Propagación de malware: un modelo basado en SEDOs

Ángel Martín del Rey
Departamento de Matemática Aplicada
Instituto de Física Fundamental y Matemáticas
Universidad de Salamanca, Salamanca, España

delrey@usal.es

Introducción

- El código malicioso (*malware*) es una de las principales amenazas a la que se ven sometidos los sistemas dependientes de las tecnologías de la información.

Introducción

Modelización Matemática: propagación de malware

- El diseño de modelos matemáticos para simular la propagación del malware es una de las principales medidas para su control.

- El propósito fundamental de estos modelos es el siguiente:
  - Entender los mecanismos de propagación del malware.
  - Predecir el impacto, tanto cualitativo como cuantitativo, de un brote de malware antes de que ocurra.
  - Evaluar posibles medidas de contención de la epidemia.

- Tipos de modelos teniendo en cuenta su naturaleza:
  - Deterministas vs. estocásticos.
  - Globales vs. individuales.
  - Continuos vs. discretos.

El Modelo de Kermack-McKendrick: Historia

- El objetivo de este taller es presentar el modelo de Kermack-McKendrick para estudiar la propagación de malware.

- El modelo de Kermack-McKendrick fue inicialmente desarrollado en 1927 para estudiar la propagación de la peste bubónica y es considerado como la piedra angular de los modelos que se propusieron posteriormente.

- Se trata de un modelo compartimental SIR en el que la población se divide en dispositivos susceptibles, infecciosos y recuperados.

El Modelo de Kermack-McKendrick: el SEDO

- La dinámica del modelo viene regida por el siguiente sistema de ecuaciones diferenciales ordinarias:

$$\begin{cases} \dfrac{dS}{dt} = -a \cdot S(t) \cdot I(t) \\ \dfrac{dI}{dt} = a \cdot S(t) \cdot I(t) - b \cdot I(t) \\ \dfrac{dR}{dt} = b \cdot I(t) \end{cases}$$

La variación del número de susceptibles es proporcional a dicho número de susceptibles

La variación del número de infectados es el balance entre una proporción de susceptibles y una de infectados

La variación del número de recuperados es proporcional al número de infectados

Coeficiente de transmisión: $a = k \cdot q$, Tasa de recuperación: $b = T^{-1}$,

k : contactos efectivos con infectados por unidad de tiempo,

q : probabilidad de que un contacto efectivo acabe en contagio,

T : duración del periodo infeccioso.

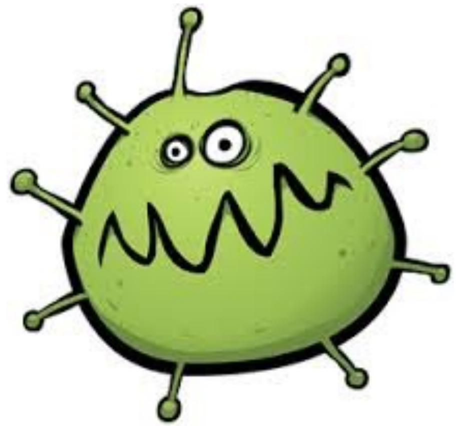
El Modelo de Kermack-McKendrick: el R_0

- El número reproductivo básico asociado al modelo propuesto es:

$$R_0 = \frac{a \cdot N}{b}.$$

de manera que se verifica que:

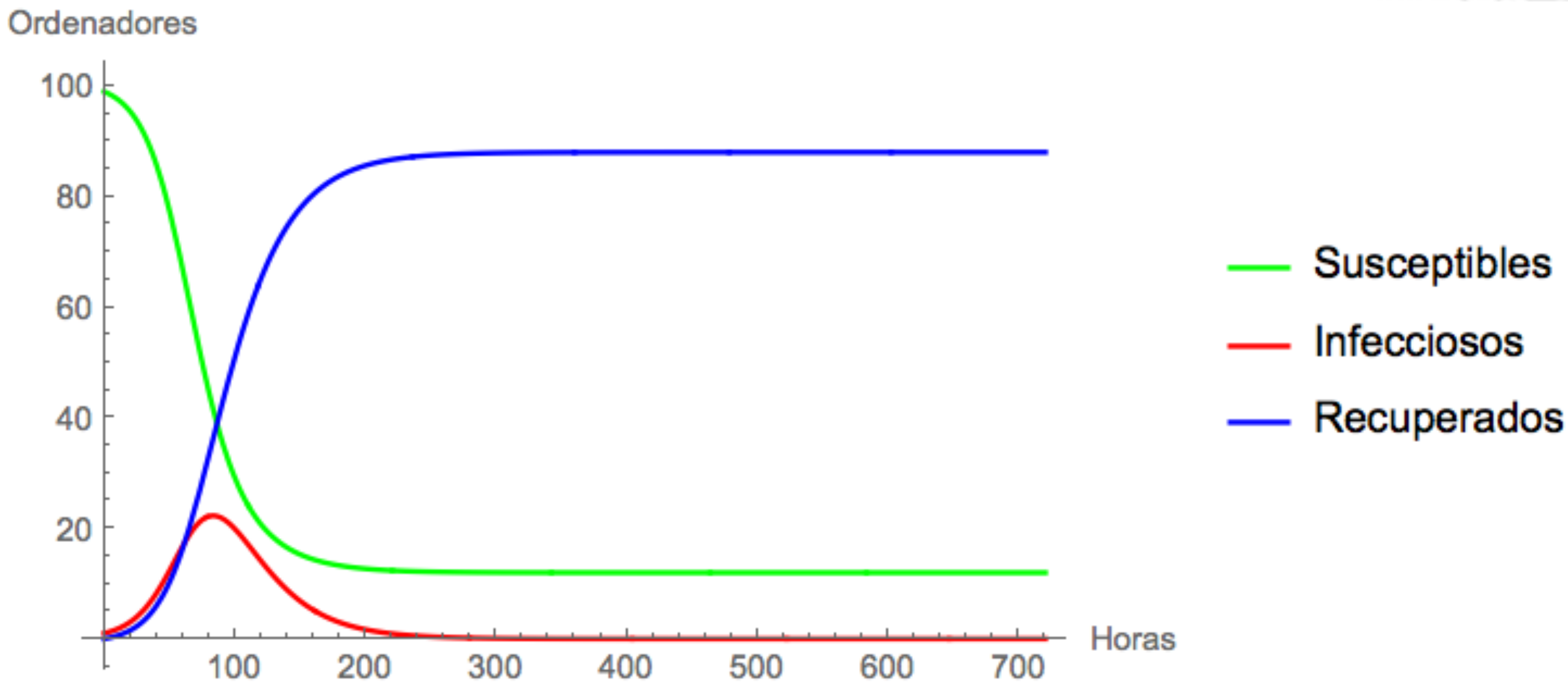
- Si $R_0 < 1$ no se producirá una epidemia.
- Si $R_0 > 1$ se producirá una epidemia.



Ángel Martín del Rey, 2015

El Modelo de Kermack-McKendrick: el R_0

Simulación I: $R_0 > 1$

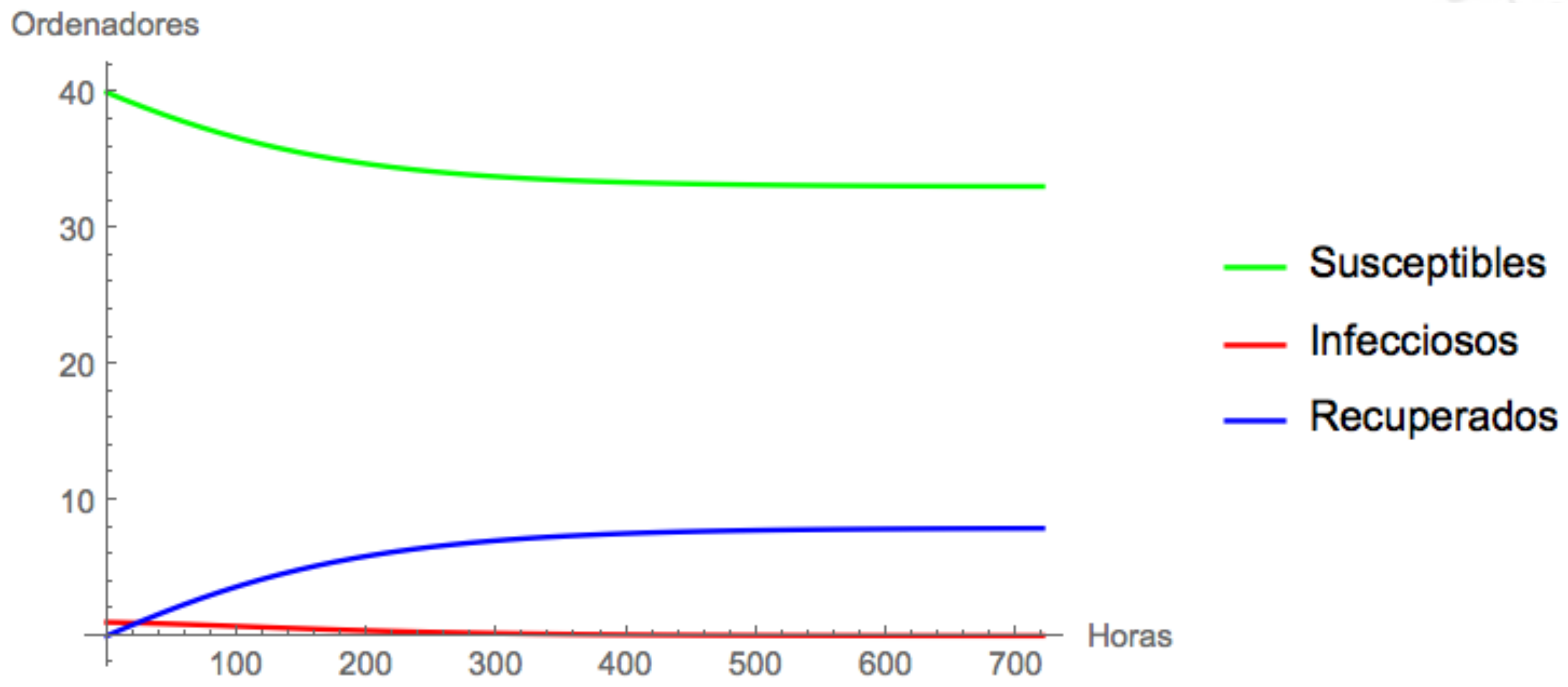


$$a = 0.001, \quad b = \frac{1}{24}, \quad S(0) = 99, \quad I(0) = 1, \quad R(0) = 0,$$

$$R_0 \approx 2.3760.$$

El Modelo de Kermack-McKendrick: el R_0

Simulación II: $R_0 < 1$



$$a = 0.001, \quad b = \frac{1}{24}, \quad S(0) = 30, \quad I(0) = 1, \quad R(0) = 0,$$

$$R_0 = 0.96.$$

El Modelo de Kermack-McKendrick: el R_0

- El número reproductivo básico se puede definir como el número esperado de dispositivos infectados causados por el contagio de un único dispositivo infeccioso en una población enteramente susceptible:

$$R_0 = \frac{a \cdot N}{b} = \frac{q \cdot k \cdot N}{1/T} = q \cdot k \cdot T \cdot N.$$

$k \cdot T =$ contactos efectivos de cada dispositivo con el dispositivo infectado durante todo su periodo infeccioso

$q \cdot k \cdot T =$ contactos efectivos de cada dispositivo con el infectado durante todo su periodo infeccioso y que acaban en contagio

$R_0 = q \cdot k \cdot T \cdot N =$ contactos efectivos totales con el dispositivo infectado durante todo su periodo infeccioso y que acaban en contagio

El Modelo de Kermack-McKendrick: el R_0

- El objetivo en el control de cualquier epidemia sería conseguir que inicialmente el número reproductivo básico fuera menor que 1 y posteriormente, si procediera, que el número reproductivo efectivo fuera también menor que 1.

- ¿Cómo se consigue ello

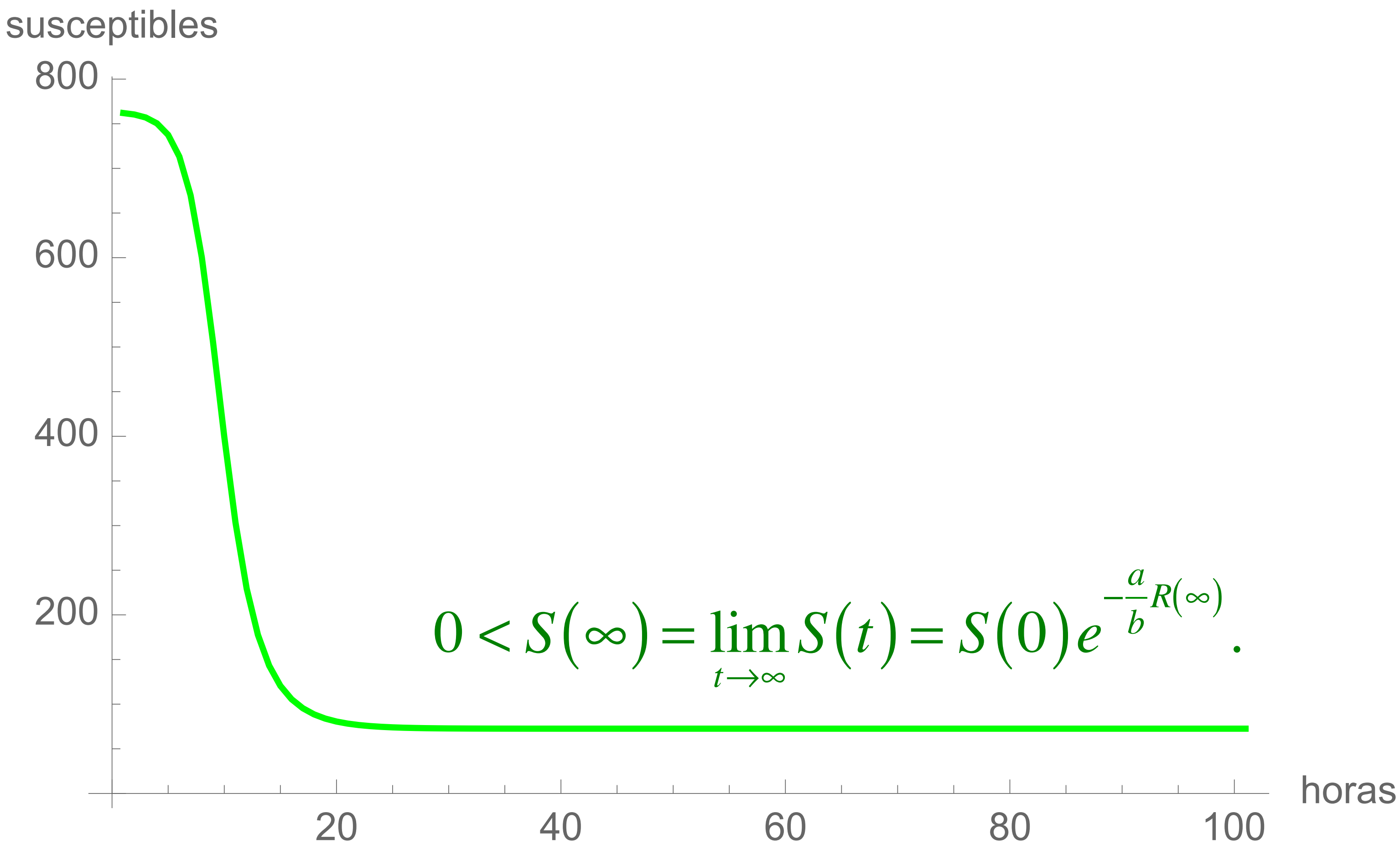
$$R_0 = \frac{a \cdot N}{b}$$

  - Disminuyendo a haciendo disminuir el número de contactos k mediante aislamiento.
  - Disminuyendo a haciendo disminuir la probabilidad q de que un infectado contagie a un susceptible.
  - Disminuyendo N aplicando programas de vacunación.
  - Aumentando b mejorando el tratamiento de los infectados.

El Modelo de Kermack-McKendrick: Soluciones

- Evolución del número de dispositivos susceptibles, $S(t)$:

$$S(t) = S(0) \cdot e^{-\frac{a}{b}R(t)}$$



$$0 < S(\infty) = \lim_{t \to \infty} S(t) = S(0) e^{-\frac{a}{b}R(\infty)}.$$

El Modelo de Kermack-McKendrick: Soluciones

- Evolución del número de dispositivos infectados, $I(t)$:

  - $I(t)$ satisface:

$$I(t)=N-S(t)+\frac{b}{a}\log\left(\frac{S(t)}{S(0)}\right).$$

  - Si $R_0<1$, entonces $I(t)$ es monótona decreciente de manera que:

$$I(\infty)=\lim_{t\to\infty}I(t)=0.$$

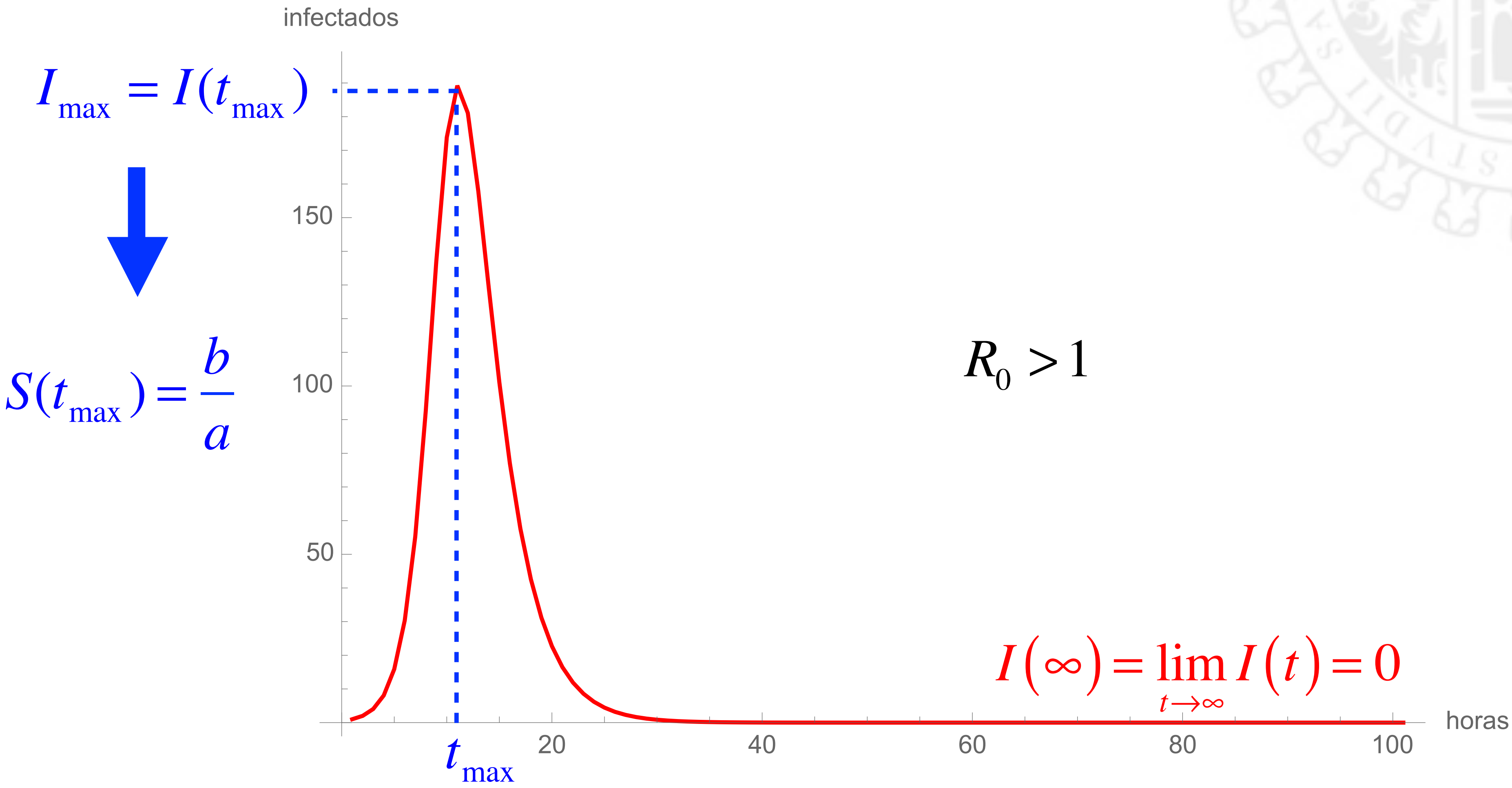
  - Si $R_0>1$, entonces $I(t)$ crece hasta alcanzar su valor máximo:

$$I_{\max}=N-\frac{b}{a}-\frac{b}{a}\log\left(\frac{a\cdot S(0)}{b}\right),$$

    y posteriormente decrece de manera que: $I(\infty)=\lim_{t\to\infty}I(t)=0.$

El Modelo de Kermack-McKendrick: Soluciones

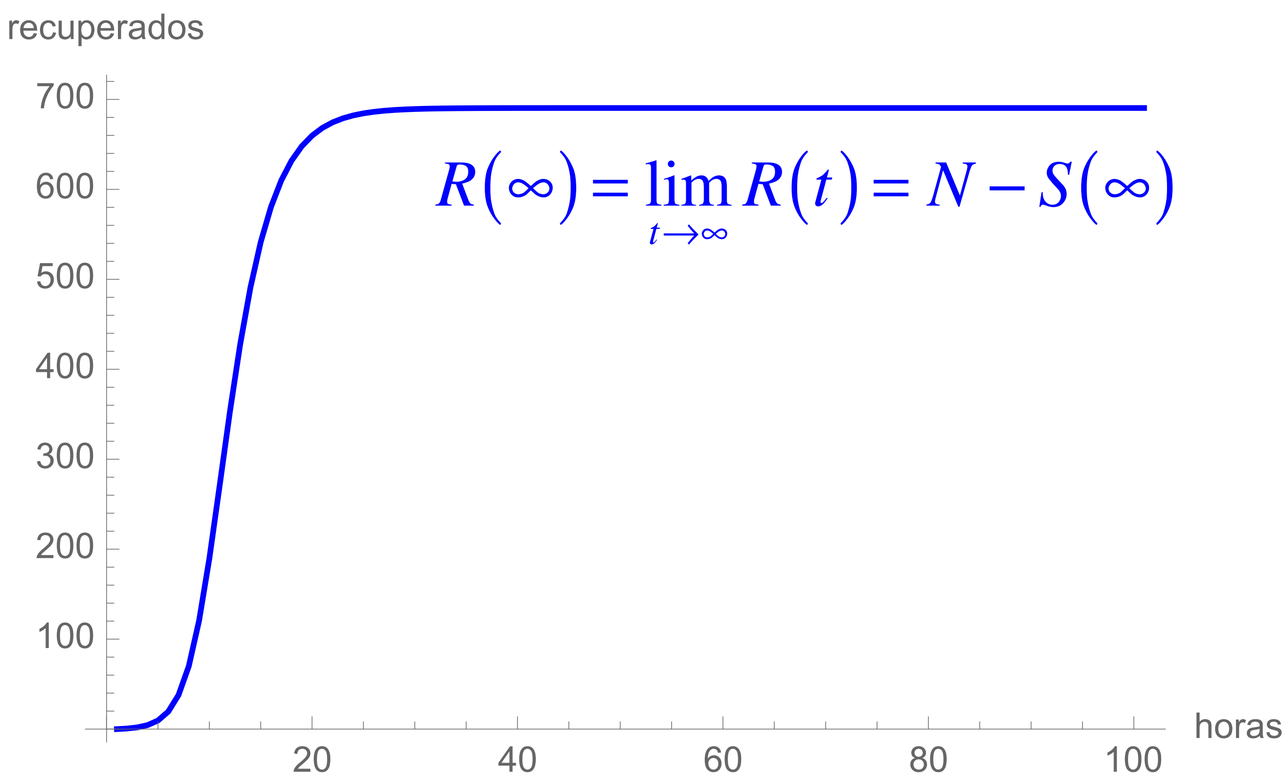
- Evolución del número de infectados, $I(t)$:



El Modelo de Kermack-McKendrick: Soluciones

- Evolución del número de dispositivos recuperados, $R(t)$:

$$R(t) = N - S(t) - I(t) = -\frac{b}{a}\log\left(\frac{S(t)}{S(0)}\right).$$



Algunas conclusiones

- La mayor parte de los modelos se encuentran basados en SEDOs (son deterministas, globales y continuos).

- Aunque desde el punto de vista matemático se encuentra bien fundamentados y gracias a la Teoría Cualitativa de SEDOs es posible analizarlos en detalle, presentan serios inconvenientes:

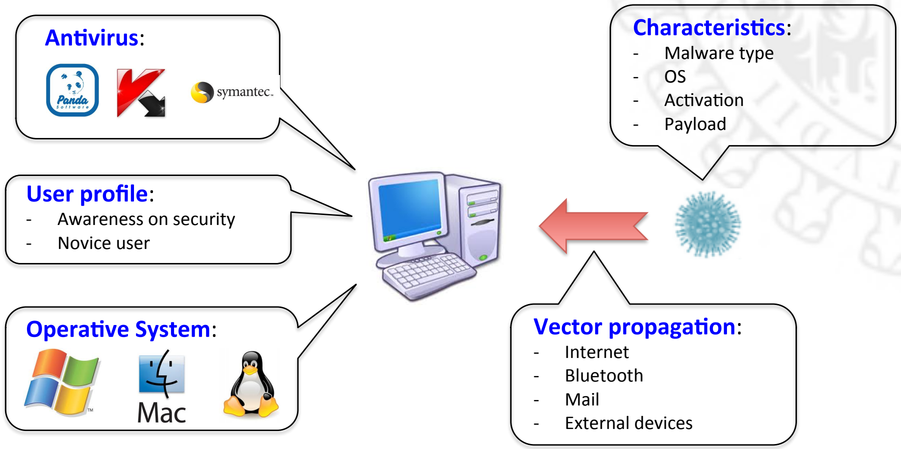
  - No tienen en cuenta las características individuales de los dispositivos.

  - La topología del sistema es homogénea.

  - No son capaces de predecir el comportamiento individual.

Algunas conclusiones

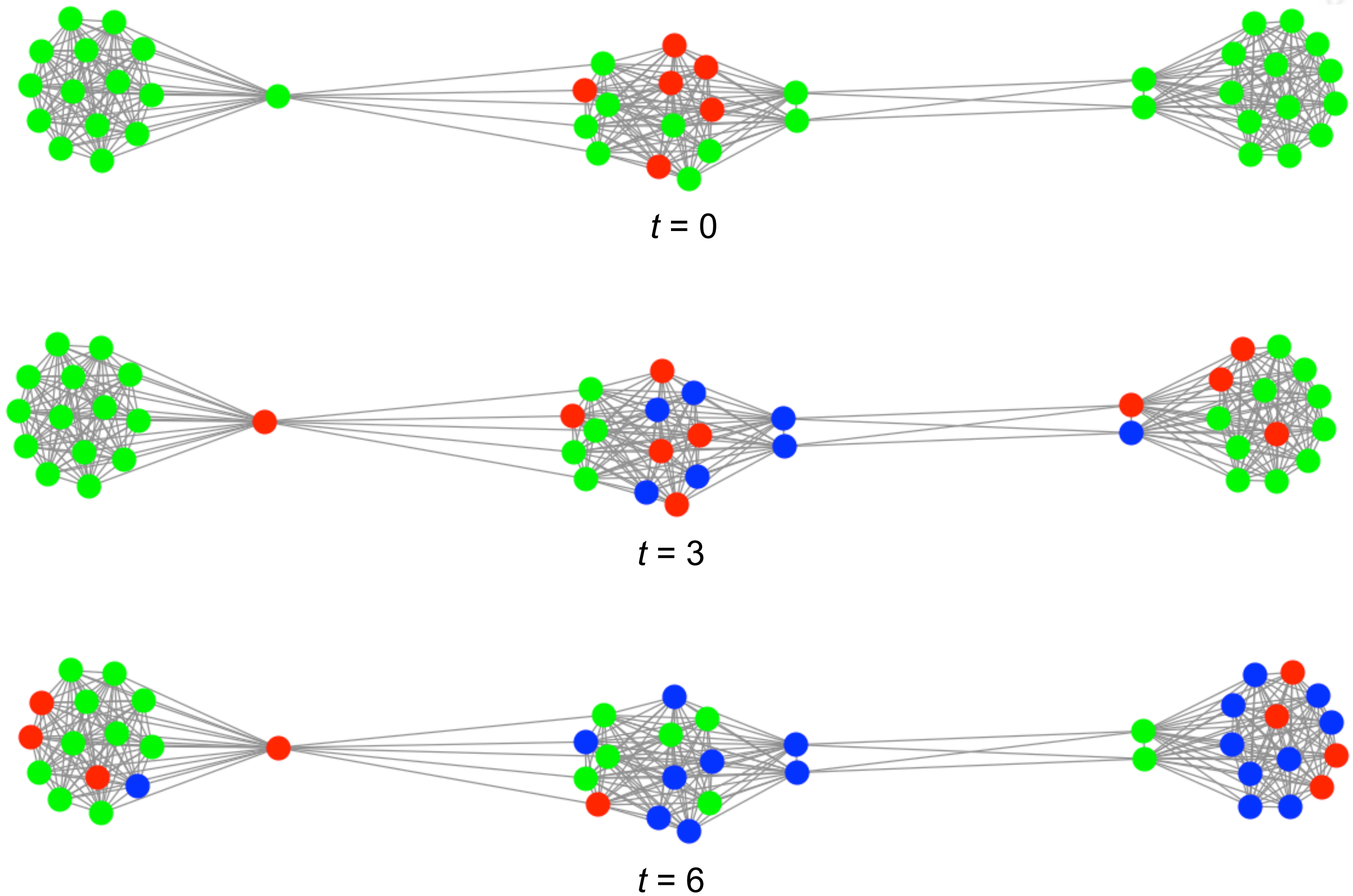
- Estos inconvenientes se pueden solventar con el uso de modelos individuales:
  - Autómatas Celulares.
  - Modelos Basados en Agentes.



- Este enfoque presenta algunos problemas:
  - Es necesario tener información de todos los actores del sistema.
  - Complejidad computacional.
  - Análisis matemático del comportamiento del sistema.

Algunas conclusiones

- Los modelos individuales nos permiten realizar simulaciones tanto a nivel individual como a nivel global:



¡Muchas gracias por vuestra atención!

¿alguna pregunta o comentario?