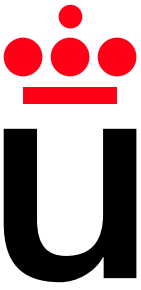

Universidad
Rey Juan Carlos

Escuela Técnica Superior
de Ingeniería Informática

Doble grado en Ingeniería Informática y Matemáticas

Curso 2022-2023

Trabajo Fin de Grado - Grado en Matemáticas

MODELOS MATEMÁTICOS DE PROPAGACIÓN DE MALWARE Y CIBEREPIDEMIAS BASADOS EN ANÁLISIS DE REDES COMPLEJAS Y ECUACIONES DIFERENCIALES

Autor: Sergio Salazar Cárdenas
Tutor: Miguel Romance del Río y David González de la Aleja Gallego

Agradecimientos

Una vez leí que lo bueno de hacer dos TFGs es que puedes agradecer dos veces a la gente que te ha acompañado en el camino hasta aquí.

A las más importantes que les agradezco todo son a mi madre y a mi hermana, son indispensables para mí, sin ellas no habría trabajo, no habría carrera y no habría nada. No habría yo.

Tengo la suerte de tener mucha gente cerca, mis amigos de Pinto, de Móstoles, mi club de Balonmano ... Gracias a ellos también estoy aquí.

Para terminar agradecer a mis profesores y compañeros de despacho del último año, ha sido fantástico trabajar cerca de ellos.

Resumen

Se presenta un análisis comparativo de la simulación de la propagación de malware y ciberepidemias en sistemas de información basados en modelos compartimentales, empleando tanto técnicas de agentes en redes complejas como modelos tipo SEIR.

Inicialmente se realiza una introducción histórica y contextual de los conceptos de modelización matemática y modelización compartimental, relacionando estos conceptos con trabajos y modelos reales.

Posteriormente se realiza un análisis de los coneptos de Redes Complejas y Teoría de Grafos como preámbulo a la implementación realizada en todo el trabajo.

Como introducción al modelo principal del trabajo se realiza un análisis previo en sendas implementaciones del conocido modelo SIR, viendo algunas de sus características similares y sacando conclusiones cómo previo a un modelo más complejo.

Finalmente se realiza un análisis profundo de uno de los modelos más modernos de previsión de la propagación de malware en ambas implementaciones, así como una comparación de las mismas. Esta parte compone el grueso del trabajo teórico y computacional. Se añade además un capítulo de investigación original en el que se usan los modelos para prever el impacto de los virus siguiendo diversas estrategias de ataque.

Palabras clave:

- Modelización Matemática
- Modelos Compartimentales
- Redes Complejas
- Ecuaciones Diferenciales
- Malware
- Python

Índice de contenidos

Índice de tablas

Índice de figuras

1 Introducción

La globalización, el auge de las nuevas tecnologías y la democratización masiva de internet han convertido al malware en una de las principales amenazas a nivel mundial. Según el *World Economic Forum* esta amenaza se ha visto muy acentuada en los últimos años, despues de la pandemia del Covid-19 [9]. Conceptos como las ciberpandemias o la ciberguerra se ponen cada vez más al día y se acrecenta un problema que la comunidad científica deberá paliar en los años venideros: la creación de malware es más rápida que la creación de software que pueda combatirlo.

La lucha contra el malware es enfrentada a través de distintas perspectivas: creación de metodologías de desarrollo seguro, desarrollo de antivirus, concienciación social y empresarial... En este trabajo se va a intentar dar perspectiva de una nueva vía que ha sido desarrollada en los últimos años y que dejaba un vacío considerable ante esta tesitura: la modelización de la propagación de malware y ciberepidemias.

Para la modelización de estos sistemas se van a utilizar los denominados modelos compartimentales. Estos modelos se basan en la idea de dividir la población de estudio (en este caso los dispositivos de una red) en diferentes compartimentos o categorías, según sus características relevantes respecto al concepto a tratar (existencia de un virus informático) y describir la dinámica de transición entre estos compartimentos.

Estos modelos nacen de la modelización de virus biológicos y han sido utilizados en situaciones reales como el COVID-19. Dadas algunas similitudes entre la propagación de virus biológicos y digitales, algunos de los primeros modelos para ciberpandemias surgen de manera casi inmediata realizando una analogía

respeto a los modelos sobre pandemias biológicas.

El desarrollo de estos modelos se ha basado históricamente en sistemas de ecuaciones diferenciales debido principalmente a su capacidad para modelar la dinámica y de esta forma identificar la evolución de las distintas poblaciones o compartimentos a lo largo del tiempo o el espacio.

Gracias a los avances informáticos surge otro sistema para interpretar los modelos compartimentales, los modelos de redes complejas, que buscan modelar a cada individuo de manera separada y no a una población entera. Cada individuo pertenecerá a uno de estos compartimentos según su estado y pudiendo evolucionar a otro según el estado general del sistema o según el conjunto de individuos con los que interactúa.

Este trabajo de fin de grado tiene dos objetivos fundamentales. Por un lado presentar y revisar el concepto de modelo compartimental y realizar una comparación entres sus variantes más identificativas: los modelos de ecuaciones diferenciales y los modelos de redes complejas.

Por otro lado desarrollar y analizar el modelo de redes complejas mencionado en [1], implementarlo en ambos esquemas y realizar un análisis detallado de los mismos, con la posibilidad de identificar características y estrategias de los actores en la red.

2

Objetivos

En este capítulo se describen los objetivos que han impulsado la realización de este trabajo. En el objetivo principal se indica la intención fundamental del trabajo, mientras que en los secundarios se establecen otros propósitos relacionados con el contexto del TFG.

2.1. Objetivo Principal

El objetivo principal del trabajo es la revisión e implementación del modelo de predicción de epidemias digitales definido en [1]. Tras el estudio inicial se busca la comparación de las distintas implementaciones, así como aprovechar las ventajas de las mismas para la realización de experimentos que permitan obtener conclusiones aplicables a escenarios reales.

2.2. Objetivos Secundarios

Los objetivos secundarios en los que se sustenta la metodología del trabajo son los siguientes:

- Revisión y contextualización de los concepto de modelización matemática y modelos compartimentales usados para la simulación de epidemias biológicas y digitales.

- Revisión y estudio de los conceptos de grafos y redes complejas, junto los conceptos derivados que surgen de estas áreas de las matemáticas.
- Revisión y estudio de los conceptos de ecuaciones diferenciales y de su resolución a través de métodos numéricos.
- Estudio y utilización del lenguaje de programación Python desde una perspectiva matemática, para la simulación de situaciones reales y resolución de modelos matemáticos.

3

Modelos matemáticos

Este capítulo está orientado a la definición y clasificación de los modelos matemáticos, así como una contextualización histórica y formal de los modelos que se van a utilizar para la descripción de la propagación de malware.

3.1. Definición y taxonomía de la modelización matemática.

La modelización matemática es una disciplina que busca trasladar problemas que surgen en un determinado campo científico al lenguaje matemático, de forma que se puedan tratar con las herramientas de esta disciplina. El propósito principal es el de entender mejor algunos de los fenómenos que se definen y poder realizar predicciones sobre los mismos.

La metodología de trabajo en modelización matemática se puede dividir en 5 etapas [4]:

1. **Identificación del fenómeno**: Se comienza con el estudio de un fenómeno o problema que se quiere definir a través de un modelo. Del mismo se deben identificar cuales son los aspectos y características que permiten su descripción: las leyes físicas o biológicas que están involucradas, los parámetros que lo influyen, sus relaciones, etc.

2. **Modelo de trabajo**: Una vez identificados estos parámetros, se debe es-

coger cuales deben considerarse para definir el modelo, tras esta fase de simplificación se obtiene un **Modelo de Trabajo**.

3. **Modelo Matemático**: Posteriormente, el modelo de trabajo se traducirá al lenguaje matemático regido por las relaciones de las distintas magnitudes y parámetros que se hayan identificado, obteniéndose el **Modelo Matemático**.

4. **Modelo Computacional**: Una vez desarrollado el modelo matemático, este debe implementarse. Esto conlleva la resolución de las ecuaciones que los describe, o de manera más reciente, una implementación computacional, donde las herramientas informáticas aportan más potencia que el desarrollo individual que se pueda realizar, definiremos este modelo como **Modelo Computacional**.

5. **Experimentación y Conclusiones**: Para finalizar se realizarán los experimentos necesarios sobre el modelo implementado y obtendremos resultados de los que obtener conclusiones. Un análisis de los mismos puede sumarse al análisis de los datos reales del fenómeno de estudio, de forma que se puede comenzar el ciclo de modelización con esta nueva información incorporada, por tanto se puede decir que el proceso de modelización no es lineal, sino que sigue una estructura cíclica.

Existen gran variedad de modelos matemáticos y existe una taxonomía que los divide según estas características:

- **Modelos deterministas contra modelos estocásticos:**
  En un modelo determinista ninguna de sus variables esta influenciada por un proceso aleatorio, mientras que en un modelo estocástico existen parámetros definidos por el azar a través de distribuciones de probabilidad.

- **Modelos continuos contra modelos discretos:**
  En los modelos continuos las variables se definen en un dominio continuo, es decir, la cantidad de valores que pueden tomar es infinita y no numerable, mientras que en un modelo discreto las variables toman valores dentro de un conjunto numerable. Ambas metodologías pueden usarse en distintas variables de un mismo sistema creando modelos mixtos.

- **Modelos globales contra modelos individuales:** En los modelos globales se trata a la población de estudio como una única masa identificando su comportamiento general y describiéndolo, en los modelos individuales se propone identificar cada individuo como único, esto es factible en grandes poblaciones gracias a la capacidad de computación moderna.

La mayoría de modelos están basados en ecuaciones diferenciales ordinarias y parciales, según la taxonomía anterior estos se identifican como modelos deterministas, continuos y globales. En cambio, gracias a la potencia de la computación,

en los últimos años crece una modelización basada en redes complejas, maquinas de estados y autómatas celulares. Su principal ventaja es poder tratar con modelos individuales, aunque suelen complicar la modelización de magnitudes continuas que normalmente requieren una simplificación en un dominio discreto. Esta comparación será troncal en el grueso del trabajo ya que los modelos a tratar serán cotejados respecto ambas implementaciones.

3.2. Modelos matemáticos para la simulación de la propagación de Malware.

El estudio de la modelización en biología asienta las bases para la simulación de la propagación de Malware. Uno de los trabajos fundamentales en este área es la modelización matemática y existen innumerable trabajos sobre la misma [14].

Los primeros modelos utilizados para la simulación de ciberepidemias nacen de la modelización de la propagación de enfermedades biológicas, debido a la similitud estructural entre las redes sociales en una comunidad y las redes de computadores. Estos modelos se denominan Modelos Compartimentales y surgen a comienzos del siglo XX con el trabajo de Kermack y McKendrick (1927) [8].

Estos modelos se basan en la separación de los elementos de estudio en compartimentos o estados según las características del individuo y se establecen las relaciones entre ellos para explicar como evoluciona la enfermedad, o en el caso de este trabajo, el estado del virus informático.

Gracias a estas características es fácil identificar estos modelos a través de diagramas de estados en forma de grafo, donde los nodos son los distintos compartimentos y las relaciones entre los mismos vienen descritas por sus enlaces, cuyos pesos hablan de los parámetros que intervienen en su evolución.

La gran mayoría de estos modelos comparten varios de los compartimentos en los que se dividen la población, el bloque más común es aquel que compone el modelo SIR [17] denominado así por sus 3 estados:

- **S: Susceptible (*susceptible*)** Aquellos individuos no infectados que no poseen inmunidad y pueden llegar a infectarse en caso de contacto.
- **I: Infectado (*infectious*)** Aquellos individuos infectados y que por tanto pueden transmitir el virus.
- **R: Recuperados y Fallecidos (*recovered*)** Individuos inmunes a la infección (o fallecidos) y que no favorecen la transmisión.

Obsérvese que las definiciones de los 3 compartimentos podría utilizarse para un modelo de pandemia biológica o computacional.

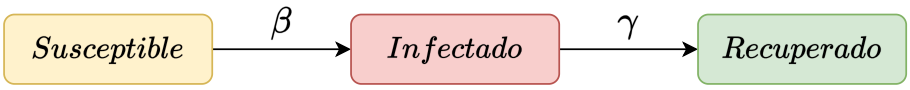

Figura 3.1: Esquema del modelo SIR.

A partir de este aparecen modelos más complejos: SEIR [11], SEIRS [2], SEIQR [7]...

Rápidamente se crea una analogía entre los compartimentos para modelos biológicos y de propagación de Malware, con ciertas variaciones en conceptos como "cuarentena" o "vacunados", pero con una gran semejanza.

La implementación de los modelos compartimentales es variada e históricamente se distinguen dos tendencias claras: modelos deterministas (*ecuaciones diferenciales*) o modelos estocásticos (*redes complejas, cadenas de Markov y ecuaciones diferenciales estocásticas*).

La mayoría de los trabajos en este ámbito modelizan la dinámica de la pandemia a través de ecuaciones diferenciales. Con estas se busca encontrar las características propias de la evolución de la enfermedad así como la comprensión de situaciones como estrategias de vacunación, prevalencia de la enfermedad, etc. Estos modelos funcionan bien cuando la población a tratar es suficientemente grande de forma que la aparición de sucesos a nivel local no afecte demasiado al estado global.

En las últimas décadas surgen nuevas maneras de simular la dinámica de estos modelos a través del uso de redes complejas. Esta perspectiva tiene como objetivo la simulación de cada individuo de la población, lo cuál demandaba un alto esfuerzo a nivel computacional. Esta es la principal razón de que estos modelos sean relativamente modernos, cuando la computación de altas prestaciones ha dejado de ser un problema.

En este caso no solo se hace un estudio a nivel macroscópico del sistema sino que se pueden analizar distintas situaciones a nivel micro, obteniendo buenos resultados para poblaciones más reducidas, por ejemplo: situaciones a nivel de comunidades, aislamiento o inserción de un conjunto de individuos del sistema, identificación de nodos críticos, ... Además permiten realizar estudios a nivel estructural, donde ya no se ve al sistema como una masa, sino que se entiende su forma.

Esta nueva perspectiva aporta grandes posibilidades a la modelización de propagación de malware, pues permite modelar algunas de las principales diferencias que existen con la propagación de virus biológicos.

A diferencia de lo que ocurre con epidemias biológicas, se puede extrapolar el concepto de ciberepidemia a un hecho cercano a la teoría de juegos, donde existe una lucha entre el ciberdelincuente y los miembros de la red. El primero buscará buenas estrategias para su ataque, maximizando el daño y su propagación, sin embargo, el resto de individuos de la red intentarán defenderse. Este hecho es diferencial para los modelos de transmisión de malware modernos.

Además, la visión de redes complejas permite modelizar la propagación atendiendo al tipo de escenario que estamos estudiando, no es lo mismo modelar el impacto de un phising en una red de correo electrónico, que el de un troyano alojado en un programa informático descargable y reproducible en internet. La estructura de la red cambia y también los medios de propagación e interactuación.

La modelización de la propagación de malware es un área relativamente moderna, su estudio de manera exhaustiva se produce en los últimos años debido a la creciente amenaza que supone [4]. En este momento comienzan a surgir modelos especializados que no suponen una simple analogía con los modelos biológicos, pueden observarse algunos de estos en [5], [12] y [20].

A parte del análisis de modelos conocidos, la motivación de este trabajo es el análisis del modelo definido en [1], bautizado en este trabajo como SIH-UA, un modelo creado específicamente para ciberepidemias que tiene en cuenta que el malware esta específicamente diseñado para atacar una red, es decir, que la evolución de la misma se ve condicionada por la intención del ciberdelincuente y de la defensa de los usuarios.

4

Redes Complejas

La aproximación esencial de este trabajo al estudio de la evolución de virus en redes informáticas es a través de la Redes Complejas, lo que nos lleva a realizar un primer estudio del concepto de Grafo y sus características. Muchas de estas definiciones y proposiciones que se van a incluir pueden encontrarse en la gran mayoría de la bibliografías sobre Redes Complejas y Teoría de Grafos, como por ejemplo [18] o [10].

4.1. Definiciones y conceptos básicos.

La definición de Red Compleja combina dos fundamentos esenciales, uno estructural y otro dinámico. De forma estructural una red compleja no es más que un grafo de cualquier tipo, en el caso de este trabajo se utilizarán fundamentalmente grafos no dirigidos y no ponderados.

Por tanto, la diferencia entre la Teoría de Redes Complejas y la Teoría de Grafos reside en el concepto de dinámica. En las Redes Complejas se entiende que existen unos sucesos que modelan los procesos deseados, se entiende que el grafo es un "terreno de juego". En cambio, históricamente, en la Teoría de Grafos, solo se han estudiado estos objetos de forma estática, atendiendo a sus propiedades combinatorias y estructurales.

Como se ha mencionado, el concepto de grafo es esencial en esta disciplina, por lo que en esta sección se van a mostrar todas las definiciones y proposiciones necesarias para la completitud del trabajo. Véase una primera definición.

Definición 1.
Un grafo no dirigido es un par $G = (\mathcal{V}, \mathcal{E})$*, formado por un conjunto finito* $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ *denominados vértices o nodos y un conjunto* $\mathcal{E} = \{e_1, e_2, ..., e_m\}$ *denominados aristas o enlaces. Los elementos de este último conjunto son pares no ordenados de elementos de* $\mathcal{V}$*, es decir,* $e_i = \{v_j, v_k\}$ *con* $i \in \{1, ..., m\}$ *y* $j, k \in \{1, ..., n\}$*.*

Notación 1.
Dado un grafo $G = (\mathcal{V}, \mathcal{E})$*, se denotará:* $N = |\mathcal{V}|$ *y* $L = |\mathcal{E}|$*, donde* $|\cdot|$ *representa el número de elementos del conjunto.*

También existen los conceptos de grafos ponderados y dirigidos, que añaden la capacidad de introducir dirección y pesos a las aristas. Sin embargo los omitimos ya que estos conceptos no van a ser utilizados en este trabajo.

Aunque son útiles, los conjuntos de la definición no suelen ser la forma usual de mostrar un grafo. Esto se puede realizar de manera gráfica o a través de su matriz de adyacencia.

De manera gráfica los nodos consistirán en puntos del plano que se verán unidos en caso de existir una relación entre los mismos. Si definimos los conjuntos $\bar{\mathcal{V}} = \{1, .., 10\}$ y $\bar{\mathcal{E}} = \{\{1,3\}, \{3,5\}, \{3,9\}, \{3,10\}, \{3,4\}, \{2,4\}, \{4,7\}, \{6,4\}, \{4,8\}\}$ podemos encontrar un ejemplo de una representación gráfica del grafo $\bar{G} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$ en la figura 4.1.

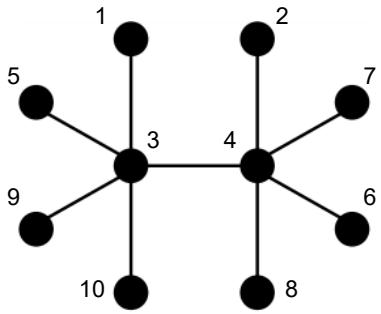

Figura 4.1: Representación gráfica del grafo $\bar{G}$.

Definición 2.
Dado un grafo $G = (\mathcal{V}, \mathcal{E})$*, se define la matriz de adyacencia de* G *como una matriz cuadrada* $A = (a_{ij}) \in M_N$ *que cumple*

$$a_{i,j} = \begin{cases} 1 & si \quad \{i,j\} \in \mathcal{E}, \\ 0 & si \quad \{i,j\} \notin \mathcal{E}. \end{cases}$$

Ahora, se introducen diversas definiciones que van a ser troncales durante el transcurso del trabajo, el primero de ellos es el de grado, que mide la cantidad de conexiones de los nodos del grafo.

Definición 3.
Se denomina grado del nodo v al número de aristas que inciden sobre v, es decir:

$$gr(v) = k_v = |\{\{v, w\} \in \mathcal{E} : w \in \mathcal{V}\}|.$$

Se define grado medio del grafo G a la media de los grados de los nodos del grafo:

$$\langle k \rangle = \frac{\sum_{v \in \mathcal{V}} gr(v)}{N} = \frac{2L}{N}.$$

Otro conceptos importantes son los de camino y ciclo, estos conceptos nos permiten identificar rutas entre los nodos de los grafos. En el caso de la trasmisión de malware, los caminos son la vía por la que el virus podrá transmitirse.

Definición 4.
Se denomina camino entre $v_1, v_{n+1} \in \mathcal{V}$ a una sucesión de vértices $v_1, v_2, ...v_{n+1} \in \mathcal{V}$ y aristas $e_1, e_2, ..., e_n \in \mathcal{E}$, tales que $e_i = \{v_i, v_{i+1}\}$

En el caso que $v_1 = v_{n+1}$ el camino se denomina ciclo.

A la cantidad de aristas que componen el camino se le denomina longitud o distancia del camino.

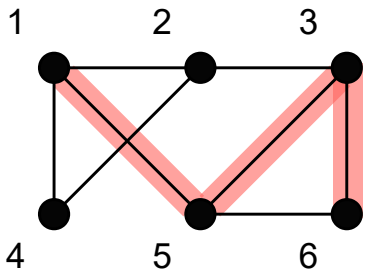

Figura 4.2: Ejemplo de un camino $C = (1, 5, 3, 6)$ entre los nodos 1 y 6.

Además, en este trabajo se trabajará con redes denominadas conexas, es decir, que todo nodo podrá conectarse con otro del grafo, véase su definición.

Definición 5.
Un grafo se denomina conexo, si para cualquier par de vértices $v, w \in \mathcal{V}$, existe al menos un camino que comience en v y termine en w.

Un tipo específico de grafo conexo es aquel cuyos nodos están conectados no solo por caminos, sino por aristas, creándose así todas las posibles, véase la definición:

Definición 6.
Un grafo se denomina completo, si para cualquier par de vértices $v, w \in \mathcal{V}$, existe la arista $\{v, w\} \in \mathcal{E}$.

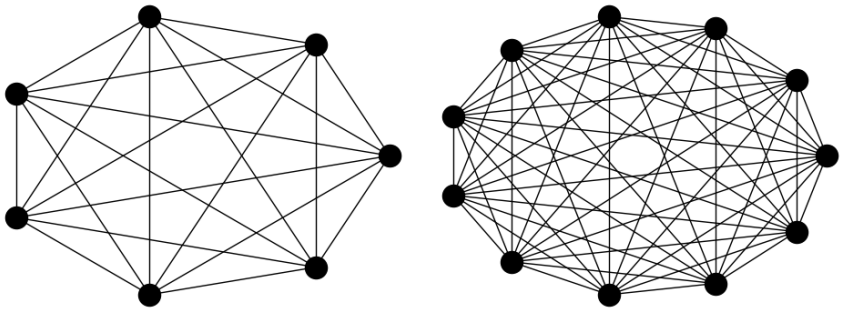

Figura 4.3: Ejemplo de grafos completos con N=7 y N=11 .

Es remarcable que no todos los caminos por los que un malware podrá propagarse son igual de eficientes, aquellos que requieran pasar por más nodos harán que el virus tarde más en propagarse, de esta idea subyace el concepto de distancia, véase su definición.

Definición 7.
Se define la distancia entre dos nodos $i, j \in \mathcal{V}$ como el mínimo entre las longitudes de todos los caminos entre i y j, es decir a la longitud del camino más corto. Se denota como d_{ij}.

Definición 8.
Se define como longitud media de un grafo conexo, como la media de las distancias entre todos los pares de nodos:

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{ij}.$$

Uno de los conceptos que se estudiará más adelante es el de redes aleatorias, debido a que se van a estudiar simulaciones de propagación de malware en las mismas, se introducen los conceptos de grafo enrejado y grafo circular para su comprensión.

Definición 9.
Se define un grafo enrejado circular como el grafo de N nodos etiquetados de $0, ..., N-1$ de tal forma que existe una conexión entre el nodo i y los nodos $i+1$ e $i-1$. El caso específico del nodo 0 se unirá con el último nodo etiquetado $N-1$ Este grafo también se puede definir como un ciclo de grado N.

Definición 10.
Se define grafo circular de N nodos y orden m par, como el grafo de N nodos etiquetados de $0, ..., N-1$ de tal forma que existe una conexión entre el nodo i y los nodos $i+1, i+2, ..., i+m/2$ e $i-1, i-2, ..., i-m/2$. En los casos específicos que se realice una conexión a un nodo con una etiqueta negativa $i < 0$, se realizará la conexión con el nodo $N+i$, es decir, se sigue el orden inverso comenzando por el nodo $N-1$. El grafo circular de grado 2 es idéntico al grafo enrejado circular.

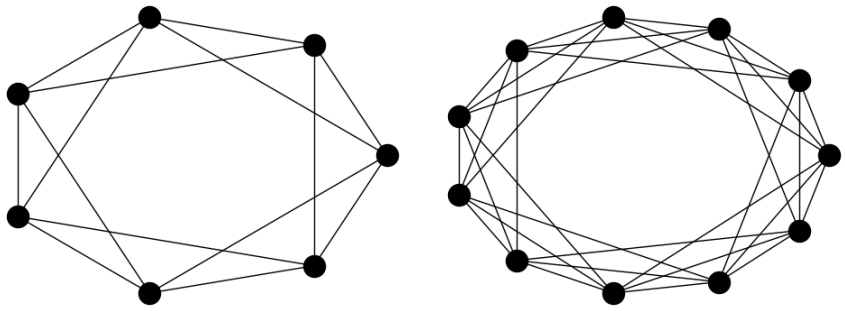

Figura 4.4: Ejemplo de grafos circulares con N=7, m=4 y N=11, m=6.

El conjunto de definiciones presentadas permitirá la exposición del trabajo de manera formal, permitiendo volver a esta sección en caso de que algún concepto no esté completamente claro cuando sea utilizado.

4.2. Redes Aleatorias

En la simulación de modelos matemáticos basados en redes la estructura del grafo que lo subyace es de vital importancia debido a que condiciona cómo se define la dinámica entre los nodos que lo conforman.

Una de los principales retos al usar este tipo de modelos es identificar los grafos que simulen de manera correcta el funcionamiento de un fenómeno real.

Debido a esto aparecen un conjunto de procesos para generar redes de manera aleatoria siguiendo un conjunto de parámetros.

4.2.1. Modelo de Erdös-Rényi

El modelo de Erdös-Rényi [16] supone el primer esfuerzo fructífero de crear grafos de manera aleatoria con un sentido claro. Este modelo crea la red de manera imparcial, es decir, no da prioridad a ningún nodo o arista, creando así redes genéricas para los parámetros buscados, estableciendo el número de nodos y la densidad de aristas en la red.

Definición 11. *El modelo de Erdös-Rényi (ER) permite crear un grafo aleatorio a través de dos parámetros:*

- *N: El número de nodos del grafo*
- *p: La probabilidad de que exista una arista entre dos nodos.*

El proceso de creación transcurre añadiendo los nodos uno a uno hasta N, por cada nodo k creado se establecerá una conexión con los $k-1$ anteriores dependiendo de la probabilidad p definida.

Es interesante mencionar entonces los casos $p = 0$ donde todos los nodos de la red quedarán aislados y $p = 1$ donde obtendremos un grafo completo.

Las redes obtenidas cada vez que se utiliza este proceso son aleatorias, por lo que en cada simulación serán distintas, podemos observar tres ejemplos de redes $ER(7, 0.3)$ en la Figura 4.5.

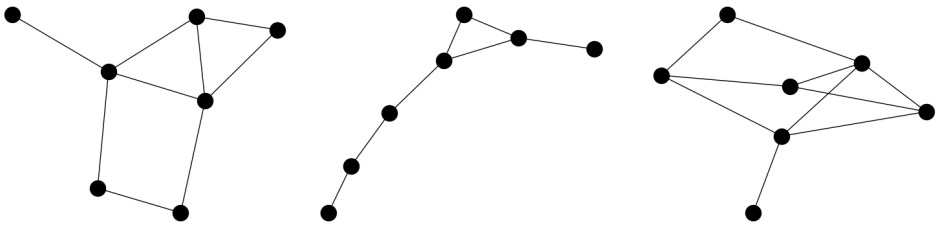

Figura 4.5: Ejemplos de redes $ER(7, 0.3)$.

Existen algunas propiedades interesantes sobre las redes de Erdös-Rényi que caben destacar:

Proposición 1. *La probabilidad de que una red de Erdös-Rényi contenga H aristas sigue una distribución binomial, con parámetro p:*

$$P(|\mathcal{E}| = L) = P(L) = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}.$$

Demostración:
El hecho de que pueda modelarse a través de una distribución binomial es dada por la definición, la existencia de cada una de las aristas es un experimento de Bernoulli independiente con una probabilidad fija p de que suceda.

Por tanto en una distribución binomial la probabilidad de obtener L éxitos se sabe que es:

$$P(X = L) = \binom{n}{L} p^L (1-p)^{n-L}$$

Siguiendo la definición del modelo, habrá tantos experimentos como aristas posibles, es decir, como aristas en un grafo conexo, es decir $\frac{N(N-1)}{2}$. La probabilidad de que cada arista existe viene definida por el propio modelo como p, asi siguiendo la notación obtenemos la ecuación:

$$P(|\mathcal{E}| = L) = P(L) = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}.$$

□

Gracias a esta fórmula podemos calcular el promedio de aristas de una red Erdös-Rényi (N, p):

Proposición 2. *El promedio de aristas de una red $ER(N, p)$ es:*

$$\langle L \rangle = p\frac{N(N-1)}{2}.$$

Demostración:
La demostración es sencilla, se trata de calcular la esperanza de esta distribución, siguiendo la notación de la proposición anterior la esperanza de la binomial es: $E[X] = p \cdot n$, por tanto:

$$\langle L \rangle = E[X] = p\frac{N(N-1)}{2}.$$

□

Gracias a la Proposición 2 y a la Definición 3, podemos obtener rápidamente el grado medio de los nodos de la red $ER(N,p)$ simplemente sustituyendo los valores conocidos:

Proposición 3. *El grado promedio de las redes $ER(N,p)$ es:*

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1).$$

La característica principal de este modelo es la homogeneidad del grado respecto a distintas métricas como la distancia entre nodos o el grado en los mismos, manteniendo una construcción imparcial para todos los nodos de la red. Las aplicaciones de este tipo de redes son realmente bastante limitadas, ya que en la realidad pocas infraestructuras tienden a comportarse como se describe en el modelo ER, no obstante, supone una base teórica fundamental para el desarrollo de redes aleatorias, lo que justifica la importancia de su estudio.

4.2.2. Modelo de Barabasi-Albert. Redes libres de escala.

El procedimiento de Barabasi-Albert [19] para crear redes aleatorias busca encontrar una mayor fidelidad entre las redes generadas y las que se presentan en situaciones reales, para ello se apoya en dos puntos clave:

- Crecimiento: Las redes no se crean estáticas, es decir no conforman un número de nodos fijo en su creación y perduran con ese mismo tamaño, sino que crecen.
- Unión Preferencial: Cuando un nuevo nodo se incorpora a la red no todas las relaciones posibles tienen la misma probabilidad de aparecer, sino que se tiende a crear aristas con aquellos nodos que más aristas tienen de por sí. Esto provoca que la distribución que siguen los grados de las aristas sea de tipo potencial:
$$P(k) = k^{-\lambda} \quad con \quad \lambda > 0.$$
A las redes complejas que siguen una distribución de este tipo se les denomina redes libres de escala. A este fenómeno también se le denomina *Efecto Mateo* [13], un concepto heredado de la Sociología.

Mientras que otros modelos de grafos aleatorios, como el de Erdös-Rényi, buscan crear un grafo ajustado a unas características estructurales específicas, los libres de escala se centran en representar la dinámica de la red.

Definición 12. *El modelo de Barabasi-Albert (BA) es un modelo que permite crear grafos de escala libre utilizando tres parámetros:* N, n_0, m *(*$m \leq n_0 << N$*). Denótese* n_t *y* l_t *el número de nodos y aristas existentes en un tiempo* t *respectivamente. El proceso de creación es el siguiente:*

0. *Se crea un grafo completo de* n_0 *nodos, etiquetados como* $1, ..., n_0$*.*
1. *Se establece* $t = 1$*.*
2. *Un nuevo nodo etiquetado como* $n = n_0 + t$ *se añade al grafo.*
3. *Se añaden* m *aristas conectando el nodo* n *con otros ya existentes. La probabilidad de que el nodo* n *se conecte con un nodo* i *se calcula como:*

$$P_{n \to i} = \frac{k_{i,t-1}}{2l_{t-1}},$$

   donde $k_{i,t-1}$ *es grado del nodo* i *a tiempo* $t - 1$*.*
4. *Si* $t \neq N - n_0$ *se repite el proceso desde el paso 2 con* $t = t + 1$*.*

Proposición 4. *Un grafo creado a través del modelo* $BA(N, n_0, m)$ *posee un total de aristas* $L = m(N - n_0) + \frac{n_0(n_0-1)}{2}$ *y el grado medio* $\langle k \rangle$ *tiende al valor* $2m$*.*

Demostración:
La demostración del número de aristas es sencilla ya que el primer sumando de la proposición corresponde al grafo completo creado inicialmente y el segundo sumando se obtiene creando m aristas por cada nuevo nodo añadido, es decir, $N - n_0$.

En el caso del grado medio este puede calcularse con la fórmula que aparece en la Definición 3, en este caso sería:

$$\langle k \rangle = \frac{2L}{N} = \frac{2m(N - n_0) + \frac{n_0(n_0-1)}{2}}{N}$$

Normalmente este proceso se simula creando un grafo completo inicial de n_0 nodos muy pequeño, por lo que se asume que este valor respecto a la cantidad de aristas que se puede crear en el proceso es despreciable, por esa razón en la literatura se puede encontrar la siguiente expresión:

$$\langle k \rangle = \frac{2L}{N} = \frac{2mN}{N} = 2m$$

Esta expresión no es totalmente fidedigna al modelo, pero es la que se suele usar en la experimentación.

□

Podemos observar algunos ejemplos de redes $BA(7, 3, 2)$ en la Figura 4.6

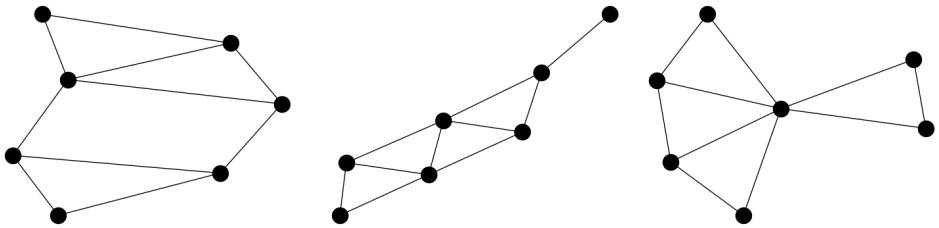

Figura 4.6: Ejemplos de redes $BA(7, 3, 2)$.

El modelo de Barabasi-Albert busca simular redes existentes en la realidad. Este fenómeno propicia la creación de nodos cuyo grado individual es altamente superior a la media del grado de la red ($k_i >> \langle k \rangle$). A estos nodos se les denomina hubs y pueden ser identificados en redes reales como redes sociales o internet. El hecho de que haya una creación imparcial de las conexiones otorga mucha heterogeneidad a las métricas típicas de los grafo como los grados de los nodos y las distancias entre pares.

4.2.3. Modelo de Watts-Strogats. Redes de mundo pequeño.

En los años 60 el psicólogo Stanley Milgran presenta un artículo utilizando a la población estadounidense donde muestra que por muy grande que sea el número de personas que residan en el país, cualquiera dos de ellas están conectadas a través de un número muy pequeño de relaciones. Posteriormente en 2003 un grupo de científicos de la Universidad de Colombia repiten una versión moderna del experimento de Milgran con más de 60 mil personas involucradas de todo el mundo, demostrando que la conexión entre ellas era realmente estrecha.

A este experimento se le bautizó como el fenómeno de los seis grados de separación, dado que cualquier persona puede llegar a otra a través de un camino de 6 conexiones.

Este concepto ha sido trasladado a otras muchas redes distintas, donde aunque el grado de separación cambia, se observa que, por muy grande que sea el número de nodos, el camino más largo que lo separa es realmente corto.

Matemáticamente podemos relacionar este concepto con el crecimiento de los nodos de una red con la longitud media de la misma, definiendo el concepto de Redes de Mundo Pequeño.

Definición 13. *Una red presenta comportamiento de mundo pequeño (small-world behaviour) si su longitud media crece de manera proporcional al logaritmo del número de nodos.*

$$\langle d \rangle \in \mathcal{O}(logN).$$

El modelo de Watts-Strogatz busca reproducir este fenómeno muy presente en redes reales y especialmente en redes informáticas, donde la conectividad es realmente alta.

Definición 14. *El modelo de Watts-Strogatz [3] es un modelo que permite crear grafos aleatorios con la característica de mundo pequeño.*

Para construir una red $WS(N, m, p)$ se comienza el proceso con un grafo circular N, m, posteriormente se consideran cada una de las aristas del grafo pudiendo cambiar sus conexiones con una probabilidad p de la siguiente manera:

1. *Visita cada nodo del grafo en sentido de las agujas del reloj, es decir, estando etiquetados, desde el nodo 1 al nodo N.*
2. *Sea i el nodo actual. Cada arista que conecta el nodo i con sus m vecinos en sentido de las agujas de reloj tiene una probabilidad p de editar su conexión.*
3. *Editar su conexión significa cambiar el extremo de la arista que no corresponde al nodo i. La elección del nuevo extremo se escoge a través de una distribución uniforme con el resto de nodos del grafo, con las limitaciones de que no existan dos aristas iguales y que un nodo se conecte a si mismo.*

Proposición 5. *Dado uno grafo $WS(N, m, p)$ se cumple que el número de aristas es $L = \frac{N \cdot m}{2}$ y el grado medio es $\langle k \rangle = m$.*

Demostración:
Como se observa en el procedimiento del modelo WS, el número de nodos y de aristas no cambia respecto al grafo circular N, m, por lo que su grado medio tampoco. En el grafo circular existen N nodos cada uno de ellos con m conexiones, lo que conforma un total de $L = \frac{N \cdot m}{2}$ aristas, ya que cada una se cuenta dos veces.

Siguiendo la Definición 3 se puede obtener:

$$\langle k \rangle = \frac{2L}{N} = \frac{2N \cdot m}{2N} = m$$

□

Aún existiendo el proceso de edición de conexiones se puede comprobar que la característica diferencial de mundo pequeño se mantiene para los distintos

valores de p. En la literatura no se ha encontrado una demostración formal de esta idea, pero si se ha realizado una experimentación con miles de valores para los parámetros N y p en [10] que muestra este resultado.

Podemos observar algunos ejemplos de redes $WS(7, 4, 0.5)$ en la Figura 4.7.

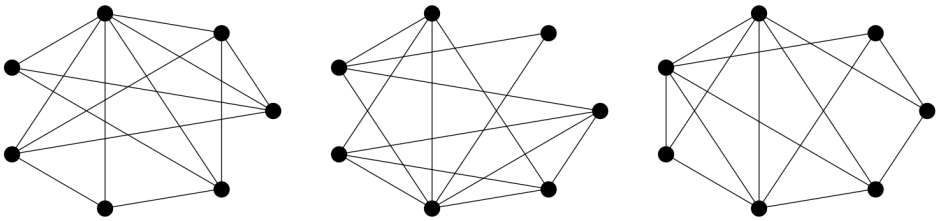

Figura 4.7: Ejemplos de redes $WS(7, 2, 0.5)$).

El modelo de Watts-Strogatz tienen múltiples aplicaciones en la representación de redes reales. El ejemplo más conocido es el asociado a la topología de internet que se puede simular con una red WS con un valor alto del parámetro p.

Por otro lado cuando el parámetro p es considerablemente bajo, las redes más similares son las denominadas *peer-to-peer* o *P2P*. En este tipo de topologías todos los usuarios se comportan como iguales, sin una diferenciación clara entre cliente y servidor, como el caso de la aplicación *µTorrent*.

La característica principal de estas redes será la de simular el comportamiento de mundo pequeño, pero se puede aportar heterogeneidad aumentando el parámetro p. Cuando el grado es muy bajo las redes son muy simétricas y son con diferencia las más homogéneas, cuando aumenta se logra mayor diversidad, pero no alcanzando a las redes Barabasi-Albert.

4.3. Medidas de Centralidad

El concepto de centralidad en una red es esencial ya que nos proporciona medidas cuantitativas para determinar la importancia o la influencia de los nodos en un sistema. Como se mostrará en la Sección 6.2, la estructura de la red influye en el comportamiento de los modelos a implementar. Adelantando el estudio que se realizará en la Sección 6.2.3 existen ciertos nodos críticos en las estructuras que influyen en los daños producidos por el malware en las redes que se estudian. Identificarlos nos permiten añadir una nueva capa de estudio, donde los atacantes y defensores podrán utilizar este conocimiento en su beneficio.

La definición de centralidad en un grafo no es única y existen distintas funciones para medirla. Las que se van a presentar en este trabajo están basadas

en dos conceptos mencionados en la Sección 4.1: el grado de los nodos y los caminos. Esta decisión es tomada de manera consciente ya que el comportamiento de los virus está influenciado tanto por la cantidad de elementos cercanos a los que se puede transmitir, como por las características de los caminos que las cepas pueden seguir entre los nodos infectados y los que no.

4.3.1. Centralidad de Grado. *Degree Centrality.*

Definición 15. *Dado un grafo $G = (\mathcal{V}, \mathcal{E})$ y nodo $i \in \mathcal{V}$, se mide la centralidad de grado del nodo i como:*

$$c_i^D = k_i.$$

Es decir, la centralidad de grado es equivalente al grado del nodo i.

Aunque el concepto y cálculo de esta medida de centralidad es realmente sencilla, intuitivamente se puede observar su importancia. Un virus que alcanza los nodos con mayor grado obtiene rápidamente muchos otros nodos candidatos a los que propagarse.

Como se ha indicado en la Sección 4.2 donde se explica el modelo de Barabasi-Albert, los nodos con una centralidad de grado mucho mayor a la media de la red se denominan *hubs*. Estos son grandes candidatos a convertirse en nodos críticos de la red, siendo muy peligrosos si el atacante es capaz de alcanzarlos.

4.3.2. Centralidad de cercanía. *Closeness Centrality.*

Definición 16. *Dado un grafo $G = (\mathcal{V}, \mathcal{E})$ conexo y un nodo $i \in \mathcal{V}$, se mide la centralidad de cercanía del nodo i como:*

$$c_i^C = \frac{1}{\sum_{j=1}^{N} d_{ij}}.$$

Es decir, el inverso de la suma de la distancia de todos los caminos mínimos entre el nodo y el resto del grafo.

Esta medida de centralidad surge de un concepto relativamente sencillo, un nodo es central si puede interactuar con cualquier otro de la red rápidamente.

El cálculo de esta centralidad está relacionado con la ejecución del algoritmo de *Dijkstra*, un algoritmo que permite calcular de manera voraz el camino más corto entre un nodo y el resto de la red en complejidad $\mathcal{O}(n \cdot log\ n)$. Dado que el trabajo se centra en grafos no ponderados, este algoritmo se puede mejorar mediante el uso de un recorrido en anchura o *BFS* (*Breadth-First Search*) que se ejecuta en complejidad $\mathcal{O}(n)$.

4.3.3. Centralidad de Intermediación. *Betweenness centrality.*

Es frecuente que la interacción entre pares de nodos de una red esté supeditada a otros individuos, especialmente ocurre entre aquellos que se encuentran más frecuentemente en los caminos que unen esos pares.

Estos nodos por tanto influencian el resto, por lo que vuelve a subyacer la idea de centralidad.

Definición 17. *Dado un grafo $G = (\mathcal{V}, \mathcal{E})$ conexo y un nodo $i \in \mathcal{V}$, se mide la centralidad de intermediación del nodo i como:*

$$c_i^B = \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\substack{k=1 \\ k \neq i,j}}^{N} \frac{\sigma_{jk}(i)}{\sigma_{jk}},$$

donde σ_{jk} representa la cantidad de caminos más cortos entre j y k; y $\sigma_{jk}(i)$ representa la cantidad de caminos más cortos entre j y k que contienen el nodo i.

Esta medida de nuevo utiliza los algoritmos de caminos mínimos *Dijkstra y BFS* para calcular su valor, dependiendo de si el grafo es o no ponderado. Vease la importancia de que los índices de los sumatorios no coincidan de forma que el uso de esta fórmula no pueda dar problemas de definición.

4.3.4. Comparativa de centralidades.

Una vez revisadas las medidas a utilizar en este trabajo, véase un ejemplo del cálculo de las centralidades de los diferentes nodos en un grafo dado. Para ilustrar este ejemplo se va a utilizar el grafo indicado en la Figura 4.8. Pueden verse el orden de importancia de los nodos según las distintas medidas en la Tabla 4.1.

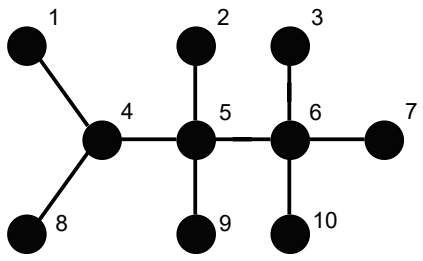

Figura 4.8: Cálculo de las medidas de centralidad de un grafo.

En las distintas medidas se observa como los nodos más centrales son los nodos 5, 6 y 4. Dependiendo de la medida utilizada se considera más la importancia de

Ranking	c_i^D	c_i^C	c_i^B
1	5,6	5	5
2	4	6	6
3	Resto	4	4
4		2,9	Resto
5		3,7,10	
6		1,8	

Tabla 4.1: Medidas de centralidad del grafo representado en la Figura 4.8.

uno o de otro. Cabe destacar que la centralidad por cercanía es capaz de clasificar el resto de nodos, mientras que el resto de medidas los toman a todos por iguales.

4.4. Redes Complejas en Modelos Compartimentales

Uno de los objetivos de esta sección es la definición de una notación común para el trabajo, esta notación es en parte original, rellenando aquellos elementos que en la literatura se dan por supuestos e introduciendo algunas especificaciones, de forma que los modelos implementados queden claros en el resto del trabajo.

Los Modelos Compartimentales en Redes Complejas tienen características generales de forma que la implementación de modelos distintos varían según la dinámica y compartimentos del fenómeno a representar, pero siguen una estructura general similar.

Para la modelización de las pandemias digitales y biológicas se sigue un proceso de discretización del tiempo, de tal forma que en un tiempo $t \in \mathbb{N}$ un individuo pertenece a un único compartimento del modelo.

Notación 2. *Si el individuo representado por $v \in \mathcal{V}$ pertenece al compartimento C en el momento t, se usará la notación de conjuntos:*

$$v \in C_t \quad ó \quad v \in C,$$

omitiendo la variable temporal si el contexto lo permite.

La evolución de los individuos que componen la red está basada en transiciones probabilísticas. Es decir, siguiendo la taxonomía explicada en la Sección 3 estamos ante un modelo estocástico, discreto e individual. El hecho de pasar de un compartimento a otro en un momento específico es un proceso aleatorio que dependerá del parámetro indicado por el modelo y del tipo de transición.

En estos modelos existen dos tipos de transiciones:

- Transiciones basadas en el contacto (*contact-based transitions*): La probabilidad de estas transiciones dependen del estado del nodo actual y del conjuntos de nodos adyacentes.
- Transiciones espontáneas (*espontaneous transitions*): La probabilidad de estas transiciones dependen únicamente del estado del nodo actual.

Actualmente y debido a su modernidad no existe un lenguaje unificado para la representación de estos modelos, por tanto, a lo largo de este trabajo se utilizará el color azul para indicar que la transición está basada en el contacto y el color naranja para indicar que la transición es espontánea.

Dado que existen las transiciones basadas en el contacto, es cómodo para el resto del trabajo definir una notación que permita referirse a los nodos adyacentes que pertenezcan a un compartimento:

Notación 3. *Dado un modelo compartimental con un conjunto de compartimentos $\mathcal{C} = \{C_1, ...C_n\}$ y una red compleja formada por un grafo $G = (\mathcal{V}, \mathcal{E})$, se denota la cantidad de nodos adyacentes a un nodo $v \in \mathcal{V}$ pertenecientes a un compartimento $C \in \mathcal{C}$ como:*

$$C(v) = |\{w \in C : \{v, w\} \in \mathcal{E}\}|$$

Como ya se ha mencionado no existe una notación unificada, por lo que se definirá una de manera que puedan explicarse sintéticamente las transiciones de los modelos:

Notación 4. *Dado el conjunto de compartimentos $\mathcal{C} = \{C, D, ...\}$ y la red compleja $G = (\mathcal{V}, \mathcal{E})$. Si un nodo v pertenece a un compartimento C en un tiempo t, se denotará sus posibles transiciones al momento $t + 1$ como:*

$$v \in C_t \Rightarrow \begin{cases} p(v \in C_{t+1}) &= 1 - \alpha, \\ p(v \in D_{t+1}) &= \alpha, \end{cases} \quad \hookrightarrow \quad v \in C \Rightarrow \begin{cases} 1 - \alpha &: C, \\ \alpha &: D. \end{cases}$$

Es decir, el nodo v en el siguiente instante $t+1$ puede transicionar al compartimento D con probabilidad α o mantenerse en el mismo compartimento C con probabilidad $1 - \alpha$.

Esta notación va a ser utilizada a lo largo del trabajo a la hora de exponer los modelos.

5

Modelo SIR para la modelización de ciberpandemias.

En este capítulo se va a estudiar el modelo SIR como preámbulo al modelo SIH-UA. Véase que parte del mismo ha sido comentado ya en la Sección 3.2 por lo que se centrará principalmente en su implementación y conclusiones que se pueden obtener de los modelos en redes complejas y en ecuaciones diferenciales.

5.1. Modelo SIR en ecuaciones diferenciales.

El modelo SIR [17] fue claramente diseñado para modelizar la transmisión de un virus biológico, pero podemos realizar una analogía con una red de ordenadores, donde los elementos susceptibles ante un virus informático son aquellos dispositivos no protegidos ante un posible malware, los infectados son aquellos que ya están infectados con el mismo y los recuperados son los que o bien han sucumbido al virus en caso de que este sea destructor o bien ya poseen un software de protección contra el mismo.

La dinámica del modelo se puede observar en la Figura 5.1:

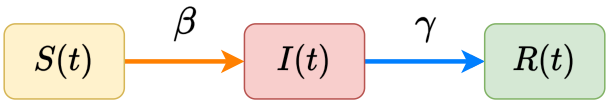

Figura 5.1: Diagrama del modelo SIR.

Cada una de las variables del modelo evolucionan con el paso del tiempo t, el parámetro β mide la velocidad con el que el virus o software malicioso se propaga y se denomina "tasa de transmisión", por otro lado el parámetro γ mide la velocidad con que el que un individuo o dispositivo infectado pasa a ser inmune y por tanto recibe el nombre de "tasa de recuperación".

El sistema de ecuaciones diferenciales que lo modela es el siguiente:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI, \\ \frac{dI}{dt} &= \beta SI - \gamma I, \\ \frac{dR}{dt} &= \gamma I. \end{aligned} \tag{5.1}$$

Este modelo asume que la población de estudio es invariante, es decir $N = S + I + R$ es constante. Por otro lado, el sistema se inicia con $S(0) = N - (N\rho_0)$, $I(0) = N\rho_0$ y $R(0) = 0$, el parámetro ρ_0 indica la fracción de población infectada al comienzo del estudio.

Se puede observar que las transiciones del modelo observado en la Figura 5.1 no se implementan de la misma manera en la ecuación (5.1). El descenso de población susceptible, dirigido a través del parámetro β, se ve influenciado no solo por si mismo, sino que también por la cantidad de infectados que existen, de esta manera se modela la interacción entre estos compartimentos, ya que la forma de transmisión de un virus es a través del contacto.

En cambio, el descenso de infectados, dirigido por el parámetro γ solo se ve influenciado por la cantidad de población del mismo compartimento, ya que la interacción no produce ningún efecto.

Estas ideas serán importantes a la hora de implementar el modelo SIR en redes complejas y el modelo descrito en [1] en ecuaciones diferenciales.

Algunas características interesantes de este modelo se pueden observar en [17],

se van a mencionar algunas de las más interesantes:

Proposición 6. *El modelo de ecuaciones diferenciales SIR se estabiliza, es decir, existen* $S(\infty), I(\infty), R(\infty) \in \mathbb{R}$ *tales que:*

$$\lim_{t\to\infty} S(t) = S(\infty)\ ,\ \lim_{t\to\infty} I(t) = I(\infty)\ ,\ \lim_{t\to\infty} R(t) = R(\infty).$$

El uso de $S(\infty), I(\infty), R(\infty)$ viene justificado por usar una notación descriptiva y más acotada.

Demostración:
Como se observa en (5.1) y se ha mencionado anteriormente:

$$\frac{dS}{dt} \leq 0\ ,\ \frac{dR}{dt} \geq 0 \quad y \quad 0 \leq S(t), R(t) \leq N.$$

Esto es suficiente para que los límites de $S(t)$ y $R(t)$ existan, lo que implica que existe el límite para $I(t)$, ya que podemos reescribir como: $I(t) = N - S(t) - R(t)$.

□

Otra proposición que cabe mencionar explica de forma muy general cómo evoluciona el modelo:

Proposición 7. *El modelo de ecuaciones diferenciales SIR no solo se estabiliza, sino que la transmisión del virus se erradica, es decir, para cualquier valores de condiciones iniciales* $N, \rho_0, \beta, \gamma > 0$*, se tiene que:*

$$I(\infty) = 0.$$

Demostración:
Se puede demostrar por reducción al absurdo, asúmase $I(\infty) \neq 0$, véase que por definición el caso $I(\infty) < 0$ no es posible, ya que el conjunto de individuos de un compartimento siempre se mantiene mayor o igual a 0. Por tanto $I(\infty) > 0$, dada esta condición se puede observar que:

$$\exists\, t_0 \in [0, \infty)\ :\ I(t) > 0\ \ \forall t \geq t_0$$

Una vez obtenido t_0, usando la tercera igualdad de (5.1), dado que $\gamma I(t) > 0$ se tiene que para esos valores de t:

$$\boxed{\frac{dR}{dt} > 0 \ \forall\ t \geq t_0}$$

Por otro lado tomando límites sobre la expresión $\frac{dR}{dt} = \gamma I$ se tiene que:

$$\lim_{t\to\infty} \frac{dR}{dt} = \lim_{t\to\infty} \gamma I(t) = \gamma I(\infty) > 0$$

$$\boxed{\lim_{t\to\infty} \frac{dR}{dt} > 0 \ \ \forall t \geq t_0}$$

Gracias ambas expresiones recuadradas se puede concluir que:

$$\lim_{t\to\infty} R(t) = R(\infty) = \infty$$

El hecho de que el límite diverja es una contradicción con la Proposición 6, por lo tanto:

$$I(\infty) = 0.$$

□

Para la resolución de este modelo se ha decidido utilizar técnicas numéricas, específicamente el método de Euler con diferencias finitas ([6]) y paso de discretización $h = 0.005$. Previo a aplicar el método de resolución se explota la característica de que la población es invariante y por tanto podemos calcular un compartimento respecto al resto: $R = N - S - I$, eliminando la última ecuación del sistema (5.1). De esta forma el sistema de ecuaciones en diferencias finitas obtenido es el siguiente:

$$\begin{aligned} S_{i+1} &= h(-\beta S_i I_i) + S_i, \\ I_{I+1} &= h(\beta S_i I_i - \gamma I_i) + I_i, \\ R_{i+1} &= N - S_{i+1} - I_{i+1}. \end{aligned} \tag{5.2}$$

Se deben incluir unas condiciones iniciales al modelo, las cuales vienen definidas por la población inicial y el número de infectados al comienzo de la simulación, de esta manera $S_0 = N - (N\rho_0)$, $I_0 = N\rho_0$ y $R_0 = 0$.

En la Figura 5.2 podemos ver la evolución de una posible ciberpandemia a través del modelo SIR, con los siguientes parámetros:

Símbolo	N	k	ρ_0	β_0	γ	β
Rango	$[1, \infty]$	$[1, \infty]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$
Valor	50000	10	0.0001	0.07	$1/5$	$\beta_o \cdot \frac{k}{N}$

Tabla 5.1: Conjunto de Valores para el modelo SIR en ecuaciones diferenciales

El parámetro k se introduce para medir el grado de interacción entre los individuos del sistema. Permite además hacer una analogía con el modelo implementado en ambos esquemas: ecuaciones diferenciales y redes complejas.

En el caso del modelo, β_0 mide la probabilidad de que un individuo infecte a cualquiera de la red. Se multiplican por la razón $\frac{k}{N}$, de forma que β mide la probabilidad de que cada individuo solo pueda infectar a la proporción de la población con la que tiene contacto, es decir el grado de conexión k entre todos los individuos de la red N.

Se puede realizar una analogía con las implementaciones en redes, donde la conexión es el grado de los nodos y el total de población el total de nodos del grafo. En la Sección 6.2.1 se profundiza en esta comparativa entre implementaciones.

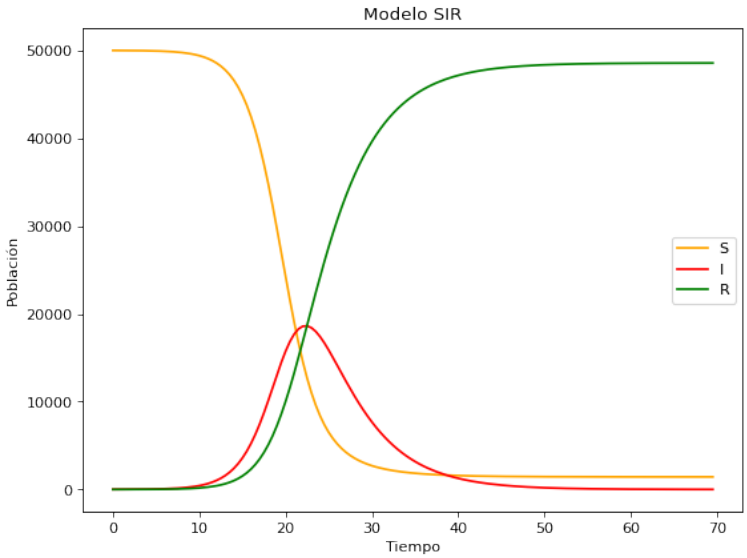

Figura 5.2: Modelo SIR con parámetros de la Tabla 5.1

En la Figura 5.2 se puede observar como se cumplen las proposiciones anteriormente mencionadas ya que las variables de Susceptibles y Recuperados tienen tendencias a valores fijos y la variable de Infectados tiende al 0.

5.2. Modelo SIR en Redes Complejas

Existe una clara analogía entre el modelo implementado en ecuaciones diferenciales y el modelo implementado a través de redes. El número de contagios variará según la tasa de transmisión y la cantidad de individuos contagiados alrededor, mientras que para la tasa de recuperación el contexto no es decisivo.

Gracias al modelo en redes complejas se pueden simular los contactos exactos entre individuos como aristas en la red, permitiendo realizar experimentos mucho más complejos, como los que se pueden observar en la Sección6.2.3.

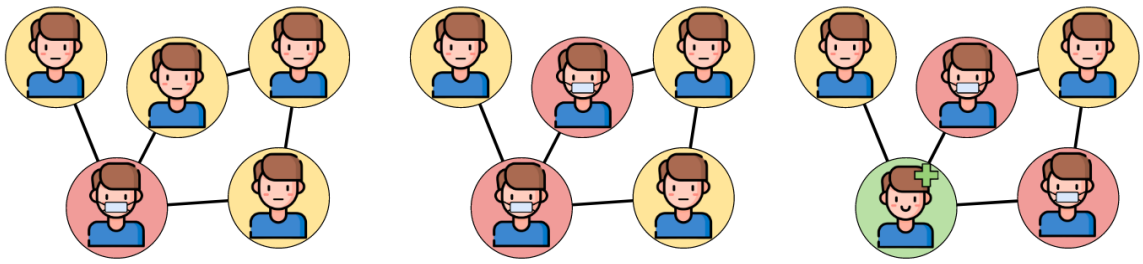

Figura 5.3: Ejemplo de transmisión del modelo SIR en una red.

En este caso en vez de tener un sistema de ecuaciones diferenciales lo que tenemos es un modelo de transiciones que nos indicará que probabilidad tiene un individuo en cambiar su estado de la siguiente forma:

$$
\begin{aligned}
v \in S \Rightarrow & \begin{cases} \beta \cdot I(v) & : \quad I, \\ 1 - \beta \cdot I(v) & : \quad S, \end{cases} \\
v \in I \Rightarrow & \begin{cases} \gamma & : \quad R, \\ 1 - \gamma & : \quad I, \end{cases} \\
v \in R \Rightarrow & \{1 \quad : \quad R.
\end{aligned}
\tag{5.3}
$$

Al igual que ocurría en el modelo de Ecuaciones Diferenciales, la población (N) es constante y definimos el parámetro ρ_0 como la fracción de individuos infectados al comienzo de la simulación.

Modelando el sistema en Python podemos obtener una comparativa respecto al modelo en ecuaciones diferenciales.

Para la primera simulación se realizan 100 ejecuciones sobre una red de Erdös-Renyi(5000,0.1). En este caso $N = 5000$, $\rho_0 = 0.001$, $\beta = 0,07$, $\gamma = 1/5$. Se puede

observar que hemos reducido en un orden la cantidad inicial de la población respecto al modelo en ecuaciones diferenciales (Figura 5.2), la razón fue que el algoritmo planteado no requiriese de un tiempo elevado de ejecución. Podemos ver el resultado en la Figura 5.4

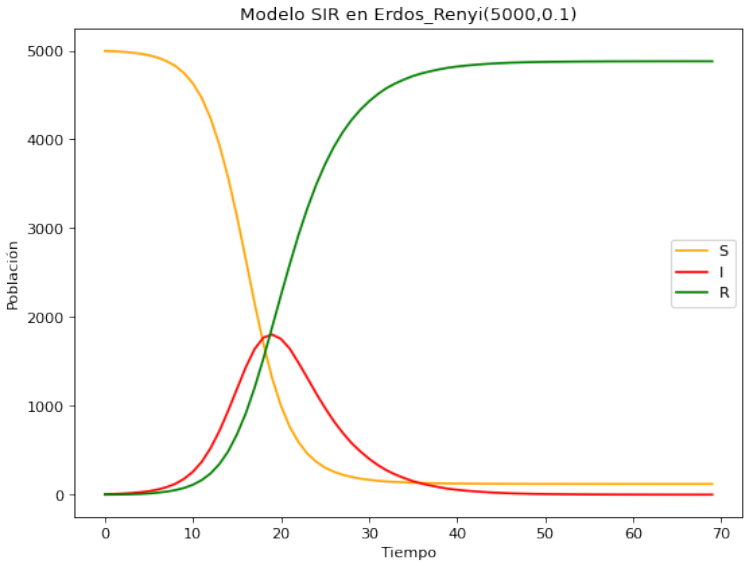

Figura 5.4: Modelo SIR en Redes Complejas.

De nuevo en este modelo se pueden observar las 3 fases que se observan en el modelo de ecuaciones diferenciales. Se puede realizar una comparativa clara entre ambas implementaciones, pero dado que el trabajo pasa por el análisis del modelo SIH-UA se va a realizar esta comparativa de manera más explicita con él. Véase la Sección 6.2.1.

6

Modelos SIH-UA

Este capítulo compone el grueso del trabajo, consiste en el análisis del modelo SIH-UA que se puede encontrar en el artículo [1]. Este es un modelo creado específicamente para la simulación de ciberepidemias, que tiene en cuentas las vicisitudes específicas que diferencian estos sucesos con las pandemias biológicas.

Se establecen dos conjuntos de estados para cada elemento de la población. El primer conjunto determina la situación del individuo respecto a la propagación del nuevo malware:

- Susceptible: S (*Susceptible*).
- Infectado: I (*Infected*).
- Curado: H (*Healed*).

Podemos observar como estos 3 compartimentos tiene una clara relación con el modelo SIR explicado en la Sección 5, únicamente cambiando el nombre del tercer compartimento por coherencia con el artículo. El segundo conjunto de estados determina si el individuo es consciente de la existencia del virus en la red o no.

- Inconsciente: u (*Unaware*).
- Consciente: a (*Aware*).

Esto establecería un total de 6 estados, pero en el modelo se establece que un dispositivo puede estar curado únicamente si este es consciente de la existencia del

malware, ya que de otra forma no habría puesto medidas contra el mismo, como un antivirus. Por tanto se establecen un total de 5 compartimentos, el diagrama que define el modelo es el siguiente:

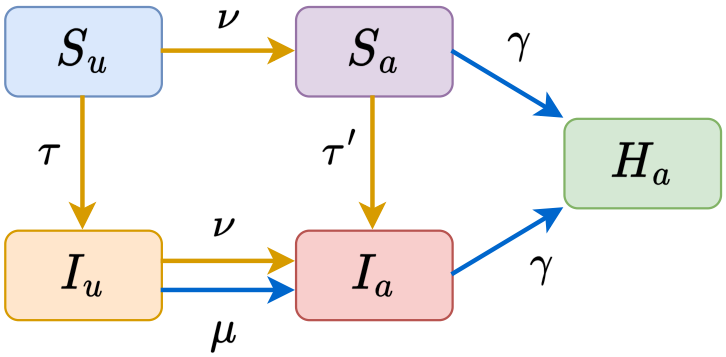

Figura 6.1: Modelo definido en [1]

Podemos observar 4 parámetros esencialmente distintos:

- Tasa de transmisión (τ) : Es equivalente al parámetro β del modelo SIR, mide la velocidad con la que el malware se propaga. En el caso que los individuos ya sean consciente de la presencia del malware en la red la tasa de transmisión se reduce a $\tau' = \tau/10$ ya que es más difícil que estos se infecten.
- Tasa de concienciación por contacto (ν) : Es equivalente al parámetro τ pero midiendo la velocidad con el que la concienciación sobre el malware se propaga.
- Tasa de concienciación espontánea (μ) : Mide la probabilidad de que un individuo se de cuenta de que esta siendo víctima de un malware, por tanto esta transición solo se produce en individuos ya infectados.
- Tasa de recuperación (γ): De nuevo es equivalente al parámetro γ del modelo SIR, mide la velocidad con la que un individuo consciente pone medidas de protección contra el virus.

Siguiendo la notación definida en la Sección 4.4 podemos observar que las transiciones de transmisión y concienciación por contacto dependen del estado de los individuos en contacto con el que se encuentra en estudio, mientras que las transiciones de concienciación espontánea y recuperación únicamente dependen del estado individual o global del sistema.

Una característica que se introduce en el modelo mencionado es el control sobre el impacto de la presencia de un malware en un sistema. Existen una gran

variedad de ataques que pueden ser programados en el malware, algunos de ellos más agresivos y dañinos. Este fenómeno se modela a través de un parámetro $d \in [0, 1]$. Este parámetro también nos permite modificar la tasa de concienciación del virus, cuánto más agresivo sea el malware, antes serán conscientes los usuarios de que están infectados. Debido a esto el, parámetro μ debe ser proporcional al daño percibido:

$$\mu = \begin{cases} \mu_0(d-\theta) & si \quad d \geq \theta, \\ 0 & si \quad d < \theta. \end{cases}$$

Donde θ se define como el límite a partir que el daño comienza a ser percibido. Un límite muy bajo puede modelizar infraestructuras de alta seguridad donde un daño leve es rápidamente monitorizado, mientras que un límite más alto modeliza sistemas con tolerancia al fallo, donde daños leves podría considerarse como errores que ignorar o solucionar.

El último parámetro que no se ha explicado todavía es ρ_0 que define la fracción de individuos infectadas al comienzo del modelo, de la misma manera que en modelos anteriores.

6.1. Modelo SIH-UA en ecuaciones diferenciales.

A diferencia del modelo SIR, que surge como un modelo de ecuaciones diferenciales y posteriormente ha podido ser implementado en redes complejas, el modelo SIH-UA es un modelo que se origina desde la perspectiva en redes complejas, pero que puede ser implementado a través de ecuaciones diferenciales.

Como se menciona en el artículo el modelo posee transiciones entre compartimentos que dependen de la interacción (véase la Figura 6.1), estos son los definidos por ν y τ.

El parámetro τ mide la tasa de infección y por lo tanto se ve influenciado por la cantidad de población existente en el compartimento de susceptibilidad (S_u *ó* S_a), y por la cantidad de población que ya está infectada, es decir, $I_u + I_a$.

En el caso del parámetro ν, que mide la tasa con la que la población se vuelve consciente del virus, se ve influenciada por la cantidad de población en el compartimento inconsciente (S_u *ó* I_u) y por la concienciación global, es decir, $S_a + I_a + H_a$.

Por tanto se puede construir un sistema de ecuaciones diferenciales análogo al modelo SIR, pero con los nuevos compartimentos. El sistema que lo modela se observa a continuación:

$$\begin{aligned}
\frac{dSu}{dt} &= -\nu Su(Sa + Ia + Ha) - \tau Su(Iu + Ia), \\
\frac{dSa}{dt} &= \nu Su(Sa + Ia + Ha) - \tau' Sa(Iu + Ia) - \gamma Sa, \\
\frac{dIu}{dt} &= \tau Su(Iu + Ia) - \nu Iu(Sa + Ia + Ha) - \mu Iu, \\
\frac{dIa}{dt} &= \tau' Sa(Iu + Ia) + \nu Iu(Sa + Ia + Ha) + \mu Iu - \gamma Ia, \\
\frac{dHa}{dt} &= \gamma Sa + \gamma Ia.
\end{aligned} \tag{6.1}$$

Para la resolución de este modelo se ha decidido utilizar técnicas numéricas, específicamente el método de Euler con diferencias finitas ([6]) y paso de discretización $h = 0.005$. De esta forma el sistema de ecuaciones en diferencias finitas obtenido es el siguiente:

$$\begin{aligned}
Su_{i+1} &= Su_i + h[(-\nu Su_i(Sa_i + Ia_i + Ha_i) - \tau Su_i(Iu_i + Ia_i))], \\
Sa_{i+1} &= Sa_i + h[(\nu Su_i(Sa_i + Ia_i + Ha_i) - \tau' Sa_i(Iu_i + Ia_i) - \gamma Sa_i)], \\
Iu_{i+1} &= Iu_i + h[(\tau Su_i(Iu_i + Ia_i) - \nu Iu_i(Sa_i + Ia_i + Ha_i) - \mu Iu_i)], \\
Ia_{i+1} &= Ia_I + h[(\tau' Sa_I(Iu_I + Ia_I) + \nu Iu_I(Sa_i + Ia_i + Ha_i) + \mu Iu_i - \gamma Ia_i)], \\
Ha_{i+1} &= Ha_i + h[(\gamma Sa_i + \gamma Ia_i)].
\end{aligned} \tag{6.2}$$

Para el experimento realizado se han utilizado los valores para los parámetros observables en la Tabla 6.1.

Los resultados del experimento se pueden observar en la Figura 6.2

Símbolo	N	τ	ν	μ_0	γ	ρ_0
Rango	$[1, \infty]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$
Valor	1000	$0.0055 \cdot \frac{10}{N}$	$0.011 \cdot \frac{10}{N}$	0.011	0.03	0.01

Tabla 6.1: Conjunto de valores para experimentos del modelo SIH-UA en ecuaciones diferenciales.

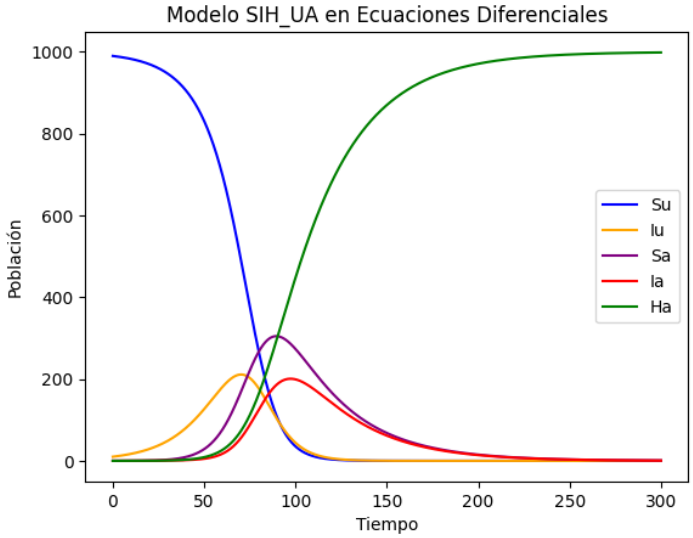

Figura 6.2: Ciclo de una ciberepidemia para el modelo implementado en ecuaciones diferenciales.

En el experimento se observan las 3 fases diferenciales de una ciber-epidemia real: Una fase preliminar de desconocimiento donde los usuarios comienzan a ser infectados, especialmente caracterizada por el descenso de usuarios Susceptibles Inconscientes (Su) y el ascenso de Infectados Inconscientes (Iu). Una segunda fase donde comienza a existir una concienciación del virus por lo que empiezan a crecer los compartimentos conscientes (Sa, Ia, Ha) y a decrecer los inconscientes (Su, Iu) y una ultima fase de erradicación donde los usuarios ponen medidas de protección, creciendo el número de usuarios curados (Ha), hasta que el virus es erradicado.

6.2. Modelo SIH-UA en Redes Complejas

Se van a realizar varios experimentos en las distintas redes que se han mencionado en la Sección 4.2. Para evitar posibles irregularidades provenientes de los procesos aleatorios se realizan 1000 iteraciones en cada uno de los experimentos realizados.

Para estos experimentos preliminares se establecen los parámetros de la siguiente manera:

Simbolo	N	θ	d	τ	ν	μ_0	γ	ρ_0	$\langle k \rangle$
Rango	$[1, \infty]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[1, N]$
Valor	1000	0.2	0.3	0.0055	0.011	0.011	0.03	0.01	10

Tabla 6.2: Conjunto de valores para experimentos del modelo SIH-UA en redes complejas.

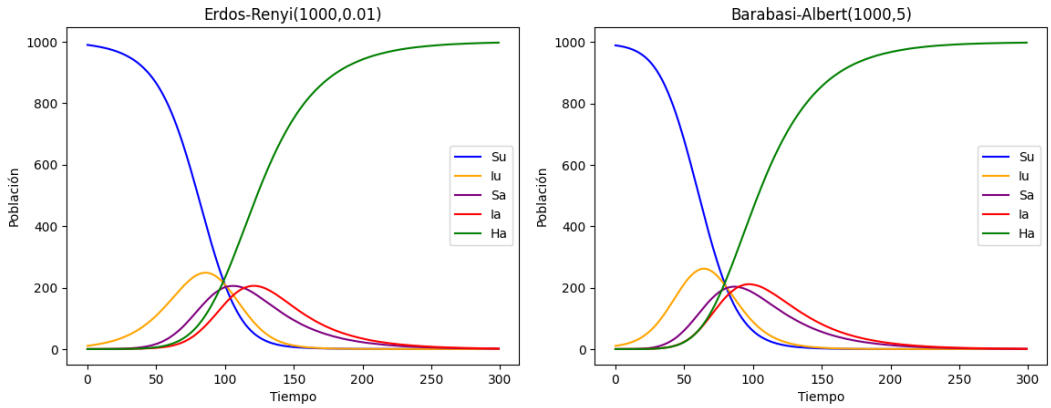

Figura 6.3: Ciclo de una ciber-epidemia para redes Erdös-Rényi y Barabasi-Albert con los parámetros de la Tabla 6.2.

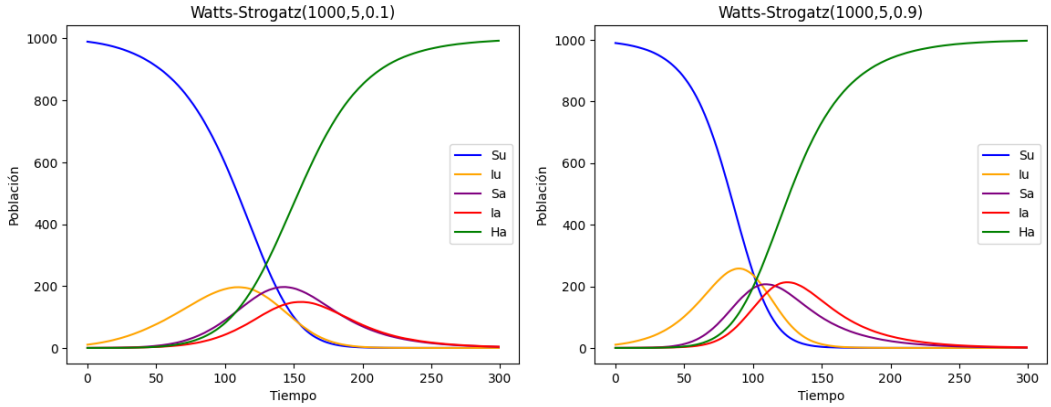

Figura 6.4: Ciclo de una ciber-epidemia para redes Watts-Strogatz con los parámetros de la Tabla 6.2.

Véase que en todas las redes se han ajustado los parámetros para que el grado medio de todas ellas sea 10.

Lo primero es que se observan en todos los experimentos las 3 fases diferenciales de una ciber-epidemia real: Una fase preliminar de desconocimientos donde los usuarios comienzan a ser infectados, especialmente caracterizada por el descenso de usuarios Susceptibles Inconscientes (S_u) y el ascenso de Infectados Inconscientes (I_u). Una segunda fase donde comienza a existir una concienciación del virus por lo que empiezan a crecer los compartimentos conscientes (S_a, I_a, H_a) y a decrecer los inconscientes (S_u, I_u) y una ultima fase de erradicación donde los usuarios ponen medidas de protección, creciendo el número de usuarios curados (H_a), hasta que el virus es erradicado.

Otra suposición que se puede realizar es como afectan cada una de las topologías de las redes al modelo descrito, aunque las diferencias no son demasiado desproporcionadas, se puede observar como en las redes $BA(1000, 5)$ las fases descritas son más estrechas, es decir, los procesos son más rápidos, mientras que en la red $WS(1000, 5, 0.1)$ ocurre lo contrario, las fases se diluyen en el tiempo.

Una posible hipótesis que justifique este hecho es cómo se distribuye el grado entre los nodos de las redes. Mientras que $WS(1000, 5, 0.1)$ es la red más regular y simétrica, existiendo pocas diferencias entre los nodos, en $BA(1000, 5)$ existen nodos *hubs* como se menciono en la Sección 4.2.2, que probablemente sean puntos críticos para los procesos en los que intervienen transiciones basadas en el contacto (τ, ν).

6.2.1. Comparativa con el modelo en ecuaciones diferenciales.

Una de las claves de este trabajo es la comparativa entre los modelos de redes complejas y ecuaciones diferenciales. Véase que siguiendo la metodología de modelización explicada en la Sección 3, ambos esquemas comparten las dos primeras fases, siendo las 3 últimas las diferenciales, el paso a la descripción de los fenómenos en los distintos modelos matemáticos.

Cuando el número de nodos es suficientemente alto, el modelo de redes de Erdös-Renyi permite la creación de una red muy equilibrada respecto al grado de los nodos de la misma. La creación de las aristas sigue una ley probabilística, lo que garantiza imparcialidad en su creación, y por tanto el promedio de las aristas tiende a la esperanza de la distribución binomial. Observando las gráficas y entendiendo el funcionamiento del modelo en ecuaciones diferenciales podemos observar que existe una clara correlación entre ambos modelos.

En la Tabla 6.2 se especificaron los parámetros de los experimentos realizados en 4 redes con grado medio $\langle k \rangle = 10$. Se puede observar que los parámetros del modelo en ecuaciones diferenciales que observamos en la Tabla 6.1 son los mismos, pero divididos entre $N = 1000$ y multiplicados por $\langle k \rangle = 10$ en caso de

que la transición este basada en el contacto.

La razón es la misma explicada en los experimentos del modelo SIR en la Sección 5. Los parámetros en el modelo de ecuaciones mide la probabilidad de transmisión de un individuo a todo el resto de la red. Multiplicar por la razón $\frac{k}{N}$ permite modificar el parámetro de forma que mida la probabilidad de transmisión a la proporción de la población con la que un individuo tiene contacto, es decir, el grado de conexión k entre todos los individuos del sistema N.

Obsérvese en la Figura 6.5 una comparativa entre sendos modelos.

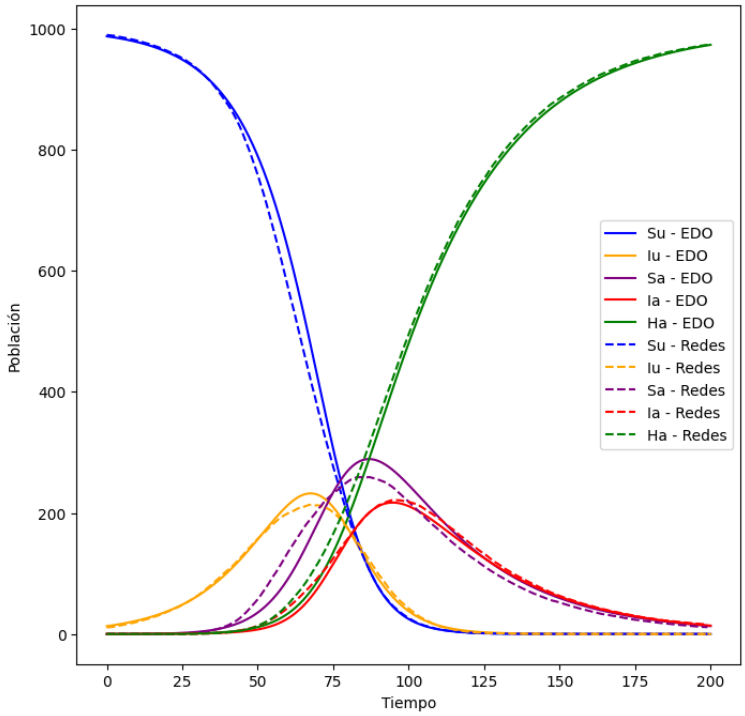

Figura 6.5: Comparativa entre modelos de redes complejas Erdös-Renyi y ecuaciones diferenciales.

Realmente podemos concluir que lo que ocurre es que el modelo en ecuaciones diferenciales se acerca a simular los sucesos en una red Erdös-Renyi. En el modelo de ecuaciones diferenciales, en el transcurso de la variable temporal los distintos compartimentos varían según el tamaño de los compartimentos relacionados, igual que ocurre en un modelo de redes complejas. La diferencia es que en el modelo de ecuaciones diferenciales se tratan los individuos como una masa y no por sus relaciones individuales; sin embargo, esto se simula en redes Erdös-Renyi cuando el número de nodos es alto y las variabilidades probabilísticas se reducen gracias a la Ley de los Grandes Números.

Cómo se puede deducir al definir los otros modelos de redes aleatorias, no existe esta correlación tan directa entre el modelo de ecuaciones diferenciales cuando se usan otros procedimientos. Como se menciona en la Sección 4.2.2, el modelo de Barabasi-Albert busca la simulación de la dinámica de algunas redes reales, caracterizadas por la presencia de hubs, en definitiva, no existe imparcialidad a la hora de la creación de las aristas. Esta diferenciación entre individuos va implicar un aumento en las diferencias respecto al modelo en ecuaciones diferenciales, como se puede observar en la Figura 6.6

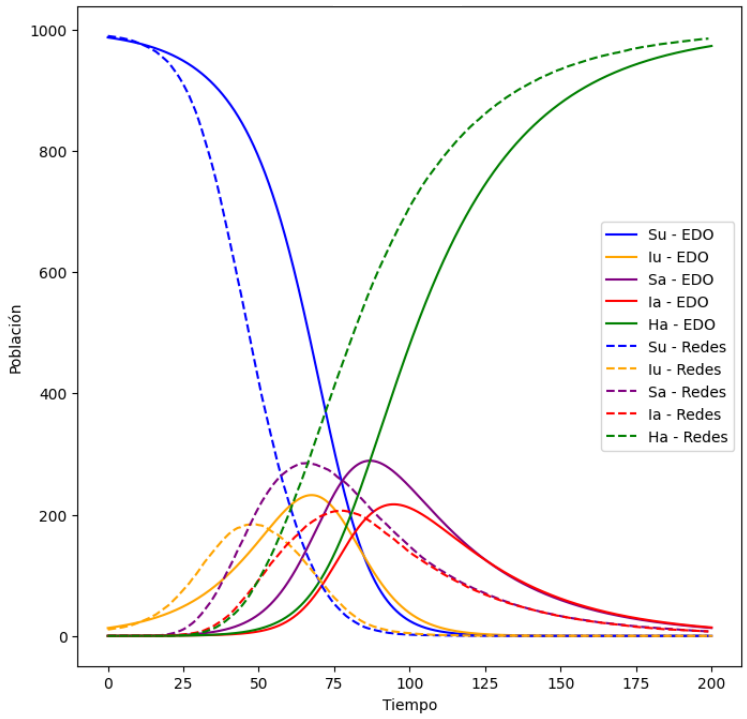

Figura 6.6: Comparativa entre modelos de redes complejas Barabasi-Albert y ecuaciones diferenciales.

Se puede realizar una comparativa entre ventajas y desventajas de sendos modelos. Por un lado el coste computacional de los modelos en ecuaciones diferenciales es mucho más bajo y además no depende del parámetro N, las gráficas se obtienen en menos de 1 segundo.

El caso de las redes complejas requiere un coste computacional mucho más elevado, se necesitan realizar varios experimentos para reducir la incertidumbre probabilística y el parámetro N es determinante, las gráficas obtenidas en esta sección que han utilizado 1000 iteraciones han tardado unos 30 segundos.

Por otro lado, los modelos en ecuaciones diferenciales son menos flexibles debido a su enfoque global, en cambio, como se observará en las siguiente secciones,

los modelos en redes complejas permiten hacer experimentos mucho más variados y obtener conclusiones más profundas, realizando cambios para distintas estrategias e incluso cambiando la estructura de la red en el tiempo de ejecución.

6.2.2. Daño percibido en la red.

Daño constante.

Uno de los parámetros que se pueden estudiar en el modelo diseñado es el impacto del virus en la red. Para ello se sigue el procedimiento introducido en [1], a través del estudio del parámetro D/N.

$$D/N = \sum_{i \in \mathcal{V}} \frac{d_i}{N},$$

donde d_i se define como el daño máximo percibido por un nodo i de la red a lo largo de un experimento.

El parámetro D/N se define como la suma normalizada de los daños individuales en cada nodo. Es decir se obtiene un promedio del daño sufrido en la red, en caso de que el parámetro d sea constante, todos los individuos infectados sufrirán el mismo daño, en caso contrario el daño en diferentes partes de la red varia y el daño de cada uno no será el mismo. Este parámetro nos permite obtener una estimación del daño global percibido por el sistema atacado.

El objetivo será observar como la variación del parámetro d afecta al daño percibido por la red. Para ello se hacen simulaciones en las mismas redes que la sección 4.2 y parámetros explicados en la Tabla 6.2, exceptuando el valor de $\langle k \rangle$ que se compararán diversos valores.

En el primer conjunto de experimentos se asume que el daño individual producido por el malware es constante ($d \in [0, 1]$). Se puede predecir el comportamiento de D/N cuando $d \leq \theta$, ya que en ese caso la presencia del malware no es detectada al no llegar al umbral definido por θ, y por tanto el virus se propagará sin limitaciones y todos los nodos recibirán el mismo daño d, por tanto:

$$D/N = \sum_{i \in \mathcal{V}} \frac{d_i}{N} = \frac{N \times d}{N} = d.$$

El caso $d > \theta$ sigue un comportamiento que no es predecible, debido a que comenzará a aparecer un conjunto de individuos conscientes del malware y otros que no llegarán a infectarse.

Se comprueba el comportamiento en el conjunto de redes definidos con distintos valores de $\langle k \rangle$ y θ y se observa su evolución. Para ello se realiza una partición

del dominio ($d \in [0, 1]$) en 60 secciones y se realizan 500 experimentos en cada uno de los puntos de la partición obteniendo el valor D/N correspondiente.

Se puede observar que se ha reducido el número de experimentos por cada valor de d respecto a la Sección 6.2. Esto se debe a limitaciones en el tiempo de ejecución ya que por cada gráfica presentada se hacen $60 \times 500 = 30000$ experimentos, aumentando considerablemente la duración para obtener cada una de las gráficas presentadas.

La notación utilizada para el conjunto de experimentos que se va a presentar es ligeramente distinta, en vez de especificar todos los parámetros se resume indicando las iniciales del modelo de la red y entre paréntesis el grado medio de la misma. En caso de las redes Watts-Strogatz el parámetro p es determinante en la topología y no modifica el grado medio, por lo que también se indica como segundo elemento en la leyenda de las figuras.

Podemos observar la comparativa para los valores de $\theta \in \{0.2, 0.4, 0.6\}$ en las figuras 6.7, 6.8 y 6.9 respectivamente.

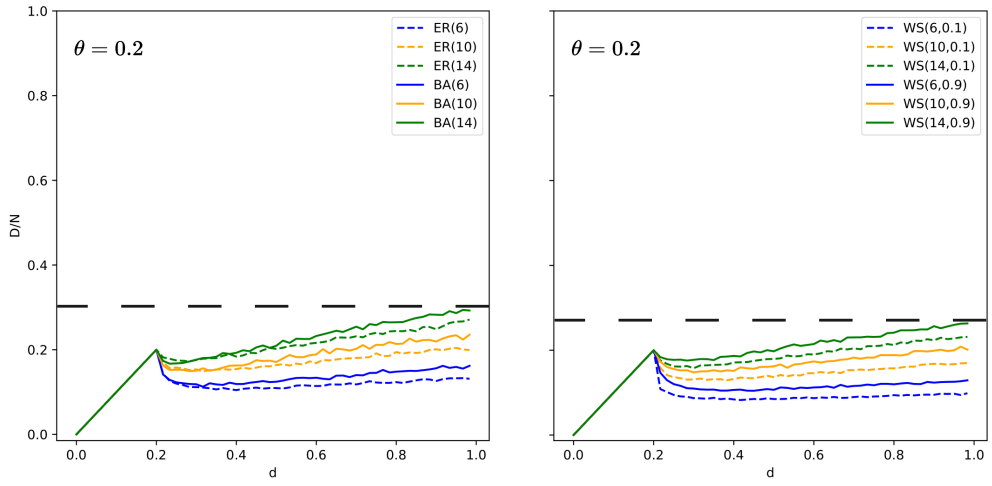

Figura 6.7: Evolución del parámetro D/N con $\theta = 0.2$ y parámetro $d \in [0, 1]$

En las figuras 6.7, 6.8 y 6.9 se pueden observar como aparecen los dos comportamientos anteriormente mencionados, uno preliminar y predecible con $d \leq \theta$ donde se sigue un crecimiento lineal y una segunda fase con $d > \theta$ donde se observa una primera caída ya que el daño aplicable es muy bajo dado que es muy cercano al límite θ y un posterior crecimiento con una tendencia clara.

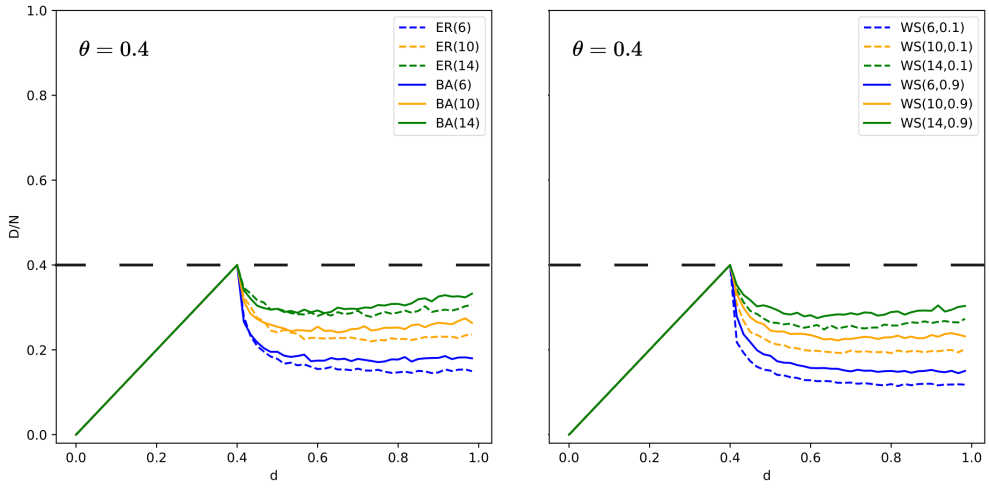

Figura 6.8: Evolución del parámetro D/N con $\theta = 0.4$ y parámetro $d \in [0, 1]$

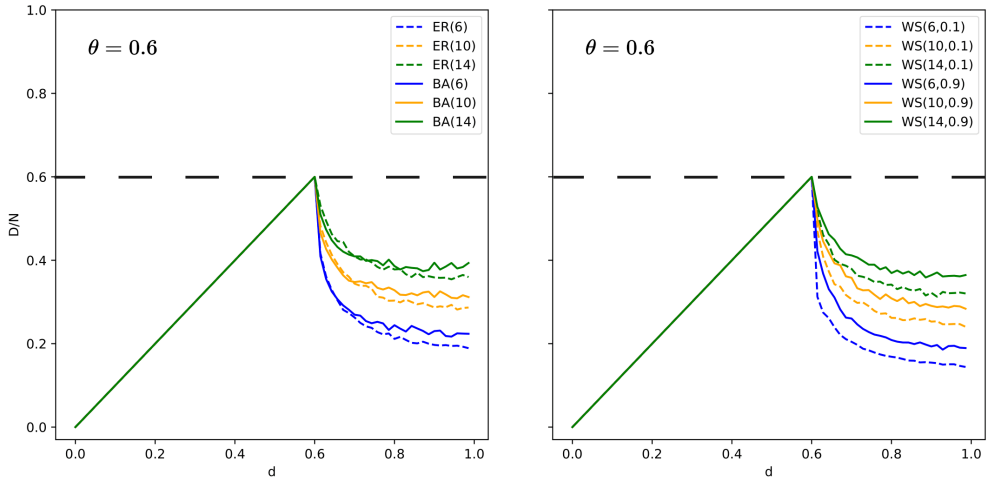

Figura 6.9: Evolución del parámetro D/N con $\theta = 0.6$ y parámetro $d \in [0, 1]$

Se observa además como aquellas redes en las que se identifican nodos con un grado mayor, anteriormente definidos como hubs, son propensas a recibir un mayor impacto negativo del malware. Esto se observa en como el daño sufrido en las redes Barabasi-Albert es mucho más alto y bastante más bajo en las redes Erdös-Rényi, donde la desviación entre los grados de los nodos es mucho más baja.

Por último es interesante comparar los valores máximos de D/N en los diferentes experimentos, ya que nos indican las estrategias a seguir por un posible

atacante. Se puede observar que en redes de alta seguridad con valores de θ bajos se maximiza el daño promedio cuando $d = 1$, es decir, desarrollando un malware que sea altamente dañino para los individuos. En cambio, cuando se tienen redes de baja seguridad o tolerantes a fallos, cuando θ es alto, el máximo de D/N se obtiene cuando $d = \theta$, es decir, la estrategia del atacante pasa por desarrollar un malware dañino pero sin sobrepasar la tolerancia de fallos de la red, de forma que este pueda camuflarse.

Daño creciente.

Aunque el conjunto de experimentos con d constante permite un análisis profundo del daño percibido por la red, existe una gran cantidad de malware que no puede modelarse a través de un daño constante. Gran parte del software malicioso desarrollado en la actualidad ataca al sistema durante todo el tiempo que persiste en la red, por lo que se vuelve más perjudicial cuanto más tiempo persiste en ella, por lo que es interesante simular el comportamiento en la red definiendo el parámetro d como una función respecto al tiempo.

$$d(t) = \frac{d_0 e^{\epsilon t}}{1 + d_0(\epsilon t - 1)}.$$

La función utilizada es la función logística con tasa de crecimiento ϵ y población inicial d_0. Esta función tiene un comportamiento exponencial ($d(t) \cong d_0 e^{\epsilon t}$) en valores bajos de t y tiende al valor máximo 1 cuando t comienza a crecer ($t \to \infty$).

Se va a comparar el daño sufrido por la red (D/N) respecto a la tasa de crecimiento ϵ, lo que nos permite distinguir aquellos valores más perjudiciales respecto a la red estudiada. El conjunto de experimentos a realizar es similar a los de la sección anterior.

Se modela el comportamiento en el conjunto de redes definidas con distintos valores de $\langle k \rangle$ y θ para observar su evolución. Se realiza una partición del dominio de estudio ($\epsilon \in [0, 1]$) en 60 secciones y se realizan 500 experimentos en cada uno de los valores obtenidos. Cabe destacar que a diferencia del caso d constante, la partición que se va a realizar de los valores de ϵ no sigue un comportamiento uniforme, sino logarítmico, siendo mucho más denso en valores cercanos a 0.

La razón de esta decisión es que se ha comprobado experimentalmente que cerca de ese rango de valores los valores de D/N varían más y es necesaria más información sobre este suceso, mientras que, como se observará en los experimentos, posteriormente sigue una tendencia mucho más clara.

Podemos observar los resultados y realizar una comparativa entre los tipos de redes en las figuras 6.10, 6.11 y 6.12.

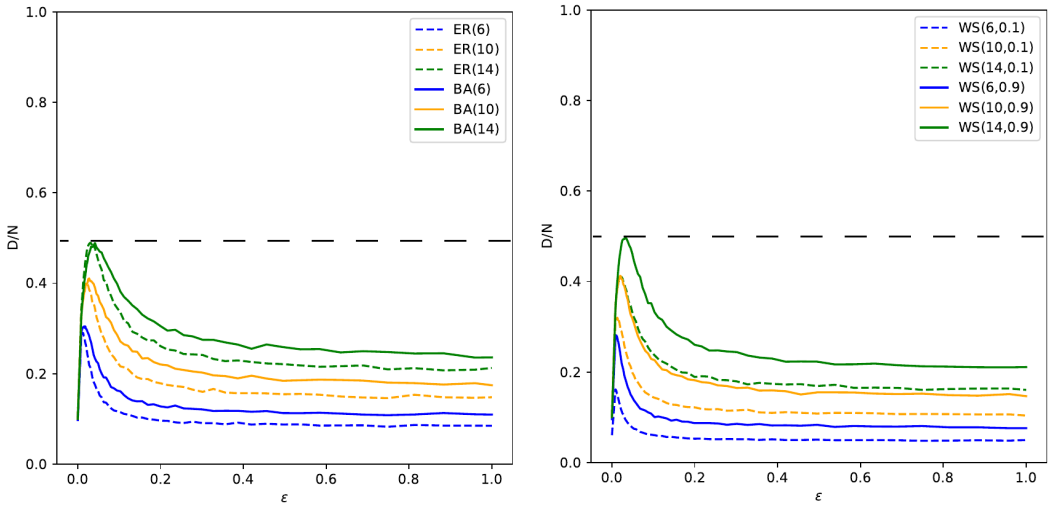

Figura 6.10: Evolución del parámetro D/N con $\theta = 0.2$ y daño variante en el tiempo.

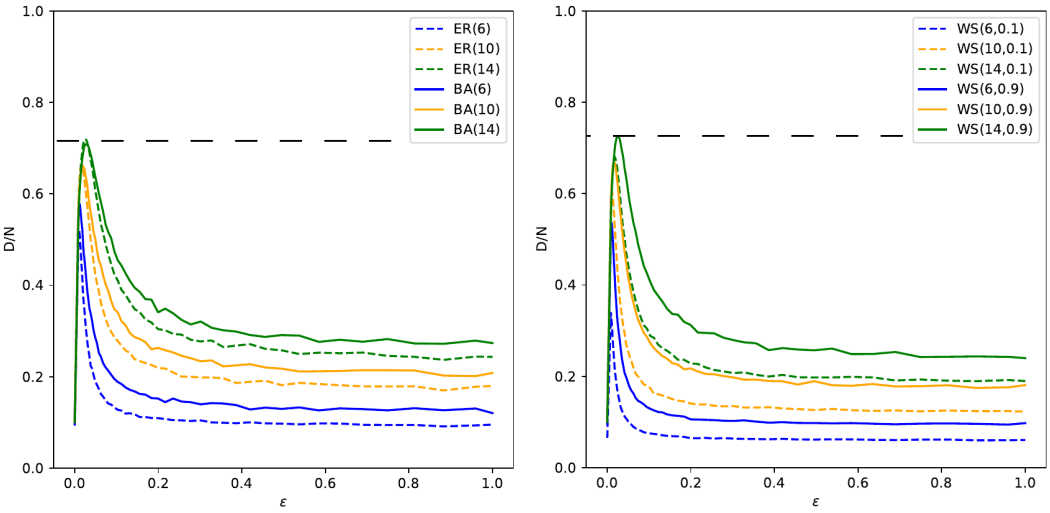

Figura 6.11: Evolución del parámetro D/N con $\theta = 0.4$ y daño variante en el tiempo.

En todos los experimentos se pueden observar tendencias claras, distinguiendo en todas las gráficas 3 fases diferenciadas.

Una primera fase con valores de ϵ muy cercanos a 0, donde aumentar el valor de la tasa de crecimiento de la función $d(t)$ implica un crecimiento del daño producido de manera muy pronunciada, esta fase finaliza de manera temprana encontrando un máximo, que en ningún caso supera el valor $\epsilon = 0.1$. Con este crecimiento el virus tiene la posibilidad de propagarse antes de ser detectado, de forma que el daño producido puede ser muy alto.

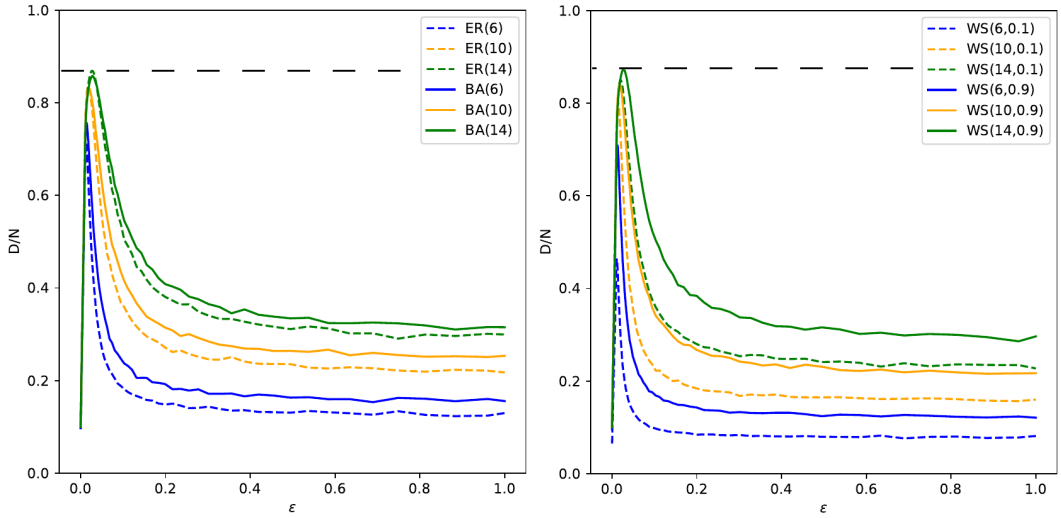

Figura 6.12: Evolución del parámetro D/N con $\theta = 0.6$ y daño variante en el tiempo.

La segunda fase tiene un comportamiento diametralmente contrario al anterior, donde el aumento de ϵ implica una caída muy pronunciada del daño producido a la red, debido principalmente al hecho de que un crecimiento demasiado rápido implica que el virus sea rápidamente detectado.

Todos los experimentos finalizan en una tercera fase donde el parámetro D/N se estabiliza respecto a los valores de ϵ, esto debido a que el crecimiento se produce tan rápido que el parámetro d tiende a comportarse rápidamente como la función constante $d = 1$, por tanto se identifica una conducta de D/N similar a los experimentos constantes.

Respecto a la comparativa entre redes se obtienen conclusiones parecidas a las que se concluyen de los experimentos constantes. Se observa que aquellas redes con topologías donde existen nodos hubs sufren más que el resto. En cambio, las redes más simétricas sufren menos la presencia del virus, observándose principalmente en los experimentos realizados con redes Watts-Strogatz con $p = 0.1$, donde se puede ver que los valores con grado medio 14, son similares a los obtenidos con redes de grado medio 10.

Por otro lado, comparando los virus de daño constante y daño creciente, se observa como la estrategia de que el parámetro d sea variante produce un mayor daño general en las redes, obteniendo valores máximos mayores a los constantes. La razón de este suceso se debe a que el comportamiento del virus en la red distingue dos fases. En la primera se le permite propagarse sin ser detectado gracias a que d permanece suficientemente bajo. En la segunda el virus es detectado, pero el crecimiento exponencial le permite causar un daño elevado en un periodo de tiempo realmente corto.

Esta idea de que como la variación del comportamiento del virus produce distintos daños a la red abre una vía de investigación interesante. Ya ha sido mencionado que la diferencia principal entre las epidemias biológicas y las epidemias digitales residen en un concepto cercano a la teoría de juegos, donde el atacante puede adoptar estrategias para aumentar su efecto en la red. Se van a estudiar algunas de estas estrategias de ataque que se puede implementar gracias al uso del modelo de redes complejas y cómo afectan al uso de distinta topologías en la red.

6.2.3. Estrategias de Ataque.

En esta sección se estudian distintas estrategias de ataque a las redes, donde el atacante va a buscar maximizar el daño que pueda producir en las mismas. Las estrategias que se van a estudiar siguen dos corrientes.

En primer lugar, se ha observado como la variabilidad del daño produce distintos valores del parámetro D/N, por lo que la primera estrategia pasa por cambiar la programación del virus y que el daño del mismo esté controlado y reaccione a la propagación del mismo.

La segunda estrategia que se va a estudiar tiene que ver con la topología de la red, introduciendo como semilla del virus en la red la infección de aquellos nodos con mayor centralidad, siguiendo los parámetros estudiados en las Sección 4.3.

Virus reactivos.

En esta estrategia se sigue una idea similar a la obtenida en la Sección 6.2.2. En este caso la función de daño no es global, sino que será explícita para cada uno de los nodos, modificando así la función que rige el parámetro d a la siguiente:

$$d(t_i) = \frac{d_0 e^{\epsilon t_i}}{1 + d_0(\epsilon t_i - 1)}.$$

Esta vez el parámetro t_i depende del histórico de cada cepa del virus, causando un daño distinto dependiendo del nodo de la red, por tanto, la variable t_i hará referencia a las veces que la cepa se ha propagado hasta infectar el nodo actual. Este cambio produce una clara variación en la medición del parámetro μ, que será determinado a través del daño base causado en el momento que el nodo es infectado.

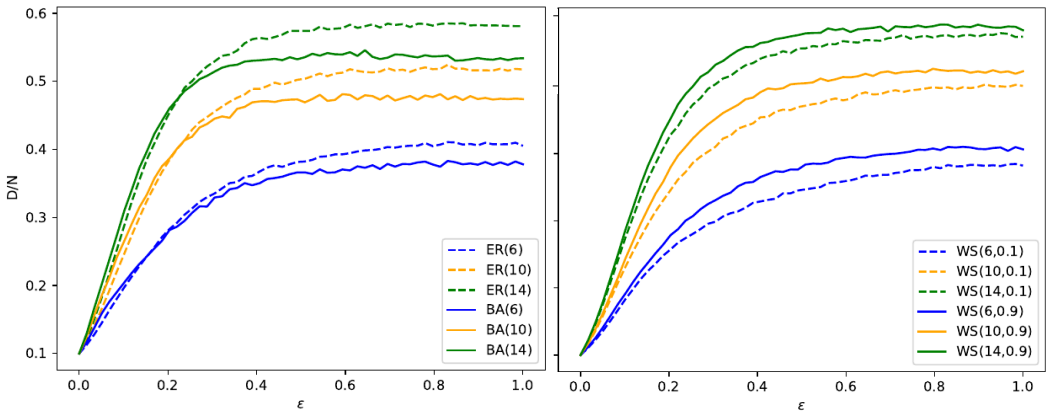

Figura 6.13: Comparación del daño producido por un virus con daño reactivo a su propagación con $\theta = 0.2$.

En la Figura 6.13 podemos observar el resultado de este experimento para redes de alta seguridad ($\theta = 0.2$). En este caso se puede observar como el valor de D/N sigue dos fases claras. Cuando $\epsilon < 0.4$ se sigue una tendencia creciente, que se inicia desde un valor inicial muy bajo en $0,1$. A partir de ese momento el valor se estabiliza habiendo alcanzado un valor máximo que se mantiene hasta el final, aunque se puede notar un leve descenso en los valores más altos de ϵ.

La ventaja principal de esta estrategia es que, dado un virus con una tasa de crecimiento alta, no es necesario identificar un valor de ϵ muy concreto para maximizar el daño, por lo que este es un planteamiento útil para un atacante que no tenga un claro control sobre el crecimiento del impacto producido.

Una de las conclusiones más importantes sobre este ataque es, y adelantando conclusiones del siguiente apartado, que es el único más efectivo sobre redes Erdös-Renyi que Barabsi-Albert. Es decir, gracias a la estrategia de propagación, consigue producir un daño más elevado sobre redes con una menor desviación sobre el grado de sus nodos. Sin embargo, como se observa en la imagen derecha con el modelo WS $p = 0.1$, una cierta variabilidad si que aporta mejores resultados, es decir la desviación favorece el ataque, pero sin ser muy alto.

Este hecho aporta una clara ventaja a los atacantes ya que no es necesario conocer la topología exacta de la red, solo obtener información sobre si tiene una estructura más centralizada o más horizontal, donde podrán elegir este ataque o algunos de los que ya se han observado, teniendo garantías de producir un daño considerable.

Obsérvese ahora el análisis de este ataque en redes de baja seguridad, donde se toma el valor $\theta = 0.6$. Las gráficas pueden observarse en la Figura 6.2.3.

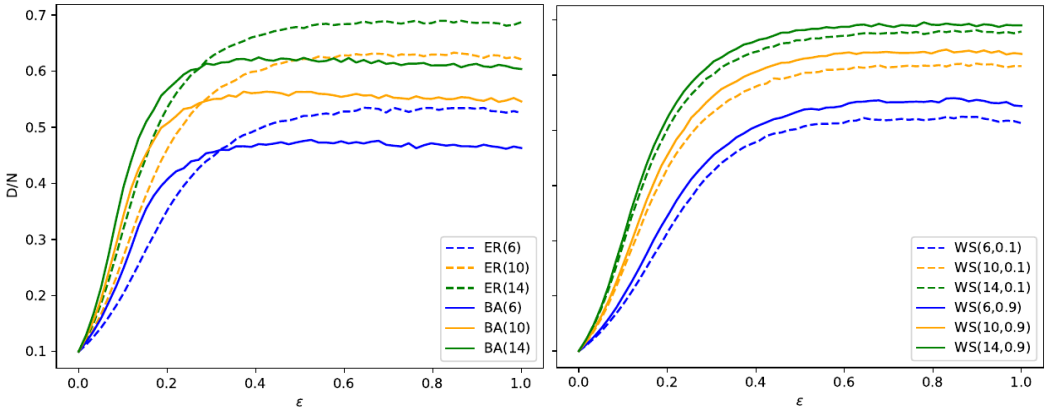

Figura 6.14: Comparación del daño producido por un virus con daño reactivo a su propagación con $\theta = 0.6$.

Se puede observar que se sigue una tendencia similar a los experimentos realizados en redes de alta seguridad. Se observan las dos fases de crecimiento y estabilidad mencionadas anteriormente y de nuevo se obtienen mayores valores de ataque en redes Erdös-Renyi. Cabe destacar, sin embargo, que los valores obtenidos en comparación a las redes de alta seguridad son relativamente cercanos, donde el máximo solo se diferencia en una subida de 0.6 a 0.7. Es decir, el parámetro θ no es tan determinante como en otros ataques ya revisados.

Como conclusión se observa que la estrategia de que el crecimiento del parámetro d reaccione a la transmisión de las cepas es una estrategia muy útil cuando el conocimiento de la red es bajo y el control sobre el daño que el virus puede causar también, produciendo daños relativamente elevados en los mejores casos.

Estrategias basadas en la topología de la red.

El impulso de la creación de esta sección surgió como consecuencia del descubrimiento del artículo [15] en el que se buscaba la obtención de los nodos críticos de una red a través de la resolución del $\alpha - Separation\ Problem$. El mecanismo utilizado para la identificación de nodos críticos está modificado debido a limitaciones computacionales. El tiempo de ejecución aumentaba considerablemente y se ha intentado mantener la misma población inicial que en los experimentos anteriores para poder realizar comparaciones fidedignas.

Como se ha mencionado en la Sección 4.3, la propia topología de la redes implica que no todos los nodos adquieran la misma importancia, los nodos que tenían más importancia se identificaban como centrales. La estrategia del atacante pasará por cambiar la semilla de la infección, comenzado el ataque por el conjunto de nodos que mayor valor de centralidad poseen.

Primeramente, se puede observar en la Figura 6.15 la comparativa del transcurso de una ciber-epidemia con un ataque donde la semilla inicial son los nodos centrales y cuando esa semilla se escoge de manera aleatoria. Para este experimento se han utilizado como medida la centralidad por cercanía o *Closeness Centrality* y se utilizan los parámetros indicados en la Tabla 6.3, véase que d sigue la estrategia de crecimiento respecto al tiempo.

Símbolo	N	θ	d	ϵ	τ	ν	μ_0	γ	ρ_0	$\langle k \rangle$
Rango	$[1, \infty]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[1, N]$
Valor	1000	0.2	$d(t)$	0.01	0.0055	0.011	0.011	0.03	0.01	10

Tabla 6.3: Conjunto de Valores para experimentos del modelo SIH-UA en redes complejas

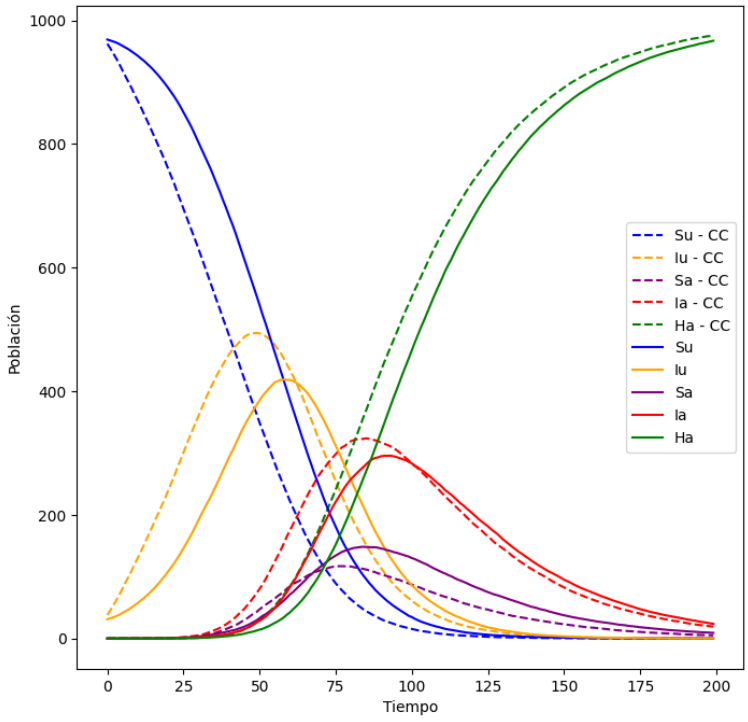

Figura 6.15: Comparación de la evolución de una ciberepidemia con ataque a los nodos centrales en una red Barabasi-Albert $N = 1000$ y $\langle k \rangle = 10$. Se distingue en linea discontinua el experimento con centralidad y en linea continua el experimento aleatorio.

Como se puede observar el crecimiento de infectados es más rápido, perdiendo las parte mas suaves del crecimiento en forma de campana. Este hecho implica que el compartimento de Susceptibles Conscientes crezca mucho menos, ya que el tamaño de Infectados de ambos tipos crece, implicando que queden menos susceptibles capaces de ser consciente del virus sin infectarse del mismo.

Con este experimento preliminar podemos concluir que existe un efecto real en las redes cuando se ataca la centralidad de los nodos, por lo que se va a realizar una experimentación sobre el daño que puede causar esta estrategia dentro de la red en las tres centralidades estudiadas.

El ataque consiste por cambiar la semilla inicial de la infección, siendo los primeros nodos infectados aquellos con mayor centralidad en vez de escogerse de manera aleatoria.

El primer experimento que se va a mostrar simula d con valores constantes sobre redes de alta seguridad, es decir, $\theta = 0.2$. Con las iniciales en ingles explicadas en la Sección 4.3 se indica el tipo de centralidad utilizada en cada caso. En la Figura 6.16 se puede observar los resultados para las distintas medidas de centralidad en redes Erdös-Renyi y Barabasi-Albert. En la Figura 6.17 se puede observar los resultados para las distintas medidas de centralidad en redes Watts-Strogatz.

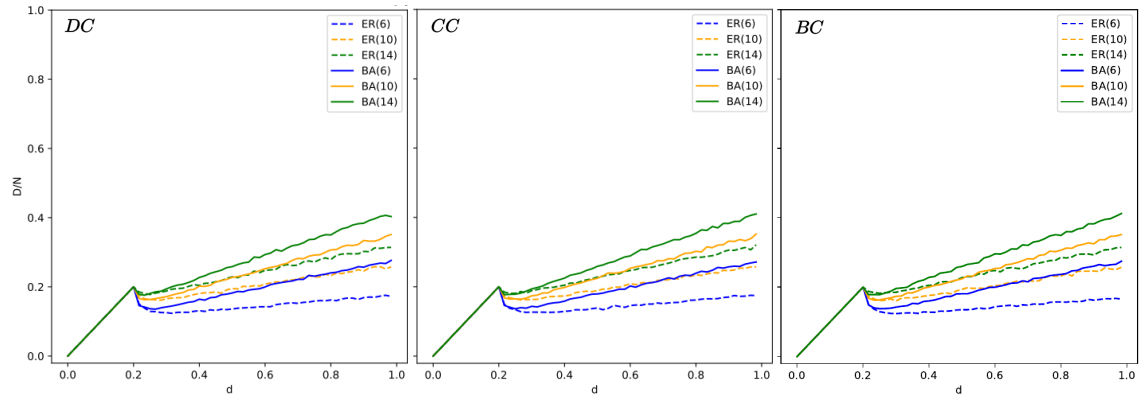

Figura 6.16: Comparación del daño en una ciberepidemia con $\theta = 0.2$ y daño constante atacando los nodos centrales en redes BA y ER.

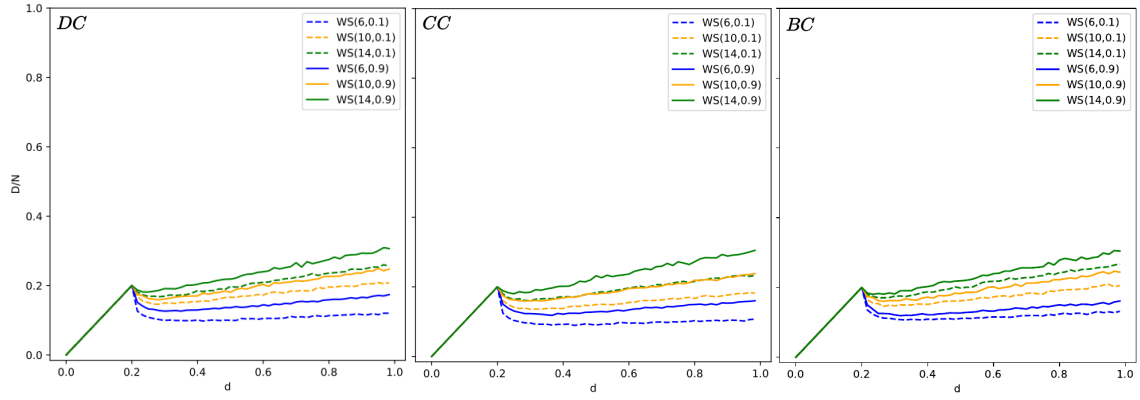

Figura 6.17: Comparación del daño en una ciberepidemia con $\theta = 0.2$ y daño constante atacando los nodos centrales en redes WS.

Estas gráficas son comparables con las de la Figura 6.7 donde se puede observar el mismo experimento con una semilla inicial aleatoria. La primera conclusión es clara, el daño que se produce con esta estrategia es mayor, alcanzando un máximo que supera el valor 0.4, cuando sin la misma apenas se alcanzaba el 0.3.

Por otro lado existen otras conclusiones interesantes que requieren una comparativa más fina. En primer lugar se puede observar como la estrategia no afecta de la misma manera a todos los tipos de redes, siendo más castigadas aquellas con centralidades máximas más altas, como las redes de Barabasi-Albert y apenas existe diferencia en las redes de Watts-Strogatz de grado inferior.

También podemos observar que existen grandes similitudes a la hora de usar distintas centralidades, no identificando patrones muy diferenciadores en ellas, por lo que la estrategia del atacante pasa por elegir aquella más difícil de identificar y explotar, normalmente la centralidad por grado.

Anticipando lo que ocurre en los experimentos que vamos a presentar, el hecho de que no existe una gran variabilidad en las 3 centralidades presentadas, no es único en el experimento anterior, por lo que por claridad en el trabajo se van a presentar los siguientes experimentos únicamente con la centralidad por grado. Las gráficas para la centralidad por cercanía e intermediación se pueden consultar en el Apéndice A, pero las conclusiones obtenidas son similares.

Una comparativa interesante pasa por ver que ocurre cuando las redes atacadas son de baja seguridad, es decir, $\theta = 0.6$. Este experimento puedes observarse en todos los tipos de redes con un ataque a la centralidad por grado en la Figura 6.18.

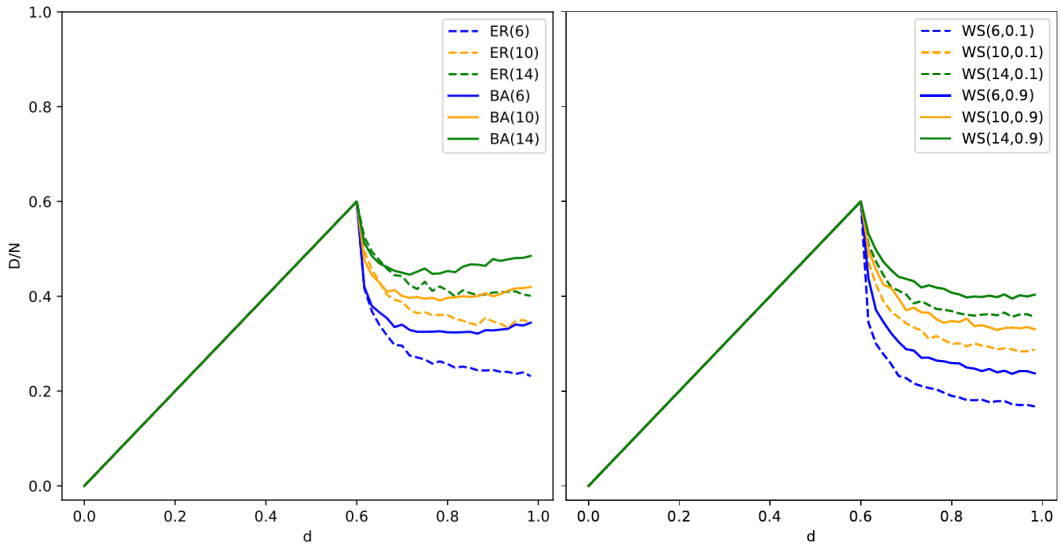

Figura 6.18: Comparación del daño en una ciberepidemia con $\theta = 0.6$ y daño constante atacando los nodos centrales por grado.

Se puede observar de nuevo que las redes Barabasi-Albert son las más condicionadas por este ataque, donde se observa un crecimiento pronunciado para altos valores de d, a diferencia de la tendencia menos creciente que observamos en el resto de redes y en la implantación de la semilla aleatoria de la Figura 6.9.

Sin embargo, podemos observar que en todos los experimentos el daño máximo es el mismo, obteniéndose con $d = 0.6$. Es decir, en busca de producir el daño más alto posible, no es rentable el coste de descubrir la topología de la red , sino que la estrategia más eficiente es la misma que con una semilla aleatoria, controlar el daño del virus para que no sea identificado debido al parámetro θ.

Una vez se han realizado los experimentos con daño constante, siguiendo el mismo esquema que el trabajo inicial, podemos observar los resultados de estas estrategias en daños crecientes en el tiempo. En la figura 6.19 podemos observar como evoluciona el daño con un crecimiento creciente del parámetro d en redes de alta seguridad ($\theta = 0.2$).

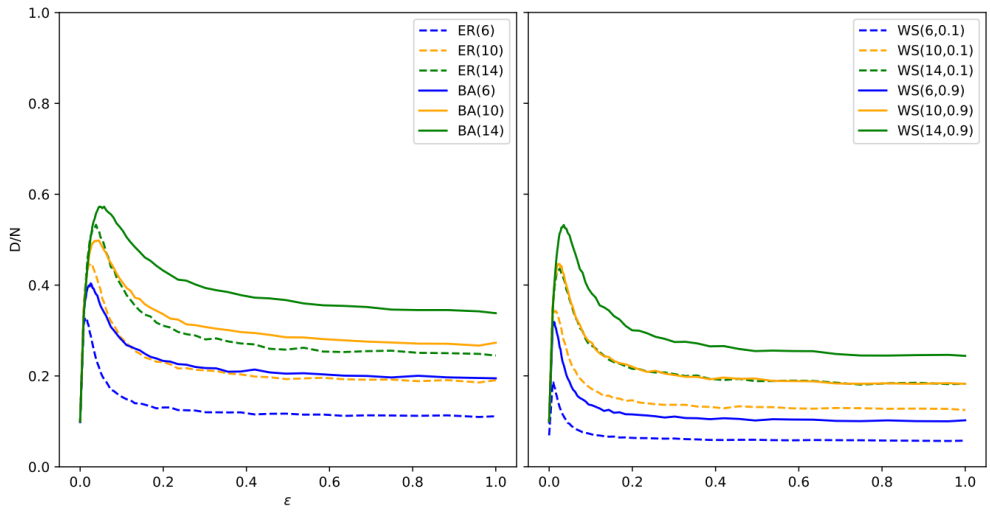

Figura 6.19: Comparación del daño en una ciberepidemia con $\theta = 0.2$ y daño creciente atacando los nodos centrales por grado.

Podemos observar de nuevo que la red más afectada es la del modelo de Barabasi-Albert, aunque esta vez el daño máximo crece para todos los tipos de redes, comparándolos con las semillas aleatorias observables en la Figura 6.10. Además se puede observar que gracias al crecimiento medio del daño, la fase más alta del parámetro D/N se ensancha, existiendo un tramo mayor de valores donde el daño del virus es relativamente alto.

Por último gracias al ataque de los nodos centrales se observa como en las redes más afectadas el decrecimiento hacia la fase de estabilidad es menos acentuado, por lo que establecer el valor de la tasa de crecimiento de la función logística para obtener un daño elevado es más sencillo.

En el Apéndice A se puede observar estos mismos experimentos en escenarios de redes de baja seguridad, donde establecemos el parámetro $\theta = 0.6$.

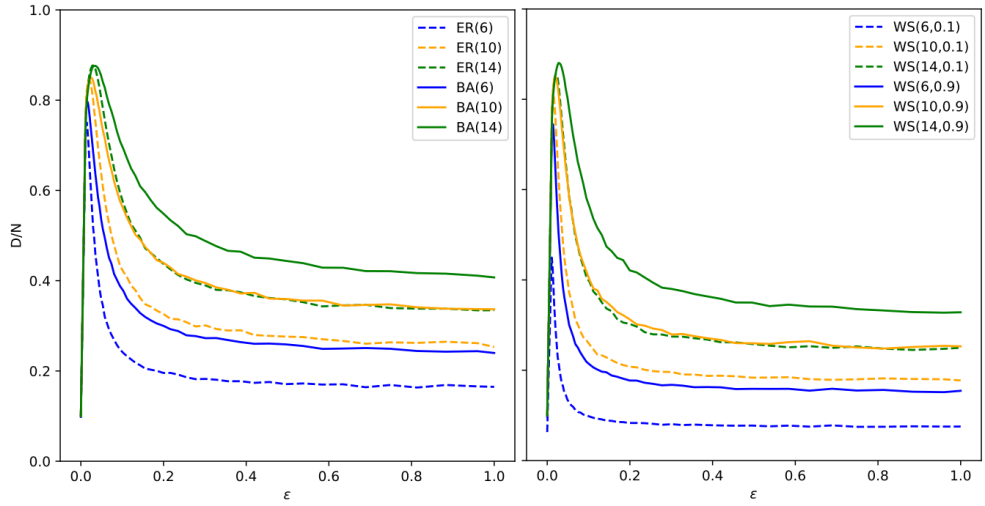

Figura 6.20: Comparación del daño en una ciberepidemia con $\theta = 0.6$ y daño creciente atacando los nodos centrales por grado.

Se puede observar como el hecho de utilizar un daño creciente, unida a la estrategia de ataque hace que la mayoría de redes alcancen un máximo de D/N muy alto y parejo, entre 0.8 y 0.9, significando que la mayoría de nodos de la red han sido atacados de forma muy virulenta. Sin embargo, estos valores ya se alcanzaban en las redes con mayor grado medio ($\langle k \rangle = 14$), por lo que realmente esta estrategía no es diferencial para esos casos.

En el extremo contrario observamos como la red $WS(6, 0.1)$ consigue resistir de manera idónea el ataque gracias a su simetría y al valor inferior del grado medio.

En las redes Barabasi-Albert se vuelve a notar el fenómeno de suavizado del máximo ayudando a que distintos valores de ϵ alcances un alto valor de D/N, no ocurre lo mismo con los otros tipos de redes estudiados. Sin embargo, si se nota como el daño promedio sube en todas ellas, en las fases que ϵ toma valores más altos.

Como conclusión general podemos observar que la estrategía de ataque a los nodos centrales es buena sobretodo en redes con alta variación en los grados de adyacencia de los nodos y con una alta seguridad, es decir, valores del parámetro θ bajos. En otras redes esta estrategia no produce un impacto negativo, pero entendiéndose que el descubrimiento de la topología de la red supone un esfuerzo para el atacante, utilizar la estrategia mencionada no merece la pena.

7

Conclusiones y trabajos futuros

Se puede afirmar que los objetivos mencionados en la Sección 2 han sido cumplidos satisfactoriamente.

El grueso del trabajo ha pasado por la implementación en ecuaciones diferenciales y en redes complejas de los dos modelos estudiados en el trabajo: el modelos SIR y el modelo presentado en [1].

Tras este estudio e implementación de ambos esquemas se concluye que, de manera general, los modelos compartimentales conforman una gran herramienta para la simulación de fenómenos, en este caso pandemias biológicas y digitales. Esta perspectiva del uso de modelos como herramientas se refuerza con la última sección del trabajo donde observamos las diferencias de experimentos en estrategias distintas, pudiendo obtener con un único modelo conclusiones de situaciones realmente variadas.

Se observa como la topología de las redes con existencia de hubs son más propensas a recibir un daño mucho mayor, específicamente las redes Barabasi-Albert, mientras que otros modelos no sufren tanto estos ataques.

Respecto a las estrategias de ataque presentadas no se observa una más dominante al resto sino que se deben utilizar respecto a las características específicas del sistema.

En caso de no tener una topología muy centralizada o tener menos información, se puede adquirir la estrategia de los virus reactivos a la infección. En cambio, si se puede conocer la forma de la red, atacar a los nodos centrales es

una buena vía, utilizando aquella centralidad de la que más información se tenga respecto al sistema en peligro.

Desde esta perspectiva el trabajo futuro tras la realización de este TFG es la explotación de estos modelos en experimentos más complejos y útiles. Gracias a la modelización de redes complejas se pueden realizar otros experimentos como por ejemplo: separación de poblaciones, estrategias para la defensa de la red o modificaciones de la red en tiempo real. Se obtendrían nuevas conclusiones útiles para la confrontación del reto de la propagación de malware en escenarios reales.

Uno de los problemas que tienen estos experimentos es la comparación de datos reales. La información sobre casos de infección de virus en redes específicas o no existe o es confidencial, la liberación de estos datos traería una perspectiva extra al uso de esto modelos pudiendo tomar una comparativa con casos reales.

Otros trabajos futuros consisten en el uso de las implementaciones utilizadas (redes complejas y ecuaciones diferenciales) en otros problemas, no solo de carácter epidemiológico, sino también en fenómenos con estructuras conocidas que puedan comportarse como redes: redes sociales, redes lingüísticas, internet o redes de trabajo, donde mucha de las técnicas y conceptos estudiados tienen una aplicación análoga.

Bibliografía

[1] D. Aleja, G. Contreras-Aso, K. Alfaro-Bittner, E. Primo, R. Criado, M. Romance, and S. Boccaletti. A compartmental model for cyber-epidemics. *Chaos, Solitons & Fractals*, 161:112310, 2022.

[2] O. N. Bjørnstad, K. Shea, M. Krzywinski, and N. Altman. The seirs model for infectious disease dynamics. *Nature methods*, 17(6):557–559, 2020.

[3] Y. W. Chen, L. F. Zhang, and J. P. Huang. The watts–strogatz network model developed by including degree distribution: theory and computer simulation. *Journal of Physics A: Mathematical and Theoretical*, 40(29):8237, 2007.

[4] A. M. del Rey. Mathematical modeling of the propagation of malware: a review. *Security and Communication Networks*, 8(15):2561–2579, 2015.

[5] A. M. del Rey, G. Hernández, A. B. Tabernero, and A. Q. Dios. Advanced malware propagation on random complex networks. *Neurocomputing*, 423:689–696, 2021.

[6] S. O. Fatunla. *Numerical methods for initial value problems in ordinary differential equations*. Academic press, 2014.

[7] D. J. Gerberry and F. A. Milner. An seiqr model for childhood diseases. *Journal of Mathematical Biology*, 59:535–561, 2009.

[8] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[9] E. Kypriotelis, G. Kolias, and P. Pappa. The growth of global risks after the covid-19 pandemic. *KnE Social Sciences*, pages 30–44, 2023.

[10] V. Latora, V. Nicosia, and G. Russo. *Complex networks: principles, methods and applications*. Cambridge University Press, 2017.

[11] M. Y. Li and J. S. Muldowney. Global stability for the seir model in epidemiology. *Mathematical biosciences*, 125(2):155–164, 1995.

[12] A. Martín del Rey, A. Queiruga Dios, G. Hernández, and A. Bustos Tabernero. Modeling the spread of malware on complex networks. In *Distributed Computing and Artificial Intelligence, 16th International Conference, Special Sessions*, pages 109–116. Springer, 2020.

[13] R. K. Merton. The matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810):56–63, 1968.

[14] J. D. Murray. *Mathematical biology: I. An introduction*. Springer, 2002.

[15] S. Pérez-Peló, J. Sánchez-Oro, and A. Duarte. Finding weaknesses in networks using greedy randomized adaptive search procedure and path relinking. *Expert Systems*, 37(6):e12540, 2020.

[16] C. Seshadhri, T. G. Kolda, and A. Pinar. Community structure and scale-free collections of erdős-rényi graphs. *Physical Review E*, 85(5):056109, 2012.

[17] H. H. Weiss. The sir model and the foundations of public health. *Materials matematics*, pages 0001–17, 2013.

[18] D. B. West et al. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, 2001.

[19] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences*, 99(21):13382–13386, 2002.

[20] Q. Zhu, X. Yang, and J. Ren. Modeling and analysis of the spread of computer virus. *Communications in Nonlinear Science and Numerical Simulation*, 17(12):5117–5124, 2012.

Apéndices

A

Experimentos de centralidad

En esta sección se muestran las gráficas de los experimentos que no se han presentado en la sección final del trabajo.

A.1. Centralidad por cercanía *Closeness Centrality*

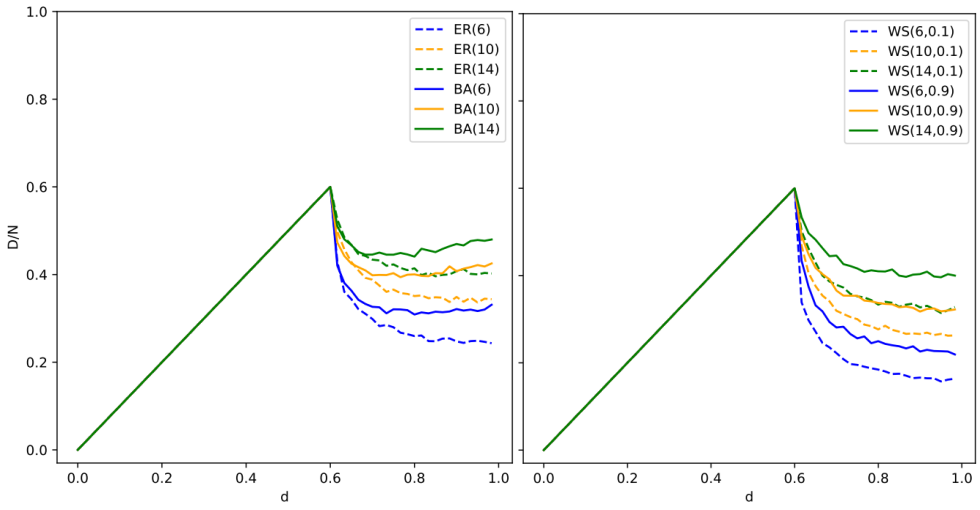

Figura A.1: Comparación del daño en una ciberepidemia con $\theta = 0.6$ y daño constante atacando los nodos centrales por cercanía.

.

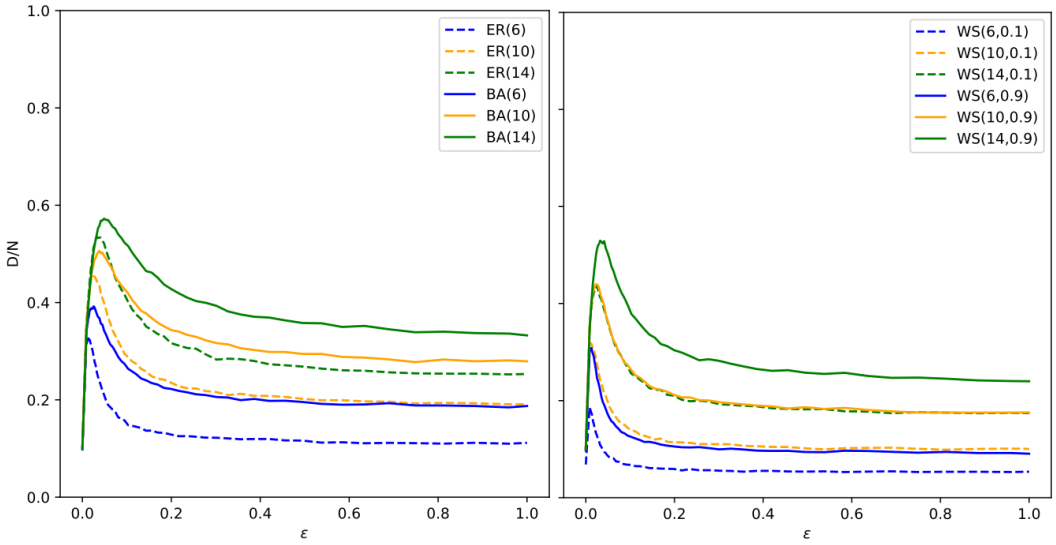

Figura A.2: Comparación del daño en una ciberepidemia con $\theta = 0.2$ y daño creciente atacando los nodos centrales por cercanía.

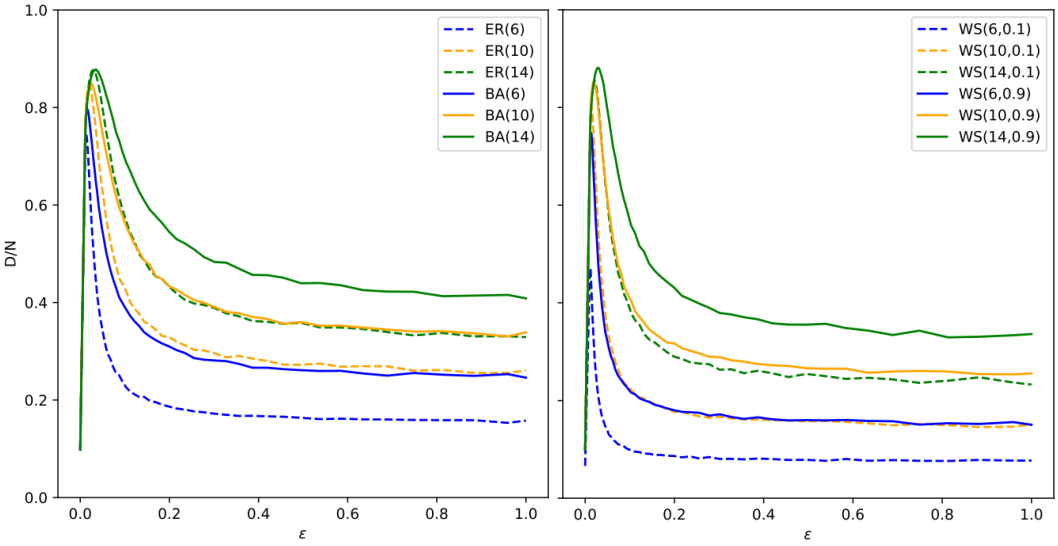

Figura A.3: Comparación del daño en una ciberepidemia con $\theta = 0.6$ y daño creciente atacando los nodos centrales por cercanía.

.

A.2. Centralidad por intermediación. *Betweeness Centrality*

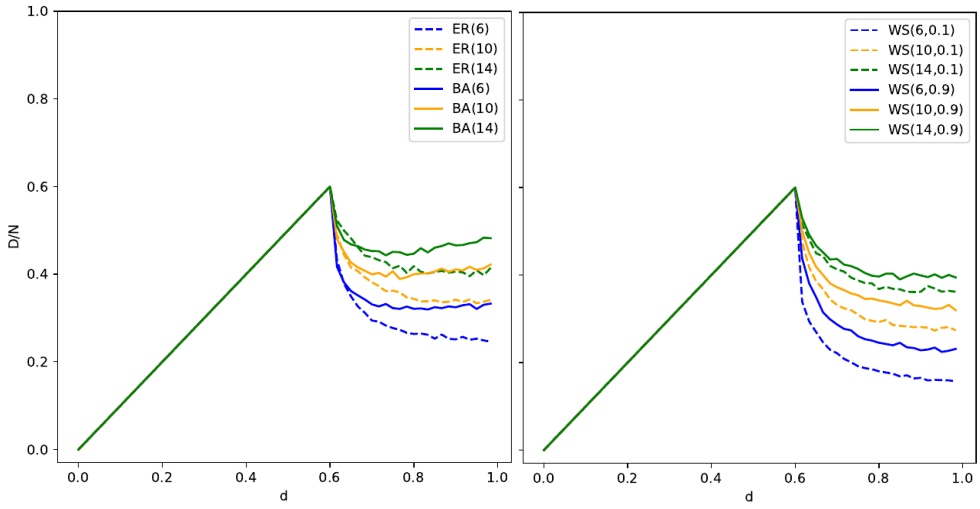

Figura A.4: Comparación del daño en una ciberepidemia con $\theta = 0.6$ y daño constante atacando los nodos centrales por intermediación.

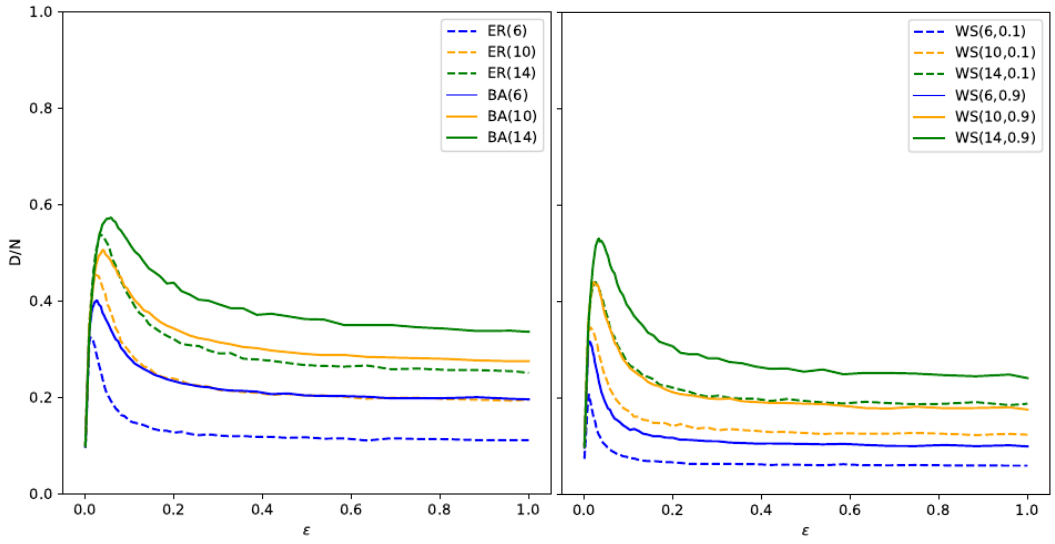

Figura A.5: Comparación del daño en una ciberepidemia con $\theta = 0.2$ y daño creciente atacando los nodos centrales por intermediación.

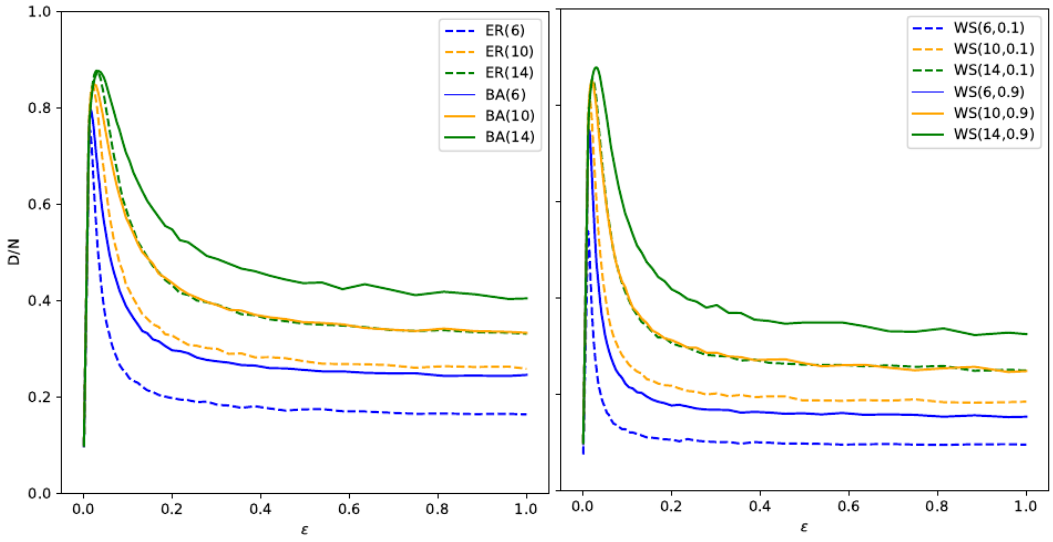

Figura A.6: Comparación del daño en una ciberepidemia con $\theta = 0.6$ y daño creciente atacando los nodos centrales por intermediación.