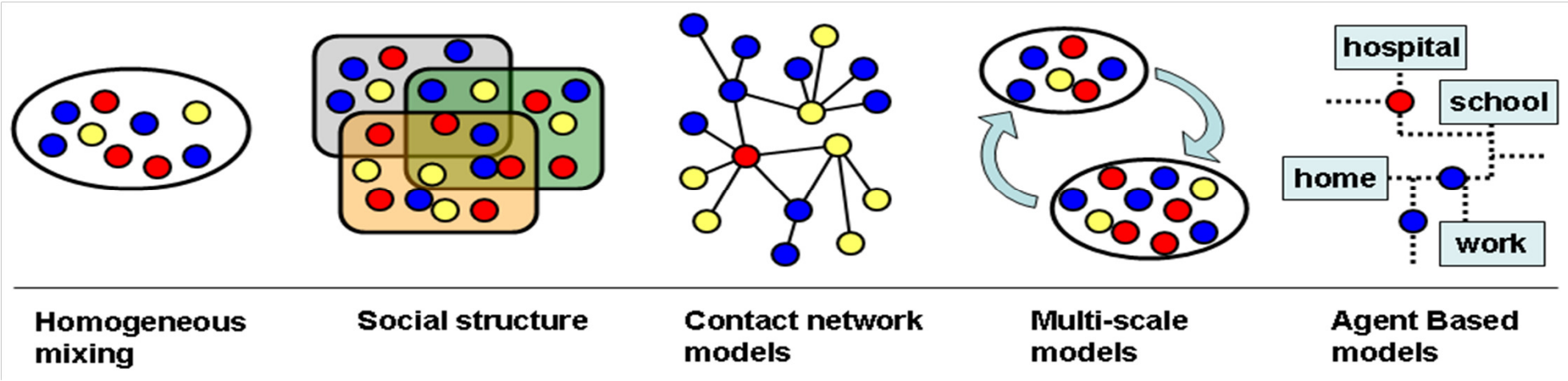
Se conocen sus **propiedades individuales**, que pueden variar de unos a otros, y se consideran las **interacciones locales** (*comportamiento emergente*)

Se modela mejor la realidad usando procesos estocásticos de actualización de estados que simulan los eventos aleatorios de los procesos dinámicos

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 (3)

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

MODELO SIR EN REDES COMPLEJAS: APLICABILIDAD

- **Modelos epidémicos y modelos de propagación de virus informáticos:** Susceptibles, Infectados y Recobrados

Pastor-Satorras y Vespignani. Epidemic spreading in scale-free networks. Physical Review Letters 86 (2001) 3200–3203

El modelo SIR es muy apropiado para las fases iniciales de un ataque de un virus de ordenador:

Pastor-Satorras y Vespignani. Evolution and structure of the Internet: A statistical physics approach. Cambridge University Press. 2004

Tabah. Literature dynamics: Studies on growth, diffusion, and epidemics. Annual Review of Information Science & Technology 34 (1999) 249-286

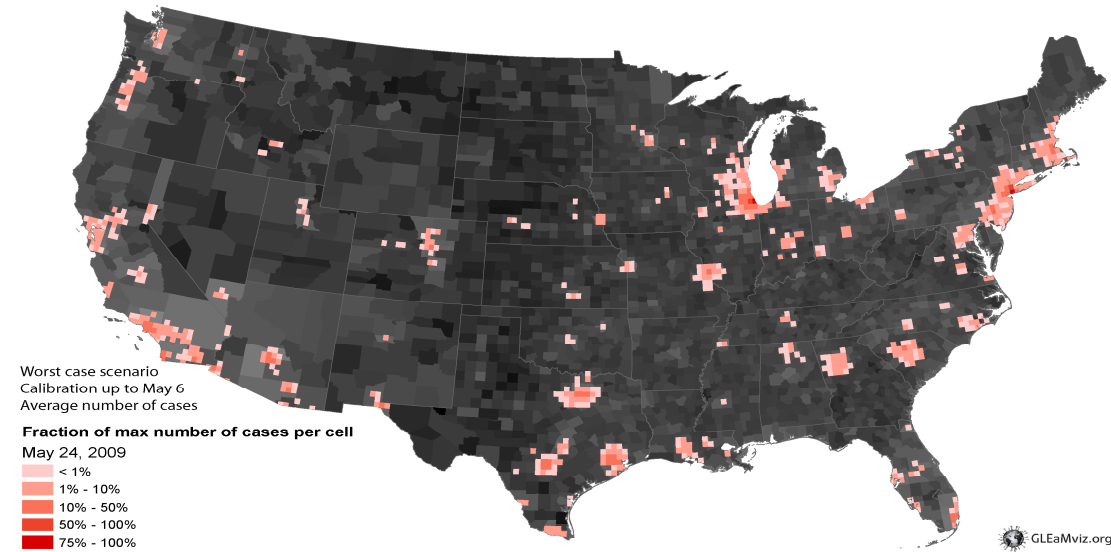
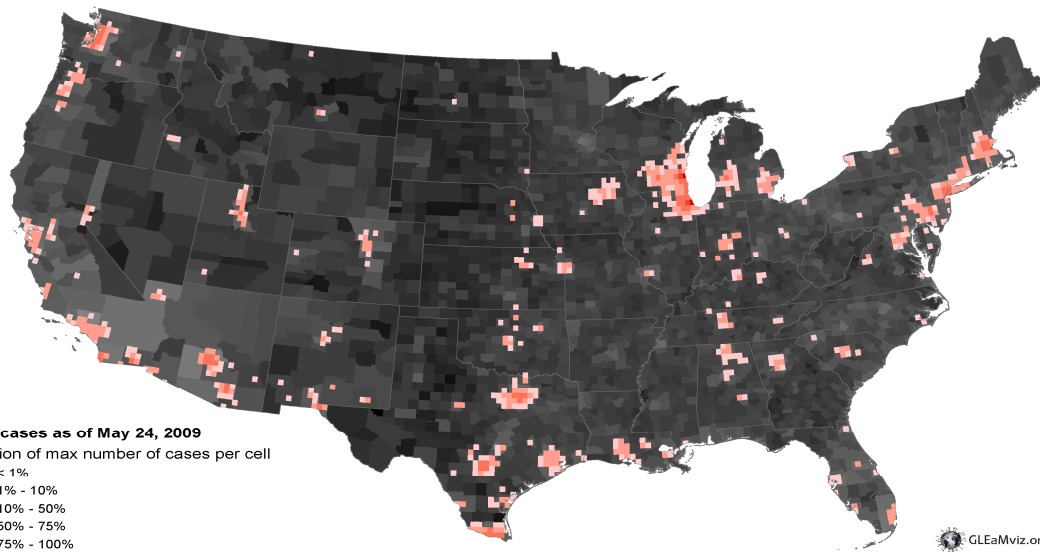
- **Modelos de propagación de rumores:** Ignorantes, Difusores y Represores

Daley, Gani y Cannings. Epidemic modeling: An introduction. Cambridge University Press. 1999

- **Modelos de difusión de conocimiento:** Innovadores, Incubadores y Adoptadores

Real

Pronosticada



Primera pandemia analizada con redes complejas y ABM

<http://www.ladamic.com/netlearn/NetLogo501/ERDiffusion.html>

setup a new network

setup with current topology

spread once spread

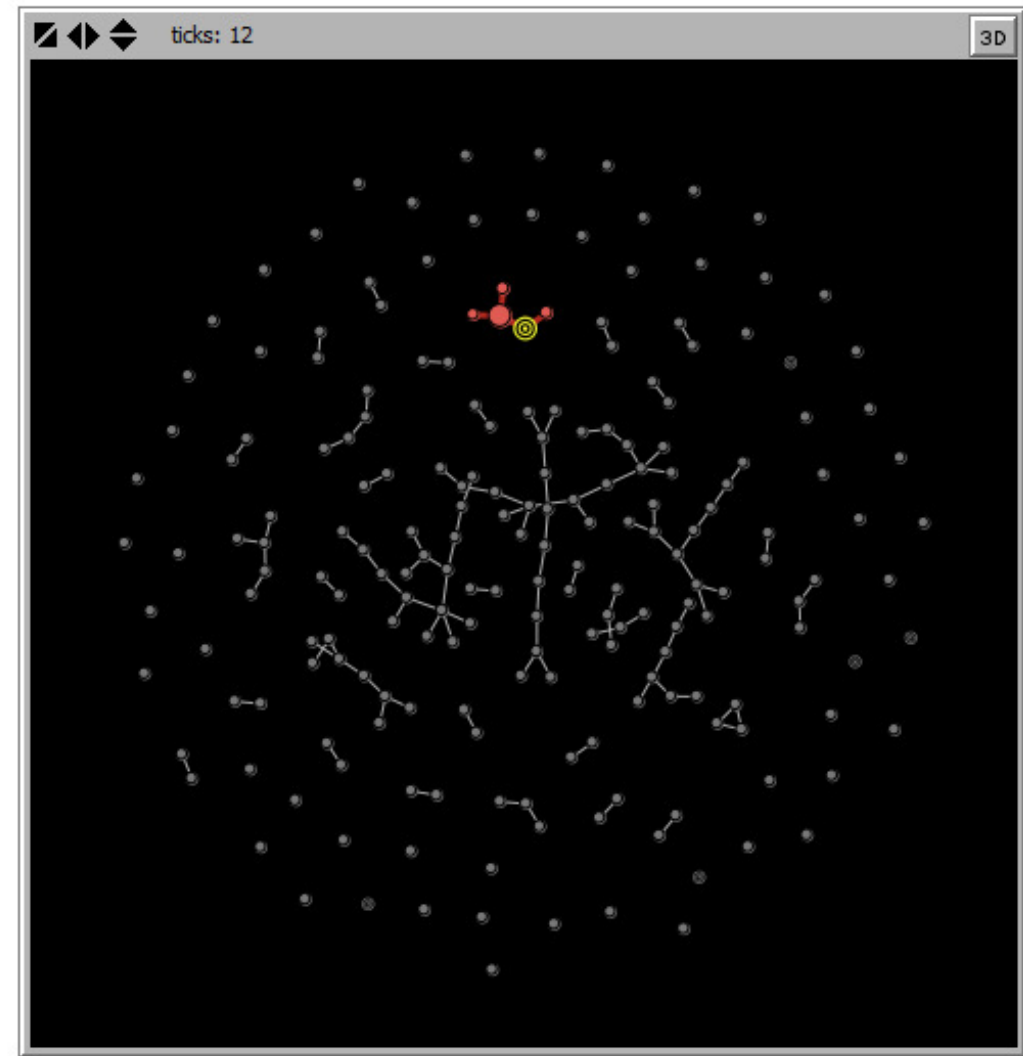
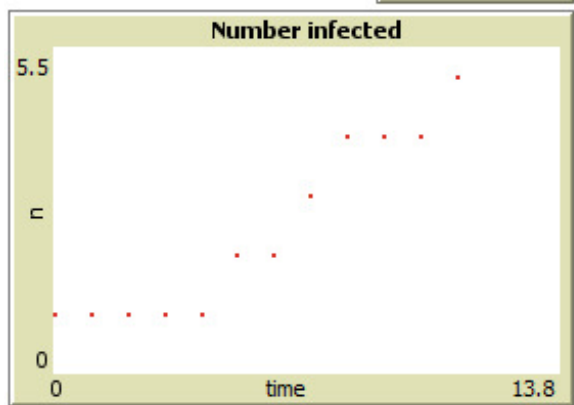
toggle infection tree layout

num-nodes 200

avg-degree 1.05

infect-rate 0.15

num-infected 5



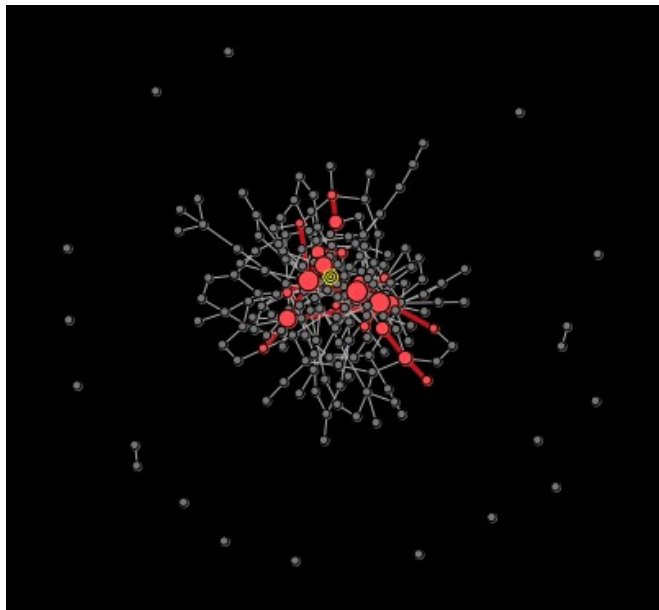
- **A diferencia del modelo clásico**, si partimos de un único nodo infectado, no todos los nodos se infectan, **sólo los que pertenecen a la misma componente conexa**
- Como la mayoría de las redes reales tienen una componente gigante y muchas componentes conexas pequeñas, **la epidemia se extiende en mayor o menor medida dependiendo de la localización del nodo inicial infectado**
- Si se escoge aleatoriamente, **la probabilidad de que pertenezca a la componente gigante y se produzca una pandemia en la población es $S=N_G/N$**
- Analizando las componentes pequeñas, se puede conocer la probabilidad de las pequeñas epidemias

INFLUENCIA DE LA CONECTIVIDAD Y LA DENSIDAD

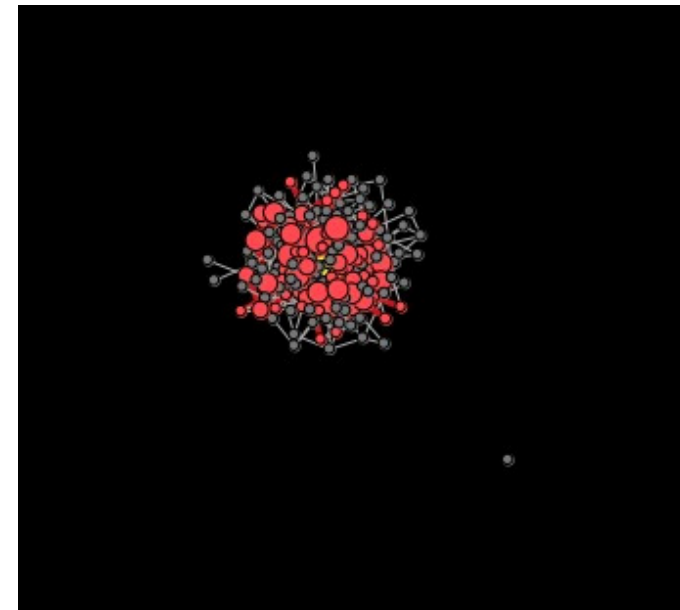
Con los mismos parámetros del modelo, ¿la difusión depende de la topología de la red!

Nodos infectados después de 10 pasos, ratio de infección = 0.15

grado medio $\langle k \rangle = 2.5$



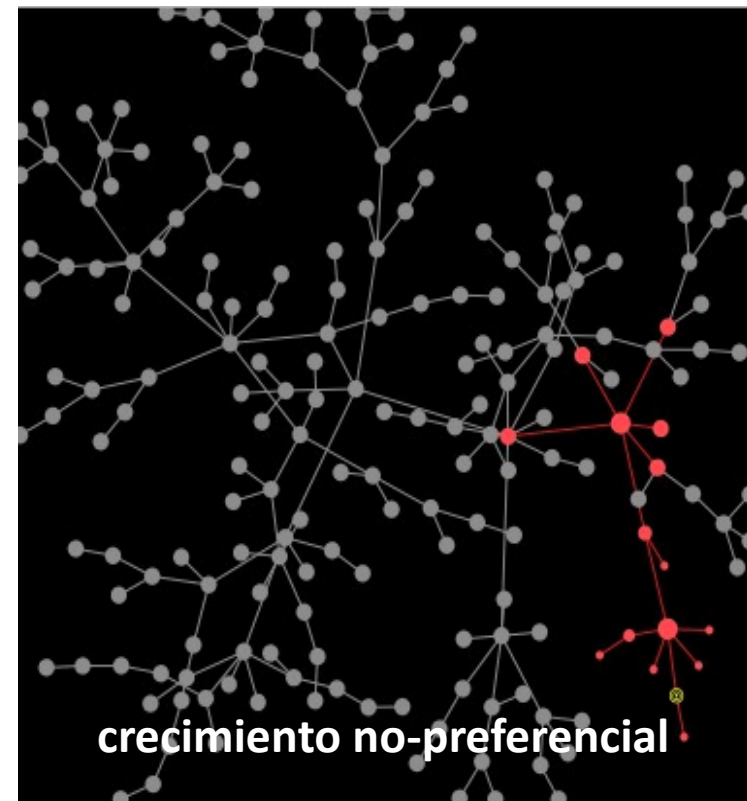
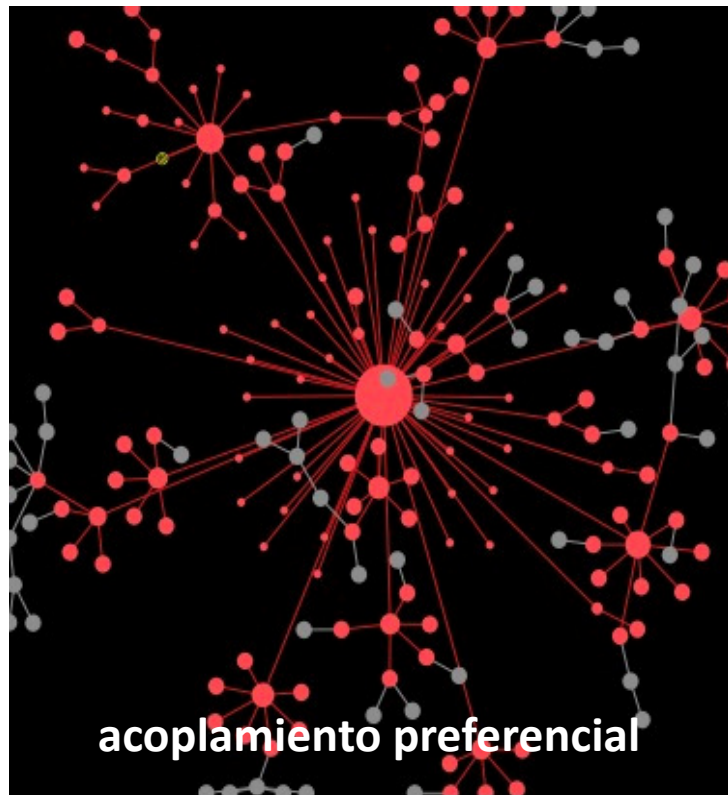
grado medio $\langle k \rangle = 10$



Quando aumenta la densidad de la red, la propagación es más rápida

<http://www.ladamic.com/netlearn/NetLogo501/BADiffusion.html>

Ratio de transmisión $\beta=1$. Nodos infectados tras cuatro unidades de tiempo:



Cuando existen hubs en la red, la velocidad de propagación es más rápida

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

The image shows a complex user interface for a simulation. At the top left, there are sliders for 'number of nodes' (set to 100), 'average node degree' (set to 6), and 'initial outbreak size' (set to 3). Below these are 'setup' and 'go' buttons. Further down are sliders for 'virus spread chance' (100%), 'virus check frequency' (1), 'recovery chance' (1), and 'gain resistance chance' (100%). A 'Network Status' graph at the bottom left shows the number of nodes over time for three categories: susceptible (green), infected (red), and resistant (blue). The central part of the interface features two network graphs. The top graph shows a network with many green nodes and a few blue nodes. The bottom graph shows a similar network but with more blue nodes. A large 'CLICK TO RUN' button with a red arrow is overlaid on the bottom graph. On the right side, there is a graph with a y-axis from 0 to 60 and an x-axis from 0 to 46.3. Below it are 'On mouse' and 'Off mouse' buttons, a 'go' button, and sliders for 'number of isolated agents' (set to 100), 'wholistic degree' (set to 1), and 'dist' (set to 10).

- **S**→**I**: Cada agente infectado puede infectar un agente susceptible conectado a él de acuerdo a la probabilidad (ratio) de infección β
- **I**→**R|S**: Cada agente infectado puede recuperarse con una probabilidad de recuperación μ . En ese caso, puede pasar a dos estados distintos:
 - **I**→**R**: El agente se vuelve inmune con probabilidad de inmunidad ι
 - **I**→**S**: El agente no se vuelve inmune y pasa de nuevo a susceptible

DIFUSIÓN EN REDES

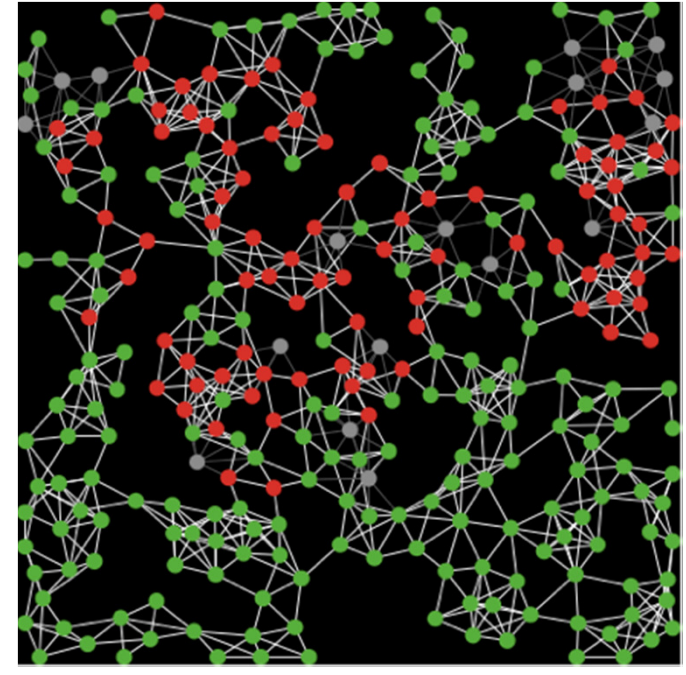
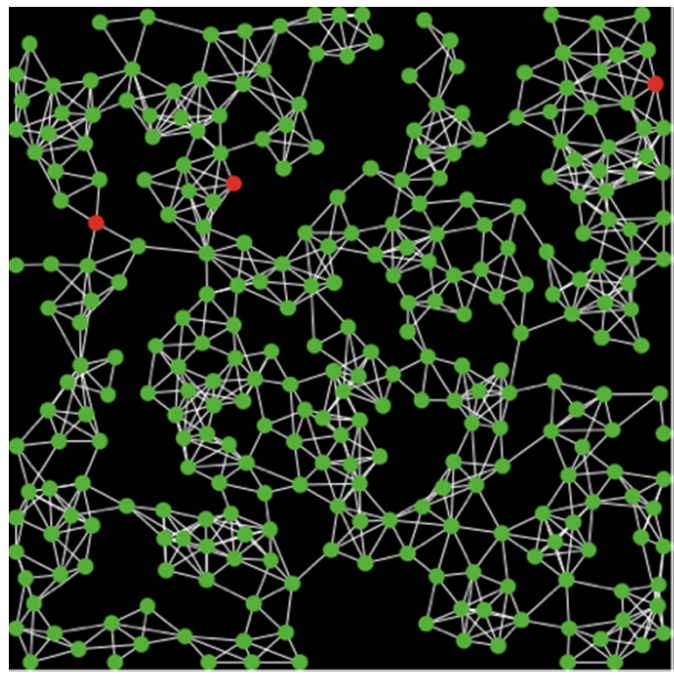
Contagio simple (Modelo SIRS) - Redes aleatorias (3)

Ejemplo:

$$\beta=0.025, \mu=0.05, \tau=0.025$$

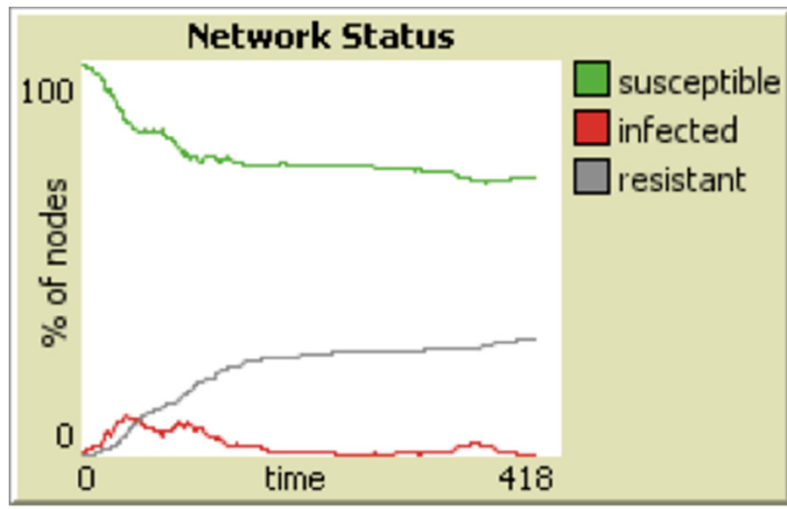
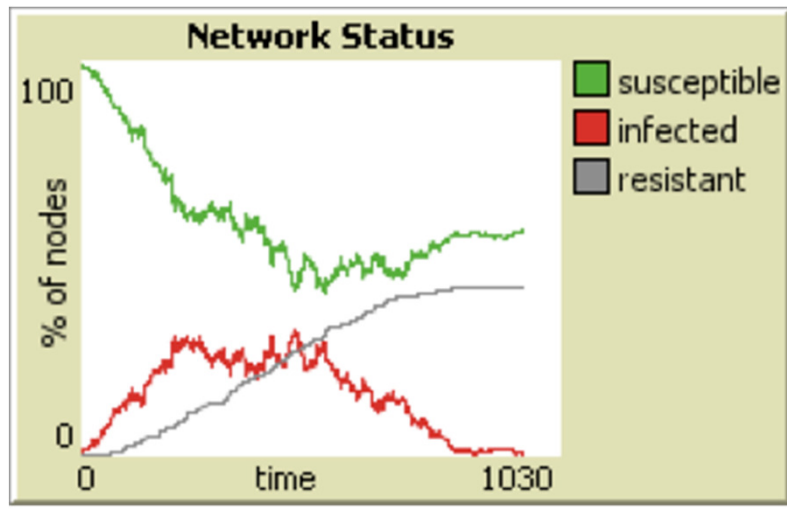
3 individuos infectados en la población inicial

214 pasos de simulación



Estudio del umbral epidemiológico:

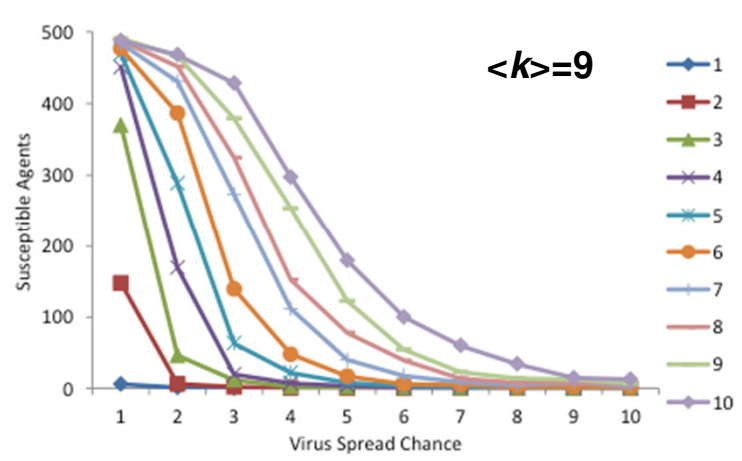
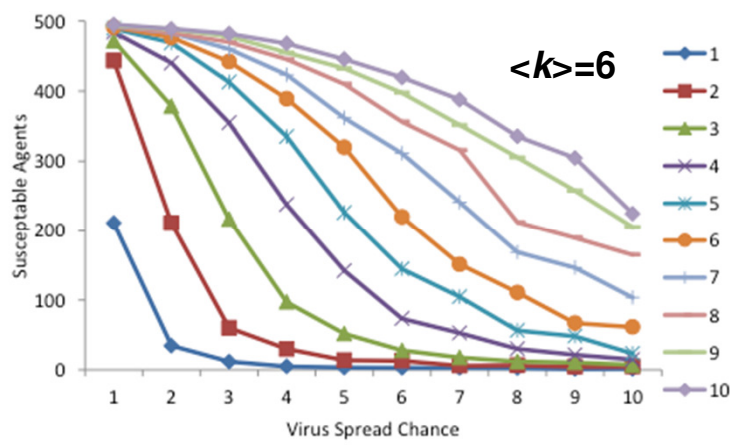
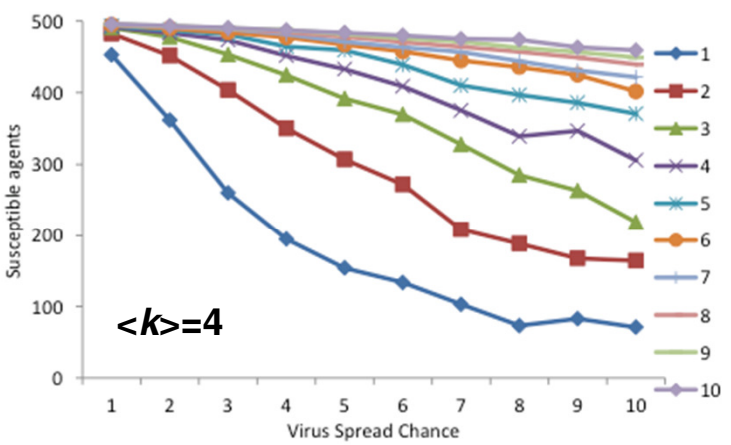
$$\beta=0.025, \mu=\{0.05, 0.5\}$$



DIFUSIÓN EN REDES

Contagio simple (Modelo SIR) - Redes aleatorias

- Estudio de la influencia de la topología de la red (grado medio $\langle k \rangle$) y de los valores de las probabilidades de infección β y recuperación μ
- Modelo SIR: Los agentes recuperados, no pueden volver a ser susceptibles (modelo SIRS anterior con $\tau=0$)
- Tres redes ER con $\langle k \rangle = \{4, 6, 9\}$. $\beta = \mu = \{0.01, 0.02, \dots, 0.1\}$. 500 agentes. 100 simulaciones con 1000 pasos (**Montecarlo**)



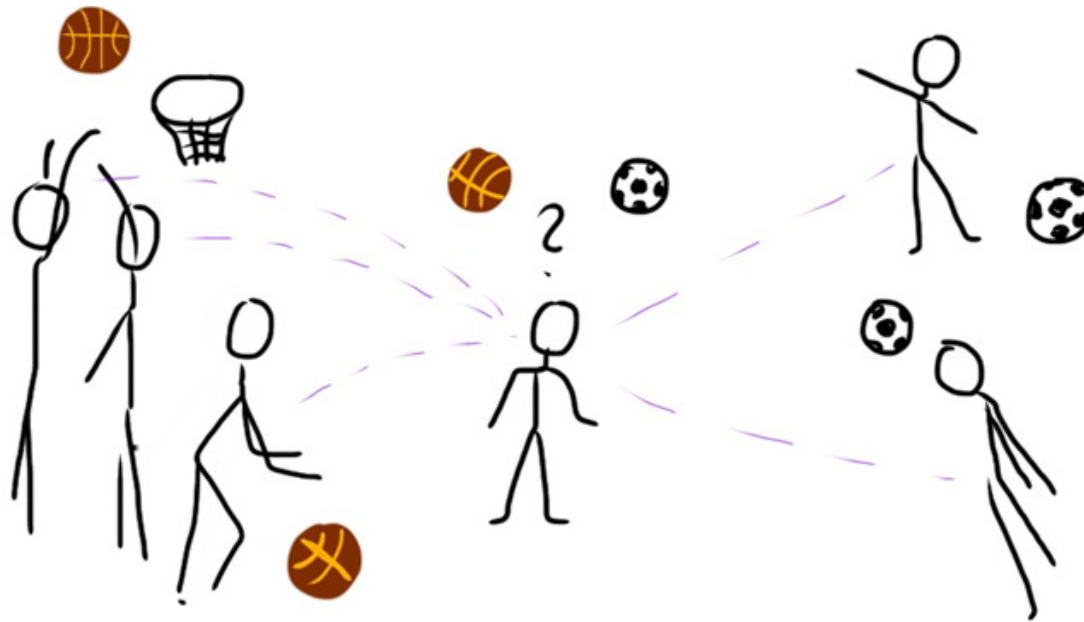
- Claramente, **a mayor grado medio, mayor facilidad para la difusión del virus**

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

- Cada individuo debe escoger entre dos opciones, A y B (p.ej. futbol o baloncesto)
- Si tus amigos escogen A , el individuo consigue una recompensa a . Si escogen B , la recompensa vale b

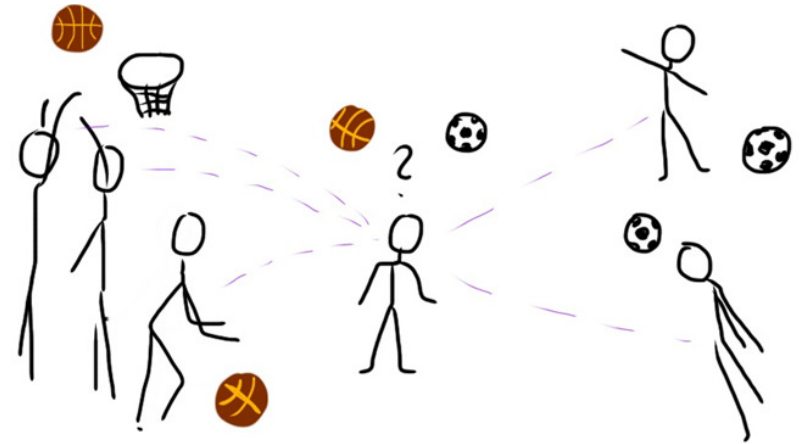
Un ratio $p = 3/5$
juegan a
baloncesto



Un ratio $p = 2/5$
juegan a fútbol

¿Qué elección proporciona una mayor recompensa?

- Un actor tiene d vecinos
- Un ratio p juegan a baloncesto (A)
- Un ratio $1-p$ juegan a fútbol (B)
- Si escoge A , recibe $p \cdot d \cdot a$
- Si escoge B , recibe $(1-p) \cdot d \cdot b$
- Por tanto, escogerá A si: $p \cdot d \cdot a \geq (1-p) \cdot d \cdot b$ o $p \geq b/(a+b)$ (**umbral**)



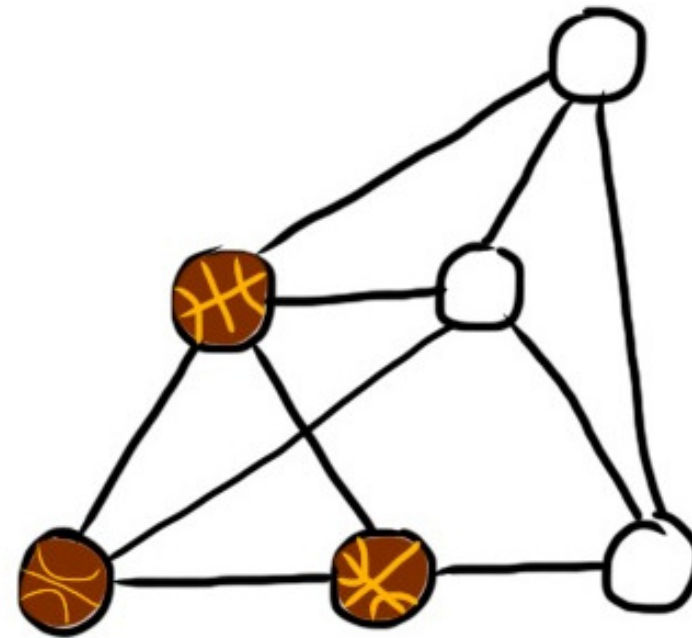
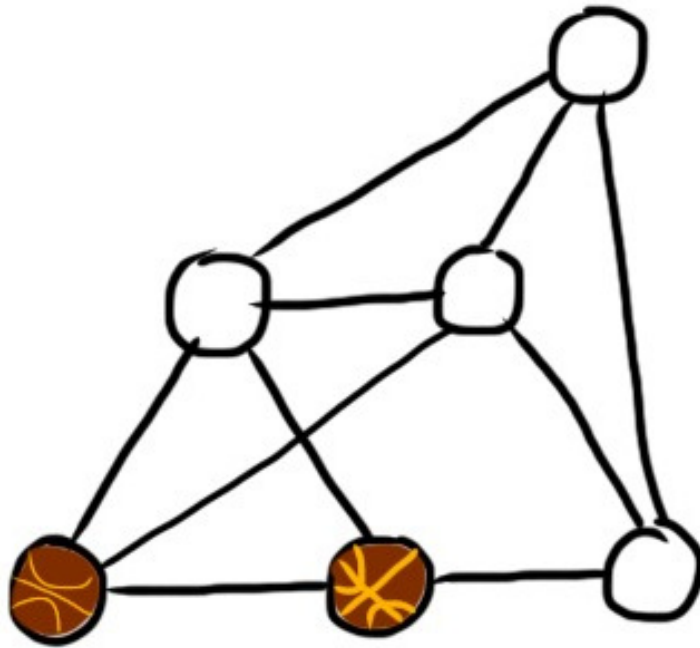
El sistema tiene dos estados de equilibrio posibles: todos adoptan A o B

¿Qué ocurre en el proceso intermedio?

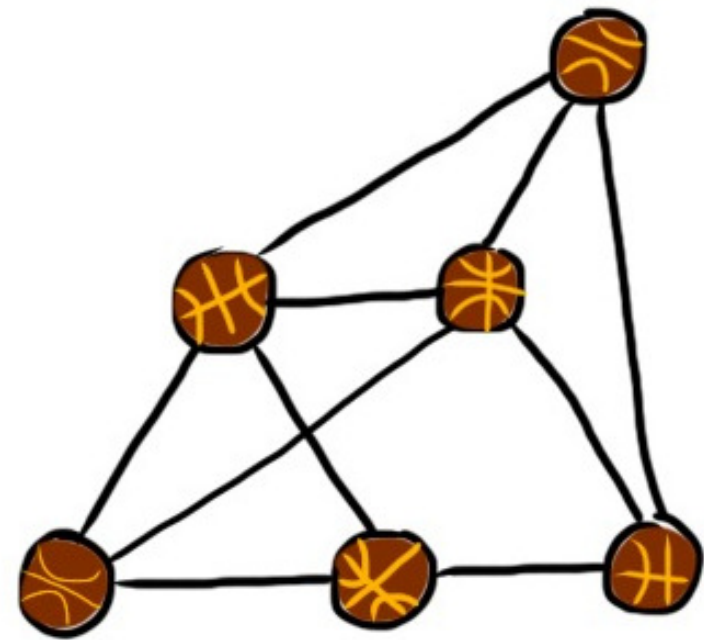
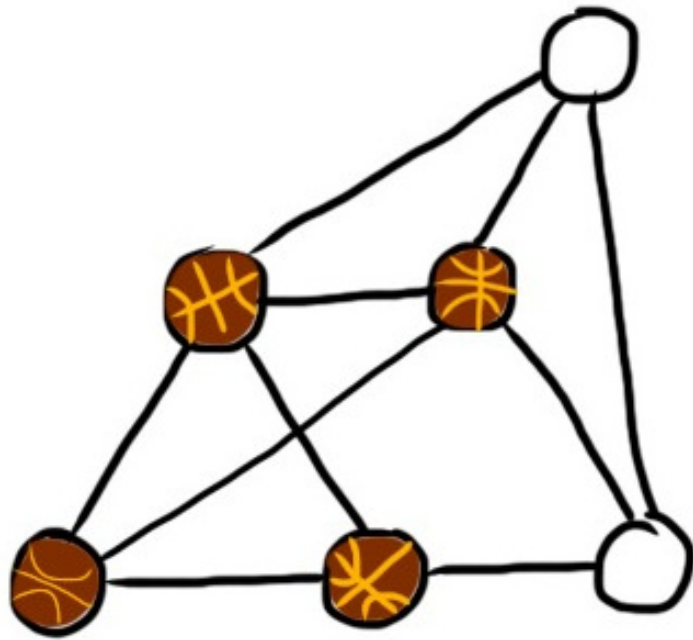
- ¿Qué pasa si dos nodos cambian su estado de forma aleatoria? ¿Se producirá una **propagación en cascada**?
- Ejemplo:
 - $a=3, b=2$
 - La recompensa para la interacción de dos nodos con comportamiento A es $3/2$ veces mayor que la que obtendrían si ambos escogieran B
 - Los nodos cambiarán de B a A si al menos $q = 2/(3+2) = 2/5$ de sus vecinos están haciendo A (**umbral de cambio**)

¿Cómo se produce una propagación en cascada?

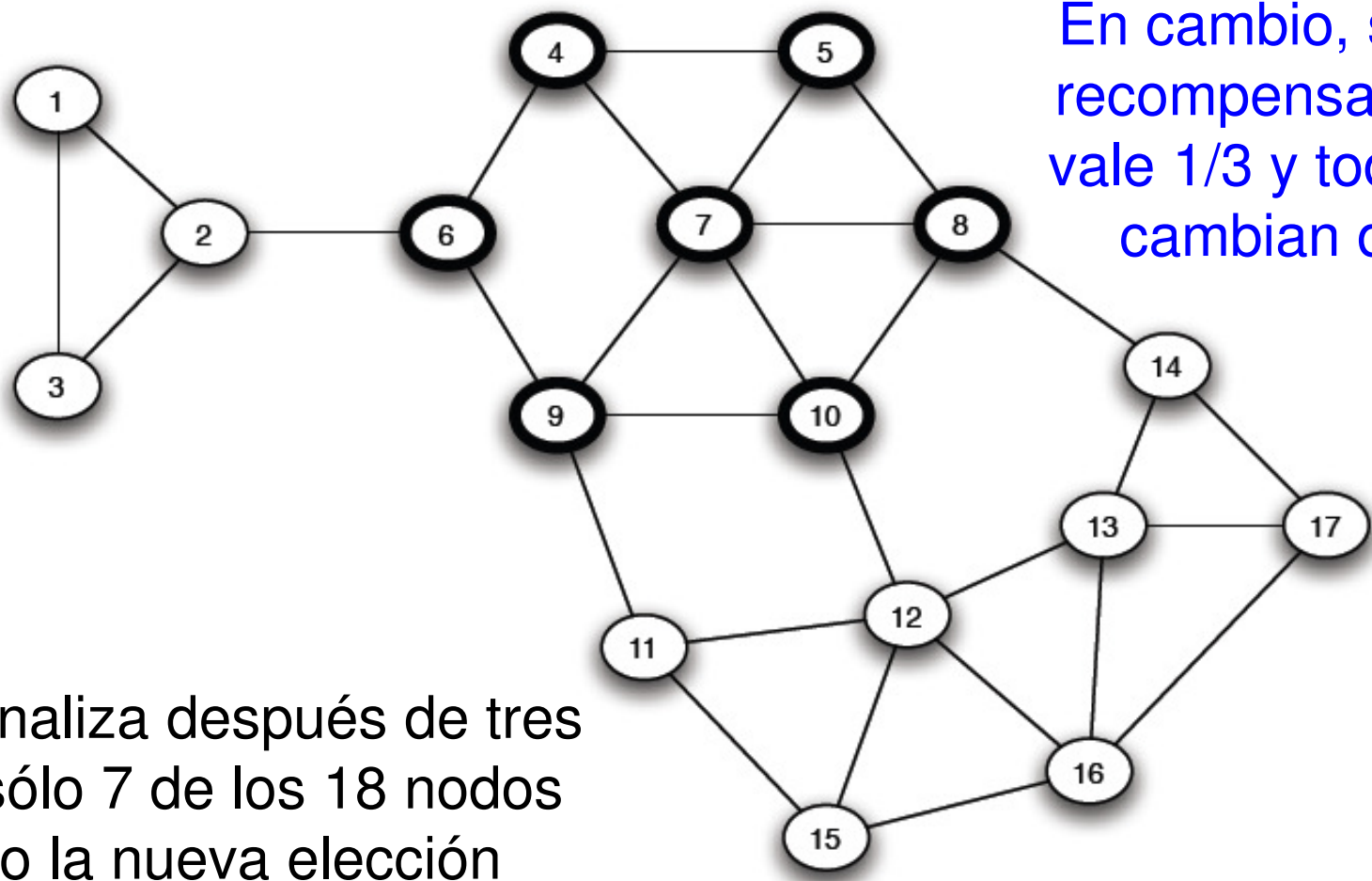
- Supongamos que dos nodos empiezan a jugar baloncesto debido a factores externos (p.ej. son “premiados” por alguna empresa con un par de zapatillas de basket)



¿Cuál será el siguiente nodo que se cambie al baloncesto?



Otro ejemplo con los mismos parámetros en el que no se converge a una única elección



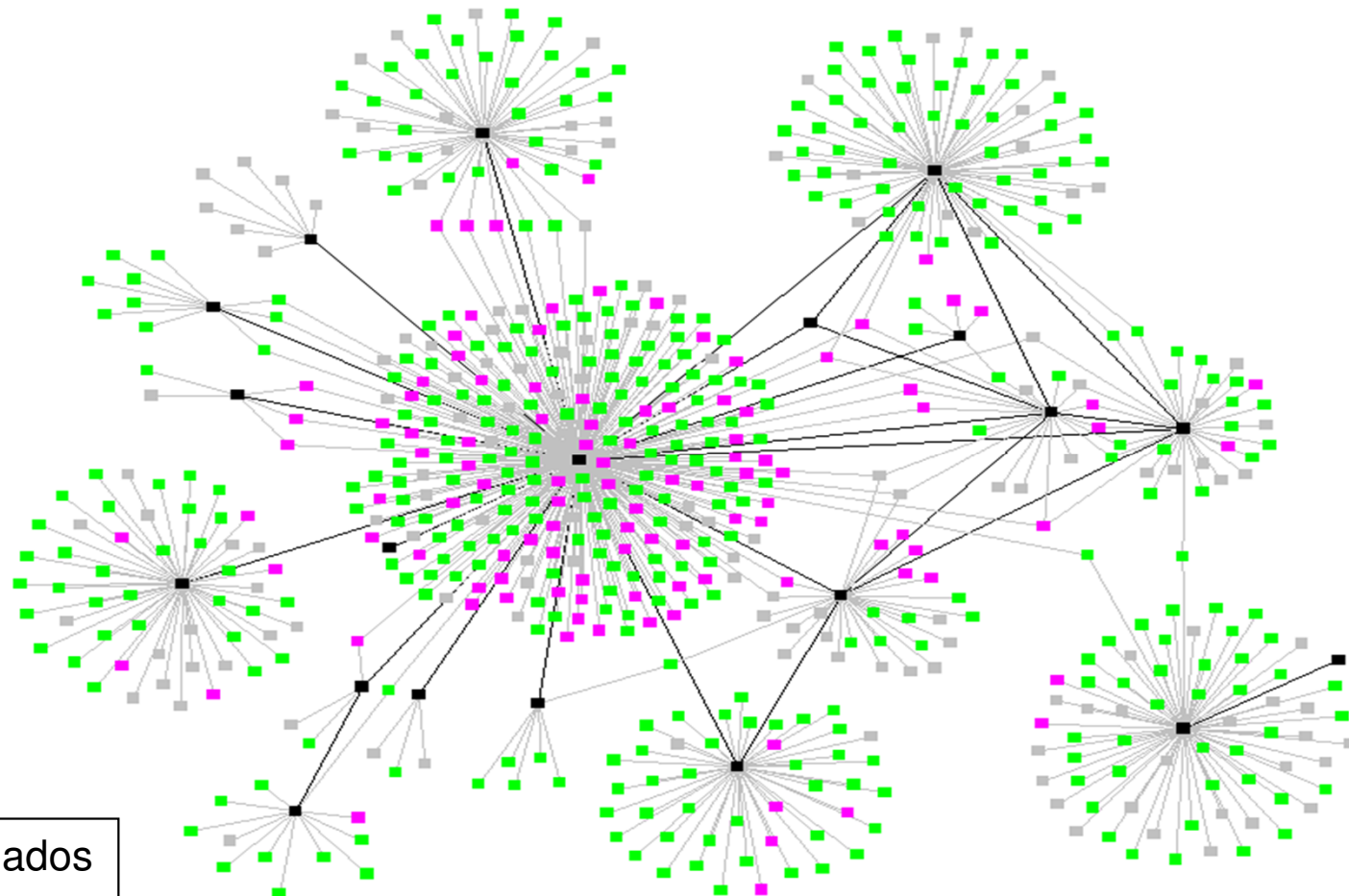
En cambio, si aumenta la recompensa a de 3 a 4, q vale $1/3$ y todos los nodos cambian de elección

El proceso finaliza después de tres pasos con sólo 7 de los 18 nodos adoptando la nueva elección

- Cuando no se puede aumentar la recompensa (calidad/precio del producto), otra estrategia es convencer a un número pequeño de individuos para usar y promocionar el producto

Hubs:

Difunden virus o ideas débilmente infecciosas

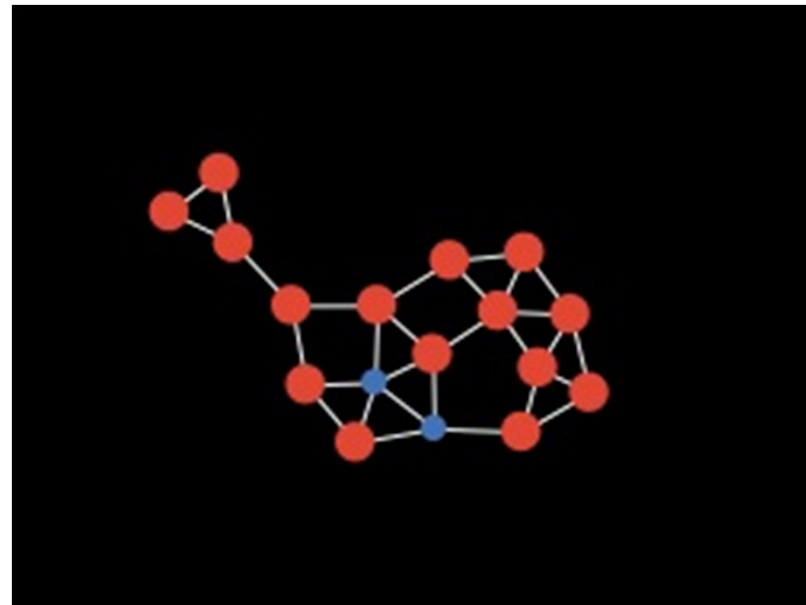


negro: líderes de opinión; rojo: influenciados
verde: no influenciados; gris: indecisos

- **La elección de los individuos es clave para producir el efecto cascada (actores influyentes de la red – *influentials*)**. En el caso anterior, partiendo de los nodos 12 y 13 se consigue convencer del 11 al 17 pero partir de 11 y 15 no produce ninguna “conversión” adicional

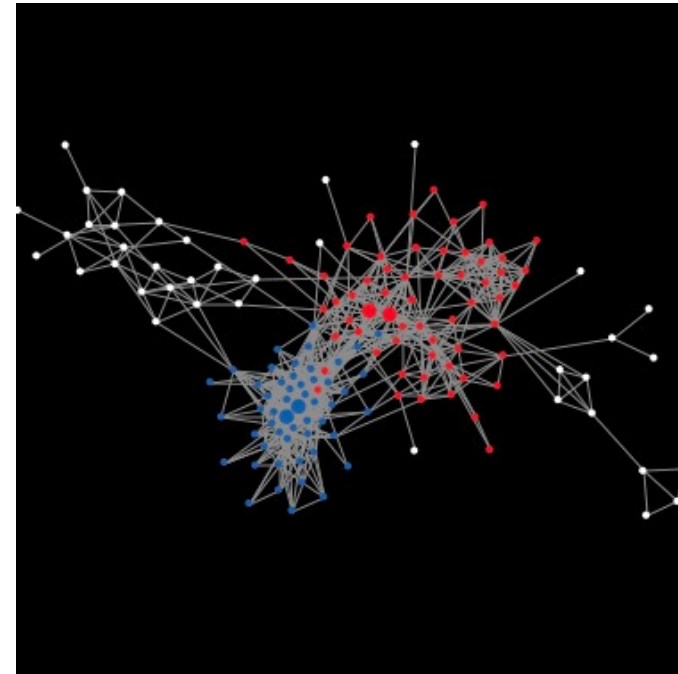
<http://www.ladamic.com/netlearn/NetLogo412/CascadeModel.htm>

¿Se producirá una propagación en cascada?



<http://www.ladamic.com/netlearn/NetLogo412/CascadeModel.html>

- **El proceso de difusión depende de la estructura de la red**
- El modelo permite usar cuatro distintas. Una de ellas en la red de Facebook de *Lada Adamic*
- Se puede usar como un juego de dos jugadores. Cada persona tiene que escoger dos nodos:
 - El primero escoge un nodo y lo marca como azul
 - El segundo escoge otro nodo y lo marca como rojo
 - El primero escoge un nodo azul adicional
 - El segundo escoge un nodo rojo adicional
 - Se ejecuta el modelo con los parámetros prefijados



MODELOS DE COMPORTAMIENTO DE CONSUMIDOR CON ABMs Y REDES SOCIALES

MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (1)

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

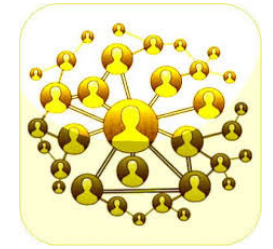
Los modelos más realistas integran reglas de decisión basadas en estas **necesidades individuales** con procesos de difusión (*Word of Mouth, WoM*) (**necesidades sociales**)

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



MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (2)

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)

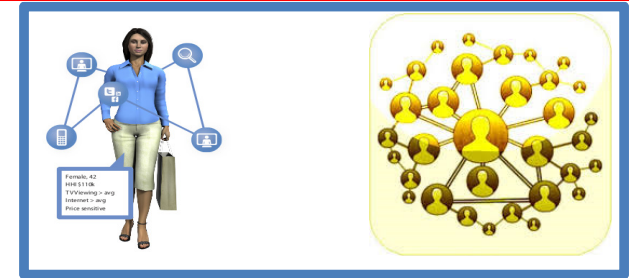
La utilidad individual se modela como la diferencia entre la preferencia del consumidor i , expresada como $p_i \in [0,1]$, y el valor de las características del producto, agrupadas en una única variable $\delta_j \in [0,1]$

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

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

MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (3)

Utilidad de producto (2):



La parte de **efecto social** de la utilidad del producto aumenta con la cantidad de amigos que consumen el producto (**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. Para cada agente se calcula el ratio de cada producto consumido por los vecinos

La utilidad total esperada de consumo del producto j por 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)**

MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (4)

Conectividad:

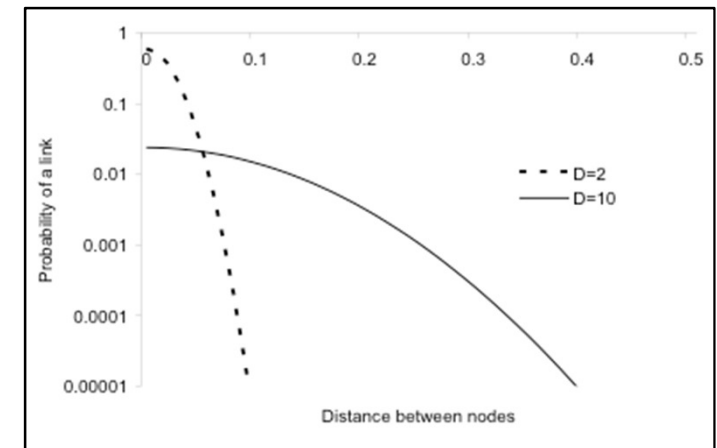
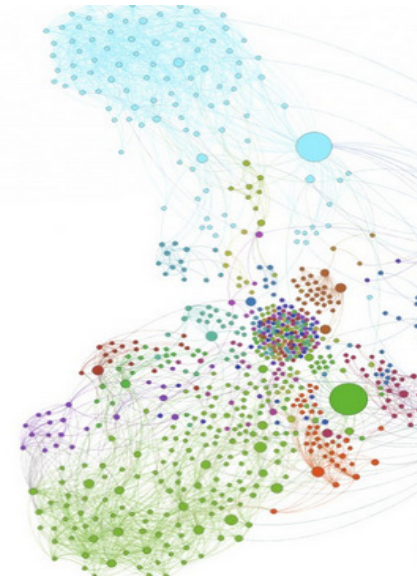
La red social se modela como una red aleatoria, que tiene en cuenta la posición espacial de los agentes

La probabilidad de conexión entre dos agentes disminuye con la distancia existente entre ellos:

$$p = \frac{n}{2\pi D^2} e^{-Nd^2/2D^2}$$

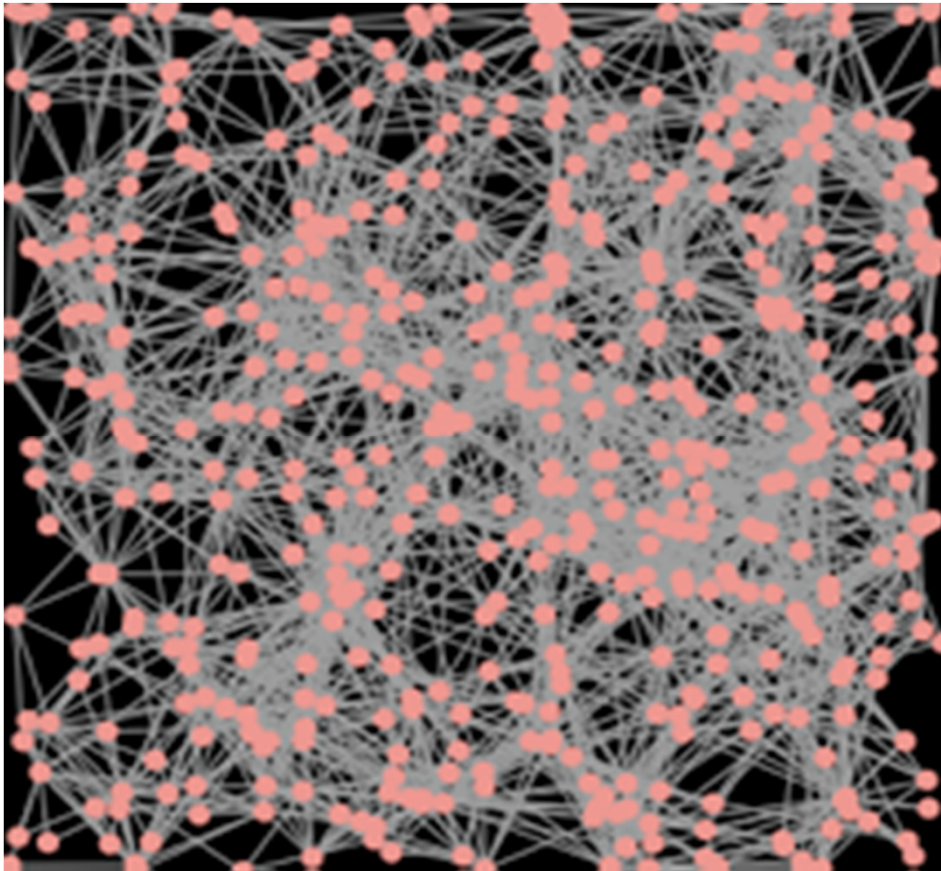
- D es la distancia del enlace “ideal”
- n es el número promedio de enlaces
- d es la densidad promedio
- N es el número de agentes

Si D es pequeño (p.ej. 2) las conexiones son preferentemente locales, entre individuos cercanos. Con D grande (p.ej. 10), aumentan las conexiones lejanas

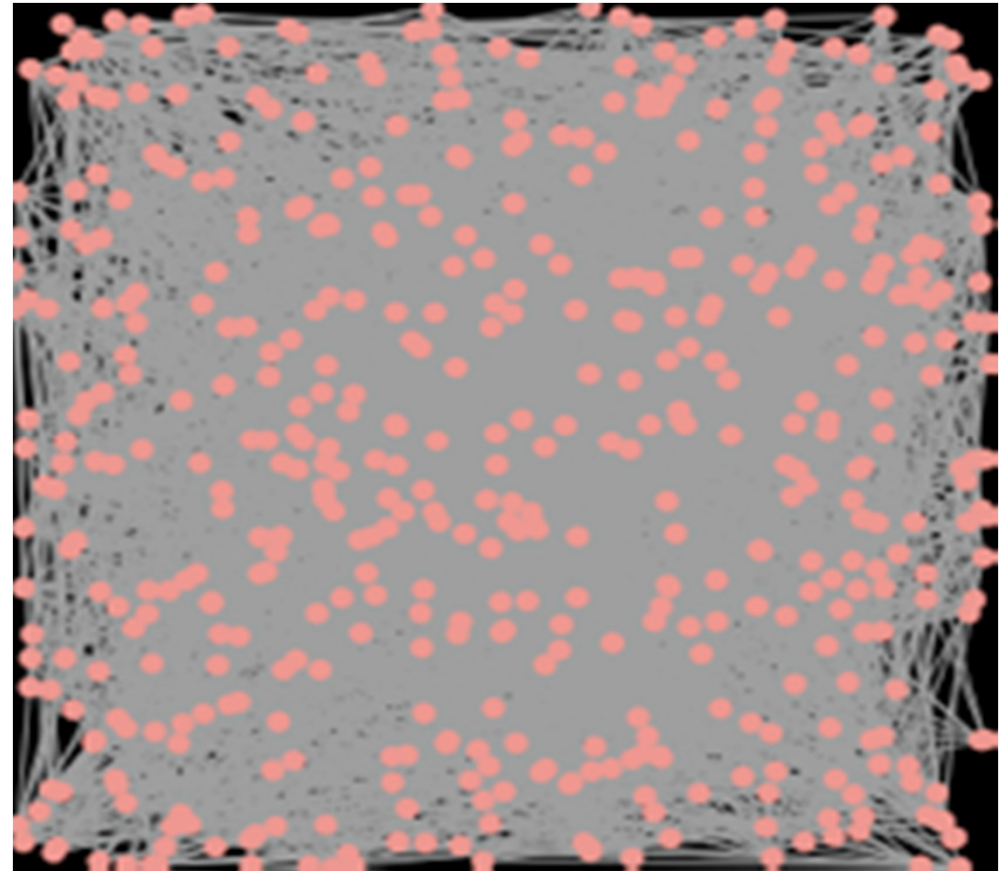


MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (5)

(a) Red social con $D=2$



(b) Red social con $D=10$



MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (6)

Proceso de compra:

Cada unidad de tiempo (p.ej. día), el consumidor decide si quiere comprar probabilísticamente y qué comprar de forma determinística

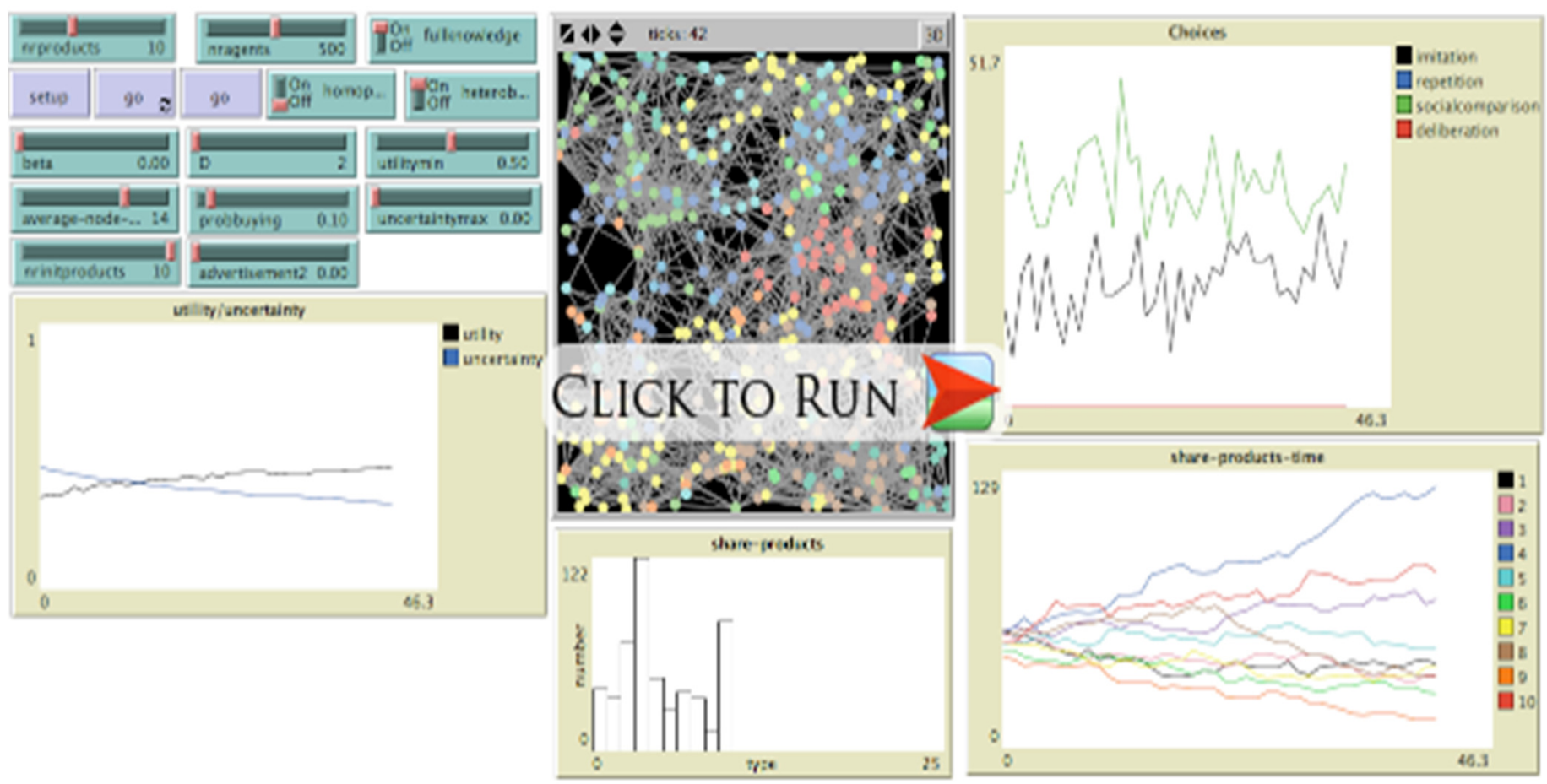
La probabilidad de comprar es un parámetro determinado p_b

En caso de comprar, el cliente mantiene su decisión anterior siempre que la utilidad del producto sea mayor que un umbral mínimo u_{\min} (**modelando así una toma de decisiones conservadora**). Al comienzo de la simulación se asigna un producto aleatorio a cada agente

Si el agente no está satisfecho, calcula la utilidad esperada para los N productos y adquiere el que maximiza dicho valor

MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (7)

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



MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (8)

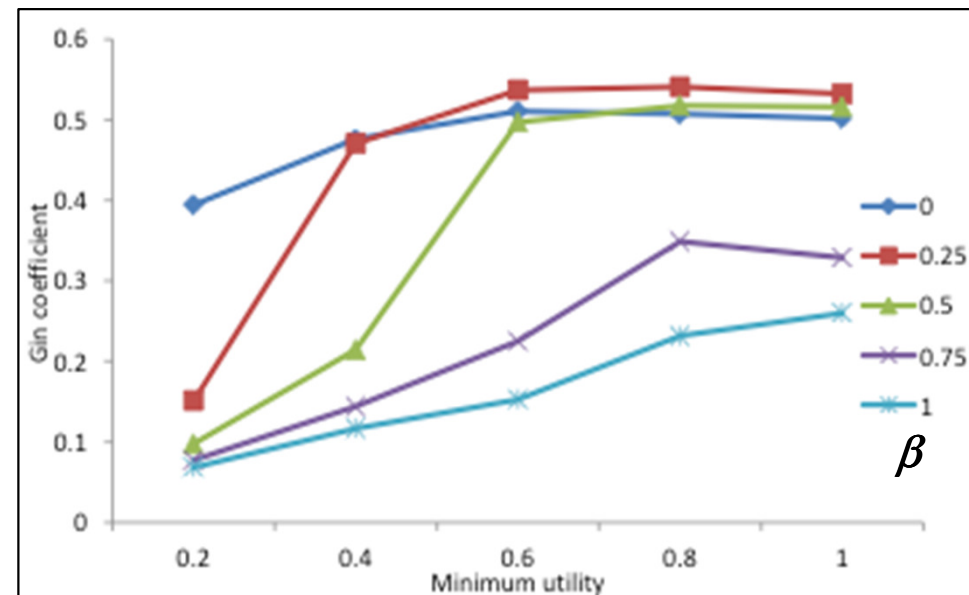
Algunas simulaciones:

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; probabilidad de comprar $p_b=0.1$; valores variables para peso social β y umbral mínimo u_{\min} , comunes para toda la población; 100 simulaciones (**Montecarlo**)

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 crece por encima de 0.5 (agentes más sociales y más exigentes)



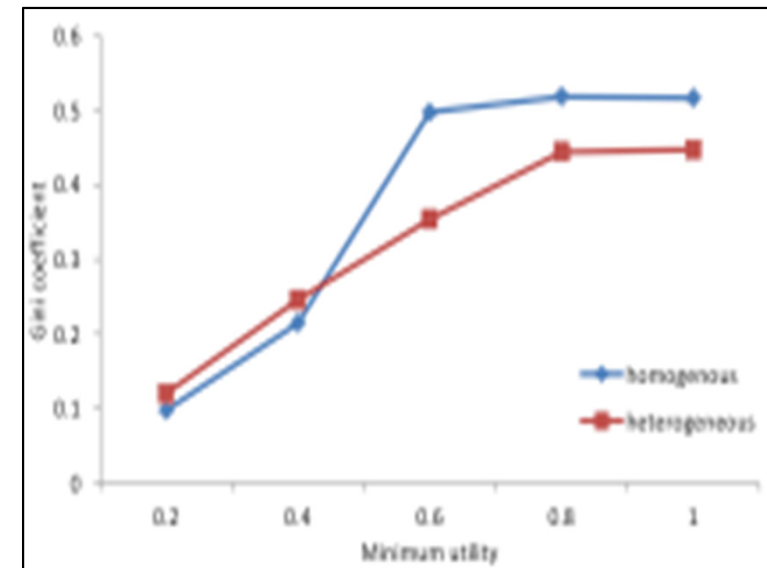
MODELADO Y SIMULACIÓN DE COMPORTAMIENTO DEL CONSUMIDOR (9)

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.

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.
- ...

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 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)

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 (3)

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 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**)

APRENDIZAJE EVOLUTIVO DE ESTRATEGIAS DE MARKETING VIRAL (4)

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

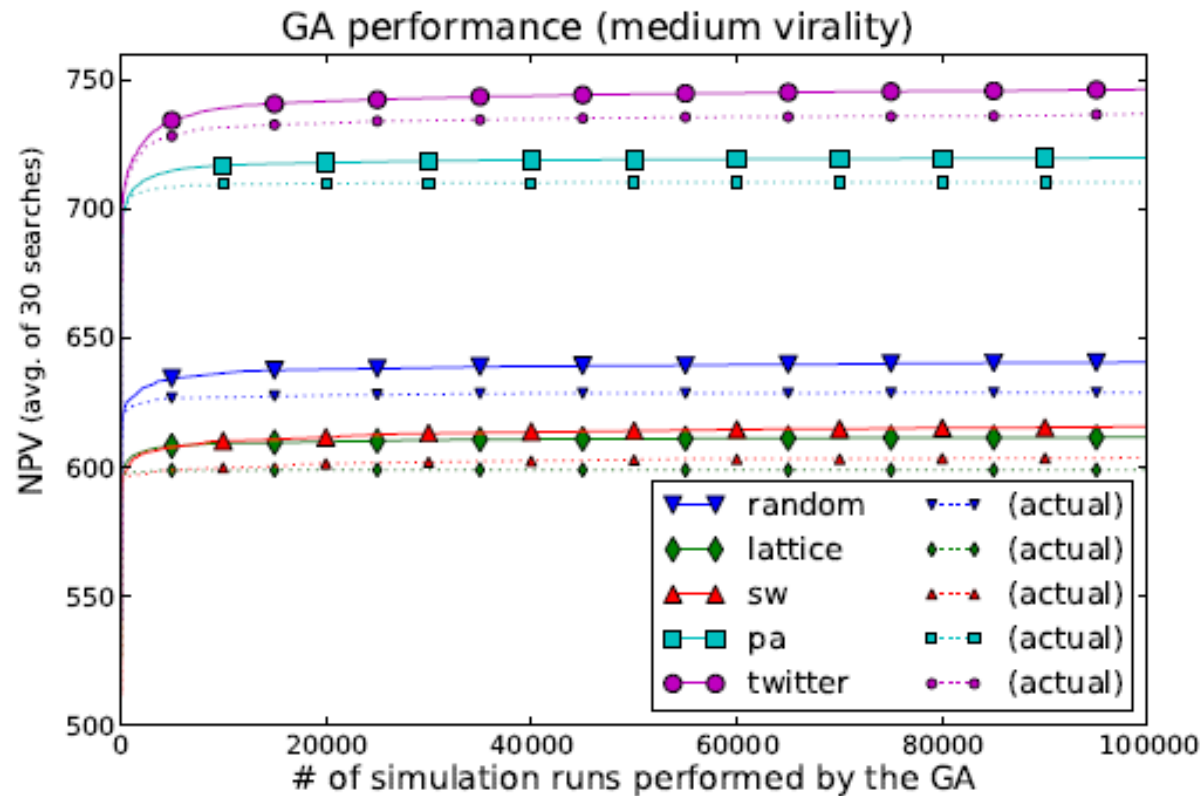
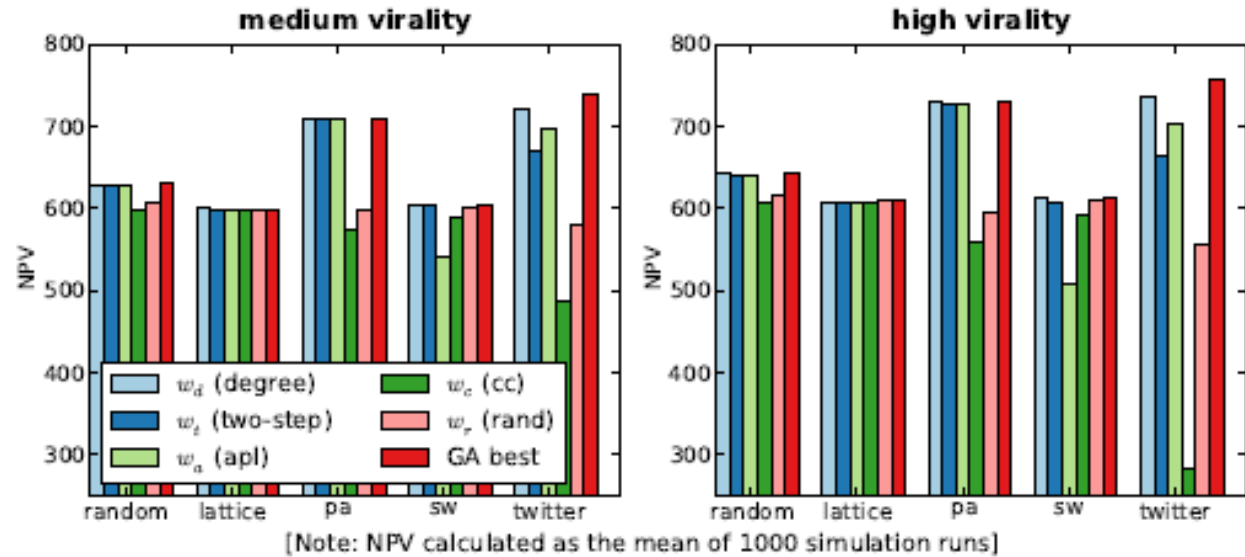


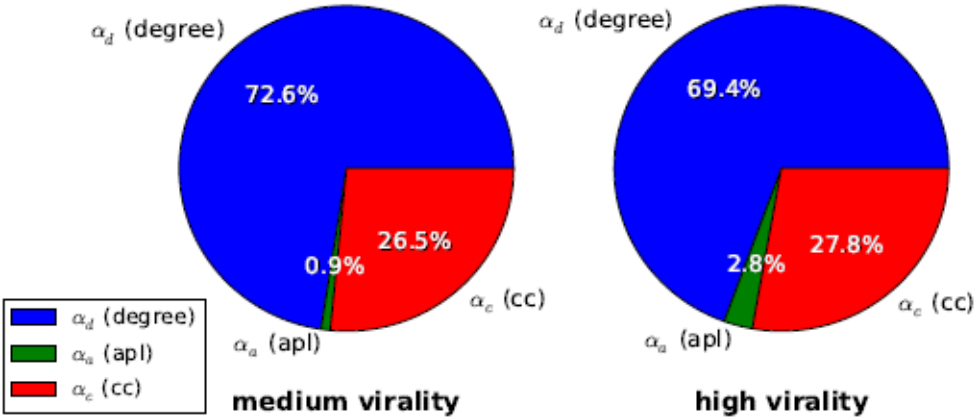
Figure 3: GA progress (averaged across 30 searches) by network topology, for the 'medium virality' scenario. GA's reported best-of-run fitness (solid lines) are compared with the actual NPV values (dotted lines), estimated by 1000 simulation runs, showing the effect of noise. (Error bars too small to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

APRENDIZAJE EVOLUTIVO DE ESTRATEGIAS DE MARKETING VIRAL (5)

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

APRENDIZAJE EVOLUTIVO DE ESTRATEGIAS DE MARKETING VIRAL (6)

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**

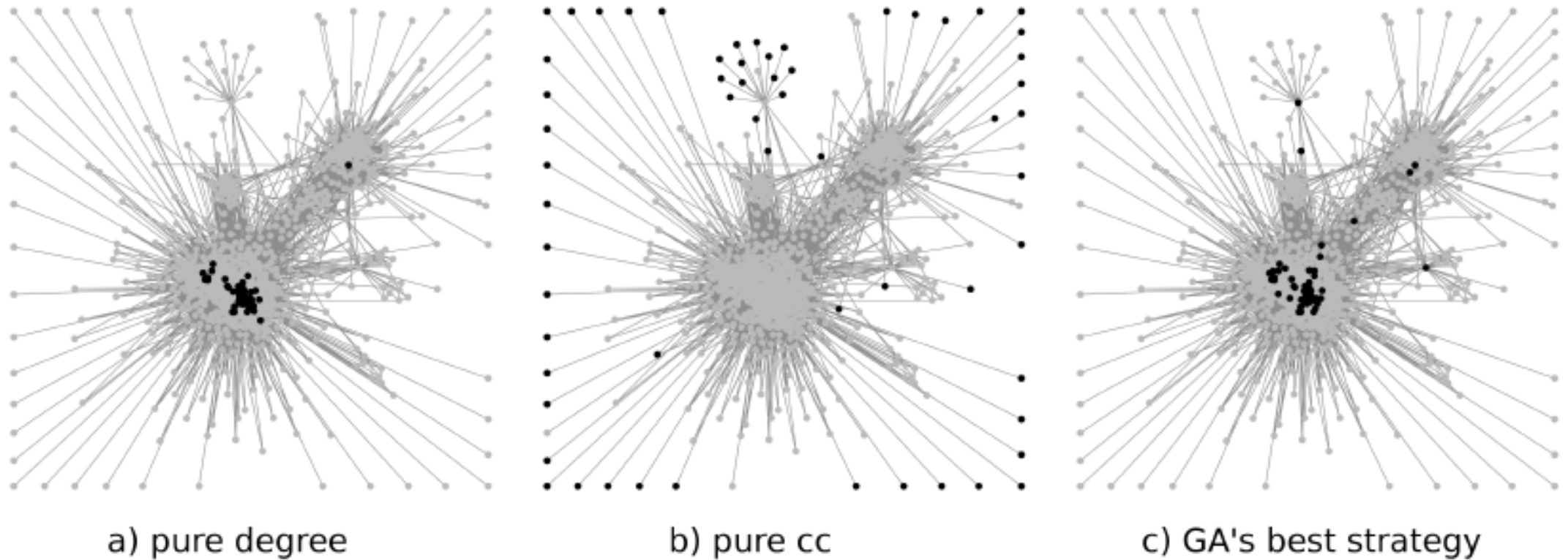


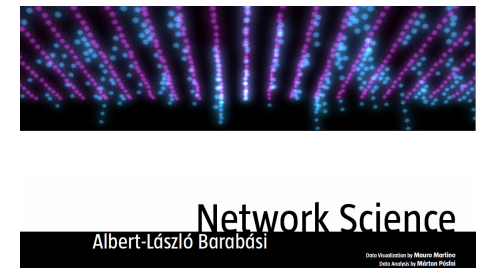
Figure 8: Visualization of three seeding strategies on the *twitter* network.

Referencias y Agradecimientos

Para elaborar las transparencias de este curso, he hecho uso de algunos materiales desarrollados por expertos en el área disponible en Internet:

- “Network Science Interactive Book Project” del Laszlo Barabasi Lab:

<http://barabasilab.com/networksciencebook>



- Curso on-line “Social Network Analysis” de Lada Adamic, Coursera, Universidad de Michigan: <https://www.coursera.org/course/sna>

