

Universidad Nacional de Educación a Distancia

Máster en Ingeniería de Sistemas y Control



TRABAJO FIN DE MÁSTER

Estudio de la dinámica de transmisión del
meningococo C mediante redes aleatorias

Autor: Víctor Sánchez Alonso

Directora: Matilde Santos Peñas

Codirector: Javier Villanueva Oller

“Una vez hecho algo, no puede valer mucho; es una obra humana con todas las imperfecciones de lo humano, pero el hecho de ejecutarla sí es interesante.”

Jorge Luís Borges

RESUMEN

El objetivo principal de este trabajo es el estudio de cómo se propaga la meningitis. Se sabe muy poco acerca de la dinámica de transmisión de la enfermedad causada por la bacteria del meningococo y mediante los modelos de redes aleatorias se realiza un estudio y modelado del patrón de transmisión en poblaciones de individuos susceptibles.

El modelo de transmisión se basa en los individuos portadores, que son aquellos individuos que han sido infectados por la bacteria del meningococo. Los datos de los que se dispone apuntan a que el comportamiento de ésta enfermedad infectocontagiosa presenta un comportamiento cíclico.

El modelo creado trata de imitar este comportamiento mediante simulaciones en un entorno de computación distribuida.

PALABRAS CLAVE

Redes aleatorias, propagación, Meningococo, población, portador, infección, computación distribuida.

ABSTRACT

The main aim on this dissertation is to study how the meningitis is spread. Very little is known about the transmission dynamics of the disease caused by meningococcal bacteria and through random networks models a study and modeling of the transmission pattern in populations of susceptible individuals is carried out.

The transmission model is based on carriers, which are those individuals who have been infected with meningococcal bacteria. The data that are available suggest that the behavior of this infectious disease is cyclical behavior.

The model is created to imitate this behavior by simulations in a distributed computing environment.

KEYWORDS

Random networks, propagation, Meningococcal, population, carrier, infection, distributed computing.

ÍNDICE

1. INTRODUCCIÓN.....	1
1.1. MOTIVACIÓN	5
1.2. OBJETIVOS	6
1.2.1. <i>Objetivos detallados</i>	7
1.3. ESTRUCTURA DE LA MEMORIA	7
2. ESTADO DE LA CUESTIÓN	9
2.1. MODELOS	9
2.1.1. <i>Modelos deterministas</i>	12
2.1.2. <i>Modelos estocásticos</i>	14
2.1.3. <i>Modelos discretos</i>	15
2.1.4. <i>Modelos continuos</i>	19
2.1.5. <i>Modelos de redes</i>	20
2.1.5.1 Una breve introducción a la teoría de grafos	22
2.1.5.2 Redes regulares.....	25
2.1.5.3 Redes de mundo pequeño	27
2.1.5.4 Redes de escala libre	29
2.1.5.5 Redes aleatorias	31
2.1.5.5.1 Uso de redes aleatorias en modelos	34
2.2. MODELOS EPIDEMIOLÓGICOS	39
2.2.1. <i>Modelos epidemiológicos continuos</i>	41
2.2.1.1 Modelo SI.....	42
2.2.1.2 Modelo SIS.....	44
2.2.1.3 Modelo SIR	45
3. PROPUESTA DE MODELIZACIÓN MEDIANTE REDES ALEATORIAS.....	49
3.1. DISEÑO DE LA RED.....	51
3.1.1. <i>Asignación de edades</i>	52
3.1.2. <i>Evolución de la pirámide poblacional</i>	53
3.1.3. <i>Reparto de portadores</i>	53
3.1.4. <i>Asignación de relaciones</i>	54
3.1.5. <i>Definición del criterio de ajuste</i>	54
3.2. ALGORITMO DE EVOLUCIÓN.....	55
3.2.1. <i>Parámetros del modelo</i>	58

3.3. GENERACIÓN DE SOLUCIONES	60
3.4. CÁLCULO Y PROCESADO DEL MODELO.....	63
4. EVALUACIÓN, ANÁLISIS DE RESULTADOS Y VALIDACIÓN DEL MODELO.....	67
4.1. GENERACIÓN DE LOS MODELOS A EVALUAR (ARCHIVOS «PROBLEMA»)	67
4.2. IMPLEMENTACIÓN DEL PROGRAMA «CALCULADOR».....	68
4.3. FICHERO «SOLUCIÓN»	69
4.4. EVALUACIÓN DE LAS SOLUCIONES	70
4.5. IMPLEMENTACIÓN DEL EVALUADOR	71
4.6. RESULTADOS	75
4.6.1. <i>Modelo correspondiente al mejor ajuste</i>	75
4.6.1. <i>Búsqueda de comportamiento cíclico en el modelo obtenido</i>	76
5. CONCLUSIONES Y TRABAJOS FUTUROS.....	79
5.1. TRABAJOS FUTUROS	79
6. BIBLIOGRAFÍA.....	81

ÍNDICE DE TABLAS

TABLA 1. CALENDARIO DE VACUNACIÓN EN LA COMUNIDAD VALENCIANA (2006)	4
TABLA 2. CALENDARIO DE VACUNACIÓN EN LA COMUNIDAD VALENCIANA (2014)	4
TABLA 3. CALENDARIO DE VACUNACIÓN EN LA COMUNIDAD DE MADRID (2014)	4
TABLA 4. CRECIMIENTO DE LA BIOMASA DE LA LEVADURA EN UN CULTIVO	17
TABLA 5. TASA DE MORTALIDAD (COMUNITAT VALENCIANA).....	53
TABLA 6. PORCENTAJE INICIAL DE PORTADORES 1997, SITUACIÓN PRE-VACUNAL	54
TABLA 7. PORCENTAJE DE PORTADORES POR GRUPO DE EDAD EN EL AÑO 2011	55
TABLA 8. DEFINICIÓN DE GRUPOS DE EDAD DEL MODELO	71
TABLA 9. AGRUPACIÓN DE EDADES POR MESES EN EL AÑO 2011	72
TABLA 10. VALORES AL AGRUPAR POR EDADES Y POR MESES EN EL AÑO 2011.....	73
TABLA 11. EJEMPLO DEL CÁLCULO DEL ECM PARA EL TURNO 162	73

ÍNDICE DE FIGURAS

FIGURA 1: NEISSERIA MENINGITIDIS, BACTERIA TIPO DIPLOCÓCCICA VISUALIZADA EN 3D	1
FIGURA 2: NEISSERIA MENINGITIDIS, MICROGRAFÍA.....	2
FIGURA 3: NEISSERIA MENINGITIDIS, ERUPCIÓN CUTÁNEA	2
FIGURA 4: LEY DE HOOKE, MODELO DETERMINISTA	13
FIGURA 5: GRÁFICO DE DATOS.....	14
FIGURA 6: EVOLUCIÓN DE LA BIOMASA FRENTE AL CAMBIO DE BIOMASA.....	18
FIGURA 7: LOS PUENTES DE KALINGRADO	22
FIGURA 8: MODELO DE EULER.....	22
FIGURA 9: EJEMPLO DE GRAFO $G(V,A)$ Y SUS CONJUNTOS V Y A (VÉRTICES Y ARISTAS).....	23
FIGURA 10: EJEMPLO DE GRAFOS REGULARES (WIKIPEDIA)	25
FIGURA 11: RED REGULAR CON REPRESENTACIÓN CUADRADA.....	26
FIGURA 12: RED REGULAR CON REPRESENTACIÓN EN ANILLO	26
FIGURA 13: GRAFOS DE PEQUEÑO MUNDO, SITUADOS ENTRE REGULARES Y ALEATORIOS	29
FIGURA 14: DISTRIBUCIÓN DE PROBABILIDAD DE AEROPUERTOS Y LÍNEAS AÉREAS EN USA	30
FIGURA 15: DISTRIBUCIÓN DE PROBABILIDAD DE ACTORES Y SUS RELACIONES	31
FIGURA 16: ESTRUCTURA DE UNA RED GENERADA POR EL MODELO DE ERDŐS Y RÉNYI	32
FIGURA 17: DISTRIBUCIÓN DE POISSON DE CONECTIVIDAD DE UNA RED ALEATORIA	34
FIGURA 18: RED GENÉTICA Y RED DE PROTEÍNAS RESPECTIVAMENTE	35
FIGURA 19: PROBABILIDAD DE CONTAGIO.....	36
FIGURA 20: CÁLCULO DEL NÚMERO R	36
FIGURA 21: RED ALEATORIA PARA MODELAR EL CONSUMO DE TABACO (CHRISTAKIS, 2008)	38
FIGURA 22: TAMAÑO DEL <i>CLÚSTER CON EL PASO DEL TIEMPO</i> (CHRISTAKIS, 2008)	39
FIGURA 23: ESQUEMA DEL MODELO SI (SUSCEPTIBLE-INFECCIOSO).....	43
FIGURA 24: SIMULACIÓN NUMÉRICA DEL MODELO SI (CHITNIS, 2011)	44
FIGURA 25: ESQUEMA DEL MODELO SIS (SUSCEPTIBLE-INFECCIOSO-SUSCEPTIBLE)	44
FIGURA 26: UNA SIMULACIÓN NUMÉRICA DEL MODELO SIS. (CHITNIS, 2011)	45

FIGURA 27: DIAGRAMA DE LA DINÁMICA DE TRANSFERENCIA DEL MODELO SIR.....	46
FIGURA 28: SIMULACIÓN SIR (CHITNIS, 2011).....	47
FIGURA 29: SIMULACIÓN SIR.....	48
FIGURA 30: PIRÁMIDE DE POBLACIÓN DE LA COMUNIDAD VALENCIANA	52
FIGURA 31: DIAGRAMA DE FLUJO DEL ALGORITMO.....	57
FIGURA 31: RED ALEATORIA CON PORTADORES	59
FIGURA 33: INTERVALOS HÍPER CUBO LATINO PARA X_2	61
FIGURA 34: ARQUITECTURA DE SISIFO	64
FIGURA 35: COMPORTAMIENTO CÍCLICO DEL MODELO EN 2000 TURNOS.....	68
FIGURA 36: EJEMPLO DE FICHERO SOLUCIÓN	69
FIGURA 37: GRAFICA DEL TURNO CON MENOR ECM EN 2011 DEL MODELO AJUSTADO	76
FIGURA 38: COMPORTAMIENTO CÍCLICO DEL MODELO A LO LARGO DE 2000 TURNOS	77
FIGURA 39: COMPORTAMIENTO NO CÍCLICO DEL MODELO EN 2000 TURNOS	78

1. Introducción

Las enfermedades han condicionado la historia de la Humanidad, siendo especialmente graves las infectocontagiosas. Asolando poblaciones enteras en todas las épocas, se sitúan entre los mayores temores del ser humano. Un ejemplo de este temor lo representan las numerosas referencias sobre las enfermedades que se realizaban en textos antiguos, a destacar las múltiples citas bíblicas sobre epidemias y plagas (Isaías 37: 36-38). El número de víctimas cobrado por estas enfermedades era tan elevado que provocaba que hubiera grandes cambios demográficos y sociales, como sucedió con la peste o el cólera (McNeill, 2010).

De entre las muchas enfermedades infectocontagiosas que hay llama nuestra atención la meningitis, que es causada por la bacteria del meningococo, nombre común que recibe la *Neisseria meningitidis* (Figura 1), la cual tiene gran importancia en la salud pública.

La meningitis es una infección del cerebro y de la médula espinal. Se transmite exclusivamente entre seres humanos y puede llegar a matar a un paciente en pocas horas si la infección alcanza la sangre (Figura 2) en lo que se denomina *sepsis meningocócica* (Blanco et al., 2004), y que causa al paciente infectado una fiebre elevada repentina, acompañada de erupciones cutáneas de color púrpura (Figura 3), hipotensión, insuficiencia suprarrenal aguda, fallo multiorgánico, shock refractario y muerte.

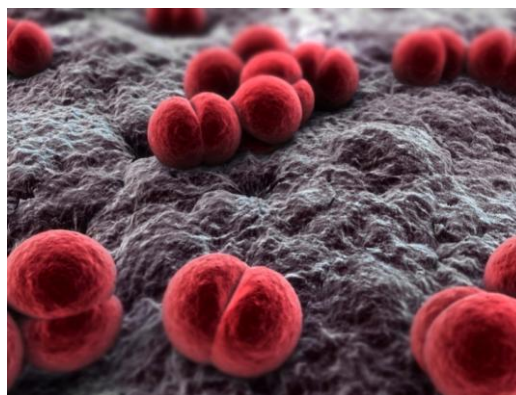


Figura 1: *Neisseria Meningitidis*, bacteria tipo diplocócica visualizada en 3D

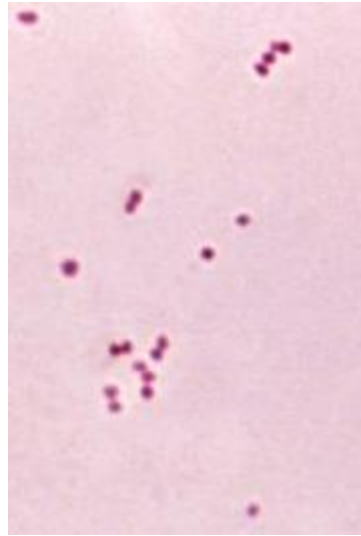


Figura 2: Neisseria Meningitidis, micrografía



Figura 3: Neisseria Meningitidis, erupción cutánea

Aunque hay varios tipos de meningitis (causadas por distintos tipos de virus y bacterias), al hablar de meningitis se entiende la producida por la bacteria del meningococo C (MenC). Esta enfermedad infecciosa tiene asociada una mortalidad que ronda el 20% de los afectados, y los que sobreviven sufren secuelas neurológicas permanentes (ONS, 1997). El contagio se realiza por contacto entre portadores de la bacteria y no portadores. En principio un portador no tiene porqué desarrollar la enfermedad (portador sano) pero, cuando esto ocurre, las consecuencias son aún más graves si cabe, ya que se multiplica la capacidad de infección de dicho portador.

La transmisión del meningococo se produce, fundamentalmente, por las secreciones rinofaríngeas emitidas a través de la tos o el estornudo, viéndose favorecida por la proximidad y el contacto físico, y factores sociales como el hacinamiento y los besos.

Aunque se sabe todavía poco de su comportamiento y dinámica de difusión, hay una política de vacunación preventiva implantada en los países desarrollados. En Estados Unidos, por ejemplo, el Centro de Control y Prevención de Enfermedades (CDC) recalca la importancia de seguir unas pautas de vacunaciones precisas, bien llevadas al día, como mejor defensa contra el meningococo.

En esta línea de actuación existe una vacuna conjugada para esta cepa, la tipo C, cuyo nombre es MCC.

El Reino Unido, en noviembre de 1999, fue el primer país en administrar la vacuna MCC (Balmer et al., 2002). Los resultados entre la población han sido cruciales a la hora de extender el uso de la MCC a otros países. De hecho desde el año 2000 la vacuna está siendo utilizada con diferentes estrategias dependiendo del país que la administra.

En España tenemos como ejemplo a la Comunidad Valenciana, donde la política de vacunación ha ido cambiando a lo largo de los años, y a la Comunidad de Madrid. El resto de comunidades también tiene su propio calendario de vacunación.

En la Comunidad Valenciana en el año 2001 la vacuna de la meningitis era inicialmente para niños menores de 6 meses. Posteriormente se renovaba con una dosis entre 1 y 6 años. En 2002 la dosis se amplía a las personas de 19 años. En 2006 cambian los plazos como se muestra en la Tabla 1, y vuelve a cambiar más tarde hasta quedar el calendario actual como se presenta en la Tabla 2.

	
Edad	Vacuna
2 meses	MenC
6 meses	MenC
18 meses	MenC

Tabla 1. Calendario de vacunación en la Comunidad Valenciana (2006)

	
Edad	Vacuna
4 meses	MenC
12 meses	MenC
12 años	MenC

Tabla 2. Calendario de vacunación en la Comunidad Valenciana (2014)

En la Comunidad de Madrid se usan otros parámetros. Se ha aumentado a cuatro dosis la vacuna, a los 2, 4, 12 meses y 12 años, según la Tabla 3 (Portal de la Salud <http://www.madrid.org/>).


	
Edad	Vacuna
2 meses	MenC
4 meses	MenC
12 meses	MenC
12 años	MenC

Tabla 3. Calendario de vacunación en la comunidad de Madrid (2014)

Debido a la escasa homogeneidad de criterios entre comunidades autónomas, el Grupo de Trabajo MenCC 2012 recomendó en el 2013 un nuevo calendario de vacunación: 2 meses, 12 meses y 12 años de edad. (Grupo de trabajo MenCC, 2013).

Como se puede observar, no hay un criterio unificado a la hora de implantar la política de vacunación, y ésta se basa no sólo en resultados de ensayos clínicos, sino también en eventos inesperados (como nuevos brotes) u objetivos de política sanitaria (distintos presupuestos en comunidades, imagen política), que no siempre están respaldadas por la comunidad médica.

Todo lo anteriormente expuesto nos hace preguntarnos si cabría la posibilidad de modelar de alguna manera el comportamiento de la enfermedad y el impacto que sobre ella tienen las distintas políticas de vacunación. Si somos capaces de definir tal modelo podríamos dar una respuesta objetiva a preguntas tales cómo qué calendario es mejor, a cuánta gente hay que vacunar, con qué coste, etc.

1.1. Motivación

El comportamiento y la dinámica de difusión de muchas enfermedades infecciosas es, por lo general, algo no totalmente entendido. Sería de gran interés para las autoridades encargadas de las políticas de salud pública tener un conocimiento más amplio de cómo se transmiten esas enfermedades. De esta forma se podrían tomar mejores decisiones respecto a qué políticas de salud son las más adecuadas y hacer un uso más optimizado de los recursos, por ejemplo no vacunando a toda la población si no es necesario sino sólo a una parte.

Aquí es donde entran en juego los modelos matemáticos, que son capaces de simular el comportamiento de sistemas complejos.

A medida que avanza la ciencia de la computación también lo hace la capacidad para modelar y simular sistemas reales más y más complejos, como por ejemplo redes de transporte (metro, tren, líneas aéreas), sistemas informáticos (redes de servidores, Internet); biología (interacción neuronal, ADN); y por último las redes epidémicas, que son las que han suscitado nuestro interés y donde cabría formular la pregunta sobre si es posible realizar la simulación y la predicción del MenC.

Como se ha mencionado, la meningitis C es bastante peligrosa, y pese a que existen campañas de vacunación y hay un conocimiento del número de vacunados por comunidad autónoma en el caso de España (Coberturas de vacunación, *Ministerio de Sanidad, Servicios Sociales e Igualdad, Junio 2014*), éste no permite verificar si las pautas que se siguen son las más acertadas. Tampoco se sabe qué puede pasar en el futuro a una población que siga una u otra pauta de vacunación, qué ocurriría en los diferentes segmentos de la población, qué características tienen los individuos, o qué comportamiento es el que define a la propagación de la enfermedad en una determinada región.

La cepa objeto de este estudio es la C, debido a que es la más extendida y a que se cuenta con datos históricos del protocolo llevado en España durante los últimos años. Nos hemos basado en concreto en el estudio llevado a cabo en la Comunidad Valenciana, desarrollado por varias universidades como la Universidad Complutense, la Universidad Politécnica de Valencia, la Universidad de Valencia; el instituto de Salud Carlos III y el Centro Superior de Investigación en Salud Pública de la Generalitat Valenciana.

En este trabajo proponemos un modelo de la dinámica de transmisión del MenC. El tipo de modelo escogido será el de uno basado en redes aleatorias discretas.

Como veremos posteriormente, el hecho de decantarnos por este tipo de modelo traerá consigo unas ventajas y unos inconvenientes que darán pie a nuevos objetivos y soluciones que iremos analizando con el debido detalle en cada una de los capítulos que componen este trabajo.

1.2. Objetivos

La realidad muestra que el modelo de transmisión del meningococo C tiene un comportamiento cíclico (Chabalier et al., 2000). Se llama comportamiento cíclico al hecho de que la cantidad de portadores no es fija si no que oscila dentro de unos límites en función del tiempo. Fuentes de información como el Centro de Control y Prevención de enfermedades (CDC) nos dice que la epidemia generalmente tiene ciclos irregulares cada 5 a 12 años en África.

El modelo de redes aleatorias para la transmisión del meningococo C que vamos a proponer debe ser capaz de reproducir este comportamiento.

El modelo tiene una serie de parámetros cuyo ajuste es fundamental para determinar si el modelo tiene un comportamiento cíclico. Por ello se deben generar muchas pruebas con unos parámetros realistas, dentro de los límites del dominio de la medicina.

Después se simularán esas pruebas y se obtendrán unas salidas que deben ser evaluadas para escoger aquellas donde los parámetros se ajusten en mayor medida a la realidad.

1.2.1. Objetivos detallados

- Implementación del modelo mediante redes aleatorias. Realizado con el lenguaje de programación Visual C++ de Microsoft para optimizar las operaciones y la estructura de datos generada en memoria que sustenta a la red aleatoria.
 - Establecimiento de las reglas de evolución de la red aleatoria.
- Generador de pruebas para ajuste (parametrización de la red). Implementado en lenguaje Mathematica (Wolfram Mathematica).
- Generador de soluciones (computación distribuida). Implementado con el sistema Sisifo.
- Evaluador de soluciones. Implementado en Mathematica.
- Visualizador de resultados. Implementado en Mathematica.

1.3. Estructura de la memoria

Esta memoria se estructura en los siguientes capítulos.

En este Capítulo 1 se ha introducido la enfermedad objeto de estudio, presentando la motivación que ha llevado a su análisis y modelización.

En el Capítulo 2 se presenta el estado de la cuestión, describiendo los modelos, donde nos encontraremos entre otros los modelos de redes aleatorias, que provienen de la teoría de grafos, cuya base teórica es el modelo Erdős.

En el Capítulo 3 se describe el problema, presentando los datos disponibles que permiten elaborar el modelo de red aleatoria. Se detalla la forma en que se genera la red aleatoria, sus reglas de evolución y la generación de las soluciones.

En el Capítulo 4 se valora el grado de bondad obtenido en el modelo que hemos creado, así como las dificultades encontradas en la simulación y cómo se abordaron las soluciones.

En el Capítulo 5 se presentan las conclusiones, donde nuestro modelo de transmisión del meningococo C cobra un valor y un propósito. También se presentan los posibles trabajos futuros.

La memoria termina con las fuentes consultadas y citadas en el trabajo.

2. Estado de la cuestión

2.1. Modelos

Kendall y Buckland (1957) definen modelo como una expresión formalizada de una teoría o la situación causal, que concierne a la generación de unos datos observados.

Hablando del mundo real, un modelo se puede definir como una descripción de un hecho o fenómeno concreto, desde el tamaño de la población, hasta fenómenos físicos como la velocidad, aceleración o densidad.

Los modelos matemáticos tratan de representar fenómenos o relaciones entre ellos a través de una formulación analítica. Son una herramienta que permite la estimación de una variable (variable dependiente) sobre la base de los valores que tomen otras (variables independientes). Se usan en amplias áreas del conocimiento; en las finanzas y en estudios de población, se hacen “predicciones” a futuro (extrapolaciones) sobre la base del comportamiento de algunos indicadores (Infante y Zárate, 1984); en epidemiología se utilizan para prever el comportamiento de las enfermedades utilizando el comportamiento de variables ambientales y la disponibilidad de inóculo como variables independientes (Dropkin, 1980; Sutton, 1996). Para la estimación de variables fisiológicas también han demostrado su utilidad los modelos matemáticos (Thornley, 1976).

Un sistema se define como una colección de partes que desarrollan funciones e interactúan juntas para el cumplimiento de algún propósito definido. En la práctica cada sistema depende de los objetivos de un estudio particular.

Eykhoff (1974) define un modelo como una representación de los aspectos esenciales de un sistema existente (o un sistema a ser construido), que presenta el conocimiento de ese sistema en forma utilizable.

El objetivo del modelo de un sistema es entender ampliamente el fenómeno y tal vez predecir su comportamiento en el futuro.

Utilizando el lenguaje matemático que nos proporciona la teoría de conjuntos podemos formalizar la noción de sistemas:

Sean a_1, a_2, \dots, a_n los elementos del sistema "S" y a_o su entorno, denotando como r_{ij} la relación existente entre a_i y a_j , el sistema "S" puede representarse como $S = \{A, R\}$

Siendo:

- A el conjunto de elementos primarios o universo del sistema (excluyendo el entorno)
- R el conjunto de relaciones entre sus elementos incluyendo el entorno o características del sistema.

Un modelo de un sistema dinámico consiste en un espacio de estados, cuyas coordenadas describen el estado en cualquier instante, y en una regla dinámica que especifica el futuro inmediato de todas las variables de estado, teniendo en cuenta sólo los valores actuales de esas mismas variables de estado (Meiss 2007).

Un sistema dinámico es un sistema físico cuyo estado evoluciona con el tiempo. El comportamiento en dicho estado se puede caracterizar determinando los límites del sistema, los elementos y sus relaciones; de esta forma se puede elaborar modelos que buscan representar la estructura del mismo.

La expresión general de un modelo dinámico se puede definir como:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_g X_{gi} + \beta_j X_{1i-1} + \beta_k Y_{i-1} + u_i \quad (1)$$

Donde la Y es la salida del modelo, con k variables explicativas y un término de error u que determina el comportamiento de la variable de salida Y , donde el subíndice i hace referencia a las diferentes observaciones, denotadas por cada X para las cuales se establece su validez. Es decir, distintos puntos de información o datos muestrales, siendo β valores de los coeficientes. β_0 representa el valor de la variable dependiente Y cuando los valores de las variables independientes X son iguales a cero.

De manera general (y simplificada), la modelización puede describirse de forma sistemática por medio de los siguientes pasos a seguir (Dreyer, 1993):

- 1) Identificar el sistema del mundo real que se quiere modelar (problema). Se trata de clarificar las preguntas que se intentan responder con el modelo, formular el problema en palabras, documentar los datos relevantes e identificar el mecanismo subyacente del problema real.
- 2) Suposiciones: el problema debe ser analizado para decidir los factores del mismo que son importantes y aquellos que pueden ser ignorados. Con todo ello deben hacerse suposiciones (o idealizaciones) lo más realistas posible.
- 3) Construcción: en este paso se “construye” el modelo, es decir, se traduce al lenguaje matemático el problema (junto con las suposiciones anteriormente realizadas) obteniéndose un conjunto de ecuaciones (o inecuaciones) después de haber identificado las variables que deben intervenir en las mismas.
- 4) Análisis o resolución: se trata de la resolución del problema. Las soluciones consistirán en general en funciones por medio de las cuales la o las variables dependientes se expresarán en términos de la o las variables independientes. Por otro lado, se obtendrá información acerca de los parámetros que intervienen en el modelo.
- 5) Interpretación: en este paso, la solución matemática debe ser comparada con la realidad para observar si se ajusta a lo conocido acerca del problema real. Se trata, en definitiva, de interrumpir el proceso si se obtienen soluciones carentes de sentido real.
- 6) Validación: una vez interpretada la solución, se comprueba numéricamente que concuerda con los datos disponibles sobre el problema. Contrastar la bondad de los resultados del modelo con datos reales. Si el grado de fidelidad obtenido no es suficiente, reformular el modelo y repetir el proceso.

Es importante mencionar que la inmensa mayoría de modelos matemáticos no son exactos per se y tienen un alto grado de idealización y simplificación, pues ante modelos muy complejos, o bien una modelización exacta es imposible, o bien no es practicable desde el punto de vista computacional,

Dada la heterogeneidad de los modelos matemáticos, es difícil dar una taxonomía completa, pero podemos agruparlos según dos grandes clasificaciones generales:

- 1) Atendiendo a la relación que tienen entre los datos de entrada y los de salida:

- a. Deterministas.
 - b. Estocásticos.
- 2) Atendiendo a sus variables de estado, por ejemplo el tiempo, tenemos:
- a. Discretos.
 - b. Continuos.

2.1.1. Modelos deterministas

Los modelos de sistemas dinámicos son deterministas si hay un consecuente único para todos los estados, o estocásticos o aleatorios si hay una distribución de probabilidad de las posibles consecuencias (Meiss 2007). Por ejemplo el lanzamiento de una moneda tiene dos posibles salidas con igual probabilidad para cada estado inicial.

Un sistema dinámico puede ser discreto o continuo. Un sistema determinista con tiempo discreto está definido por,

$$x_1 = f(x_0), \tag{2}$$

que devuelve el estado x_1 obtenido a partir del estado inicial x_0 en el siguiente valor en el tiempo. Después de un tiempo n tenemos

$$x_n = f^n(x_0), \tag{3}$$

donde f^n es la n -ésima iteración de f .

Un sistema determinista con tiempo continuo se define por un flujo,

$$x(t) = \varphi_t(x(0)), \tag{4}$$

que devuelve el estado en el tiempo t , dado que el estado fue $x(0)$ en el tiempo 0.

Por ejemplo, la planificación de una línea de producción en cualquier proceso industrial es posible realizarla con la implementación de un sistema de gestión que incluya un modelo determinista en el cual estén cuantificadas las materias primas, la mano de obra, los tiempos de producción y los productos finales asociados a cada proceso.

Otro ejemplo sería el estado de un péndulo, su ángulo y la velocidad angular, cuya ecuación se basa en la ley de Newton $F = ma$.

La mayoría de los modelos de la mecánica clásica son deterministas. Por ejemplo, el movimiento de un oscilador armónico simple (un muelle sujeto a la Ley de Hooke) es un modelo determinista (Figura 4).

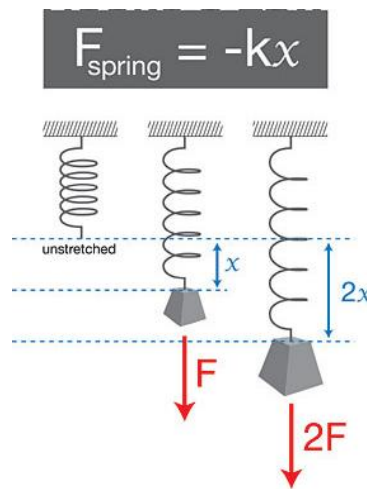


Figura 4: Ley de Hooke, modelo determinista

Imaginemos que deseamos determinar la constante k de un resorte que sigue la ley de Hooke:

$$F = -k x \quad (5)$$

donde F es la fuerza elástica y x la elongación del resorte. Para determinar k se procede a cargar al resorte con diferentes pesos P y medir la elongación que producen. Es fácil reconocer a estas cargas como el estímulo externo, el cual provoca como respuesta del "sistema resorte" la elongación observada x . Sin embargo, no es necesario que en la representación gráfica de los pares de datos (P, x) , la carga P ocupe el lugar de la variable independiente sobre el eje de las abscisas. Por comodidad en el manejo de los datos, es más adecuado representar P en función de x , y de la pendiente de la recta obtener directamente la constante k buscada. En el presente caso, y como regla casi general, si representamos gráficamente los datos experimentales, éstos no caerán exactamente sobre una recta, sino que presentarán cierta dispersión como se ilustra en la Figura 5.

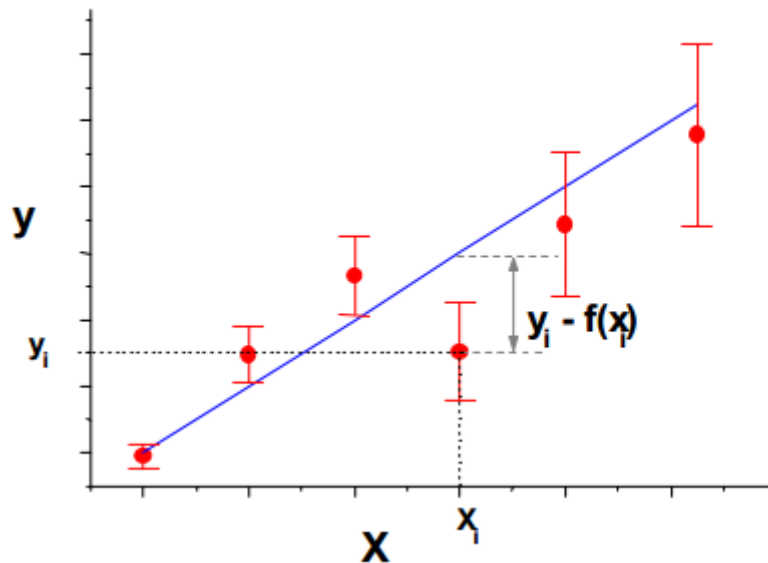


Figura 5: Gráfico de datos.

La cantidad $y_i - f(x_i)$ representa la desviación de cada observación de y_i respecto del valor predicho por el modelo $f(x_i)$.

2.1.2. Modelos estocásticos

El modelado estocástico asume la incertidumbre del mundo real y trata de modelarlo mediante el uso de la probabilidad. En ellos al menos una variable del sistema se toma como un dato al azar y las relaciones entre variables se describen por medio de funciones probabilísticas.

Es importante aclarar que el uso de este tipo de modelos no implica que el sistema bajo estudio se comporte de manera “aleatoria”. Por ejemplo, el comportamiento de un individuo puede parecer impredecible, pero si lo entrevistamos y observamos unas preferencias, puede revelar un comportamiento predecible.

De una manera más formal, un proceso estocástico se define como una familia de variables aleatorias $\{\xi_s\}_{s \in S}$, definidas sobre un mismo espacio de probabilidad (Ω, B, P) , que representa el comportamiento aleatorio del fenómeno que se estudia. En este contexto ξ_s suele representar el valor de una variable de interés ξ medida en el instante s , o el valor que toma esa variable sobre el s -ésimo objeto que interviene en el fenómeno analizado. Por ejemplo:

- ξ_t = número de llamadas telefónicas recibidas en una centralita hasta cierto instante t .
- ξ_k = Tiempo que permanece en el buffer de un conmutador el k -ésimo paquete de una conexión desde que llega hasta que es transmitido.

Asimismo el espacio de probabilidad (Ω, B, P) representa el azar. Los elementos $\omega \in \Omega$ son los posibles resultados aleatorios de la observación del fenómeno; P es la función que asigna a los ω su probabilidad de ocurrencia $P(\omega)$, y B es la colección de conjuntos de Ω a los que se puede asignar de modo consistente un valor de probabilidad.

Dentro del conjunto de procesos estocásticos se encuentran, por ejemplo, el tiempo de funcionamiento de una máquina entre avería y avería, su tiempo de reparación y el tiempo que necesita un operador humano para realizar una determinada operación.

Podemos simular las colas que se forman en un mostrador utilizando una variable estocástica que indique el número (aleatorio) de clientes que entran en la oficina por minuto, y luego usando un algoritmo que use el tiempo de atención que se da a cada cliente, que también puede ser estocástico.

2.1.3. Modelos discretos

En un modelo discreto los cambios en el tiempo son predominantemente discontinuos o instantáneos, es decir, que las propiedades que describen su comportamiento cambian en momentos determinados de tiempo, y entre esos instantes no sucede variación alguna.

Un modelo discreto se formula mediante:

- la especificación del estado del sistema a modelar usando un conjunto discreto de variables x_1, x_2, \dots, x_n (donde las variables x_i no tienen por qué ser numéricas, necesariamente),
- y un algoritmo que permite calcular las variables de estado y de salida, a partir de las variables de entrada: $x_k = f(x_{k+1}, x_{k+2}, \dots, x_n)$.

Las variables de entrada pueden incluir al tiempo u otras variables independientes. Pero, por ser el algoritmo discreto, sólo nos interesará conocer los valores de salida del sistema en un conjunto discreto (finito o numerable) de valores de dichas variables independientes.

El cambio en las variables de entrada puede estar ocasionado por la evolución temporal del modelo (esto ocurre solamente en los modelos dinámicos), o por la intervención del usuario sobre la simulación, experimentando con el modelo y cambiando alguno de sus valores de entrada. Esto puede pasar en todo tipo de modelos, pues ocurre a resultas del deseo del usuario de experimentar con el modelo.

La estructura del modelo de un sistema discreto en general está compuesta por:

- Entidades: son objetos de interés en el sistema para un determinado objetivo.
- Atributos: son las propiedades o características de las entidades. Permiten describir cuantitativamente al sistema.
- Estado: son los valores de los atributos de las entidades en un instante de tiempo. Los atributos describen el estado de una entidad, y los estados de las entidades más importantes definen el estado del sistema.
- Evento: es el cambio instantáneo de estado. Los eventos pueden ser continuos si los cambios se producen continuamente en el tiempo, o discretos si se producen a intervalos de tiempo.
- Actividad: es la operación que produce transformaciones en los estados del sistema. Las actividades comienzan y terminan con eventos.
- Funciones: Establecen las relaciones entre los atributos.

Ejemplos:

- El tamaño de una población de insectos en el año i .
- La proporción de individuos de una población que lleva un gen particular en la generación i -ésima.
- El número de células en un cultivo bacteriano en i día.
- La concentración de un gas tóxico en el pulmón después de la respiración i -ésima.
- La concentración de fármaco en la sangre después de la dosis i -ésima.

Pongamos como ejemplo el modelado del crecimiento de un cultivo de levadura, partimos de los siguientes datos observados (Tabla 4).

Tiempo (horas) n	Biomasa de la levadura p_n	Cambio en la biomasa $\Delta p_n = p_{n+1} - p_n$
0	9.6	8.7
1	18.3	10.7
2	29.0	18.2
3	47.2	23.9
4	71.1	48.0
5	119.1	55.5
6	174.6	82.7
7	257.3	

Tabla 4. Crecimiento de la biomasa de la levadura en un cultivo

Para modelar tenemos que:

$$x_{n+1} = x_n + \Delta x_n \tag{6}$$

donde la meta del modelado es encontrar una aproximación razonada de Δx_n que reproduzca un conjunto de datos del fenómeno observado.

El cambio en la biomasa es proporcional a la población:

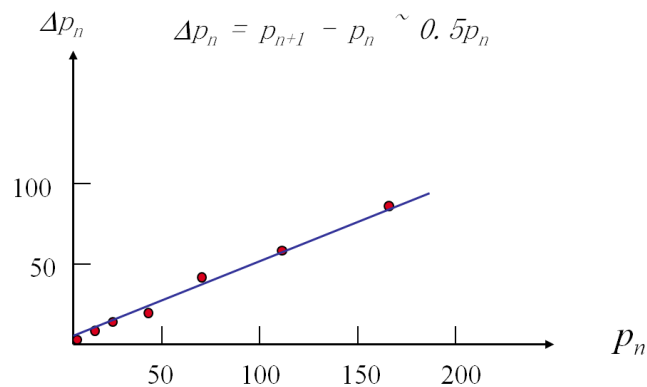


Figura 6: Evolución de la biomasa frente al cambio de biomasa

A juzgar por los resultados obtenidos podemos estimar que:

$$\Delta p_n = p_{n+1} - p_n \sim 0.5p_n \quad (7)$$

donde obtenemos el modelo:

$$p_{n+1} = p_n + 0.5p_n = 1.5p_n \quad (8)$$

La solución por tanto es:

$$p_{n+1} = 1.5(1.5p_{n-1}) = 1.5[1.5(1.5p_{n-2})] = \dots = (1.5)^{n+1} p_0 \quad (9)$$

Obteniendo entonces:

$$p_n = (1.5)^n p_0. \quad (10)$$

Este modelo predice una población que se incrementa siempre. Este resultado nos invita a volver a examinar los datos observados para tratar de obtener un modelo mejor. Por ejemplo tomando más datos podríamos encontrar que la biomasa de la levadura tiene un límite en 665, por lo que nuestro modelo quedaría:

$$\Delta p_n = k(665 - p_n) p_n \quad (11)$$

$$p_{n+1} = p_n + k(665 - p_n) p_n \quad (12)$$

2.1.4. Modelos continuos

El modelo es continuo si permite conocer los valores de salida en todos los instantes de un intervalo de tiempo. Los modelos de sistemas dinámicos continuos suelen estar basados en ecuaciones diferenciales, tanto ordinarias como parciales.

En un modelo continuo, se desea conocer el estado del sistema para valores de una o más variables independientes que pueden tomar cualquiera de los valores de un intervalo de la recta real.

Dentro de los continuos, el caso más frecuente es el de modelos dinámicos, que son los basados en ecuaciones diferenciales ordinarias EDO. En ellos la variable independiente es el tiempo t y la ecuación que gobierna la variación del estado viene dada por una EDO. Por ejemplo, en la ecuación de la oscilación forzada de una masa en el extremo de un muelle:

$$m \ddot{x}(t) = -k((t) - L) - b\dot{x}(t) + f(t) \quad (13)$$

donde $f(t)$ es la fuerza externa que se ejerce sobre la masa.

Paradójicamente, en muchos de estos sistemas, el tiempo puede no aparecer explícitamente en la ecuación diferencial (sistema dinámico). Esto ocurre habitualmente en las ecuaciones de la mecánica clásica, como en la atracción gravitatoria entre planetas.

En la construcción de un modelo para un sistema continuo se deben distinguir:

- Componentes: son partes constituyentes del sistema, también llamados elementos o subsistemas.
- Variables: están asociadas al concepto matemático de variable. Pueden clasificarse en:
 - Exógenas, de entrada o independientes: son las que afectan al sistema, pero éste no puede modificarlas. Pueden modificarse arbitrariamente desde el medio ambiente.
 - Endógenas o Dependientes: Son variables del sistema que se modifican de acuerdo a relaciones, no pueden ser modificadas arbitrariamente.

- De estado: Es el conjunto mínimo de variables dependientes que permiten describir el sistema en $t + \Delta t$, si se conocen sus valores más los valores de las independientes en t .
- De salida: Es el conjunto mínimo de variables de estado que permiten evaluar los objetivos del modelo.
- Parámetros: son los valores de entrada del modelo. Están asociados al concepto de constante matemática. No cambian su valor en una ejecución o escenario de simulación, pero pueden ser modificados entre ejecución o escenarios.
- Relaciones funcionales: es una relación entre parámetros de uno o más componentes y las variables de salida.
- Restricciones: Son las limitaciones que tiene el sistema y pueden ser:
 - Naturales: Son las que físicamente no pueden violarse. Ejemplo: si el recipiente está herméticamente cerrado en ningún momento puede variar la cantidad de agua total.
 - Autoimpuestas: se imponen para acotar el estudio. Ejemplo: la temperatura inicial será siempre de 30°C. Incluye las hipótesis.
- Funciones Objetivo: son los objetivos y las metas del modelo. Implica definir para qué se modela y simula y cómo se evaluarán los factores que intervendrán en la experiencia.

El mayor problema para implementar un modelo continuo en un ordenador digital consiste en realizar una discretización que permita aproximar con la suficiente fidelidad la solución del problema continuo.

2.1.5. Modelos de redes

El National Research Council define a la ciencia de las redes como “el estudio de las representaciones de las redes de fenómenos físicos, biológicos y sociales que conducen a modelos de predicción de estos fenómenos”.

La ciencia de las redes es un campo interdisciplinario y académico que estudia las redes complejas, como las redes de ingeniería, redes de información, redes biológicas, cognitivas, redes semánticas y redes sociales. Se basa en las teorías y métodos de la teoría de grafos, así como las matemáticas, la mecánica estadística, la física, la

representación y codificación de datos, la minería de datos, la de inferencia estadística.

Las redes de comunicación, la World Wide Web (www), el genoma humano, las redes de proteínas, las redes neuronales, las de transportes, las redes sociales, las redes de colaboración científica o las redes terroristas son algunos ejemplos. Incluso el lenguaje que nos sirve para escribir este trabajo es una red, compuesta por palabras unidas por relaciones sintácticas y semánticas (Barabási y Albert, 1999a; Barabási et al., 1999b), Bonabeau, 2003; Börner et al., 2005).

Gracias a los avances en las ciencias de la computación, el estudio y el análisis de las redes ha mejorado. La adquisición y manipulación de datos por ordenador ha permitido manejar voluminosas bases de datos, y el aumento del potencial de computación ha permitido la exploración de redes con millones de nodos (Albert y Barabási, 2002).

Se considera que la teoría de redes tuvo su inicio con el matemático suizo Leonhard Euler que planteó el curioso problema de los siete puentes sobre el río Pregel de la actual ciudad prusiana de Kaliningrado, que hasta 1936 se llamó Königsberg (Figura 7).

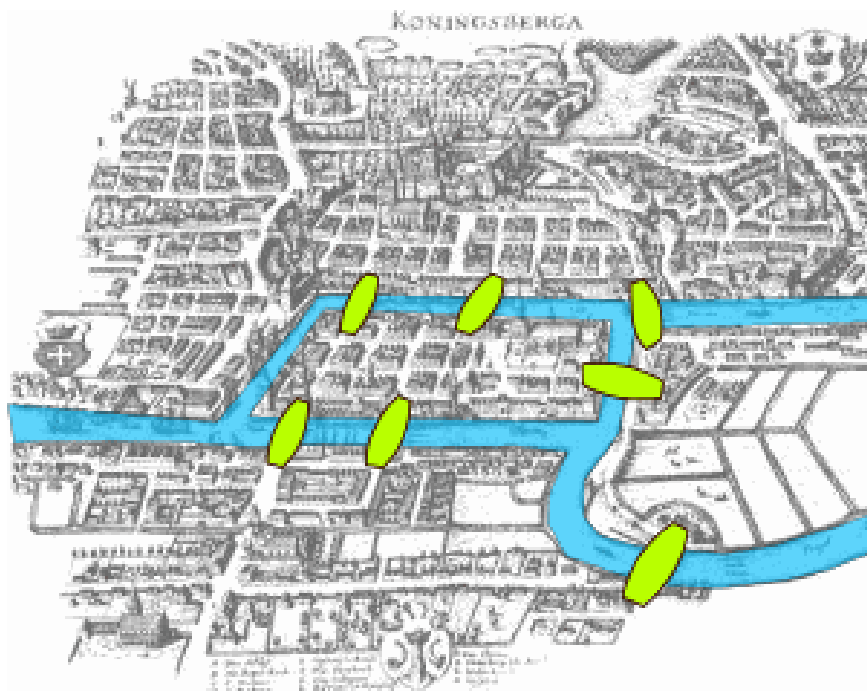


Figura 7: Los puentes de Kalingrado

¿Es posible dar un paseo comenzando por cualquiera de las cuatro partes de tierra, cruzando cada puente una sola vez y volviendo al punto de partida? Euler representó cada parte de tierra por un punto y cada puente por una línea, haciendo la siguiente pregunta: ¿se puede recorrer el dibujo sin repetir las líneas? (Figura 8). La respuesta es no.

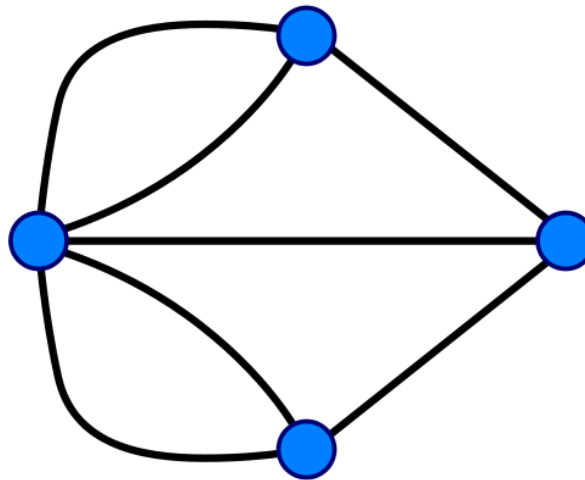


Figura 8: Modelo de Euler

A continuación se revisará de manera sucinta la teoría de grafos y la estructura y diferencias que se dan entre los distintos tipos de redes del mundo real, basadas en los análisis que tratan de explicar y modelar sus propiedades.

2.1.5.1 Una breve introducción a la teoría de grafos

Una grafo (en literatura matemática) o red (en literatura computacional), formalmente hablando, es un par de elementos $\{v, w\} \in (V(G), E(G))$ donde los elementos $v \in V(G)$ son llamados vértices y los elementos w de $E(G)$, son las aristas (Albert y Barabási, 2002).

Para describirlo gráficamente, comúnmente $v \in V(G)$ se muestran como puntos y se dibuja un arco entre $v \in V(G)$ y w cuando son adyacentes (Figura 9). Se le llama a $V(G)$ conjunto de vértices de G y a $E(G)$ el conjunto de aristas de G . Cuando se sobreentiende lo que significa G en el contexto, simplemente escribimos V y E respectivamente. También es común escribir $v \in G$ o $\{v, w\} \in G$, en lugar de la notación formal $v \in V(G)$.

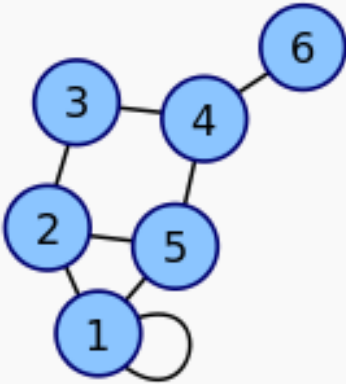
Grafo $G(V,A)$	Conjuntos
	$V = \{ 1, 2, 3, 4, 5, 6 \}$ $A = \{ \{1,1\}, \{1,2\}, \{1,5\}, \{2,3\}, \{2,5\}, \{3,4\}, \{4,5\}, \{4,6\} \}$

Figura 9: Ejemplo de grafo $G(V,A)$ y sus conjuntos V y A (vértices y aristas)

Algunas definiciones básicas sobre grafos de la literatura consultada y es necesario tener en cuenta a la hora de describir las redes aleatorias son:

- **Vértices:** unidad fundamental de la red, también llamados individuos.
- **Arista:** línea que conecta dos vértices.
- **Direccionado o no direccionado:** una arista está direccionada si va en un solo sentido y no lo está si puede ir en ambos sentidos. Un grafo direccionado es aquel donde todas sus aristas están direccionadas.
- **Grado o valencia:** es el número de aristas que están conectadas a un vértice.
- **La distribución de conexiones (o vecinos) $P(k)$:** Es la probabilidad de que un nodo escogido al azar tenga k conexiones (o vecinos). Por ejemplo, en una red de contactos sexuales $P(k)$ es la probabilidad de que una persona escogida al azar en una sociedad haya tenido k parejas sexuales distintas a lo largo de su vida.
- **Conectividad:** grado de conexión interna de la red. Nos dice como se distribuyen las relaciones de los individuos dentro de la red.
- **Clique:** un conjunto $S \subseteq A$, siendo A el conjunto de aristas, de forma que si todos los pares $(x,y) \in S$ son adyacentes. El número de clique, a menudo denotado por $\omega(G)$, es el número del mayor clique de G .
- **Coeficiente de agrupamiento (*clustering*):** una propiedad común de las redes sociales es que forman cliques, representado círculos de amistad o relación.

Ésta tendencia inherente de agrupamientos ha sido cuantificada como coeficiente de agrupamiento (Watts y Strogats, 1998).

De manera más formal, tomamos i como nodo en la red, que contiene k aristas que lo conectan a otros k_i nodos. Si los primeros vecinos del nodo original forman parte de un clique, habrá $k_i(k_i-1)/2$ aristas entre ellos. La relación entre el número de aristas E_i que realmente existe entre esos k_i nodos y el número total $k_i(k_i-1)/2$ resulta el valor del coeficiente de agrupamiento del nodo i ,

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (14)$$

El coeficiente de agrupamiento de toda la red viene dado por la media de todos los C_i 's. En un grafo aleatorio, como las aristas están distribuidas aleatoriamente, el coeficiente de agrupamiento es $C = p$, siendo p la probabilidad de que dos nodos vecinos estén conectados.

- Un conjunto independiente S es aquel en el que no existen pares $x, y \in S$ que sean adyacentes. Al mayor tamaño del conjunto independiente se le denomina el número independiente y se denota por $\alpha(G)$.
- Otra propiedad, la coloración de grafos, es un caso especial de etiquetado de grafos. Sea G un grafo, la k -coloración de G se denota como un mapa $f: V \rightarrow \{1, 2, \dots, k\}$ tal que si x, y son adyacentes tenemos que $f(x) \neq f(y)$. Al menor valor de k , para el que la k -coloración existe se le denomina el número cromático de G y se denota por $\chi(G)$.

Se ha de tener en cuenta que $\omega(G) \leq \chi(G)$ debido a que todos los vértices de un clique deben estar coloreados de distinta forma.

- El complementario de G , denotado por \bar{G} , tiene exactamente el mismo conjunto de vértices que G . Además x, y son adyacentes en \bar{G} si y solo si x, y no lo son en G .
- Un grafo regular es un grafo donde cada vértice tiene el mismo grado. Un grafo regular con vértices de grado k es llamado grafo k -regular o grafo regular de grado k .

- Un grafo completo de n vértices tiene $n(n-1)/2$ aristas, y se denota K_n . Es un grafo regular con todos sus vértices de grado $n-1$. La única forma de hacer que un grafo completo se torne desconexo a través de la eliminación de vértices sería eliminándolos todos.
- Un grafo vacío compuesto por k vértices, denotado por I_k consiste en un conjunto de vértices de tamaño k , donde no hay pares x,y adyacentes.

2.1.5.2 Redes regulares

Diremos que un grafo es regular, cuando todos los vértices tienen el mismo grado. Un grafo con vértices de grado k , se denomina k -regular (Figura 10).

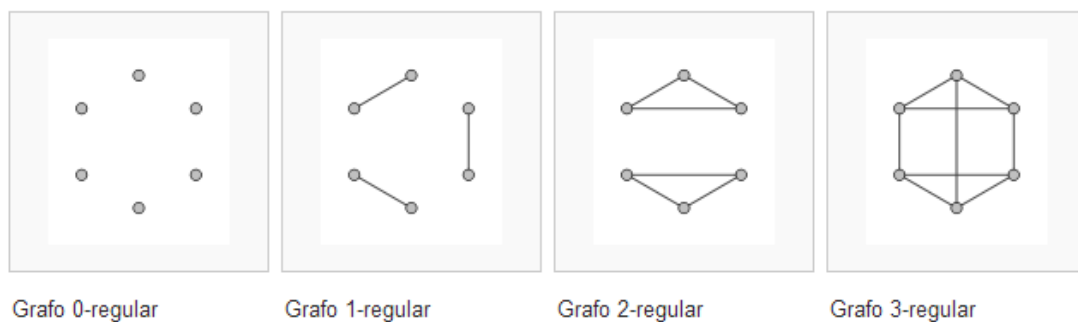


Figura 10: Ejemplo de grafos regulares (Wikipedia)

Una red regular no aleatoria es aquella donde cada nodo se conecta a todos sus vecinos más próximos. Las redes regulares o de rejilla (por su nombre *lattice* en inglés) también pueden ser representadas como anillos, árboles y estrellas. La red puede caracterizar enjambres, escuelas, rebaños, etc. en los que el comportamiento de cada individuo (nodo) depende del comportamiento de sus vecinos más cercanos.

Por ejemplo, una red de rejilla representada como un cuadrado, se muestra en la Figura 11:

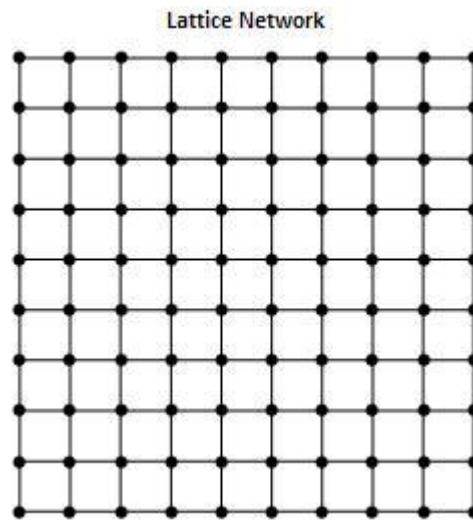


Figura 11: Red regular con representación cuadrada

Esta red (Figura 11) tiene una topología en la que cada nodo está vinculado a la totalidad de sus cuatro vecinos más cercanos. No se necesita ninguna regla estadística para definir el grado de distribución de la red debido a que el número de grados (4) es el mismo para cada nodo.

Si estudiamos la topología en anillo de la Figura 12, es exactamente el mismo modelo no aleatorio y regular que el de la red cuadrada anterior. Simplemente se adaptó a la forma circular doblando la red cuadrada y uniendo sus extremos. Tenemos que tener en cuenta que, al igual que su representación cuadrada, cada nodo se agrupa formando un clúster con los cuatro nodos vecinos.

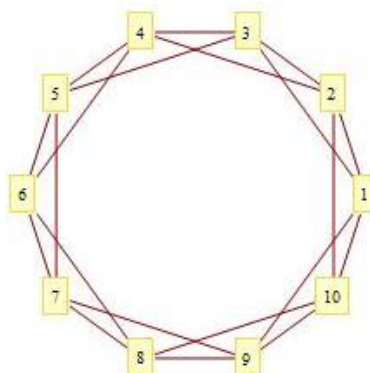


Figura 12: Red regular con representación en anillo

Por ejemplo, los nodos 1, 2, 3, 10, y 9 forman un grupo, ya que todos se conectan al nodo 1. La forma circular hace que sea fácil demostrar los efectos de los caminos con distancias cortas y la agrupación, *clustering*, que son características propias de las redes regulares.

Dos nodos situados en lados opuestos de la Figura 12 (ver nodos 1 y 6) son miembros de *clústeres* separados que no se intersectan. Además hay caminos relativamente largos (baja conectividad) entre estos dos nodos.

Dos nodos que están a una distancia de tres o menos nodos el uno del otro en la Figura 12 (nodos 1 y 4) pertenecen a *clústeres* distintos, pero pueden intersectarse. Esto es equivalente a dos grupos de amigos donde hay algunas personas pertenecientes a ambos grupos. Este tipo de intersección entre *clústeres* mejora la conectividad dentro de la red.

Una arquitectura de red regular, tal como la de rejilla, no ofrece una alta conectividad debido a que habrá un camino largo para llegar a muchos de los nodos.

En términos prácticos, por ejemplo, la red de rejilla circular regular puede producir una cadena de reacciones químicas para transformar el nodo 5 a un compuesto químico en el nodo 10.

2.1.5.3 Redes de mundo pequeño

Los grafos pequeño mundo son un tipo especial donde cohabitan propiedades tanto del mundo ordenado como del completamente aleatorio. Este concepto fue introducido en el año 1998 por los investigadores Duncan Watts y Steven Strogatz en su artículo en la revista Nature titulado Collective dynamics of 'small-world' networks.

Este tipo de redes tienen simultáneamente las propiedades de alta conectividad en vértices cercanos (grafos regulares) y distancia pequeña entre vértices alejados (grafos aleatorios), lo que permite una alta eficiencia en el flujo de información entre nodos.

Para definir su modelo, Watts y Strogatz definieron dos parámetros: el coeficiente de agrupamiento (*clustering coefficient*), que indica cómo de 'acumulados' están los nodos de la red, y la distancia media entre vértices, que cuantifica, en media, la distancia entre dos vértices cualesquiera.

Utilizando estos conceptos propusieron un modelo de redes de pequeño mundo cuyas características más sobresalientes son:

- Un trayecto mínimo promedio entre vértices pequeño. Es decir, que cualesquiera dos vértices de la red se comunican por un camino de vértices intermedios relativamente pequeño.
- Un coeficiente de agrupamiento medio elevado: si dos vértices o nodos no están conectados directamente entre sí, entonces existe una gran probabilidad de que se conecten mediante la intervención de otros nodos.

La primera descripción del modelo de Watts y Strogatz mostró que había una graduación entre lo que se puede denominar un mundo grande (un grafo regular) y un grafo aleatorio (conexiones desordenadas entre los vértices); entre estos dos extremos se hallaban las redes de mundo pequeño.

Se han encontrado ejemplos de redes de pequeño mundo en la naturaleza, desde la que define las proteínas que interaccionan en el metabolismo de las bacterias hasta las redes neuronales.

En las últimas décadas se ha observado que la estructura de muchas redes reales (redes sociales, tráfico aéreo, redes neuronales) se puede clasificar de manera efectiva usando los parámetros introducidos por Watts y Strogatz en su estudio.

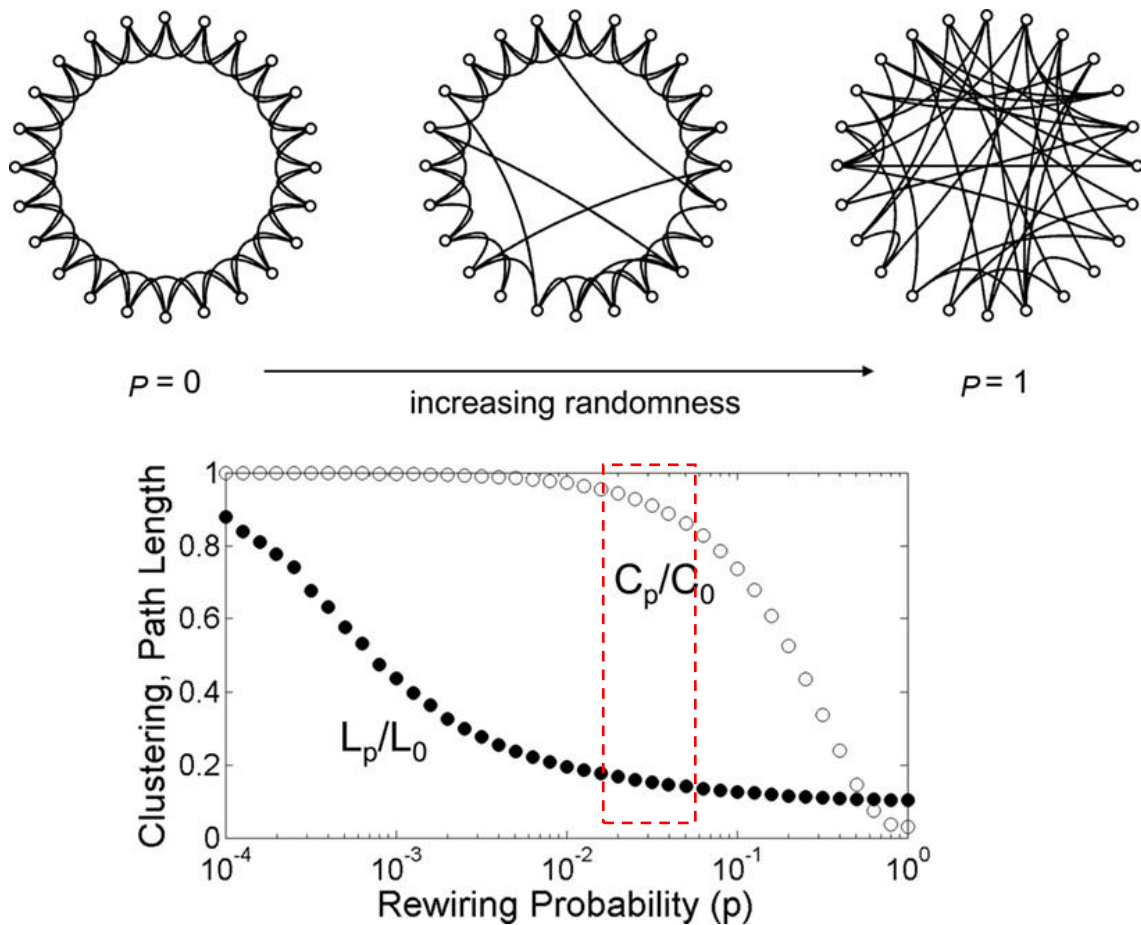


Figura 13: Grafos de pequeño mundo, situados entre regulares y aleatorios

En la Figura 13 se representa en el eje de ordenadas el grado de conectividad normalizado junto con la distancia entre nodos normalizada, y en el eje de abscisas podemos ver la probabilidad p . A medida que se incrementa la distancia y la conectividad se modifican, habiendo una zona remarcada con un rectángulo discontinuo en la que tenemos una alta conectividad y una distancia pequeña entre nodos. Esa área sería la zona en la que quedan definidas las redes de mundo pequeño.

2.1.5.4 Redes de escala libre

En una red de escala libre, algunos nodos están altamente conectados, es decir, poseen un gran número de enlaces a otros nodos, aunque el grado de conexión de casi todos los nodos es bastante bajo.

Por ejemplo podemos ver en la Figura 14 las aerolíneas en Estados Unidos, donde los aeropuertos principales, estarían altamente conectados y el resto presentan un grado de distribución muy bajo.

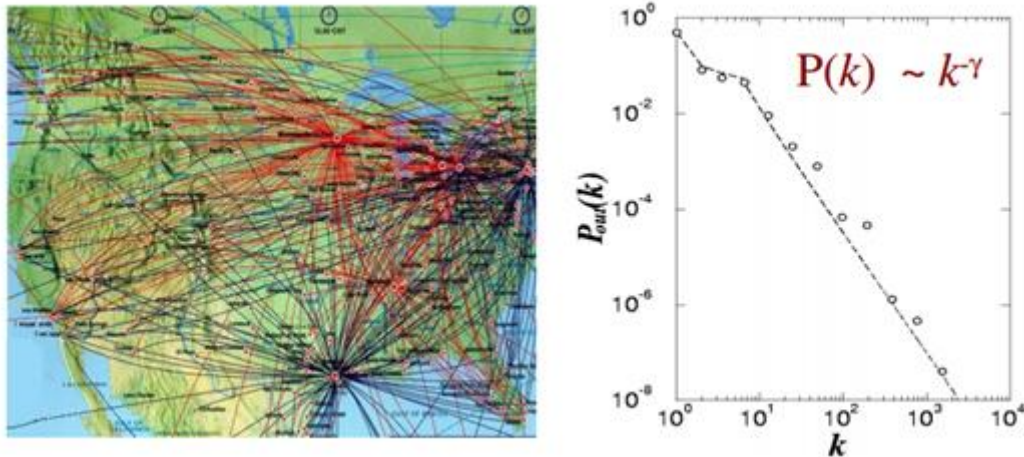


Figura 14: Distribución de probabilidad de aeropuertos y líneas aéreas en USA

Barabási y Albert-Laszlo (1999b) estudiaron un modelo de red altamente conectado que llegó a ser conocido como la red "libre de escala".

En comparación con una red aleatoria, la red de escala libre tiene un tipo muy diferente de conectividad debido a que el grado de distribución se define por una distribución que sigue una ley de potencia en lugar de la distribución de Poisson asociada a la red aleatoria (Figura 14, derecha). La ley de potencia es un miembro de la familia de distribuciones sesgadas hacia los valores extremos, de tal forma que describe eventos en los cuales una variable aleatoria alcanza valores altos con poca frecuencia, mientras que los valores medianos o bajos son mucho más comunes (Newman 2005).

En una distribución de ley de potencia la mayoría de los nodos tienen relativamente pocas conexiones, pero unos pocos nodos (llamados *hubs*) tienen un alto número de enlaces. La contribución a la conectividad de los *hubs* en general es muy alta. La contribución a la conectividad de los nodos con un menor número de enlaces es mucho menor. Un ejemplo de este modelo sería Internet donde la mayoría de las páginas web tienen sólo unos pocos enlaces que conectan con ellos, pero sitios como Google y Yahoo tienen un gran número de enlaces que apuntan a ellos.

Una aplicación curiosa de este tipo de redes es la de colaboración de actores, donde extrayendo la información del sitio web IMDB, podemos encontrar cómo se relacionan los actores entre sí, entendiendo como relación el que hayan aparecido en la misma película.

Con un número de nodos $N= 212,250$ y una conectividad promedio $\langle k \rangle$ de 28.78, un valor de $P(k) \sim k^{-\gamma}$ donde la escala $\gamma = 2.3 \pm 0.1$ obtenemos la siguiente gráfica (Barabási et al., 1999b):

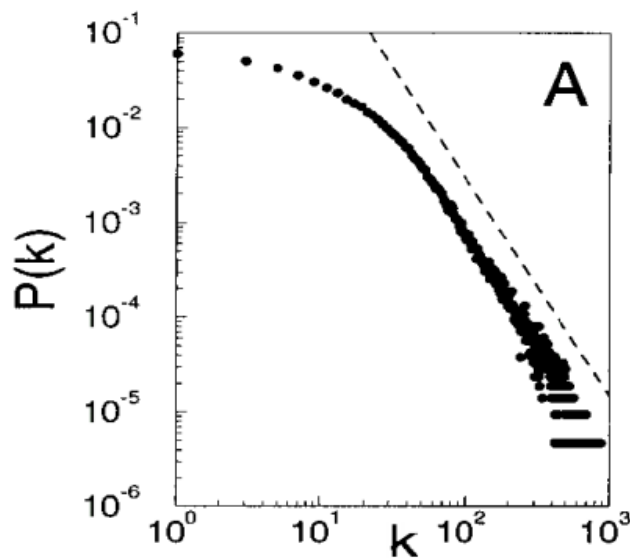


Figura 15: distribución de probabilidad de actores y sus relaciones

En la figura 15 podemos apreciar que hay un pequeño número de actores altamente relacionados con otros, y muchos actores poco relacionados con el resto.

2.1.5.5 Redes aleatorias

Se denomina red aleatoria a un grafo que es generado por algún tipo de proceso aleatorio. La teoría de los grafos aleatorios cae en la intersección entre la teoría de grafos y la teoría de probabilidades y se fundamenta en el estudio de ciertas propiedades de los grafos aleatorios.

Su origen se remonta a la década de los 50 y culmina en 1960 con el trabajo pionero *On the evolution of random graphs* de Erdős y Rényi. Más actual, es el libro *Random Graphs* de Bollobás (2001) que es el referente principal en este tipo de grafos. Erdős y Rényi propusieron un modelo de grafos aleatorios. En este modelo

cada par de los n vértices de un grafo tiene una probabilidad p de estar conectado por una arista, siendo esta probabilidad independiente de la existencia o no de otras aristas con otros nodos.

La idea subyacente es sencilla si ponemos un símil con botones. Imaginemos un conjunto de N botones de pantalón distribuidos aleatoriamente sobre una mesa e inicialmente desconectados. En $t = 0$ escogemos aleatoriamente una pareja de botones y los hilvanamos con un hilo. Después de haber enlazado a esta pareja, la dejamos sobre la mesa y escogemos aleatoriamente otra pareja para hilvanar. Podemos escoger botones que están conectados con otros botones, pero si la pareja que escogemos ya está conectada entre sí, la descartamos y escogemos otra pareja. Lo que no está permitido es hilvanar más de una vez a la misma pareja de botones. Repetimos este proceso sucesivamente M veces, escogiendo aleatoriamente una pareja de botones cada vez. Al final del proceso habremos establecido M enlaces entre M parejas diferentes de botones, generando así una red de botones (Figura 16).

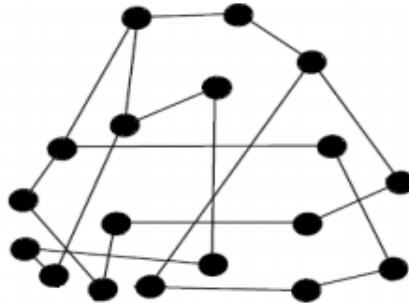


Figura 16: Estructura de una red generada por el modelo de Erdős y Rényi

Intuitivamente podemos calcular que si M (el número total de enlaces) es pequeño comparado con N (el número total de botones), entonces la red resultante estará desmembrada en varios subgrafos o redes pequeñas. Dentro de cada subred aislada los botones estarán hilvanados entre sí, pero estarán desconectados de las otras subredes. Sin embargo, si M es muy grande comparado con N , terminaremos con casi todos los botones hilvanados unos con otros. Probablemente haya subredes muy pequeñas desconectadas de la red principal, pero seguramente la gran mayoría de botones formarán parte de una red principal.

Después de haber hilvanado M parejas en un conjunto total de N botones, ¿cuál es la distribución de conexiones $P(k)$ en la red resultante? Como veremos en un momento, la red que resulta de este proceso tiene una distribución $P(k)$ de Poisson.

Calculemos ahora la probabilidad $P(k)$ para nuestra red de botones hilvanados. Para comenzar notemos que el número total N_p de parejas que se pueden formar en un conjunto de N botones es

$$N_p = \frac{1}{2}N(N - 1) \quad (15)$$

Como enlazamos M parejas de botones, la probabilidad p_e de que una pareja arbitraria seleccionada al azar esté enlazada es

$$p_e = \frac{M}{N_p} = \frac{2M}{N(N - 1)} \quad (16)$$

Ahora enfoquemos nuestra atención sobre un nodo particular v_j de la red, escogido al azar. El número total de parejas que podrían contener a v_j es $N-1$, ya que v_j se podría haber hilvanado con los $N-1$ nodos restantes de la red. Sin embargo, en los M enlaces que se llevaron a cabo, no necesariamente escogimos al nodo v_j todas las veces posibles que se podría haber escogido.

Supongamos entonces que de las M parejas que se escogieron, el nodo v_j estaba solamente en k de ellas. La probabilidad de que v_j esté contenido en k parejas de las $N-1$ posibles es

$$P(k) = \binom{N-1}{k} (p_e)^k (1 - p_e)^{N-1} \quad (17)$$

Esta es una distribución binomial para N y M finitas. Pero si consideramos ahora que la red es muy grande y tomamos el límite $N \rightarrow \infty$ y $M \rightarrow \infty$ de tal forma que la cantidad

$$z = \frac{2M}{N} \quad (18)$$

permanezca finita, entonces la distribución se transforma en

$$P(k) = e^{-z} \frac{z^k}{k!} \quad (19)$$

lo que resulta ser la distribución de Poisson con promedio z (Figura 17).

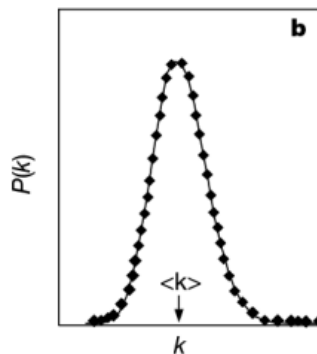


Figura 17: Distribución de Poisson de conectividad de una red aleatoria

Cualquier aleatoriedad en la red ampliará la forma de este pico, y en el caso límite de una red completamente aleatoria, la distribución de grados sigue una distribución de Poisson, donde la forma de la distribución cae de manera exponencial a medida que nos

2.1.5.5.1 *Uso de redes aleatorias en modelos*

Hoy en día los grafos aleatorios sirven como modelo en varios campos de investigación y son muy utilizados en el estudio de redes sociales, biológicas, redes de comunicación, física estadística etc.

En el análisis de sistemas complejos, las redes aleatorias además proporcionan una alternativa al enfoque tradicional llevado a cabo mediante ecuaciones diferenciales.

Existen trabajos basados en modelos de redes aleatorias, que han pasado a ser importantes en el análisis de sistemas complejos, por ejemplo en la biología evolutiva (Proulx et al., 2005).

El artículo de Proulx, *Network thinking in ecology and evolution*, empieza hablando de la propagación del síndrome de inmunodeficiencia en los 70 como introducción a

los modelos de redes. Su estudio versa sobre las redes de moléculas, las interacciones metabólicas, regulación de genes e interacciones entre proteínas.

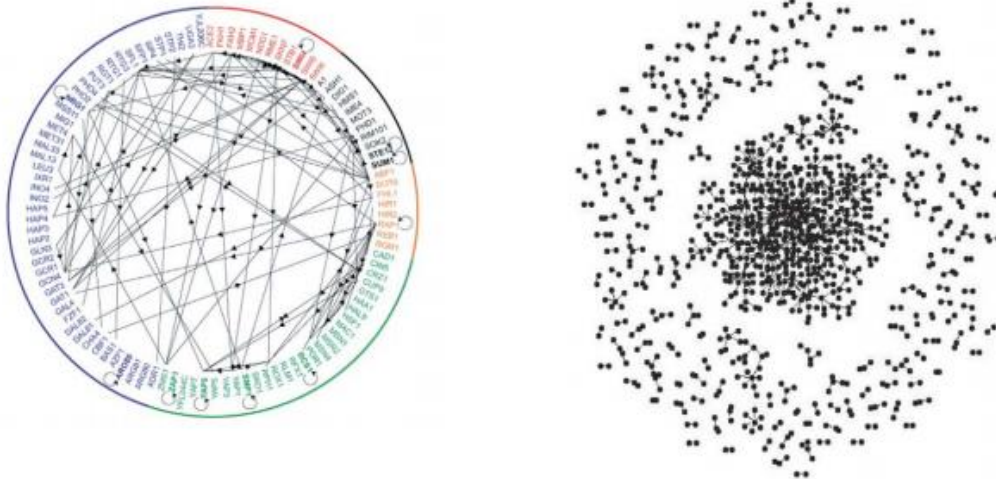


Figura 18: Red genética y red de proteínas respectivamente

A medida que los ecologistas y los biólogos evolutivos desarrollan nuevos enfoques de estos problemas. Los enfoques basados en redes, utilizando datos acumulados, podrían servir para revelar los patrones globales detrás a gran escala los procesos ecológicos y evolutivos.

Otro área que va en línea con el trasfondo de nuestro trabajo, es el de la epidemiología, donde los modelos de redes han servido por ejemplo para debatir sobre si el control de la viruela ha de realizarse mediante campañas de vacunación masivas o dirigidas (Halloran et al., 2002) (Koopman, 2002).

El artículo *Containing Bioterrorist Smallpox* de Halloran describe un modelo estocástico del contagio de la viruela en comunidades de gente que interactúa en grupos conocidos de contacto. Para cada simulación, se genera de manera aleatoria una comunidad de 2000 personas basándose en los datos censales de los hogares estadounidenses en el año 2000. Cada comunidad tenía cuatro vecindades, una universidad, una escuela de educación primaria y secundaria, y dos guarderías. Las probabilidades de transmisión persona a persona, son distintas dentro de cada grupo conocido de contacto. Cada día las personas susceptibles tienen una probabilidad de quedar contagiadas basándose en su estado de vacunación, quién era infeccioso en

su grupo de contacto y sus estados de vacunación, junto con la probabilidad de transmisión específica de ese grupo.

La tabla siguiente muestra la probabilidad de contagio por grupos de edad y grupos de contacto.

Contact group	Children					Adults	
	Pre-School		School			Younger	Older
	Small play group	Large daycare	Elementary	Middle	High		
Small play groups	0.09	-	-	-	-	-	-
Large daycare centers	-	0.05	-	-	-	-	-
Elementary School	-	-	0.006	-	-	-	-
Middle School	-	-	-	0.005	-	-	-
High School	-	-	-	-	0.002	-	-
Family							
child	0.12	0.12	0.12	0.12	0.12	0.05	0.05
adult	0.05	0.05	0.05	0.05	0.05	0.06	0.06
Neighborhood	0.00004	0.00004	0.00005	0.00005	0.00005	0.00014	0.00014
Community	0.00004	0.00004	0.00005	0.00005	0.00005	0.00014	0.00014

Figura 19: Probabilidad de contagio

En 200 simulaciones, la viruela fue introducida por uno o cinco adultos no vacunados en cada comunidad de 2.000 personas.

El simulador recoge el tiempo de generación típico en epidemias de la viruela (14 a 16 días). Empíricamente exploran el número básico reproductivo R_0 : el número promedio de nuevos individuos infectados que producirá en una población un infectado inicial donde los individuos son completamente susceptibles.

Infectado x R_0 = nuevos infectados

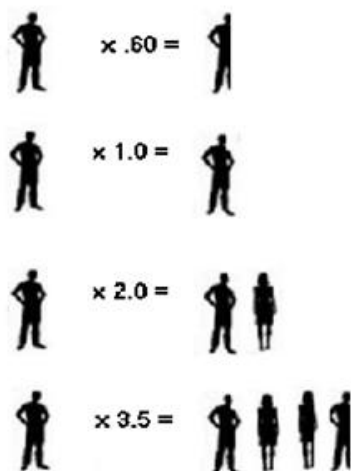


Figura 20: Cálculo del número R

En la figura 20 podemos entender cómo actúa el número reproductivo R_0 .

Los autores invitan a continuar con su modelo, con posteriores calibraciones cambiando tasas de contagio a través del mayor o menor conocimiento de las comunidades establecidas en el modelo. Llegando en un caso a concluir que la vacunación dirigida (*targeted*) es más óptima si tenemos conocimiento objetivo de las comunidades involucradas en la epidemia o por el contrario vacunación en masa cuando no tenemos un conocimiento de las comunidades que componen nuestra población.

En muchas situaciones, únicamente la topología de la red es el factor relevante para el modelo de transmisión de la epidemia (Shirley et al., 2005). Modelos más refinados que usan las redes aleatorias como elemento principal, son igualmente considerados (Witten y Poulter, 2007) (Acedo et al., 2011) (Barret et al., 2008), donde los autores describen algoritmos para la simulación del modelo de transmisión de las epidemias en unas redes de tamaño considerable.

Estos modelos de redes aleatorias se han validado con datos reales y se han utilizado para estudiar pandemias sociales como el consumo de tabaco (Christakis, 2008) o la obesidad (Christakis, 2007).

Christakis, en su artículo sobre el descenso del consumo de tabaco en Estados Unidos, se basa en un modelo de redes aleatorias, donde trata de estudiar en qué medida afecta volverse fumador cuando otra persona es fumadora y en qué medida los grupos conectados de personas dejan de fumar de manera conjunta.

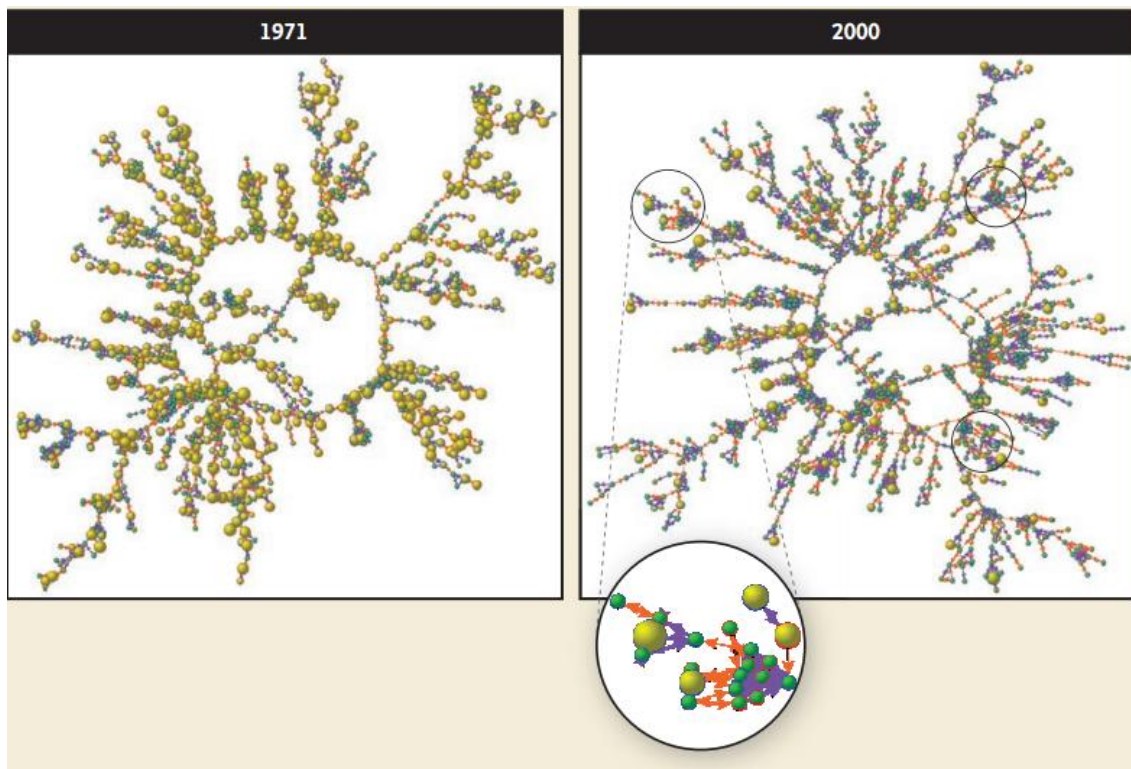


Figura 21: Red aleatoria para modelar el consumo de tabaco (Christakis, 2008)

El estudio que trata la evolución desde 1971 hasta el año 2000 partiendo de datos de la *Framingham Heart Study* (un proyecto de la universidad de Boston), nos hace ver en la Figura 21 una muestra aleatoria de 1000 sujetos de la red social del *Framingham Heart Study* escogida por ser la subred más extensa, donde en la izquierda tenemos una instantánea, una de 1971 y otra del año 2000.

Cada círculo (nodo) representa una persona. Los círculos con los bordes rojos representa mujeres, y los círculos con los bordes azules representan hombres. El interior del nodo de los círculos representa cuanto fuman usando una gama de colores indicando que, por ejemplo, amarillo es más de uno al día, mientras que verdes significa que no fuman. Los colores de las aristas nos dicen como es la relación entre dos sujetos: naranja denota amistad o matrimonio y morado denota relación familiar. En el año 2000, los fumadores solían aparecer con más probabilidad en la periferia de la red. Además, los fumadores se suelen agrupar en pequeños grupos al contrario que los no fumadores. Las áreas localizadas mediante un círculo negro en la parte del año 2000, identifican clústeres altamente conectados de nodos verdes (no fumadores) donde además los que son fumadores se establecen por el borde.

Otra parte del estudio que resulta interesante comentar es acerca de los clústeres y la información que podemos obtener de ellos. La siguiente figura 22 muestra como el tamaño medio de clúster de fumadores totalmente conectado cambia con el tiempo. También muestra el tamaño medio que habríamos observado en esos clústeres si hubiésemos preservado la topología de la red y la prevalencia de los fumadores reasignando sujetos a fumadores o no fumadores de manera aleatoria.

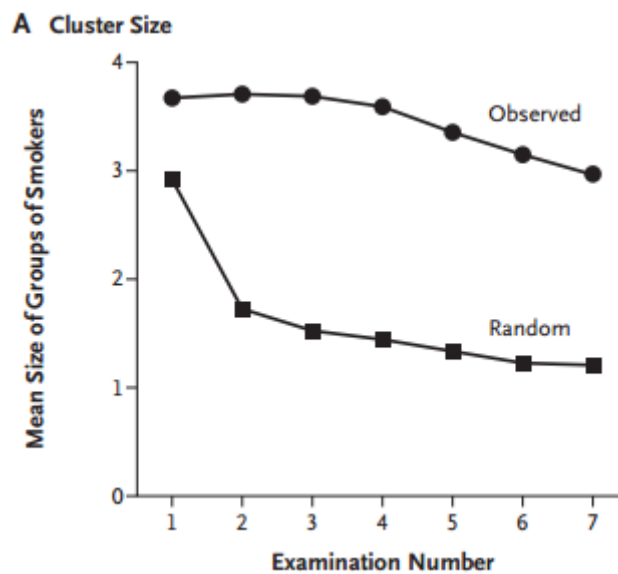


Figura 22: Tamaño del clúster con el paso del tiempo (Christakis, 2008)

Estos datos proporcionan evidencia que sugiere que los que dejan de fumar no lo hacen paulatinamente en los márgenes de los clústeres de fumadores. En cambio, parece que grupos completos de fumadores se vuelven no fumadores, haciendo que el tamaño promedio del grupo se mantenga alto, a pesar de la disminución general de prevalencia.

Todos estos artículos de los cuales se han comentado algunos, prueban el uso extendido de modelos de redes y en particular modelos de redes aleatorias entre los investigadores y en particular sobre los que trata el trabajo: en epidemiología.

2.2. Modelos epidemiológicos

El modelado de epidemias consiste en el uso del lenguaje y herramientas matemáticas para explicar y predecir el comportamiento de agentes infecciosos y potencialmente dañinos a poblaciones humanas o animales, llamándose de manera

genérica «epidemiológicos» (Hethcote, 1989). Estos modelos se usan como instrumento para evaluar cómo evolucionan las enfermedades (u otros eventos asimilables de tipo transmisible) en el tiempo y cómo se comportan ante estímulos del entorno (cambios en la población, campañas de vacunación, etc.). Es esto precisamente, en su capacidad de aplicar hipótesis a los modelos y extrapolar el resultado a la vida real lo que los hace tan interesantes. Permiten a los responsables de la Salud Pública u otras autoridades el tomar decisiones sabiendo de antemano qué resultados se van a obtener y qué impacto tendrán las estrategias de control llegado el caso.

Por supuesto el modelado de un problema implica que dicho modelo debe ajustarse al propósito buscado y tiene que poder ser verificado y validado de forma apropiada. Dicha verificación es a menudo uno de los aspectos más delicados del proceso, y depende de los datos obtenidos con estudios experimentales y la experiencia práctica.

Existen algunos conceptos básicos que se manejan en la literatura sobre el modelado epidemiológico.

- Número R_0 : Número medio de sujetos infectados que aparecen en el momento en el que un primer individuo infectado (paciente cero) entra en contacto con la población de individuos susceptibles (Diekmann et al., 1990). Este número define de manera general cómo va a evolucionar el proceso de contagio.
 - Si $R_0 > 1$ los individuos infecciosos al inicio de la enfermedad incrementarán el número de nuevas infecciones, es decir, se producirá un brote epidémico.
 - $R_0 < 1$ los individuos enfermos serán, como población, transmisores poco eficientes y la enfermedad acabará por desaparecer.
- Tamaño de la epidemia: es una de las propiedades asintóticas más importantes en los modelos epidemiológicos. Se define como el número total de individuos infectados en una epidemia (Dietz, 1967). Es una medida cuantitativa muy importante porque se relaciona de forma estrecha con los costes de la epidemia.
- Intervalo de serie: es el tiempo transcurrido entre la aparición de la enfermedad en un caso y el siguiente.

- **Incidencia:** es el número de casos nuevos de una enfermedad, síntomas, muertes o lesiones que se presentan durante un periodo de tiempo específico. La incidencia muestra la probabilidad de que una persona en esa población resulte afectada por la enfermedad.
- **Infectividad:** Se denomina infectividad en Epidemiología a la capacidad de un agente patógeno (bacteria, virus...) para invadir un organismo y provocar en él una infección. La infectividad viene dada por la expresión $(N^{\circ} \text{ infectados} / N^{\circ} \text{ susceptibles}) \times 100$.

Una vez vistos algunos conceptos básicos, pasaremos a describir varios modelos matemáticos que se han utilizado para estudiar la propagación de enfermedades en el ámbito de la epidemiología.

2.2.1. Modelos epidemiológicos continuos

La dinámica de transmisión de las enfermedades infecciosas ha sido tradicionalmente modelizada con ecuaciones diferenciales. Las ecuaciones diferenciales son una herramienta potente y bien conocida para el análisis de sistemas dinámicos desde que aparecieron en los trabajos de Kermack y McKendrick (Kermack, McKendrick, 1927), Eldestein-Keshet (Eldestein-Keshet, 1988), y Murray (Murray, 2002).

Kermack y McKendrick presentaron un modelo epidemiológico definido por ecuaciones diferenciales basadas en la edad de la infección, donde se establece una población agrupada por compartimentos en los que las personas pueden ser susceptibles a la epidemia o recuperadas de la misma.

En estos modelos se parte del supuesto de que los individuos se encuentran en uno de varios estados posibles. En 1927, W.O. Kermack y A.G. McKendrick definieron los siguientes estados:

- individuos susceptibles (S)
- infectados (I)
- recuperados (R)

y bautizaron a su modelo como SIR (Susceptible – Infectado – Recuperado) indicando la forma en la que los individuos pasaban de un estado a otro.

Con el tiempo surgieron otros modelos que seguían esta nomenclatura, siendo los más importantes los, junto con SIR, los siguientes:

- SI (Susceptible – Infectado)
- SIS (Susceptible – Infectado – Susceptible)
- SIR (Susceptible – Infectado – Recuperado)

Pueden modelarse en forma determinista o estocástica. En estos modelos se asume que la población es homogénea (todo la gente es igual) y la única distinción se encuentra en su pertenencia a una u otra categoría, además de que la interacción entre individuos de la población es aleatoria (Anderson, 1992).

2.2.1.1 Modelo SI

Cuando tenemos una enfermedad con un patógeno infeccioso, el modelo susceptible-infectado (SI) divide a la población en dos grupos: sujetos susceptibles (que pueden ser infectados por el agente patógeno pero no pueden infectar), S , y los sujetos infecciosos (que son los que están infectados con el patógeno) I . Si aceptamos por válida la ley de acción de masas para el modelo, la tasa mediante la cual un sujeto susceptible pasa a ser infectado es el producto del número de contactos que el sujeto tiene por unidad de tiempo, r la probabilidad de transmisión de la infección por contacto es β , y la proporción de sujetos de la población que son infecciosos es I/N , donde $N = S + I$ es el tamaño total de la población (Figura 23). En éste modelo los sujetos infectados permanecen infectados de por vida.

Una de las enfermedades que representa este modelo es el patógeno causante del síndrome de inmunodeficiencia humana (VIH), donde no hay sujetos recuperados. A continuación en la figura 23 mostramos un esquema.

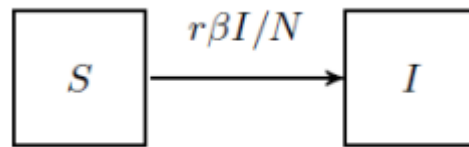


Figura 23: Esquema del modelo SI (susceptible-infeccioso)

La siguiente ecuación presenta el modelo SI en su versión determinística, es decir, como un modelo continuo en su forma más simple, consistente en un sistema de dos ecuaciones diferenciales:

$$\frac{dS}{dt} = -r\beta S(t) \frac{I(t)}{N} \quad (20)$$

$$\frac{dI}{dt} = r\beta S(t) \frac{I(t)}{N} \quad (21)$$

Dado que asumimos que el tamaño de la población es fijo, podemos reducir el sistema a una dimensión con la sustitución $S(t) = N - I(t)$ para obtener la ecuación logística:

$$\frac{dI}{dt} = r\beta(N - I(t)) \frac{I(t)}{N} \quad (22)$$

Se puede resolver analíticamente con la condición inicial $I(0) = I_0$ obteniendo que,

$$I(t) = \frac{I_0 N}{(N - I_0)e^{r\beta t} + I_0} \quad (23)$$

La Figura 24 muestra una simulación del modelo SI.

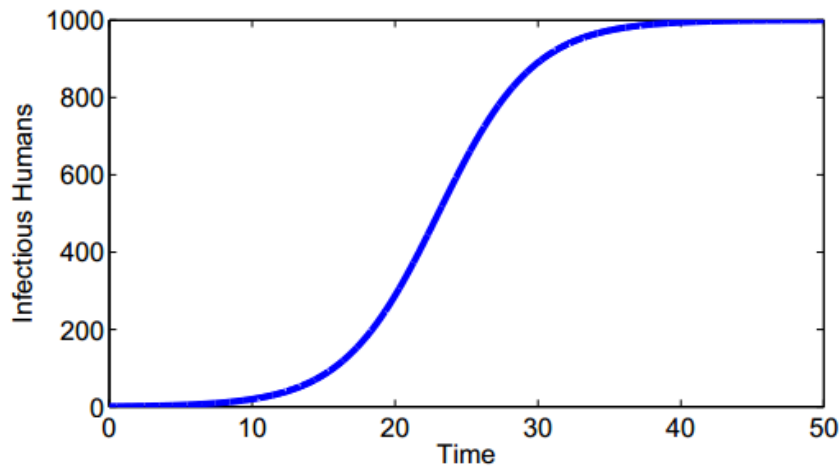


Figura 24: Simulación numérica del modelo SI (Chitnis, 2011)

En la Figura 24 hemos simulado con $r = 6$ y $\beta = 0.05$; y unas condiciones iniciales de $S(0) = 999$ e $I(0) = 1$.

2.2.1.2 Modelo SIS

El modelo susceptible-infectado-susceptible es similar al SI pero permite que los infectados se recuperen de la infección y vuelvan al estado de susceptible (donde pueden volver a ser infectados). Los sujetos infectados se recuperan a una tasa constante γ , de modo que $1/\gamma$ es la duración del periodo de infección con un tiempo de espera exponencial (Figura 25). Éste modelo sirve para infecciones bacterianas, o para infecciones con una evolución rápida tales como el resfriado común, donde la infección no proporciona inmunidad.

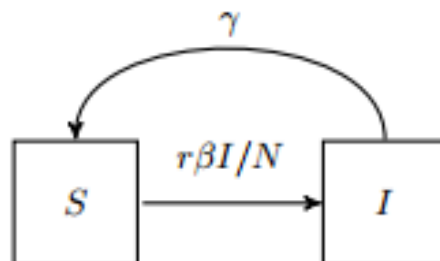


Figura 25: Esquema del modelo SIS (susceptible-infeccioso-susceptible)

Las ecuaciones del modelo SIS son muy parecidas, pero incluyen el parámetro γ .

$$\frac{dS}{dt} = -r\beta S \frac{I}{N} + \gamma I \quad (24)$$

$$\frac{dI}{dt} = r\beta S \frac{I}{N} - \gamma I \quad (25)$$

Como el tamaño de la población es fijo, de forma similar a como ocurre en el modelo SI, podemos reducir el sistema a una ecuación si sustituimos, $S = N - I$ y nos queda,

$$\frac{dI}{dt} = r\beta(N - I) \frac{I}{N} - \gamma I \quad (26)$$

Una simulación de las ecuaciones diferenciales puede verse en la Figura 26.

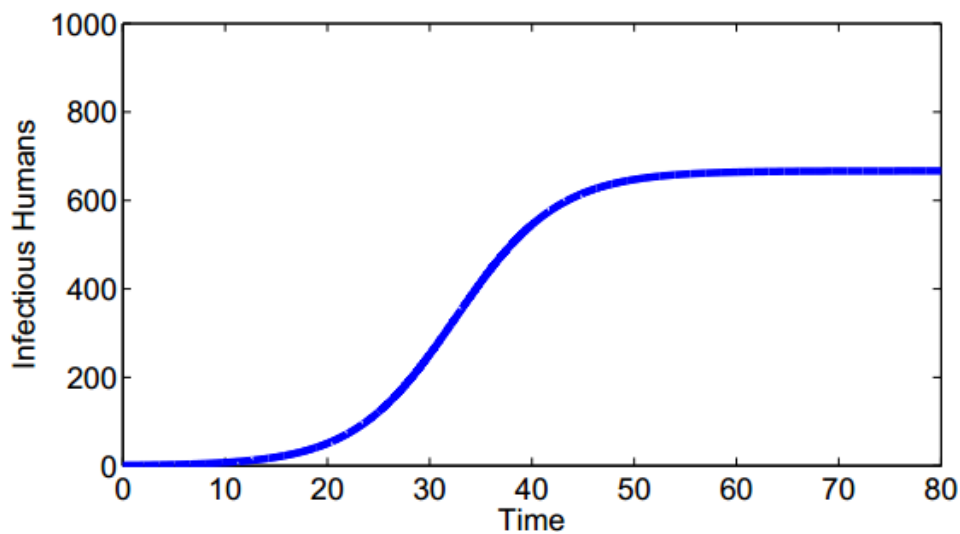


Figura 26: Una simulación numérica del modelo SIS. (Chitnis, 2011)

En la Figura 26 simulamos el modelo susceptible-infeccioso-susceptible con $r=6$, $\beta = 0.05$, $\gamma=0.1$; y las condiciones iniciales son $S(0)=999$ e $I(0)=1$.

2.2.1.3 Modelo SIR

Consideremos una enfermedad en donde la población queda dividida en tres clases diferentes: individuos susceptibles (S), que pueden contraer la enfermedad; los individuos infectados (I), que tienen la enfermedad y pueden transmitirla; y los individuos recuperados (R), aquellos que tuvieron la enfermedad y se recuperaron

volviéndose inmunes. En la evolución de la enfermedad los individuos quedan de manera esquemática representados por (Figura 27):

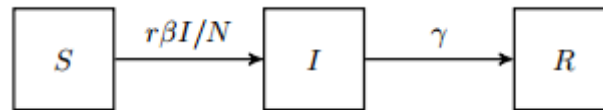


Figura 27: Diagrama de la dinámica de transferencia del modelo SIR

En su forma más simple puede formularse como un conjunto de ecuaciones diferenciales, tal y como se muestra a continuación.

$$\frac{dS}{dt} = -r\beta S(t) \frac{I(t)}{N} \quad (27)$$

$$\frac{dI}{dt} = r\beta S(t) \frac{I(t)}{N} - \gamma I(t) \quad (28)$$

$$\frac{dR}{dt} = \gamma I(t) \quad (29)$$

Si asumimos que el tamaño de la población es fijo, podemos reducir el sistema a otro con dos ecuaciones, con $R(t) = N - S(t) - I(t)$. Para centrarnos en la proporción de individuos que son infecciosos, $i(t)$ y susceptibles, $s(t)$, podemos normalizar las variables dividiendo por el tamaño de la población,

$$s(t) = \frac{S(t)}{N} \quad (30)$$

$$i(t) = \frac{I(t)}{N} \quad (31)$$

Con $r(t) = 1 - s(t) - i(t)$. Realizando algunos cálculos obtenemos que,

$$\frac{ds}{dt} = \frac{1}{N} \frac{dS}{dt} = -r\beta s(t)i(t), \quad (32)$$

$$\frac{di}{dt} = \frac{1}{N} \frac{dl}{dt} = r\beta s(t)i(t) - \gamma i(t), \quad (33)$$

El modelo SIR, al contrario que los otros modelos no tiene solución analítica, pero podemos llevar a cabo simulaciones numéricas y hallar un punto de equilibrio.

En la Figura 28 simulamos con unas condiciones iniciales de $S(0)=800$, $I(0)=1$, $R(0)=1990$ y con parámetros $r=6$, $\beta =0.05$, $\gamma=0.1$.

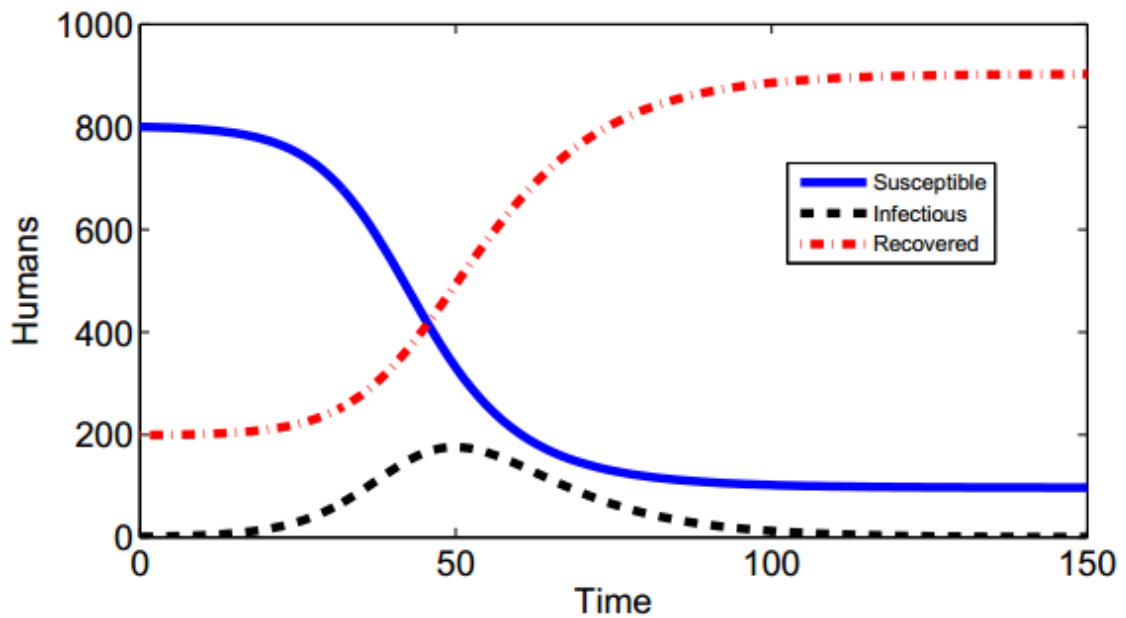


Figura 28: Simulación SIR (Chitnis, 2011)

En la Figura 29 simulamos con unas condiciones iniciales de $S(0)=900$, $I(0)=100$, $R(0)=0$ y con parámetros $r=6$, $\beta =0.05$, $\gamma=0.1$.

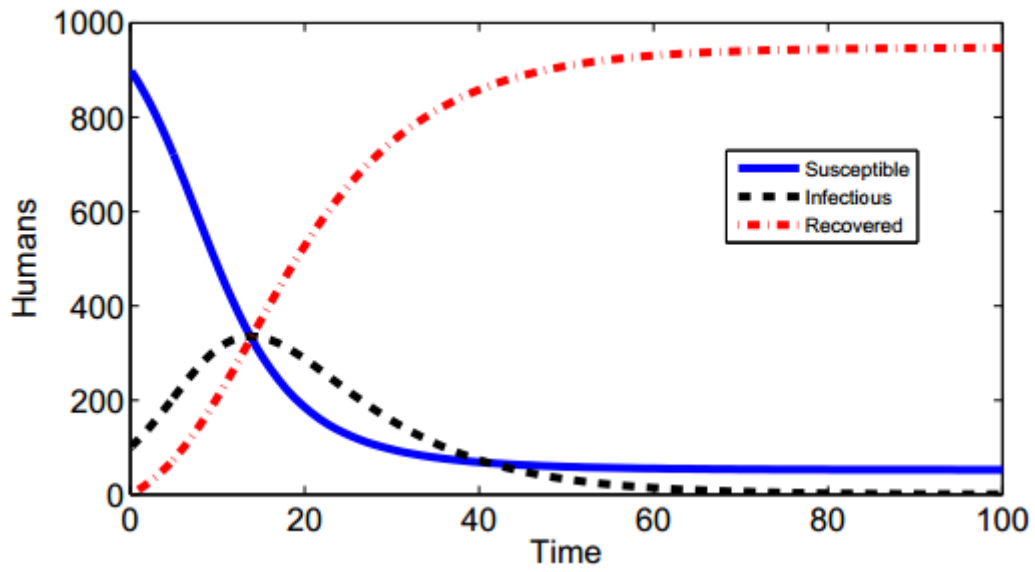


Figura 29: Simulación SIR

3. Propuesta de modelización mediante redes aleatorias

El interés por los modelos de redes procede de las dificultades que presenta el enfoque continuo para el tipo de fenómeno que nos ocupa. El modelo continuo no puede hacer distinciones entre individuos o manejar atributos particulares tales como la edad, enfermedades pasadas, vacunación, y otros parámetros que tienen repercusión sobre el modelo de transmisión de la epidemia bajo estudio haciéndolas muy difíciles de representar. Un estudio para la prevención de la onda epidémica de gripe A/H1/N1 (Donado-Campos et al., 2010) es una prueba de las complicaciones que pueden tener estos modelos. Trataron de predecir el número de casos de gripe A mediante un modelo dinámico con seis grupos de poblaciones. Los problemas que tuvieron se deben a la complejidad que añade comparar distinta información (la del modelo y de los médicos), falta de conocimiento de la historia natural (casos asintomáticos, patrón de contactos entre susceptibles, etc.), no se consideraron medidas de prevención primaria, secundaria o terciaria.

El objeto de nuestro estudio va a ser un modelo capaz de simular la transmisión del MenC. A la hora de elegir el tipo de modelo, optamos por uno de redes en lugar de un modelo continuo más convencional con el fin de evitar la limitación antes mencionada de usar únicamente compartimentos para agrupar todos los individuos en clases, y así ser capaces de procesar de manera individual y con propiedades específicas a cada individuo.

Dentro de los modelos de redes presentados en el capítulo anterior, analizamos los pros y contras de cada uno para decidir la mejor red.

- 1) Las redes de rejilla no las consideramos adecuadas porque una cantidad de relaciones entre usuarios fija e igual para todos los nodos no se ajusta a la naturaleza de las relaciones en las sociedades humanas. Parece poco realista fijar a priori el mismo número k para el caso que nos compete. Una persona que realiza su vida cotidiana y sale por ejemplo, a comprar el pan, se relacionará con un número variable de personas distinto al de otra persona que vaya a otra panadería o al colegio a recoger a sus hijos.

- 2) Las redes de escala libre (*scale-free*) se suelen usar para la propagación de virus informáticos y enfermedades de transmisión sexual. Pero como el número de individuos es dinámico y no creemos que existan grupos en lo que a la forma de transmitir el MenC se refiere, pensamos que las redes de escala libre no se asemejan a una enfermedad que se transmite compartiendo fluidos. Lo ideal sería tener una red donde los nodos se conectan por intimidad (adolescentes, familiares, parejas), pero no hay una categoría semejante por el momento.
- 3) Las redes de mundo pequeño presentan una alta conectividad y una distancia corta entre nodos. No obstante estas redes no han tenido una difusión y uso tan extendido como las aleatorias, ni se han llegado a estudios decisivos más allá del experimento de los 6 grados de conocimiento o los actores de Hollywood.

Por lo tanto parece que las mejores candidatas son las redes aleatorias para implementar nuestro modelo de propagación del MenC.

Para obtener la evolución a corto y medio plazo de una enfermedad necesitamos construir una red que represente una estructura de la población real. Además es necesario etiquetar cada nodo respecto al estado de la enfermedad (y otras propiedades como el sexo, edad, estado civil, si fuese necesario), y establecer reglas de evolución, es decir, cómo se transmite la enfermedad, cuánto tiempo dura un individuo infectado, etc. Luego se puede simular, evaluar la bondad del modelo y ver si se pueden tener en consideración las conclusiones obtenidas.

En las reglas que rigen la evolución de una enfermedad en una población se puede introducir el azar para modelar escenarios más realistas. De ahí las llamadas redes aleatorias.

A pesar de estas ventajas este enfoque conlleva la dificultad de determinar parámetros desconocidos del modelo de la enfermedad. Su estimación se suele realizar mediante técnicas que los generan en el espacio de soluciones de esos parámetros dentro de unos márgenes razonables, o si no se dispone de ningún tipo de información se toman unos valores cualesquiera. En el caso de las redes aleatorias el ajuste o calibración del modelo se realiza mediante un número elevado de simulaciones.

Finalmente evaluaremos si los parámetros generados dan la validez necesaria a esos modelos para que finalmente puedan resultar útiles.

Una vez establecido tanto el modelo de la red como las reglas que hacen evolucionar la enfermedad, es posible simular su evolución de la enfermedad en los nodos de la red a lo largo del tiempo y estudiar el modelo de transmisión entre la población.

Además, como una red aleatoria es un conjunto de nodos que representan individuos, hay etiquetas y propiedades que son asignadas a cada nodo tales como la edad, el sexo, y el estado con respecto a la enfermedad (susceptibilidad, infección, recuperación, latencia, etc.) que permiten obtener y controlar mucho mejor la información de la población, haciendo que se asemeje en mayor o menor grado a la población real si hay datos disponibles (por ejemplo acceso al registro de empadronamiento).

En nuestro modelo, siendo cada uno de los individuos de la población un nodo que caracteriza por unos atributos determinados, también sucede que se relacionan con otros individuos en mayor o menor medida, haciendo de este modo evolucionar a nuestra población.

3.1. Diseño de la red

La red que vamos a generar para representar a la población objeto de estudio se va a hacer siguiendo las siguientes consideraciones:

- 1) El conjunto de individuos estará estratificado siguiendo una pirámide poblacional que evolucionará con valores reales de mortalidad y natalidad tomados del Instituto Nacional de Estadística. Por tanto hemos de establecer:
 - a. La pirámide poblacional.
 - b. La mortalidad por edades.
 - c. El reparto de portadores por edades.
- 2) Una vez definidas sus características, la red se construirá como sigue:
 - 1) En primer lugar se generan a los individuos (nodos de la red).
 - 2) A continuación se establecen las relaciones entre los nodos al azar según el grado medio definido.

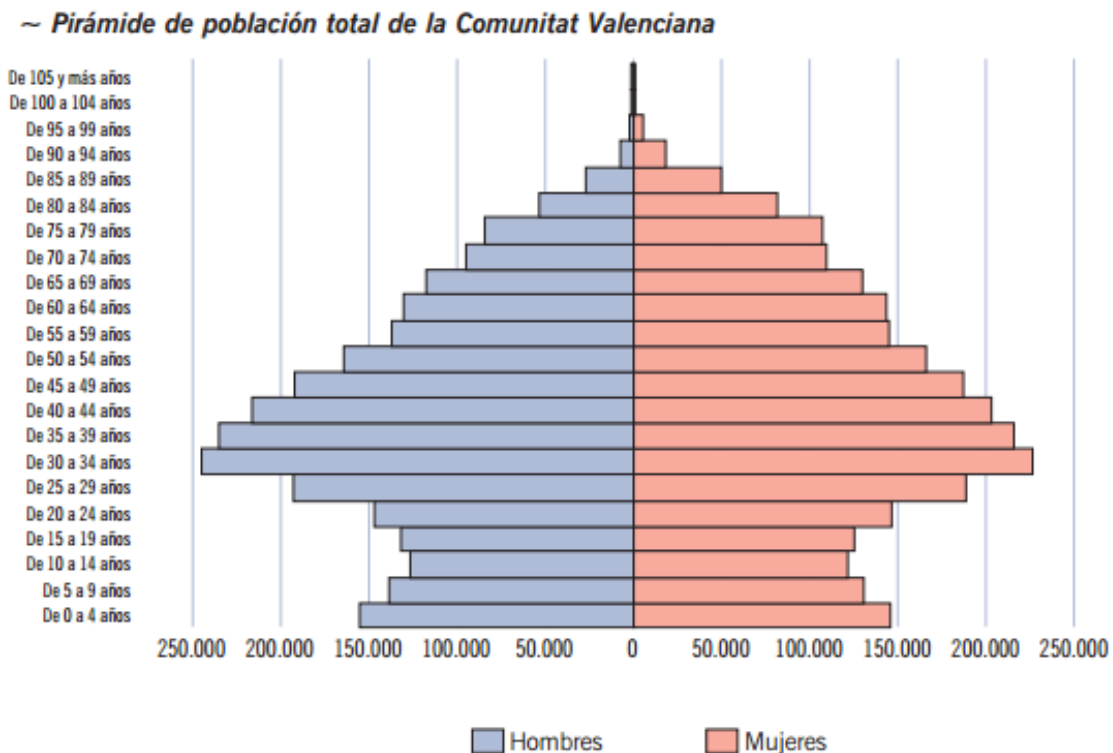
3) Por último se ha de definir el criterio de ajuste del modelo.

3.1.1. Asignación de edades

La red va a estar estructurada por edades, ya que es un hecho documentado (Trotter et al., 2006) que la incidencia y la transmisión del MenC depende en gran medida de esta variable. Concretamente los adolescentes son los que presentan una mayor incidencia y tasa de transmisión debido a su mayor cantidad de contactos sociales y relaciones.

Puesto que la edad es relevante, nuestra red en el instante de partida tendrá unos grupos de edad realistas acorde a las proporciones de la pirámide poblacional de la Comunidad Valenciana del año 2000 según la Consejería de Sanidad de la Comunidad Valenciana (www.san.gva.es, 2014) que se muestra en la Figura 30.

Población total de la Comunitat Valenciana



Fuente de datos SIP a fecha 1/1/2010

Figura 30: Pirámide de población de la Comunidad Valenciana

3.1.2. Evolución de la pirámide poblacional

Por simplificar, y dado que no tiene un impacto especial en la población, vamos a usar un modelo con población estática. Es decir, sustituiremos a cada persona fallecida por un recién nacido, de forma que la población será constante. Además hacemos evolucionar el modelo demográfico hasta alcanzar una cierta estabilidad por edades.

Por ello sólo necesitamos considerar la tasa de mortalidad, la cual toma en consideración los datos obtenidos de la tabla 5, donde se recoge la mortalidad en la Comunidad Valenciana desde 1997 para ambos sexos:

Units: Deaths per 1000 inhabitants

	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997
Comunitat Valenciana																
Both sexes																
0 years old	2,62	2,79	3,47	3,18	3,64	3,58	3,21	3,47	2,99	3,70	3,86	3,75	3,59	4,01	4,51	4,67
1 to 4 years old	0,15	0,16	0,20	0,19	0,17	0,27	0,23	0,28	0,21	0,25	0,27	0,20	0,21	0,26	0,20	0,38
5 to 9 years old	0,08	0,09	0,07	0,04	0,08	0,07	0,13	0,10	0,13	0,13	0,15	0,14	0,14	0,16	0,15	0,18
10 to 14 years old	0,10	0,11	0,13	0,09	0,09	0,15	0,09	0,15	0,13	0,08	0,18	0,15	0,15	0,12	0,13	0,20
15 to 19 years old	0,20	0,22	0,21	0,22	0,38	0,33	0,36	0,41	0,37	0,38	0,37	0,43	0,53	0,54	0,53	0,50
20 to 24 years old	0,20	0,24	0,30	0,34	0,30	0,34	0,43	0,49	0,43	0,54	0,59	0,52	0,61	0,65	0,60	0,54
25 to 29 years old	0,35	0,31	0,31	0,35	0,40	0,48	0,49	0,48	0,50	0,57	0,67	0,70	0,70	0,68	0,81	0,87
30 to 34 years old	0,49	0,46	0,48	0,45	0,60	0,61	0,63	0,66	0,66	0,79	0,99	0,94	0,94	0,96	0,96	1,32
35 to 39 years old	0,66	0,71	0,74	0,77	0,82	0,91	0,98	1,17	1,08	1,30	1,22	1,35	1,32	1,34	1,31	1,46
40 to 44 years old	1,17	1,22	1,31	1,27	1,45	1,60	1,50	1,80	1,72	1,81	1,77	1,78	1,91	1,91	1,77	1,91
45 to 49 years old	1,99	2,16	2,12	2,18	2,41	2,36	2,27	2,66	2,71	2,71	2,76	2,52	2,69	2,79	2,57	2,78
50 to 54 years old	3,27	3,26	3,31	3,45	3,59	3,56	3,64	3,64	3,69	3,86	4,03	3,94	4,02	4,06	3,67	4,00
55 to 59 years old	4,72	4,98	5,14	5,23	4,92	5,36	5,41	5,47	5,48	5,43	5,49	5,80	5,76	6,62	5,87	6,17
60 to 64 years old	7,25	7,17	6,89	7,21	7,52	7,81	7,93	8,10	8,05	8,87	8,65	9,09	9,23	9,51	9,30	9,84
65 to 69 years old	9,75	10,54	10,47	11,31	11,42	11,68	11,81	12,81	13,23	13,92	13,54	14,14	14,63	15,66	15,18	15,64
70 to 74 years old	17,20	17,55	17,92	18,17	19,20	19,79	19,46	21,10	21,33	23,47	23,50	24,70	24,83	26,46	26,10	27,18
75 to 79 years old	31,37	31,06	31,81	32,82	34,53	35,60	36,53	39,97	38,80	42,31	42,45	43,69	43,96	47,78	46,34	46,37
80 to 84 years old	60,40	61,18	60,48	64,05	64,27	67,28	67,28	74,40	72,32	77,32	75,98	76,04	75,79	83,26	83,11	84,53
85 to 89 years old	115,83	110,53	114,21	117,29	120,52	124,11	123,34	135,50	133,91	143,31	136,97	138,38	139,02	151,22	148,51	148,98
90 to 94 years old	200,96	189,19	198,78	203,35	211,24	213,74	217,74	229,51	226,88	251,59	235,71	222,17	229,49	248,41	249,17	246,08
95 years old and over	305,83	297,76	315,27	323,44	333,53	348,79	334,76	365,48	342,99	383,74	340,57	321,15	353,31	415,54	426,70	467,07

Tabla 5. Tasa de mortalidad (Comunitat Valenciana)

3.1.3. Reparto de portadores

Partimos de los datos conocidos de Trotter et al. (2006). Éstos se presentan en la Tabla 6, donde pueden apreciarse los grupos de edades y los porcentajes de portadores en agosto de 1997, en una situación pre-vacunal:

Grupo de edad	% portadores
0-9 años	0%
10-14 años	1%
15-19 años	1.6%
20-24 años	0.8%
25-29 años	0.2%
30-34 años	0%
35-39 años	0.8%
40-44 años	0.6%
45-49 años	0.8%
50-54 años	0.4%
55-59 años	0.8%
60-64 años	0.4%
65 años y más	0%

Tabla 6. Porcentaje inicial de portadores 1997, situación pre-vacunal

Nos vamos a centrar en el periodo comprendido entre los 15 y 19 años aproximadamente, donde la tasa es la más elevada, del 1.6%.

3.1.4. Asignación de relaciones

Una vez establecido el reparto demográfico crearemos la red aleatoria con los T nodos. Para crear la red consideraremos K (parámetro que tendremos que ajustar) el grado medio de los T vértices del grafo. Entonces, el grafo poseerá $e = KT/2$ aristas generadas de manera aleatoria. El número de aristas determinado por e , nos dirá el número de relaciones que un individuo posee con los nodos vecinos. El número K por tanto determina el número de relaciones que se crearán en nuestra red.

3.1.5. Definición del criterio de ajuste

Definiremos como "factor de ajuste" o "bondad" del modelo a la diferencia de los porcentajes de portadores del modelo y los definidos en la Tabla 7 que se corresponden con valores reales de Ibarz-Pavón et al. (2011). Estos valores son con campaña de vacunación.

Grupo de edad	% portadores
0-9 años = 0-119 meses	0%
10-14 años = 120-179 meses	0.3125%
15-19 años = 180-239 meses	0.5%
20-24 años = 240-299 meses	0.25%
25-29 años = 300-359 meses	0.0625%
30-34 años = 360-419 meses	0%
35-39 años = 420-479 meses	0.25%
40-44 años = 480-539 meses	0.1875%
45-49 años = 540-599 meses	0.25%
50-54 años = 600-659 meses	0.125%
55-59 años = 660-719 meses	0.25%
60-64 años = 720-779 meses	0.125%
65 años y más >= 780 meses	0%

Tabla 7. Porcentaje de portadores por grupo de edad en el año 2011

Al vector correspondiente de la segunda columna de la Tabla 7 lo vamos a llamar abreviadamente «PORT» (de de portador). Tiene 13 componentes, una por cada grupo de edades. De nuevo observamos que para la franja de edad que nos interesa en este estudio, el porcentaje según los datos que tenemos decrece un tercio los datos de Trotter para los adolescentes.

Con la red así creada estamos por fin en disposición de empezar el proceso de evolución del modelo.

3.2. Algoritmo de evolución

Se va a analizar a continuación la evolución del modelo, prestando atención a los nodos. Para ello definimos el siguiente algoritmo de evolución:

1. El bucle principal itera siguiendo una unidad de tiempo o “turno” (fijada en este caso en meses).

2. Para cada turno hemos de llevar la cuenta de los contagios que se producen, los portadores que hay y el total de individuos de la población, todos por cada grupo de edad.
3. El bucle secundario itera sobre todos los nodos de la red, aplicando a cada uno de ellos las reglas de evolución.
 - a. A cada individuo se le añade un mes a su edad.
 - b. Si el individuo no muere aplicamos las reglas de vacunación y portador.
 - i. La regla de vacunación comprueba si le toca recibir una vacuna y en su caso, vacunarlo.
 - ii. La regla de portador comprueba si el individuo es portador, cuáles de sus vecinos no portadores resultan contagiados por su vecindad y luego vemos si cesa en el estado de portador.
 - c. Si se muere se sustituye por un nodo con atributos de recién nacido (edad 0, sin vacunar y susceptible).

En la Figura 31 podemos un diagrama de flujo aclaratorio:

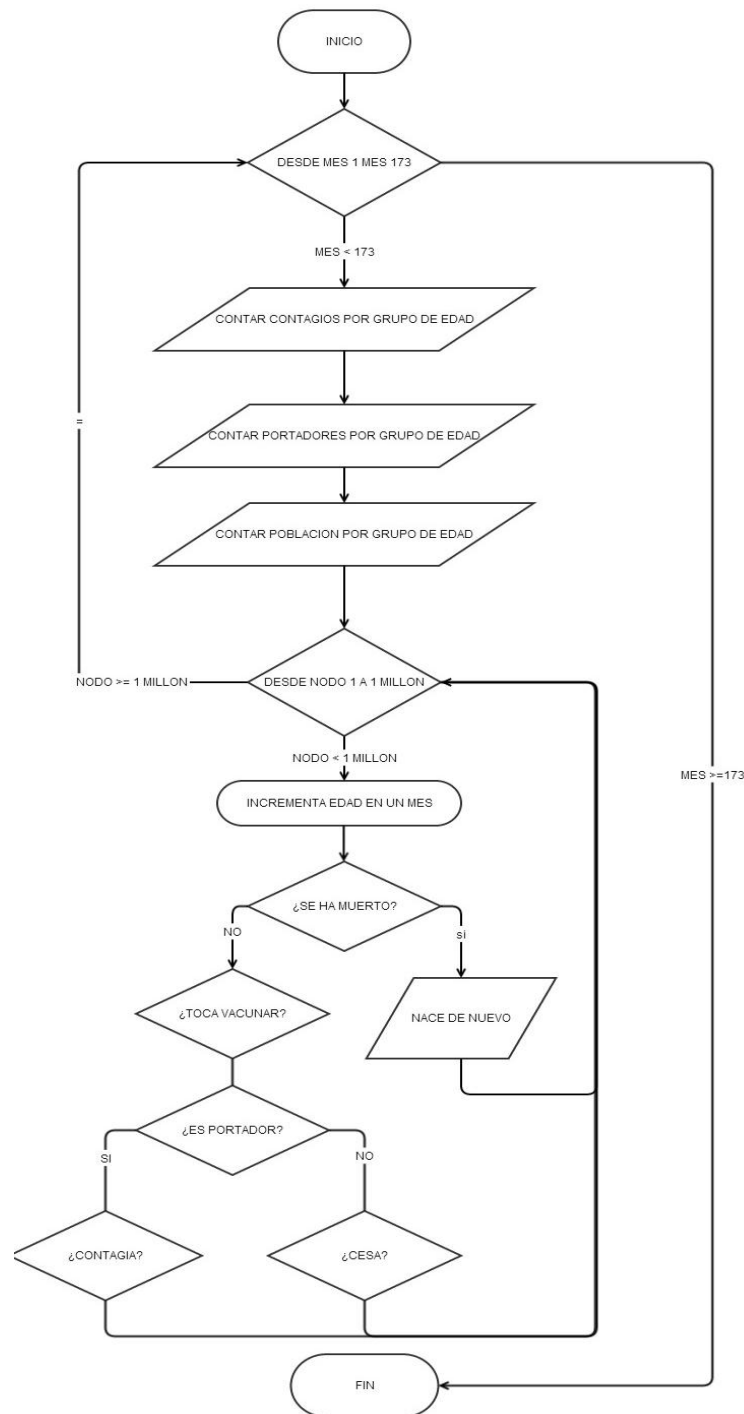


Figura 31: Diagrama de flujo del algoritmo

Cada simulación requiere un tiempo de computación relativamente largo (cerca de 30 minutos en un PC de altas prestaciones), por tanto la realización de sucesivas simulaciones, del orden de decenas de miles, supone un tiempo de cómputo inviable

en la práctica. Como veremos, el problema de la complejidad computacional se soluciona llevando a cabo las simulaciones en un entorno de computación distribuida.

3.2.1. *Parámetros del modelo*

Cada modelo, que a su vez está representado por una red, queda definido por una serie de parámetros característicos, conocidos o no, como la cantidad de nodos, la conectividad, la probabilidad de contagio, etc.

Los parámetros conocidos y que serán fijos para todas las simulaciones son en nuestro caso:

- T , el número de nodos de la red (tamaño de la población).
- t , el tiempo en meses a simular.
- $\mu(i)$, tasa de mortalidad de la edad i , que se obtiene de datos conocidos de la Comunidad Valenciana [Instituto nacional de estadística, 2012].
- cob , la cobertura vacunal actual, que es del 96% en la Comunidad Valenciana [Grupo de trabajo MenCC, 2012].

Los parámetros desconocidos sobre los que tendremos que probar hasta dar con aquellos valores que definan el modelo óptimo son los siguientes:

- K , *grado medio de la red, que es el número medio de contactos que tienen los individuos.*
- γ , *la tasa de pérdida de estado de portador.*
- β_a , *tasa de contagio entre adolescentes.*
- β_o , *tasa de contagio donde, al menos, uno de los dos no es adolescente.*

Como establecimos, la condición inicial se sitúa en agosto de 1997 (situación antes de vacunar) y los niveles (porcentajes) de portadores son los que se describen en la Tabla 6. La condición de ajuste se obtiene en cualquier mes de 2011 (situación con vacuna) donde los niveles (porcentajes) de portadores son los que aparecen en la Tabla 7.

El objetivo es generar múltiples simulaciones de nuestra red explorando diferentes combinaciones de los valores desconocidos analizando para cada simulación cómo de parecido es su resultado a nuestros valores de referencia de los datos de la Tabla 7.

Partimos del modelo de MenC que hemos descrito al comienzo de ésta sección estableciendo parámetros con las hipótesis de que:

1. La adolescencia se considera como el período comprendido entre los 10 años (121 meses) y los 19 años (229 meses). En nuestro modelo no distinguimos entre los hombres y las mujeres porque la incidencia de la enfermedad meningocócica C o la transmisión no depende del sexo de los individuos (Trotter et al., 2006).
2. La unidad de tiempo es de un mes debido a que el tiempo medio de contagio se considera inicialmente como de 3 meses, y parece razonable establecer una escala de tiempo suficientemente pequeña como para seguir la evolución del portador, y a la vez lo suficientemente grande como para disminuir los cálculos asociados a cada actualización en cada paso de tiempo discreto, que consumen tiempo computacional.
3. Nuestro mes de partida es agosto de 1997 ($t = 0$), y la simulación finaliza en 2011 (dado que la unidad t es en meses, 2011 acaba con $t = 173$).
4. Consideramos una red con $T = 5.000.000$ nodos ya que, aproximadamente, es la población de la Comunidad Autónoma de Valencia.

La red está formada por la conexión al azar entre los T nodos con un número medio de K aristas o enlaces por nodo. Por lo tanto la red incluye un total de $e = K T/2$ aristas (dividimos por 2 debido a que la conexión de enlace del nodo I con el nodo J es recíproca y única). Un ejemplo de una red aleatoria se representa en Figura 32.

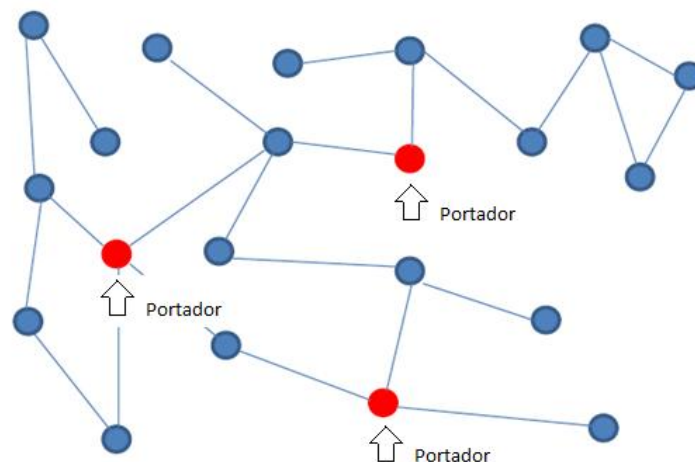


Figura 32: Red aleatoria con portadores

La Figura 32 representa esquematizada una red aleatoria. Los nodos no portadores (azules) son los susceptibles de ser portadores. Los marcados con una flecha son los nodos portadores (rojos). Las líneas representan las conexiones entre los individuos. Una conexión entre un nodo portador y otro susceptible provocará o no dependiendo de la tasa de contagio definida más adelante que el nodo susceptible pase también a ser portador.

- K , indica el número medio de personas con las que un individuo se relaciona. Este valor estará en el intervalo: [5; 60] que definimos de manera arbitraria dentro de unos márgenes que nos parecen empíricamente razonables (Acedo et al., 2011).
- β_a , tasa de contagio entre adolescentes, estará en el intervalo: [0.00075; 0.08], intervalo basado en cómo es el contagio de la enfermedad (Acedo et al., 2011).
- β_o , tasa de contagio donde al menos uno de los dos nodos no es adolescente. En este caso ocurre que $\beta_a > \beta_o$ ya que la tasa de contagio entre adolescentes es mayor que la otra debido a que la adolescencia es la edad de máximo contagio por el tipo de relaciones físicas más cercanas propias de esa edad. Este valor estará en el intervalo: [0.00075; 0.08] (Acedo et al., 2011).
- γ , tiempo de pérdida del estado de portador, define cuándo un individuo deja de ser portador. Este valor será entero y estará en el intervalo: [2; 4] dado que pese a ser desconocida, se sabe que es corto (Trotter et al., 2006).

A modo de introducción al capítulo siguiente, recordemos que tenemos que averiguar los parámetros desconocidos K , β_a , β_o y γ en esos intervalos

3.3. Generación de soluciones

Una vez definidos los parámetros de la red y las reglas de evolución, se generan tuplas de estos parámetros con distintos valores que usamos para definir el modelo, evaluar dicho modelo durante un periodo de 173 meses, y tomar como solución la cantidad de portadores que haya en la población al final de la simulación.

Aquella tupla que produzca la red cuya solución tenga el mejor ajuste (se parezca más a los datos reales de los que disponemos del año 2011) será la que defina nuestro modelo de manera más acertada.

Puesto que las tuplas que vamos a probar se han de generar cubriendo el espacio de parámetros desconocidos dentro del intervalo que estimamos realista hemos de escoger alguna técnica de generación.

El muestreo por *hipercubo latino* (LHS) es un método estadístico para la generación de una muestra de colecciones plausibles de valores de los parámetros de una distribución multidimensional, tipo Montecarlo. El método de muestreo se utiliza a menudo para construir experimentos hechos mediante simulación computacional (McKay et al., 2000).

Se muestrean aleatoriamente valores de los parámetros siguiendo distribuciones uniformes en los intervalos que se deciden para cada parámetro. Luego, mediante unas transformaciones algebraicas, se hace que los valores muestreados se repartan uniformemente por el espacio de parámetros, evitando así que haya zonas de aglomeración.

Para nuestro modelo lo realizamos con el programa informático Mathematica. El cálculo del hipercubo latino se puede ilustrar con un ejemplo. Supongamos que queremos generar 5 muestras para 2 variables de entrada. Asumimos que las variables de entrada X_1 y X_2 siguen una distribución discreta uniforme. Creamos 5 intervalos como vemos en la Figura 33 donde mostramos X_2 (X_1 es igual pero los intervalos tendrían letras de la A a la G).

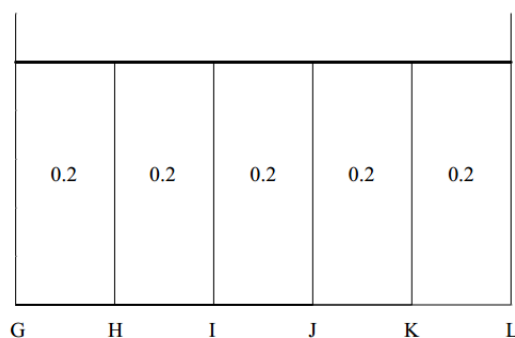


Figura 33: Intervalos híper cubo latino para X_2

El siguiente paso es elegir aleatoriamente un valor dentro de cada intervalo con respecto al valor de la función de densidad de cada uno. Para una función de distribución uniforme, como es nuestro caso, la altura de cada intervalo es la misma (no ocurriría así con una función de distribución normal, puesto que la altura de la función de densidad variaría).

Después, para cada variable de entrada, los valores son puestos en tuplas de k elementos y se asocian según McKay mediante permutaciones de los n números enteros que componen las muestras que tomamos. Para nuestro ejemplo con 5 muestras una permutación daría un conjunto del tipo: (3,2,5,4,1) para X_1 y (4,3,1,5,2) para X_2 .

En cada ejecución el algoritmo forma vectores siguiendo el orden descrito por las permutaciones. Para el primer vector de los 5 que formaría, toma como primer valor el tercero de X_1 y lo empareja con el segundo de X_2 ; para el segundo vector tomaría como primer valor el segundo de X_1 y el tercero de X_2 , y así sucesivamente.

Para encontrar los valores específicos se calculan números aleatorios escogidos de la distribución uniforme estándar (distribuidos uniformemente entre 0 y 1). Se denotan como U_m donde $m = 1, 2, 3, 4, 5$, y se usan para calcular valores aleatorios de la distribución para cada uno de los 5 intervalos.

Además hemos de escalar estos números para obtener la probabilidad acumulada correspondiente, P_m , de modo que cada P_m , cae dentro del intervalo m -ésimo.

Para nuestro ejemplo tenemos:

$$P_m = \frac{1}{5}U_m + \frac{m-1}{5} \quad (34)$$

Esto nos asegura que exactamente una probabilidad P_m va a caer dentro de cada uno de los 5 intervalos (0,0.2), (0.2,0.4), (0.4,0.6), (0.6,0.8), (0.8,1).

Aplicamos entonces la inversa de la función de distribución aplicada a los valores P_m (función cuantil) para producir los valores finales que usamos en las muestras del Hiper Cubo Latino.

3.4. Cálculo y procesado del modelo

Con lo descrito anteriormente el modelo se puede simular. Para ello alimentamos el modelo con cada una de las tuplas que recorren el espacio de parámetros a explorar, generando un total de 50000 modelos distintos, y por tanto 50000 soluciones distintas asociadas a estos modelos. Luego veremos cuál de las soluciones se ajusta mejor a la situación definida por la Tabla 7.

El procesado es lento y costoso en términos computacionales, pese a haberse realizado en código optimizado en C++. Las funciones fueron diseñadas para operaciones complejas específicas, ahorro de espacio de la estructura de la red en la memoria, uso de tipo de datos ajustado al propósito, etc.

El tiempo medio de ejecución de una prueba depende de los turnos a simular, siendo mayor cuantos más turnos simulemos. Nos movemos en un orden de media hora en los casos donde hemos simulado 173 meses (se realizó un *script* que obtuvo la media de ejecución medida en segundos de las 50.000 simulaciones).

Si hacemos el cálculo de media hora multiplicado por 50000 soluciones, obtenemos un tiempo de casi 3 años para terminar nuestro experimento usando equipos con las siguientes características:

- Xeon X3230 2.66Ghz 6GB RAM (2 cores) X 10. Windows 7 64 bits.
- Xeon X3430 2.4Ghz 16GB (4 cores) X 5. Windows 7 64 bits.

Para abordar el problema haremos uso de la computación distribuida. Con esta técnica empleamos un grupo de ordenadores dentro de una red de cálculo distribuido que nos permitirá procesar grupos de redes de 40 en 40 a la vez.

El entorno de cálculo distribuido más extendido es el software BOINC de código abierto (BOINC, 2014). Inicialmente creado para el proyecto SETI@HOME, goza hoy en día de gran popularidad y se usa en multitud de proyectos científicos (matemáticas, medicina, biología molecular, climatología, medio ambiente, y la astrofísica) incluyendo estudios epidemiológicos como MalariaControl o enfermedades infecto-contagiosas como el VRS o Virus Respiratorio Sincitial (Acedo et al., 2011).

Hay, sin embargo, otras plataformas de cómputo distribuido como EGI, WCG, o SISIFO.

En concreto, para nuestro problema usaremos SISIFO. Conceptualmente muy similar a BOINC, SISIFO es capaz de asignar tareas a un conjunto de PCs, esperar a que las tareas se completen y recoger los resultados para su posterior análisis. Sin embargo, y en contraste con BOINC (que requiere meses de trabajo de configuración sólo para instalarlo y ponerlo en marcha), SISIFO es bastante simple, dando como resultado un sistema que prácticamente no requiere mantenimiento, necesita muy poca configuración, y puede ser configurado para nuestros fines en cuestión de minutos.

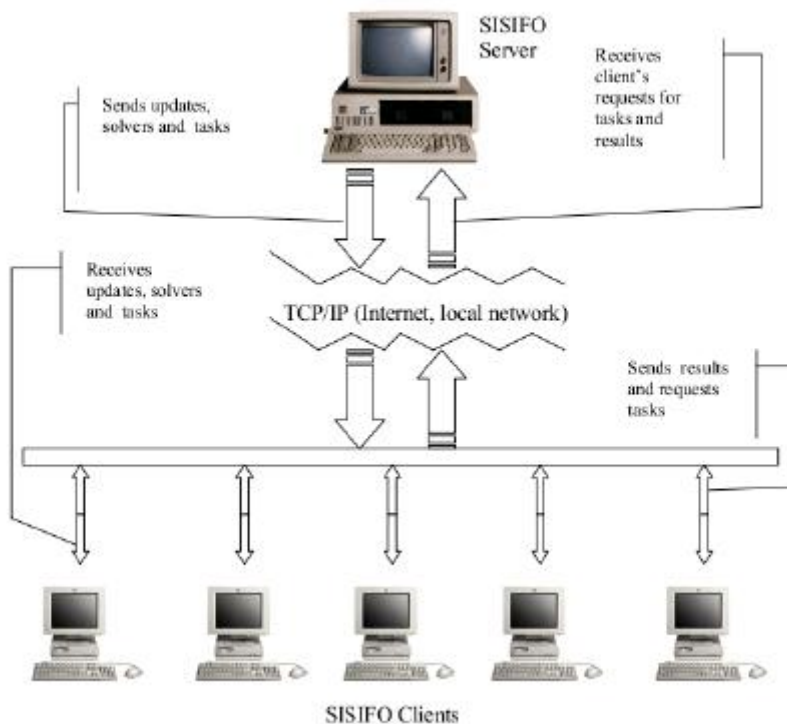


Figura 34: Arquitectura de SISIFO

Ventajas

- Esfuerzo de mantenimiento casi cero (ninguna base de datos, no hay un servidor web).

- Esfuerzo casi cero de instalación. (copiar los archivos en una carpeta, configurar algunas rutas y ejecutar).
- Puede trabajar en múltiples S.O. (actualmente disponible para las plataformas Windows y Linux).
- Consume muy pocos recursos (sólo un único programa que puede ejecutarse en un PC bajo especificaciones con conexión Ethernet).

Desventajas

- No es adecuado para su uso con clientes que no son de confianza.
- No maneja los usuarios independientes, estadísticas individuales, foros y otras características Web.
- No maneja redundancia de soluciones ni decide que una solución es igual a otra ("*matching*").
- Carece de una interfaz amigable.

Cliente SISIFO

Dentro de la arquitectura de SISIFO (Figura 34), el llamado "Cliente" es un programa almacenado en uno (o varios) equipos PC. Este cliente se conecta al servidor, y pide un paquete de trabajo. Este paquete de trabajo se compone de dos elementos: un "problema" (un archivo que contiene los datos que definen el Problema a resolver) y un "ejecutor" (el archivo ejecutable capaz de resolver el Problema dado). El cliente, una vez que se recibe el paquete de trabajo, extrae el Ejecutor, el problema, y se llama al Ejecutor para que pueda ejecutar, el acceso al problema, leer los datos, resolverlo y almacenar de nuevo una solución. Una vez que el Ejecutor termina la operación, el cliente recoge la solución y lo envía de vuelta al servidor.

Servidor SISIFO

El llamado "Servidor" se ejecuta en un ordenador de IP fija, y mantiene la escucha de solicitud de los clientes. El servidor ha almacenado uno o más ejecutables, un conjunto de problemas que hay que resolver, y un conjunto de soluciones enviadas. El servidor atiende a un puerto TCP / IP configurado específica para ese propósito, y es capaz de ejecutar de manera concurrente con muchos otros servidores, por lo que es posible configurar varios servidores, cada uno atendiendo a un puerto y sirviendo

diferentes problemas. La configuración típica del servidor contiene, entre otros, los siguientes archivos:

- Servidor
- Monitor
- config.cfg
- IP_filtradas.txt
- .Ejecutor\exe.exe (para sistemas Windows).
- .Ejecutor\exe.lin (para sistemas Linux).

Cada cliente utiliza un núcleo de CPU. Cuantos más núcleos tiene el PC, más clientes se pueden ejecutar al mismo tiempo. Sin embargo, hay que tener en cuenta la cantidad de RAM que cada problema requiere a la hora de decidir el número de clientes a ejecutar en un PC específico. En nuestro *cluster* de servidores de cuatro núcleos tenemos la siguiente configuración:

- 4 clientes están configurados para trabajar en el puerto A.
- 2 clientes están configurados para trabajar en el puerto B.
- 1 cliente está configurado para funcionar en el puerto C.

Si los problemas requieren poca memoria RAM y los PCs pueden gestionar 4 problemas al mismo tiempo, entonces se utiliza el servidor en el puerto de A. Si se pueden gestionar 2 problemas en un PC al mismo tiempo, se utiliza el servidor en el puerto B. Por último, para problemas que exigen mucha memoria RAM haciendo que sólo se pueda ejecutar uno a la vez en los ordenadores, se ejecuta en el puerto del servidor C.

Además contamos con el apoyo de 16 equipos de altas prestaciones que usamos como capacidad mínima de cálculo, en el Instituto Universitario de Matemática Multidisciplinar (IMM), de la Universidad Politécnica de Valencia.

4. Evaluación, Análisis de Resultados y Validación del Modelo

4.1. Generación de los modelos a evaluar (archivos «problema»)

Llamaremos «problema» al conjunto de valores que definen una red cualquiera que queremos evaluar. Esos valores se gestionan como archivos de texto (uno por cada «problema») con el siguiente contenido:

```

identificador_de_prueba
Prueba21578
consumo_de_RAM
3235000
generador_de_red
0
turnos_a_simular
2000
numero_de_realizaciones
1
numero_de_nodos
1000000
grado_medio_de_los_nodos
58
tasa_de_vacunacion_p
0.96
edad_vacunacion_1
2
edad_vacunacion_2
6
edad_vacunacion_3
18
tasa_contagio_adolescentes_betaA
0.0729749
tasa_contagio_no_adolescentes_beta0
0.00433782
tasa_perdida_de_portador
4
tramo_inicio_nueva_estrategia_vacunacion_(ENE2014)
28
numero_dosis_vacuna
3
    
```

Para generar estos archivos «problema» utilizamos un programa escrito en Mathematica acorde al método del Hipercubo Latino descrito en la sección 3.3, y que se encarga de explorar el espacio de valores de cada parámetro y generar un fichero con la combinación de valores correspondiente.

4.2. Implementación del programa «calculador»

La evolución del modelo definido por cada uno de los ficheros «problema» se hace mediante un programa escrito en C++ al que llamamos «calculador» y que se encarga de abrir el fichero, leer los parámetros, generar una red acorde a dichos parámetros, hacerla evolucionar según las reglas descritas en la sección 3.2 y finalmente guardar el resultado de dicha evolución en otro archivo de texto cuyo contenido se explicará en la sección siguiente.

Podemos ver la salida de la ejecución del modelo en la Figura 35 cuando se ejecuta.

```

C:\Users\vsancheza\Downloads>C:\Users\vsancheza\Downloads\Meningococo@Home64PFM.exe Prueba21578.pro
*****
Simulador de susceptibilidad para Meningococo@Home MGC_20140623_W01
Autores: J. Villanueva-Oller y R. J. Villanueva
Departamento de Matematica Aplicada, Universidad Politecnica de Valencia
CE3 Felipe II, Universidad Complutense de Madrid
Direccion de contacto: jvillanueva@pdi.ucm.es
*****
Comienza la simulacion de la inmunidad
Estimamos 407 MB de RAM
Realizacion 1
Asignando edades
Asignando vacunas y factores de proteccion al sujeto 999999
57999942 de 57999942 (100%) (835 colisiones) 5272722 ar/seg
Tiempo de generacion: 00:00:11
Velocidad de generacion: 5272722 aristas/segundo
Valores iniciales:
Entrada: Prueba21578.pro Salida: Prueba21578.sol ID Prueba21578
RAM: 3235000 Generador: 0 Turnos: 2000
Nodos: 999999 Grado: 58 Tasa vacunacion: 0.96
Calendario oficial: 2 6 18
Nuevo calendario: 2 6 18 Empieza nuevo calendario el turno: 28
Simulacion:
T:1/2000 C:2236 M:774 U:205812 F:0 (00:00:17)
T:2/2000 C:3127 M:792 U:205812 F:0 (00:00:20)
T:3/2000 C:4668 M:861 U:205812 F:0 (00:00:23)
T:4/2000 C:5763 M:843 U:205812 F:0 (00:00:26)
T:5/2000 C:7082 M:801 U:205812 F:0 (00:00:29)
T:6/2000 C:9111 M:805 U:205812 F:0 (00:00:32)
T:7/2000 C:11241 M:844 U:205812 F:0 (00:00:36)
T:8/2000 C:13517 M:825 U:205812 F:0 (00:00:39)
T:9/2000 C:16576 M:820 U:205812 F:0 (00:00:42)
T:10/2000 C:19453 M:845 U:205812 F:0 (00:00:45)
T:11/2000 C:22753 M:874 U:205812 F:0 (00:00:48)
T:12/2000 C:25603 M:827 U:205812 F:0 (00:00:51)
T:13/2000 C:28162 M:916 U:205812 F:0 (00:00:54)

```

Figura 35: Comportamiento cíclico del modelo en 2000 turnos

Podemos ver en la Figura 35 un extracto de la consola de ejecución del modelo, donde a *grosso modo* se crea la red, los nodos y se empieza a simular cada turno con los parámetros establecidos.

4.3. Fichero «solución»

El resultado de la simulación de cada modelo tiene la forma de un fichero de texto con una extensión que lo identifica como tal («.sol», abreviatura de «solución»).

Dicho resultado contiene tres partes claramente diferenciadas: parámetros del problema evaluado, resultado de la simulación de dicho problema (la «solución» propiamente dicha), y un conjunto de valores auxiliares que se usan para información accesoria. El tamaño del fichero de texto depende del número de turnos a simular, y en nuestro caso los 173 turnos simulados hacen que el fichero solución tenga un peso aproximado de 563 Kilobytes.

La Figura 36 muestra un ejemplo de la estructura del fichero solución. Están resaltados los dos primeros de los elementos antes descritos.

```
(* ID de prueba, RAM, generador de red, turnos, realizaciones, nodos, grado medio, tasa de vacunacion,
calendario 1, calendario 2, calendario 3, betaA, betaO, perdida de portador, inicio nueva estrategia,
dosis nueva estrategia,vacuna especial, tramo inicio vacuna especial, tramo fin vacuna especial, edad
inicio vacuna especial, edad fin vacuna especial, tasa de vacunacion especial*)
problema={Prueba24016,3235000,0,2000,1,999999,56,0.96,2,6,18,0.0453606,0.00490022,3,28,3,0,0,0,0,0,0};
(*{{turno,{edad,poblacion total, no portadores con menor SBA, no portadores con NO menor SBA, portadores
con menor SBA, portadores con NO menor SBA}...*) solucion={
{0,{0,12347,3220,9127,0,0},{1,12394,3020,9374,0,0},{2,12543,2333,10210,0,0},{3,12585,7612,4973,0,0},{4,1253
9,7518,5021,0,0},{5,12496,8297,4199,0,0},{6,12661,8427,4234,0,0},{7,12505,11074,1431,0,0},{8,12467,10986,14
81,0,0},{9,12359,11200,1159,0,0},{10,12477,11177,1164,125,11},{11,12506,11232,1147,110,17},{12,12402,10117,
```

Figura 36: Ejemplo de fichero solución

Los valores del «problema» son una lista numérica de elementos separados por comas y llaves, con la información propia de la prueba simulada (número de turnos, nodos, probabilidad de contagio, etc.) ya descritos en capítulos anteriores. Se incluye por comodidad, para evitar tener que consultar el archivo de problema original que dio resultado a la solución.

Los valores de la « solución» son una lista numérica de valores separados por comas y llaves, para cada uno de los 173 turnos (recordemos que son meses). Estos valores numéricos son en concreto la cantidad de portadores y susceptibles para cada uno de los 13 grupos de edad definidos en Tabla 6.

La parte «auxiliar» (que no aparece en la figura) contiene en este caso el tiempo total de ejecución de la simulación y una comprobación de integridad de los datos.

Así, por cada una de las 50.000 simulaciones realizadas, tendremos un fichero de solución correspondiente que hemos de evaluar a fin de descubrir cuál de ellos parametriza de manera más fiel el comportamiento ideal del modelo.

4.4. Evaluación de las soluciones

La evaluación se hará, por cuestiones de eficiencia, de una manera automática que nos permita decidir según un criterio definido cuál de las tuplas de parámetros hace que nuestro modelo logre el mejor ajuste a los datos de la Tabla 7.

Una vez encontrada la mejor solución, re-evaluaremos el modelo elegido durante un periodo de tiempo suficientemente largo, 2000 turnos, para comprobar si tiene un comportamiento cíclico.

Nuestro criterio para decidir qué solución es mejor es comparar las salidas del modelo en los meses 162, 162,..., 173 (el año 2011 completo, desde enero a diciembre) con los datos de la Tabla 7 y ver cuál tiene el menor error cuadrático medio (ECM).

Del fichero «solución», tenemos que analizar la parte que contiene las 173 tuplas (una por cada turno) con la siguiente estructura:

Solución
<ul style="list-style-type: none"> • Turno • Grupo de edad. • Número de individuos del grupo de edad. • Número de individuos que no son portadores. • Número de individuos que si son portadores.

Estos datos hay que agruparlos según los grupos de edad definidos en la Tabla 6 o en la Tabla 7. Teníamos 13 grupos de edad (Tabla 8):

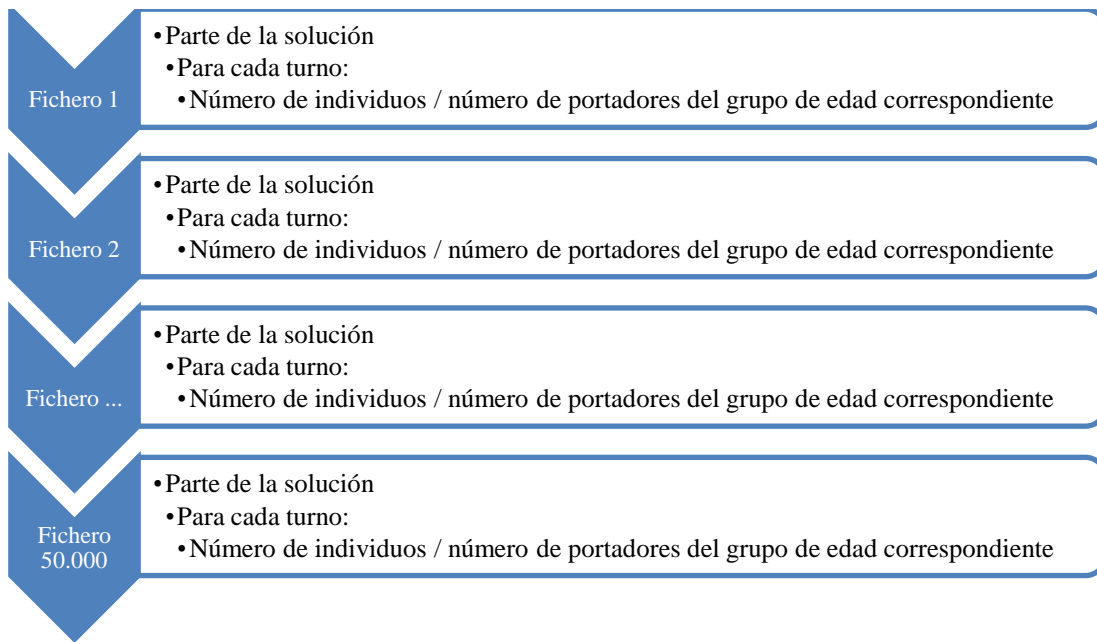
Grupo	Edad
1	0-9
2	10-14
3	15-19
4	20-24
5	25-29
6	30-34
7	35-39
8	40-44
9	45-49
10	50-54
11	55-59
12	60-64
13	65-119

Tabla 8. Definición de grupos de edad del modelo

4.5. Implementación del evaluador

Una vez descrito el contenido de la tupla solución pasamos a detallar la forma en que funciona el evaluador de soluciones.

El procedimiento de búsqueda recorre los 50.000 ficheros solución y para cada turno comprendido en el año 2011 (mes 162 a 173 incluido) tomamos:



Al final tendremos para cada fichero «solución» una matriz donde las filas representan el turno en meses y las columnas el grupo de edades correspondiente. La matriz tiene la forma expuesta en la Tabla 9.

Grupo \ Meses	1	2	...	13
162				
163				
164				
...				
173				

Tabla 9. Agrupación de edades por meses en el año 2011

Cada valor de la matriz de la Tabla 9 representa la relación entre individuos portadores e individuos totales en cada grupo de edad que se almacena en un vector

$MO(t, i)$, donde t es el turno en el que nos encontramos en 2011 e i el grupo de edad actual.

En la tabla 10 a continuación se detalla un ejemplo de la entrada $MO(162;i)$ para un fichero solución ficticio:

Grupo \ Meses	1	2	3	4	5	6	7	8	9	10	11	12	13
162	0.04	0.03	0.001	0.03	0.03	0.03	0.001	0.02	0.01	0.01	0.01	0.01	0.002

Tabla 10. Valores al agrupar por edades y por meses en el año 2011

Como vamos a calcular el ECM siguiendo la ecuación

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (MO_{t,i} - PORT_i)^2} \tag{35}$$

para cada turno t tenemos el vector $MO(t) = \{MO(t; 1), \dots, MO(t; 13)\}$, donde cada componente del vector $MO(t; i)$ será restada por cada componente del vector $PORT_i$, que recordemos que es la segunda columna de la Tabla 7 (situación de portadores de 2011).

Por ejemplo supongamos que queremos hallar el ECM del turno 162 de un fichero solución. Siguiendo con el ejemplo de la tabla 10 haríamos:

Grupo \ Meses	1	...	13
162	$\sqrt{(0.04 - PORT_i)^2}$	$\sqrt{(MO_{t,i} - PORT_i)^2}$	$\sqrt{(0.002 - PORT_i)^2}$

Tabla 11. Ejemplo del cálculo del ECM para el turno 162

Se suman los 13 valores del turno 162 y se dividen entre 13, que es el número de grupos de edad totales.

Con el objetivo de quedarnos con el menor error ECM en el año 2011 de cada fichero solución, guardamos el valor menor de los hallados en los 12 meses del año 2011.

Finalmente el evaluador se realiza en código Mathematica y se muestra a continuación el código fuente:

```

PORT = {0,0.3125,0.5,0.25,0.0625,0,0.25,0.1875,0.25,0.125,0.25,0.125,0} / 100

trotter = { 0, 1, 1.6, 0.8, 0.2, 0, 0.8, 0.6, 0.8, 0.4, 0.8, 0.4, 0} / 100

(* Gupos de edad de interés para la visualización de los datos *)

geCAR = { {0, 9}, {10, 14}, {15, 19}, {20, 24}, {25, 29}, {30, 34}, {35, 39}, {40,
44}, {45, 49}, {50, 54}, {55, 59}, {60, 64}, {65, 119}};

(* toma los archivos .sol del directorio *)

FicherosSol = FileNames["*.sol"];

turnoFin          = 174

turnoIni          = 163

ListaFicherosMinECM = {}

(* bucle principal de 1 a 50.000 *)

For[ficheroSol = 1, ficheroSol <= Length[FicherosSol], ficheroSol++,

    (* Vamos cargando ficheros solucion. Tiene una lista problema y una lista
solucion*)

    Get[FicherosSol[[ficheroSol]]];

    Print["FicherosSol -> ",FicherosSol[[ficheroSol]]];

    ListaECM = {};

    For[t = turnoIni, t <= turnoFin, t++,

        (* Nos quedamos con la solucion del turno indicado *)

        sol = solucion[[ t ]] // Rest;

        (* Agrupamos por grupos de edad *)

        agrupal = Map[ sol[[ #[[1]] + 1 ;; #[[ 2 ]] + 1 ]] &, geCAR ];

        (* Sumamos los datos de cada grupo de edad *)

```

```

agrupa2 = Map[ Total, agrupa1 ];

(* Quitamos el primer elemento de la lista que no sirve *)

agrupa3 = Map[ Rest, agrupa2 ];

(* Calculamos porcentajes *)

aux = Map[ ( #[[ 4 ]] + #[[ 5 ]] ) / #[[ 1 ]], agrupa3 ];

(* Error cuadrático medio *)

error = Sqrt[Total[(PORT-aux)^2]];

ListaECM = Append[ListaECM, {error,t}];

];

ECM = Min[ListaECM];

ListaFicherosMinECM = Append[ListaFicherosMinECM, {ECM, FicherosSol[[ficheroSol]]}]

]

SortBy[ListaFicherosMinECM, First]

```

4.6. Resultados

4.6.1. Modelo correspondiente al mejor ajuste

De las 50.000 «pruebas» o modelos simulados, la que mejor ajusta (menor ECM) tiene como parámetros de ajuste:

- Grado medio de la red -> 42 (K)
- Tasa de contagio entre adolescentes -> 0.0722085 = 722 de cada 10.000 (β_a)
- Tasa de contagio entre NO adolescentes -> 0.00589516 = 58 de cada 10.000 (β_o).
- Tiempo medio de portación -> 3 meses (γ).

De los 12 meses, dibujamos el mes de abril, que es el que tiene de todos un mejor ajuste, como se muestra en la Figura 37.

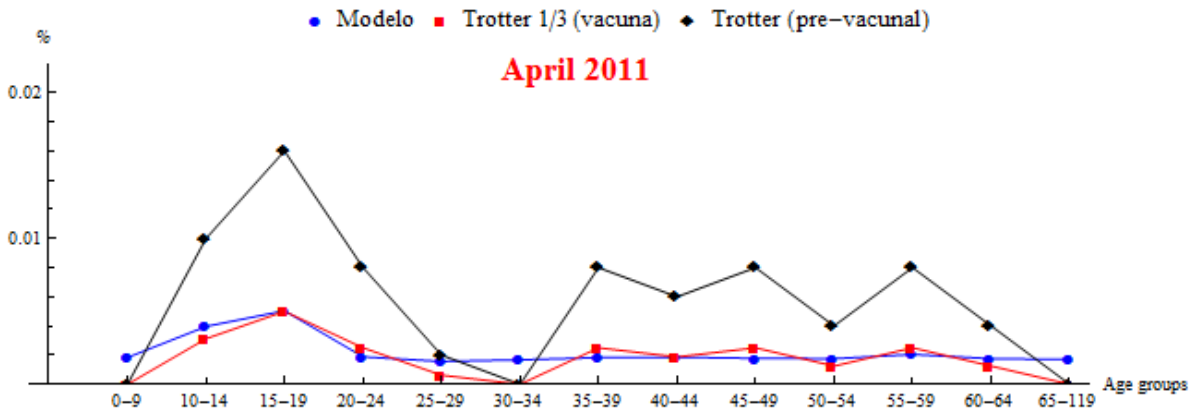


Figura 37: Grafica del turno con menor ECM en 2011 del modelo ajustado

La Figura 37 representa los porcentajes por grupo de edad de portadores. En rojo tenemos lo establecido en la tabla 7, en negro lo establecido en la tabla 6 y en azul los portadores obtenidos por el modelo para la simulación descrita con los parámetros obtenidos.

Se puede observar que el comportamiento del modelo es muy satisfactorio y se ajusta bastante a los datos esperados.

4.6.1. Búsqueda de comportamiento cíclico en el modelo obtenido

Llegados a este punto, estamos en disposición de comprobar si hemos logrado nuestro objetivo inicial. Ése era, recordémoslo, dar con un modelo del contagio de la enfermedad que logre caracterizar su comportamiento cíclico respecto a la cantidad de portadores en función del tiempo, característica ya identificada en los estudios de (Chabalier et al., 2000).

Para ello lo que haremos será re-evaluar el modelo definido por los parámetros que hemos obtenido en la sección anterior, pero no ya durante 173 meses sino por un periodo de tiempo arbitrariamente largo, por ejemplo 2000 meses.

Los resultados, que conllevan un tiempo de ejecución de 3 horas, se muestran en la Figura 38. Como vemos, la cantidad de portadores aumenta y disminuye cíclicamente, que es lo que queríamos conseguir.

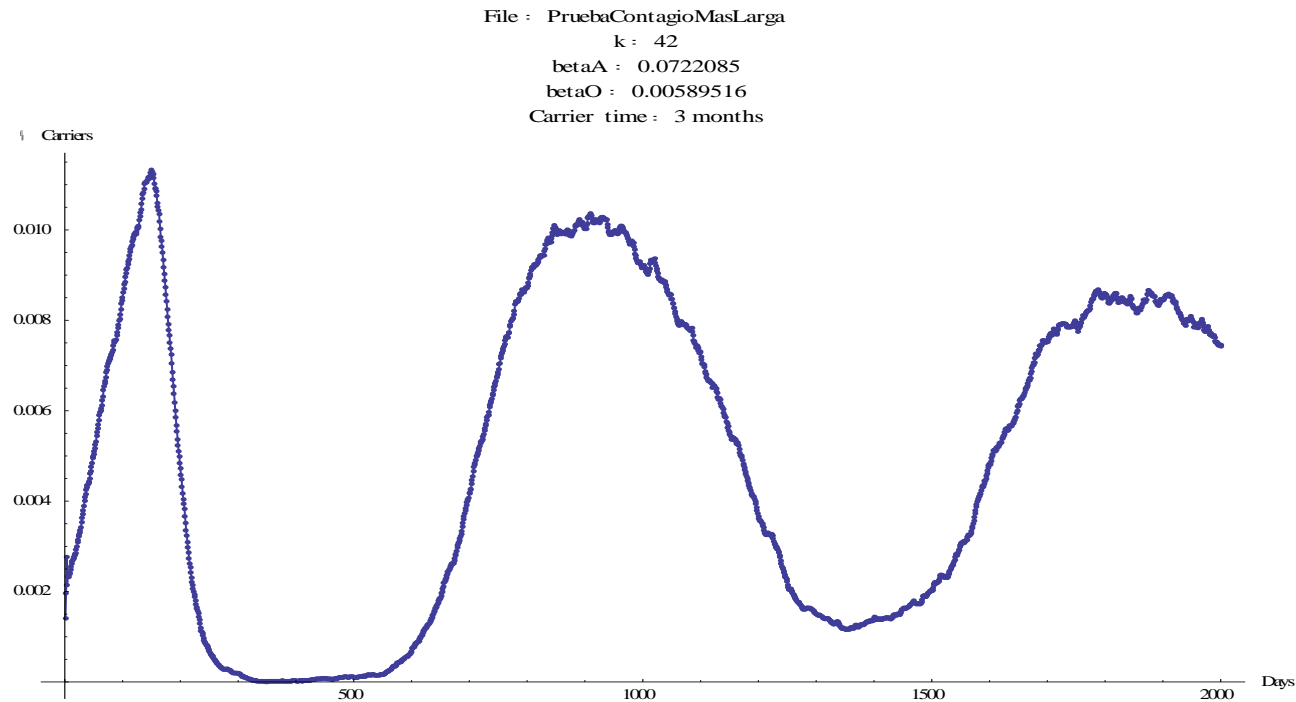


Figura 38: Comportamiento cíclico del modelo a lo largo de 2000 turnos

A título de ejemplo, indicar que si hacemos la misma simulación pero usando parámetros que den un peor ajuste, el número de portadores termina desapareciendo, como se muestra en la Figura 39.

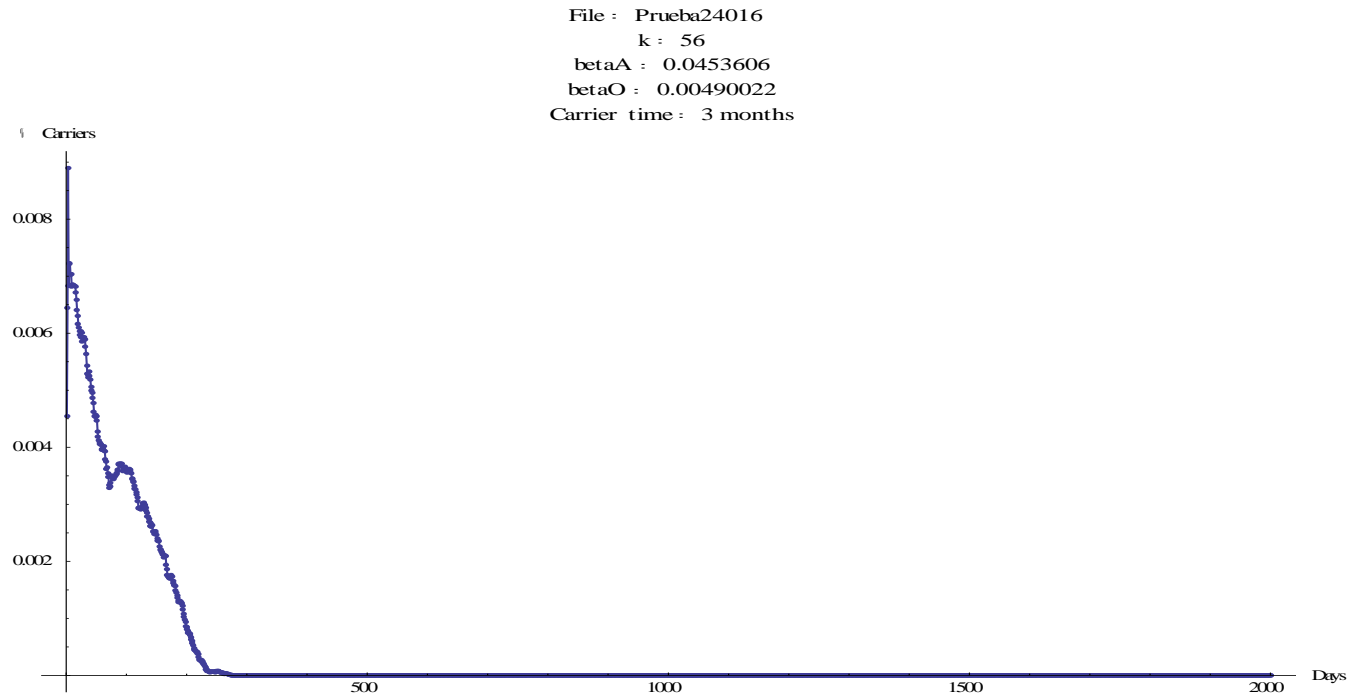


Figura 39: Comportamiento no cíclico del modelo en 2000 turnos

5. Conclusiones y Trabajos Futuros

En el presente trabajo hemos abordado la problemática de modelar el comportamiento de un sistema complejo como es la forma en la que se transmite una enfermedad infectocontagiosa en una población, incorporando factores tales como la edad, mortalidad, probabilidad de contagio, vacunaciones, etc.

De entre las distintas alternativas que tenemos a nuestra disposición para efectuar dicho modelado, tras el análisis pertinente, hemos optado por el uso de redes aleatorias. Aunque no exentos de inconvenientes pues los modelos de este tipo son, desde el punto de vista computacional, muy costosos y bastante más complejos de implementar que otras alternativas más clásicas, tienen por otra parte una serie de ventajas que nos han impulsado a solventar estos escollos utilizando herramientas adicionales, en concreto la computación distribuida, sin la cual el experimento no hubiera sido realizable en un plazo de tiempo realista.

El resultado de la experiencia ha sido positivo. Hemos sido capaces de definir un modelo de la enfermedad, y luego dar con una parametrización del mismo con un ajuste respecto a los datos del mundo real de los que disponemos, para que sea capaz de evolucionar con el tiempo y comportarse espontáneamente como lo hace la enfermedad en la población real: manteniendo un grupo permanente de población portadora de la enfermedad cuyo número oscila en función del tiempo, no lo suficientemente grande como para contagiar a todos y no lo suficientemente pequeña como para que desaparezca, comportamiento que se sabe es lo que ocurre en África según (Chabalier et al., 2000).

5.1. Trabajos Futuros

Son varias las puertas que se nos abren con el uso de las redes aleatorias para modelar este tipo de sistemas dirigidos por contagio.

- Con respecto al Meningococo C, cabe preguntarse cómo impactan las distintas políticas de vacunación en la dinámica de contagio, y si hay alguna que sea mejor que otra.

- Con respecto a otras enfermedades (varicela, gripe, SIDA, etc.), parece factible también diseñar modelos para ellas que permitan estudiar su evolución ante influencias externas (campañas de concienciación, prevención, estratificación por edades, vacunación, etc.).

6. Bibliografía

Acedo, L., Moraño, J. A., & Díez-Domingo, J. (2011). Cost analysis of a vaccination strategy for respiratory syncytial virus (RSV) in a network model. *Mathematical and Computer Modelling*, 52(7), 1016-1022.

Acedo, L., Moraño, J. A., Villanueva, R. J., Villanueva-Oller, J., & Díez-Domingo, J. (2011). Using random networks to study the dynamics of respiratory syncytial virus (RSV) in the Spanish region of Valencia. *Mathematical and Computer Modelling*, 54(7), 1650-1654.

Albert, R., and A. L. Barabási. "Review of modern physics." *Rev. Mod. Phys* 74 (2002): 47.

Anderson, Roy M., Robert M. May, and B. Anderson. *Infectious diseases of humans: dynamics and control*. Vol. 28. Oxford: Oxford university press, 1992.

Balmer P, Borrow R, Miller E. Impact of meningococcal C conjugate vaccine in the UK. *J Med Microbiol* 2002; 51: 717-722

Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999a): 509-512.

Barabási, Albert-László, Réka Albert, and Hawoong Jeong. "Mean-field theory for scale-free random networks." *Physica A: Statistical Mechanics and its Applications* 272.1 (1999b): 173-187.

Barrett, C. L., Bisset, K. R., Eubank, S. G., Feng, X., & Marathe, M. V. (2008, November). EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing* (p. 37). IEEE Press.

Blanco Quirós, A., Casado Flores, J., Nieto Moro, M., Garrote Adrados, J. A., Arranz Sanz, E., & Asensio Antón, J. (2004, December). Sepsis meningocócica en pediatría. Parámetros asociados a mala evolución. In *Anales de Pediatría* (Vol. 61, No. 4, pp. 305-313).

Bollobás, B., FULTON, W., KATOK, A., KIRWAN, F., & SARNAK, P. (2001). Cambridge Studies in Advanced Mathematics. In Random graphs. New York: Cambridge University Press.

BOINC, Programas de código abierto para computación voluntaria y computación en red (grid). Consultado en junio de 2014 en <http://boinc.berkeley.edu/>

Chabalier, F., Djingarey, M. H., Hassane, A., & Chippaux, J. P. (2000). Meningitis seasonal pattern in Africa and detection of epidemics: a retrospective study in Niger, 1990–1998. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(6), 664-668..

Chitnis, Nakul. "Einführung in die Mathematische Epidemiologie: Introduction to Mathematical Epidemiology: Deterministic Compartmental Models."

Christakis, Nicholas A., and James H. Fowler. "The collective dynamics of smoking in a large social network." *New England journal of medicine* 358.21 (2008): 2249-2258.

Christakis, Nicholas A., and James H. Fowler. "The spread of obesity in a large social network over 32 years." *New England journal of medicine* 357.4 (2007): 370-379.

Coberturas de vacunación, Ministerio de Sanidad, Servicios Sociales e Igualdad. Enlace consultado en junio 2014 <http://www.msssi.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/coberturas.htm>

Consejería de Sanidad. Enlace de mayo de 2014 http://www.san.gva.es/documents/153218/167583/06_Analisis.pdf.

Diekmann, Odo, J. A. P. Heesterbeek, and Johan AJ Metz. "On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations." *Journal of mathematical biology* 28.4 (1990): 365-382.

Dietz, Klaus. "Epidemics and rumours: A survey." *Journal of the Royal Statistical Society. Series A (General)* (1967): 505-528.

Donado-Campos, F. Morilla, F. Martínez-Navarro. Dynamic model for prediction of the epidemic wave of flu A/H1/N1. Why we have made a mistake? 2010 European Scientific Conference on Applied Infectious Disease Epidemiology (ESCAIDE).

Dreyer, Thomas Prins. *Modelling with ordinary differential equations*. CRC Press, 1993.

Dropkin, Victor Harry. Introduction to plant nematology. No. Ed 2nd. John Wiley and Sons Inc., 1989.

Edelstein-Keshet, Leah. *Mathematical models in biology*. Vol. 46. Siam, 1988.

Erdős, Paul, and Alfréd Rényi. "On random graphs." *Publicationes Mathematicae Debrecen* 6 (1959): 290-297.

Erdős, Paul, and Alfréd Rényi. "On the evolution of random graphs." *Magyar Tud. Akad. Mat. Kutató Int. Közl* 5 (1960): 17-61.

Eykhoff, Pieter. "System identification parameter and state estimation." (1974).

Grupo de trabajo MenCC 2012, "Revisión del programa de vacunación frente a enfermedad meningocócica por serogrupo C"

<http://www.msssi.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/docs/MenC.pdf>

Halloran, M. E., Longini, I. M., Nizam, A., & Yang, Y. (2002). Containing bioterrorist smallpox. *Science*, 298(5597), 1428-1432. Halloran, M. E., Longini, I. M., Nizam, A., & Yang, Y. (2002). Containing bioterrorist smallpox. *Science*, 298(5597), 1428-1432.

Hethcote, Herbert W. "Three basic epidemiological models." *Applied mathematical ecology*. Springer Berlin Heidelberg, 1989. 119-144.

Ibarz-Pavón, A. B., MacLennan, J., Andrews, N. J., Gray, S. J., Urwin, R., Clarke, S. C., ... & Maiden, M. C. (2011). Changes in serogroup and genotype prevalence among carried meningococci in the United Kingdom during vaccine implementation. *Journal of Infectious Diseases*, 204(7), 1046-1053.

Infante Gil, Said, and G. Zárate de Lara. "Métodos estadísticos." *Trillas México*(1984).

INSTITUTO NACIONAL DE ESTADÍSTICA. 2009. Instituto Nacional de estadística. Madrid: INE. Consulta: junio 2012. Disponible en: <http://www.ine.es/jaxiBD/menu.do?L=0&divi=IDB&his=0&type=db>

James Meiss (2007) Dynamical systems. Scholarpedia, 2(2):1629.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, 407(6804), 651-654.

Keeling, Matt J., and Ken TD Eames. "Networks and epidemic models." *Journal of the Royal Society Interface* 2.4 (2005): 295-307.

Kendall, M. G., & Buckland, W. R. (1971). A dictionary of statistical terms: prepared for the International Statistical Institute by Maurice G. Kendall and William R. Buckland. Oliver & Boyd.

Kermack, M., & McKendrick, A. G. (1927). Contributions to the mathematical theory of epidemics. Part I. In *Proc. R. Soc. A* (Vol. 115, No. 5, pp. 700-721).

Koopman, Jim. "Controlling smallpox." *Science* 298.5597 (2002): 1342-1344.

Lou, Jie, and Tommaso Ruggeri. "The dynamics of spreading and immune strategies of sexually transmitted diseases on scale-free network." *Journal of Mathematical Analysis and Applications* 365.1 (2010): 210-219.

MalariaControl. Sitio consultado en Junio de 2014 en <http://www.malariacontrol.net/>.

Martin, J. Villanueva-Oller, I. Hidalgo, M. Alberquilla, and I. Contreras, *Anales de Ingenieria Técnica Informática de Sistemas* 3, CES Felipe II, Virtual supercomputers and distributed computing with BOINC, Berkeley, Calif, USA, 2010.

McKay, M. D., R. J. Beckman, and W. J. Conover. "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code." *Technometrics* 42.1 (2000): 55-61.

McNeill, William. *Plagues and peoples*. Random House LLC, 2010.

Murray J.D., *Mathematical Biology: I. An Introduction*, Springer, Berlin, Germany, 2002.

Newman, Mark EJ. "Power laws, Pareto distributions and Zipf's law." *Contemporary physics* 46.5 (2005): 323-351.

Office for National Statistics. Mortality Statistics: Childhood, infant and perinatal, England and Wales. 1997, Series DH3, no.30, p37.

Proulx, Stephen R., Daniel EL Promislow, and Patrick C. Phillips. "Network thinking in ecology and evolution." *Trends in Ecology & Evolution* 20.6 (2005): 345-353.

Santamaría González, Fernando. "Posibilidades pedagógicas. Redes sociales y comunidades educativas." *Telos: Cuadernos de Comunicación, Tecnología y Sociedad* 76 (2008): 99-109.

Schneeberger, A., Mercer, C. H., Gregson, S. A., Ferguson, N. M., Nyamukapa, C. A., Anderson, R. M., ... & Garnett, G. P. (2004). Scale-free networks and sexually transmitted diseases: a description of observed patterns of sexual contacts in Britain and Zimbabwe. *Sexually transmitted diseases*, 31(6), 380-387.

Schwarzacher, Walther. *Sedimentation models and quantitative stratigraphy*. Elsevier, 1975.

Shirley, Mark DF, and Steve P. Rushton. "The impacts of network topology on disease spread." *Ecological Complexity* 2.3 (2005): 287-299.

Sutton, Turner B. "Changing options for the control of deciduous fruit tree diseases." *Annual review of phytopathology* 34.1 (1996): 527-547.

Thornley, John HM. *Mathematical models in plant physiology*. Academic Press (Inc.) London, Ltd., 1976.

Traud, Amanda L., Peter J. Mucha, and Mason A. Porter. "Social structure of Facebook networks." *Physica A: Statistical Mechanics and its Applications* 391.16 (2012): 4165-4180.

Trotter, N.J. Gay, W.J. Edmunds, *The natural history of meningococcal carriage and disease*, *Epidemiol Infect.* 134(3) (2006) 556-66.

Van der Leij, M. J. *The Economics of Networks: theory and empirics*. Diss. Erasmus University Rotterdam, 2006.

Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." *nature* 393.6684 (1998): 440-442.

Witten, G., and G. Poulter. "Simulations of infectious diseases on networks." *Computers in Biology and Medicine* 37.2 (2007): 195-205.

Zhang, Juping, and Zhen Jin. "The analysis of epidemic network model with infectious force in latent and infected period." *Discrete Dynamics in Nature and Society* 2010 (2010).

