

Minería de Medios Sociales

Máster en Ciencia de Datos e Ingeniería de Computadores

Bloque I: Redes Sociales y Ciencia de Datos en Redes

Sesión I.5: Procesos en Redes

Oscar Cordón García

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

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 y la Etnografía, que se trasladan así a la Minería de Medios Sociales

Nos centraremos en técnicas que pueden modelar los procesos de difusión de información

EEUU, Febrero 2013, 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*”

La propagación fue desmesurada, alcanzando **más de 15,000 retweets y 20,000 likes en menos de 2 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



Torrejón, 28 Febrero 2017: una profesora de 3º de ESO diseña un experimento para concienciar de la rapidez de difusión de imágenes en redes

Dibujan un monigote con el texto “*Ayúdame a recorrer el mundo. Soy Nico*” y lo envían por sus perfiles (Instagram y Twitter) pidiendo su difusión. La profesora lo manda por Facebook y Whatsapp.

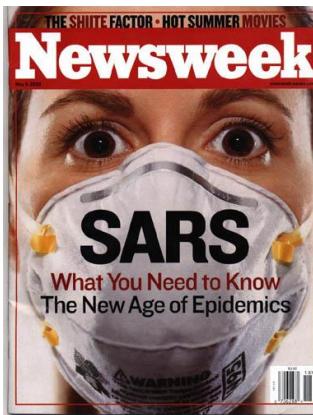
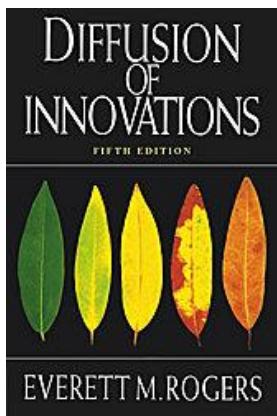
El dibujo lleva una maleta de una conocida marca deportiva. Muchos internautas pensaron que era una campaña de marketing

Nico se hizo viral en Whatsapp gracias a usuarios de 30 a 40 años. Llegó a los perfiles de Twitter de Guardia Civil y Policía Nacional. **Pasó por EEUU, Nicaragua, Venezuela, Honduras, Costa de Marfil, Italia, Francia, ...**



PROCESOS EPIDÉMICOS Y DE DIFUSIÓN

¿POR QUÉ ES TAN IMPORTANTE EL PROCESO DE PROPAGACIÓN?



***** SMART 11:37 AM

Search Twitter

#ALDubEBforLOVE 11.5M Tweets about this trend

#ShowtimeKapamilyaDay 3.45M Tweets about this trend

#NewAmericana 40.9K Tweets about this trend

#LarrysPureLove 180K Tweets about this trend

#2030NOW 3.76M Tweets about this trend

Epi + demos sobre pueblo



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

Biología:

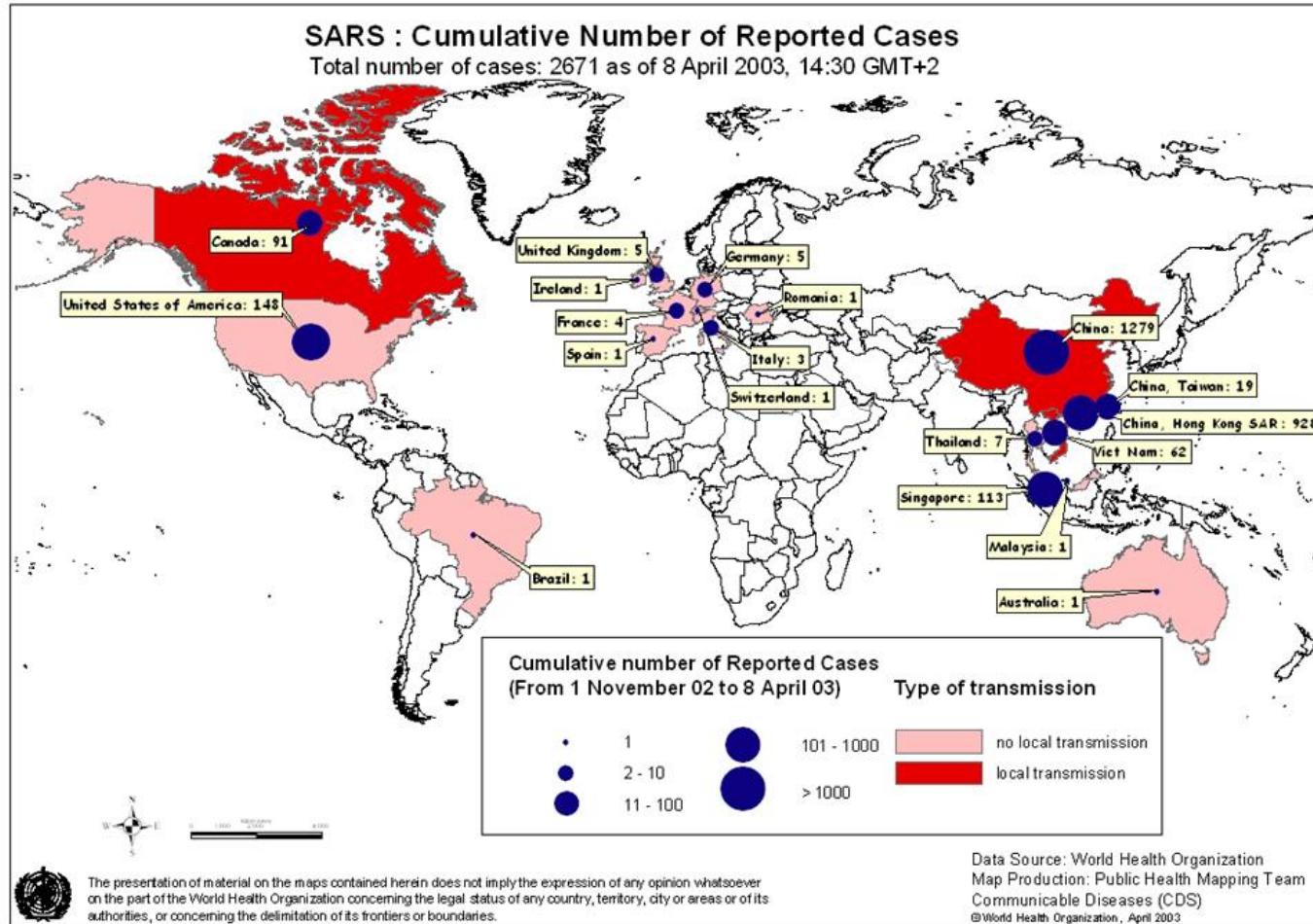
- Enfermedades transmitidas por el aire (gripe, gripe aviar, ...)
- Enfermedades venéreas (VIH, ...)
- Otras enfermedades infecciosas, incluidos algunos cánceres (VPH (virus del papiloma humano), ...)
- Parásitos (chinches, malaria, ...)

TIC:

- Virus de ordenador, gusanos
- Virus de teléfonos móviles

Conceptos/Aspectos intelectuales:

- **Difusión de información**
- **Difusión de innovaciones**
- **Rumores**
- **Memes (tweets, whatsapp, ...)**
- **Prácticas empresariales**



Fuente: Organización Mundial de la Salud

https://elpais.com/elpais/2019/09/25/planeta_futuro/1569435266_953355.html



PATRICIA PEIRÓ

SALUD

El mundo tiene que prepararse para la próxima gran pandemia letal

Un comité de expertos entrega a la ONU un análisis sobre el riesgo de una emergencia sanitaria global y qué hay que hacer para prevenirla. ¿La mayor amenaza? Una gripe masiva y mortal



Dos trabajadores sanitarios contra el ébola en RDC. UNICEF

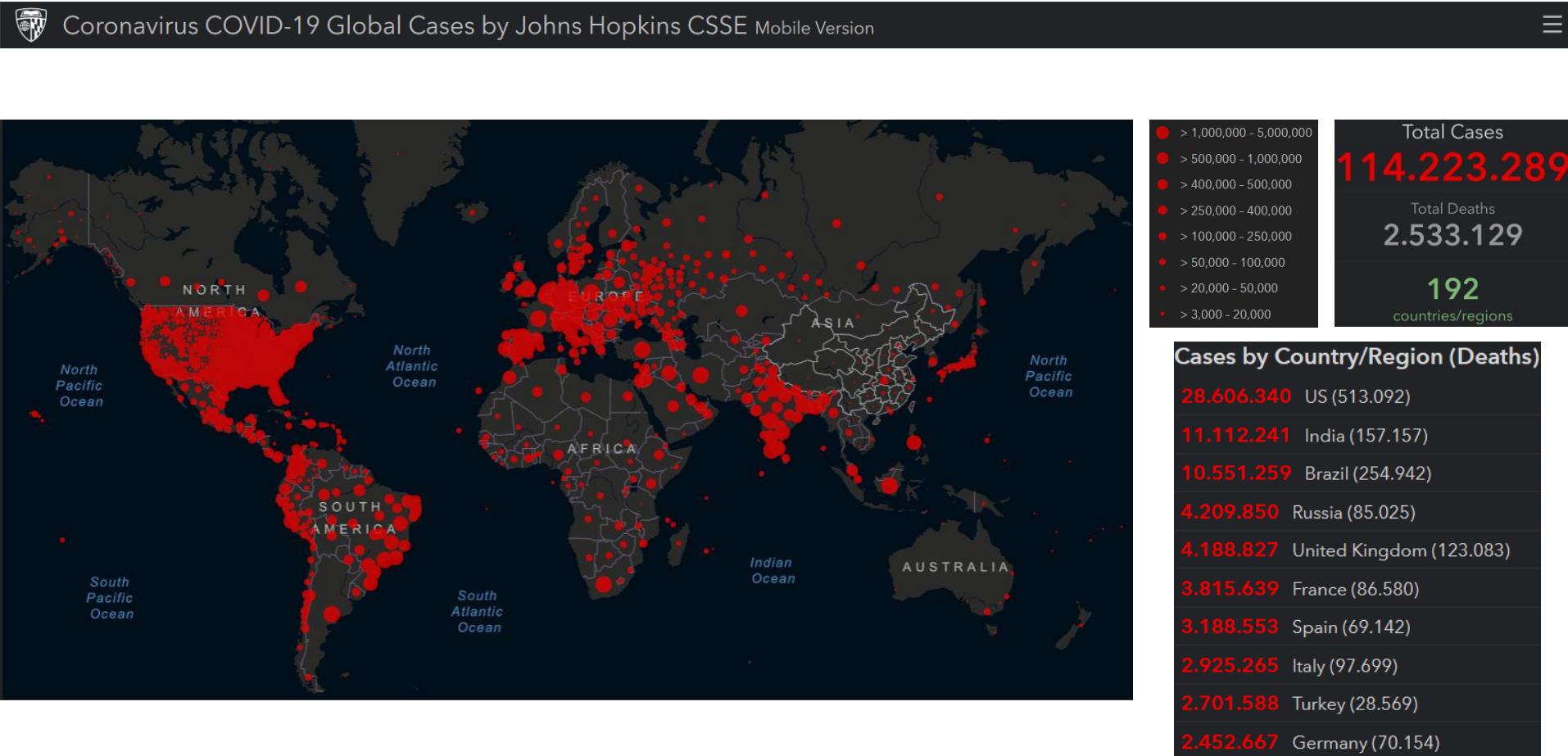
Nueva York 5 OCT 2019 - 17:38 CEST

Si un brote de un nuevo y agresivo tipo de gripe estallara mañana, el mundo no tendría herramientas para evitar la devastación. Morirían entre 50 y 80 millones de personas y liquidaría el 5% de la economía global. No contamos con las estructuras suficientes para hacer frente a la próxima pandemia letal. Esta es la cruda realidad sobre la que alerta un grupo de expertos de la OMS y el Banco Mundial, reunidos en una junta recién creada y llamada The Global Preparedness Monitoring Board (GPMB), a los que la ONU encargó una evaluación tras la última epidemia de ébola en África subsahariana, con el objetivo de aprender de los errores del pasado.

INTRODUCCIÓN

Siglo XXI: Hoy mismo... COVID-19 a 01/03/2021

<https://www.arcgis.com/apps/opsdashboard/index.html#/85320e2ea5424dfa75ae62e5c06e61>



INTRODUCCIÓN

Siglo XXI: Hoy mismo... COVID-19 a 25/11/2022

<https://www.arcgis.com/apps/opsdashboard/index.html#/85320e2ea5424dfa75ae62e5c06e61>



Coronavirus COVID-19 Global Cases by Johns Hopkins CSSE Mobile Version



Sorted by Cases in the Past 28 Days



INTRODUCCIÓN

Siglo XXI: Hoy mismo... COVID-19 a 25/11/2022

<https://www.covidvisualizer.com/>



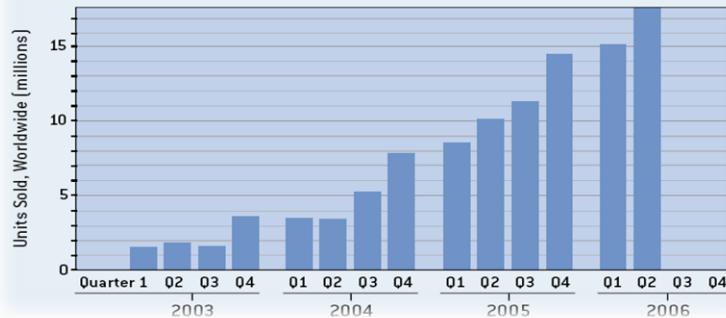
Data last updated a minute ago by [Worldometers](#).

This website was developed by [Navid Mamoon](#) ([@navidmx](#)) and [Gabriel Rasskin](#), ([@gabrielrasskin](#)) two students at Carnegie Mellon University.

The goal of this project is to provide a simple, interactive way to visualize the impact of COVID-19. We wanted people to be able to see this as something that brings us all together. It's not one country, or another country; it's one planet – and this is what our planet looks like today.

The data is from [Worldometer's](#) real-time updates, utilizing reliable sources from around the world. The *TODAY* cases/deaths are based on GMT (+0). The website pulls new data every 2 minutes, refresh to see any changes.

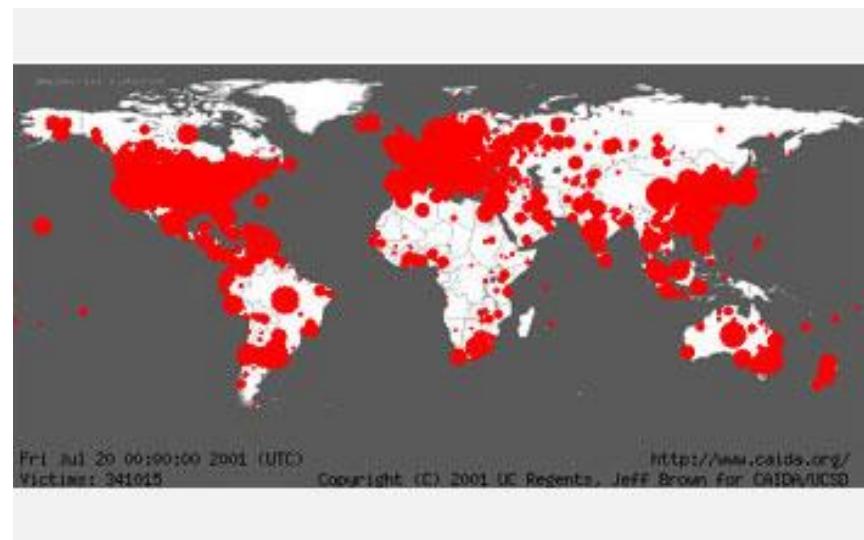
SMARTPHONES ON THE RISE



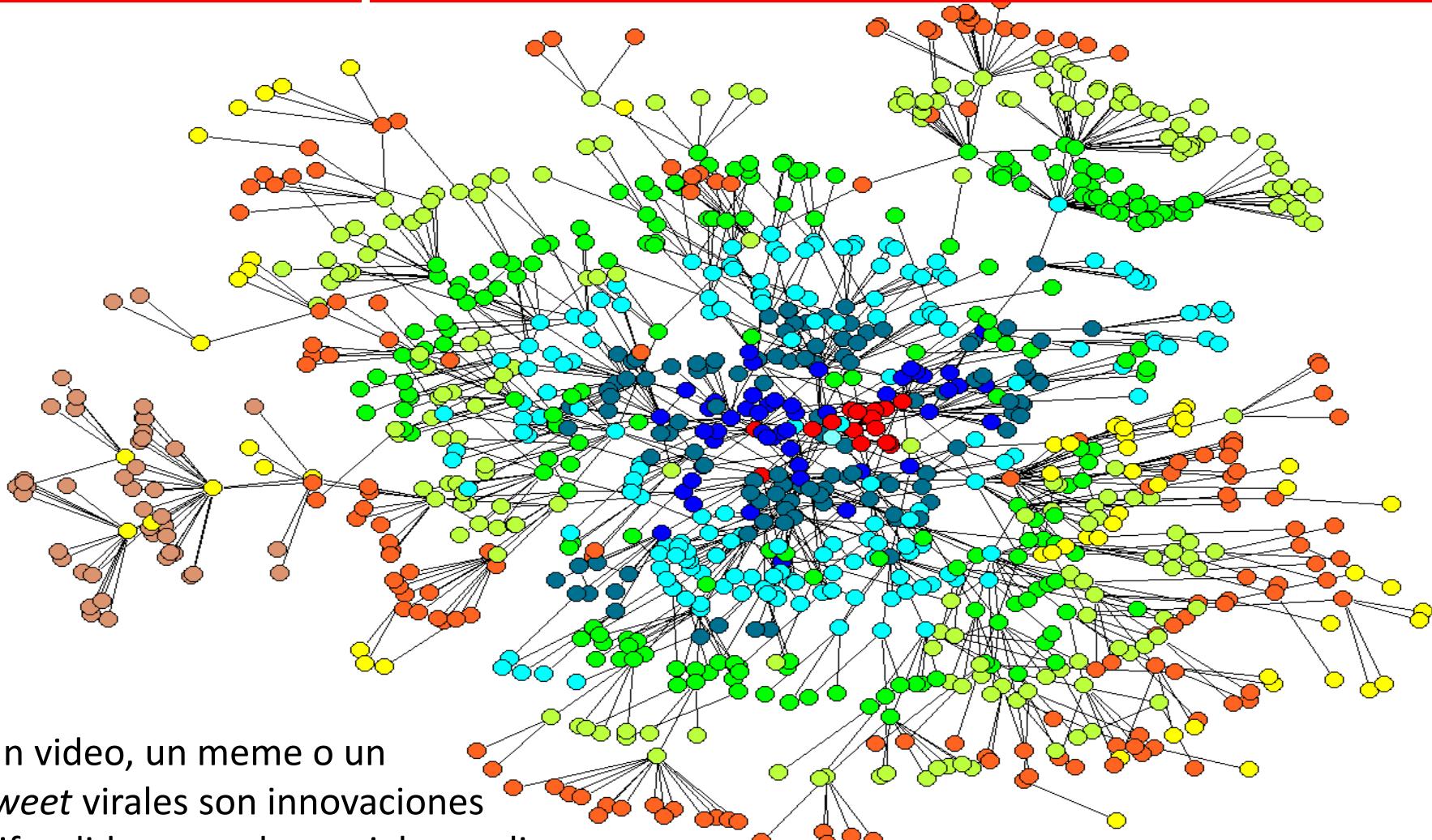
GROWTH IN MOBILE MALWARE



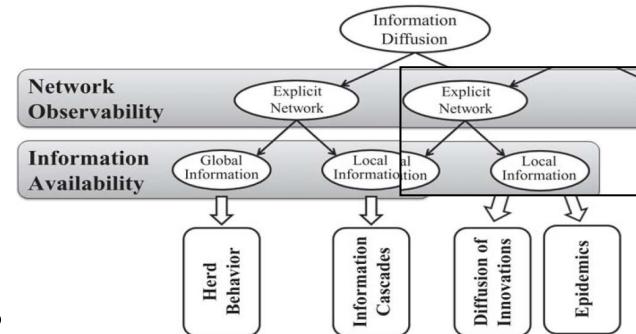
El código del *Red Worm* paralizó Internet en muchos países

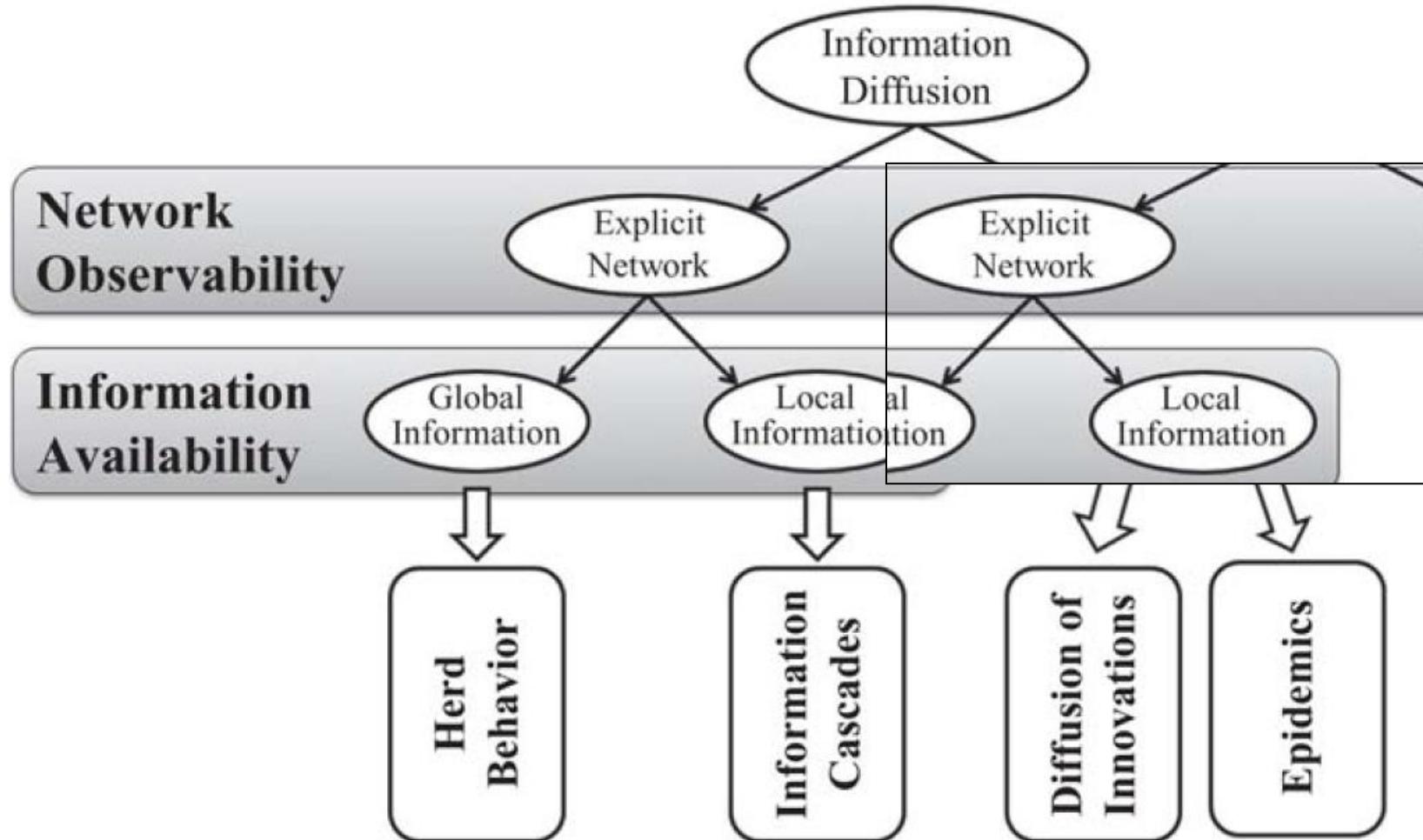


<http://www.caida.org/publications/visualizations/>



- Los modelos de difusión de información 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)
- En los modelos epidémicos y de cascada *centrados en el emisor*, los individuos no toman la decisión por si mismos. En el resto sí
- Los **contagios** pueden ser **simples** (individuales por probabilidad) y **complejos** (umbrales)





MODELOS CLÁSICOS DE PROPAGACIÓN DE EPIDEMIAS

[http://es.wikipedia.org/wiki/Modelaje matemático de epidemias](http://es.wikipedia.org/wiki/Modelaje_matem%C3%A1tico_de_epidemias)

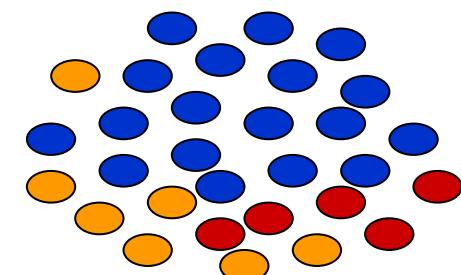
http://www.uni-tuebingen.de/modeling/Mod_Pub_Software_SIR_en.html

La **Epidemiología** describe el proceso mediante el que se difunden las enfermedades. Se basa en:

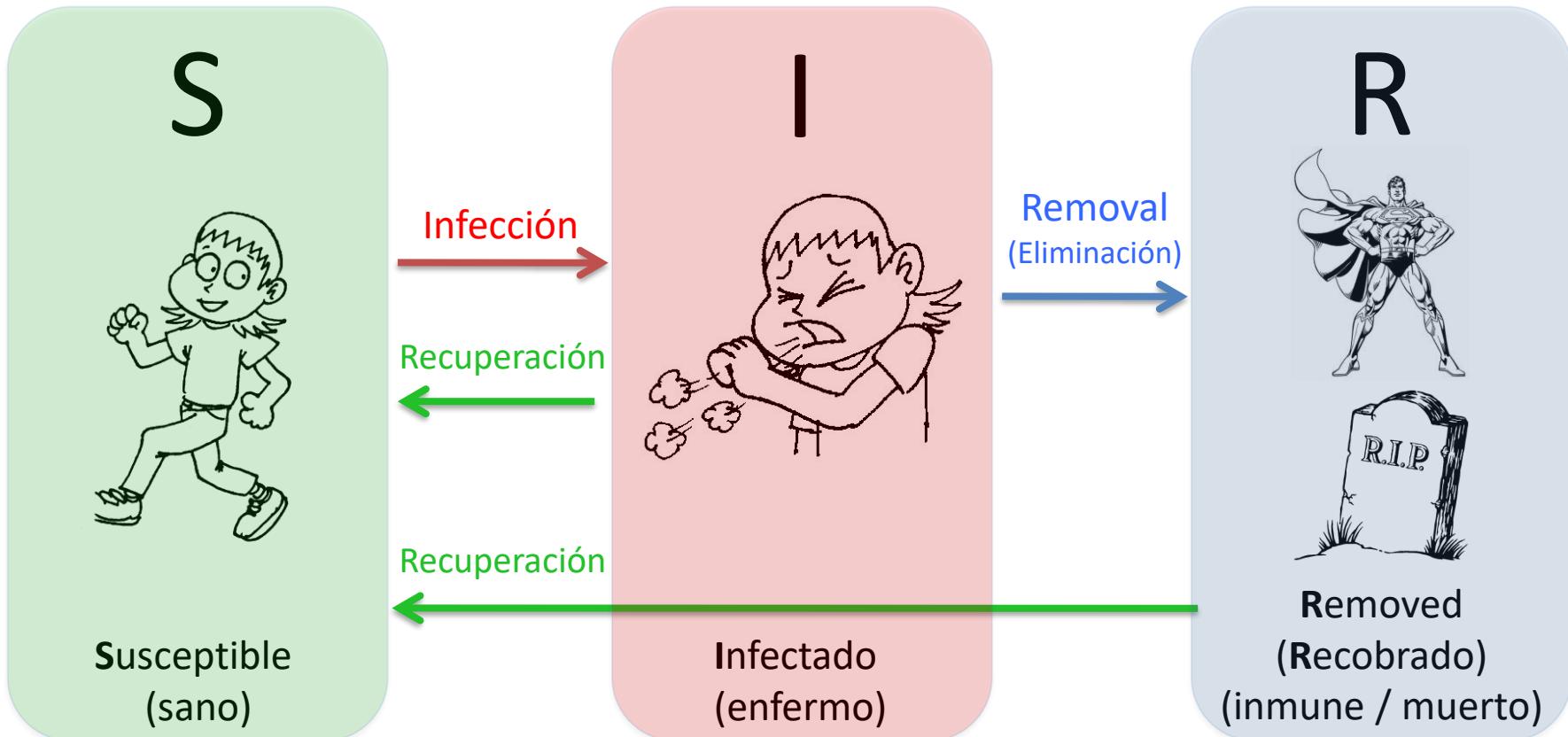
- Un **patógeno** (la enfermedad que se está transmitiendo),
- Una **población de huéspedes** (humanos, animales, plantas, etc.) en **distintos estados**
- Un **mecanismo de transmisión** (respiración, bebida, actividad sexual, etc.)

Los individuos se clasifican en varios estados según la etapa de desarrollo de la enfermedad (**compartimentación**):

- *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

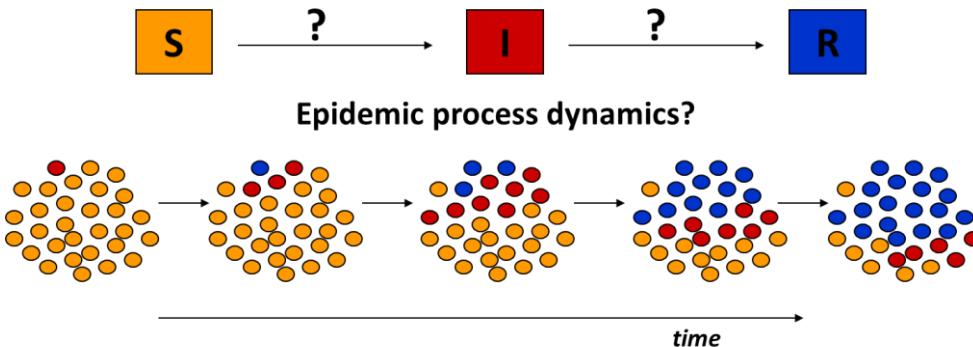


ESTADOS BÁSICOS Y TRANSICIONES DEL MODELO CLÁSICO SIR

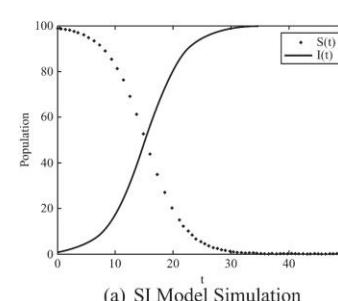


Los modelos clásicos (1927) asumen una red de contactos implícita y que no se conocen las conexiones entre los individuos → **No hay red = mezclado homogéneo (cada individuo puede infectar a cualquier otro en cualquier momento):**

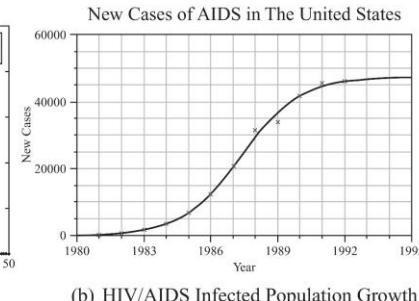
- Soluciones analíticas basadas en ecuaciones diferenciales
- Sólo interesa la obtención de **patrones globales** (tendencias y ratios de población infectada) y no quién infecta a quién
- Estimación errónea de la dinámica de contagio (velocidad, picos, ...)



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

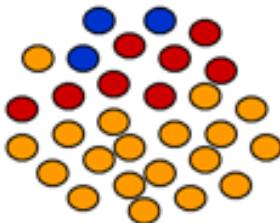


(a) SI Model Simulation



(b) HIV/AIDS Infected Population Growth

Logistic growth function compared to the HIV/AIDS growth in the United States



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

La probabilidad de que quede infectado por esos contactos es $\beta \in [0,1]$ (**CONTAGIO SIMPLE**). El ratio de transmisión de la enfermedad es $\beta \cdot \langle k \rangle$ y determina el contagio

Los infectados se recuperan, volviéndose inmunes, o mueren con una probabilidad μ

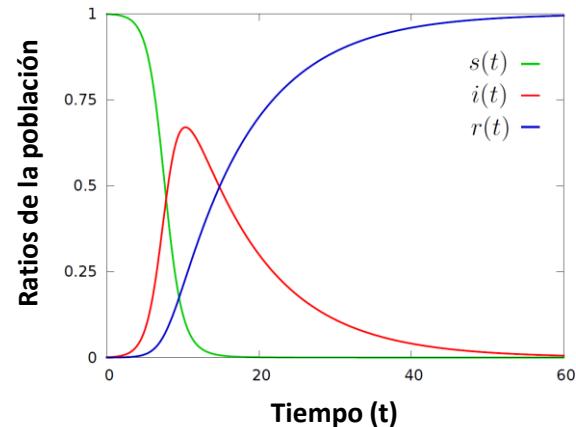
En tiempo t , la fracción de individuos infectados en una población de tamaño N es $i = I(t)/N$, la de individuos sanos $s = S(t)/N$ y la de individuos recobrados es $r = R(t)/N$. Lógicamente, $N = i + s + r$

Al haber i infectados, el **ratio medio de nuevas infecciones** en tiempo t es $\beta \cdot \langle k \rangle \cdot s \cdot i$ y el **ratio medio de recuperaciones/fallecimientos** es $\mu \cdot i$

$$\frac{ds(t)}{dt} = -\beta \langle k \rangle i(t) [1 - r(t) - i(t)]$$

$$\frac{di(t)}{dt} = -\mu i(t) + \beta \langle k \rangle i(t) [1 - r(t) - i(t)]$$

$$\frac{dr(t)}{dt} = \mu i(t).$$



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/contener la epidemia es la **cuarentena y/o la vacunación tempranas**

Un parámetro de estudio importante es el **tiempo característico τ , que indica el tiempo necesario para que se contagie una fracción $1/e$ de la población** (aprox. un 36%):
$$\tau = \frac{1}{\beta \cdot \langle k \rangle}$$

Es la inversa de la velocidad con la que el patógeno se extiende por la población. Aumenta con el número de contactos o la probabilidad de contagio

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.

MODELOS DE PROPAGACIÓN DE EPIDEMIAS BASADOS EN REDES

NO CONSIDERAR LA RED PROVOCA ERRORES GARRAFALES: H1N1 2009

EL PAÍS

[https://elpais.com/diario/2010/06/05/
sociedad/1275688803_850215.html](https://elpais.com/diario/2010/06/05/sociedad/1275688803_850215.html)

SOCIEDAD

EDUCACIÓN MEDIO AMBIENTE IGUALDAD SANIDAD CONSUMO LAICISMO COMUNICACIÓN ÚLTIMAS NOTICIAS

La OMS ocultó que sus expertos en gripe A cobraron de farmacéuticas

El organismo no informó a los Gobiernos cuando les instó a reservar antivirales

JAVIER SAMPEDRO

Madrid - 05 JUN 2010 - 07:00 CEST

Aparte de los intereses económicos, se estimó erróneamente el pico, pensando que se comportaría como una gripe normal

Los países hicieron acopio de fármacos por valor de 4.900 millones de euros

Pese al temor inicial, el H1N1 ha sido menos letal que el virus convencional

España ha usado 2 de los 13 millones de dosis de vacunas que compró



La gestión de la pandemia de gripe A por la [Organización Mundial de la Salud \(OMS\)](#) fue poco transparente. Según [un estudio del British Medical Journal \(BMJ\)](#), una de las revistas médicas de referencia, un informe clave de la OMS ocultó los vínculos financieros entre sus expertos y las farmacéuticas Roche y Glaxo, fabricantes de Tamiflu y Relenza, los fármacos antivirales contra el virus H1N1. Ese fue el informe que instó a los Gobiernos a apilar reservas de esos medicamentos, por valor de unos 6.000 millones de dólares (4.900

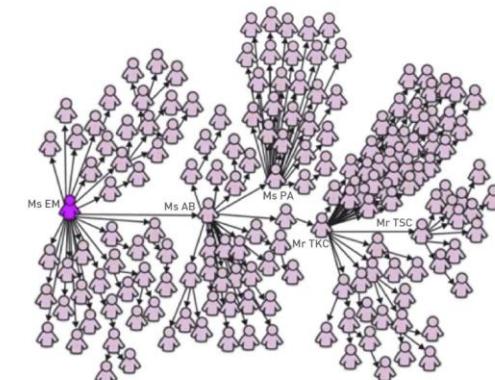
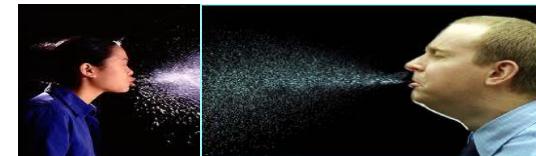
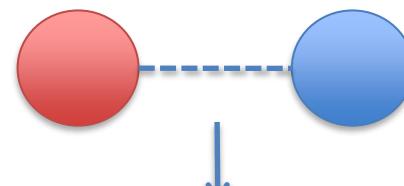
REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (1)

¡La difusión de epidemias (y cualquier otro proceso de propagación) implica siempre una estructura de red subyacente!

P. ej. las epidemias se propagan a través de las **redes de contactos**. Los nodos son huéspedes y los enlaces indican interacciones entre ellos:

- HIV/SIDA: los enlaces indican relaciones sexuales (redes sexuales)
- Gripe: los huéspedes son gente que respira el mismo aire y se mueve por el mundo (redes de contactos y de transporte)

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
Chinches	Red de hoteles – viajeros	Chinches
Malaria	Red de mosquitos – humanos	<i>Plasmodium</i>
Propagación de rumores	Red de comunicaciones	Información, memes
Difusión de innovaciones	Red de comunicaciones	Ideas
Gusanos de Internet	Internet	Malware (códigos binarios)
Virus de teléfonos móviles	Red social / Red Bluetooth de proximidad	Malware (códigos binarios)

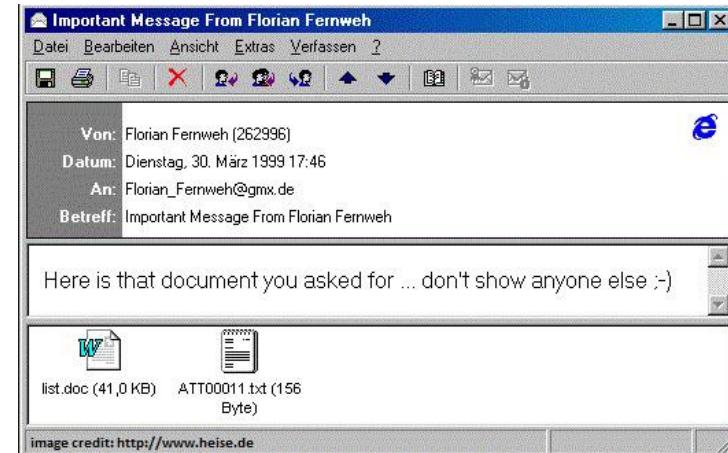
REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (3): Ejemplo

Marzo 1999: El virus Melissa infecta ordenadores a través del gestor de e-mail MS Outlook

El usuario recibía un correo con un documento Word con virus en el attachment. Al abrirlo, el virus se reenviaba a los primeros 50 contactos de la libreta de direcciones

Se detectó por primera vez el viernes 26 de Marzo. El lunes 29 de Marzo el virus había infectado **más de 100.000 ordenadores**

Compañías como Microsoft, Intel o Lucent Technologies tuvieron que bloquear sus conexiones a Internet a causa de Melissa



REDES COMPLEJAS Y MODELADO DE EPIDEMIAS (4)

El enfoque clásico de modelado de epidemias **no tiene en cuenta explícitamente que la propagación se produce en una red compleja**

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 te contagias si una cantidad/porcentaje de tus “amigos” lo están (**umbráles**)

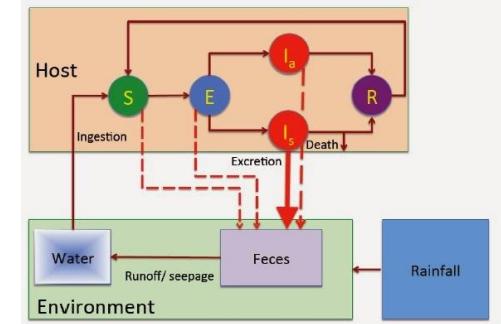
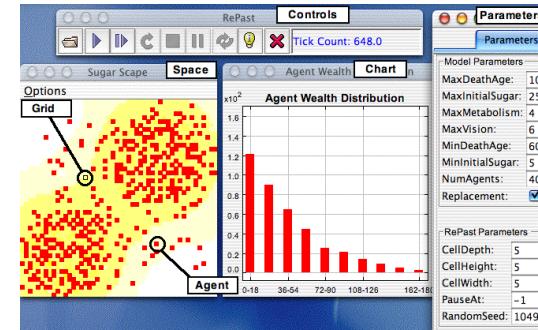
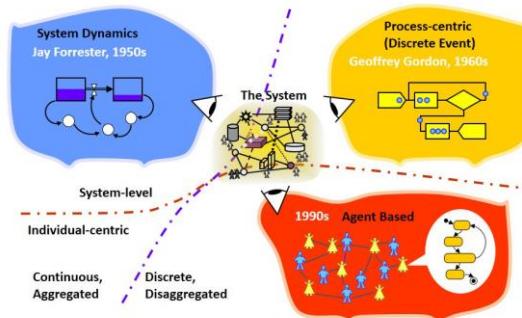
MODELO SIR EN REDES COMPLEJAS: FUNDAMENTOS (1)

Los modelos epidemiológicos modernos (>2000) consideran el papel de la red de contactos en el proceso epidémico. Se basan en una metodología de Inteligencia Artificial muy potente llamada Modelos Basados en Agentes (ABMs)

Los ratios de cada estado se calculan de forma agregada. Se trabaja a nivel de cada individuo (agente). Se consideran sus características individuales, que pueden variar entre ellos, las interacciones locales (*comportamiento emergente*) y la aleatoriedad (Montecarlo)

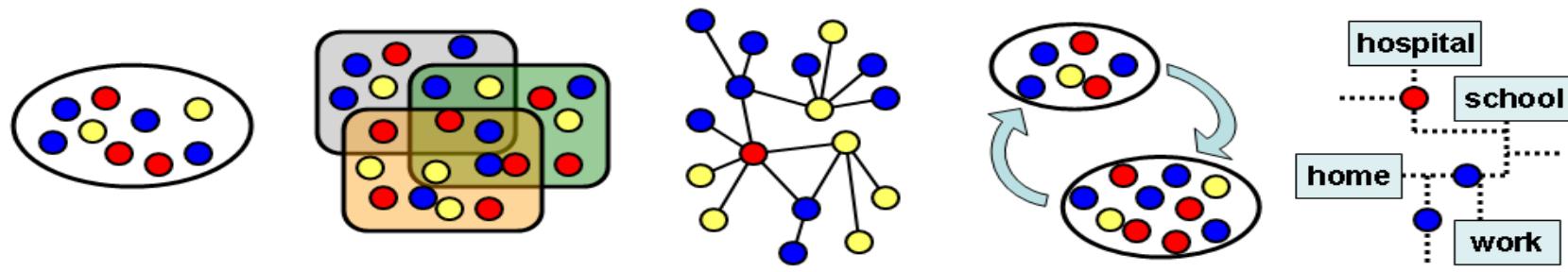
P.ej. se puede modelar el progreso del brote en distintos escenarios con individuos de distintos tipos, distintas conexiones entre ellos, etc.

De hecho, las redes de contactos reales sólo pueden emplearse con ABMs. Los métodos basados en expresiones analíticas sólo pueden usar redes implícitas



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:



Homogeneous mixing

Social structure

Contact network models

Multi-scale models

Agent Based models

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

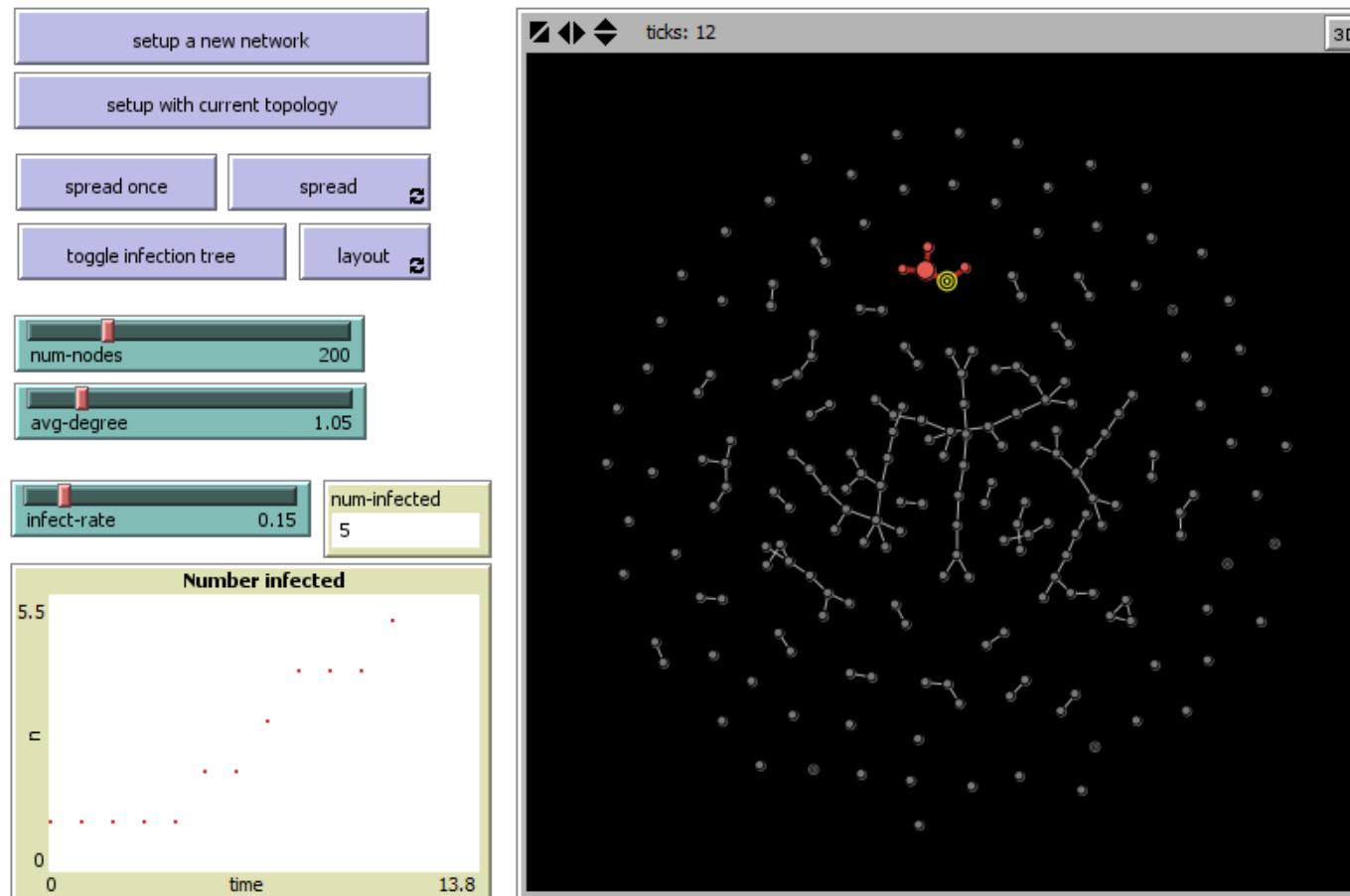
El modelo SIR es muy apropiado para las fases iniciales de un ataque de un virus de ordenador:

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

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



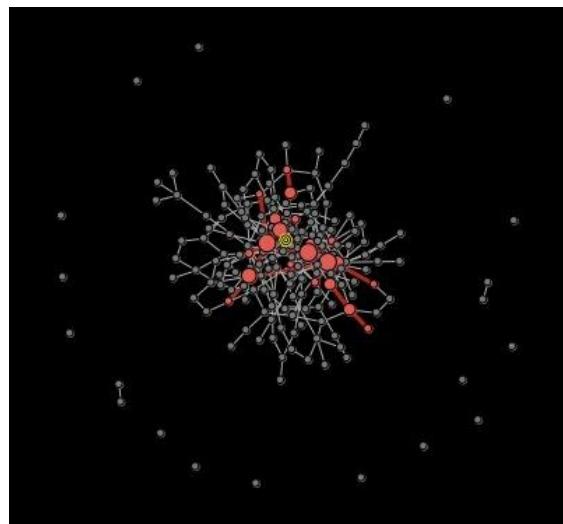
- Cuando $t \rightarrow \infty$, cada nodo susceptible con posibilidad de infectarse queda infectado
- La única condición es que exista un camino en la red entre dicho nodo y cualquier otro nodo infectado, de forma que la enfermedad pueda alcanzarlo
- **A diferencia del modelo clásico**, si partimos 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

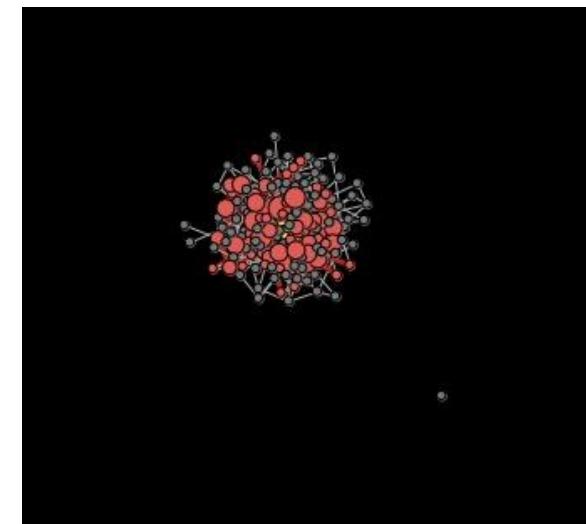
Con los mismos parámetros del modelo, ¡la difusión depende de la topología de la red!

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

grado medio $\langle k \rangle = 2.5$



grado medio $\langle k \rangle = 10$



PREGUNTA: Cuando aumenta la densidad de la red,
¿la propagación es igual, más rápida o más lenta?

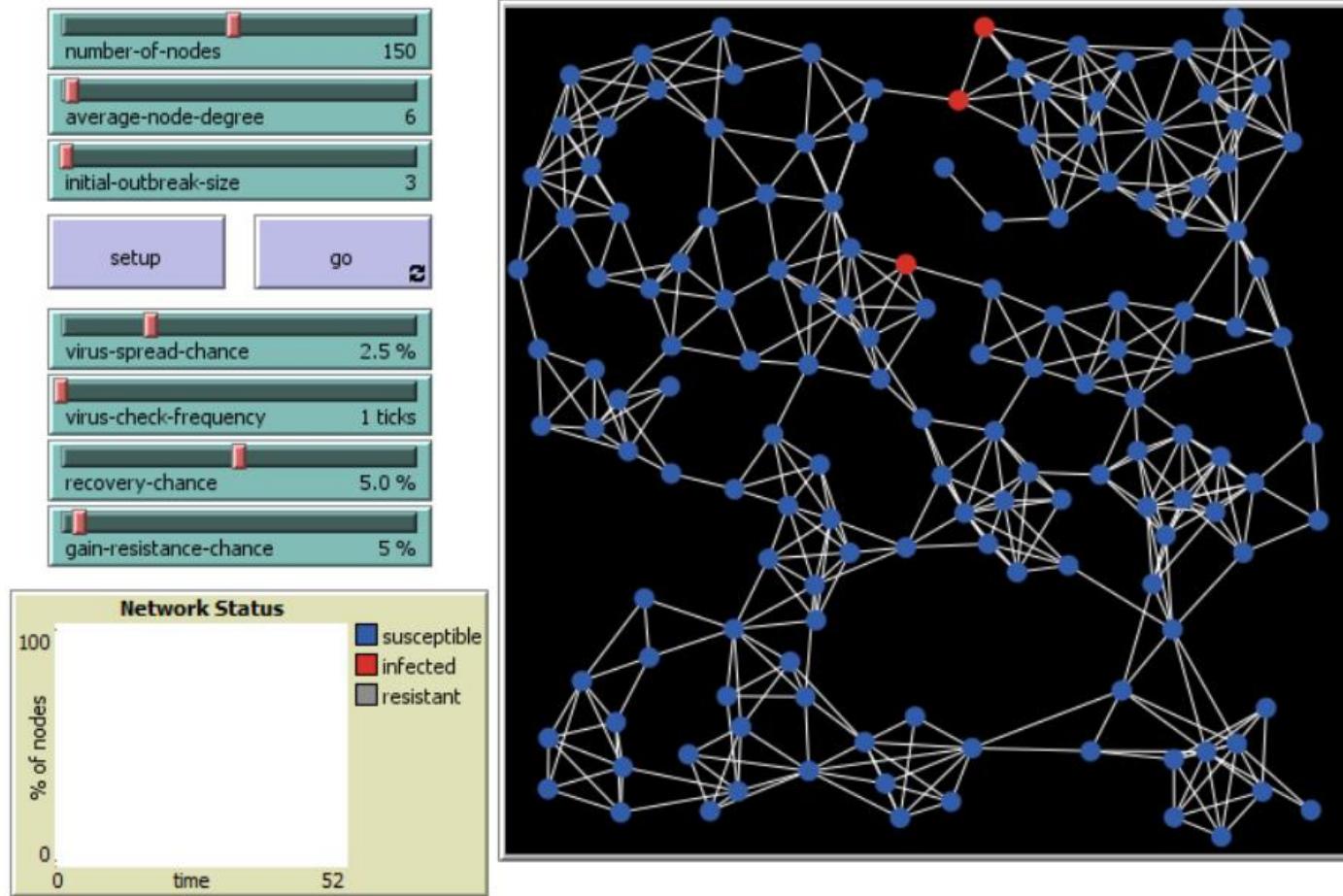
<http://www.ladamic.com/netlearn/NetLogo501/BADiffusion.html>

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



La presencia de hubs en la red aumenta la velocidad de difusión

<https://ccl.northwestern.edu/netlogo/models/VirusonaNetwork>

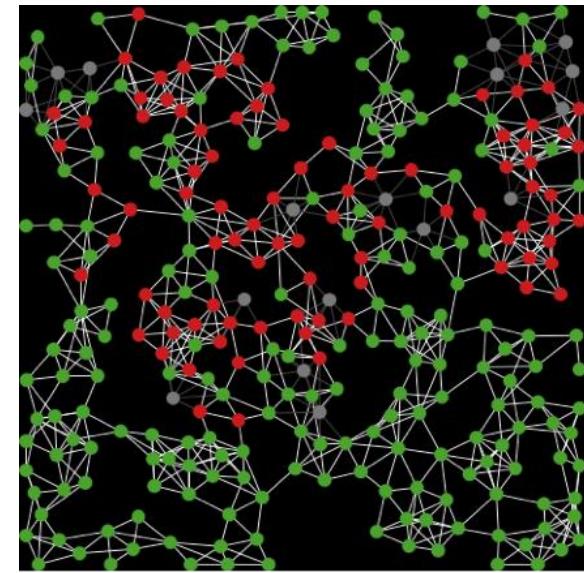
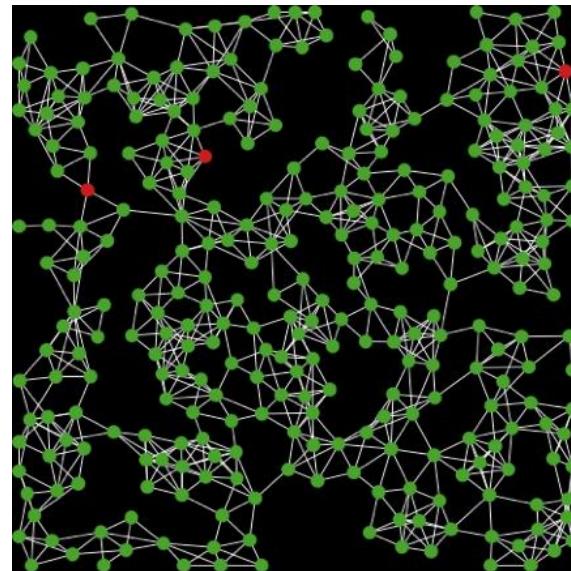
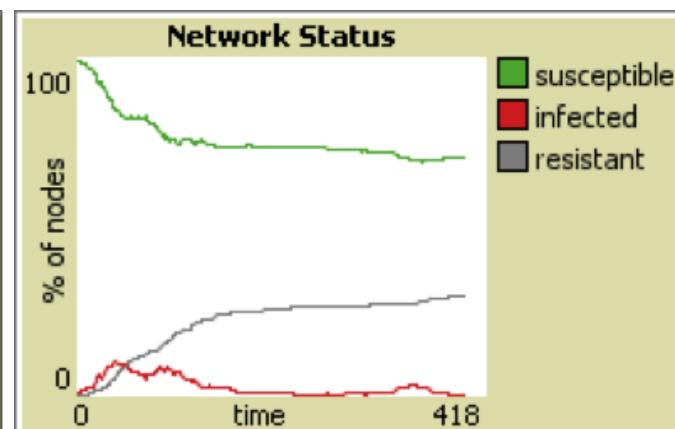
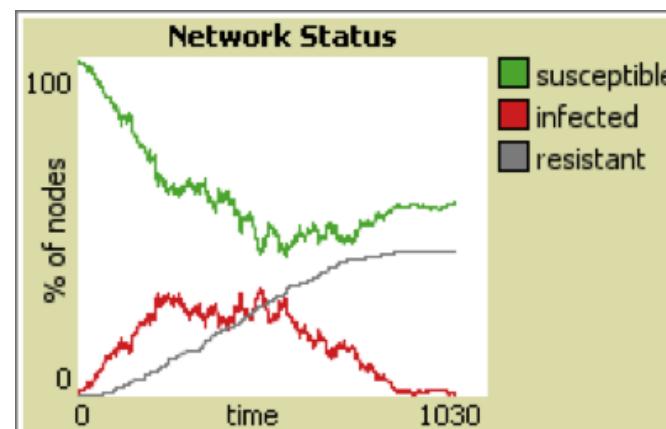


- **S→I:** Cada agente infectado puede infectar un agente susceptible conectado a él de acuerdo a la probabilidad (ratio) de infección β
- **I→R|S:** Cada agente infectado puede recuperarse con una probabilidad de recuperación μ . En ese caso, puede pasar a dos estados distintos:
 - **I→R:** El agente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

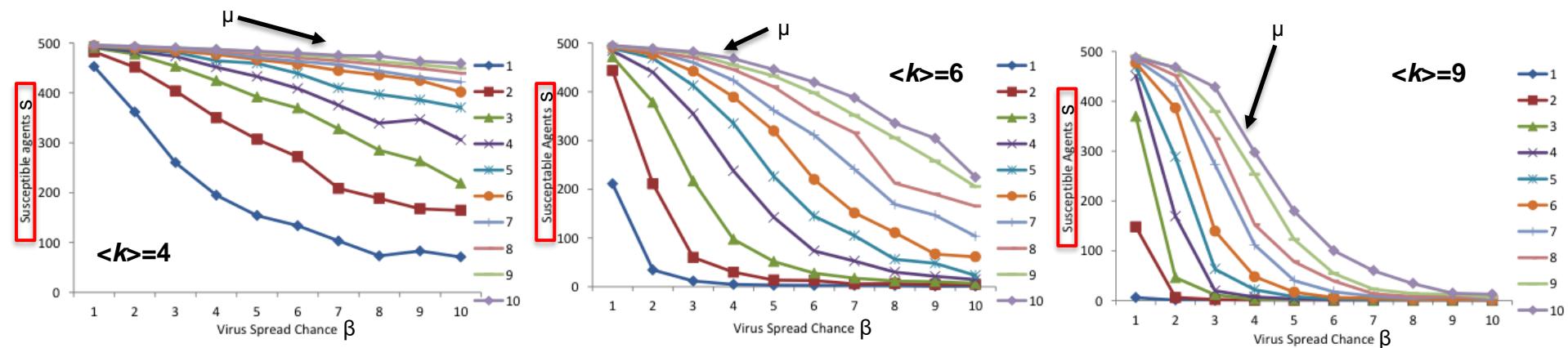
Ejemplo: $\beta = 0.025$ (2.5%) $\mu = 0.005$ (0.5%) $\tau = 0.05$ (5%)Grado medio $\langle k \rangle = 6$

3 individuos infectados en la población inicial

214 pasos de simulación

**Estudio del umbral epidemiológico:** $\beta = 0.025$ (2.5%) $\mu = \{0.005, 0.05\}$
(0.5%, 5%)

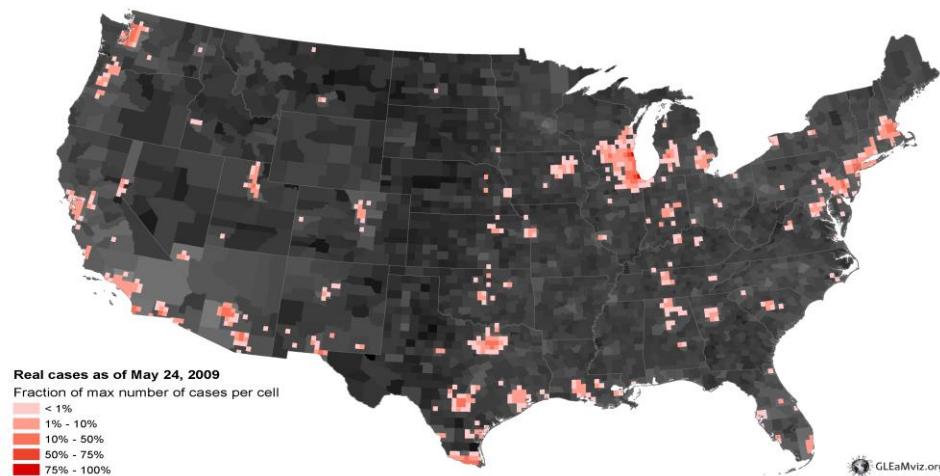
- Estudio de la influencia de la topología de la red (grado medio $\langle k \rangle$) y de los valores de las probabilidades de infección β y recuperación μ
- Modelo SIR: Los agentes recobrados no pueden volver a ser susceptibles (modelo SIRS anterior con $\iota=1$)
- Tres redes ER con $\langle k \rangle=\{4, 6, 9\}$. $\beta=\mu=\{0.01, 0.02, \dots, 0.1\}$. 500 agentes. 100 simulaciones con 1000 pasos (**Montecarlo**)



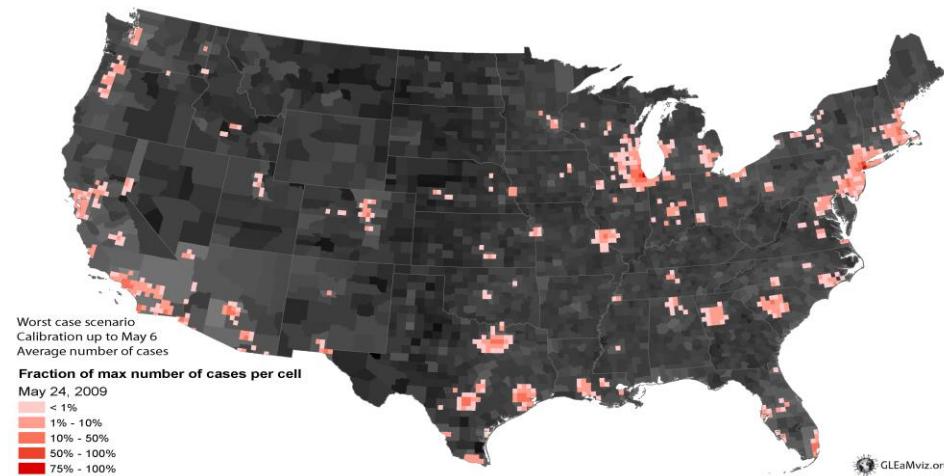
- Claramente, a mayor grado medio, mayor facilidad para la difusión del virus

Predicción de la pandemia de la gripe A (H1N1) de 2009:

Real



Pronosticada



Primera pandemia analizada con redes complejas y ABM



GLEAMviz The Global Epidemic and Mobility Model

NEWS TEAM PUBLICATIONS PRESS

VISION CHALLENGES APPROACH MODEL SIMULATOR CASE STUDY

PUTTING POWERFUL TOOLS
IN THE HANDS OF EXPERTS



SIMULATOR

The GLEAM Simulator system consists of the GLEAM Server and the GLEAMviz Client application.

The GLEAM Server uses GLEAM as the engine to perform the simulations. This server runs on high-performance computers managed by the GLEAM project.

The GLEAMviz Client is a desktop application through which users interact with the GLEAM Server. It provides a simple, intuitive and visual way to set up simulations, develop disease models, and evaluate simulation results using a variety of maps, charts and data analysis tools.

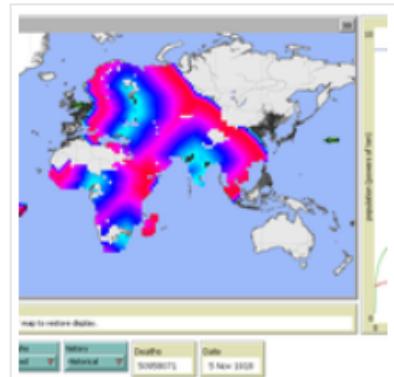
GLEAM IN ACTION

DOWNLOAD OUR SIMULATOR CLIENT
AND RUN YOUR OWN SIMULATIONS!



<http://www.bleamviz.org>

Kahn, Ken (2013, August 05). "A Modelling4All/NetLogo model of the Spanish Flu Pandemic" (Version 1.1.0). *CoMSES Computational Model Library*. Retrieved from: <https://www.comses.net/codebases/3890/releases/1.1.0/>



A Modelling4All/NetLogo model of the Spanish Flu Pandemic

Ken Kahn | Published Mon Aug 5 15:17:09 2013 | Last modified Mon Aug 5 15:21:25 2013

A global model of the 1918-19 Influenza Pandemic. It can be run to match history or explore counterfactual questions about the influence of World War I on the dynamics of the epidemic. Explores two theories of the location of the initial infection.

<https://elephantinthelab.org/predictions-on-the-spread-of-covid-19-bear-many-uncertainties/>

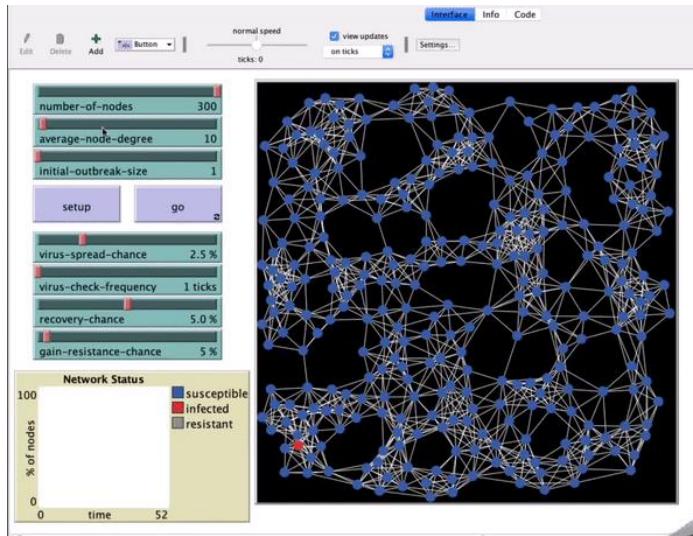
WANDER JAGER



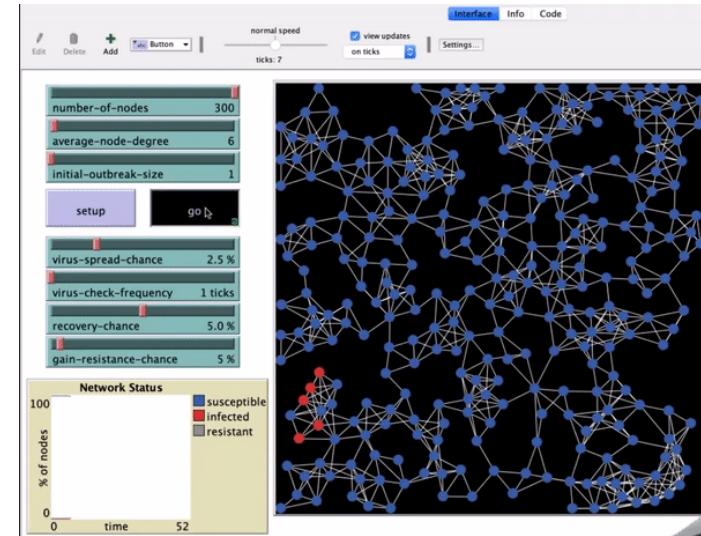
ELEPHANT

Bold ideas and critical thoughts on science. IN THE LAB

22 March 2020 | doi:10.5281/zenodo.3724048 | No Comments



Pic.1. Each person has ten contacts on average: the virus spreads very fast.



Pic.2. Each person has six contacts on average: the virus spreads slowly.

nature<https://www.nature.com/articles/s41586-020-2923-3>

Article | Published: 10 November 2020

Mobility network models of COVID-19 explain inequities and inform reopening

Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky & Jure Leskovec

Abstract

The coronavirus disease 2019 (COVID-19) pandemic markedly changed **human mobility patterns**, necessitating epidemiological models that can capture the effects of these changes in mobility on the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Here we introduce a **metapopulation susceptible–exposed–infectious–removed (SEIR) model that integrates fine-grained, dynamic mobility networks to simulate the spread of SARS-CoV-2 in ten of the largest US metropolitan areas**. Our mobility networks are derived from mobile phone data and map the hourly movements of **98 million people** from neighbourhoods (or census block groups) to points of interest such as restaurants and religious establishments, connecting 56,945 census block groups to 552,758 points of interest with 5.4 billion hourly edges. We show that by integrating these networks, a relatively simple SEIR model can accurately fit the real case trajectory, despite substantial changes in the behaviour of the population over time. **Our model predicts that a small minority of ‘superspread’ points of interest account for a large majority of the infections, and that restricting the maximum occupancy at each point of interest is more effective than uniformly reducing mobility.** Our model also correctly predicts higher infection rates among disadvantaged racial and socioeconomic groups solely as the result of differences in mobility: we find that disadvantaged groups have not been able to reduce their mobility as sharply, and that the points of interest that they visit are more crowded and are therefore associated with higher risk. By capturing who is infected at which locations, our model supports detailed analyses that can inform more-effective and equitable policy responses to COVID-19.

nature
PUBLIC HEALTH

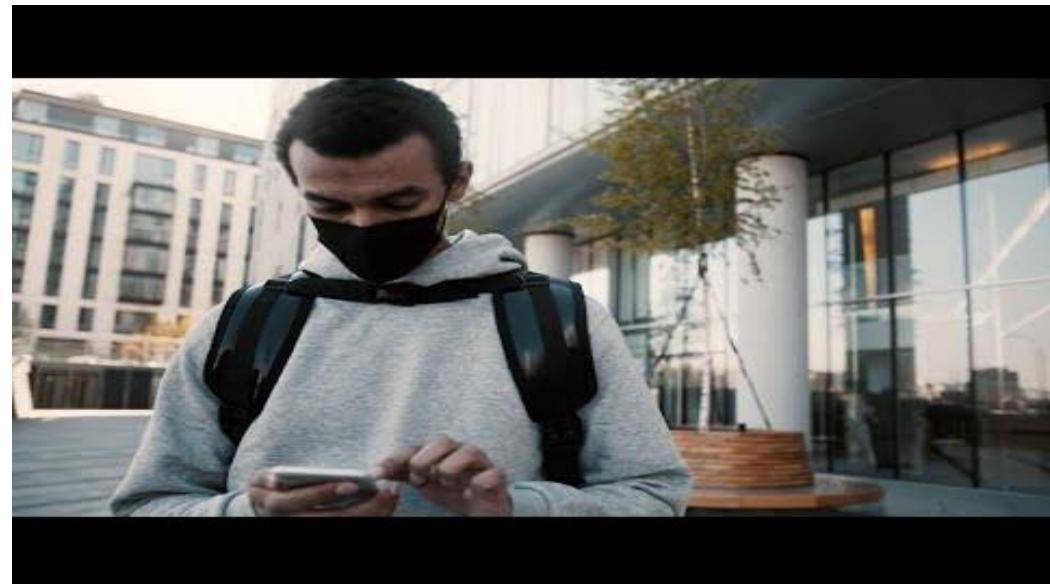
How to Stop Restaurants from Driving COVID Infections

Mobile phone data suggest restaurants, gyms and cafés can be COVID hotspots—and reveal strategies for limiting spread

By David Cyranoski, Nature magazine on November 10, 2020

<https://www.scientificamerican.com/article/how-to-stop-restaurants-from-driving-covid-infections/>

<https://youtu.be/FZG85TN1nhE>



COVID-19 MOBILITY NETWORK MODELING

Mobility network models of COVID-19
explain inequities and inform reopening.

If you are writing about or interested in our work, we highly recommend reading
our [FAQ](#) for the main takeaways and for non-technical answers to common
questions!

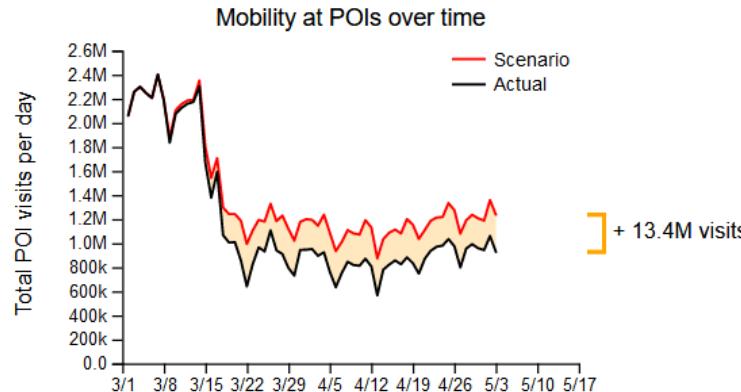


<https://covid-mobility.stanford.edu/>

SIMULATION

What if we hadn't socially distanced? What was the impact of mobility to different types of places on the overall epidemic curve? Find out in our interactive simulation below!

MSA: San Francisco Chicago New York



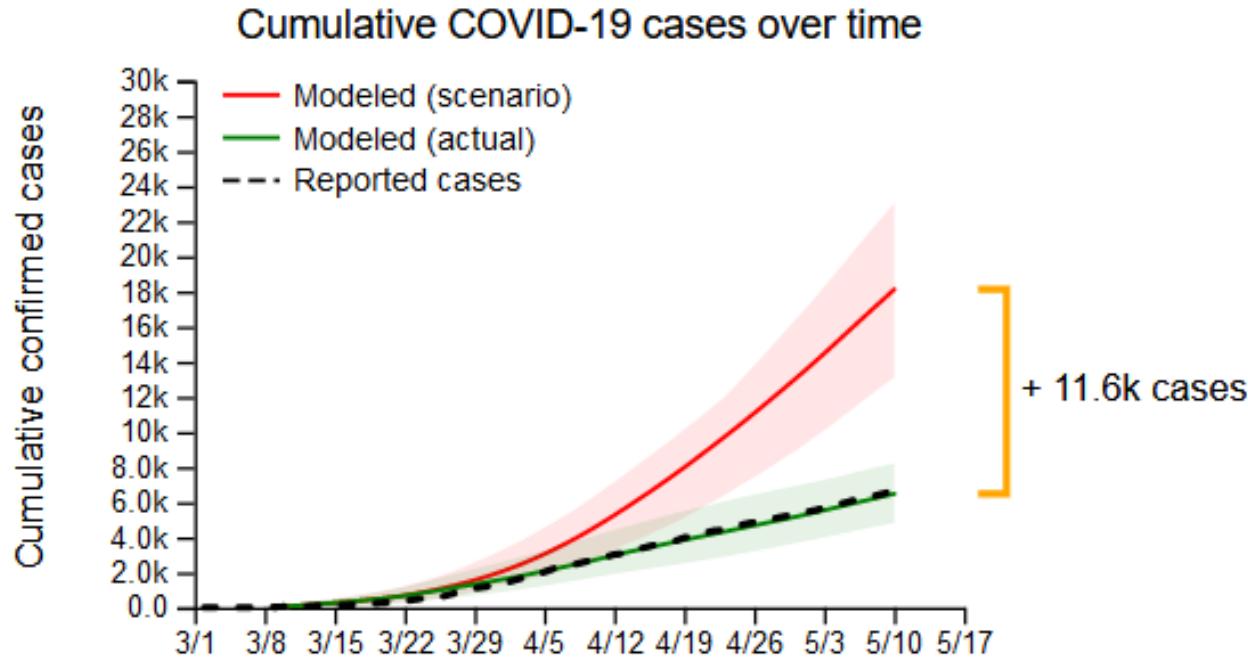
This graph illustrates the amount of mobility you would have had under your scenario (red), compared to what actually happened (black). Mobility is represented as daily total number of visits to points of interest (any non-residential place), based on aggregated geolocation data from SafeGraph.

SHOW MOBILITY AND INFECTIONS WITH THE FOLLOWING OPTIONS:

- Full-service restaurants
- Fitness centers
- Takeout restaurants
- Religious organizations
- Grocery stores
- Department stores
- Pharmacies

Scenario: What if we hadn't reduced mobility at full-service restaurants?

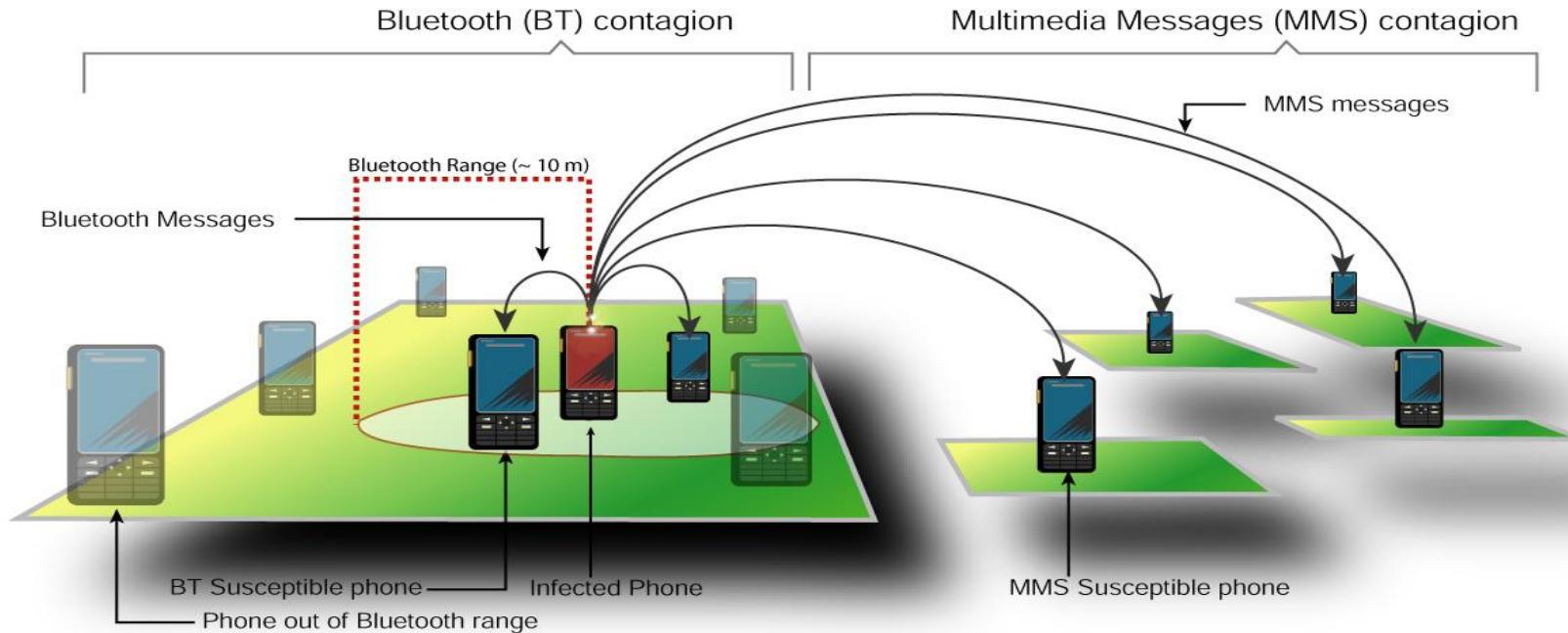
<https://covid-mobility.stanford.edu/>



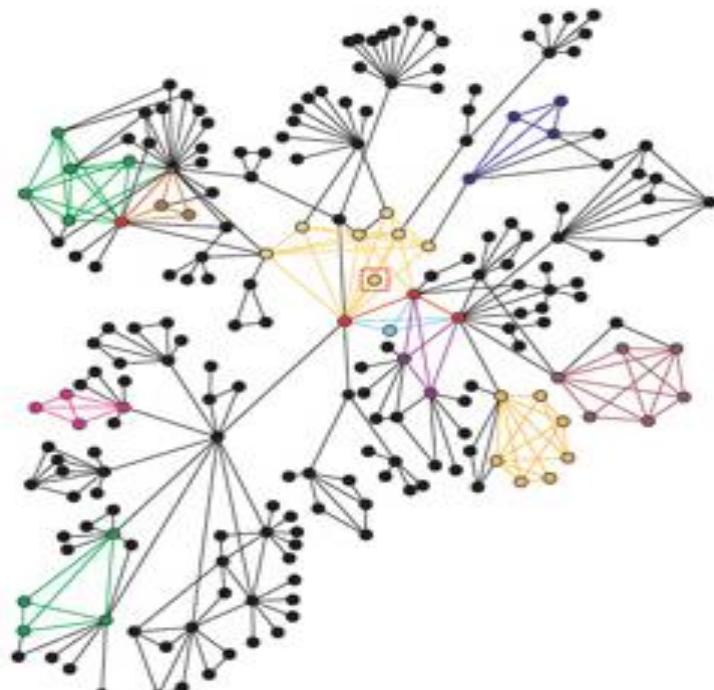
This graph depicts the cumulative COVID-19 case counts predicted by our model under your scenario (red), compared to our model predictions when run with actual mobility data (green), which closely track real case counts (as reported by The New York Times).

<https://covid-mobility.stanford.edu/>

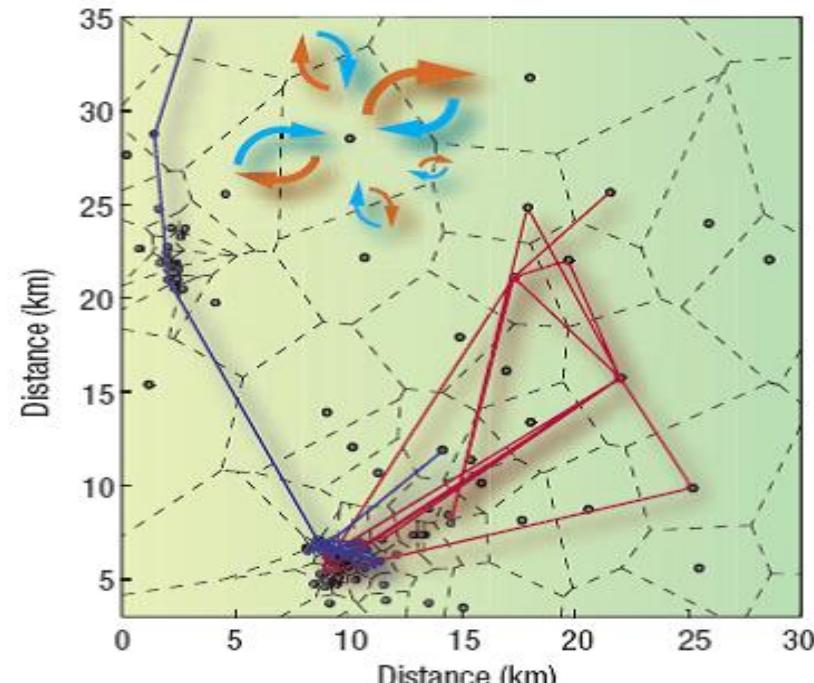
¿Cómo se transmiten los virus de móviles? Virus Bluetooth y MMS



Virus Bluetooth y MMS:



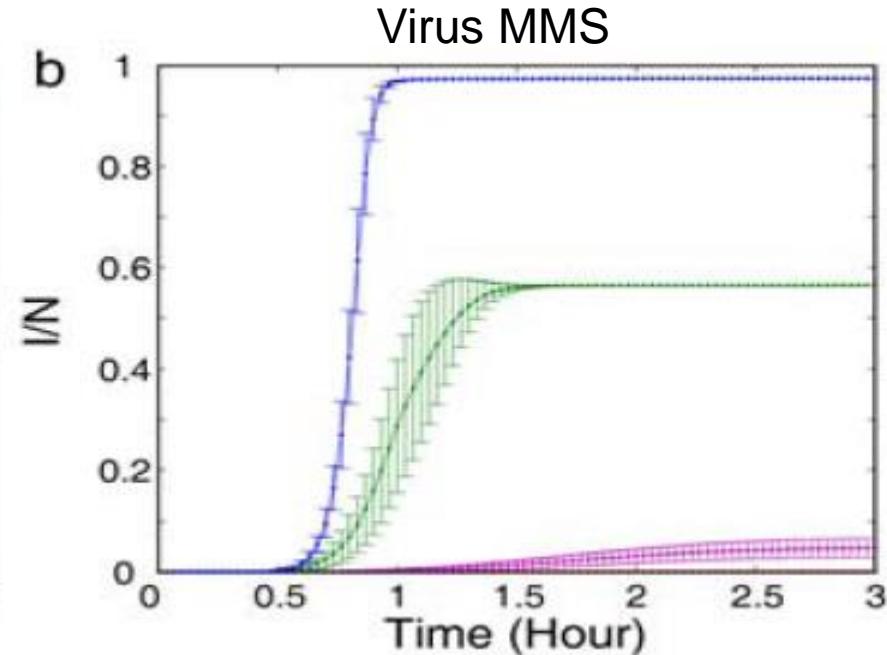
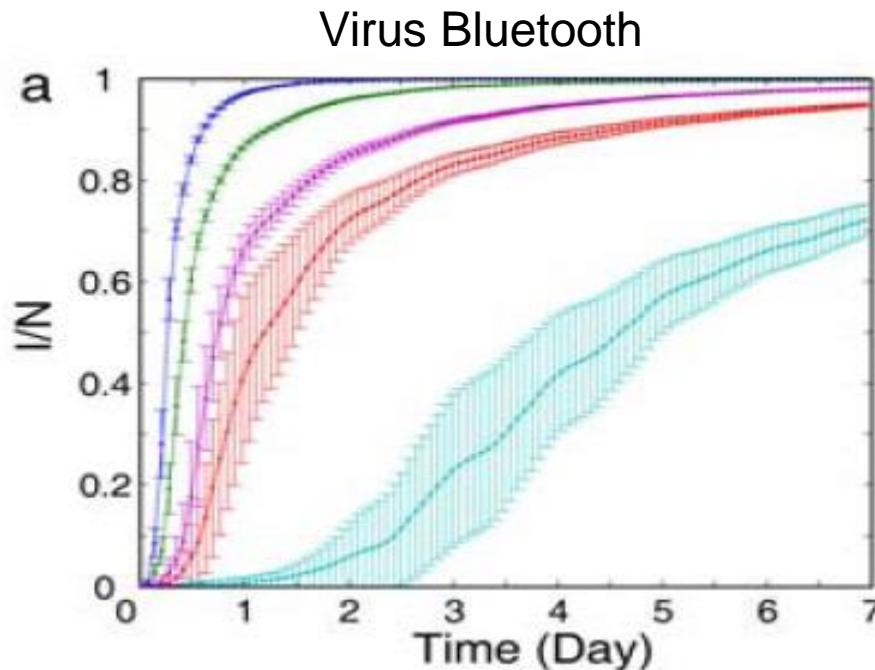
Red social (virus MMS)



Movilidad humana (virus Bluetooth)

Onella et al. PNAS (2007). Palla et al. Nature 446: 664 (2007)
González, Hidalgo y Barabasi. Nature 453: 779 (2008)

Patrones de difusión espacial de los virus Bluetooth y MMS:



Guiado por la movilidad humana:

Lento, pero puede alcanzar a todos los usuarios con el tiempo

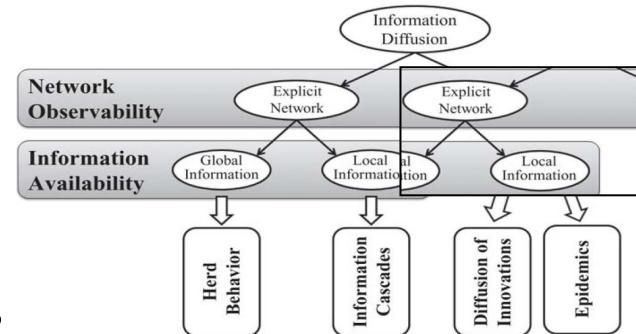
Guiado por la red social:

Rápido, pero sólo puede alcanzar una fracción finita de usuarios (la componente gigante)

OTROS MODELOS DE DIFUSIÓN DE INFORMACIÓN EN REDES

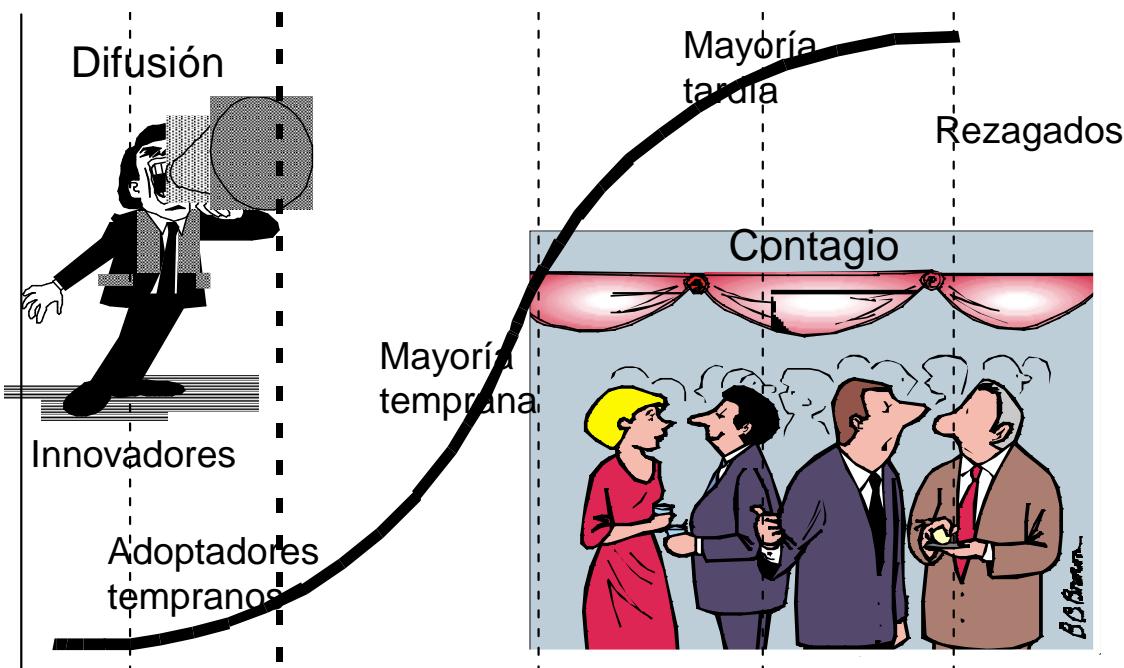
**Easley y Kleinberg. Cascading Behavior in Networks. Networks, Crowds, and Markets:
Reasoning about a Highly Connected World (Cap. 19). Cambridge University Press, 2010**

- Los modelos de difusión de información 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)
- En los modelos epidémicos y de cascada *centrados en el emisor*, los individuos no toman la decisión por si mismos. En el resto sí
- Los **contagios** pueden ser **simples** (individuales por probabilidad) y **complejos** (umbrales)



DIFUSIÓN DE LA INNOVACIÓN

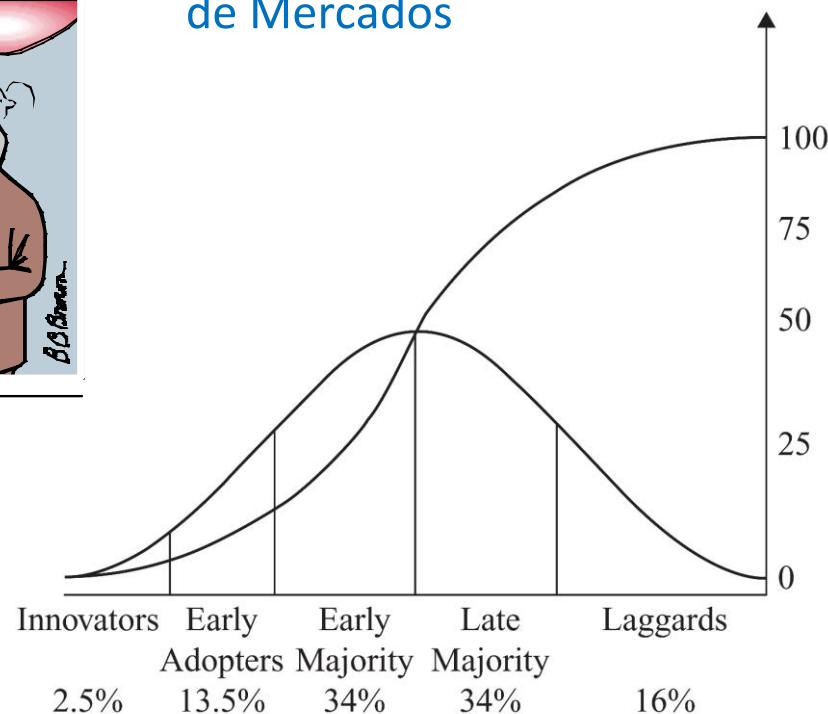
Curva de Adopción



La **Teoría de la Difusión de Innovaciones** estudia cómo y por qué se difunden esas innovaciones

Una **innovación** es una idea, práctica u objeto percibido como nuevo por un individuo

Muy estudiado en **Investigación de Mercados**



- 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)

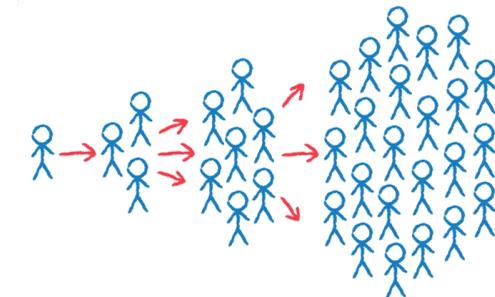
β : factor de imitación (influencia de los adoptadores, boca-a-oreja, local)

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 consideran los adoptadores de entre los contactos del nodo (red social personal). Están centrados en el emisor

- En los medios sociales, los individuos suelen reenviar el contenido publicado por otros usuarios, recibido directamente de los vecinos directos (*amigos*)
- Las **cascadas de información** se producen cuando **la información se propaga únicamente a través de los amigos**
- No hay ninguna información global externa que influya en la propagación (p.ej. la viralidad de un tweet (*trending topics*), la publicidad de un producto, ...)
- Las decisiones son binarias, los nodos pueden ser activos o inactivos
- Los nodos vecinos **influyen** al individuo para que adopte el comportamiento, la innovación o la decisión
- **Hay modelos centrados en el emisor y en el receptor**



1996-97: Hotmail fue uno de las primeras compañías en usar marketing viral con gran éxito

Sólo tuvo que añadir la frase “*Get your free Email at Hotmail*” en el pie de cada correo enviado por sus usuarios

Consiguió **12 millones de usuarios en 18 meses**

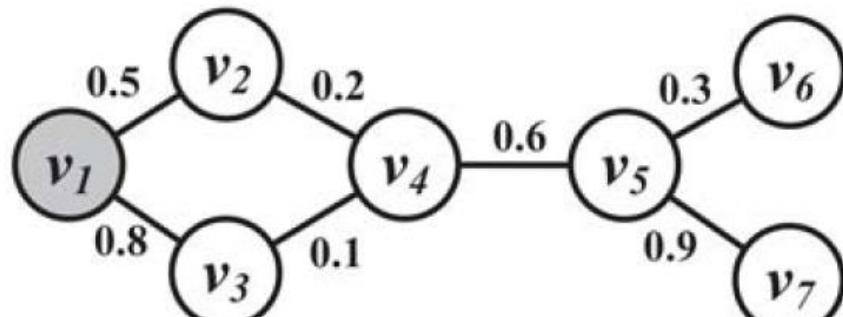
Fue el crecimiento mayor y más rápido de una empresa basada en usuarios en la época. Cuando alcanzó los 66 millones de usuarios, **la compañía creaba 270.000 nuevas cuentas cada día**



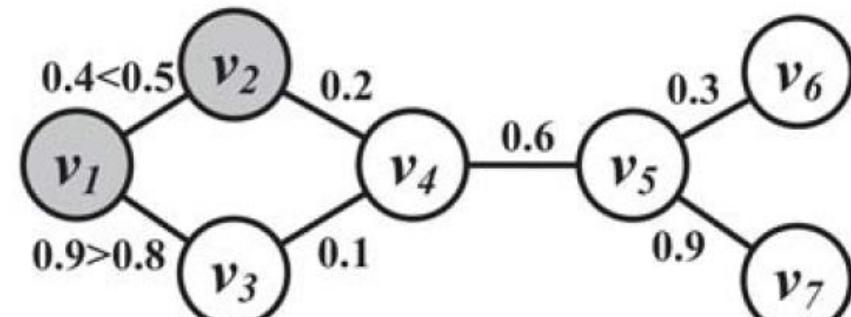
Hotmail®

--
Get your free Email at [Hotmail](#)

- El **Independent Cascade Model** es uno de los modelos más conocidos
- Es un **modelo centrado en el emisor** en el que cada individuo tiene una probabilidad de influir a cada uno de sus vecinos
- Una vez activado, el nodo puede influir a sus vecinos **una sola vez**, en un proceso progresivo
- **En principio**, no existe la posibilidad de desactivarse



Step 1

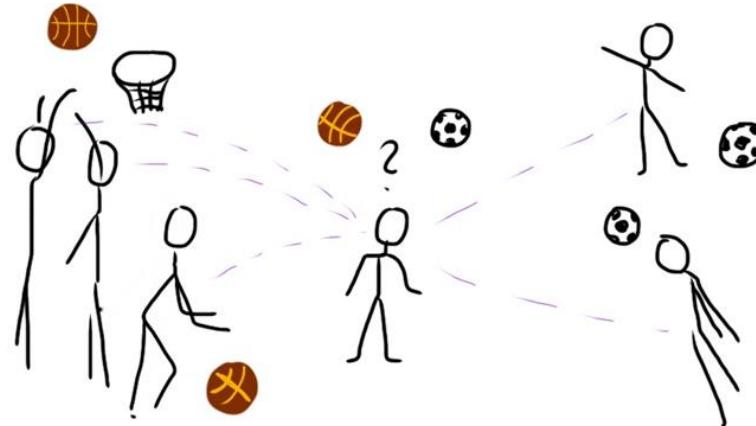


Step 2

- **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)
- **Modelos de umbral:** la adopción requiere superar un umbral asociado a un porcentaje de contactos infectados (20%, 30%, ...), que puede ser común a toda la población o particular de cada agente
- Pueden incluir información sobre las ventajas de la adopción (**recompensas**). Ej: cada individuo tiene d amigos y debe escoger entre dos opciones (p.ej. futbol o baloncesto)
- Si sus amigos escogen B , el individuo recibe una recompensa r_b . Si escogen F , la recompensa es r_f

Ejemplo: $d=5$

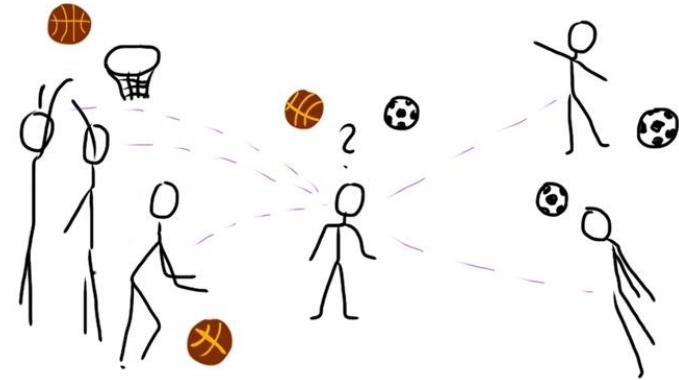
Una ratio $p = 3/5$
juegan al
baloncesto



Una ratio $1-p = 2/5$
juegan al fútbol

¿Qué elección proporciona una mayor recompensa?

- Un actor tiene d vecinos
- Una ratio p juegan al baloncesto (B)
- Una ratio $1 - p$ juegan al fútbol (F)
- Si escoge B , recibe $p \cdot d \cdot r_b$
- Si escoge F , recibe $(1 - p) \cdot d \cdot r_f$
- Así, escogerá B si: $p \cdot d \cdot r_b \geq (1 - p) \cdot d \cdot r_f$ o $p \geq \frac{r_f}{r_b + r_f}$ (**umbral adopción q**)



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

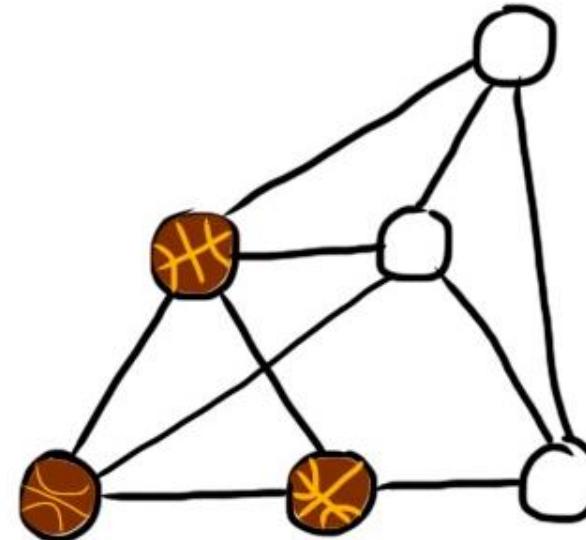
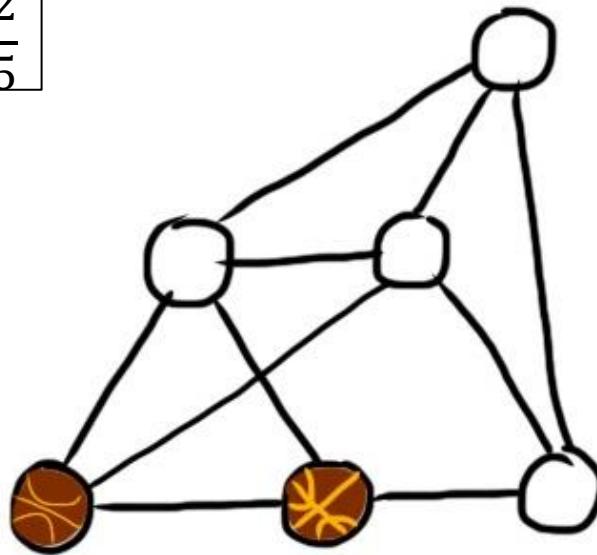
¿Qué ocurre en el proceso intermedio?

- ¿Qué pasa si dos nodos cambian su estado de forma aleatoria? ¿Se producirá una **propagación en cascada**?
- Ejemplo:
 - Valores de las recompensas: $r_b = 3, r_f = 2$
 - La recompensa para la interacción de dos nodos con comportamiento B es $3/2$ veces mayor que la que obtendrían si ambos escogieran F
 - Los nodos cambiarán de F a B si al menos $q = \frac{r_f}{r_b+r_f} = \frac{2}{3+2} = \frac{2}{5}$ de sus vecinos están jugando a B (**umbral de adopción q de B**)

¿Cómo se produce una propagación en cascada?

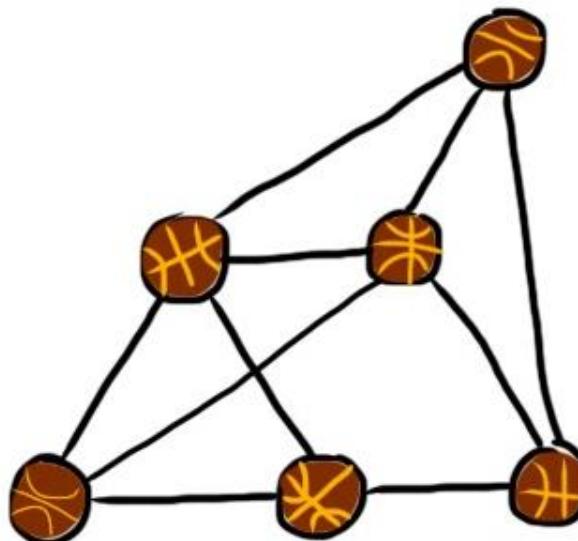
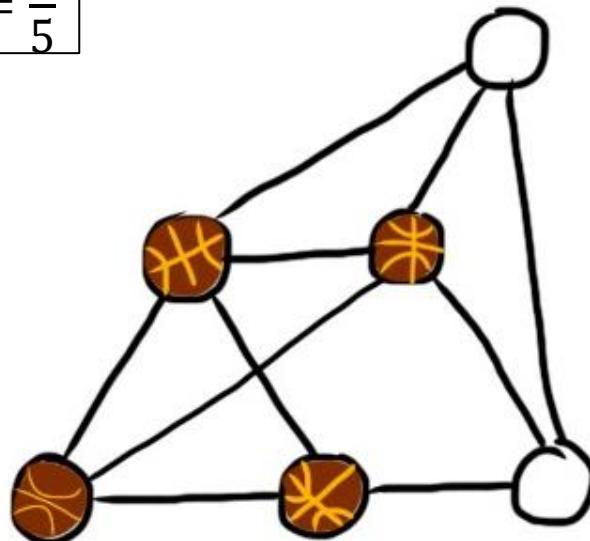
- Supongamos que dos nodos empiezan a jugar baloncesto:

$$q = \frac{2}{5}$$



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

$$q = \frac{2}{5}$$



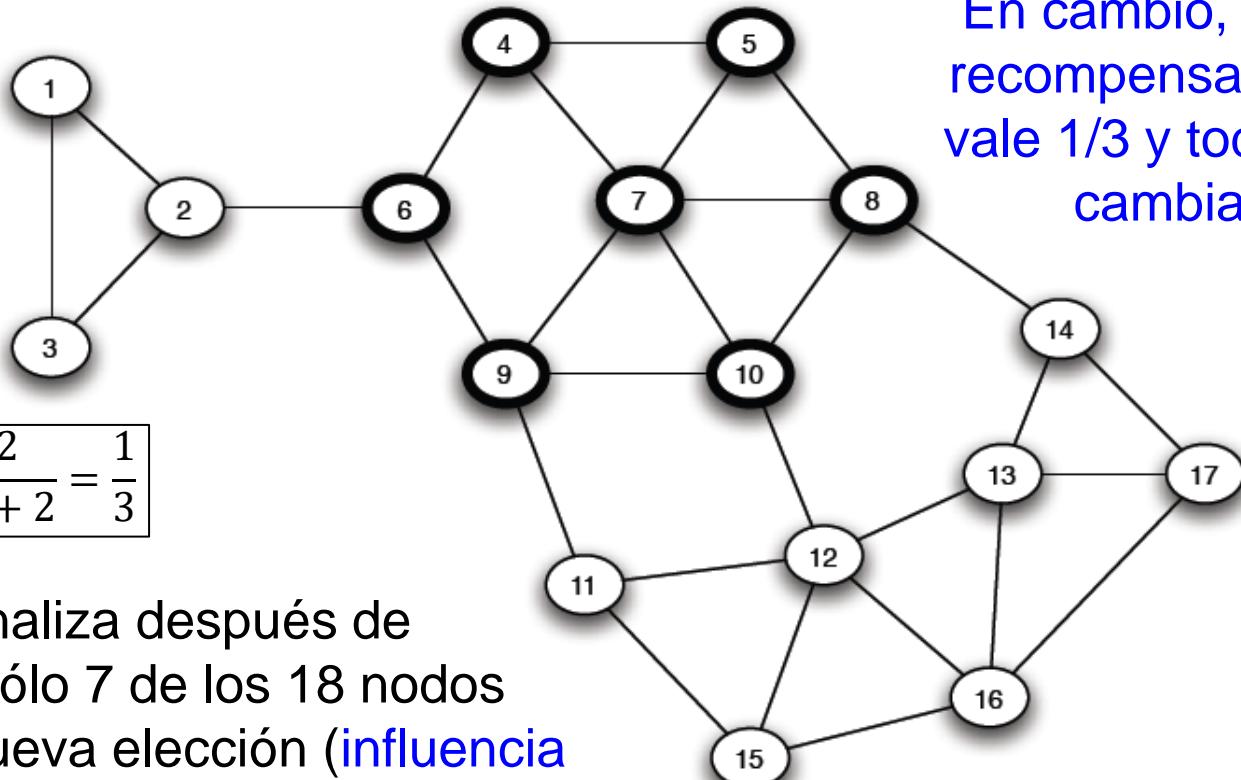
Otro ejemplo con los mismos parámetros en el que no se converge a una única elección (no se produce una cascada completa)

$$q = \frac{2}{5}$$



$$q = \frac{r_f}{r_b + r_f} = \frac{2}{4 + 2} = \frac{1}{3}$$

El proceso finaliza después de tres pasos. Sólo 7 de los 18 nodos adoptan la nueva elección (influencia de la estructura de comunidades)



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

odos son los
ores iniciales
de la nueva elección

En procesos epidémicos, las intervenciones van orientadas a interferir con la propagación del patógeno en la red de contactos para retrasarla/detenerla

En otros procesos de difusión/adopción podemos estar interesados en **acelerar la propagación**. Un ejemplo claro es el **marketing viral (WoM)**

Se basa en que las conversaciones de consumidores sobre un producto son una herramienta más poderosa que la publicidad tradicional

Cuando no se puede aumentar la recompensa (calidad/precio del producto), una estrategia habitual es “recompensar” a unos pocos individuos (con un pequeño presupuesto de campaña) para promocionar el producto entre sus amigos **buscando una adopción grande y rápida** (una cascada que redunde en un aumento de ventas)



El problema consiste en decidir **qué individuos escoger** para maximizar cantidad y ratio de adopción del producto. Es similar a la decisión de qué individuos vacunar pero a la inversa

Cuantos más se recompensen en la campaña promocional, mayor será la velocidad de adopción pero también mayor el coste de la campaña

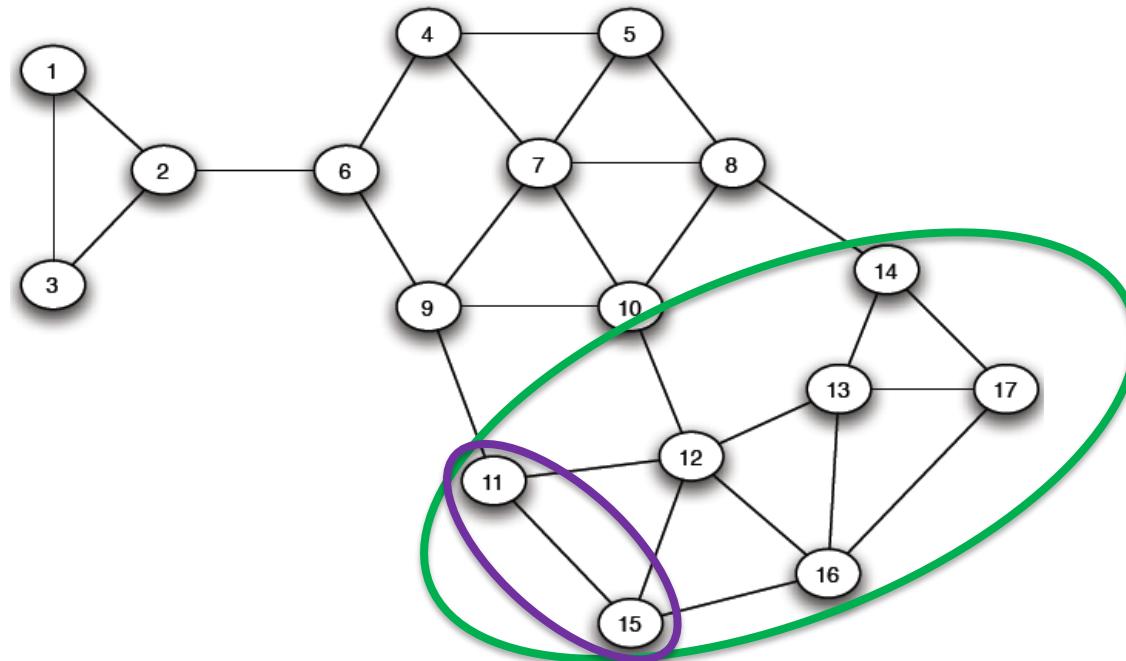
Los hubs difunden virus o ideas débilmente infecciosas.
Son una buena opción como semillas pero no la única
(estructura de comunidades)

Los **actores influyentes de la red** (**influentials**) según distintas medidas de análisis de redes sociales son una buena opción



- La elección de los individuos que inician el cambio (semillas) 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

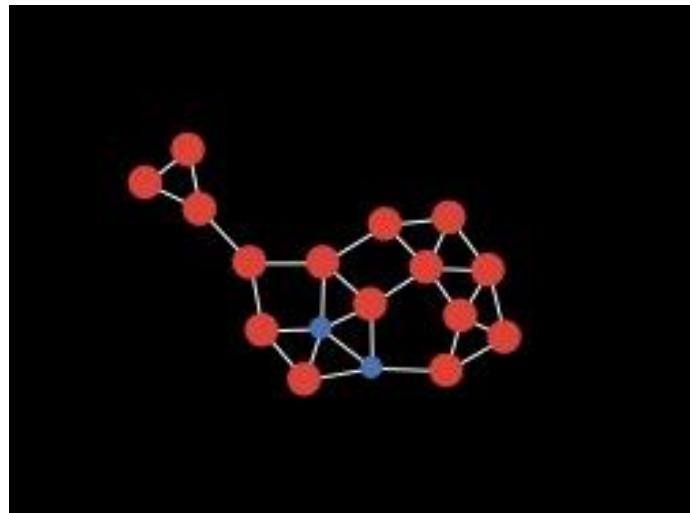
$$q = \frac{2}{5}$$



- La elección de los individuos que inician el cambio (semillas) es clave para producir el efecto cascada (actores influyentes de la red – *influentials*)

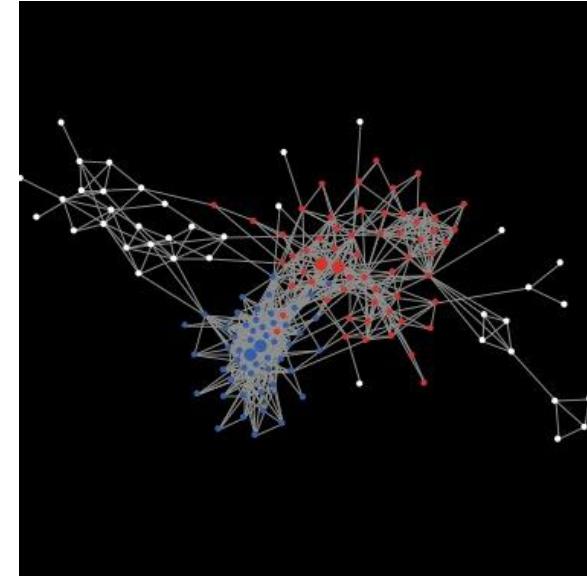
<http://www-personal.umich.edu/~ladamic/netlearn/NetLogo412/CascadeModel.html>

¿Se producirá una propagación en cascada?



<http://www-personal.umich.edu/~ladamic/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



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

PREGUNTA: ¿Cuál es el papel de las comunidades en los procesos de contagio complejo?

- a) Posibilitar que las ideas se difundan en presencia de umbrales
- b) Crear “bolsas aisladas” impermeables a las ideas externas
- c) Permitir que distintas opiniones coexistan en distintas partes de la red

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.
Northeastern University: <http://networksciencebook.com/>
- Curso on-line “Social Network Analysis” de Lada Adamic, Coursera,
Universidad de Michigan
- Transparencias del Capítulo 7 del libro “Social Media Mining” de
R. Zafarini, M.A. Abassi y H. Liu. <http://socialmediamining.info/>



Northeastern University
Center for Complex Network Research

