**ESTUDIO DE DISTINTOS MODELOS DE CONTAGIO SOBRE UNA RED DE AEROPUERTOS**

GRUPO 3

EDERSON ALDAIR FUNES CASTILLO

JOSE LUIS GOMEZ ALONSO

ZHIHAO ZHENG

# Índice

1. Descripción del Proyecto
2. Acerca de la Red
3. Acerca de los Casos Base
4. Acerca de los Modelos/Algoritmos
5. ¯\\_(ツ)\_/¯
6. Simulaciones con Distintas Tasas de Infección
7. Simulaciones con Distintas Tasas de Recuperación
8. Simulaciones con Distintas Cuarentenas
9. Simulaciones con Distinto Paciente Cero
10. Resumen de Conclusiones
11. Bibliografía

# Descripción del Proyecto

Nuestro proyecto final consiste en simular los distintos modelos de contagio (SI, SIS y SIR) sobre una red real de aeropuertos, estudiar la evolución de estos modelos a lo largo del tiempo, y comparar los resultados obtenidos en las simulaciones con los resultados esperados de aplicar las formulas ya definidas para estos modelos.

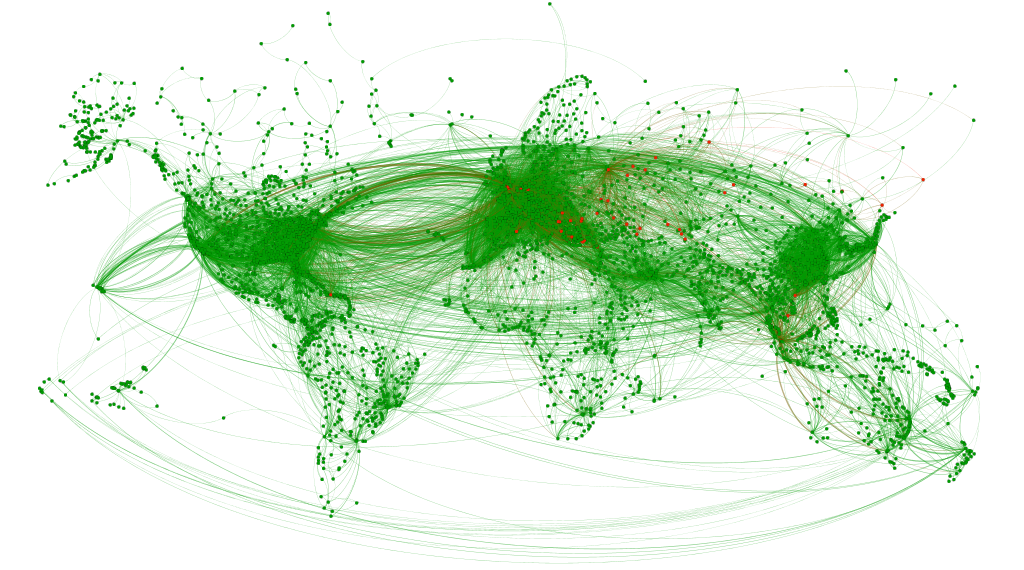
Además, estudiaremos como afecta la introducción de distintos eventos (eventos aleatorios, posibles cuarentenas, distintos focos desde los cuales empezar la infección, etc.) a la propagación de nuestro agente sobre la red, ya sea extendiendo/disminuyendo el tiempo que nuestro agente tarda en infectar toda la red en el modelo SI, aumentando/disminuyendo la brecha entre la población susceptible e infectada en el modelo SIS, acelerando la extinción de la enfermedad en el modelo SIR, etc.

# Acerca de la Red

La red que decidimos utilizar en nuestras simulaciones es la red mundial de aeropuertos, y los datasets utilizados para fueron obtenidos de <http://openflights.org/data.html>.

El primer dataset (airports.dat) contiene distinta información (País, Ciudad, ID del aeropuerto, etc.) de más de 8000 aeropuertos activos a nivel mundial. Tras filtrar los datos innecesarios, obtuvimos un archivo que solo contenía el ID del aeropuerto, su ciudad, y su latitud y longitud. Estos aeropuertos representan los nodos de nuestra red.

El segundo dataset (routes.dat) contiene información de más de 67000 rutas que conectan los distintos aeropuertos, aunque muchas de estas rutas se repiten en muchas ocasiones. La única información que utilizamos de este dataset fueron los ID de los aeropuertos de origen y destino. Estas rutas representan las aristas que unen los distintos nodos de nuestra red.



# Acerca de los Cases Base

Debido a la gran cantidad de parámetros modificables por el usuario a la hora de simular los modelos de contagio, decidimos generar una simulación de cada modelo con unos parámetros fijos. Estas simulaciones nos servirán como base para estudiar los diferentes efectos que tienen estos parámetros a la hora extender la infección sobre nuestra red.

Los parámetros que utilizamos fueron:

|  |  |
| --- | --- |
| Tasa de Infección | 0.5 |
| Tasa de Recuperación (\*) | 0 / 0.5 |
| Número de Iteraciones | 150 |
| ¿Random Walks? | Si |
| Frecuencia de los Random Walks | 5 |
| ¿Modo Especial? | Si |
| % Necesario Para Activar la Cuarentena | 0.85 |
| Grado Mínimo para entrar en Cuarentena | 75 |
| El 1.º Nodo infectado será… | Aleatorio |

\*Los casos base de los modelos SIS y SIR cuentan con una tasa de recuperación igual a la de infección. Para el modelo SI, esta tasa es igual a 0.

Los parámetros relacionados con los eventos aleatorios (tanto su activación como su frecuencia) y el modo especial son fijos para todas nuestras simulaciones, tanto las base como las que no lo son.

La razón detrás de esta decisión es que los existen nodos de la red que no están conectados al componente gigante, sino que forman sus propias componentes conexas, aisladas del resto. Estos nodos (28 en total) forman un total de 7 componentes conexas pequeñas.

De no existir eventos aleatorios en la red, nuestro patógeno sería incapaz de infectar a todos los nodos de la red.

El parámetro ‘Modo Especial’ condiciona el comportamiento de los eventos aleatorios. Cuando esta desactivado, nuestro algoritmo escogerá de forma aleatoria un nodo de la red e intentará infectarlo, si no está infectado ya, pero cuando está activado, el algoritmo realizara la misma selección aleatoria, pero esta vez sobre una lista que contiene solo nodos susceptibles. Este parámetro es muy importante a la hora de extender nuestro patógeno por la red ya que, de no estar activado, tardaríamos muchísimo más tiempo en infectar los nodos aislados, por que tendríamos que esperar que la selección aleatoria no solo escoja un nodo susceptible, sino que este forme parte de una de las componentes conexas aisladas de la componente gigante.

Otro parámetro fijo en todas nuestras simulaciones es el grado mínimo que deben de tener los nodos para ser considerados para la cuarentena. Escogimos el número 75 porque el número de nodos que superaba esta barrera (110 nodos) suponía un 3.3 % del total, y muchos de ellos eran aeropuertos ubicados en Europa y EE. UU, los cuales creemos que serían algunos de los primeros en cerrarse en caso de brote epidémico.

# Simulaciones con Distintas Tasas de Infección

Para estudiar la velocidad de la propagación del contagio bajo diferentes tasas de infección hemos decidido realizar simulaciones con los modelos SI, SIS y SIR. Y hemos tomado los siguientes datos como una base, cambiando solo la tasa de infección para poder realizar una comparación horizontal entre diferentes tasa de infección.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Tasa Infec. | Tasa Recup. | Nº Iter. | RW? | Frec. RW | Modo Especial | Grado Mínimo para Cuarentena | 1º Nodo Infectado |
| 0.5 | 0 / 0.5 | 150 | Si | 5 | Si | 75 | Aleatorio |

Las columnas representan la tasa de infección 0.5 que hemos tomado como base, y las líneas son los diferentes parámetros con los que hemos tomado para simulaciones.

**Modelo SI:**

Con este gráfico podemos observar, la tasa de infección está directamente relacionado con la velocidad del contagio, cuando mayor sea la tasa, más rapidez propagara.

La diferencia más significativa esta entre tasa de infección 0.1 y 0.25, ya que con tasa de infección 0.25, a partir de instante de tiempo 4, se propagará de manera brutal, cuando llega al instante de tiempo 9, la 80% de aeropuertos están infectados, sin embargo, con 0.1, se crece muy lenta, hasta instante de tiempo 20, solo ha infectado 80% de aeropuertos.

La tasa de infección entre 0.25-0.9, aunque hay diferencia entre la velocidad de propagación pero es menor.

**Modelo SIS:**

Parecido al modelo SI, la tasa de infección está directamente relacionado con la velocidad de propagación, pero al tener una tasa de recuperación 0.5, los afectados llega a un punto de máximo, y a partir de entonces, los afectados y susceptibles empiezan a variar en un intervalo de (-0.1,+0.1).

Al subir tasa de infección, la velocidad de propagación aumenta proporcionalmente desde 0.5 hasta 0.75, pero no hay mucha diferencia entre 0.75 y 0.9, ambos tienen como limite 40

**Modelo SIR:**

Con los dos gráficos anteriores podemos saber, al cambiar la tasa de infección, en principio el porcentaje de infectados puede aumentar entre 3-4 instante de tiempo, pero se recupera muy rápido en ambos casos, la diferencias serian cuando la tasa de infección es más alta, el número de recuperados crece muy rápido. Y una vez que el número de infectados sea 0, entonces ninguno aeropuerto está infectado, y a partir a ese instante, el grafico no cambia

Podemos observar que dentro del gráfico, al cambiar la tasa de infección, el crecimiento de los aeropuertos infectados puede llegar un punto máximo 0.2, a partir de entonces, el número de infectados se van a recuperando y los susceptibles se infectan poco. Y en instante 15-20, todos los aeropuertos infectados han sido recuperados.

# Simulaciones con Distintas Cuarentenas

Ante la duda de si la cuarentena afectaba o no a la velocidad con la que se transmite nuestro agente sobre la red, decidimos realizar distintas simulaciones sobre los 3 modelos, en los que la cuarentena se activaría en distintas fases del contagio. Las simulaciones se generaron utilizando los siguientes parámetros:

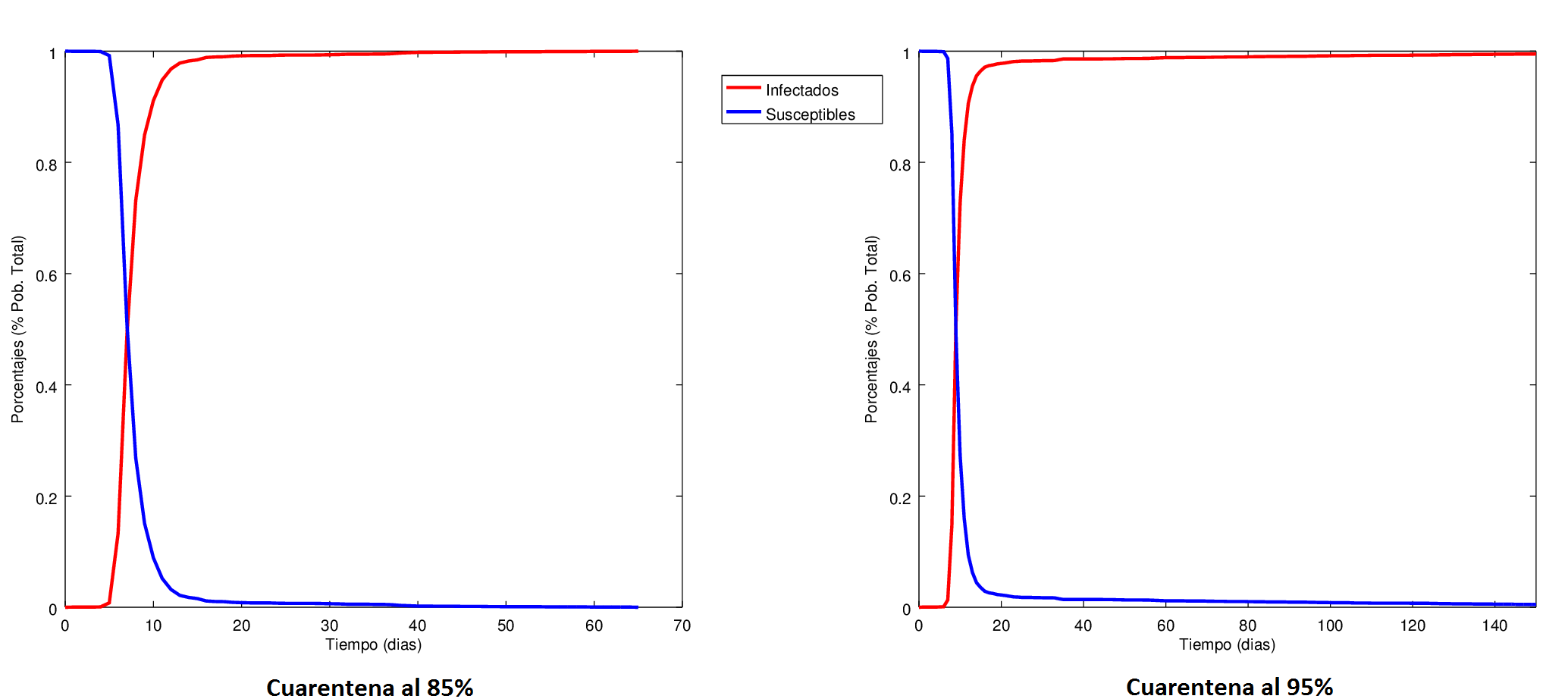
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Tasa Infec. | Tasa Recup. | Nº Iter. | RW? | Frec. RW | Modo Especial | Grado Mínimo para Cuarentena | 1º Nodo Infectado |
| 0.5 | 0 / 0.5 | 150 | Si | 5 | Si | 75 | Aleatorio |

En las simulaciones que usamos como casos base, el porcentaje al cual debía disminuir la población susceptible para que la cuarentena se activase era de un 85%. Las siguientes comparaciones se han realizado con simulaciones en los que este factor tomaba 2 valores: Un porcentaje de un 95%, lo cual provocaba que se activase un poco más pronto que en el caso base, y un porcentaje de 0%, lo cual provoca que la cuarentena nunca llega a activarse.

Los resultados que esperábamos obtener, antes de realizar las nuevas simulaciones, eran distintas evoluciones para los modelos SIS y SIR, ya que en estos pensamos que era posible la erradicación del agente si la cuarentena se activaba de forma más prematura.

1. **Cuarentena al 95%**

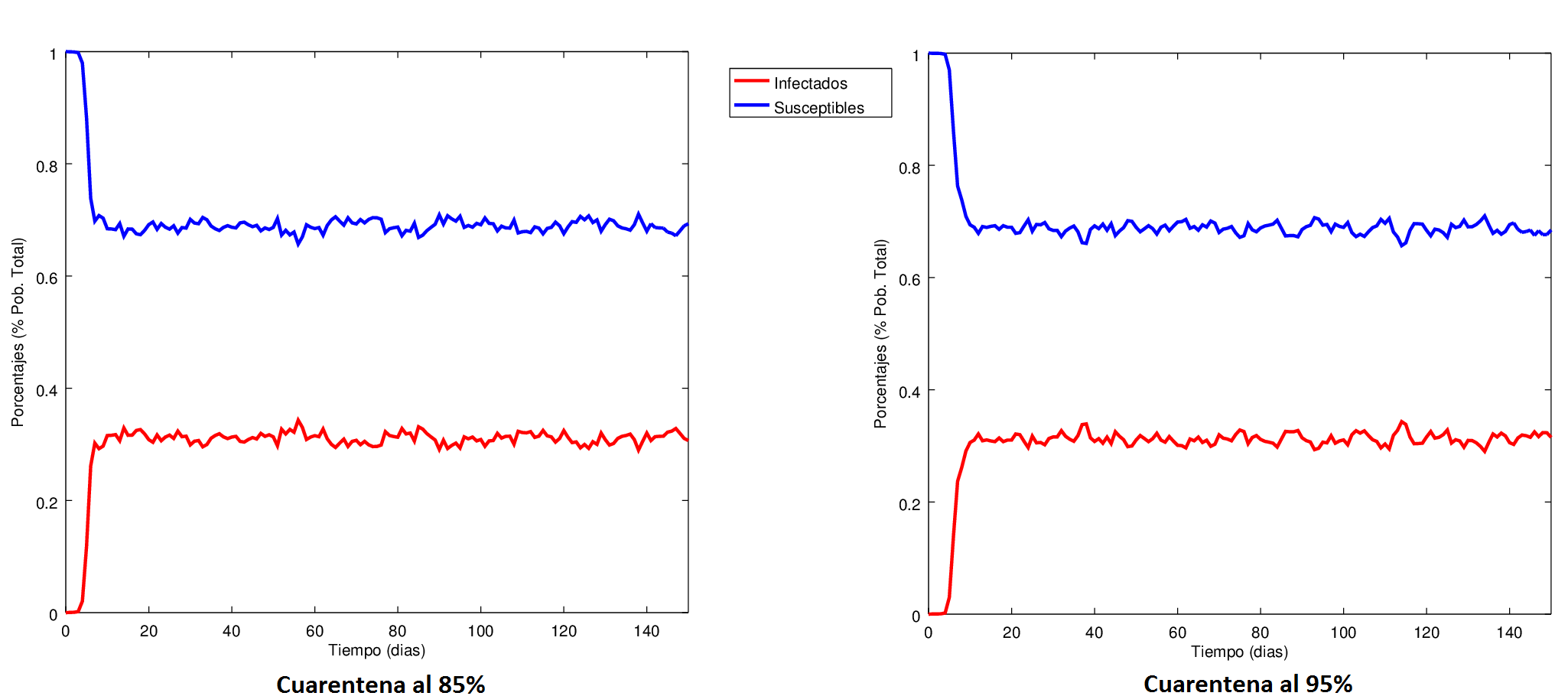
**Modelo SI:** Las simulaciones nos permitieron ver que, en la mayoría de casos, el tiempo que tarda nuestro agente en infectar a toda la población viene determinado por cuan pronto se activa la cuarentena.



En el caso base, nuestro agente tardaba menos de 70 días en infectar toda la red, mientras que nuestra simulación con la cuarentena al 95%, la infección tarda más de 150 días en infectar al 99% de la población. Para confirmar que la activación de la cuarentena era el verdadero factor que provocaba la ralentización del contagio, decidimos efectuar otras 4 simulaciones con los mismos parámetros, de los cuales 3 imitaban el comportamiento observado en la primera simulación.

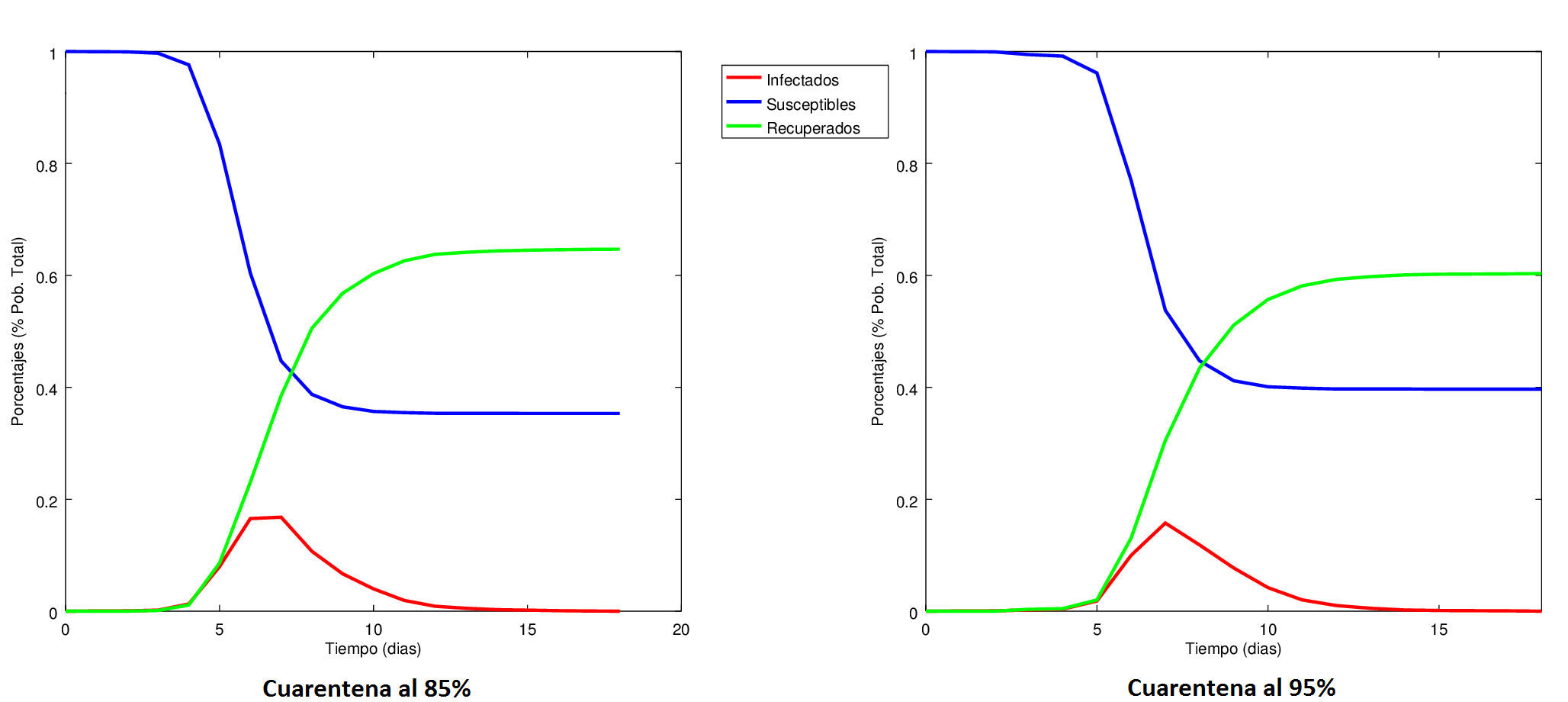
**La conclusión a la que llegamos es que, cuanto más pronto se active la cuarentena tras las primeras infecciones, mayor será el tiempo que el agente tarde en contagiar al resto de nodos de la red. Debido al crecimiento exponencial del desarrollo de un virus SI (en sus primeras fases sobretodo) hay una alta probabilidad de que con una cuarentena como la inicial (85%) consiga infectar alguno sino la mayor parte de los hubs de la red al ser los mayores interconectados de la red, pero con solo aumentar un 10% la tasa de cuarentena no da un margen a la infección como para llegar a infectar estos hubs a tiempo antes de activar la cuarentena. Esto hace que para infectar el resto de nodos, la infección tenga que “rodear” a los hubs y tarda mucho más tiempo (más del doble de días en este caso) en poder infectar la totalidad de la población.**

**Modelo SIS:** Como podemos observar en las gráficas de la evolución de la población, la gráfica correspondiente a la activación de la cuarentena al 95% no presenta ninguna diferencia con la evolución del caso base del modelo SIS que hemos estado viendo hasta ahora.



Los porcentajes de población susceptible y población infectada se mantienen estables, rondando entre el 70% y el 30% respectivamente, en ambas simulaciones. **Por lo tanto, podemos asegurar que una activación temprana de la cuarentena no tiene ningún efecto significativo sobre la transmisión de nuestro agente.La recuperación al modelo susceptible (y viceversa) se da de la misma manera continuamente pueda la infección utilizar o no los hubs debido a que estos solo sirven en cuarentena para la infección de la totalidad poblacional en un largo espacio de tiempo.**

**Modelo SIR**: Tras algunas simulaciones observamos que, al igual que en el modelo SIS, la cuarentena no es un factor decisivo a la hora de reducir el tiempo en el cual la población infectada se reduce a 0.

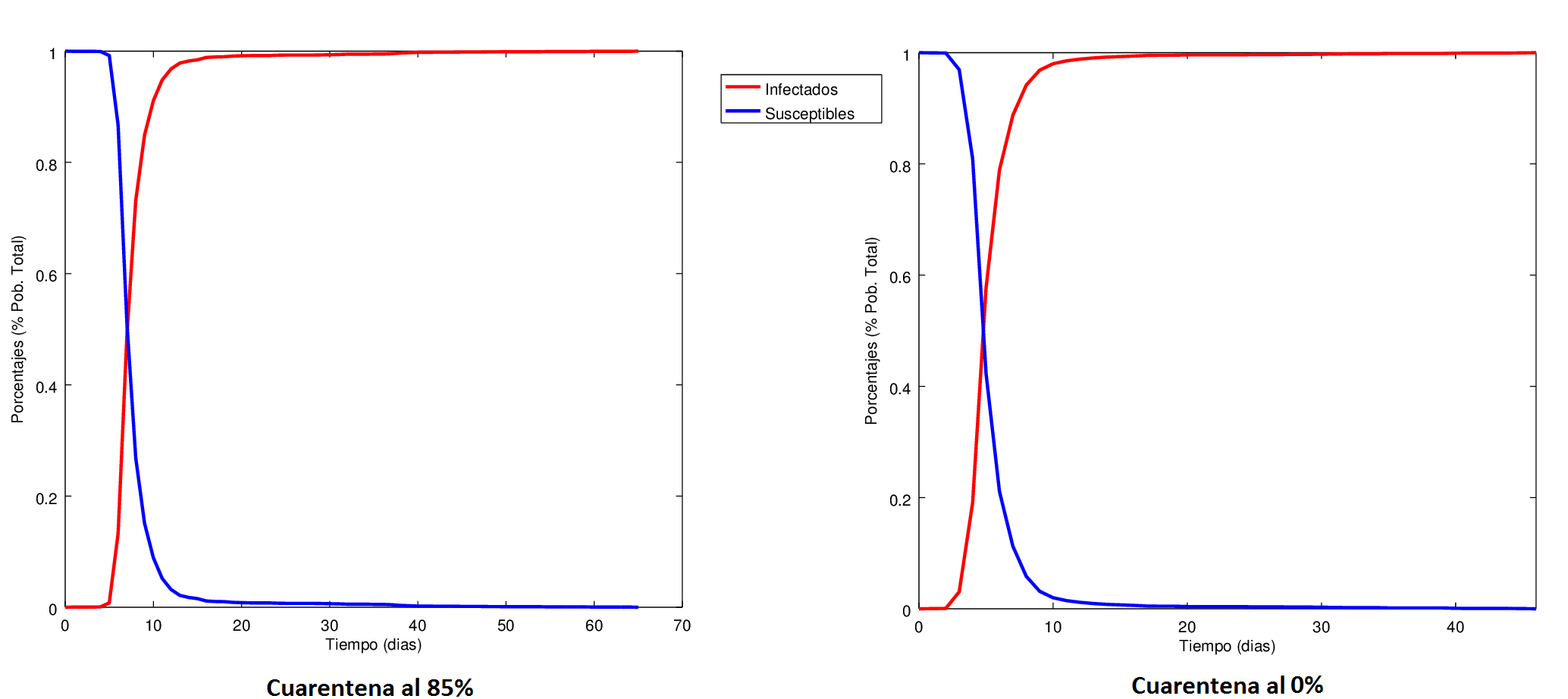


Pero algo interesante que si pudimos observar es que, en más de 4 simulaciones de este modelo con la cuarentena al 95%, la población recuperada y la población susceptible siempre rondaban el 60% y 40% respectivamente. De hecho, el porcentaje más alto de población recuperada al que llegaron estas simulaciones fue de un 61%, mientras que el porcentaje de nuestro caso base del modelo SIR ronda el 64% ~ 65% de la población total.

**Creemos que esta pequeña diferencia en los porcentajes se debe a que la cuarentena se activó de manera prematura con respecto al caso base.**

1. **Cuarentena al 0%**

**Modelo SI:** Para las simulaciones en las que nunca activamos la cuarentena, podemos observar que hay una diferencia en los tiempos que tarda el agente en infectar toda la red.

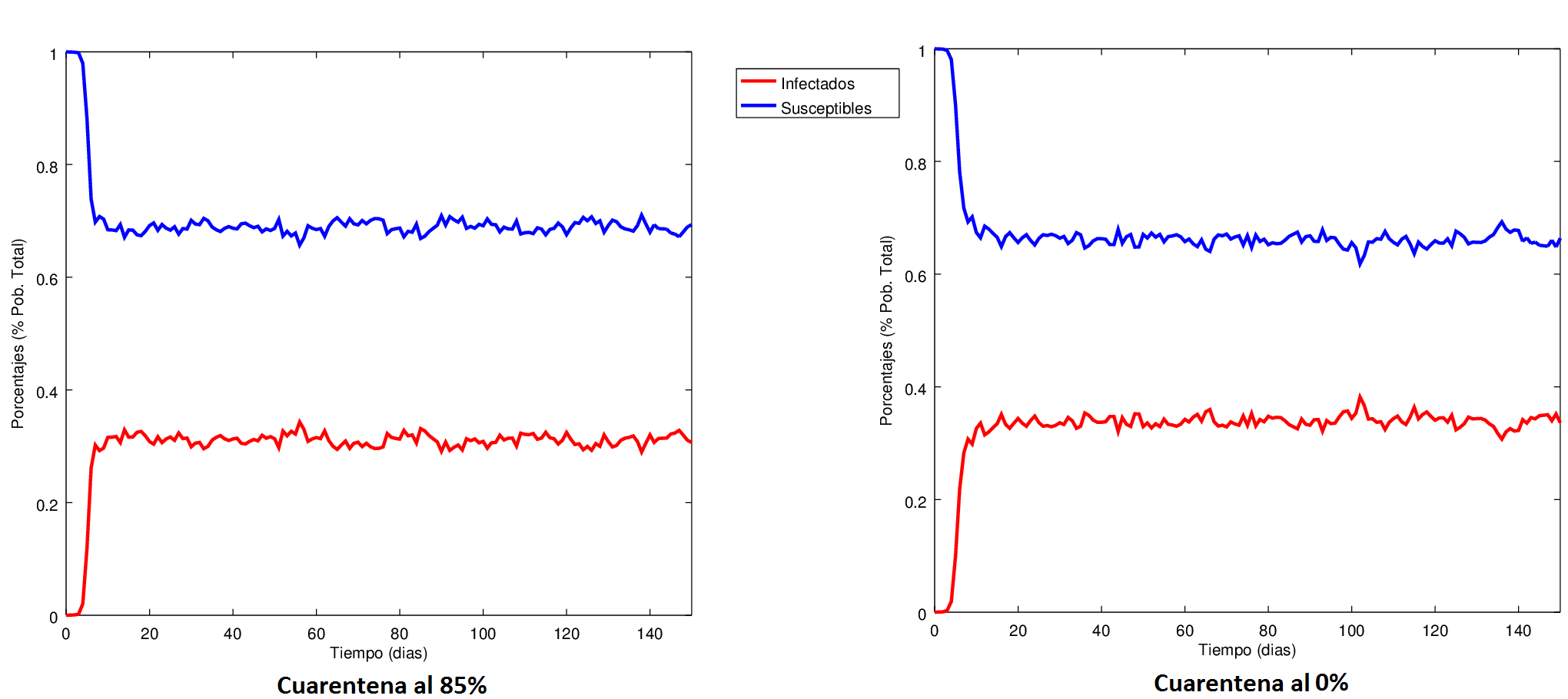


En el caso base, en los que sí existe cuarentena (al 85%), observamos que los días que tarda en extenderse por toda la red es mayor a 60 días (65 días en concreto), mientras que en la simulación en la que no hay cuarentena, este tiempo se reduce a menos de 50 días (46 días en concreto).

Tras repetir la simulación con los mismos parámetros otras 4 veces, comprobamos que, en efecto, el tiempo para lograr la infección total en las redes sin cuarentena no sobrepasaba los 50 días, siendo 51 días el máximo de 5 simulaciones.

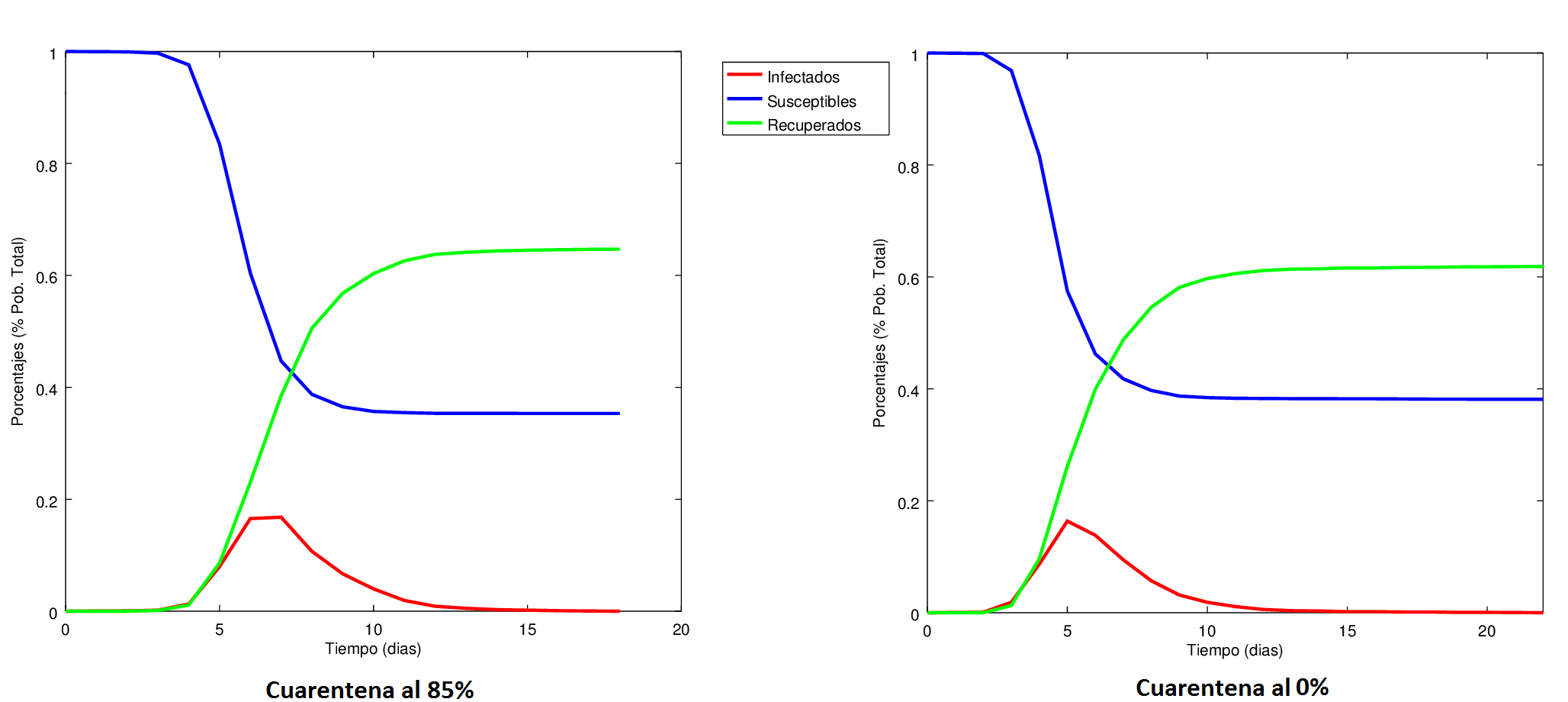
**Por lo tanto, podemos asegurar que la ausencia de cuarentena reduce ligeramente el tiempo total que tarda el agente en infectar todos los nodos debido a que tiene acceso total y sin restricción a todos y cada uno de los hubs disponibles donde extender su influencia.**

**Modelo SIS:** En todas las simulaciones sobre este modelo, no hemos encontrado ninguna diferencia significativa con respecto al caso base.



**La conclusión es que la cuarentena no tendrá ninguna repercusión en este modelo, tanto si se activa con la población susceptible al 85%, como si nunca se activa. La recuperación al modelo susceptible (y viceversa) se da de la misma manera continuamente pueda la infección utilizar o no los hubs debido a que estos solo sirven en cuarentena para la infección de la totalidad poblacional en un largo espacio de tiempo.**

**Modelo SIR:** Para este modelo, gran parte de las simulaciones eran casi idénticas, y todas ellas eran similares al caso base del modelo SIR. La conclusión a la que llegamos es igual a la del modelo SIS antes mencionado: l**a cuarentena no tendrá ninguna repercusión en este modelo, tanto si se activa con la población susceptible al 85%, como si nunca se activa por la misma razón que en el modelo SIS.**



# Simulaciones con Distinto Paciente Cero

Para comprobar si la selección del primer nodo es importante a la hora de disminuir o incrementar la velocidad a la que se transmite nuestro agente infeccioso, decidimos generar distintas simulaciones sobre los 3 modelos, en los que la cuarentena se activaría en distintas fases del contagio. Las simulaciones se generaron utilizando los siguientes parámetros:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Tasa Infec. | Tasa Recup. | Nº Iter. | RW? | Frec. RW | Modo Especial | % Cuarentena | Grado Mínimo para Cuarentena |
| 0.5 | 0 / 0.5 | 150 | Si | 5 | Si | 0.85 | 75 |

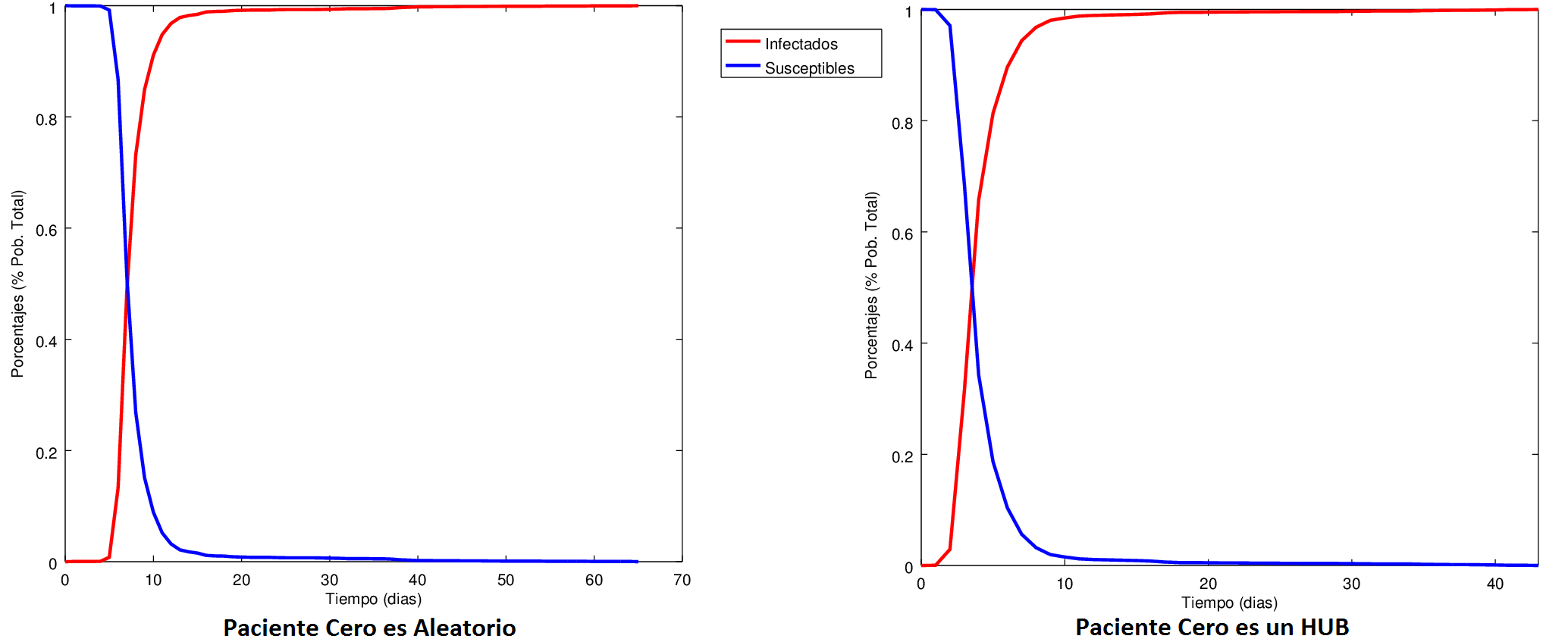
En las simulaciones que usamos como casos base, el primer nodo infectado era un nodo escogido de forma aleatoria de entre todos los nodos de la red. Las siguientes comparaciones se han realizado con simulaciones en los que nuestro primer infectado (paciente cero) era, o un nodo con un grado muy bajo (< 6 vecinos), o un hub con un grado muy alto (> 170 vecinos).

Los resultados que esperábamos obtener, antes de realizar las nuevas simulaciones, eran distintas evoluciones para los modelos SI, SIS y SIR. Para el modelo SI esperábamos que al escoger un hub como primer infectado, la infección tardaría mucho menos que al escoger un nodo de grado bajo, y para los modelos SI y SIR esperábamos que al escoger un nodo de grado bajo como paciente cero, la infección no llegaría a extenderse demasiado, y terminaría siendo erradicada a los pocos días de haber empezado.

Para comprobar si nuestras predicciones son correctas, procederemos a comparar nuestros casos base con las simulaciones que hemos generado.

1. **Cuando el Paciente Cero es un HUB**

**Modelo SI:** En las simulaciones de este modelo, observamos que cuando el primer infectado es un HUB con más de 170 vecinos, el tiempo que nuestro agente tarda en infectar a toda la población se reduce casi 20 días (43 días) con respecto al tiempo conseguido en nuestro caso base de este modelo (65 días).

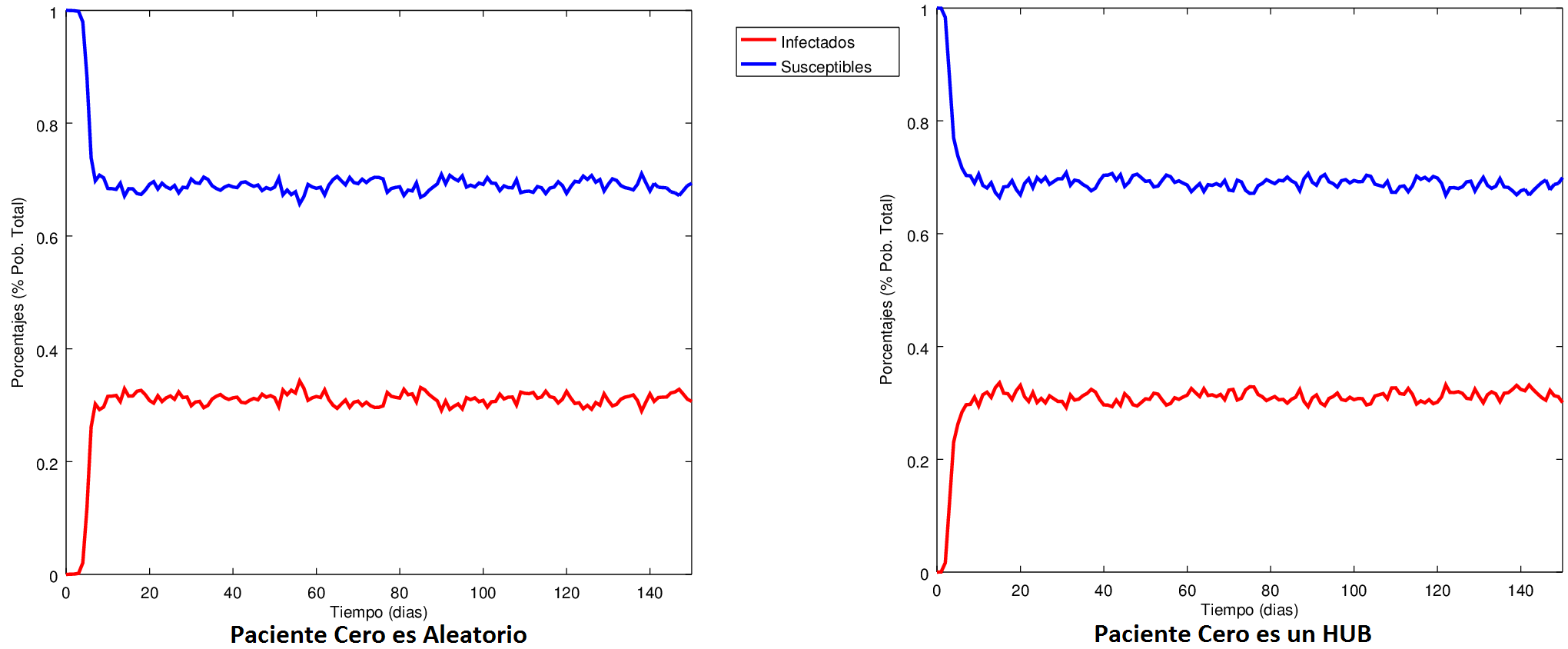


Los tiempos finales de las nuevas 4 simulaciones que generamos con los mismos parámetros para este modelo siempre eran similares, rondando entre los 40 y 50 días. Por lo tanto, podemos asegurar que el grado / n.º de nodos vecinos que tiene el primer nodo infectado tienen un papel importante a la hora de reducir los tiempos que tardaran nuestros agentes en infectar a toda la población.

**Modelo SIS:** Para este modelo, las distintas simulaciones que generamos no presentaron ninguna diferencia significativa con respecto al caso base.

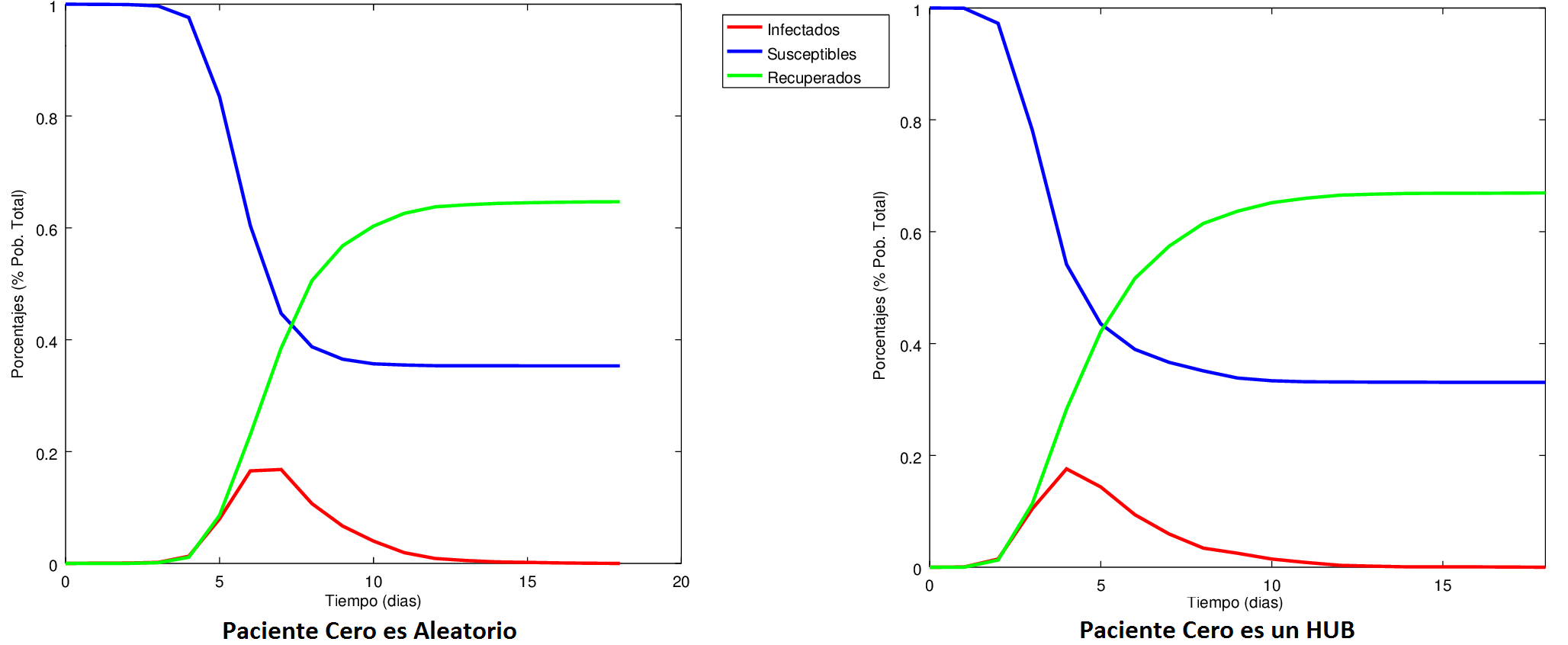
Existe la posibilidad de que la ausencia de diferencias se deba a que el paciente cero de nuestro caso base, que se escogió de forma aleatoria, resulto ser un nodo con un grado muy alto.

Si este es el caso, no podremos conocer la importancia de la selección del primer nodo hasta realizar comparaciones con las simulaciones en las que el primer nodo infectado tenga un grado muy bajo.



**Modelo SIR:** Al igual que con las simulaciones del modelo SIS, no encontramos ninguna diferencia importante con respecto el caso base.

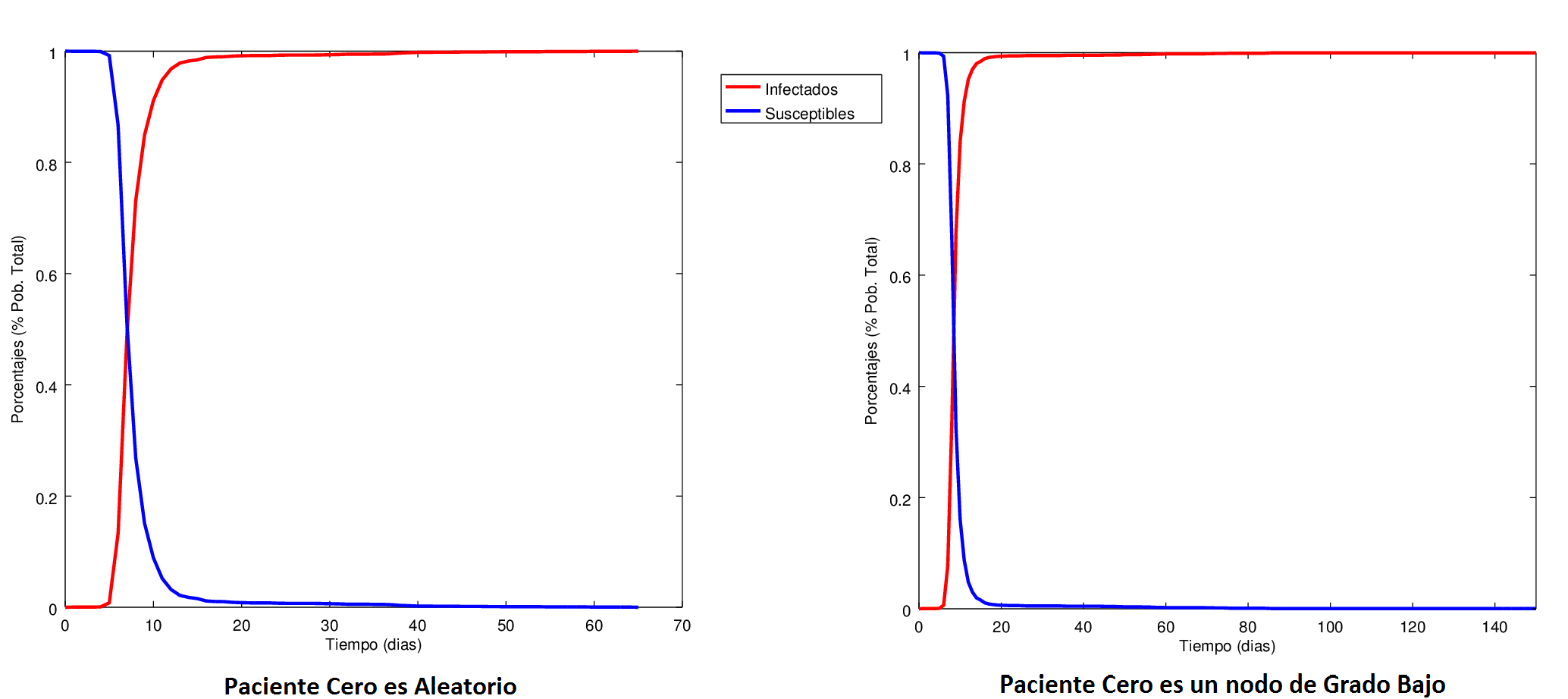
No podremos llegar a una conclusión hasta realizar las comparaciones con las simulaciones en los que el primer nodo infectado es uno de grado muy bajo.



1. **Cuando el Paciente Cero tiene un Grado Bajo**

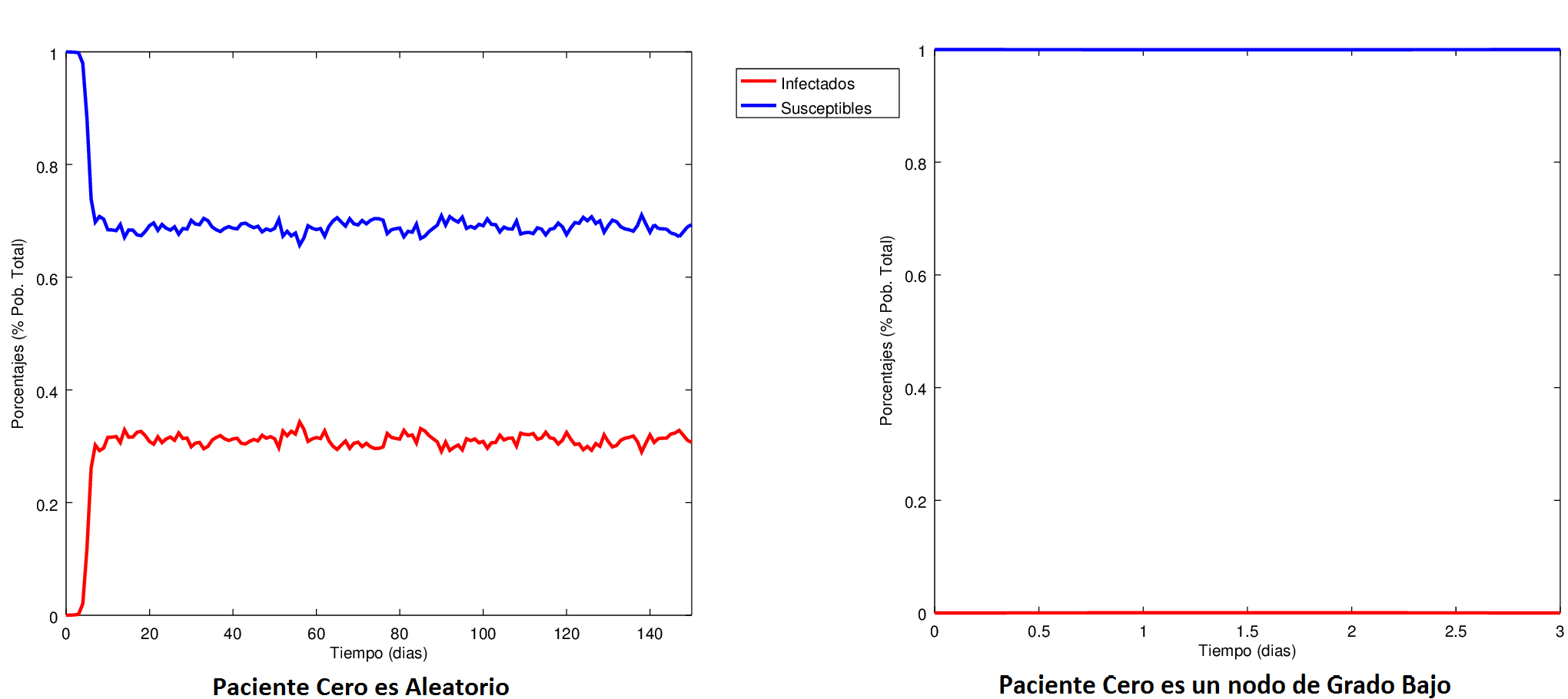
**Modelo SI:** En las nuevas simulaciones observamos que el tiempo que suele tardar la infección en extenderse a lo largo de la red suele superar, casi siempre, la barrera de los 150 días. Esto nos hace suponer dos cosas:

* El paciente cero de nuestro caso base fue un nodo de grado medio-alto, lo que explica su acelerado crecimiento.
* El agente tarda más tiempo en infectar toda la red cuantos menos vecinos tenga el paciente cero.



**Por lo tanto, podemos concluir que el grado de nuestro paciente cero, ya sea alto o bajo, siempre afecta al tiempo total de la infección sobre la red.**

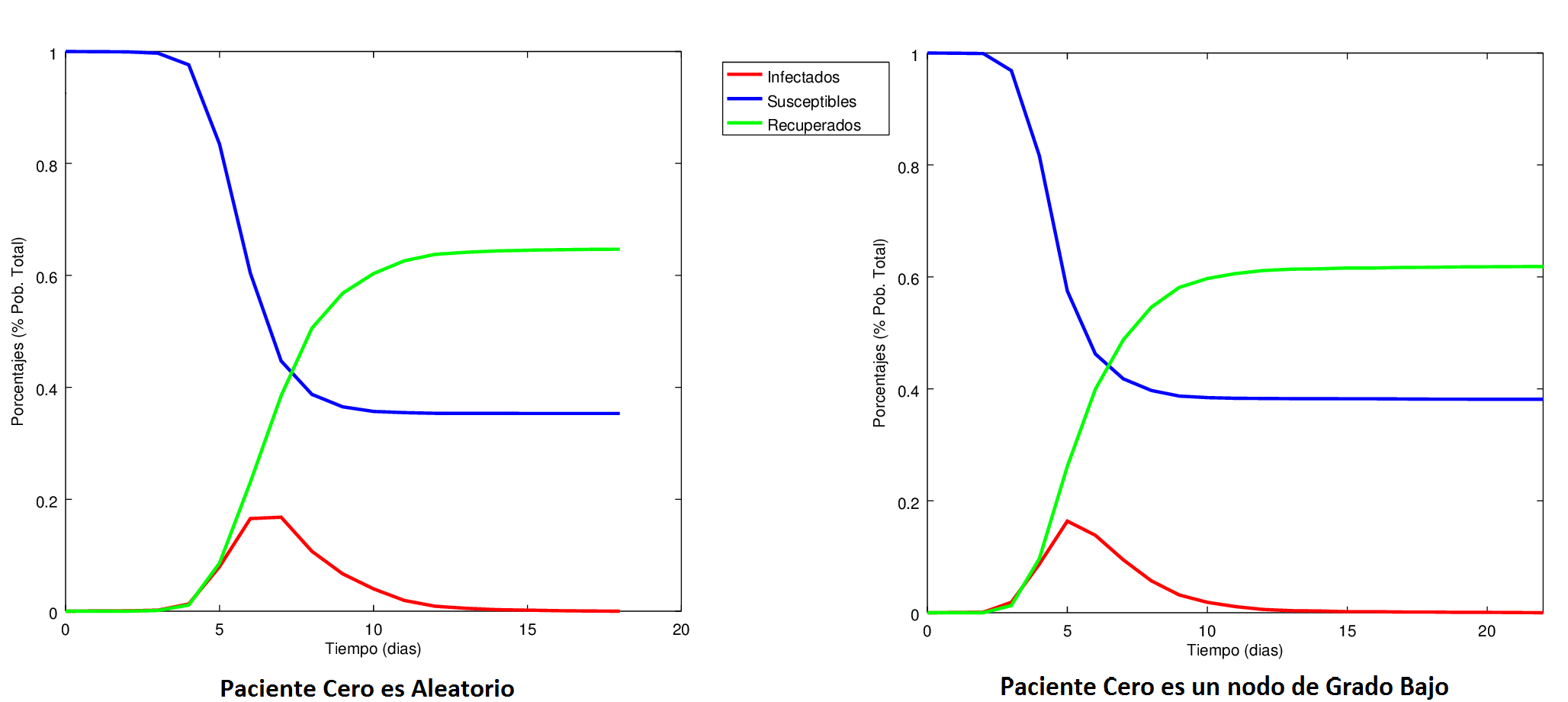
**Modelo SIS:** El cambio de tipo de paciente cero tuvo un gran impacto sobre la evolución del contagio, y es que, como podemos ver en la gráfica, nuestro agente nunca llega a contagiar a un porcentaje significativo de la población, sino que la infección muere pasado unos días del inicio.



Decidimos generar las mismas simulaciones con los mismos parámetros, y un 80% de ellas terminaba de la misma manera: con la infección siendo erradicada a los pocos días de haber iniciado.

**Podemos concluir que el tipo de paciente cero en los modelos SIS afecta de manera directa al desarrollo de la infección, siendo capaz de provocar la erradicación de esta cuando el paciente cero está aislado / tiene pocos vecinos.**

**Modelo SIR:** Las simulaciones de este modelo devolvían 2 tipos de resultados: Un escenario en el cual la infección se erradicaba a los pocos días de empezar, sin afectar ni siquiera a un 1% de la población, al igual que en el modelo SIS, y otro escenario en el cual la infección continuaba durante muchos más días (más de 15), y la población recuperada superaba el 60%.



**Debido a esta diferencia de resultados la única conclusión que podemos sacar es, que en una situación la infección no es capaz de infectar nodos de mayor grado y esta muere por la recuperación antes de poder seguir extendiéndose, o una situación en la que uno de los pocos vecinos del paciente cero es un hub y este logra a tiempo infectarle ya que nos deja una grafica con bastante similitud a la infección de inicial de un hub con alto grado.**

# Resumen de Conclusiones:

# Simulaciones con Distintas Cuarentenas:

La inserción de la cuarentena en nuestra de red de aeropuertos internacional es un completo fracaso. ¿A qué es debido esto? Aunque nuestra densidad de la red es muy baja, rozando el 0,001, lo que hace que usar una cuarentena en los nodos con grados mas alto sea inútil es el coeficiente de agrupamiento o clustering de esta. Este es muy alto, rozando el 0,63, lo que quiere decir que nuestros nodos están muy interconectados entre si y no les hacen falta pasar por los hubs de la red para conectarse en muy pocos pasos con la mayor parte de los nodos. Entonces, es completamente inútil poner una cuarentena a los nodos de grado 75 o mayor cuando existen muchos nodos con mucha interconectividad de grados 70, 64, 56, etc.

Esto cambiaria si nuestra red tuviera un tipo de clustering bajo, entonces las cuarentenas a los hubs más grandes serian de una gran utilidad y puede que en algunos casos pudiéramos incluso llegar a parar de forma permanente la infección aislándola en un conjunto de nodos.

Aunque bien es verdad que en los modelos SIS y SIR la ayuda de la cuarentena es casi inobservable, es cierto que en modelo SI la visibilidad de la ayuda de la cuarentena puede verse a largo plazo. Para ello es vital la rápida actuación (%Cuarentena alto en nuestro programa, <90% es casi inservible) de la utilización de la cuarentena, ya que, si podemos poner en cuarentena a los nodos con mayor grado antes de que se infecten (cosa que no tarda en las primeras fases de la infección debido a la alta conectividad de estos) podremos frenar a largo plazo (a corto plazo es imposible por la velocidad de propagación exponencial y el alto coeficiente de agrupamiento) la extensión total de la infección en nuestra red duplicando y con suerte triplicando el tiempo que esta tarda.

# Simulaciones con Distintos Pacientes Cero:

Definitivamente dependiendo de cuan mayor o menor sea el grado del paciente cero el desarrollo de nuestra infección en la red de aeropuertos tendrá una diferente forma u otra.

En el caso de del modelo SIS