

**UNIVERSIDAD CARLOS III DE MADRID**

**ESCUELA POLITÉCNICA SUPERIOR**



**PROYECTO FIN DE CARRERA**

**INGENIERÍA TÉCNICA EN INFORMÁTICA DE GESTIÓN**

**DEPARTAMENTO DE ARQUITECTURA DE COMPUTADORES,  
COMUNICACIONES Y SISTEMAS**

**SIMULACIÓN DE LA TRANSMISIÓN DE UNA  
ENFERMEDAD CONTAGIOSA EN ENTORNOS URBANOS**

**Autor:** Gonzalo Martín Cruz  
**Tutores:** David Expósito Singh  
Maria-Cristina Marinescu

Leganés, a 5 de Julio de 2010



## Proyecto Fin de Carrera

**TÍTULO:** Simulación de la transmisión de una enfermedad contagiosa en entornos urbanos

**AUTOR:** Gonzalo Martín Cruz

**TUTORES:** David Expósito Singh  
Maria-Cristina Marinescu

La defensa del presente Proyecto Fin de Carrera se realizó el día 12 de Julio de 2010, siendo calificada por el siguiente tribunal:

**PRESIDENTE:** Jesús Carretero Pérez

**SECRETARIO:** Isabel Segura Bedmar

**VOCAL:** M<sup>a</sup> Soledad Escolar Díaz

Que acuerda otorgarle la calificación de:

**PRESIDENTE:**

**SECRETARIO:**

**VOCAL:**





# Agradecimientos

En primer lugar gracias a mi familia. A los que están y estarán, a los que estuvieron y tristemente se fueron. Especialmente a mis padres, Alicia y Jose Antonio, a mi hermano, Alexis, y a mi novia, Romina, por todos los esfuerzos y sacrificios que habéis realizado en vuestras vidas, vuestros apoyos y ánimo para ayudarme a ser quien soy y de que hoy pueda estar escribiendo estas líneas de agradecimiento.

Quiero acordarme en particular de mis abuelos Josefa, Gonzalo, Sacramento y Enrique. Gracias al resto de mi familia, a mis abuelos, tíos y primos, desde el primero hasta el último, Hugo, el pequeñajo que más quiero: Gracias.

Gracias a todos mis amigos, a los que conocí en mi infancia hace ya más de 20 años y que aún permanecemos unidos, a los que se unieron más tarde pero han sido igual de importantes.

Gracias a todas esas grandes personas que con las que he compartido algunos de estos últimos cinco años de mi vida en la universidad: Julio, Quique, Borja, Almudena, Bea, Javi, Elena, Helen, David, Felipe y a todos aquellos que en algún momento se cruzaron en mi camino.

Gracias también a los tutores de este proyecto, David y Maria-Cristina, por vuestra ayuda, paciencia y comprensión a lo largo de este último año, en los que habéis sido unos verdaderos compañeros, dispuestos siempre a la consecución del mismo y permitiéndome trabajar junto a vosotros como un equipo.

*“Stay hungry, Stay foolish”*  
Stewart Brand (1938)

# Tabla de contenidos

<b>1.Introducción</b>	<b>1</b>
1.1.Motivación	1
1.2.Objetivos	2
1.2.1.Objetivo principal	2
1.2.2.Objetivos específicos	2
1.3.Estructura del documento	3
<b>2.Modelos de representación</b>	<b>4</b>
2.1.Modelo social	4
2.1.1.Estructura de datos	4
2.1.2.Generación del modelo de población	7
2.1.3.Parámetros del modelo social	15
2.2.Modelo de enfermedad	18
2.2.1.Modelo de transición de estados de una enfermedad infecciosa	18
2.2.2.Estructuras de datos	21
2.2.3.Algoritmo de simulación de la transmisión	23
2.2.4.Parámetros del modelo de enfermedad	35
<b>3.Implementación MPI</b>	<b>37</b>
3.1.Modelo de programación de paso de mensajes	37
3.2.Proceso de paralelización	39
3.2.1.Descomposición	41
3.2.2.Asignación	44
3.2.3.Orquestación	46
3.2.4.Mapeo	50
<b>4.Caso de estudio: Gripe A (H1N1) en Leganés</b>	<b>51</b>
4.1.Estudio del problema	51
4.1.1.Selección de la enfermedad contagiosa del estudio	51
4.1.2.Selección de la población del estudio	54

4.2.Parámetros del caso de estudio	56
4.2.1.Parámetros generales	56
4.2.2.Parámetros del modelo social	57
4.2.3.Parámetros del modelo de enfermedad	59
<b>5.Evaluación</b>	<b>61</b>
5.1.Descripción del entorno de evaluación	61
5.2.Evaluación de los supuestos	63
5.2.1.Supuesto 1	63
5.2.2.Supuesto 2	67
5.2.3.Supuesto 3	71
5.2.4.Supuesto 4	75
5.2.5.Supuesto 5	80
5.3.Evaluación del rendimiento de la aplicación	85
5.3.1.Evaluación de rendimiento para el supuesto 1	86
5.3.2.Evaluación de rendimiento para el supuesto 2	90
5.3.3.Evaluación de rendimiento para el supuesto 3	94
5.3.4.Análisis de speedup, eficiencia y escalabilidad	98
<b>6.Conclusiones y trabajo futuro</b>	<b>104</b>
6.1.Conclusiones	104
6.2.Trabajo futuro	106
<b>7.Presupuesto</b>	<b>107</b>
7.1.Cálculo de costes	107
7.1.1.Descripción del proyecto	107
7.1.2.Costes de personal	107
7.1.3.Costes de equipos y tecnología	108
7.1.4.Costes de software	108
7.1.5.Costes de material fungible	108
7.2.Presupuesto	109
<b>8.Anexo I: Manual de usuario</b>	<b>110</b>

8.1.Contenido	110
8.2.Requisitos hardware y software	110
8.2.1.Requisitos hardware de compilación	110
8.2.2.Requisitos software de compilación	110
8.2.3.Requisitos hardware de ejecución	111
8.2.4.Requisitos software de ejecución	111
8.3.Compilación e instalación de la aplicación	111
8.4.Guía rápida sobre el uso y funcionamiento de la aplicación	112
8.5.Resolución de problemas	114
<b>9.Referencias</b>	<b>115</b>

# Lista de Figuras

<i>Ilustración 1.</i> Resultados de [Google Trends] para las búsquedas sobre la Gripe A H1N1.	1
<i>Ilustración 2.</i> Ejemplo de distribución de colectivos sobre una población de 200.000 habitantes.	8
<i>Ilustración 3.</i> Aspecto general del grafo de [Enron Corporation].	9
<i>Ilustración 4.</i> Aspecto general del grafo de [Facebook].	9
<i>Ilustración 5.</i> Vista ampliada del grafo de [Facebook].	10
<i>Ilustración 6.</i> Generación de conocidos dentro del ámbito del colectivo con la función de escalado.	11
<i>Ilustración 7.</i> Modelo de transición de estados del individuo.	19
<i>Ilustración 8.</i> Aspecto general de un grafo completo de población.	25
<i>Ilustración 9.</i> Ampliación [x2] de un grafo completo de población.	25
<i>Ilustración 10.</i> Aspecto general de un grafo tipo 1 de población.	27
<i>Ilustración 11.</i> Ampliación [x4] de un grafo tipo 1 de población.	27
<i>Ilustración 12.</i> Aspecto general de un grafo tipo 2 de población.	29
<i>Ilustración 13.</i> Ampliación [x2] de un grafo tipo 2 de población.	29
<i>Ilustración 14.</i> Aspecto general de un grafo tipo 3 de población.	31
<i>Ilustración 15.</i> Ampliación [x2] de un grafo tipo 3 de población.	31
<i>Ilustración 16.</i> Etapas del proceso de paralelización. [Culler et al, 1999] y [ARCOS, 2009].	39
<i>Ilustración 17.</i> Diagrama de flujo de la aplicación.	40
<i>Ilustración 18.</i> Distribución de tareas sobre un grafo de una población de ejemplo.	41
<i>Ilustración 19.</i> Distribución de un array <i>occupation</i> de ejemplo.	42
<i>Ilustración 20.</i> Distribución de arrays locales.	43
<i>Ilustración 21.</i> Asignación de tareas sobre un grafo de una población de ejemplo.	44
<i>Ilustración 22.</i> Asignación de tareas sobre un array <i>occupation</i> de ejemplo.	45
<i>Ilustración 23.</i> Asignación de arrays locales.	45
<i>Ilustración 24.</i> Diagrama de comunicaciones del modelo social.	46
<i>Ilustración 25.</i> Diagrama de flujo del paso de mensajes del modelo social.	47
<i>Ilustración 26.</i> Diagrama de comunicaciones del modelo enfermedad.	48
<i>Ilustración 27.</i> Ejemplo de operación MPI_Reduce.	49

<i>Ilustración 28.</i> Gráfico de casos positivos reportados en Estados Unidos entre las semanas 35 de 2009 y 23 de 2010. [CDC, 2009].	52
<i>Ilustración 29.</i> Gráfico de hospitalizaciones y defunciones reportadas en Estados Unidos entre las semanas 35 de 2009 y 13 de 2010. [CDC, 2009].	53
<i>Ilustración 30.</i> [Kasukabe] por [ARCOS, 2009].	61
<i>Ilustración 31.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en estado sano para el supuesto 1.	64
<i>Ilustración 32.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en estado sano para el supuesto 1.	64
<i>Ilustración 33.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 1.	65
<i>Ilustración 34.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 1.	65
<i>Ilustración 35.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en estado sano para el supuesto 2.	68
<i>Ilustración 36.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en estado sano para el supuesto 2.	68
<i>Ilustración 37.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 2.	69
<i>Ilustración 38.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 2.	69
<i>Ilustración 39.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en estado sano para el supuesto 3.	73
<i>Ilustración 40.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en estado sano para el supuesto 3.	73
<i>Ilustración 41.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 3.	75
<i>Ilustración 42.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 3.	75
<i>Ilustración 43.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor $media\_sp = 30$ .	76

<i>Ilustración 44.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el valor $media\_sp = 30$ .	76
<i>Ilustración 45.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor $media\_sp = 90$ .	78
<i>Ilustración 46.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el valor $media\_sp = 90$ .	78
<i>Ilustración 47.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor $media\_sp = 120$ .	80
<i>Ilustración 48.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el valor $media\_sp = 120$ .	80
<i>Ilustración 49.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación A.	82
<i>Ilustración 50.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación A.	82
<i>Ilustración 51.</i> Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación B.	83
<i>Ilustración 52.</i> Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación B.	83
<i>Ilustración 53.</i> Análisis de speedup para el supuesto 1.	98
<i>Ilustración 54.</i> Análisis de speedup para el supuesto 2.	100
<i>Ilustración 55.</i> Análisis de speedup para el supuesto 3.	102

# Lista de Tablas

<i>Tabla 1. "Distribución de grupos quinquenales según sexos" del [Ayto de Leganés, 2010]</i>	55
<i>Tabla 2. Distribución de población según edad sobre colectivos de interés, con datos del [Ayto de Leganés, 2010].</i>	56
<i>Tabla 3. Distribución de población sobre colectivos de interés con datos del [Ayto de Leganés, 2010] e [INE, 2001-2009].</i>	56
<i>Tabla 4. Resultados estadísticos de la evaluación del supuesto 1.</i>	64
<i>Tabla 5. Resultados estadísticos de la evaluación del supuesto 2.</i>	68
<i>Tabla 6. Resultados estadísticos de la evaluación del supuesto 3.</i>	72
<i>Tabla 7. Resultados estadísticos de la evaluación del supuesto 4.</i>	76
<i>Tabla 8. Resultados estadísticos de la evaluación del supuesto 5.</i>	82
<i>Tabla 9. Tiempo total de ejecución para el supuesto 1 con un proceso.</i>	86
<i>Tabla 10. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con un proceso.</i>	86
<i>Tabla 11. Tiempo total de ejecución para el supuesto 1 con dos procesos.</i>	87
<i>Tabla 12. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con dos procesos.</i>	87
<i>Tabla 13. Tiempo total de ejecución para el supuesto 1 con cuatro procesos.</i>	88
<i>Tabla 14. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con cuatro procesos.</i>	88
<i>Tabla 15. Tiempo total de ejecución para el supuesto 1 con ocho procesos.</i>	89
<i>Tabla 16. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con ocho procesos.</i>	89
<i>Tabla 17. Tiempo total de ejecución para el supuesto 2 con un proceso.</i>	90
<i>Tabla 18. Tiempo de ejecución de la rutina de propagación para el supuesto 2 con un proceso.</i>	90
<i>Tabla 19. Tiempo total de ejecución para el supuesto 2 con dos procesos.</i>	91
<i>Tabla 20. Tiempo de ejecución de la rutina de propagación para el supuesto 2 con dos procesos.</i>	91
<i>Tabla 21. Tiempo total de ejecución para el supuesto 2 con cuatro procesos.</i>	92



<i>Tabla 22.</i> Tiempo de ejecución de la rutina de propagación para el supuesto 2 con cuatro procesos.	92
<i>Tabla 23.</i> Tiempo total de ejecución para el supuesto 2 con ocho procesos.	93
<i>Tabla 24.</i> Tiempo de ejecución de la rutina de propagación para el supuesto 2 con ocho procesos.	93
<i>Tabla 25.</i> Tiempo total de ejecución para el supuesto 3 con un proceso.	94
<i>Tabla 26.</i> Tiempo de ejecución de la rutina de propagación para el supuesto 3 con un proceso.	94
<i>Tabla 27.</i> Tiempo total de ejecución para el supuesto 3 con dos procesos.	95
<i>Tabla 28.</i> Tiempo de ejecución de la rutina de propagación para el supuesto 3 con dos procesos.	95
<i>Tabla 29.</i> Tiempo total de ejecución para el supuesto 3 con cuatro procesos.	96
<i>Tabla 30.</i> Tiempo de ejecución de la rutina de propagación para el supuesto 3 con cuatro procesos.	96
<i>Tabla 31.</i> Tiempo total de ejecución para el supuesto 3 con ocho procesos.	97
<i>Tabla 32.</i> Tiempo de ejecución de la rutina de propagación para el supuesto 3 con ocho procesos.	97
<i>Tabla 33.</i> Tabla de análisis de speedup y eficiencia del supuesto 1.	98
<i>Tabla 34.</i> Tabla de análisis de speedup y eficiencia de la rutina de propagación del supuesto 1.	99
<i>Tabla 35.</i> Tabla de análisis de speedup y eficiencia del supuesto 2.	100
<i>Tabla 36.</i> Tabla de análisis de speedup y eficiencia de la rutina de propagación del supuesto 2.	101
<i>Tabla 37.</i> Tabla de análisis de speedup y eficiencia del supuesto 3.	102
<i>Tabla 38.</i> Tabla de análisis de speedup y eficiencia de la rutina de propagación del supuesto 3.	103
<i>Tabla 39.</i> Costes de personal.	107
<i>Tabla 40.</i> Costes de equipos y tecnología.	108
<i>Tabla 41.</i> Costes de software.	108
<i>Tabla 42.</i> Costes de material fungible.	108
<i>Tabla 43.</i> Presupuesto total.	109

# Introducción

*En este primer capítulo introductorio se ofrece una visión global del proyecto realizado y documentado en la presente memoria, incluyendo las motivaciones que conllevaron a su desarrollo y los objetivos perseguidos.*

## Motivación

En la primavera de 2009 se produce la alerta sanitaria en Norteamérica a causa del brote del virus de la gripe porcina o, como es científicamente conocida, gripe A (H1N1). El 11 de junio de 2009 la OMS eleva el nivel de alerta por la gripe porcina a fase 6, la más elevada, lo que indica la expansión del virus por contagio en el contacto humano y presencia del mismo en al menos un país distinto de la región en la que se inició el brote.

La declaración de la fase 6 significa la puesta en marcha de una pandemia mundial del virus, produciéndose una rápida propagación geográfica y provocando temores en la población debido a que el ataque del virus no se producía sobre los grupos de riesgo habituales, los ancianos, sino sobre individuos jóvenes que gozan de buena salud, mujeres embarazadas o enfermos crónicos.

A causa de estos hechos se generó un clima de alarma social y un problema de salud pública, dado que las autoridades sanitarias se encontraban desconcertadas ante la actividad del virus y la falta de medios para controlar su propagación. Un vistazo sobre el historial de búsquedas de [GoogleTrends] nos ofrece estadísticas ligadas al estado de preocupación de la población durante las estaciones de primavera y verano de 2009:

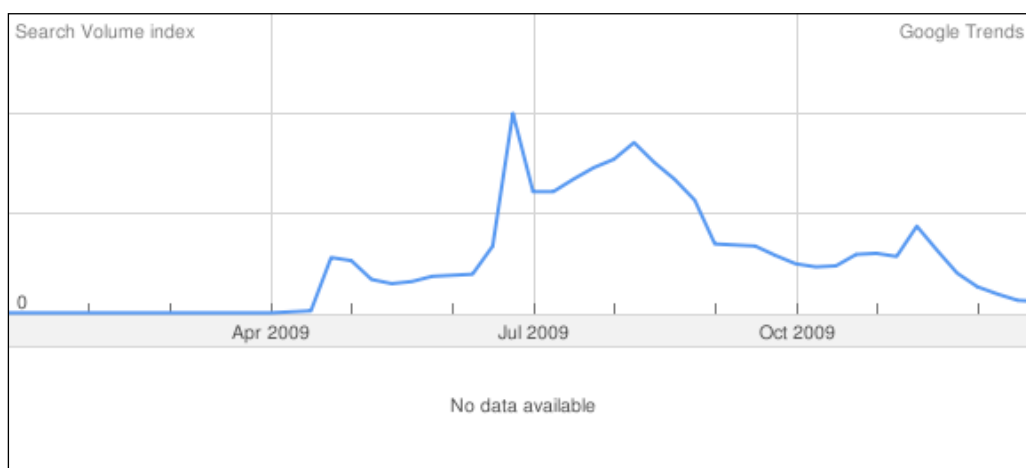


Ilustración 1. Resultados de [GoogleTrends] para las búsquedas sobre la Gripe A H1N1.

Debido al estado global existente en el verano de 2009 se propone el desarrollo del presente proyecto, para tratar de dar respuesta y generar previsiones futuras sobre el comportamiento del virus de la gripe A (H1N1) en entornos urbanos.

## Objetivos

Una vez expuestos los motivos que conllevaron al desarrollo del proyecto se establecen una serie de objetivos generales y específicos que deben proporcionar las conclusiones adecuadas a la finalización del mismo.

### Objetivo principal

El objetivo principal del proyecto es modelar la propagación del virus de la gripe A (H1N1) en un entorno de población urbana, con el fin de comprender el modelo de propagación bajo distintos escenarios.

### Objetivos específicos

Además del objetivo principal surgen una serie de objetivos específicos a tratar durante el desarrollo del proyecto:

- **Modelado del colectivo social y de la enfermedad:** se debe modelar matemáticamente el colectivo social de una población, lo que incluye individuos, colectivos e interacciones entre los mismos. Por otro lado se debe modelar la enfermedad contagiosa, lo que incluye definición del virus y propagación del mismo sobre la población.
- **Desarrollo de una herramienta de simulación distribuida:** se debe desarrollar una herramienta sobre la que se realicen simulaciones de la propagación del virus sobre el colectivo social de la población. La herramienta, además, ha de ser distribuida. Dicha herramienta ha de presentar las siguientes características:
  - **Flexibilidad:** ha de ser una herramienta flexible que se adapte a diferentes poblaciones y enfermedades contagiosas, es decir, ser capaz de modelar distintos entornos.
  - **Escalabilidad:** la aplicación debe ser ejecutada en paralelo de forma eficiente, permitiendo reducir su tiempo de ejecución conforme más recursos de cómputo son empleados. Del mismo modo, debe ser escalable, es decir, debe permitir simular de forma eficiente entornos de mayor complejidad y tamaño.
  - **Realista:** se pretende que la herramienta realice simulaciones de manera realista, hecho que se comprobará en función de los resultados obtenidos.
- **Evaluación de la simulación de la propagación del virus:** por último se debe realizar la correspondiente evaluación de la simulación de la propagación del virus sobre la población considerada, Leganés.

## Estructura del documento

Para facilitar la lectura del documento se ofrece a continuación un breve resumen descriptivo del contenido de cada una de las secciones:

- **Capítulo 1: Introducción.** Capítulo introductorio que sirve como presentación general del proyecto, motivaciones que han conllevado a su desarrollo y objetivos del mismo.
- **Capítulo 2: Modelos de representación.** Descripción de los modelos de representación de la aplicación, modelo social y modelo de enfermedad.
- **Capítulo 3: Implementación MPI.** Descripción del modelo de programación paralela seleccionado para el desarrollo.
- **Capítulo 4: Caso de estudio: Gripe A en Leganés.** Descripción del caso de estudio analizado, con agente infeccioso y municipio concretos.
- **Capítulo 5: Evaluación.** Evaluación de la aplicación sobre diferentes supuestos de simulación.
- **Capítulo 6: Presupuesto.** Descripción detallada del presupuesto del proyecto.
- **Anexo I: Manual de usuario.** Manual de usuario de la aplicación, que contiene una guía rápida sobre compilación, instalación, configuración y ejecución de la herramienta distribuida.

# Modelos de representación

*En este apartado se detallan los modelos de representación existentes: el modelo social que da forma a la población y el modelo de enfermedad utilizado para simular el comportamiento de una infección.*

## Modelo social

### Estructura de datos

La estructura de datos elegida para representar los grafos que maneja la aplicación es una matriz de adyacencia dispersa no estructurada en esquema *Compressed Sparse Column (CSC o CCS)*. Por matriz dispersa se conoce a las matrices generalmente de grandes dimensiones cuya mayoría de elementos son nulos, se encuentran diseminados sobre la matriz, y sobre las que es posible aplicar una serie de técnicas especiales para obtener ventaja de su principal característica, la gran cantidad de elementos nulos que contienen. En el caso concreto de las matrices dispersas no estructuradas nos referimos a ellas cuando los elementos no nulos que contienen se distribuyen sobre la matriz de forma irregular, sin formar bloques o patrones regulares.

### Esquema de representación coordinado

La estructura de datos elemental de almacenamiento de una matriz es un array bi-dimensional en el que cada entrada del mismo representa un elemento  $a_{ij}$  de la matriz y al que se accede a través de los índices de fila y columna del mismo. Para una matriz de dimensiones  $m$  filas y  $n$  columnas será necesario disponer en memoria de  $m \times n$  entradas para representar toda la información, entre la que estamos desperdiciando la necesaria para almacenar los elementos nulos de la matriz. En cambio las matrices dispersas utilizan un esquema de representación coordinado que utiliza tres vectores de dimensión  $nnz$  (*number of non zero elements*) o número de elementos no nulos:

- **AA**: almacena los elementos no nulos de la matriz. Vector de elementos en punto flotante de dimensión  $nnz$ .
- **IA**: almacena los índices de fila asociados a los elementos no nulos del vector AA. Vector de elementos enteros de dimensión  $nnz$ .
- **JA**: almacena los índices de columna asociados a los elementos no nulos del vector AA. Vector de elementos enteros de dimensión  $nnz$ .

Tomemos la siguiente matriz de ejemplo  $A$  representada en formato convencional:

$$A = \begin{pmatrix} 1 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 \\ 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 2 \\ 0 & 0 & 3 & 7 & 9 \end{pmatrix}$$

El almacenamiento de dicha matriz en formato de representación coordinado en los vectores  $AA$ ,  $IA$  y  $JA$  es el siguiente:

$$AA = [ 1 \ 4 \ 5 \ 3 \ 7 \ 8 \ 7 \ 2 \ 9 ]$$

$$IA = [ 1 \ 3 \ 1 \ 5 \ 2 \ 4 \ 5 \ 4 \ 5 ]$$

$$JA = [ 1 \ 1 \ 3 \ 3 \ 4 \ 4 \ 4 \ 5 \ 5 ]$$

Este esquema de representación ya presenta suficientes ventajas respecto a la estructura de datos elemental de almacenamiento de matrices dispersas, ya que reduciendo el número de elementos de la matriz que es necesario almacenar se está reduciendo de forma consecuente desde el punto de vista computacional la cantidad de memoria necesaria para mantener dichos elementos.

Para este ejemplo se ha realizado una ordenación de los elementos de acuerdo al índice de columna al que pertenecen (según se puede observar en el vector  $JA$ ), aunque esta condición no es necesaria ya que los elementos pueden introducirse en los vectores de modo arbitrario, pero esta condición sirve como punto de partida al formato de representación *Compressed Sparse Column* o *CSC*.

### Esquema de representación *Compressed Sparse Column*

El formato *CSC* realiza una compresión eliminando información redundante del vector  $JA$  del esquema coordinado siempre y cuando todos sus elementos estén ordenados por columnas:

- **AA**: almacena los elementos no nulos de la matriz. Vector de elementos en punto flotante de dimensión  $nnz$  (número de elementos no nulos).
- **IA**: almacena los índices de fila asociados a los elementos no nulos del vector  $AA$ . Vector de elementos enteros de dimensión  $nnz$  (número de elementos no nulos).
- **JA**: es reemplazado por un vector de elementos enteros de dimensión  $n+1$  (siendo  $n$  el número de columnas de la matriz) que almacena punteros al comienzo de los elementos de cada columna con la siguiente estructura:

$$\begin{aligned} JA(1) &= 1, \\ JA(i+1) - JA(i) &= \text{número de elementos no nulos de la columna } i. \end{aligned}$$

Para obtener el valor de los elementos nos guiaremos por los punteros  $JA(i)$  y  $JA(i+1)$ , que delimitan el comienzo y final de los elementos de la columna  $i$  en los vectores  $AA$  e  $IA$ . Es por lo tanto condición necesaria para este esquema de representación que los elementos no nulos de la matriz se introduzcan de forma ordenada por columnas en los vectores de almacenamiento. De esta manera, la representación de la matriz  $A$  anterior en esquema  $CSC$  sería:

$$AA = [ 1 \ 4 \ 5 \ 3 \ 7 \ 8 \ 7 \ 2 \ 9 ]$$

$$IA = [ 1 \ 3 \ 1 \ 5 \ 2 \ 4 \ 5 \ 4 \ 5 ]$$

$$JA = [ 1 \ 3 \ 3 \ 5 \ 8 \ 10 ]$$

Así podemos obtener los elementos de cada una de las columnas de la matriz con la función expresada anteriormente. Por otra parte, existe un esquema de compresión análogo a  $CSC$  pero que realiza la compresión en función de las filas de la matriz, denominado *Compressed Sparse Row (CSR o CRS)*.

### **Ventajas del esquema *Compressed Sparse Column***

Las principales ventajas que se pretenden potenciar utilizando el esquema  $CSC$  respecto al formato convencional y esquema coordenado tratan de preservar:

- Minimización del consumo de memoria.
- Minimización del uso de procesador.
- Minimización del tiempo de acceso a los datos.
- Localidad de los datos.

### **Matrices dispersas en lenguaje de programación C: librería '*CSparse*'**

Para manejar las matrices dispersas comprimidas en formato  $CSC$  desde el lenguaje de programación C se ha recurrido a la librería de código libre bajo licencia *GNU Lesser General Public License 'CSparse'* [Davis, 2006] que implementa métodos directos para manejar sistemas lineales dispersos.

Las rutinas de esta librería utilizadas para gestionar las matrices dispersas son las siguientes:

- ***CS\_ADD***: realiza la suma de dos matrices dispersas  $A$  y  $B$ .
- ***CS\_TRIPLET***: convierte una matriz dispersa en esquema coordenado a  $CSC$ .

## Generación del modelo de población

### Generación de conexiones

Los individuos de las poblaciones que se modelan para realizar las evaluaciones de las simulaciones se distribuyen en varios colectivos que pertenecerán a uno de los siguientes tipos: trabajadores (*T*), estudiantes (*E*), ancianos (*A*) y amas de casa (*AC*). Para cada uno de estos tipos de colectivos existirán varios parámetros: límites de edad mínima y máxima de los individuos pertenecientes al colectivo; límites de tamaño mínimo y máximo entre los que se comprenderá el del colectivo generado.

Dicha información está contenida en un fichero de configuración de entrada, "*config\_grafos.xml*", desde el cuál es posible cargar los parámetros de configuración de los tipos de colectivos para modelar distintas poblaciones y que cuenta además con una medida del porcentaje de población que pertenecerá a cada tipo. Para generar el grafo resultado del modelado de una determinada población es necesario proveer a la aplicación de dicho archivo de configuración además del tamaño de la misma. Con dicha información se genera un fichero intermedio que contiene la distribución de la población que se va a simular: número de colectivos, tipo de cada uno de los colectivos y tamaño de cada uno de los colectivos.

Una vez que se dispone del fichero intermedio de colectivos la aplicación se encarga de generar el grafo o matriz de adyacencia en el que se representan las diferentes conexiones entre los individuos de la población. En este caso la existencia de una conexión entre dos elementos del grafo (representado como una entrada en el mismo) significa que existe una relación entre ambos: son conocidos. Los conocidos de cada uno de los individuos se distribuyen a su vez en dos grupos: conocidos dentro del ámbito del colectivo y conocidos fuera del ámbito del colectivo, entre los que se incluyen los familiares.

En la ilustración 2 se muestra el modelado de una población de 200.000 habitantes distribuida en dos colectivos de trabajadores (*T*) y uno de amas de casa (*AC*), ancianos (*A*) y estudiantes (*E*).

Notar que en esta representación cada colectivo aparece como una submatriz densa, es decir, en la que todos los miembros del colectivo tienen conectividad con todos y en la que no existe conectividad entre miembros de distintos colectivos. En un caso real estas submatrices no son densas sino dispersas y existe conectividad entre miembros de distintos colectivos. En las siguientes secciones se describe el procedimiento empleado.



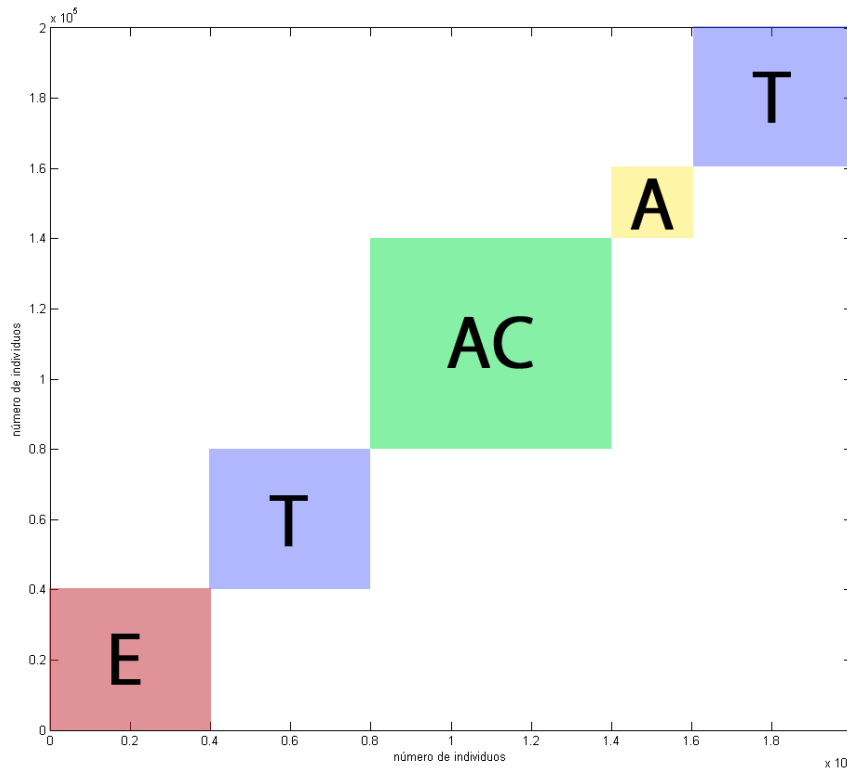


Ilustración 2. Ejemplo de distribución de colectivos sobre una población de 200.000 habitantes.

### Generación de conexiones dentro del ámbito del colectivo

Para generar los conocidos dentro del ámbito del colectivo para cada uno de los individuos de la población de la simulación se realiza un escalado sobre varios grafos que contienen datos de poblaciones reales. Dichos grafos son extraídos de la empresa energética [Enron Corporation] y de la red social [Facebook] y contienen las conexiones entre sus individuos en formato CSC. A partir de este momento nos referiremos a dichos grafos como *grafos origen*. Según la población que se desee modelar se puede emplear cada uno de los grafos origen para generar unos u otros tipos de colectivos.

En la ilustración 3 se puede observar la matriz de adyacencia de [Enron Corporation] mientras que en las ilustraciones 4 y 5 se muestra la de [Facebook] con diferentes ampliaciones:

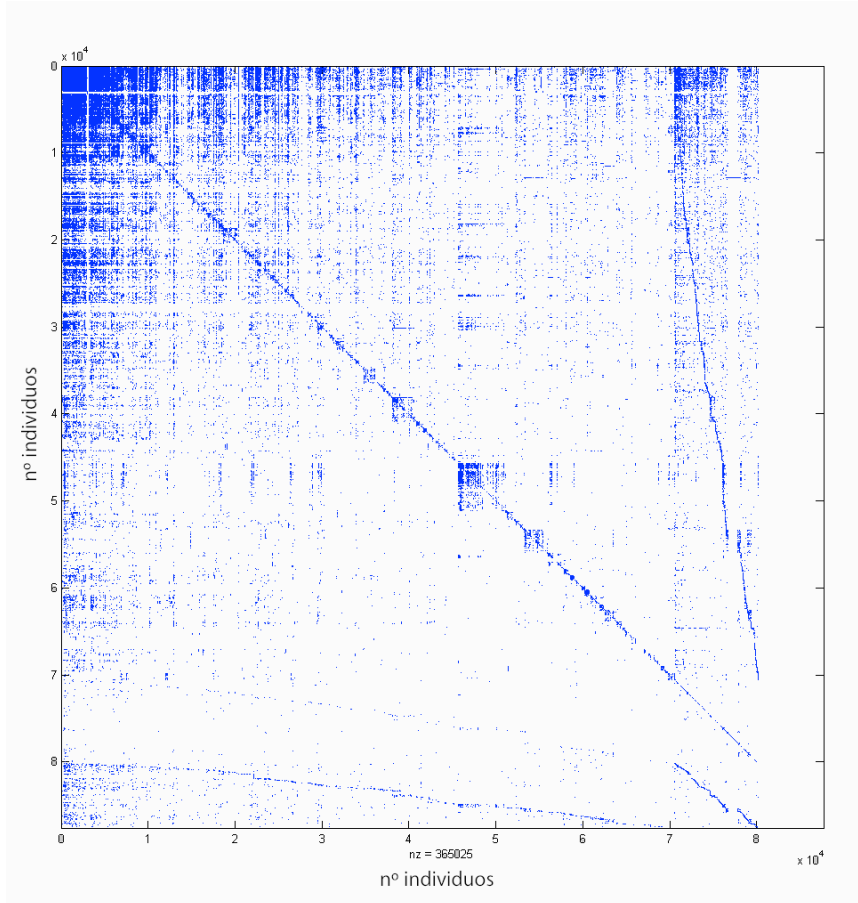


Ilustración 3. Aspecto general del grafo de [Enron Corporation].

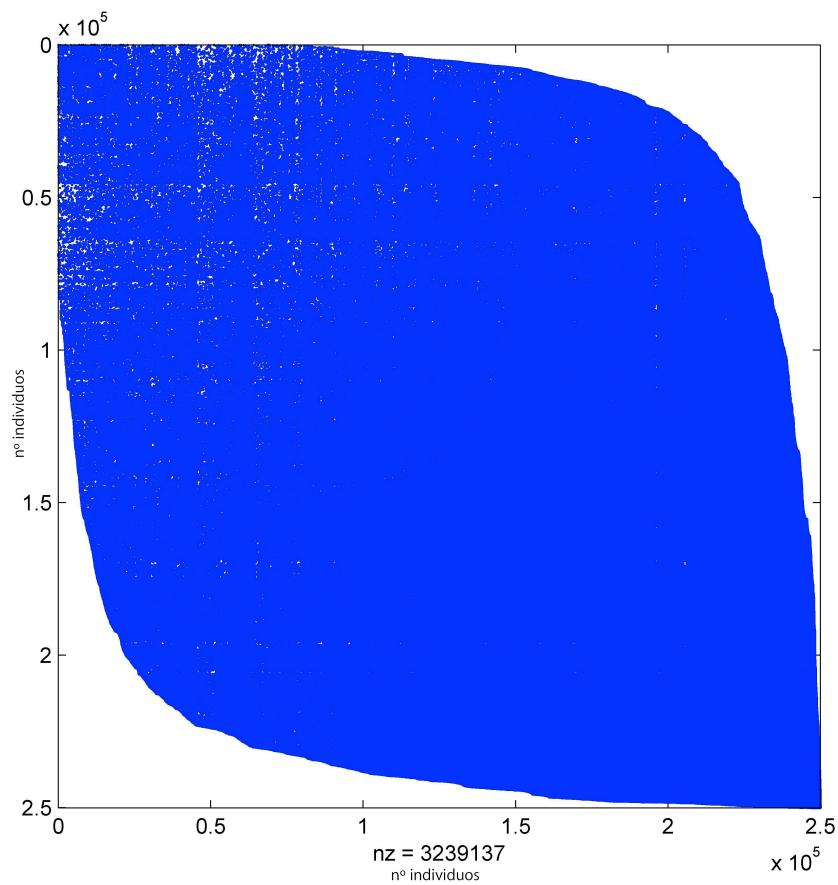


Ilustración 4. Aspecto general del grafo de [Facebook].

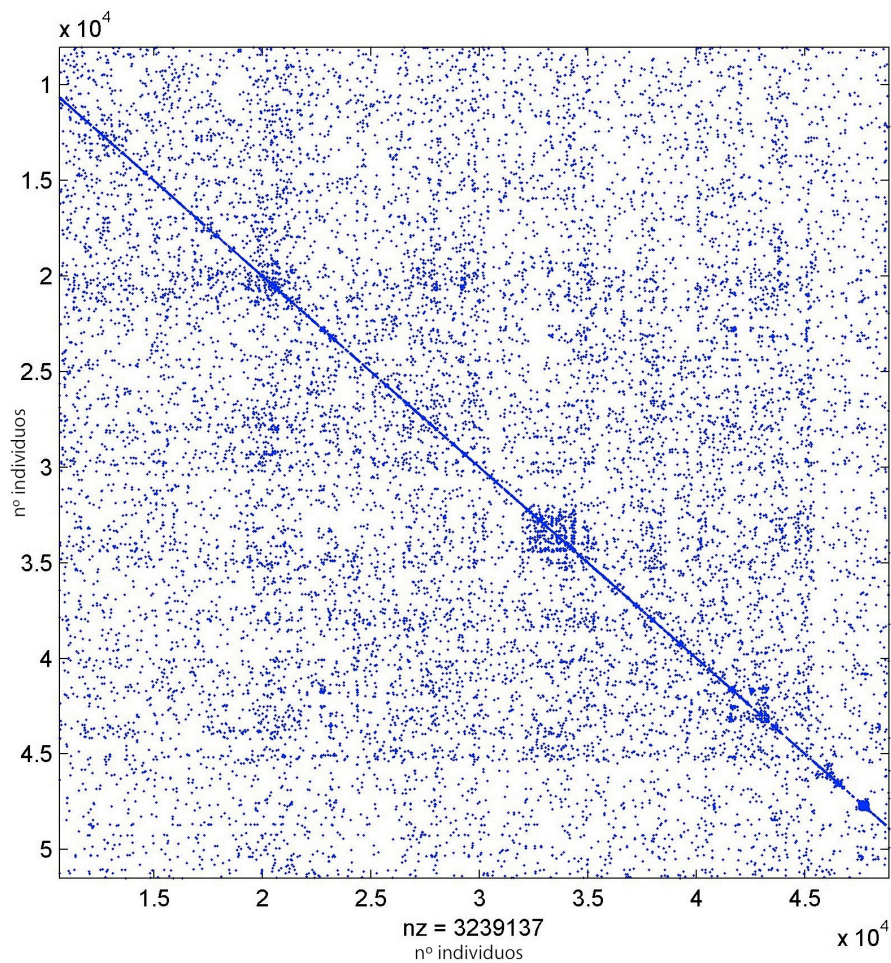


Ilustración 5. Vista ampliada del grafo de [Facebook].

Los parámetros  $t\_grafo$ ,  $e\_grafo$ ,  $a\_grafo$  y  $ac\_grafo$  pertenecientes al fichero de configuración "*config\_grafos.xml*", y del cuál se hablará más adelante, nos permiten definir qué grafo origen utilizar para generar cada tipo de colectivo, pudiendo seleccionar uno de los dos disponibles, Enron (1) o Facebook (2).

El algoritmo de generación de conocidos dentro del ámbito del colectivo toma cada uno de los individuos de la población y, en función del tipo del colectivo al que pertenece, obtiene mediante una función de probabilidad uno de los individuos del grafo origen correspondiente y su número de conocidos. Para escalar el número de conocidos del grafo origen al colectivo del individuo se comprueba si dicho número es menor o igual que el tamaño del colectivo multiplicado por un *umbral* (ver descripción de parámetros del modelo social) y en caso contrario se obtiene un nuevo individuo distinto del grafo origen.

El pseudocódigo del algoritmo de generación de conocidos dentro del ámbito del colectivo para cada uno de los colectivos existentes en la población es el que sigue a continuación:

**//Generar conexiones dentro del ámbito del colectivo**

```

Desde individuo = 0 hasta individuo = num_individuos_colectivo
    Hacer
        Seleccionar individuo del grafo origen al azar
        Obtener el número de conocidos del individuo del grafo origen(num_con)
        Mientras ( num_individuos_colectivo * umbral < num_con )
    Fin Desde
    
```

Posteriormente, cada uno de los conocidos del individuo del grafo origen se escala desde este al tamaño del colectivo al que pertenece el individuo destino. El grafo origen contiene la representación de los individuos pertenecientes a un único colectivo y sus respectivas conexiones con el resto de individuos del mismo.

En la ilustración 6 se muestra con un ejemplo gráfico cómo se realizaría la función de escalado de un determinado individuo desde el grafo origen. En el lado izquierdo tenemos situada la representación de los dos grafos origen disponibles y en el lado derecho la población objeto de simulación que estamos generando. Notar que, al igual que ocurría anteriormente, en esta representación se muestra cada colectivo como una submatriz densa, cuando en realidad se trata de matrices dispersas y, en el caso de la población generada, con conexiones entre los miembros de distintos colectivos. En este ejemplo se puede observar cómo se utiliza cada uno de los grafos origen para generar los colectivos que componen la población, proceso en el que la función cuyo algoritmo ha sido previamente descrito es la que se encarga de tomar de forma aleatoria un individuo del grafo origen y escalar sus conexiones al tamaño del colectivo al que pertenece el individuo que se está generando.

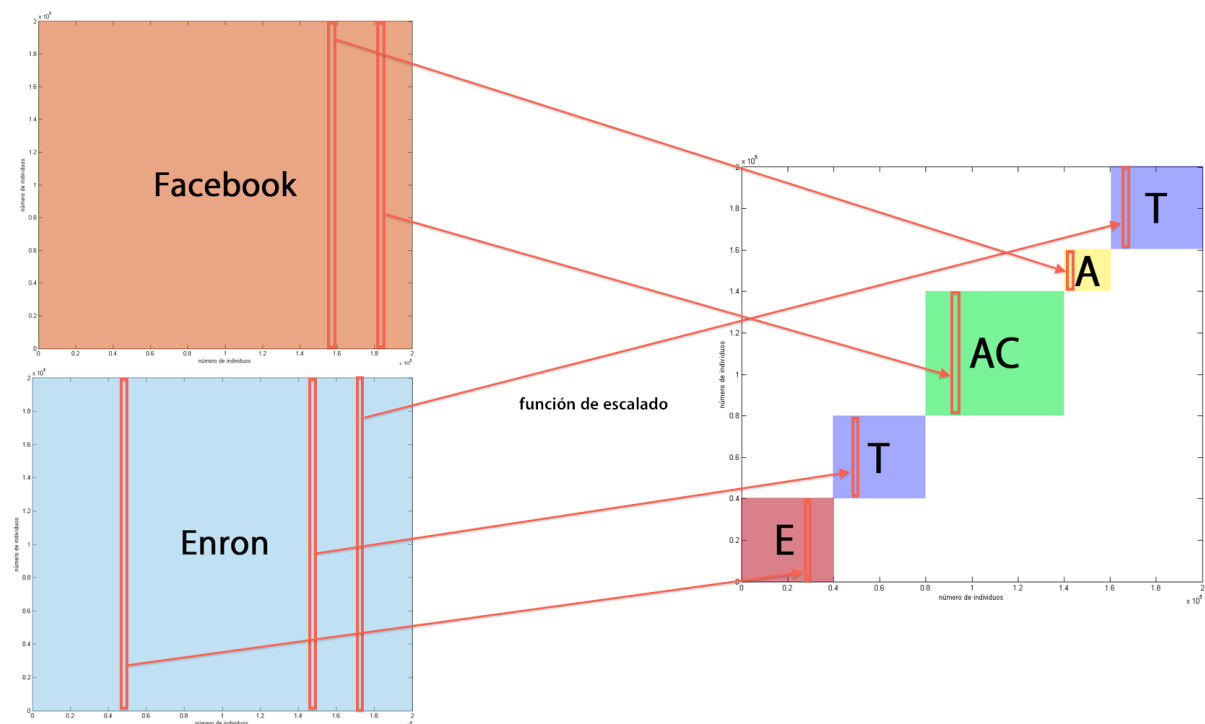


Ilustración 6. Generación de conocidos dentro del ámbito del colectivo con la función de escalado.

## Generación de conexiones fuera del ámbito del colectivo

La generación de los conocidos fuera del ámbito del colectivo se realiza a nivel de colectivo en base a una serie de valores porcentuales que representan el tanto por ciento de conexiones existentes entre dos colectivos de un determinado tipo. Tendremos dos grados de conexiones que se lanzarán de manera aleatoria entre pares de colectivos de toda la población, conexiones **fuertes** y **débiles**.

Los colectivos unidos por conexiones de grado fuerte indican que los individuos de un par de colectivos están altamente ligados, es decir, existe un elevado número de conexiones entre los individuos que los componen. En cambio, las conexiones débiles indican que un par de colectivos están débilmente ligados, existiendo un determinado número base de conexiones entre sus individuos siempre inferior al de una conexión fuerte.

Lo que se pretende modelar es la existencia de distintos vínculos entre ciertos colectivos. Por ejemplo, entre dos grupos de alumnos que comparten el mismo colegio existen muchos vínculos entre ambos colectivos, por lo que existirán conexiones fuertes. Otro ejemplo sería el de dos grupos de trabajadores que comparten las mismas instalaciones.

Por contra, las conexiones débiles reflejan pequeños vínculos entre distintos colectivos. Un ejemplo sería un vínculo débil entre dos grupos de estudiantes que pertenecen a distintos colegios, en el que el número de vínculos existentes se reduce a unos pocos miembros de cada colectivo.

El algoritmo de generación de conexiones liga pares de colectivos en función de unos parámetros de probabilidad existentes para conexiones débiles y fuertes. Debido a dicha probabilidad un par de colectivos determinado puede estar ligado por una conexión fuerte, una conexión débil o simplemente no estar ligado, es decir, no existen conexiones entre los individuos de ambos colectivos.

La probabilidad de que un par de colectivos esté fuerte o débilmente ligado está definida por los parámetros `prob_grupo_fuerte` y `prob_grupo_debil` respectivamente, pertenecientes al fichero de configuración "`config.xml`" y que se detallará más adelante. El algoritmo se encarga de generar una probabilidad para cada uno de los pares de colectivos existentes en la población y determinar, en función del valor de dicha probabilidad y de los de los parámetros indicados, si estarán fuerte ó débilmente ligados.

El número de conexiones entre los pares de colectivos en función del grado de conectividad (fuerte o débil) está definido por el tamaño de los colectivos y una serie de valores porcentuales que indican qué tanto por ciento de individuos de cada tipo de colectivo permanecerán conectados.

Esta serie de valores están almacenados en formato *CSC* en dos matrices, para valores de conexiones fuertes y débiles respectivamente, en las que se representa cada uno de los tipos de colectivos (estudiantes, trabajadores, ancianos y amas de casa) con una fila y columna de la misma. De esta forma, para conocer el porcentaje de conexiones entre dos colectivos se obtiene el tipo al que pertenecen, que nos permitirá acceder a una determinada fila y columna de la matriz para recoger el valor porcentual. Este valor porcentual se multiplicará por el tamaño del colectivo con menor número de individuos, lo que nos ofrecerá un valor entero que será el número de conexiones existentes entre ambos colectivos.

**//Determinar conectividad fuerte**

```
Desde colectivoA = 0 hasta colectivo = total_colectivos
    Desde colectivoB = 0 hasta colectivo = total_colectivos
        //Nota: si (colectivoA != colectivoB)
        Si ( probabilidad < prob_grupo_fuerte ) {
            Obtener grafo de tamaño mínimo (colectivoA, colectivoB)
            Obtener porcentaje de conexiones entre tipo_colectivoA y
            tipoColectivoB para conexión tipo fuerte
            Número de conexiones = porcentaje de conexiones * tamaño de
            grafo mínimo
            Generar número de conexiones al azar entre individuos de
            ambos colectivos
        }
    Fin Desde
Fin Desde
```

**//Determinar conectividad débil**

```
Desde colectivoA = 0 hasta colectivo = total_colectivos
    Desde colectivoB = 0 hasta colectivo = total_colectivos
        //Nota: si (colectivoA != colectivoB)
        Si ( probabilidad < prob_grupo_débil ) {
            Obtener grafo de tamaño mínimo (colectivoA, colectivoB)
            Obtener porcentaje de conexiones entre tipo_colectivoA y
            tipoColectivoB para conexión tipo débil
            Número de conexiones = porcentaje de conexiones * tamaño de
            grafo mínimo
            Generar número de conexiones al azar entre individuos de
            ambos colectivos
        }
    Fin Desde
Fin Desde
```



### **Generación de conexiones familiares**

La generación de los conocidos familiares de cada uno de los individuos se realiza obteniendo otros tres individuos de la población de manera aleatoria ya sean de dentro o fuera del ámbito del colectivo. Con esto se pretende representar una unidad familiar compuesta por cuatro miembros.

### **Grado de riesgo de las conexiones**

Existen dos tipos de conexiones entre los individuos de los distintos colectivos que componen la población de la simulación. Estos dos tipos de conexiones atienden a dos diferentes grados de riesgo, alto y bajo respectivamente. Anticipándonos al modelo de representación de la enfermedad diremos que el riesgo representa el grado de vulnerabilidad de la conexión a servir como canal de transmisión de la infección, que será mayor en las conexiones de riesgo alto y menor en las conexiones de riesgo bajo.

Con esta característica se pretende dar forma a la idea de que entre todas las conexiones que tiene un determinado individuo sobre el total de la población existirá un porcentaje concreto de ellas (conexiones de riesgo alto) que serán con las que el individuo coincida durante más horas del día, ya sea por cuestión de afinidad, familiar, laboral, etc. Por otro lado existirá un porcentaje mayor de conexiones que serán de riesgo bajo, los conocidos del individuo con los que mantiene contacto más o menos esporádico al cabo de su actividad diaria.

## Parámetros del modelo social

Los parámetros que intervienen en la creación del modelo social de la aplicación pueden distribuirse según el ámbito de aplicación al que pertenezcan:

### Parámetros generales del modelo

Existe una serie de parámetros generales cuya finalidad es definir la composición del modelo social de la población que se encuentran contenidos en el fichero de configuración "*config.xml*" y son los siguientes:

Identificador	Descripción
pop_sim	tamaño total de la población de la simulación
avgn	estimación del número medio de conexiones de cada individuo de la población
umbral	umbral de número de conocidos utilizado para realizar el escalado desde el grafo origen (en valor porcentual)

### Parámetros de generación del fichero intermedio de colectivos

Los parámetros que intervienen en la generación del fichero intermedio de colectivos se encuentran contenidos en el fichero de configuración "*config\_grafos.xml*" y son los siguientes:

Identificador	Descripción
num_grafos	número de grafos origen utilizados
t_grafo	identificador del grafo origen del tipo de colectivo de Trabajadores
t_edad_min	edad mínima de los individuos del colectivo de Trabajadores
t_edad_max	edad máxima de los individuos del colectivo de Trabajadores
t_perc	porcentaje de población que pertenecerá al colectivo Trabajadores
t_grupo_min	tamaño mínimo de un determinado colectivo de tipo Trabajadores
t_grupo_max	tamaño máximo de un determinado colectivo de tipo Trabajadores
e_grafo	identificador del grafo origen del tipo de colectivo de Estudiantes
e_edad_min	edad mínima de los individuos del colectivo de Estudiantes
e_edad_max	edad máxima de los individuos del colectivo de Estudiantes
e_perc	porcentaje de población que pertenecerá al colectivo Estudiantes
e_grupo_min	tamaño mínimo de un determinado colectivo de tipo Estudiantes
e_grupo_max	tamaño máximo de un determinado colectivo de tipo Estudiantes



<b>Identificador</b>	<b>Descripción</b>
<code>a_grafo</code>	identificador del grafo origen del tipo de colectivo de Ancianos
<code>a_edad_min</code>	edad mínima de los individuos del colectivo de Ancianos
<code>a_edad_max</code>	edad máxima de los individuos del colectivo de Ancianos
<code>a_perc</code>	porcentaje de población que pertenecerá al colectivo Ancianos
<code>a_grupo_min</code>	tamaño mínimo de un determinado colectivo de tipo Ancianos
<code>a_grupo_max</code>	tamaño máximo de un determinado colectivo de tipo Ancianos
<code>ac_grafo</code>	identificador del grafo origen del tipo de colectivo de Amas de casa
<code>ac_edad_min</code>	edad mínima de los individuos del colectivo de Amas de casa
<code>ac_edad_max</code>	edad máxima de los individuos del colectivo de Amas de casa
<code>ac_perc</code>	porcentaje de población que pertenecerá al colectivo Amas de casa
<code>ac_grupo_min</code>	tamaño mínimo de un determinado colectivo de tipo Amas de casa
<code>ac_grupo_max</code>	tamaño máximo de un determinado colectivo de tipo Amas de casa

Además, es necesaria la creación de los siguientes ficheros que contendrán grafos el formato CSC con la información relativa a las conexiones de grado fuerte y débil:

<b>Identificador fichero</b>	<b>Descripción</b>
<code>conocidos_debil.xml</code>	valores de los elementos <i>ncol</i> , <i>nrow</i> y <i>nnz</i> del grafo de conocidos de conexión débil
<code>conocidos_ncol_debil.xml</code>	valores <i>ncol</i> del grafo en formato CSC
<code>conocidos_nrow_debil.xml</code>	valores <i>nrow</i> del grafo en formato CSC
<code>conocidos_nnz_debil.xml</code>	valores <i>nnz</i> del grafo en formato CSC
<code>conocidos_fuerte.xml</code>	valores de los elementos <i>ncol</i> , <i>nrow</i> y <i>nnz</i> del grafo de conocidos de conexión fuerte
<code>conocidos_ncol_fuerte.xml</code>	valores <i>ncol</i> del grafo en formato CSC
<code>conocidos_nrow_fuerte.xml</code>	valores <i>nrow</i> del grafo en formato CSC
<code>conocidos_nnz_fuerte.xml</code>	valores <i>nnz</i> del grafo en formato CSC

### Parámetros de generación de conocidos dentro del ámbito del colectivo

No existen parámetros que intervengan en la generación de conocidos dentro del ámbito del colectivo de los individuos de la población, ya que se generan a través de la función de escalado previamente explicada.

### Parámetros de generación de conocidos fuera del ámbito del colectivo

Los parámetros que intervienen en la generación de conocidos fuera del ámbito del colectivo se encuentran contenidos en el fichero de configuración "*config.xml*" y son los siguientes:

Identificador	Descripción
<code>prob_grupo_debil</code>	probabilidad de conectar un par de colectivos con conexión débil (en valor porcentual)
<code>prob_grupo_fuerte</code>	probabilidad de conectar un par de colectivos con conexión fuerte (en valor porcentual)

### Parámetros de generación de conocidos familiares

No existen parámetros que intervengan en la generación de conocidos familiares de los individuos de la población.

## Modelo de enfermedad

### Modelo de transición de estados de una enfermedad infecciosa

Según el [WSDH] se conoce por enfermedad infecciosa la evidencia clínica de la presencia de una determinada enfermedad en humanos o animales como resultado de una infección provocada por el agente infeccioso que la transmite.

La presencia de un agente infeccioso genera una perturbación en el organismo que lo hospeda lo que provoca que éste sufra un periodo de contagio constituido por las siguientes etapas [Leyland]:

- *Incubación*: el periodo de incubación es el intervalo de tiempo que existe entre el contacto inicial con el agente infeccioso hasta la aparición de los primeros síntomas asociados con la infección. Según la enfermedad en esta etapa el organismo huésped puede ser infeccioso, es decir, ser transmisor de la misma.
- *Prodrómico*: en dicha etapa aparecen los síntomas generales de la enfermedad, que no permiten determinar la patología concreta que afecta al huésped. En esta etapa el organismo en el que reside el agente es infeccioso.
- *Clínico* (también denominada pico de la enfermedad): la aparición de los síntomas específicos de la enfermedad determina el periodo clínico. La aparición de dichos síntomas permite determinar la patología que afecta al organismo huésped, que durante esta etapa alcanza el máximo grado de transmisión.
- *Convalecencia*: alcanzado el periodo de convalecencia el organismo puede sufrir la recuperación por eliminación del agente infeccioso, la enfermedad puede volverse crónica o incluso provocar la muerte del huésped.

Partiendo de las etapas que componen el proceso patológico se construye para el uso de la aplicación un modelo de transición de estados válido para simular el comportamiento de cualquier tipo de enfermedad infecciosa en el organismo de un individuo y compuesto por los siguientes estados:

- *Sano*: estado en el que el paciente no sufre la presencia de cualquier agente infeccioso en su organismo y puede ser infectado. Este estado no aparecía en el modelo propuesto por [Layland] pero se incluye para definir el estado inicial de la población que va a ser víctima de la propagación.
- *Incubación*: estado en el que el paciente ha entrado en contacto con el agente infeccioso pero aún no presenta los primeros síntomas resultado de la infección. Durante este periodo el organismo puede o no ser contagioso dependiendo de las características propias de la enfermedad.
- *Infección*: estado en el que el paciente presenta los síntomas tanto generales como característicos de la enfermedad que sufre, con lo que podría tratarse como la unión de los periodos *prodómico* y *clínico*. Durante este periodo se presenta el mayor grado de transmisión de la enfermedad.

- *Inmune*: estado en el que el paciente ha superado la enfermedad tras el periodo de infección y su organismo es inmune a nuevas infecciones de la misma patología. Durante este periodo el organismo puede o no ser contagioso dependiendo de las características propias de la enfermedad.
- *Hospitalización*: estado en el que el paciente en periodo de infección requiere de atención e internamiento hospitalario hasta que finalice el mismo.
- *Fallecido*: estado en el que el paciente es derrotado por la presencia del agente infeccioso. Este estado puede ser alcanzado directamente desde el estado de infección o el de hospitalización.

Estos tres últimos estados son los que [Layland] define en uno sólo como estado de *convalecencia*, y que se ha creído conveniente discretizar por su interés desde el punto de vista computacional y de simulación de la propagación, ya que permite medir el estado concreto en el que finaliza la infección en cada uno de los individuos de la población.

El siguiente diagrama muestra la representación gráfica de las etapas y las diferentes transiciones entre las mismas que componen el proceso patológico de la siguiente manera:

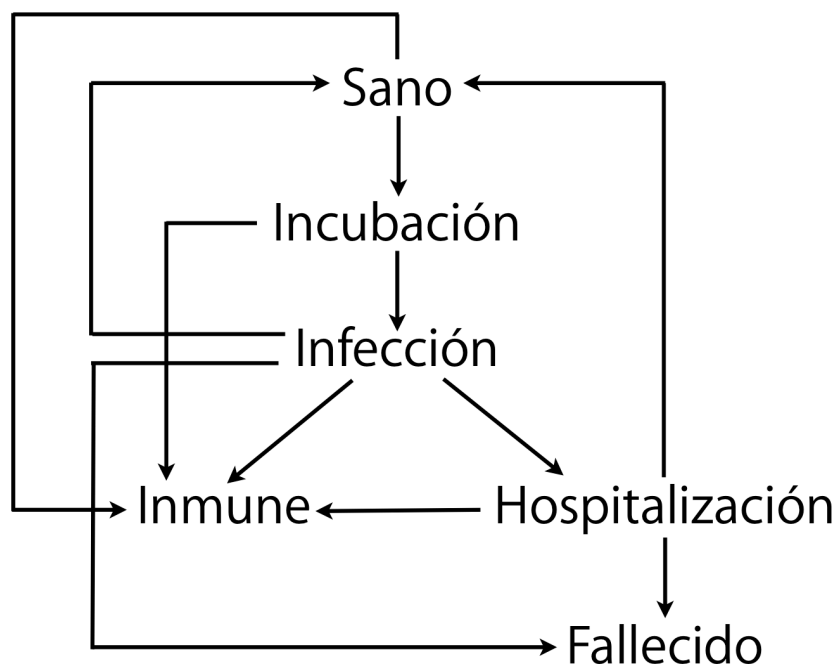


Ilustración 7. Modelo de transición de estados del individuo.

Las transiciones que se producen entre los diferentes estados del modelo vienen determinadas por una serie de parámetros definidos en apartado de descripción de parámetros del modelo de enfermedad. A continuación se describe la situación médica que provoca la transición entre los pares de estados y se nombran los parámetros que intervienen en la misma y que serán definidos más adelante:

- Transición *Sano - Incubación*: se produce cuando un individuo sano es contagiado por un agente infeccioso. Los parámetros que intervienen en la transición son: `prob_inf_ini`, `prob_inf_inc`, `prob_inf_rbajo`, `prob_inf_ralto`, `prob_inf_media`, `prob_inf_inm`.

- Transición *Sano - Inmune*: se produce cuando un individuo sano recibe una vacunación para el agente infeccioso. Los parámetros que intervienen en la transición son:  $vac\_time$ ,  $vac\_perc$ ,  $vac\_san\_acc$ .
- Transición *Incubación - Infección*: se produce cuando un individuo en periodo de incubación comienza a mostrar los primeros síntomas de infección. Los parámetros que intervienen en la transición son:  $\sigma\_tinc$ ,  $\mu\_tinc$ .
- Transición *Incubación - Inmune*: se produce cuando un individuo en periodo de incubación recibe una vacunación para el agente infeccioso. Los parámetros que intervienen en la transición son:  $vac\_time$ ,  $vac\_perc$ ,  $vac\_inc\_acc$ .
- Transición *Infección - Sano*: se produce cuando un individuo que ha superado la infección vuelve al estado inicial, es decir, puede ser infectado por el mismo agente nuevamente. Los parámetros que intervienen en la transición son:  $\sigma\_tinc$ ,  $\mu\_tinc$ ,  $prob\_inf\_a\_san$ .
- Transición *Infección - Inmune*: se produce cuando un individuo que ha superado la infección se vuelve inmune ante la enfermedad, es decir, no puede volver a ser infectado por el mismo agente infeccioso. Los parámetros que intervienen en la transición son:  $\sigma\_tinc$ ,  $\mu\_tinc$ ,  $prob\_inf\_a\_inm$ .
- Transición *Infección - Hospitalización*: se produce cuando un individuo en periodo de infección requiere de atención e internamiento hospitalario. Los parámetros que intervienen en la transición son:  $\sigma\_tinc$ ,  $\mu\_tinc$ ,  $prob\_inf\_a\_hos$ .
- Transición *Infección - Fallecido*: se produce cuando un individuo en periodo de infección fallece a consecuencia de la misma. Los parámetros que intervienen en la transición son:  $\sigma\_tinc$ ,  $\mu\_tinc$ ,  $prob\_inf\_a\_fal$ .
- Transición *Hospitalización - Sano*: se produce cuando un individuo que finaliza el periodo de internamiento hospitalario vuelve al estado inicial en el que puede volver a ser contagiado nuevamente por el agente infeccioso. Los parámetros que intervienen en la transición son:  $\sigma\_thos$ ,  $\mu\_thos$ ,  $prob\_hos\_a\_san$ .
- Transición *Hospitalización - Inmune*: se produce cuando un individuo que finaliza el periodo de internamiento hospitalario se vuelve inmune ante la enfermedad, es decir, no puede volver a ser contagiado por el mismo agente infeccioso. Los parámetros que intervienen en la transición son:  $\sigma\_thos$ ,  $\mu\_thos$ ,  $prob\_hos\_a\_inm$ .
- Transición *Hospitalización - Fallecido*: se produce cuando un individuo que finaliza el periodo de internamiento hospitalario fallece a causa de la infección. Los parámetros que intervienen en la transición son:  $\sigma\_thos$ ,  $\mu\_thos$ ,  $prob\_hos\_a\_fal$ .

## Estructuras de datos

El conjunto de estructuras de datos necesarias para gestionar la representación del modelo de enfermedad propuesto mantienen para cada individuo de la población que interviene en la simulación la siguiente información:

- Conexiones con el resto de individuos de la población, es decir, conocidos. Se incluye conexiones diferenciadas con individuos de riesgo alto y bajo.
- Edad del individuo.
- Ocupación del individuo, es decir, tipo de colectivo al que pertenece: trabajador, estudiante, anciano o ama de casa.
- Condición del individuo, es decir, estado del modelo de transición de estados en el que se encuentra en el instante actual.
- Cantidad de tiempo del periodo de incubación en caso de contraer la enfermedad.
- Cantidad de tiempo del periodo de infección en caso de contraer la enfermedad.
- Cantidad de tiempo del periodo de hospitalización en caso de necesitar asistencia hospitalaria tras contraer la enfermedad.

La gestión de las conexiones con el resto de individuos de la población se realiza con el grafo definido en la generación del modelo social de la simulación. El resto de información personal requiere de las siguientes estructuras de datos:

- Array `age`: vector unidimensional de tipo entero y compartido por todos los procesos que contiene un elemento por individuo de la población en el que se almacena la edad del mismo. La edad de cada individuo se genera de forma aleatoria con los valores de edad máxima y mínima del colectivo al que pertenece el mismo.
- Array `occupation`: vector unidimensional de tipo entero y compartido por todos los procesos que contiene un elemento por individuo de la población en el que se almacena el tipo de colectivo al que pertenece.
- Array `condition`: vector unidimensional de tipo entero y compartido por todos los procesos que contiene un elemento por individuo de la población en el que se almacena el estado del mismo (se codifica cada estado con un valor entero). La condición inicial de los individuos es el estado sano.
- Array `tinc`: vector unidimensional de entero que contiene un elemento por individuo de la población en el que se almacena el tiempo de incubación expresado en minutos. El tiempo de incubación de cada individuo se genera con una distribución Gaussiana con los valores de los parámetros `sigma_tinc` y `mu_tinc`.
- Array `tinf`: vector unidimensional de entero que contiene un elemento por individuo de la población en el que se almacena el tiempo de infección expresado en minutos. El tiempo de infección de cada individuo se genera de forma aleatoria con valores centrados en el parámetro `sigma_tinf` con un desplazamiento positivo o negativo de valor del parámetro `mu_tinf`.

- 
- Array `thos`: vector unidimensional de entero que contiene un elemento por individuo de la población en el que se almacena el tiempo de hospitalización expresado en minutos. El tiempo de hospitalización de cada individuo se genera con una distribución Gaussiana con los valores de los parámetros `sigma_thos` y `mu_thos`.

## Algoritmo de simulación de la transmisión

El algoritmo de simulación de la transmisión se lleva a cabo utilizando de forma conjunta el modelo social y el modelo de enfermedad de la aplicación. Mediante el modelo social es posible simular la transmisión del agente infeccioso entre los diferentes individuos que componen la población usando las conexiones entre ellos como vías de propagación del agente. Mediante el modelo de enfermedad se simulan los diferentes estados que atraviesa cada individuo de la población a lo largo de la simulación de la infección.

La función que realiza el algoritmo de simulación se encuentra en el fichero fuente de la aplicación y se denomina `Spread`. Dicha función recibe los parámetros necesarios para su funcionamiento y realiza las siguientes rutinas:

### Discretización de conexiones en riesgo alto y bajo

Una vez que se alcanza la función de simulación de la transmisión de la infección la primera rutina consiste en discretizar las conexiones de cada individuo en los dos grupos de riesgo existentes, riesgo alto y bajo. Recordemos que el riesgo de cada conexión representa el grado de vulnerabilidad de dicha conexión para servir como canal de transmisión del agente infeccioso entre los dos individuos que la componen.

El porcentaje de conexiones de cada uno de los individuos que pertenecerán a cada uno de los tipos de riesgo está definido por el valor de los parámetros de infección `prob_inf_rbajo`, `prob_inf_ralto` y `prob_inf_media`, con los que se construye la siguiente ecuación:

$$\frac{(\text{prob\_inf\_media} - \text{prob\_inf\_rbajo})}{(\text{prob\_inf\_ralto} - \text{prob\_inf\_rbajo})}$$

El resultado de dicha ecuación es el porcentaje de conexiones de un determinado individuo que pertenecerán a sus conexiones de riesgo bajo, siendo el porcentaje restante el que representa las conexiones de riesgo alto. Finalmente, aplicando dichos porcentajes al número de conexiones (conocidos) de cada individuo se obtiene el número de cuáles serán de bajo y alto riesgo.

### //Algoritmo de discretización de conexiones

```
Desde individuo = 0 hasta individuo = total tamaño población
    Obtener número de conexiones de riesgo bajo y de riesgo alto
    Seleccionar al azar qué conexiones serán de riesgo bajo
    //El resto de conexiones serán de riesgo alto
Fin Desde
```



## Transmisión de la infección

La rutina de transmisión se mantiene activa durante el tiempo de simulación establecido como parámetro  $ntime$  y se encarga de recorrer las conexiones de cada uno de los individuos de los distintos colectivos que componen la población generando una probabilidad de propagación de la infección para cada una de las conexiones.

La probabilidad de propagación depende de los siguientes factores:

- **Tipo de conexión y hora actual.** El tipo de cada una de las conexiones del individuo hace referencia al modelo social de la población. Recordemos que los tipos de conexiones existentes son: conexiones dentro del ámbito del colectivo, conexiones fuera del ámbito del colectivo y conexiones familiares.

Con esto se pretende representar que según las diferentes franjas horarias establecidas durante el día el grafo de la población se dividirá en tres diferentes subgrafos, uno por cada tipo de conexión existente. Por lo tanto, se puede expresar el grafo de la población como la suma de los tres subgrafos.

Las siguientes ilustraciones (de 8 a 15) muestran la representación de grafos y subgrafos en el software matemático MatLab. Notar que dicho software invierte el sentido del eje Y, por lo que la representación aparece en la diagonal inversa en la que hasta ahora se había representado la distribución de los colectivos sobre el grafo.

Un grafo de población extraído de una simulación real tiene el siguiente aspecto:

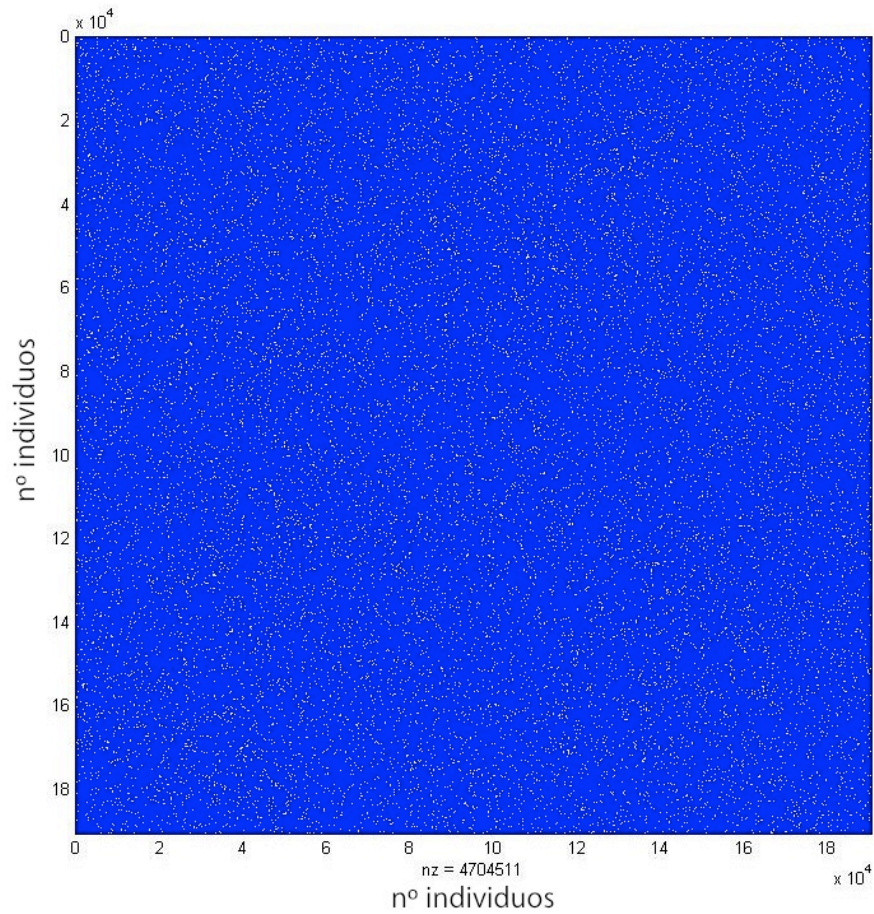


Ilustración 8. Aspecto general de un grafo completo de población.

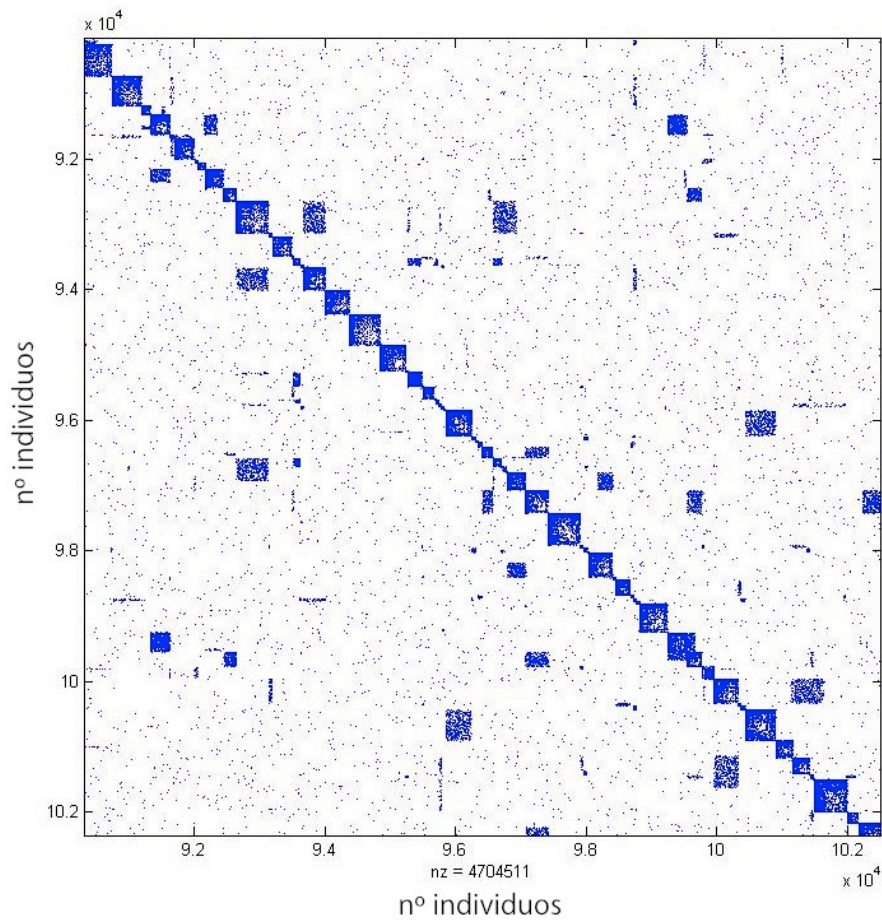


Ilustración 9. Ampliación [x2] de un grafo completo de población.

---

La función de la discretización de las conexiones en diferentes tipos se basa en las diferentes franjas del día en que dichas conexiones están activas:

- Las conexiones de tipo dentro del ámbito del colectivo se producirán dentro de la franja horaria dedicada a la actividad propia del colectivo, que se corresponde de 9 horas de la mañana a 17 horas de la tarde.

En esta franja horaria el subgrafo que interviene es el de las conexiones entre miembros del mismo colectivo (tipo 1):

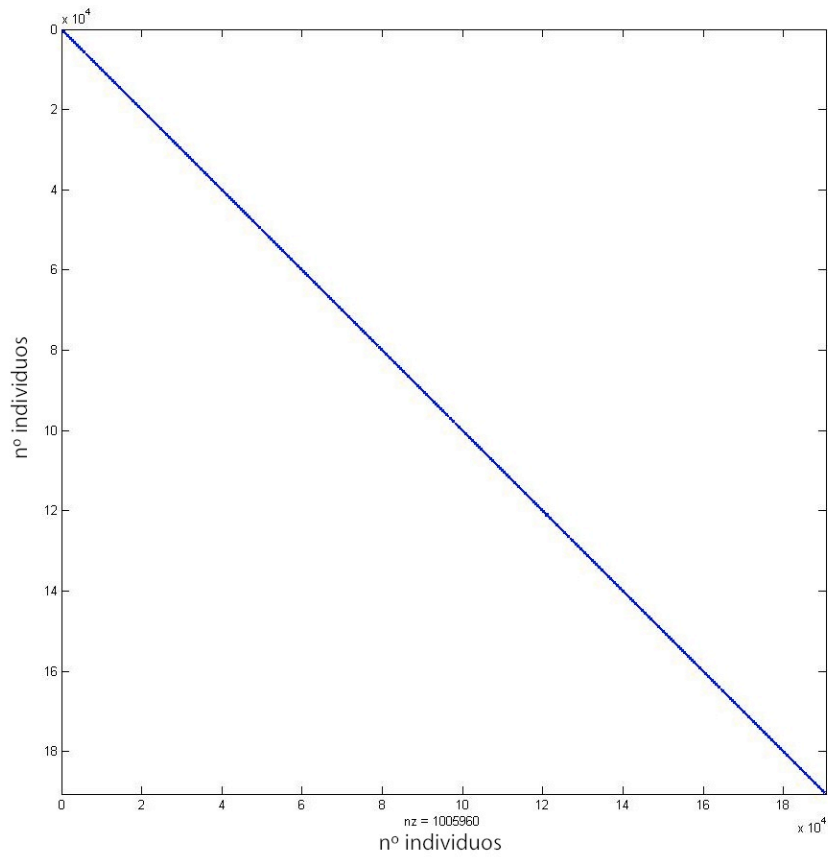


Ilustración 10. Aspecto general de un grafo tipo 1 de población.

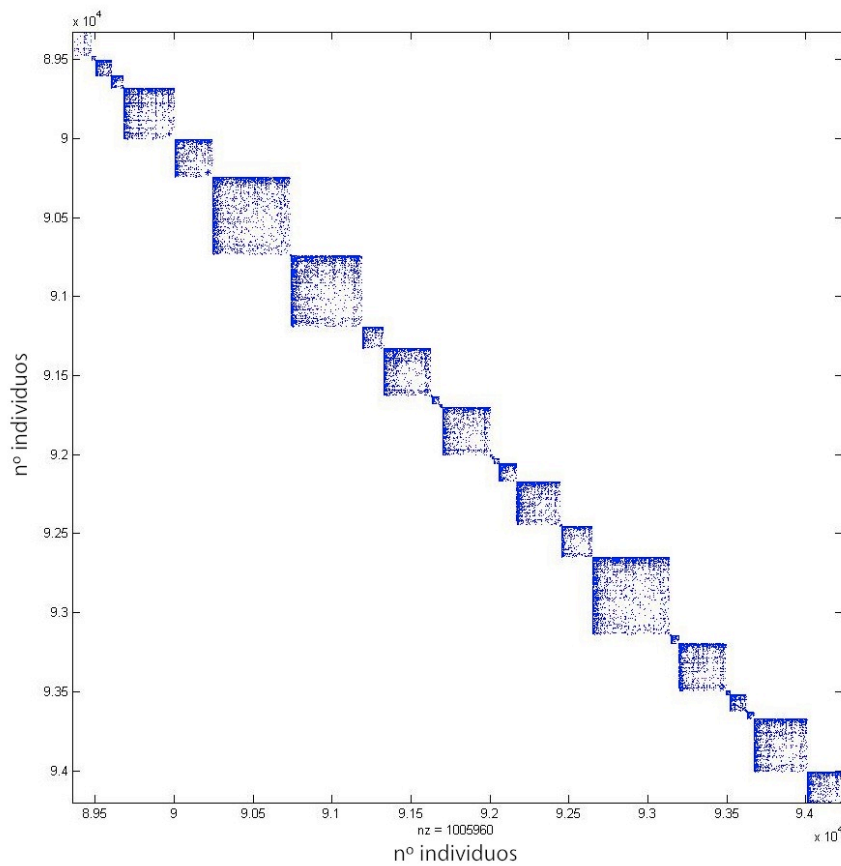


Ilustración 11. Ampliación [x4] de un grafo tipo 1 de población.

- 
- Las conexiones de tipo fuera del ámbito del colectivo se producirán dentro de la franja horaria dedicada a actividades externas al colectivo, que se corresponde de 17 horas de la tarde a 20 horas de la tarde.

En esta franja horaria el subgrafo que interviene es el de las conexiones entre individuos que no pertenecen al mismo colectivo y no son familiares (tipo 2):



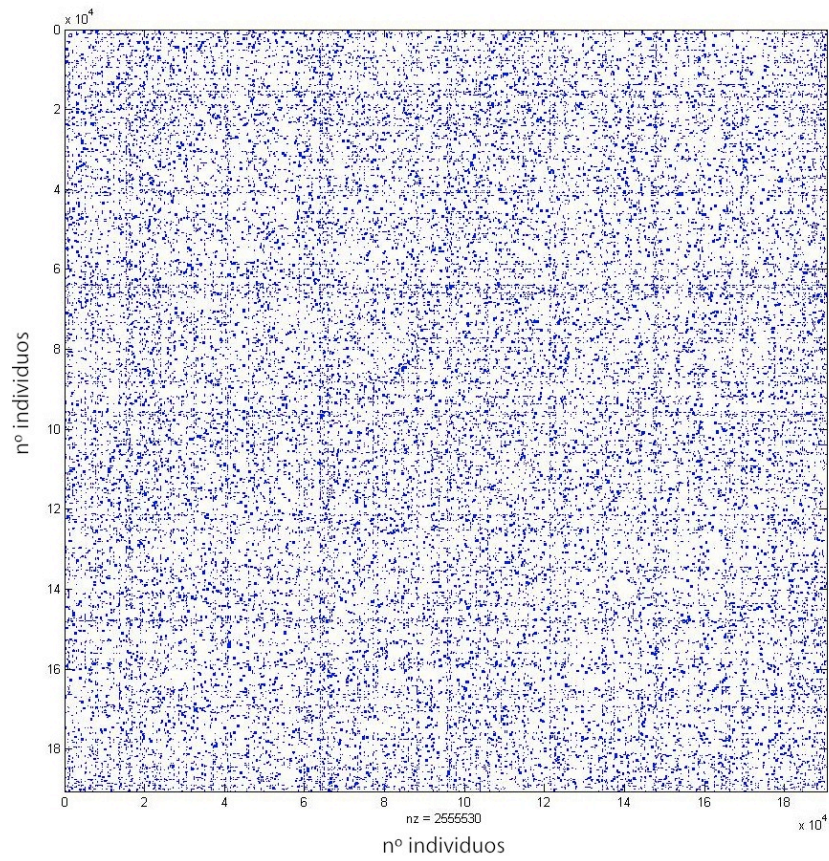


Ilustración 12. Aspecto general de un grafo tipo 2 de población.

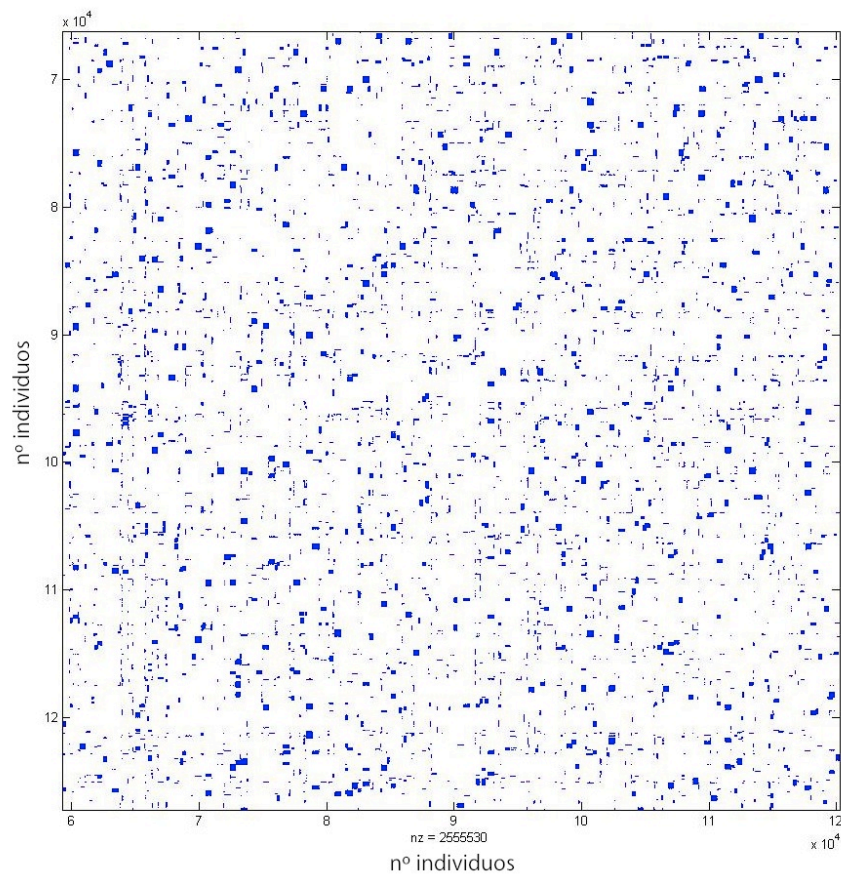


Ilustración 13. Ampliación [x2] de un grafo tipo 2 de población.

- Las conexiones de tipo familiar se producirán dentro de la franja horaria en que el individuo permanece en contacto directo con la familia, que se corresponde de 20 horas de la tarde a 9 horas de la mañana.

En esta franja horaria el subgrafo que interviene es el de las conexiones entre familiares (tipo 3):



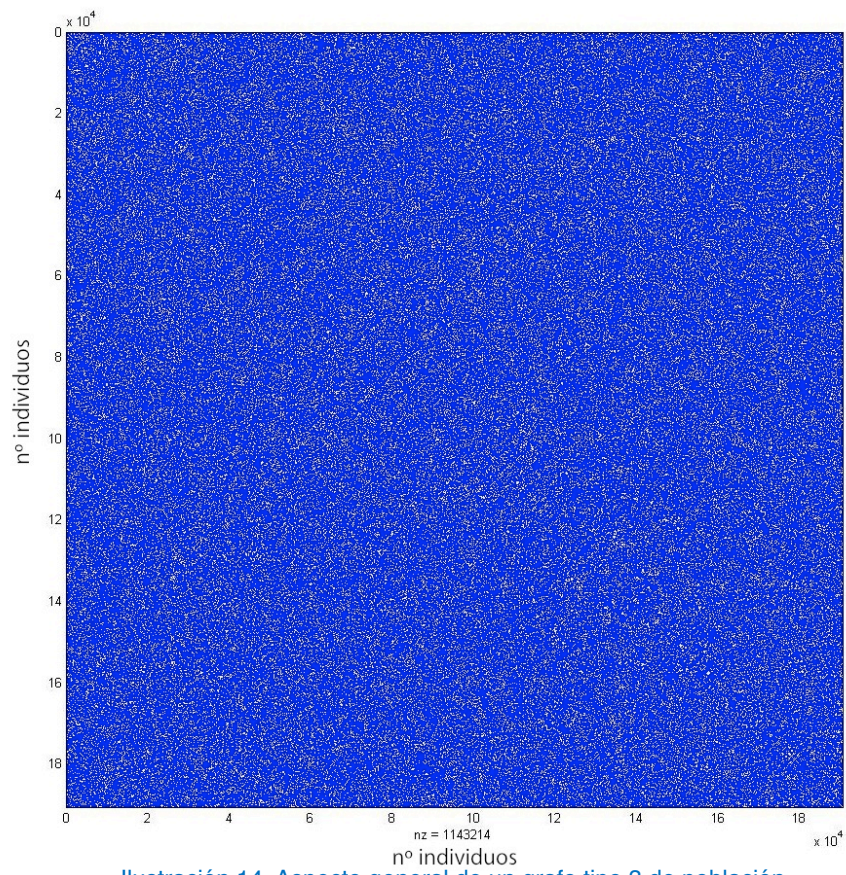


Ilustración 14. Aspecto general de un grafo tipo 3 de población.

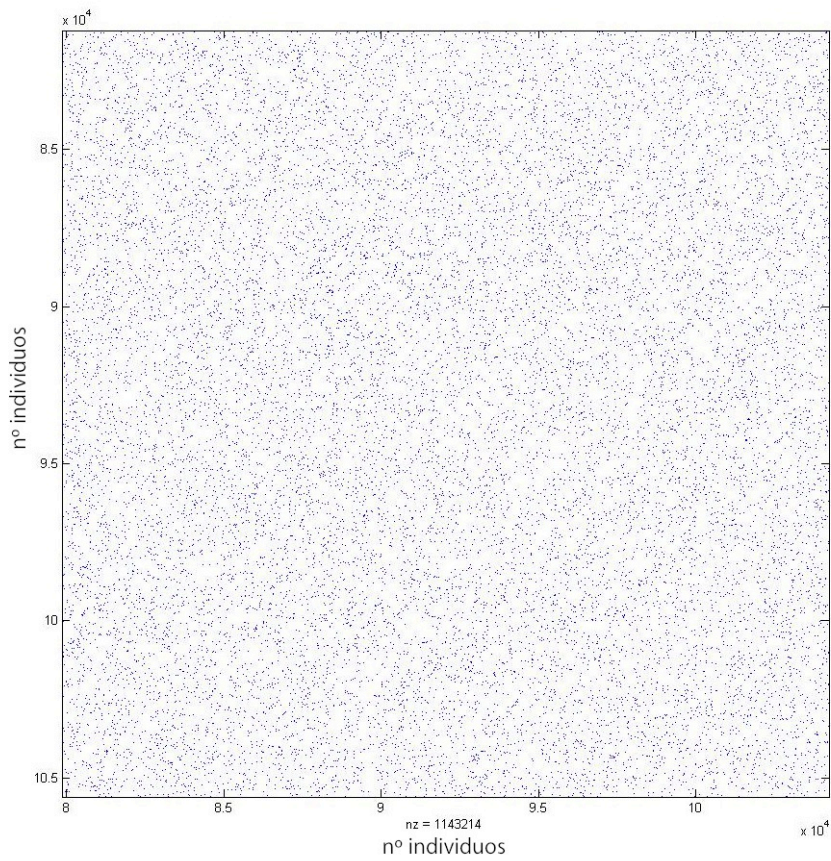


Ilustración 15. Ampliación [x2] de un grafo tipo 3 de población.



- **Condición del individuo.** La condición (o estado del modelo de transición de estados en que se encuentra) del individuo portador del agente infeccioso en su organismo determina la aplicación de diferentes probabilidades en función de la misma.

La probabilidad de infección se obtiene mediante la siguiente fórmula en la que intervienen los parámetros `prob_inf_rbajo`, `prob_inf_ralto`, `ttrans`, `media_sp` y el tiempo de infección del individuo almacenado en el array `tinf`:

$$\text{Probabilidad de infección} = \frac{(\text{prob\_inf}(\text{ralto o rbajo}) * \text{ttrans})}{(\text{media\_sp} * \text{tinf}[\text{individuo}])}$$

Lo que se pretende representar con esta fórmula es que la probabilidad de infección (para conocidos de riesgo alto o bajo, según) depende de los pasos de tiempo en los que se simula la propagación (`ttrans`), de la media de conocidos en la población del grafo estándar (`media_sp`) y del tiempo de infección del individuo (`tinf[individuo]`).

### //Rutina de transmisión del virus por contagio

Desde individuo = 0 hasta individuo = total tamaño población

Si estado = INCUBANDO, INFECTADO o INMUNE y según tipo de conexión y hora actual

Desde conocido = 0 hasta conocido = total conocidos individuo

Ejecutar probabilidad

Si conocido es de riesgo bajo

Calcular probabilidad de infección de riesgo bajo e infectar si procede

Si conocido es de riesgo alto

Calcular probabilidad de infección de riesgo alto e infectar si procede

Fin Desde

Fin Desde

## Actualización de cambios del estado de los individuos de la población

Una vez aplicado el algoritmo de transmisión se realiza una evaluación del estado en el que se encuentra cada uno de los individuos para comprobar si debe realizar una transición a otro estado. La situación y parámetros que intervienen en las diferentes transiciones fueron descritos en el apartado de definición del modelo de transición de estados de una enfermedad infecciosa.

En la rutina de actualización no se realizan transiciones de estado para los individuos que se encuentran en estado sano, inmune o fallecido, ya que la transición desde sano se realiza por contagio del virus y los estados inmune y fallecido son estados finales.

A continuación se muestra el pseudocódigo correspondiente a la rutina de transición entre los distintos estados del modelo:

### //Rutina de actualización de transición de estados

Desde individuo = 0 hasta individuo = total tamaño población

    Si estado = INCUBANDO y superado tiempo\_incubación

        Transitar a estado = INFECTADO

    Si estado = INFECTADO y superado tiempo\_infección

        Ejecutar probabilidad

        Según probabilidad transitar a estado = SANO, INMUNE,  
        HOSPITALIZADO, FALLECIDO

    Si estado = HOSPITALIZADO y superado tiempo\_hospitalización

        Ejecutar probabilidad

        Según probabilidad transitar a estado = SANO, INMUNE,  
        FALLECIDO

Fin Desde

## Aplicación de vacunas

Caso particular de una actualización de cambio de estado de los individuos las vacunaciones se aplican de manera probabilística sobre la población de la simulación. La implementación de las vacunaciones permite realizar estimaciones del impacto de campañas de vacunación en poblaciones concretas y cómo se refleja en la propagación del virus.

El sistema de vacunaciones está constituido por cuatro parámetros que permiten definir el instante de tiempo en el que se aplica la campaña de vacunación, el porcentaje de población que se vacuna y el porcentaje de éxito en individuos que se encuentren en estado sano o de incubación.

Cuando un individuo es vacunado se refleja con una transición en el modelo de estados de la enfermedad desde el estado en que se encuentre (sano o incubando) hasta el estado inmune.

### //Rutina de aplicación de vacunas

Desde individuo = 0 hasta individuo = total tamaño población

Si estado = SANO o INCUBANDO

Ejecutar probabilidad

Si probabilidad ok

Ejecutar probabilidad de éxito de la vacuna en función del estado del individuo (SANO o INCUBANDO)

Si probabilidad ok

Aplicar vacuna, el individuo transita al estado INMUNE

Fin Desde

## Parámetros del modelo de enfermedad

Los parámetros que intervienen en la creación del modelo de enfermedad son los siguientes:

Identificador	Descripción
agente_patogeno	nombre identificador del agente infeccioso que se simula
ttrans	tiempo que transcurre entre las transmisiones de la enfermedad entre los individuos de la población (en minutos)
sigma_tinc	componente $\sigma$ del tiempo de incubación de la enfermedad
mu_tinc	componente $\mu$ del tiempo de incubación de la enfermedad
sigma_tinf	componente $\sigma$ del tiempo de infección de la enfermedad
mu_tinf	componente $\mu$ del tiempo de infección de la enfermedad
sigma_thos	componente $\sigma$ del tiempo de hospitalización de la enfermedad
mu_thos	componente $\mu$ del tiempo de hospitalización de la enfermedad
vac_time	tiempo de aplicación de vacunas sobre la población (en minutos)
vac_perc	porcentaje de individuos de la población que accede a la vacuna contra el agente infeccioso (en valor porcentual)
vac_inc_acc	efectividad de la vacuna sobre individuos que estén incubando la enfermedad (en valor porcentual)
vac_san_acc	efectividad de la vacuna sobre individuos sanos (en valor porcentual)
prob_inf_ini	probabilidad de población infectada al inicio de la simulación (en valor porcentual)
prob_inf_inc	probabilidad en estado de incubación de transmitir la infección a otros individuos (en valor porcentual)
prob_inf_ralto	probabilidad en estado de infección de transmitir la infección a otros individuos mediante una conexión de riesgo alto (en número de individuos)
prob_inf_rbajo	probabilidad en estado de infección de transmitir la infección a otros individuos mediante una conexión de riesgo bajo (en número de individuos)
prob_inf_media	probabilidad media en estado de infección de transmitir la infección a otros individuos (en número de individuos)
prob_inf_inm	probabilidad en estado inmune de transmitir la enfermedad a otros individuos (en valor porcentual)

---

<b>Identificador</b>	<b>Descripción</b>
prob_inf_a_san	probabilidad de evolucionar de estado infectado a sano transcurrido el periodo de infección (en valor porcentual)
prob_inf_a_inm	probabilidad de evolucionar de estado infectado a inmune transcurrido el periodo de infección (en valor porcentual)
prob_inf_a_hos	probabilidad de evolucionar de estado infectado a hospitalización transcurrido el periodo de infección (en valor porcentual)
prob_inf_a_fal	probabilidad de evolucionar de estado infectado a fallecido transcurrido el periodo de infección (en valor porcentual)
prob_hos_a_san	probabilidad de evolucionar de estado hospitalizado a sano transcurrido el periodo de hospitalización (en valor porcentual)
prob_hos_a_inm	probabilidad de evolucionar de estado hospitalizado a inmune transcurrido el periodo de hospitalización (en valor porcentual)
prob_hos_a_fal	probabilidad de evolucionar de estado hospitalizado a fallecido transcurrido el periodo de hospitalización (en valor porcentual)

# Implementación MPI

*En este apartado se realiza una breve introducción al modelo de programación de paso de mensajes con la librería Interfaz de Paso de Mensajes (MPI) para posteriormente describir el proceso de paralelización seguido para construir la aplicación.*

## Modelo de programación de paso de mensajes

El modelo de programación de paso de mensajes se implementa con la librería Interfaz de Paso de Mensajes, más conocida por sus siglas en inglés MPI (Message Passing Interface), que es [Culler et al, 1999] una especificación de protocolo de comunicaciones estándar para computación paralela con memoria distribuida.

Las principales ventajas del establecimiento del estándar MPI son la portabilidad, el rendimiento y la facilidad de uso que ofrece:

- Portabilidad ya que existen numerosas implementaciones del estándar para multitud de lenguajes de programación, entre los que se encuentran C, C++, Fortran o Java.
- Rendimiento ya que cada implementación de la biblioteca ha sido previamente optimizada para el hardware en el cuál se ejecuta.
- Facilidad de uso, como resultado de su portabilidad y característica estándar.

El principal objetivo de MPI es el establecimiento de un estándar ampliamente desarrollado que pueda ser utilizado para escribir aplicaciones de paso de mensajes bajo una arquitectura de memoria distribuida. Además, se describen otra serie de objetivos:

- Diseñar una interfaz de programación de aplicaciones.
- Permitir comunicaciones de forma eficiente.
- Permitir el desarrollo de implementaciones que puedan ser utilizadas en entornos heterogéneos.
- Permitir enlaces a la interfaz desde diversos lenguajes de programación, tales como C, C++, o Fortran.
- Establecer una interfaz de comunicaciones fiable.
- Definir una interfaz que pueda ser implementada por múltiples desarrolladores sin cambios significativos en la capa de comunicación.
- La semántica de la interfaz es independiente del lenguaje.
- La interfaz es diseñada para proporcionar seguridad a los subprocesos.

La versión más actual de la librería MPI es la MPI-2.2, una actualización menor que corrige errores y ambigüedades presentes en el estándar MPI-2.1. MPI-2 añade a las características de MPI-1 nativo (paso de mensajes y entorno de ejecución estático) operaciones de entrada salida E/S, gestión dinámica de los procesos y operaciones remotas de memoria.

MPI proporciona funciones de comunicación punto a punto entre un par de procesos así como una serie de funciones colectivas en las que toman parte todos los procesos presentes en la ejecución paralela. Entre las comunicaciones punto a punto se distingue entre las operaciones síncronas (en las que el proceso que envía no espera a que el mensaje sea recibido) y asíncronas (en las que el proceso detiene su ejecución hasta que el mensaje que ha enviado sea recibido por el proceso destino). Dentro de las funciones colectivas contamos con funciones para difusión (de uno al resto de procesos) y de recolección (del resto a un proceso) de información.

Una de las principales características de MPI es que antes de comenzar la ejecución del programa se asigna a cada nodo un identificador (denominado *rank*) que permitirá definir secciones de código individualizadas para cada uno de ellos. A excepción de ello MPI es capaz de abstraer la topología a nivel físico para lo que se define un comunicador que actuará como colección de procesos que formarán de parte conjunta los procesos activos en ejecución.

De entre las muchas implementaciones de MPI disponibles se ha utilizado MPICH2, correspondiente a la implementación MPI-2.1 [MPICH2] y presente en el clúster donde se ha llevado a cabo la evaluación de los supuestos del proyecto.

## Proceso de paralelización

Se denomina computación paralela a la existencia de varios procesos en ejecución y además las instrucciones se ejecutan de forma simultánea. La computación paralela pretende dividir un problema secuencial de grandes dimensiones en diferentes tareas que se ejecuten de forma concurrente. El proceso de paralelización [Culler et al, 1999] es un proceso secuencial compuesto por cuatro etapas:

- *Descomposición* del problema en tareas.
- *Asignación* de las tareas a los procesos.
- *Orquestación* de los accesos a datos necesarios, comunicaciones y sincronización entre los procesos.
- *Mapeo* para asignar los procesos a los procesadores.

La siguiente imagen corresponde con la ilustración del proceso de paralelización descrito [Culler et al, 1999] y [ARCOS, 2009]:

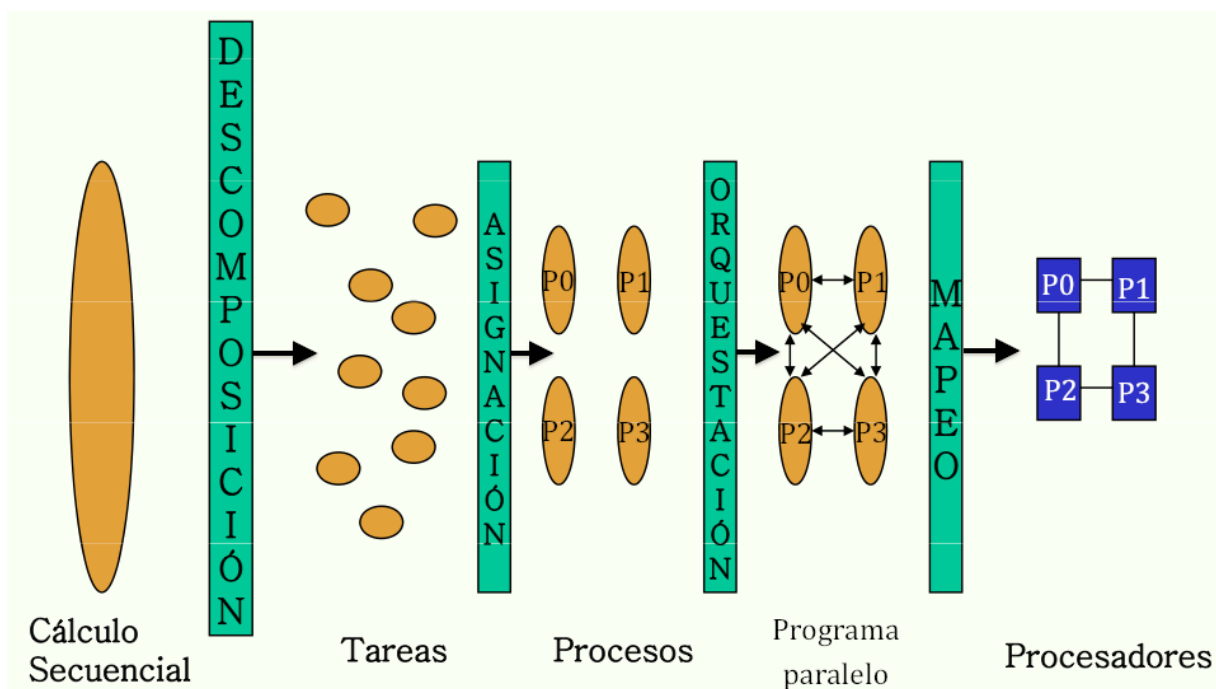


Ilustración 16. Etapas del proceso de paralelización. [Culler et al, 1999] y [ARCOS, 2009].



En la ilustración 17 se representa un diagrama de la aplicación en el que se detalla las principales acciones que realiza cada sección, las funciones que intervienen en cada una de ellas y las zonas de cálculo secuencial o paralelo:

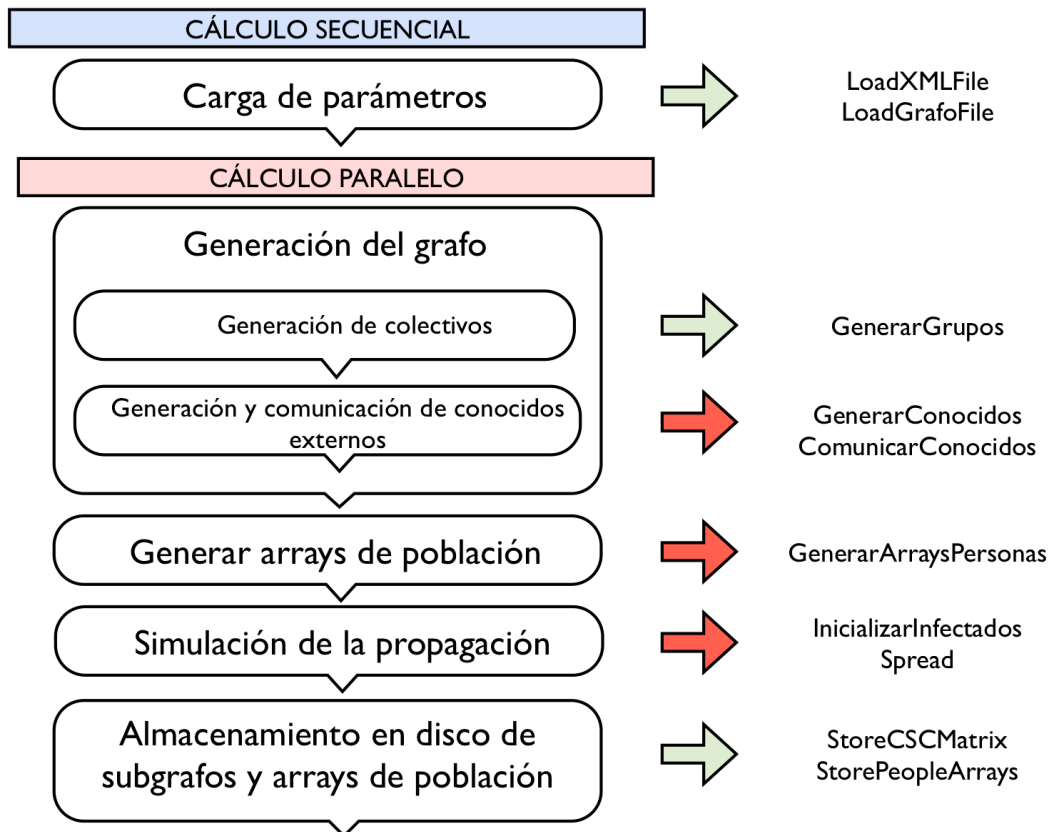


Ilustración 17. Diagrama de flujo de la aplicación.

En la figura se denota con una flecha en color rojo las secciones que requieren comunicación y sincronización mediante el paso de mensajes (MPI) entre los distintos procesos en ejecución, y en verde las secciones en las que cada proceso trabaja de forma autónoma.

## Descomposición

La primera etapa del proceso de paralelización consiste en dividir el cómputo secuencial en una colección de tareas. Las tareas son partes del trabajo secuencial que pueden ser asignadas de forma individual a cada uno de los procesos. Las dos principales funciones con cómputo secuencial del proyecto se corresponden con la construcción de cada uno de los modelos de representación existentes:

- *Modelo social*: la construcción del grafo es una función que requiere de elevada potencia de cálculo y tamaño de memoria y que es posible dividir en tareas siendo cada una de ellas la construcción de una porción del mismo.
- *Modelo de enfermedad*: ligado al caso anterior es posible dividir la simulación en función de los individuos que queden distribuidos sobre cada proceso respecto al grafo del modelo social.

### Modelo social

En cuanto al modelo social, la construcción del grafo dividida en tareas y que posteriormente es distribuida sobre cada uno de los procesos que intervienen en la ejecución en la fase de *Asignación* tendría un aspecto similar al siguiente:

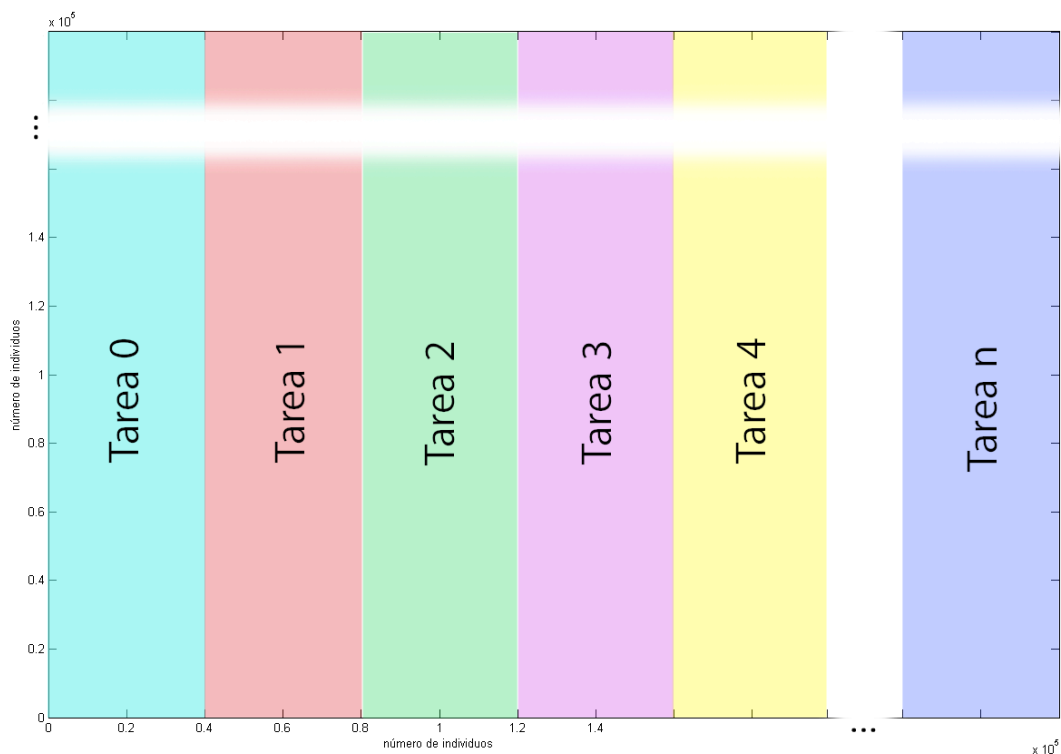


Ilustración 18. Distribución de tareas sobre un grafo de una población de ejemplo.

En esta ilustración se representa una población de  $4 * 10^4 * n$  individuos, donde  $n$  es el número de tareas en las que se divide el problema de forma equitativa. Notar que no se pretende representar cada una de las secciones del grafo como matrices densas, sino que el coloreado simplemente sirve para indicar que cada sección pertenecerá a tareas distintas.

Se puede apreciar en la imagen cómo el grafo queda distribuido por columnas en función del tamaño de la población y el número de procesos de la ejecución. El número de columnas asignado a cada una de las tareas resulta de dividir el tamaño de la población entre el número de procesos. De este modo se realiza un balanceo de carga entre los distintos procesos preservando que la cantidad de trabajo de las tareas asignadas a cada uno de los mismos sea equitativa con el objetivo de alcanzar el mejor rendimiento posible de la aplicación.

Con la distribución prevista cada uno de los procesos podrá trabajar de forma autónoma sobre el grafo ya que dispondrá de las conexiones de todos los individuos, a excepción de las conexiones fuera del ámbito del colectivo y familiares que se encuentren fuera de la población asignada al proceso, en cuyo caso deberá transferirse, al igual que en caso contrario en el que dicho proceso disponga de información que pertenezca a conexiones externas del resto de procesos. En la fase de *Orquestación* se describe cómo se realiza la transferencia de los datos compartidos entre procesos.

### Modelo de enfermedad

En cuanto al modelo de enfermedad se presentan dos tipos de comportamientos en función del carácter local o distribuido de los diferentes vectores de información del modelo. Recordemos que existían una serie de vectores de carácter distribuido, lo que significa que todos los procesos requieren de la información almacenada en todo el vector aún cuando éste contiene la información de todos los individuos de la simulación, incluso de los asignados en las tareas a otros procesos. Los vectores de carácter local pertenecen a cada uno de los procesos y los gestionan de forma autónoma, sin requerir que su información sea compartida al resto de procesos en ejecución.

Los vectores de carácter distribuido son aquellos que almacenan la edad (array `age`), condición (array `condition`) y ocupación (array `occupation`) de los individuos. En la ilustración 19 (y continuando con el ejemplo de la población mostrada en el grafo del apartado anterior) se representa el estado del array `occupation` en un determinado instante de tiempo, siendo el indicador entero contenido en cada posición el valor de `occupation` para un determinado individuo de la población. La sección de array gestionado por cada proceso será de  $4 * 10^4$  posiciones:

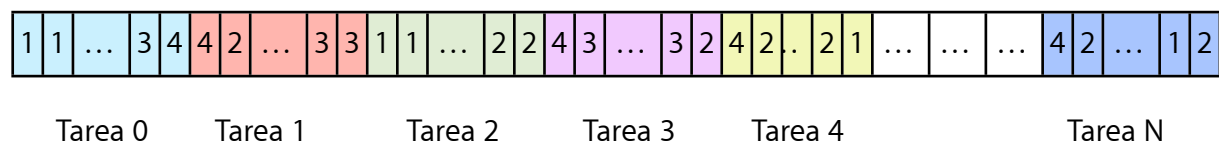


Ilustración 19. Distribución en tareas de un array `occupation` de ejemplo.

Se descompone la información de cada uno de los vectores comentados entre los  $n$  procesos que se ejecutan de forma paralela obteniendo un número de tareas  $n$  que posteriormente son asignadas a cada uno de ellos con un tamaño de población determinado. Cada proceso al que es asignada la tarea es el encargado de generar la información del vector perteneciente a la misma. La información del resto de individuos de la población (generada por el resto de procesos) es combinada por todos los procesos mediante una operación de comunicación que se describe en la fase de *Orquestación*.

Por contra, los vectores de carácter local son aquellos que almacenan el tiempo de incubación (array `tinc`), tiempo de infección (array `tinfect`) y tiempo de hospitalización (array `thos`) de los individuos.

Ejemplo en el que aparece cada uno de los arrays sin información en sus elementos:

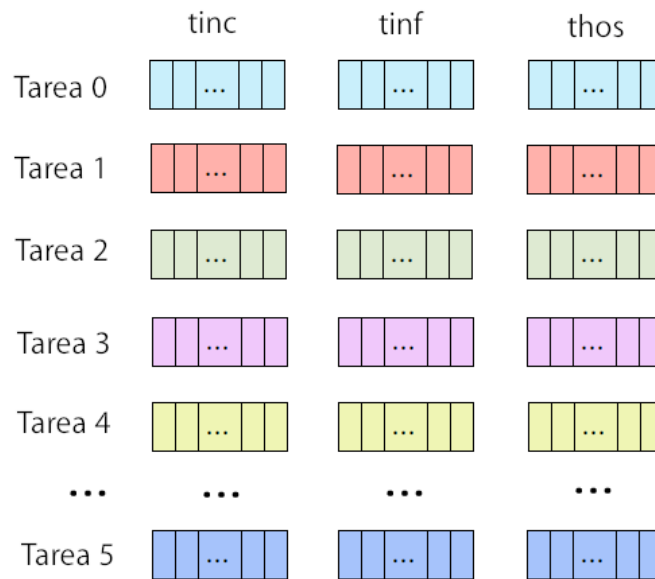


Ilustración 20. Distribución de arrays locales.

Para una situación con  $n$  tareas como el ejemplo anterior, cada una de las tareas sería asignada a cada uno de los procesos que posteriormente genera y mantiene la información única y exclusivamente de los individuos que pertenecen a su asignación. De este modo, y siempre según el ejemplo anterior ilustrado gráficamente, cada uno de los  $n$  procesos tendría tres vectores de tamaño cinco elementos: array de tiempo de incubación, array de tiempo de infección y array de tiempo de hospitalización.

La información del resto de individuos de la población (generada por el resto de procesos) no es necesaria para que los procesos puedan trabajar de forma autónoma, por lo que no se requiere comunicación para transmitirla.

## Asignación

La etapa de asignación tiene como objetivo distribuir las tareas entre los procesos que realizan la ejecución paralela. Con el objetivo de reducir la comunicación y sincronización entre los distintos procesos la principal meta de la etapa de asignación es obtener un balanceo de la carga que se distribuye entre los mismos. Balanceo de carga se refiere a un reparto equitativo del trabajo asignado a los procesos así como que la información asignada a cada uno de ellos les permita trabajar de forma autónoma sin requerir de información extra que solicitar al resto de procesos.

Con dicha premisa se realiza una asignación dinámica del trabajo, lo que significa que la cantidad de trabajo (tamaño de las tareas) que posteriormente se distribuye sobre los procesos se realiza en tiempo de ejecución. Ya se ha explicado antes cómo tanto los datos del modelo social como del modelo de enfermedad se distribuían en tareas en función del tamaño total del problema (tamaño de la población) y el número de procesos que intervienen en la ejecución paralela. De este modo obtenemos un balanceo de la carga de trabajo: cada proceso adquiere una tarea de tamaño equitativo a la que reciben el resto de procesos pero que requiere de comunicaciones para obtener la información que le permita funcionar de forma autónoma.

## Modelo social

Cada una de las tareas del modelo social definidas en la fase de *Descomposición* es asignada a uno de los procesos en ejecución, de forma que el grafo queda distribuido sobre los diferentes procesos como muestra la imagen inferior.

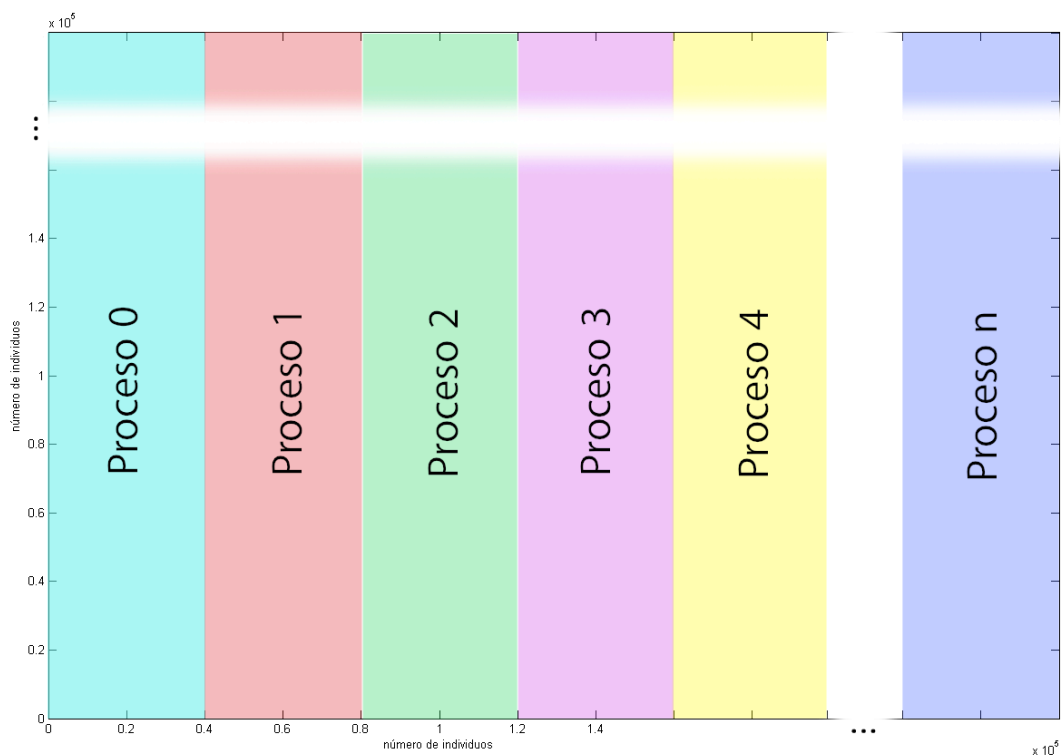


Ilustración 21. Asignación de tareas sobre un grafo de una población de ejemplo.

En la imagen observamos cómo cada uno de los procesos adquiere de forma ordenada una de las tareas en las que se descompone el problema del grafo inicialmente. En este contexto entendemos por tarea la parte de procesamiento de cada uno de los subgrafos generados.

### Modelo de enfermedad

En cuanto al modelo de enfermedad los vectores de carácter distribuido que se dividía en tareas en la fase de *Descomposición*. Siguiendo con el ejemplo de la fase de *Descomposición* en la ilustración 22 se muestra cómo cada una de dichas tareas queda asignada a uno de los procesos en ejecución de la siguiente manera:

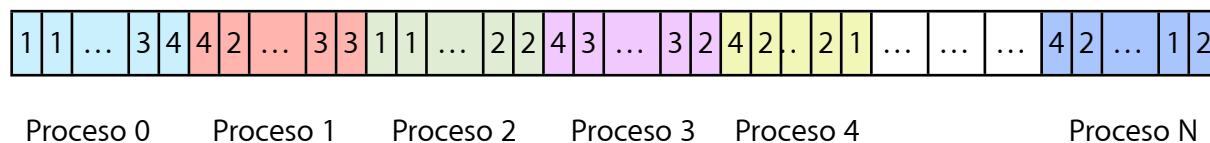


Ilustración 22. Asignación de tareas sobre un array *occupation* de ejemplo.

Y por último los vectores de carácter local quedan distribuidos sobre los procesos tal y como muestra la siguiente imagen:

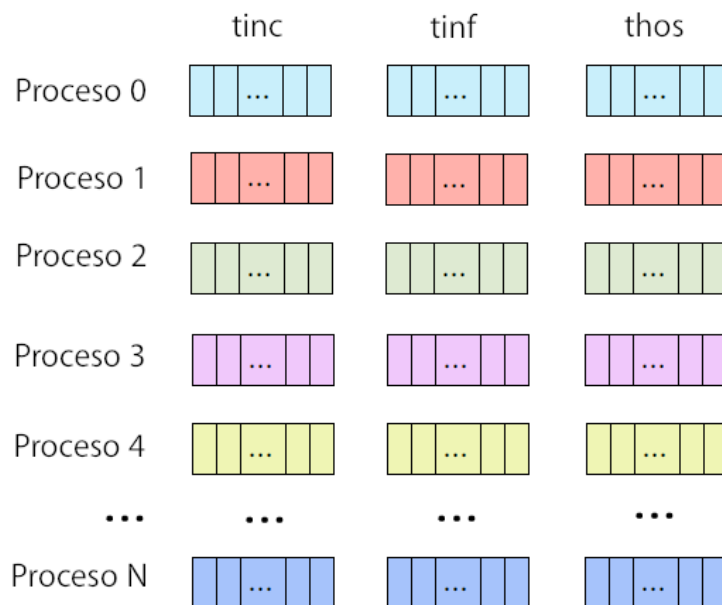


Ilustración 23. Asignación de arrays locales.

Igualmente, cada una de las tareas definidas en la fase previa son asignadas a uno de los procesos en ejecución, quedando de esta manera concluida la fase de *Asignación* de tareas.

## Orquestación

La fase de *Orquestación* es en la que la arquitectura y el modelo de programación, al igual que el lenguaje de programación, juegan un importante rol cuya meta es establecer mecanismos para acceder a los datos de intercambio con otros procesos, además de estructurar la comunicación y sincronización entre los distintos procesos.

### Modelo social

En el modelo social las comunicaciones se establecen entre pares de procesos con el objetivo de transferir la información del grafo compartida entre ambos y que se encuentra asociada a uno y otro respectivamente. El paso de mensajes se realiza de forma síncrona.

La información que se transfiere entre los pares de procesos es la ubicación en el grafo de los conocidos entre pares de colectivos y los cuáles pertenecen uno a cada uno de los procesos. Esta información se transmite en un buffer que contiene la ubicación de las conexiones en formato coordinado: índice de fila, índice de columna y valor del elemento.

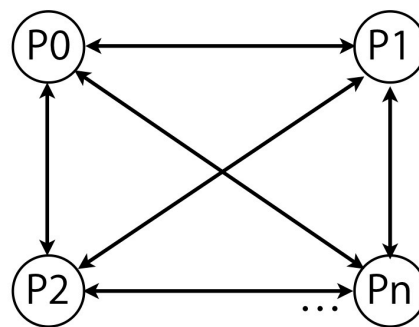


Ilustración 24. Diagrama de comunicaciones del modelo social.

La rutina de comunicación entre los procesos ejecutada por cada uno de ellos se encarga de comprobar qué información necesita compartir con qué otro proceso, realizando los envíos y recepciones de forma ordenada:

**//Rutina de comunicación de conocidos entre distintos procesos**

```

Desde proceso = 0 hasta proceso = total procesos en ejecución
    Si proceso != id_proceso (proceso actual)
        Recibir conocidos desde proceso
        Insertar conocidos en el grafo
    Si No
        Desde proceso = 0 hasta proceso = total procesos en ejecución
            Si proceso != id_proceso (proceso actual)
                Obtener conocidos con proceso y mandarlos al buffer
                Enviar conocidos a proceso
        Fin Desde
    Fin Desde
Fin Desde

```

En un ejemplo con tres procesos en ejecución [P0, P1, P2] y en el que cada uno de los mismos contiene un único colectivo respectivamente [C0, C1, C2] el intercambio de información entre los mismos se realizaría como se muestra en la ilustración 25. En la descripción aparecen los datos transferidos y el sentido de las flechas indica el proceso emisor y receptor de cada mensaje:

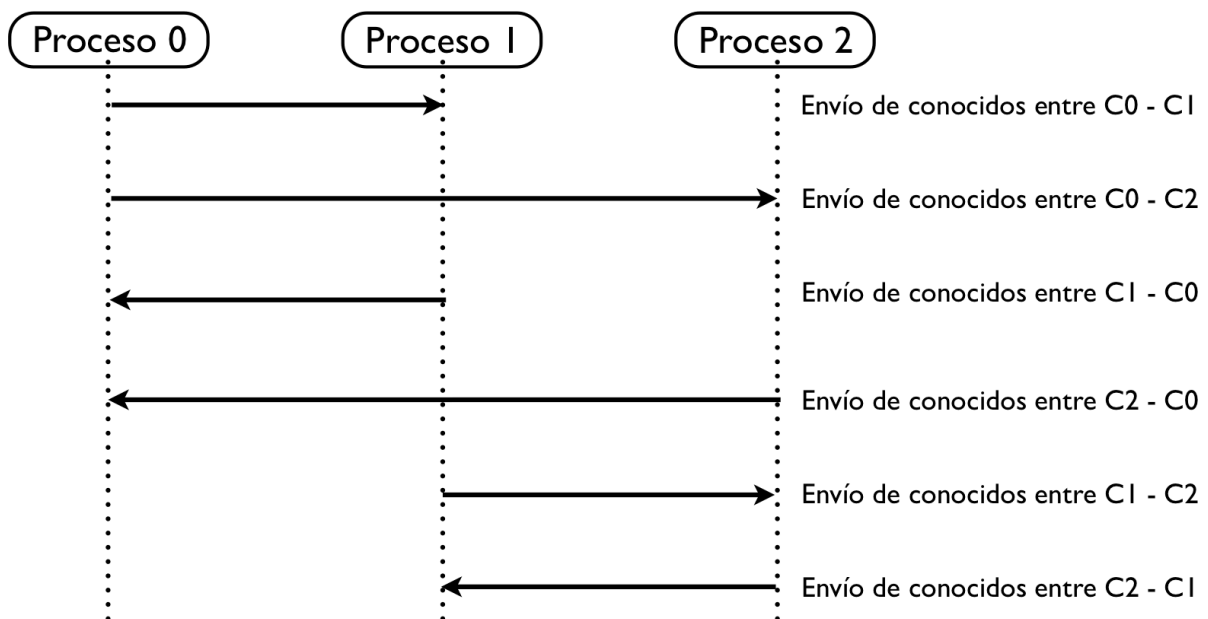


Ilustración 25. Diagrama de flujo del paso de mensajes del modelo social.



## Modelo de enfermedad

En el modelo de enfermedad las comunicaciones se establecen entre todos los procesos en ejecución ejecutando primitivas de reducción. Mediante dichas reducciones se mantiene una copia de los datos de todos los procesos en cada uno de ellos. El paso de mensajes se realiza de forma síncrona.

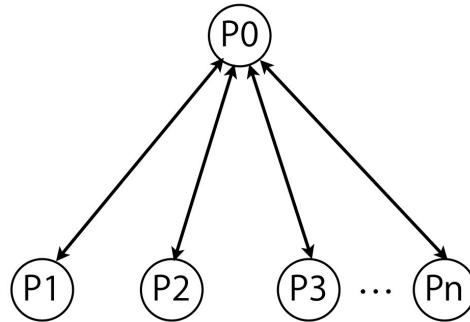


Ilustración 26. Diagrama de comunicaciones del modelo enfermedad.

## Primitivas de MPI para la comunicación de procesos

La aplicación ha sido desarrollada en lenguaje de programación C y el modelo de programación paralelo utilizado es el paso de mensajes (MPI). MPI posee una gran cantidad de funciones dedicadas a la comunicación y sincronización entre procesos, distinguiendo entre funciones adecuadas para transferir datos entre un par de procesos y funciones adecuadas para transferir datos colectivamente (generalmente todos los que se encuentren en ejecución). De entre las funciones existentes para dichos fines han sido necesarias las siguientes:

### Funciones MPI para la comunicación entre dos procesos

Las principales funciones para la comunicaciones entre pares de procesos son las primitivas que operan el envío y recepción de mensajes en modo síncrono:

- `MPI_Send (void *buf, int count, MPI_Datatype datatype, int dest, int tag, MPI_Comm comm)`

La primitiva `MPI_Send` gestiona el envío de un mensaje contenido en el buffer `buf` de tamaño `count` y tipo `datatype` desde un proceso origen al proceso destino `dest`.

- `MPI_Recv (void *buf, int count, MPI_Datatype datatype, int tag, MPI_Comm comm, MPI_Status *status)`

La primitiva `MPI_Recv` gestiona la recepción de un mensaje en el buffer `buf` de tamaño `count` y tipo `datatype`.

Las funciones `MPI_Send` y `MPI_Recv` operan sobre comunicación síncrona, es decir, ambas son funciones bloqueantes ya que el proceso que invoca el envío permanece bloqueado hasta que la comunicación y recepción del mensaje por parte del proceso destino se complete.

## Funciones MPI para comunicaciones colectivas

La primitiva `MPI_AllReduce` permite realizar una operación de reducción sobre los datos almacenados en cada procesador para ser compactados por uno de los procesos en un sólo mensaje que es reenviado al resto de procesos.

- `MPI_AllReduce (void *sendbuf, void *recvbuf, int count, MPI_Datatype datatype, MPI_Op op, int root, MPI_Comm comm)`

La función `MPI_Reduce` combina los datos del buffer `sendbuf` de cada uno de los procesos mediante la operación `op` y devuelve los valores combinados en el buffer `recvbuf` del proceso `root`. Ejemplo operación `MPI_SUM`.

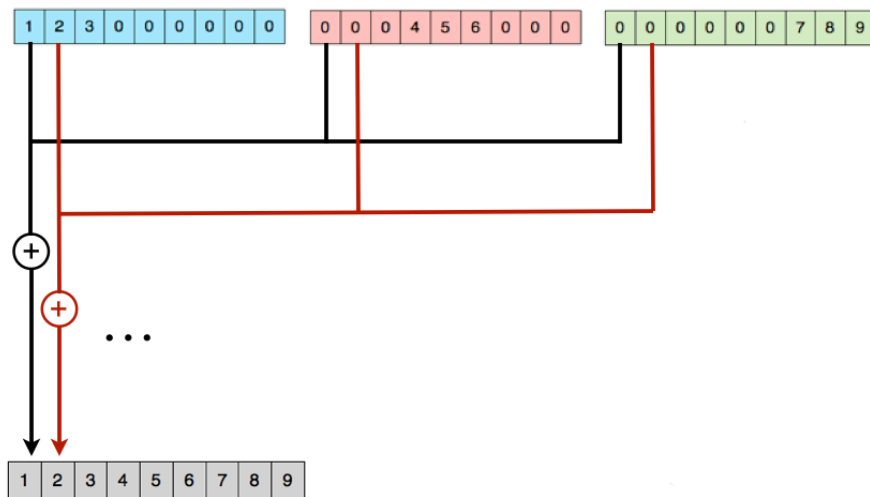


Ilustración 27. Ejemplo de operación `MPI_Reduce`.

## Mapeo

La fase de *Mapeo* hace referencia a la asignación de los procesos sobre los procesadores físicos disponibles. Un escenario ideal es la asignación de un proceso por cada uno de los procesadores existentes, aunque MPI permite gestionar la asignación de más de un proceso por procesador abstrayendo dicha función y haciéndola invisible para el usuario. Es por ello que salvo el escenario ideal, que será el que de forma teórica puede proporcionar el mayor rendimiento a la aplicación ya que permite dotar de toda la potencia de cómputo del procesador al cálculo de las operaciones de un único proceso, la fase de *Mapeo* no presenta ningún tipo de restricción.

# Caso de estudio: Gripe A (H1N1) en Leganés

*Una vez definidos los diferentes modelos de representación así como la implementación seleccionada para llevar a cabo el desarrollo del proyecto se presenta el estudio del problema elegido.*

## Estudio del problema

El problema concreto seleccionado como objeto de estudio en el proyecto es el de la simulación de la transmisión del virus de la Gripe A (H1N1) en el municipio madrileño de Leganés, España.

### Selección de la enfermedad contagiosa del estudio

La enfermedad contagiosa Gripe A (H1N1) o Gripe porcina está producida por un subtipo del virus de la gripe estacional frente al cual los humanos no poseen inmunidad. El pasado 11 de Junio de 2009 la Organización Mundial de la Salud (OMS) declaró el nivel de alerta de pandemia de gripe en fase 6, lo que indica la actividad de una pandemia de carácter mundial, por lo que existe una mayor diseminación del virus.

El contagio del virus de la Gripe A (H1N1) se produce de manera similar al de la gripe estacional: por el contacto persona a persona, principalmente por vía aérea o secreciones nasales, de sudor o al toser, hablar o estornudar. Indirectamente, en caso de que dichas secreciones permanezcan en manos, superficies o dispersas en el aire, el contagio puede producirse por tocarse los ojos, nariz o boca después haber entrado en contacto con ellas.

La transmisión del virus puede producirse desde un día antes de comenzar a percibirse los síntomas de la gripe hasta los siete días posteriores, siendo los tres primeros los de capacidad máxima de contagio. De forma general el individuo infectado evoluciona de forma favorable en dos o cuatro días aunque puede extenderse a una o dos semanas. En España se dispone de dos preparados de vacunación seguros y eficaces que mantienen el virus inactivo en el individuo vacunado.

El Gobierno de España proporcionó informes de situación en [Ministerio de Sanidad y Política Social del Gobierno de España - Gripe A] que fueron discontinuados a partir del 30 de Diciembre de 2009. El último informe de situación constataba hasta el momento:

- La tasa estimada de gripe clínica en atención primaria en la semana del 20 al 26 de Diciembre fue de 40,80 casos por cada 100.000 habitantes.
- 18.232 casos de gripe clínica de los cuales el 100% de los analizados ha resultado positivo para el nuevo virus de la Gripe A (H1N1).
- 271 fallecimientos relacionados con la Gripe A (H1N1).

- La tasa de letalidad desde la semana 27 es de 0,21 fallecidos por cada mil afectados de Gripe A (H1N1).

El [CDC, 2009] aún ofrece de forma semanal información detallada de la actividad de la Gripe A en el territorio de los Estados Unidos:

Influenza Positive Tests Reported to CDC by U.S. WHO/NREVSS Collaborating Laboratories, National Summary, 2009-10

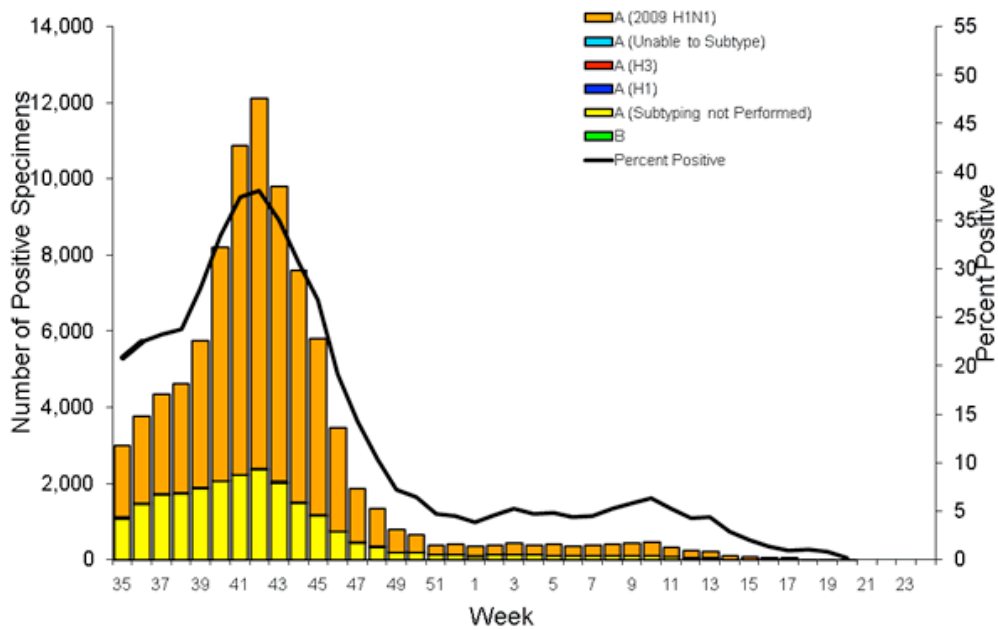


Ilustración 28. Gráfico de casos positivos reportados en Estados Unidos entre las semanas 35 de 2009 y 23 de 2010. [CDC, 2009]

El gráfico superior muestra el número de casos positivos que se producen cada semana en Estados Unidos y su evolución desde la semana 35 de 2009 hasta la 23 de 2010. Estos datos nos serán de gran utilidad a la hora de realizar la evaluación del proyecto, ya que los resultados que se obtengan deben ser similares en cierta medida a los que aquí se presentan.

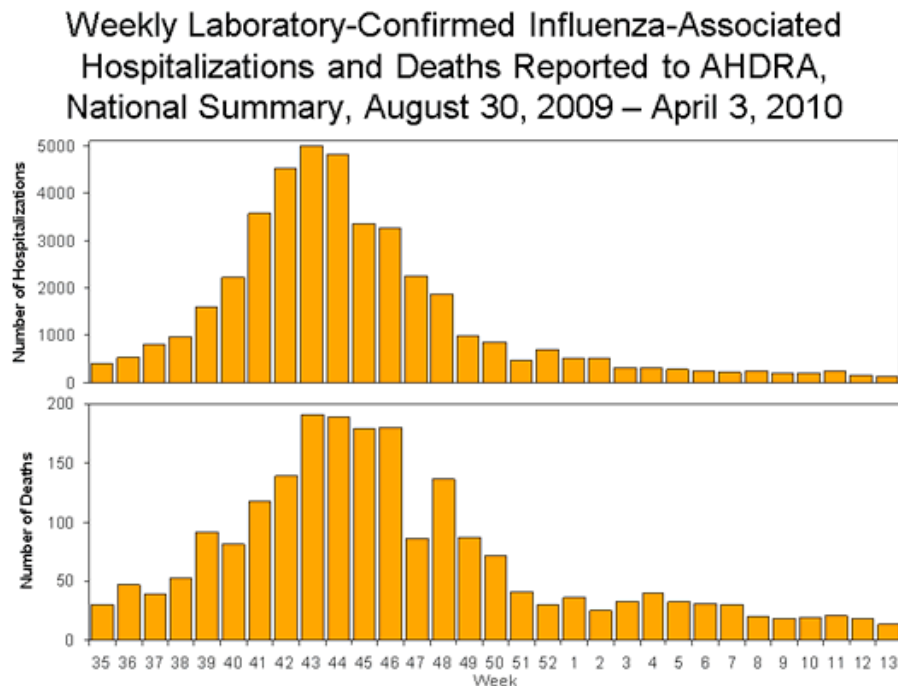


Ilustración 29. Gráfico de hospitalizaciones y defunciones reportadas en Estados Unidos entre las semanas 35 de 2009 y 13 de 2010. [CDC, 2009]

Los siguientes dos gráficos muestran el número de hospitalizaciones y defunciones registradas en Estados Unidos en el periodo desde la semana 35 de 2009 a la semana 13 de 2010. Dichos datos también serán relevantes para determinar el grado de realismo de los resultados obtenidos en la evaluación de la herramienta.

Algunas otras importantes características sobre el comportamiento del virus de la Gripe A (H1N1) a nivel mundial imprescindibles para el estudio son las siguientes:

- Cada individuo infectado con el virus contagia a un número de entre 1.4 y 1.6 individuos sanos.
- Un individuo que ha superado el período de infección (estado de Convalecencia) se vuelve inmune al virus.
- El porcentaje de individuos infectados que requieren de internamiento hospitalario en países desarrollados es del 11%.
- El porcentaje de individuos infectados que fallecen a causa del virus es del 0,007%.

Con las estadísticas e información recopilada es posible construir los parámetros del modelo de enfermedad necesarios para realizar el caso de estudio de la Gripe A, que se describen en el apartado de parámetros del caso de estudio que nos ocupa y posteriormente obtener datos fiables de los supuestos planteados para su simulación.

## Selección de la población del estudio

La población seleccionada como objeto del caso de estudio es el municipio de Leganés, situada al sur de la Comunidad de Madrid, España. El [Ayto de Leganés, 2010] facilita en su página web una serie de estadísticas anuales sobre la población del municipio, entre los que destaca la “Distribución en grupos quinquenales según sexos” de sus 190.540 habitantes censados a fecha de Enero de 2010:

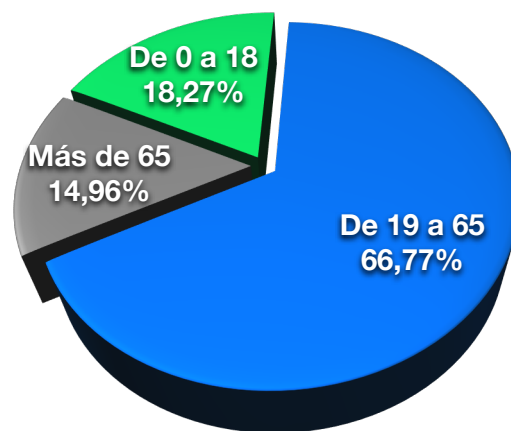
Grupos de edad	Hombres	%	Mujeres	%	Total	%
De 0 a 4	5.072	2,66%	4.906	2,57%	9.978	5,24%
De 5 a 9	4.471	2,35%	4.128	2,17%	8.599	4,51%
De 10 a 14	3.875	2,03%	3.684	1,93%	7.559	3,97%
De 15 a 19	4.371	2,29%	4.304	2,26%	8.675	4,55%
De 20 a 24	5.493	2,88%	5.410	2,84%	10.903	5,72%
De 25 a 29	7.746	4,07%	7.240	3,80%	14.986	7,87%
De 30 a 34	10.535	5,53%	9.693	5,09%	20.228	10,62%
De 35 a 39	9.049	4,75%	8.467	4,44%	17.516	9,19%
De 40 a 44	7.315	3,84%	6.948	3,65%	14.263	7,49%
De 45 a 49	6.271	3,29%	6.411	3,36%	12.682	6,66%
De 50 a 54	5.096	2,67%	5.521	2,90%	10.617	5,57%
De 55 a 59	5.246	2,75%	6.719	3,53%	11.965	6,28%
De 60 a 64	6.709	3,52%	7.355	3,86%	14.064	7,38%
De 65 a 69	5.329	2,80%	5.420	2,84%	10.749	5,64%
De 70 a 74	3.339	1,75%	3.519	1,85%	6.858	3,60%
De 75 a 79	2.216	1,16%	2.781	1,46%	4.997	2,62%
De 80 a 84	1.238	0,65%	1.861	0,98%	3.099	1,63%
De 85 a 89	545	0,29%	1.230	0,65%	1.775	0,93%
De 90 a 94	202	0,11%	511	0,27%	713	0,37%
Más de 94	103	0,05%	211	0,11%	314	0,16%
<b>TOTALES</b>	<b>94.221</b>	<b>49,45%</b>	<b>96.319</b>	<b>50,55%</b>	<b>190.540</b>	<b>100,00%</b>

Tabla 1. “Distribución de grupos quinquenales según sexos” del [Ayto de Leganés, 2010].

Realizando la distribución de la población del municipio sobre cada uno de los colectivos de interés definidos en el modelo social se extraen los siguientes datos:

Edades	Total	%
De 0 a 18 años	34.811	18,27%
De 19 a 65 años	127.224	66,77%
Más de 65 años	28.505	14,96%
<b>TOTAL</b>	<b>190.540</b>	<b>100,00%</b>

Tabla 2. Distribución de población según edad sobre colectivos de interés, con datos del [Ayto de Leganés, 2010].



Si suponemos para el estudio que el 100% de los individuos del primer grupo, de 0 a 18 años, son estudiantes y que el 100% de los individuos del grupo de más de 65 años son jubilados nos encontramos con que el grupo de 19 a 65 años quedará distribuido entre individuos trabajadores y amas de casa. Este grupo tiene un tamaño de 127.224 individuos, de los cuáles según [Ayto de Leganés, 2010] el 50,11% son mujeres, por lo que la distribución de individuos entre 19 y 65 años es de 63.764 mujeres y 63.455 hombres.

Según elaboración de [INE, 2001-2009] en su "Clasificación de la población de 16 y más años según su relación con la actividad económica" el 44,14% de las mujeres de España se encuentran en estado activo. Considerando tal como el porcentaje de mujeres trabajadoras y extrapolándolo a la población del municipio que nos ocupa obtenemos que 28.145 de las mujeres pertenecerán al grupo de trabajadores, junto con los 63.455 hombres que ya había. El número restante de mujeres, 35.624, quedarán agrupadas en el colectivo de amas de casa:

Colectivo	Edades	Total	%
<b>Trabajadores</b>	De 19 a 65 años	91.600	48,07%
<b>Estudiantes</b>	De 0 a 18 años	34.811	18,27%
<b>Ancianos</b>	Más de 65 años	28.505	14,97%
<b>Amas de casa</b>	De 19 a 65 años	35.624	18,69%
<b>TOTAL</b>		<b>190.540</b>	<b>100,00%</b>

Tabla 3. Distribución de población sobre colectivos de interés con datos del [Ayto de Leganés, 2010] e [INE, 2001-2009].



## Parámetros del caso de estudio

Los parámetros del caso de estudio quedan distribuidos en tres diferentes secciones según el ámbito de aplicación de los mismos. Junto al nombre identificativo del parámetro se muestra una breve descripción del mismo junto con el valor obtenido del caso de estudio. Notar que en la fase de evaluación algunos de los parámetros variarán según los distintos supuestos, con el objetivo de construir el escenario adecuado para cada uno de los mismos:

### Parámetros generales

Los parámetros generales se encuentran contenidos en el archivo de configuración "*config.xml*" y son necesarios para el funcionamiento general de la aplicación:

Identificador	Descripción	Valor
<code>ntime</code>	tiempo total de la simulación (en minutos)	525600
<code>tcomm</code>	tiempo que transcurre entre las comunicaciones de los distintos procesos (en minutos)	60
<code>tact</code>	tiempo que transcurre entre las actualizaciones del estado de los individuos de la población (en minutos)	60
<code>test</code>	tiempo que transcurre entre la impresión de estadísticas por pantalla (en minutos)	60
<code>ttrans</code>	tiempo que transcurre entre las transmisiones de la enfermedad entre los individuos (en minutos)	1
<code>tfich</code>	tiempo que transcurre entre el almacenamiento en fichero del estado de los individuos (en minutos)	10000

## Parámetros del modelo social

Los parámetros del modelo social se encuentran distribuidos en dos diferentes archivos de configuración. Los primeros de ellos se encuentran en el archivo *“config\_grafos.xml”* y son necesarios para la generación del archivo intermedio de colectivos. El resto de parámetros se encuentran contenido en el archivo *“config.xml”* y son necesarios para la generación de las conexiones entre los diferentes colectivos del grafo:

### Parámetros del archivo de configuración *“config\_grafos.xml”*

Identificador	Descripción	Valor
num_grafos	número de grafos origen utilizados	2
t_grafo	identificador del grafo origen del tipo de colectivo de Trabajadores	1
t_edad_min	edad mínima de los individuos del colectivo de Trabajadores	19
t_edad_max	edad máxima de los individuos del colectivo de Trabajadores	64
t_perc	porcentaje de población que pertenecerá al colectivo Trabajadores (en valor porcentual)	0,4807
t_grupo_min	tamaño mínimo de un determinado colectivo de tipo Trabajadores	20
t_grupo_max	tamaño máximo de un determinado colectivo de tipo Trabajadores	500
e_grafo	identificador del grafo origen del tipo de colectivo de Estudiantes	2
e_edad_min	edad mínima de los individuos del colectivo de Estudiantes	0
e_edad_max	edad máxima de los individuos del colectivo de Estudiantes	18
e_perc	porcentaje de población que pertenecerá al colectivo Estudiantes (en valor porcentual)	0,1827
e_grupo_min	tamaño mínimo de un determinado colectivo de tipo Estudiantes	20
e_grupo_max	tamaño máximo de un determinado colectivo de tipo Estudiantes	60
a_grafo	identificador del grafo origen del tipo de colectivo de Ancianos	1
a_edad_min	edad mínima de los individuos del colectivo de Ancianos	65

Identificador	Descripción	Valor
a_edad_max	edad máxima de los individuos del colectivo de Ancianos	100
a_perc	porcentaje de población que pertenecerá al colectivo Ancianos (en valor porcentual)	0,1497
a_grupo_min	tamaño mínimo de un determinado colectivo de tipo Ancianos	5
a_grupo_max	tamaño máximo de un determinado colectivo de tipo Ancianos	20
ac_grafo	identificador del grafo origen del tipo de colectivo de Amas de casa	2
ac_edad_min	edad mínima de los individuos del colectivo de Amas de casa	19
ac_edad_max	edad máxima de los individuos del colectivo de Amas de casa	64
ac_perc	porcentaje de población que pertenecerá al colectivo Amas de casa (en valor porcentual)	0,1869
ac_grupo_min	tamaño mínimo de un determinado colectivo de tipo Amas de casa	2
ac_grupo_max	tamaño máximo de un determinado colectivo de tipo Amas de casa	15

### Parámetros del archivo de configuración "config.xml"

Identificador	Descripción	Valor
prob_grupo_debil	probabilidad de conectar un par de colectivos con conexión débil (en valor porcentual)	0,13
prob_grupo_fuerte	probabilidad de conectar un par de colectivos con conexión fuerte (en valor porcentual)	0,01
pop_sim	tamaño total de la población de la simulación	
avgn	estimación del número medio de conexiones de cada individuo de la población	
umbral	umbral de número de conocidos utilizado para realizar el escalado desde el grafo origen (en valor porcentual)	

## Parámetros del modelo de enfermedad

Los parámetros del modelo de enfermedad se encuentran contenidos en el archivo de configuración "*config.xml*" y son necesarios para simular el comportamiento de la transmisión del agente infeccioso:

Identificador	Descripción	Valor
agente_patogeno	nombre identificador del agente infeccioso que se simula	GripeA H1N1
sigma_tinc	componente sigma del tiempo de incubación de la enfermedad	0,8
mu_tinc	componente $\sigma$ del tiempo de incubación de la enfermedad	5
sigma_tinf	componente $\mu$ del tiempo de incubación de la enfermedad	3
mu_tinf	componente $\sigma$ del tiempo de infección de la enfermedad	5
sigma_thos	componente $\mu$ del tiempo de infección de la enfermedad	0,3
mu_thos	componente $\sigma$ del tiempo de hospitalización de la enfermedad	3
vac_time	componente $\mu$ del tiempo de hospitalización de la enfermedad	-
vac_perc	tiempo de aplicación de vacunas sobre la población (en minutos)	0,3
vac_inc_acc	porcentaje de individuos de la población que accede a la vacuna contra el agente infeccioso (en valor porcentual)	0
vac_san_acc	efectividad de la vacuna sobre individuos que estén incubando la enfermedad (en valor porcentual)	0,9
prob_inf_ini	efectividad de la vacuna sobre individuos sanos (en valor porcentual)	0,00527
prob_inf_inc	probabilidad de población infectada al inicio de la simulación (en valor porcentual)	0
prob_inf_ralto	probabilidad en estado de incubación de transmitir la infección a otros individuos (en valor porcentual)	1,8
prob_inf_rbajo	probabilidad en estado de infección de transmitir la infección a otros individuos mediante una conexión de riesgo alto (en número de individuos)	0,3

<b>Identificador</b>	<b>Descripción</b>	<b>Valor</b>
prob_inf_media	probabilidad en estado de infección de transmitir la infección a otros individuos mediante una conexión de riesgo bajo (en número de individuos)	1,5
prob_inf_inm	probabilidad media en estado de infección de transmitir la infección a otros individuos (en número de individuos)	0
prob_inf_a_san	probabilidad en estado inmune de transmitir la enfermedad a otros individuos (en valor porcentual)	0
prob_inf_a_inm	probabilidad de evolucionar de estado infectado a sano transcurrido el periodo de infección (en valor porcentual)	0,883
prob_inf_a_hos	probabilidad de evolucionar de estado infectado a inmune transcurrido el periodo de infección (en valor porcentual)	0,11
prob_inf_a_fal	probabilidad de evolucionar de estado infectado a hospitalización transcurrido el periodo de infección (en valor porcentual)	0,007
prob_hos_a_san	probabilidad de evolucionar de estado infectado a fallecido transcurrido el periodo de infección (en valor porcentual)	0
prob_hos_a_inm	probabilidad de evolucionar de estado hospitalizado a sano transcurrido el periodo de hospitalización (en valor porcentual)	0,98
prob_hos_a_fal	probabilidad de evolucionar de estado hospitalizado a inmune transcurrido el periodo de hospitalización (en valor porcentual)	0,02

# Evaluación

*En este capítulo se realiza la evaluación de la aplicación construida sobre cuatro diferentes supuestos para obtener los resultados de las simulaciones y posteriormente evaluar el rendimiento de la misma.*

## Descripción del entorno de evaluación

El entorno en el que se lleva a cabo la evaluación de la aplicación está constituido por un clúster controlado por PBS (Portable Batch System), un software de gestión de planificación de los trabajos que se lanzan sobre el clúster del dominio de la Universidad Carlos III de Madrid denominado [Kasukabe]. El clúster está compuesto por 16 computadoras con las siguientes características:

- Procesador: AMD Athlon XP 1700+
- Velocidad del procesador: 1467 MHz
- Tamaño de la memoria RAM: 1,5 GB
- Tamaño del disco duro: 40 GB
- Sistema Operativo: Linux Debian Kernel 2.6.26

El gestor de trabajos PBS se encarga de garantizar el uso exclusivo de los recursos del clúster [Kasukabe] a los distintos usuarios, proporcionando comandos para enviar, monitorizar y eliminar trabajos. La conexión al clúster se realiza por SSH, siendo cada uno de los 16 nodos accesibles únicamente desde el sistema de colas.

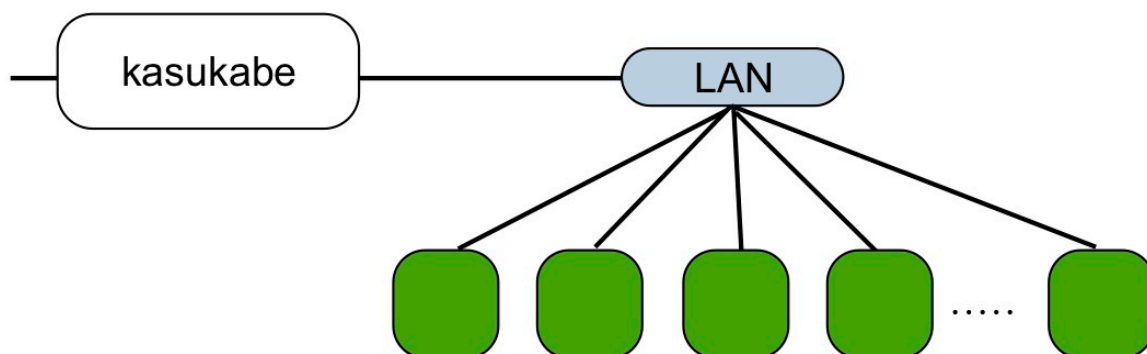


Ilustración 30. [Kasukabe] por [ARCOS 2009].

---

Para llevar a cabo la evaluación de cada uno de los supuestos que se presentan la aplicación ha sido previamente compilada e instalada, para ser posteriormente configurada individualmente para cada una de las diferentes evaluaciones. En el “Anexo 1: Manual de usuario” de este mismo documento se detalla el proceso de compilación, instalación y ejecución de la aplicación.

## Evaluación de los supuestos

### Supuesto 1

En el primer supuesto se propone la evaluación con el conjunto de parámetros estándar definidos en el capítulo de caso de estudio para la Gripe A [H1N1] sobre la población de Leganés. Los valores de los principales parámetros que intervienen son los siguientes:

- Tamaño de la población [pop\_sim]: 190.540 habitantes.
- Tiempo simulado [ntime]: 525.600 minutos (365 días).
- Probabilidad de conectividad débil entre colectivos [prob\_grupo\_debil]: 13 %.
- Probabilidad de conectividad fuerte entre colectivos [prob\_grupo\_fuerte]: 1 %.
- Probabilidad de infección para conexiones de riesgo alto [prob\_inf\_ralto]: 1,8 individuos de media.
- Probabilidad de infección para conexiones de riesgo bajo [prob\_inf\_rbajo]: 0,3 individuos de media.
- Probabilidad de infección media [prob\_inf\_media]: 1,5 individuos.
- Porcentaje de individuos infectados iniciales [prob\_inf\_ini]: 0,527 %.
- Sin vacunaciones.

Con este conjunto de valores se pretende generar un grafo que se asemeje lo más posible a la realidad de la población objeto de estudio y que el comportamiento de la propagación del virus sea lo más realista posible.

El tamaño del grafo en términos de número de conexiones entre individuos es  $\approx 12.000.000$ , por lo que existe una media de  $\approx 63$  conocidos por individuo de la población, lo que se traduce en una media de  $\approx 63$  posibles enlaces de propagación del virus.

Una vez realizada la simulación anual del supuesto los resultados del estado final de la población son los siguientes:

Estado	Número de individuos	%
Sanos	188.873	99,13%
Incubando	0	0,00%
Infectados	0	0,00%
Inmunes	1.653	0,87%
Hospitalizados	0	0,00%
Fallecidos	14	0,01%
<b>Total</b>	<b>190.540</b>	<b>100,00%</b>

Tabla 4. Resultados estadísticos de la evaluación del supuesto 1.



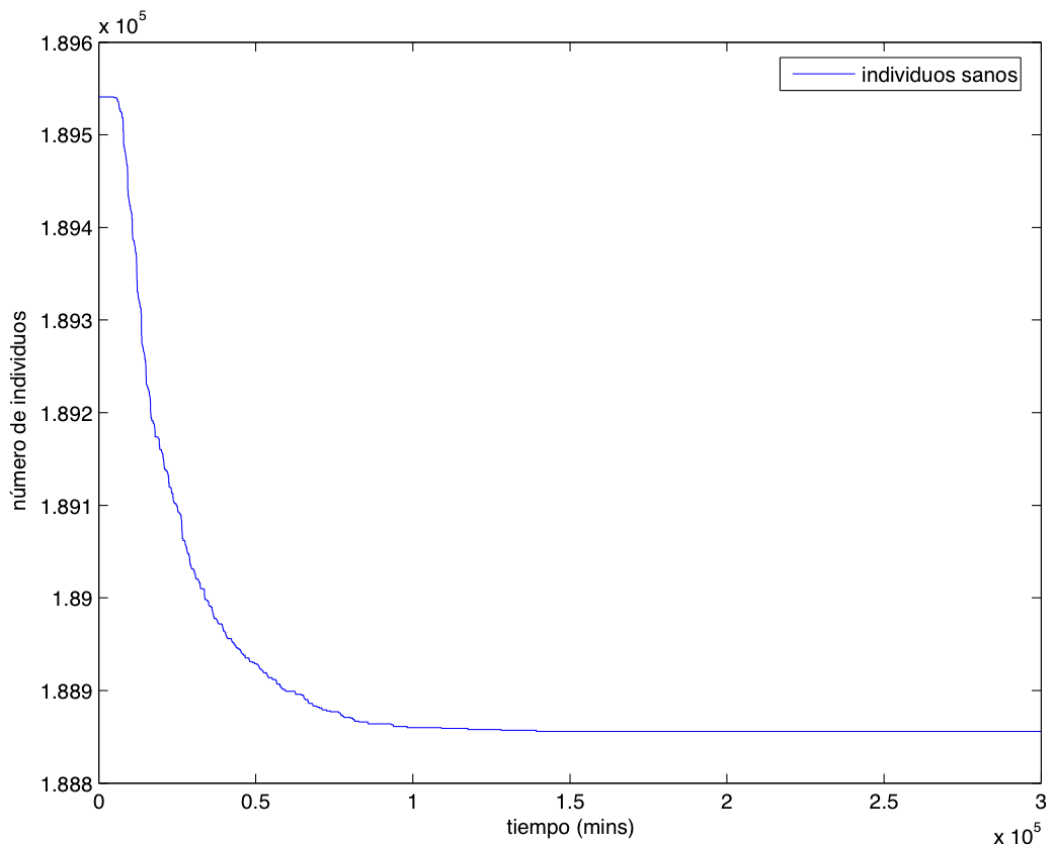


Ilustración 31. Gráfica en escala decimal de la evolución en el tiempo de los individuos en estado sano para el supuesto 1.

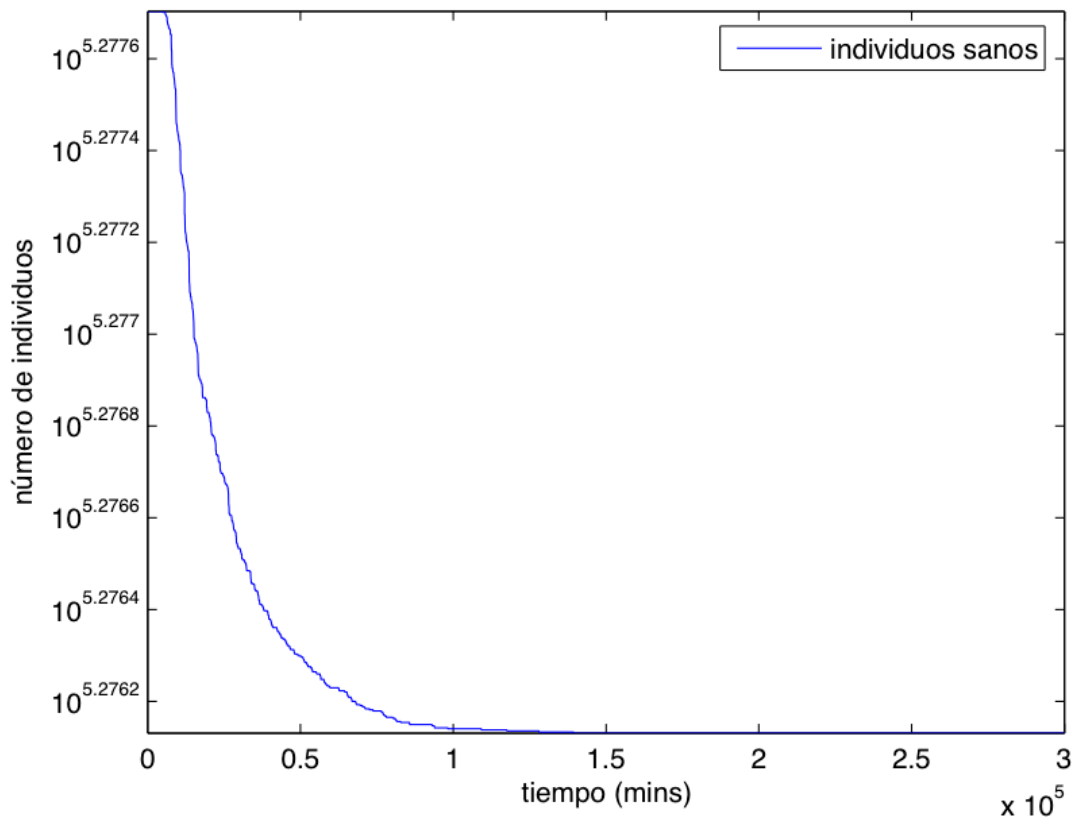


Ilustración 32. Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en estado sano para el supuesto 1.

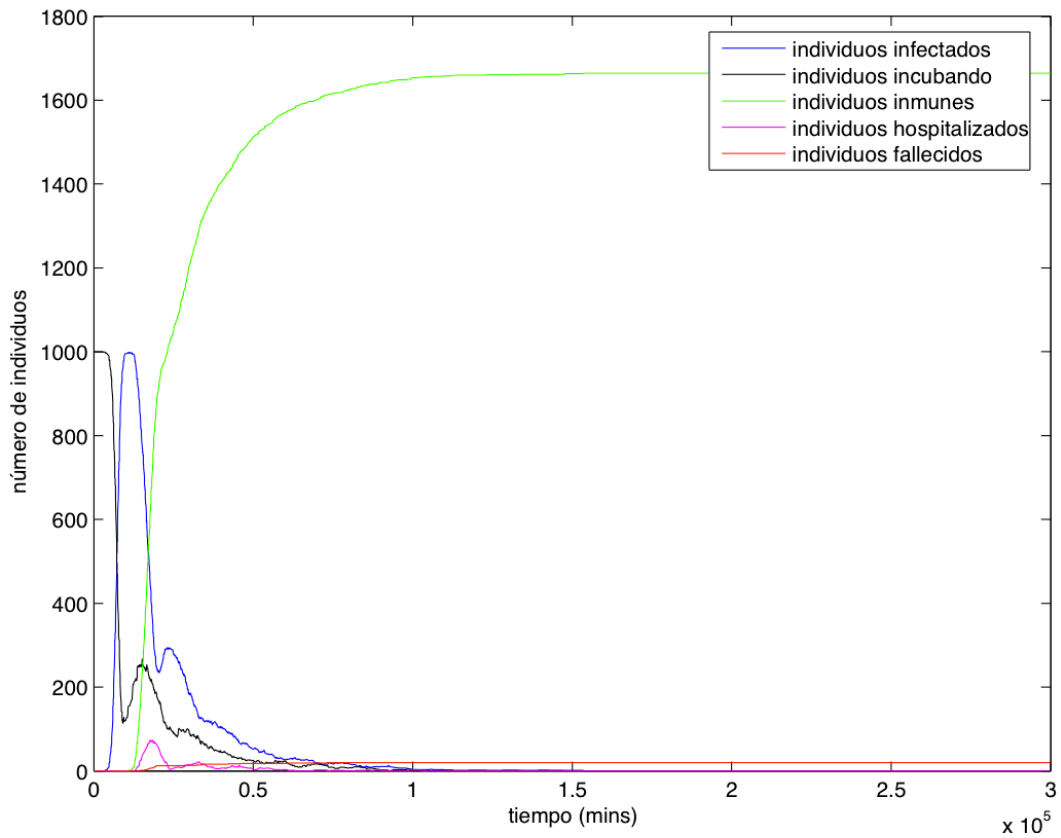


Ilustración 33. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 1.

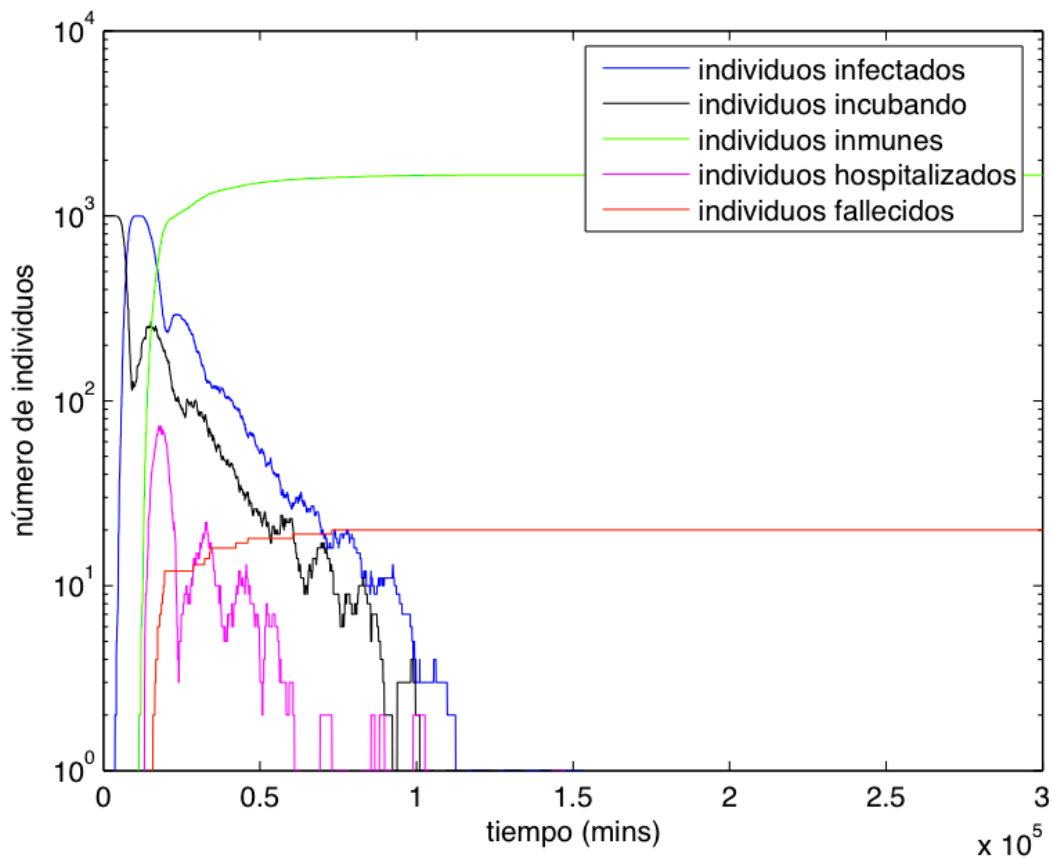


Ilustración 34. Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 1.

En las gráficas se aprecia la evolución en el tiempo de los individuos sanos e infectados, incubando, inmunes, hospitalizados y fallecidos del supuesto.

Una de las principales valoraciones que se puede realizar a la hora de analizar los resultados gráficos del supuesto es que la propagación del virus sobre la población no se prolonga más allá de los 88 días (aproximadamente 125.000 minutos) desde el punto inicial en el que se produce el foco de infección.

La probabilidad de infección inicial provoca  $\approx 1.000$  individuos contagiados con el virus, que lo propagan sobre otros  $\approx 70\%$ , lo que indica que de cada  $\approx 63$  conocidos de media por persona y con una probabilidad media de 1,5 contagios la propagación es inferior a un individuo por cada contagio inicial.

Observando la relación entre individuos en estado de incubación, infectados y hospitalizados es interesante comprobar que el primer pico de individuos infectados producido alrededor de los 8 días de comenzar la infección ( $\approx 11.000$  minutos) provoca una segunda oleada de contagios que se refleja en un repunte de los individuos en estado de incubación y hospitalización alrededor del décimo día ( $\approx 15.000$  minutos). Este dato es de importante relevancia para realizar previsiones de infecciones y hospitalizaciones cuando un determinado porcentaje de individuos es contagiado por el virus.

## Supuesto 2

El segundo supuesto que se propone es la evaluación del conjunto de parámetros estándar a excepción de la modificación de los valores de conectividad entre los colectivos, los cuáles se duplican. Los valores de los principales parámetros que intervienen son los siguientes:

- Tamaño de la población [pop\_sim]: 190.540 habitantes.
- Tiempo simulado [ntime]: 525.600 minutos (365 días).
- Probabilidad de conectividad débil entre colectivos [prob\_grupo\_debil]: 26 %.
- Probabilidad de conectividad fuerte entre colectivos [prob\_grupo\_fuerte]: 2 %.
- Probabilidad de infección para conexiones de riesgo alto [prob\_inf\_ralto]: 1,8 individuos de media.
- Probabilidad de infección para conexiones de riesgo bajo [prob\_inf\_rbajo]: 0,3 individuos de media.
- Probabilidad de infección media [prob\_inf\_media]: 1,5 individuos.
- Porcentaje de individuos infectados iniciales [prob\_inf\_ini]: 0,527 %.
- Sin vacunaciones.

El efecto que produce la modificación de los valores de los parámetros descritos es el notable aumento en el tamaño del grafo cifrado en número de conexiones. Si en el supuesto con valores estándar (supuesto 1) el tamaño del grafo era de  $\approx 12.000.000$  para este segundo supuesto es de  $\approx 22.000.000$ , por lo que ahora existe una media de  $\approx 115$  conocidos por individuo de la población, lo que se traduce en que los posibles enlaces de propagación del virus por cada uno de los individuos es prácticamente el doble.

El efecto que se produce analizado en materia de resultados finales de una simulación anual del supuesto es:

Estado	Número de individuos	%
Sanos	187.567	98,44%
Incubando	0	0,00%
Infectados	0	0,00%
Inmunes	2.944	1,55%
Hospitalizados	0	0,00%
Fallecidos	29	0,02%
<b>Total</b>	<b>190.540</b>	<b>100,00%</b>

Tabla 5. Resultados estadísticos de la evaluación del supuesto 2.

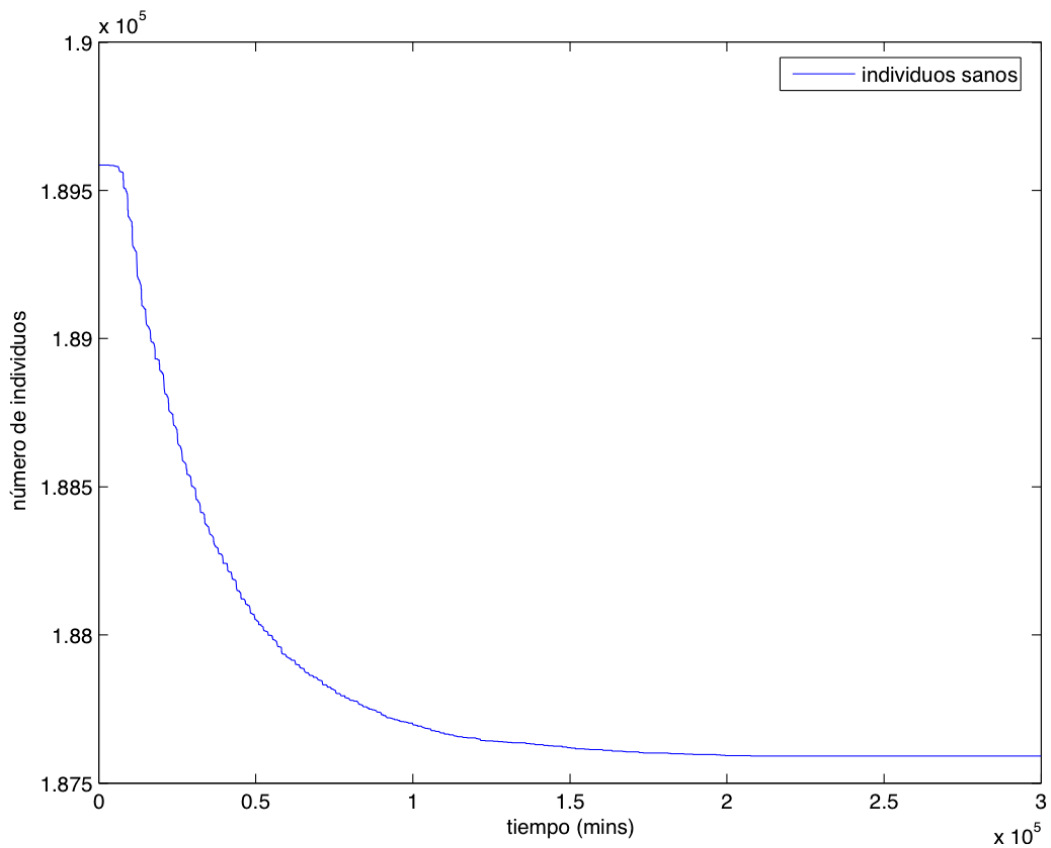


Ilustración 35. Gráfica en escala decimal de la evolución en el tiempo de los individuos en estado sano para el supuesto 2.

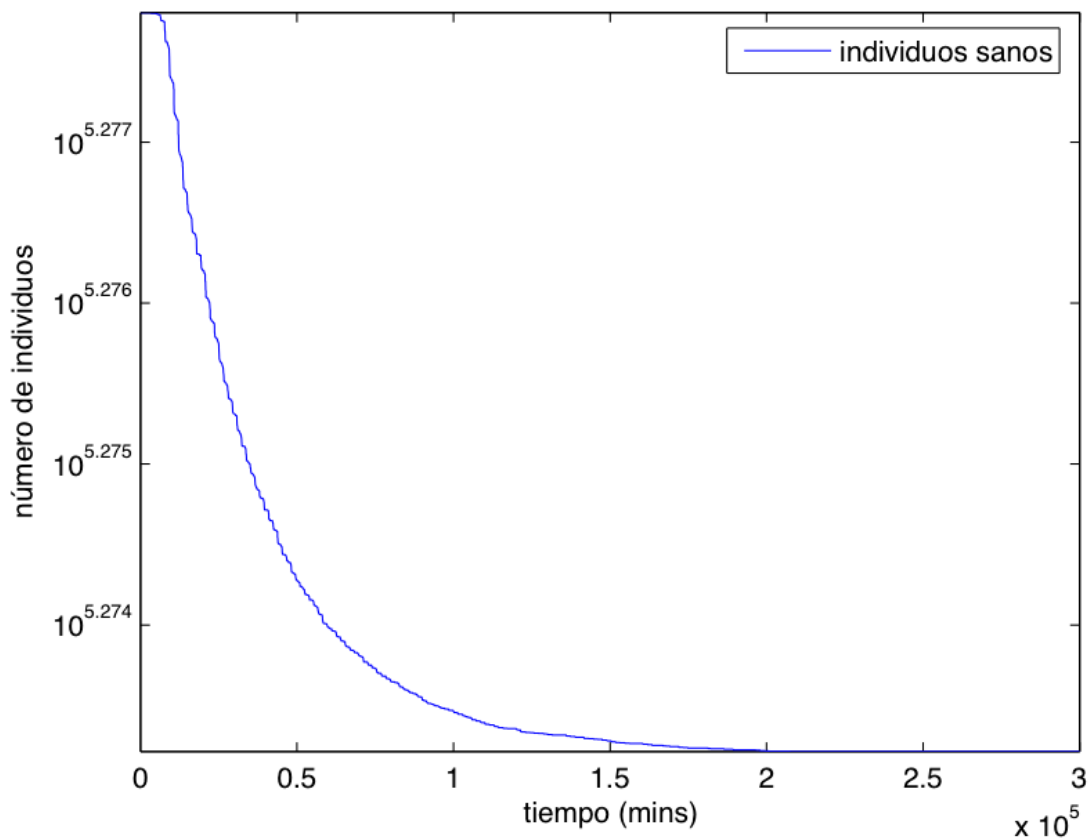


Ilustración 36. Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en estado sano para el supuesto 2.

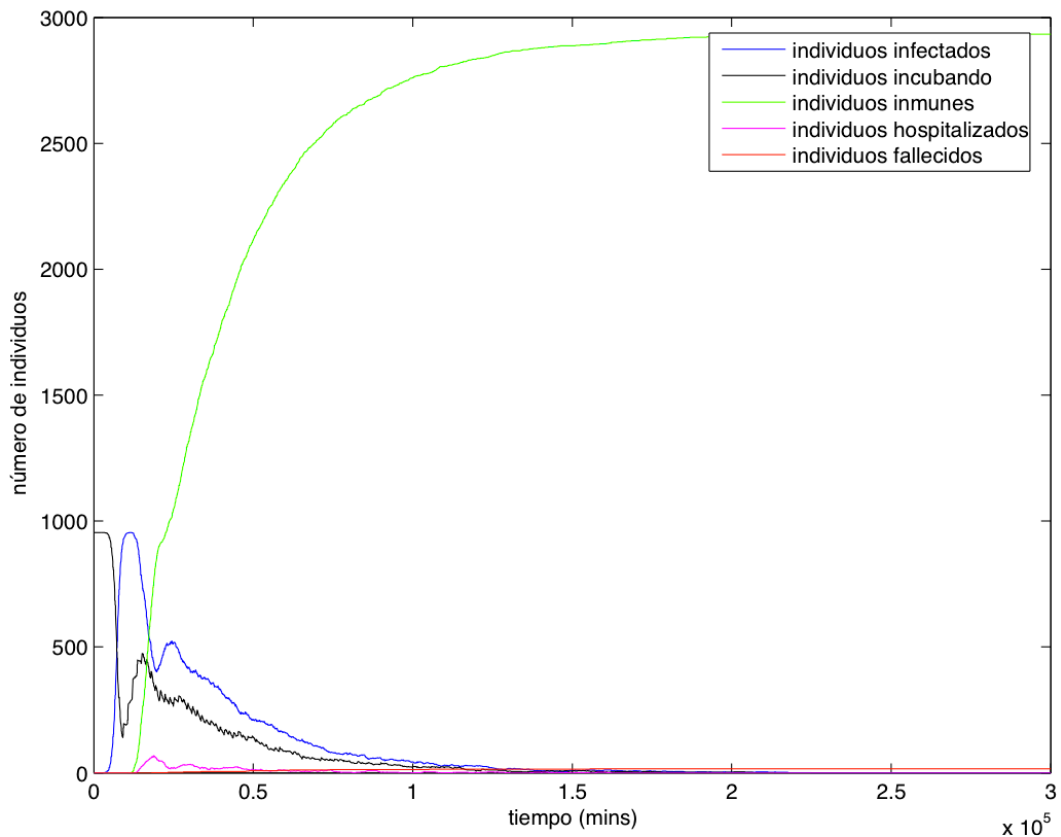


Ilustración 37. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 2.

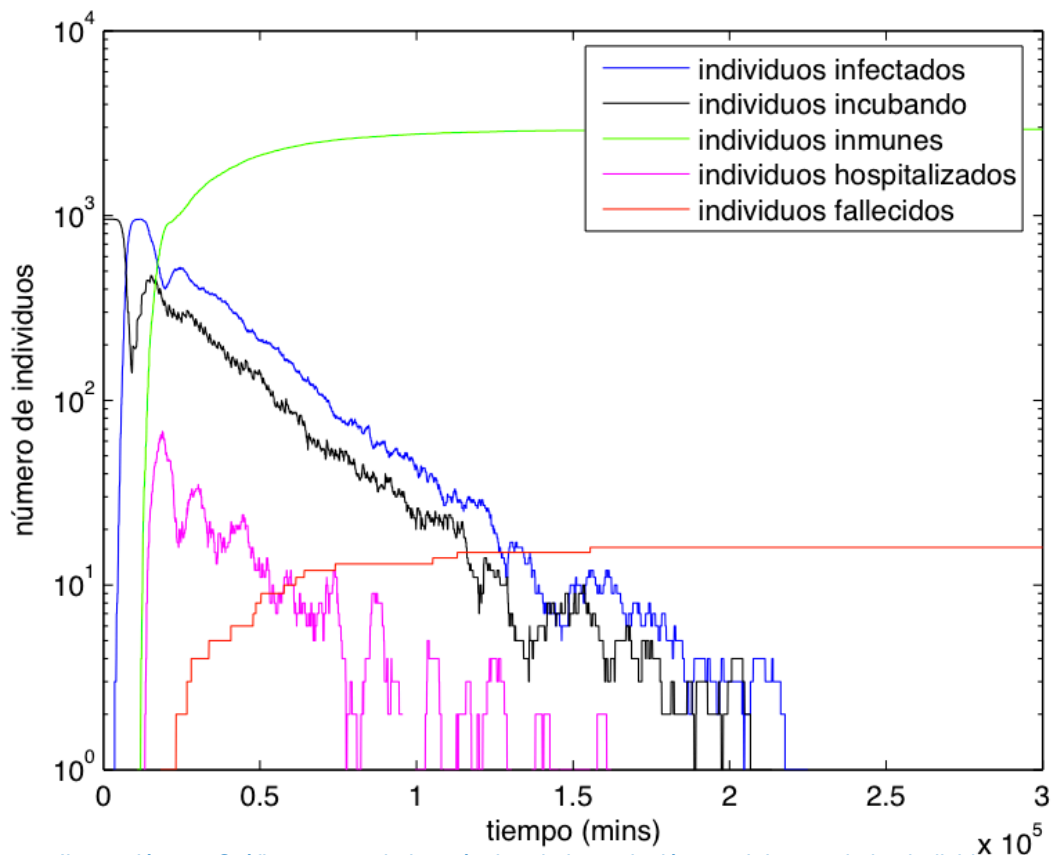


Ilustración 38. Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 2.

Al efecto anteriormente citado producido sobre el tamaño del grafo se añade principalmente el comportamiento de los individuos en estado de incubación e infección respecto al primer supuesto. Si bien se observa que en ambos casos se produce un repunte del número de individuos en cada estado proporcional al tamaño del grafo, en el caso del segundo supuesto la disminución de personas infectadas e incubando disminuye más lentamente ya que se produce una mayor cantidad de contagios.

Especialmente significativo es el hecho de que el número de individuos infectados y fallecidos por causa del virus es directamente proporcional al tamaño del grafo respecto al primer supuesto. Si duplicamos el tamaño del grafo se duplica el número de individuos en los estados citados.

En este caso la propagación del virus se prolonga durante más de 152 días (aproximadamente 220.000 minutos), también proporcional respecto al primer supuesto.

La probabilidad de infección inicial provoca igualmente  $\approx 1.000$  individuos contagiados por el virus, que lo propagan sobre otros  $\approx 300\%$ , lo que indica que de cada  $\approx 115$  conocidos de media por persona y una media de 1,5 contagios la propagación es de 2 infectados por cada individuo contagiado inicialmente.

También se observa cómo existe relación entre los individuos que se encuentran en los estados de incubación, infección y hospitalización, ya que la evolución de la gráfica de individuos en incubación se replica en los otros dos estados.

### Supuesto 3

En el tercer supuesto se propone es la evaluación del conjunto de parámetros estándar a excepción de la modificación de los valores de propagación del virus, los cuáles se duplican para dotar al mismo de mayor agresividad. Los valores de los principales parámetros que intervienen son los siguientes:

- Tamaño de la población [pop\_sim]: 190.540 habitantes.
- Tiempo simulado [ntime]: 525.600 minutos (365 días).
- Probabilidad de conectividad débil entre colectivos [prob\_grupo\_debil]: 13 %.
- Probabilidad de conectividad fuerte entre colectivos [prob\_grupo\_fuerte]: 1 %.
- Probabilidad de infección para conexiones de riesgo alto [prob\_inf\_ralto]: 3,6 individuos de media.
- Probabilidad de infección para conexiones de riesgo bajo [prob\_inf\_rbajo]: 0,6 individuos de media.
- Probabilidad de infección media [prob\_inf\_media]: 3 individuos.
- Porcentaje de individuos infectados iniciales [prob\_inf\_ini]: 0,527 %.
- Sin vacunaciones.

En este supuesto se espera un incremento del número de individuos infectados debido a las nuevas características de propagación del mismo. El tamaño del grafo permanece como en el primer supuesto en los  $\approx 12.000.000$ , por lo que existe una media de  $\approx 63$  conocidos por individuo de la población, lo que se traduce en una media de  $\approx 63$  posibles enlaces de propagación del virus.

El efecto que se produce analizado en materia de resultados finales de una simulación anual del supuesto es:

Estado	Número de individuos	%
Sanos	180.896	94,94%
Incubando	0	0,00%
Infectados	0	0,00%
Inmunes	9.553	5,01%
Hospitalizados	0	0,00%
Fallecidos	91	0,05%
<b>Total</b>	<b>190.540</b>	<b>100,00%</b>

Tabla 6. Resultados estadísticos de la evaluación del supuesto 3.



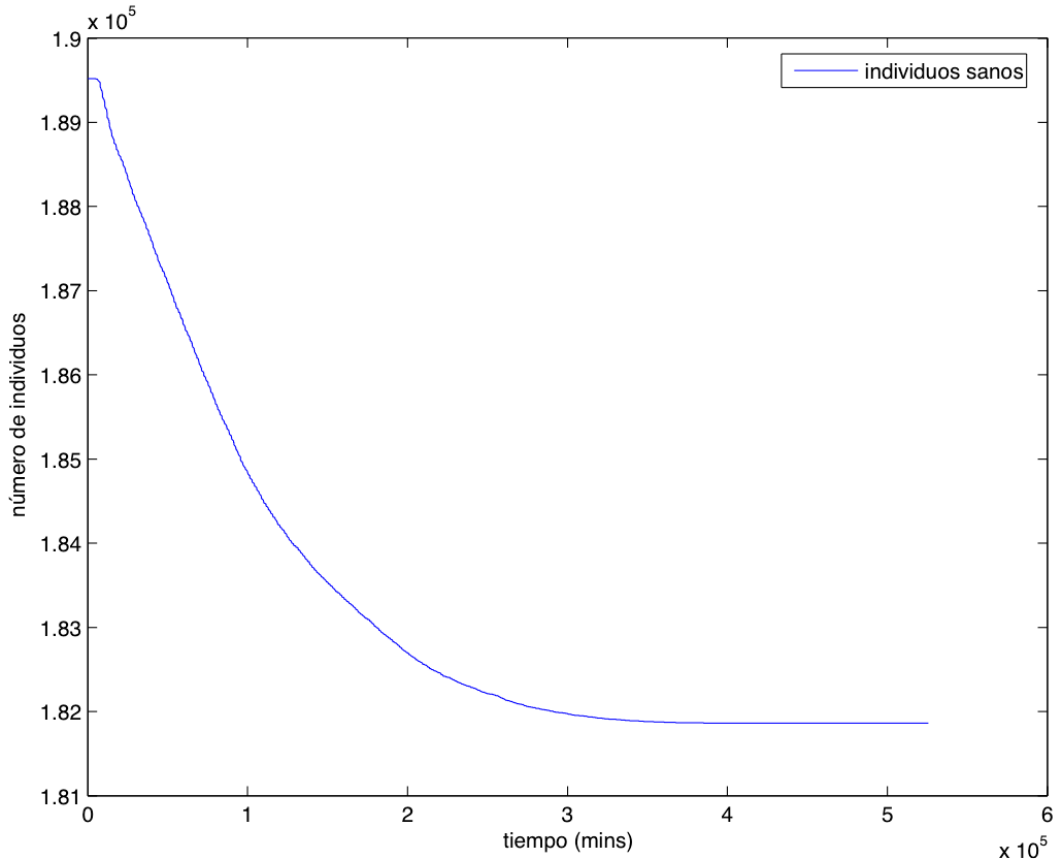


Ilustración 39. Gráfica en escala decimal de la evolución en el tiempo de los individuos en estado sano para el supuesto 3.

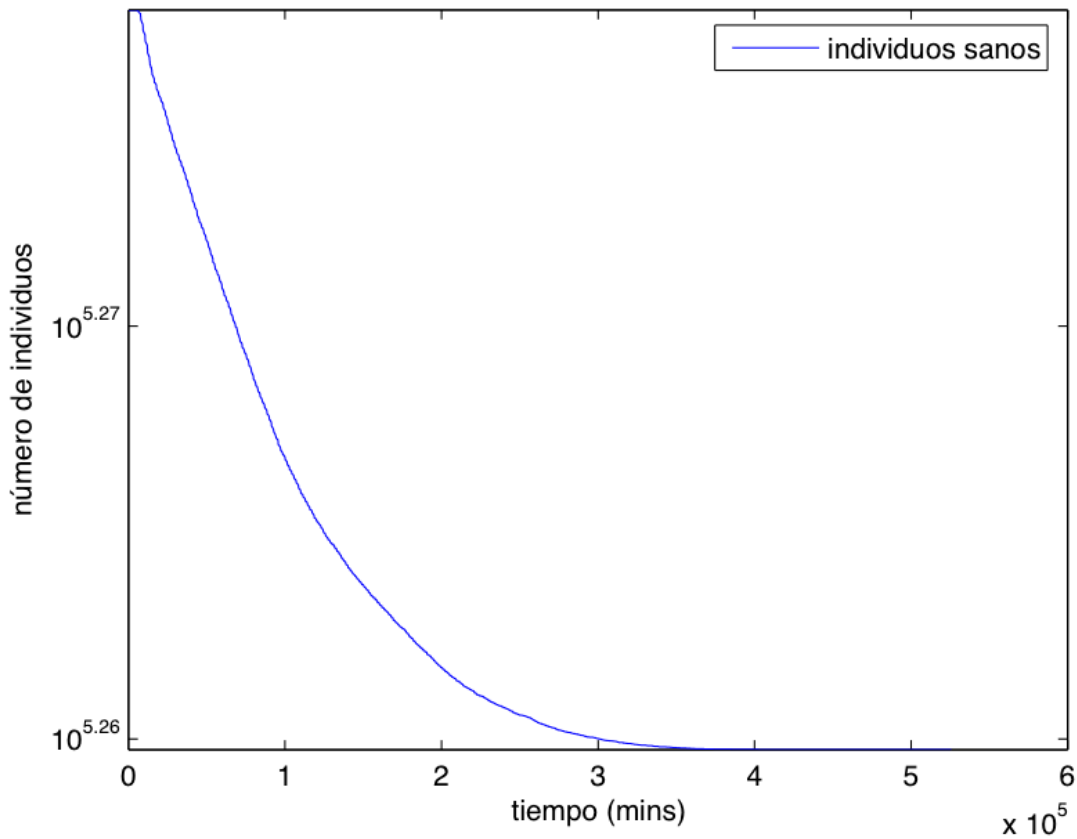


Ilustración 40. Gráfica en escala decimal de la evolución en el tiempo de los individuos en estado sano para el supuesto 3.

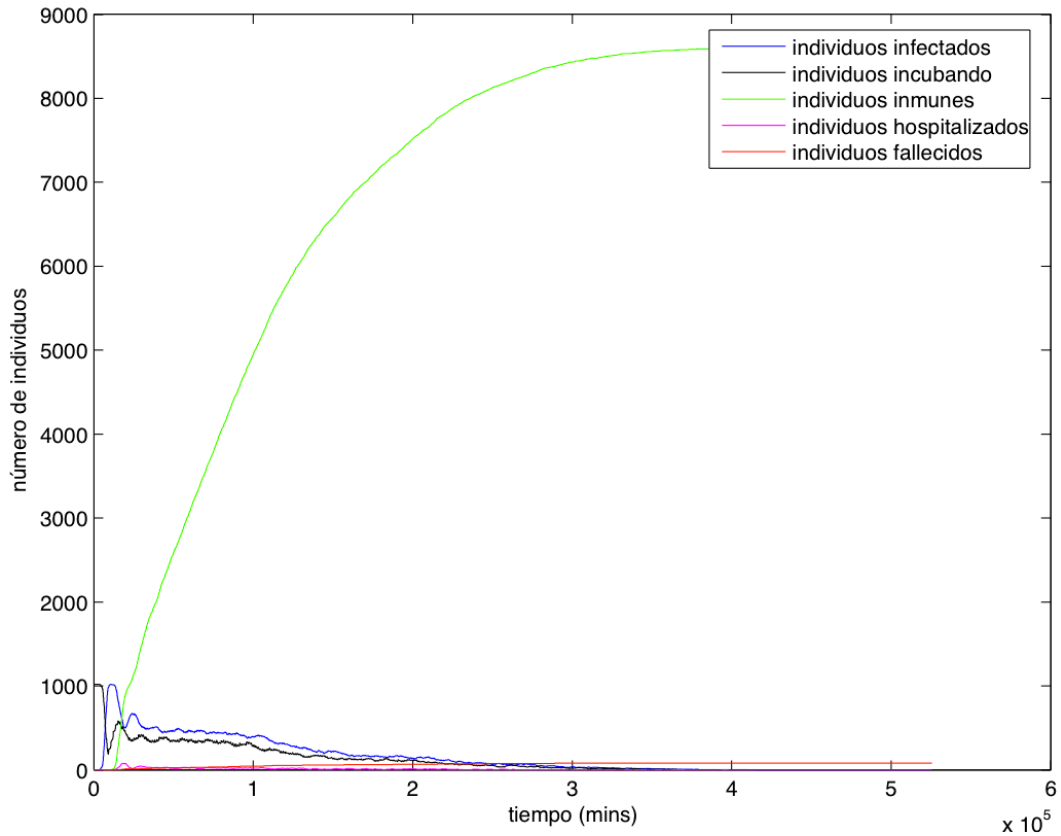


Ilustración 41. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 3.

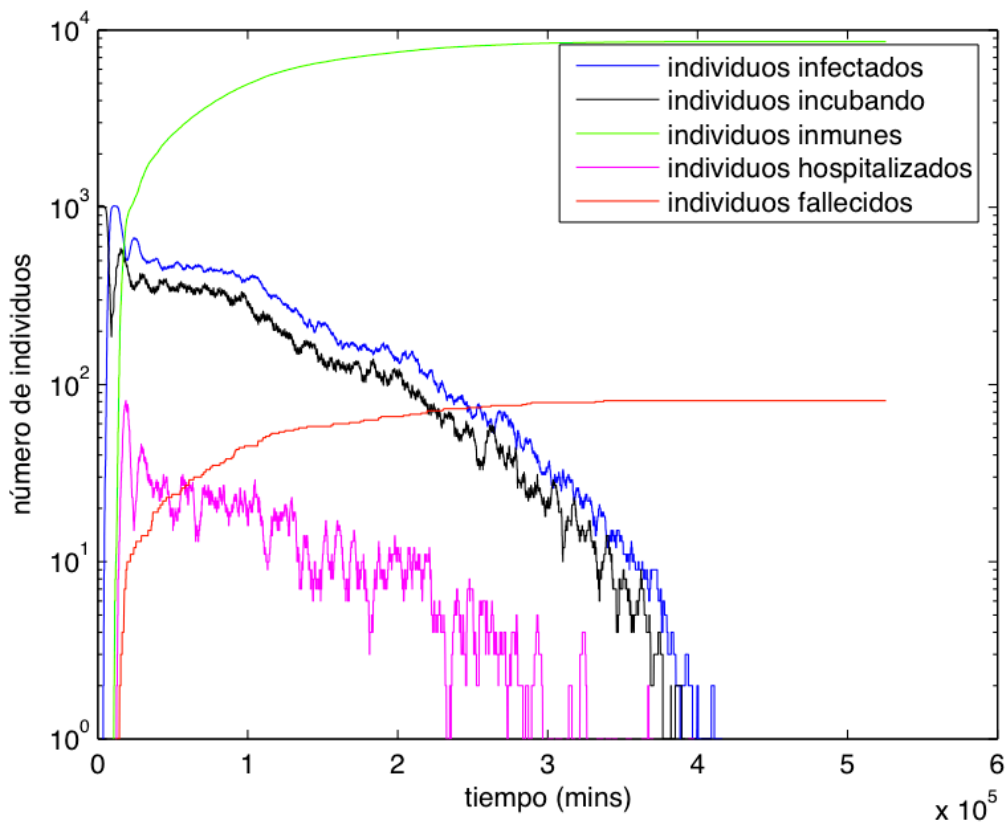


Ilustración 42. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el supuesto 3.

Nada más observar las estadísticas de individuos se comprueba cómo las nuevas características de propagación del virus tienen un efecto directo en las infecciones y defunciones.

En este supuesto la propagación del virus se prolonga durante más de 284 días (aproximadamente 410.000 minutos), y se observa cómo el comportamiento de la gráfica es de una pendiente mucho menos inclinada que en los supuestos anteriores.

La probabilidad de infección inicial provoca como en casos anteriores  $\approx 1.000$  individuos contagiados por el virus, que lo propagan sobre otros  $\approx 900\%$ , lo que indica que de cada  $\approx 63$  conocidos de media por persona y una media de 3 contagios la propagación es de 8 infectados por cada individuo contagiado inicialmente.

Al igual que en supuestos anteriores se observa la relación que existe entre los individuos que se encuentran en los estados de incubación, infección y hospitalización, replicando la evolución de la gráfica de individuos que incuban la enfermedad en los otros dos.

## Supuesto 4

El cuarto supuesto propone la evaluación del conjunto de parámetros estándar a excepción de la modificación del valor de media de conexiones del grafo, con el objetivo de valorar los diferentes efectos que provocan los diferentes valores. Los valores de los principales parámetros que intervienen son los siguientes:

- Tamaño de la población [pop\_sim]: 190.540 habitantes.
- Tiempo simulado [ntime]: 525.600 minutos (365 días).
- Probabilidad de conectividad débil entre colectivos [prob\_grupo\_debil]: 13 %.
- Probabilidad de conectividad fuerte entre colectivos [prob\_grupo\_fuerte]: 1 %.
- Probabilidad de infección para conexiones de riesgo alto [prob\_inf\_ralto]: 1,8 individuos de media.
- Probabilidad de infección para conexiones de riesgo bajo [prob\_inf\_rbajo]: 0,3 individuos de media.
- Probabilidad de infección media [prob\_inf\_media]: 1,5 individuos.
- Porcentaje de individuos infectados iniciales [prob\_inf\_ini]: 0,527 %.
- Sin vacunaciones.
- Número medio de conexiones del grafo [media\_sp]: 30, 90 y 120.

En este supuesto se esperan diversas variaciones en la propagación del virus ya que las modificaciones del parámetro *media\_sp* deben producir un aumento (en el caso del valor 30) o disminución (en los casos de los valores 90 y 120) de la probabilidad de infección entre los individuos. El tamaño del grafo es el estándar,  $\approx 12.000.000$ , por lo que existe una media de  $\approx 63$  conocidos por individuo de la población, lo que se traduce en una media de  $\approx 63$  posibles enlaces de propagación del virus.

Los efectos producidos por las variaciones del parámetro *media\_sp* sobre la propagación del virus analizado en materia de resultados finales de una simulación anual del supuesto es:

Valor parámetro	media_sp [30]		media_sp [90]		media_sp [120]	
	Número de individuos	%	Número de individuos	%	Número de individuos	%
<b>Sanos</b>	178.145	93,49%	189.153	99,27%	189.265	99,33%
<b>Incubando</b>	4	0,00%	0	0,00%	0	0,00%
<b>Infectados</b>	8	0,00%	0	0,00%	0	0,00%
<b>Inmunes</b>	12.260	6,43%	1.374	0,72%	1.262	0,66%
<b>Hospitalizados</b>	0	0,00%	0	0,00%	0	0,00%
<b>Fallecidos</b>	123	0,06%	13	0,01%	13	0,01%
<b>Total</b>	190.540	100,00%	190.540	100,00%	190.540	100,00%

Tabla 7. Resultados estadísticos de la evaluación del supuesto 4.

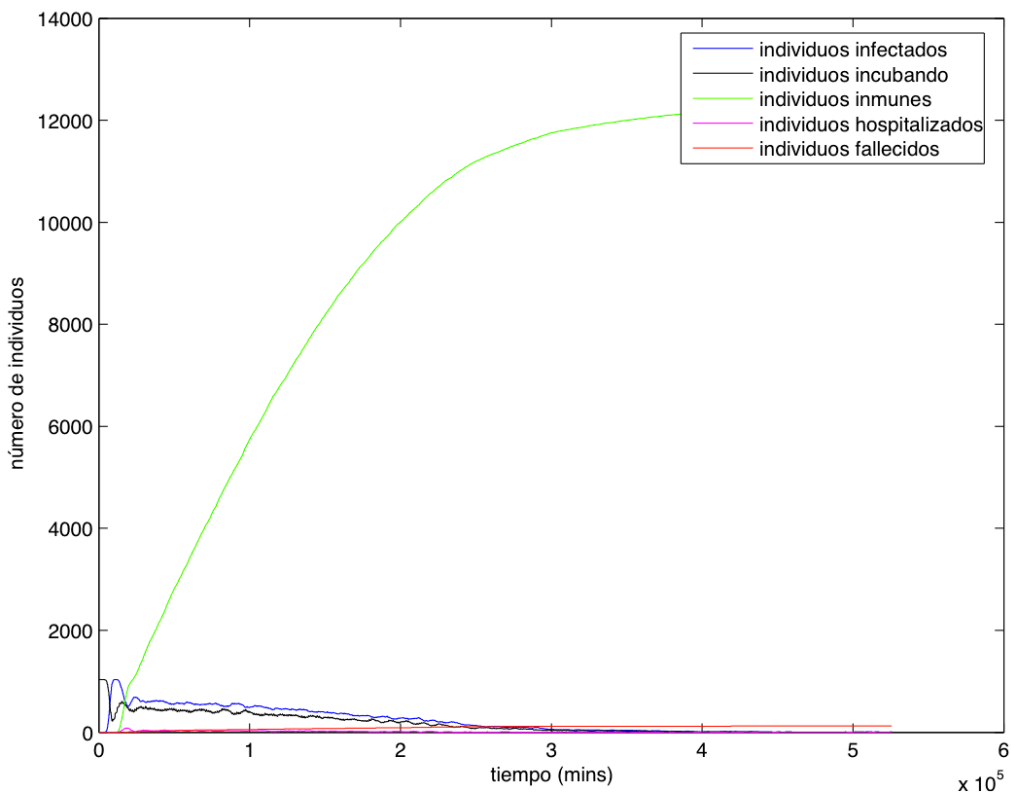


Ilustración 43. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor  $media\_sp = 30$ .

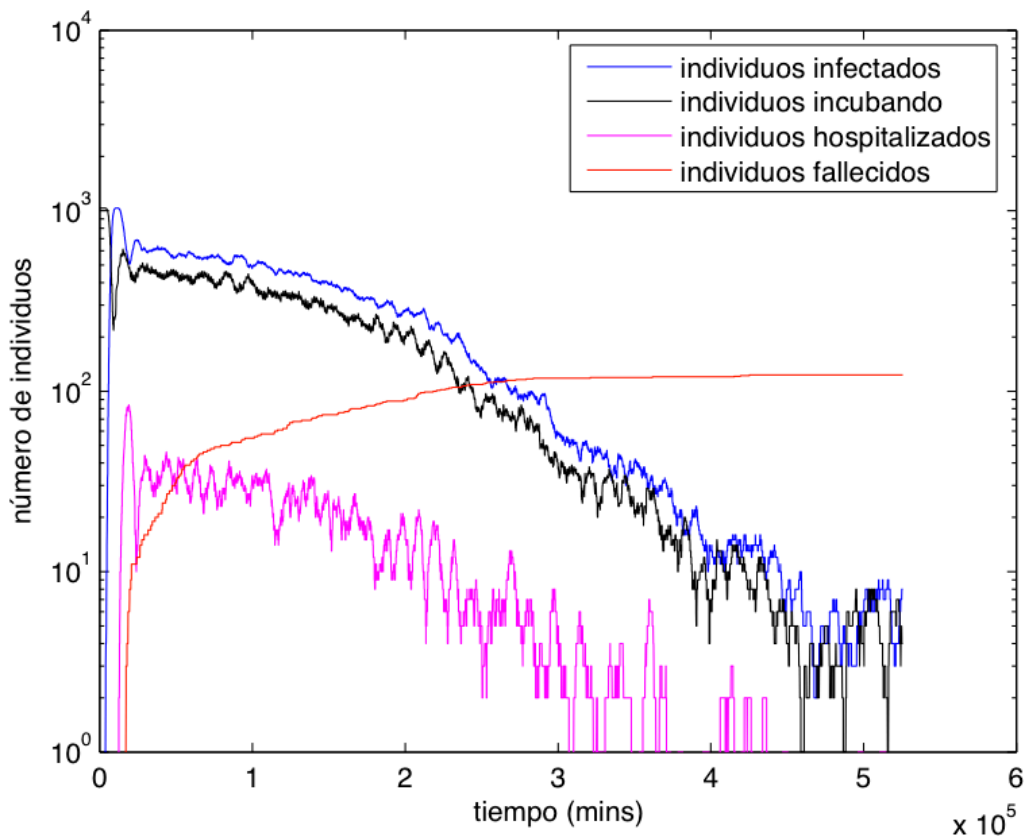


Ilustración 44. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor  $media\_sp = 30$ .

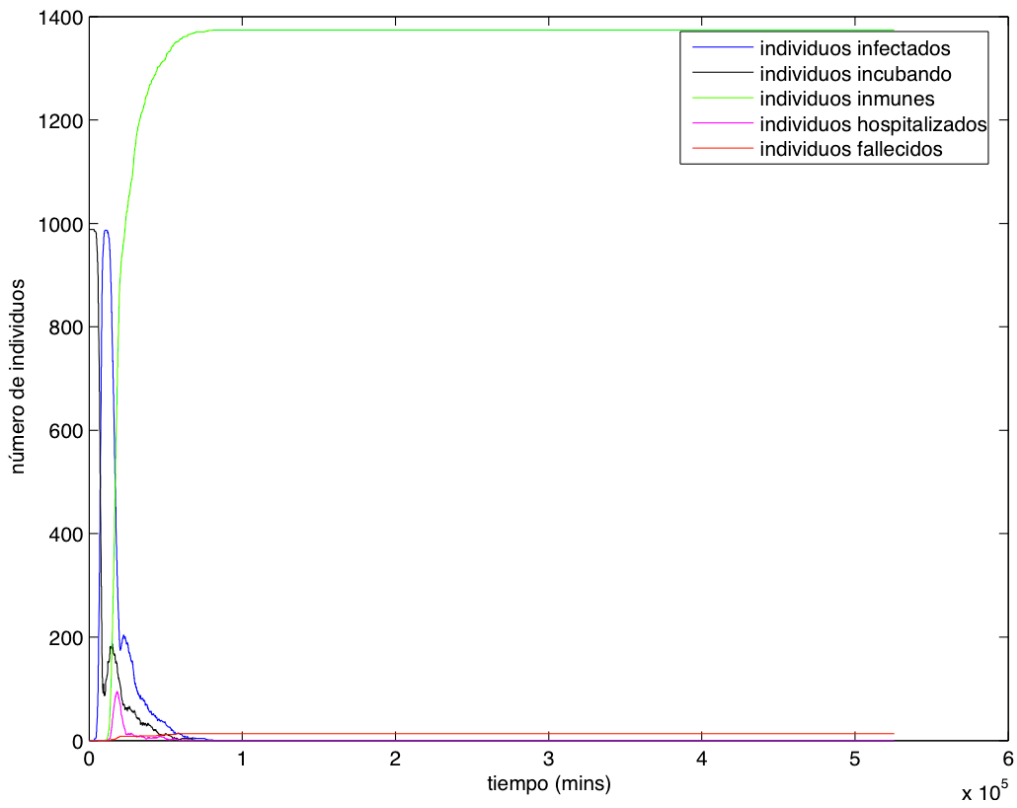


Ilustración 45. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor  $media\_sp = 90$ .

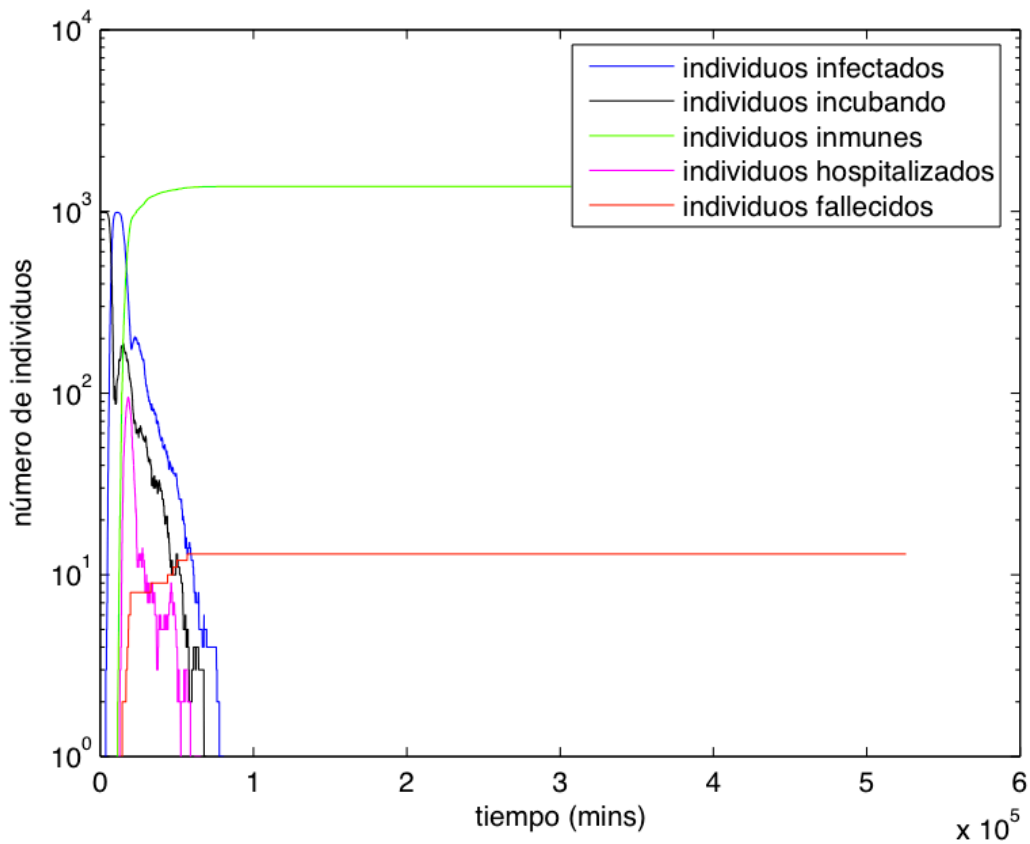


Ilustración 46. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor  $media\_sp = 90$ .

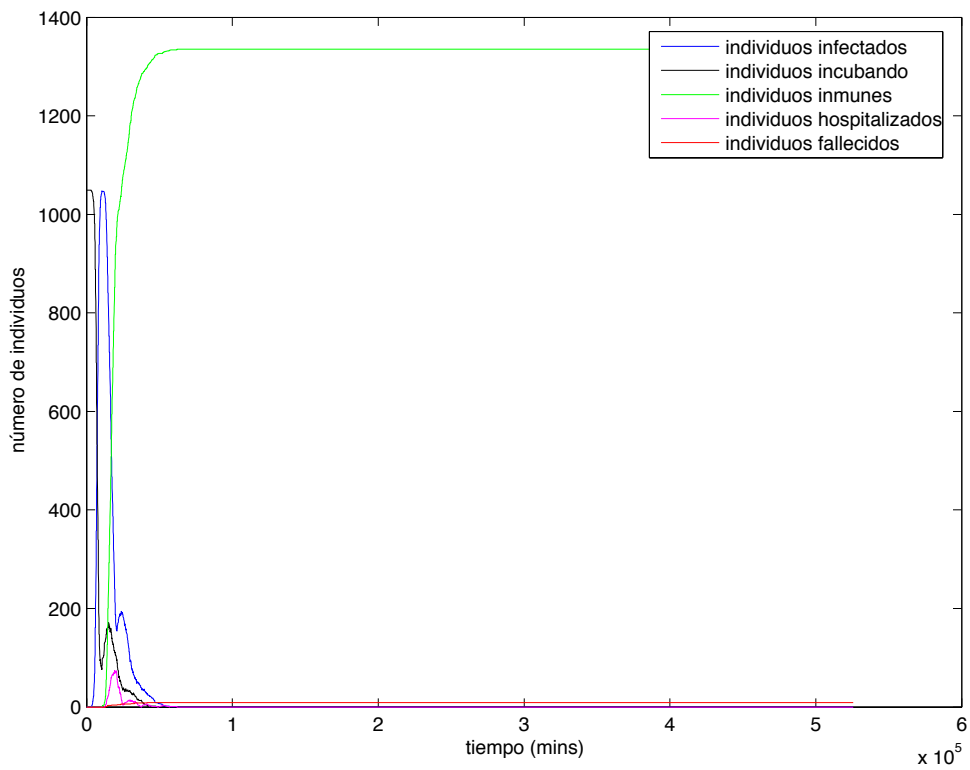


Ilustración 47. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor  $media\_sp = 120$ .

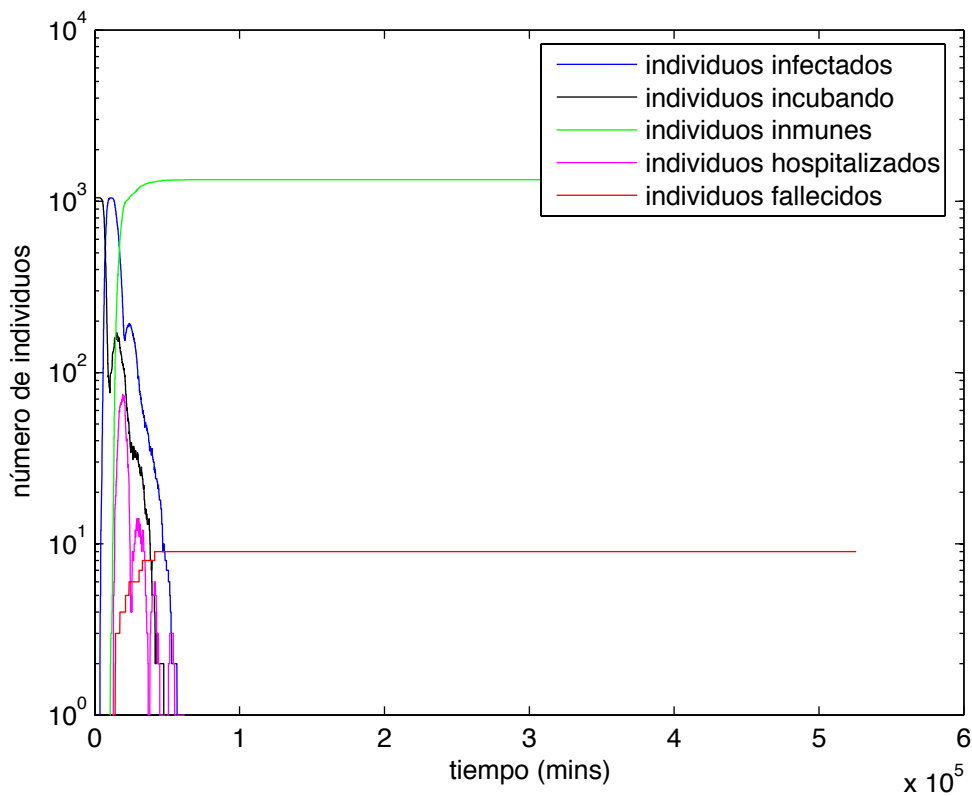


Ilustración 48. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para el valor  $media\_sp = 120$ .

Si comparamos los datos obtenidos por cada una de las modificaciones realizadas con los resultados de la simulación del supuesto 1 se comprueba que el efecto producido por la variación del parámetro es el esperado, ya que según se incrementa su valor el número de individuos contagiados ha de disminuir incrementalmente.

Especialmente significativo es el hecho de que en el caso del valor 30 el número de infectados sea 6 veces mayor que en caso del valor por defecto, 60. En caso contrario, en el que se duplica el valor y se lleva hasta 120, el número de infectados tiene un descenso del 30% aproximadamente.

Resulta interesante en los casos de valor 30 y 120 cómo se prolonga en el tiempo la propagación del virus, ya que en el primer caso se mantiene activa durante todo el tiempo de simulación (un año), mientras que en el segundo caso apenas alcanza los 40 días (60.000 minutos.)

Respecto a las gráficas de resultados, y tal y como ocurría en anteriores supuestos, se comprueba cómo el número de contagios iniciales provoca una reacción en cadena en el número de infectados y hospitalizados.



## Supuesto 5

En el quinto y último supuesto se propone la evaluación del efecto de la aplicación de campañas de vacunación de individuos sobre el supuesto 1. Los valores de los principales parámetros que intervienen son los siguientes:

- Tamaño de la población [pop\_sim]: 190.540 habitantes.
- Tiempo simulado [ntime]: 525.600 minutos (365 días).
- Probabilidad de conectividad débil entre colectivos [prob\_grupo\_debil]: 13 %.
- Probabilidad de conectividad fuerte entre colectivos [prob\_grupo\_fuerte]: 1 %.
- Probabilidad de infección para conexiones de riesgo alto [prob\_inf\_ralto]: 1,8 individuos de media.
- Probabilidad de infección para conexiones de riesgo bajo [prob\_inf\_rbajo]: 0,3 individuos de media.
- Probabilidad de infección media [prob\_inf\_media]: 1,5 individuos.
- Porcentaje de individuos infectados iniciales [prob\_inf\_ini]: 0,527 %.
- Porcentaje de la población que accede a la vacuna [vac\_perc]: 30 %.
- Probabilidad de éxito de la vacuna sobre un individuo en estado de incubación [vac\_inc\_acc]: 0 %.
- Probabilidad de éxito de la vacuna sobre un individuo en estado sano [vac\_san\_acc]: 90 %.
- Instante de aplicación de la vacuna [vac\_time]: 11.000 y 15.000 minutos.

Para llevar a cabo la simulación del efecto de una campaña de vacunación en la propagación del virus se toma como referencia el supuesto 1, que se corresponde con el escenario estándar. Partiendo de dicho supuesto se decide aplicar una campaña de vacunación en dos diferentes momentos: cuando se produce el primer gran volumen de infectados y cuando se produce el primer gran volumen de hospitalizaciones.

De esta forma, resultará interesante y comprobar cómo se desarrolla la propagación del virus respecto a un escenario sin vacunaciones, del que ya se poseían datos anteriores. La primera campaña de vacunación, que coincide con el primer gran volumen de infectados en el primer supuesto, se producirá a los 11.000 minutos. En otro supuesto la campaña de vacunación se producirá a los 15.000 minutos, coincidiendo con el primer gran volumen de hospitalizaciones.

La campaña de vacunación afectará al 30% de la población sana o en estado de incubación, con diferentes probabilidades de surtir efecto en cada uno de ellos. Será interesante comprobar cuál es el momento más adecuado para lanzar una campaña de tal calado y no desaprovechar inoculantes que, en el caso real de la Gripe A (H1N1), fueron escasos durante los primeros meses de la propagación.

La comparativa entre los resultados del supuesto estándar y de las campañas de vacunación son los siguientes, siendo "Vacunación A" la campaña aplicada a los 11.000 minutos y "Vacunación B" los resultados si aplicamos la vacunación a los 15.000 minutos:

Escenario	Supuesto 1		Vacunación A		Vacunación B	
Estado	Número de individuos	%	Número de individuos	%	Número de individuos	%
<b>Sanos</b>	188.873	99,13%	137.964	72,41%	137.781	72,31%
<b>Incubando</b>	0	0,00%	0	0,00%	0	0,00%
<b>Infectados</b>	0	0,00%	0	0,00%	0	0,00%
<b>Inmunes</b>	1.653	0,87%	52.566	27,59%	52.743	27,68%
<b>Hospitalizados</b>	0	0,00%	0	0,00%	0	0,00%
<b>Fallecidos</b>	14	0,01%	10	0,01%	16	0,01%
<b>Total</b>	190.540	100,00%	190.540	100,00%	190.540	100,00%

Tabla 8. Resultados estadísticos de la evaluación del supuesto 5.

Como se puede observar en la tabla sólo con los datos estadísticos no se puede emitir una valoración objetiva sobre los efectos de las vacunaciones ya que el número de individuos inmunes incluye tanto a los individuos contagiados como vacunados. El único dato en claro del que disponemos es el número de personas fallecidas, con un número similar en los tres casos, por lo que en este aspecto se puede concluir que el número de defunciones no sufre variaciones de relevancia.

Es por ello que resulta necesario el análisis de las gráficas de evolución en el tiempo para confirmar si se produce algún tipo de efecto en la población en cada uno de los casos:

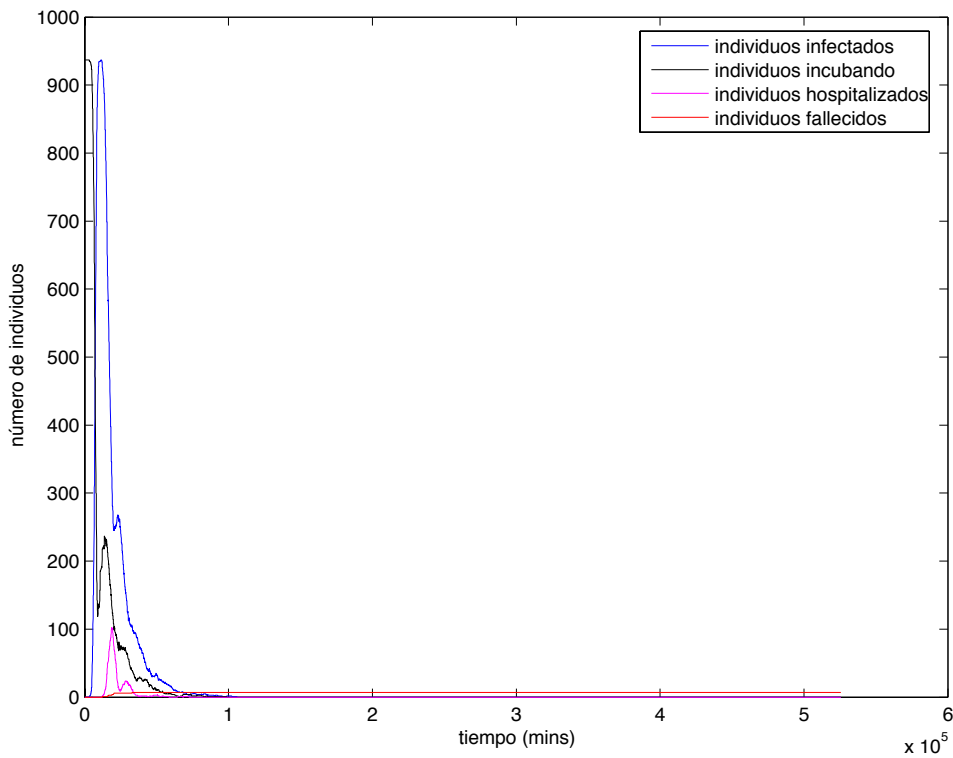


Ilustración 49. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación A.

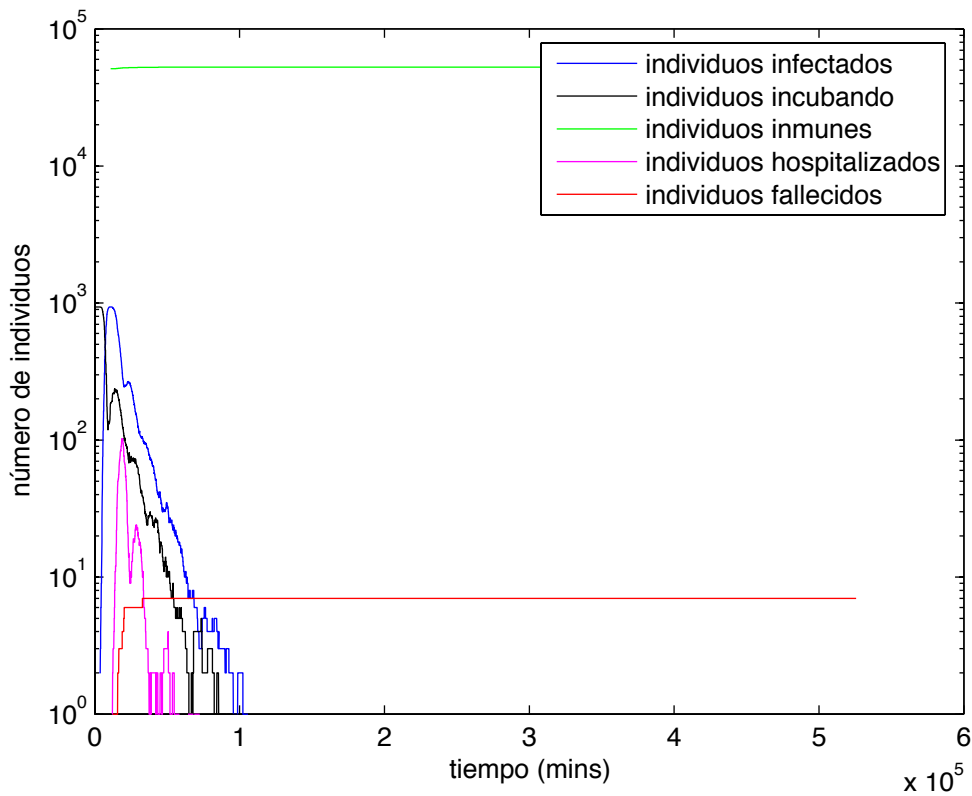


Ilustración 50. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación A.

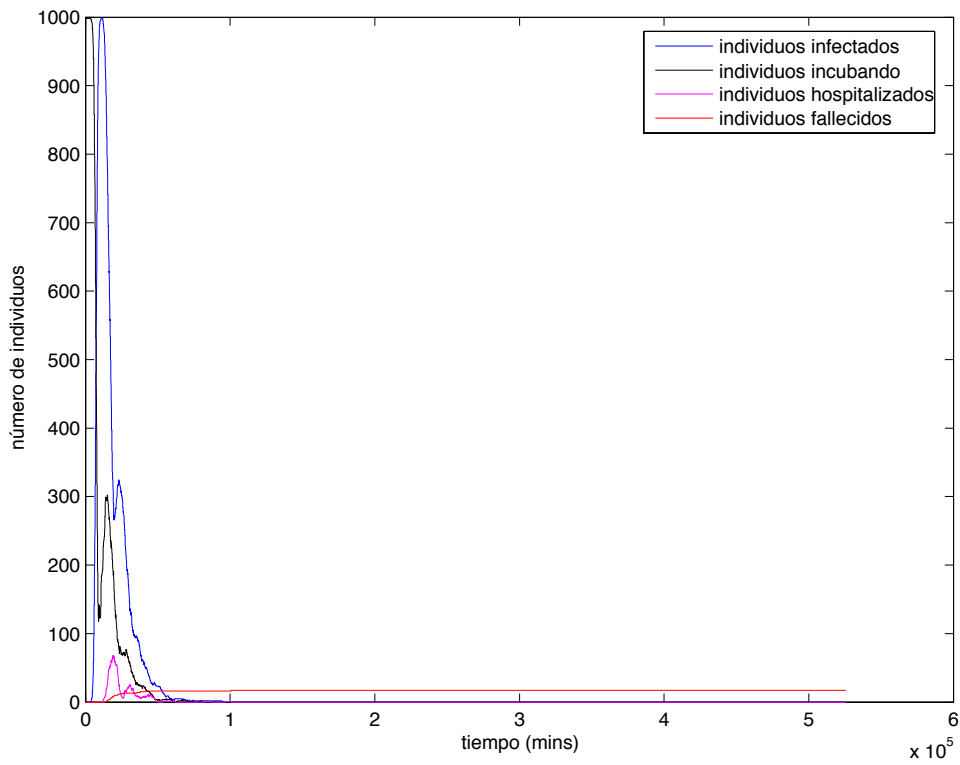


Ilustración 51. Gráfica en escala decimal de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación B.

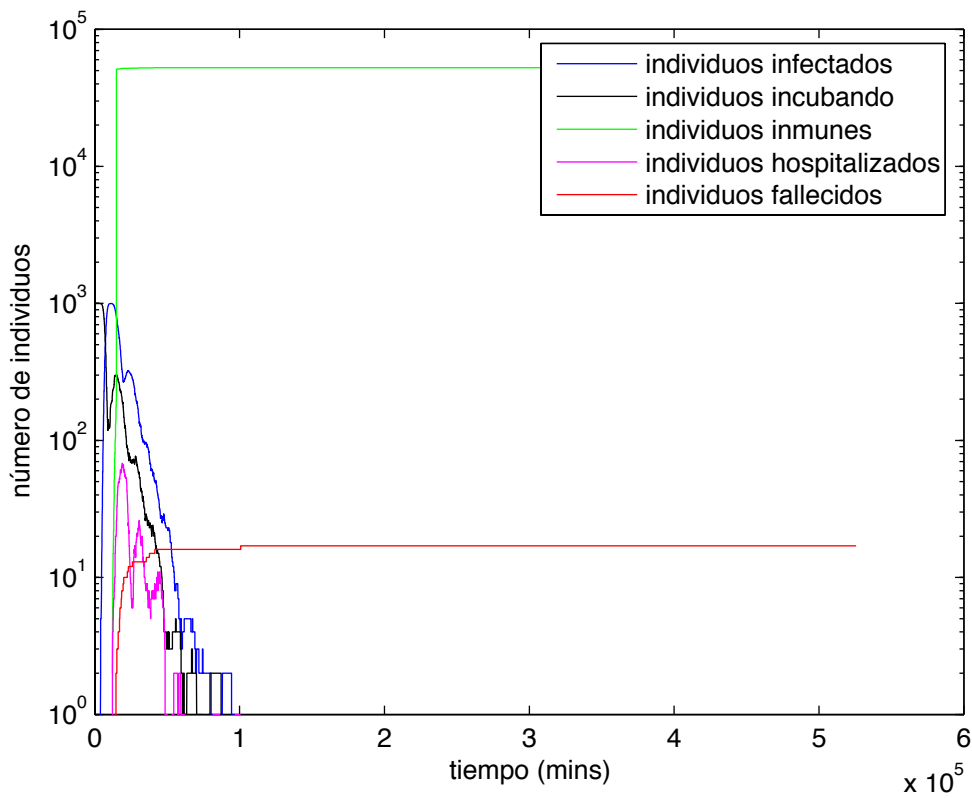


Ilustración 52. Gráfica en escala logarítmica de la evolución en el tiempo de los individuos en el resto de estados para la campaña de vacunación B.

En el caso supuesto la vacunación se realiza cuando la enfermedad se ha propagado y empieza a remitir. Lo que estos resultados muestran es que no tiene un impacto significativo en la reducción de la mortalidad.

Para realizar conclusiones más definitivas sería necesario realizar un estudio más exhaustivo sobre el efecto que tiene la vacunación de la población en la propagación de la enfermedad, especialmente estudiando la vacunación en distintos estadios de desarrollo de la misma.

Analizando el proceso de vacunación desde el punto de vista de conectividad del grafo, un individuo vacunado es incapaz de ser infectado, lo que equivale a la eliminación del nodo (y vértices asociados) del grafo. Esto ha sido evaluado en el segundo supuesto para el caso contrario: un grafo con mayor conectividad.

## Evaluación del rendimiento de la aplicación

A partir de la realización de los supuestos detallados en el apartado anterior se han obtenido una serie de datos que sirven para analizar el rendimiento de la aplicación. Estos datos son:

- Tiempo total de ejecución (medido en segundos) para cada uno de los procesos.
- Tiempo de ejecución (medido en segundos) de la rutina de propagación del virus para cada uno de los procesos.

La evaluación de los distintos supuestos se ha realizado sobre 1, 2, 4 y 8 procesos, siendo mapeados cada uno de estos sobre un equipo (un procesador) del clúster [Kasukabe] con el objetivo de obtener el mejor rendimiento posible.

Para cada una de las evaluaciones de los tres primeros supuestos se han realizado tres ejecuciones con el objetivo de recoger datos suficientes para tomar valores medios y de desviación estándar de los tiempos de ejecución.

Se realiza la medida de tiempos, cálculo de aceleración y eficiencia en cada caso y análisis de la escalabilidad de la herramienta.

## Evaluación de rendimiento para el supuesto 1

### Rendimiento con un proceso

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.694,65	-	-	-	-	-	-	-
2ª	1.690,74	-	-	-	-	-	-	-
3ª	1.696,94	-	-	-	-	-	-	-
Tiempo medio	1.694,11	-	-	-	-	-	-	-
Desviación estándar	3,13	-	-	-	-	-	-	-

Tabla 9. Tiempo total de ejecución para el supuesto 1 con un proceso.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.660,82	-	-	-	-	-	-	-
2ª	1.656,72	-	-	-	-	-	-	-
3ª	1.662,84	-	-	-	-	-	-	-
Tiempo medio	1.660,13	-	-	-	-	-	-	-
Desviación estándar	3,12	-	-	-	-	-	-	-

Tabla 10. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con un proceso.

## Rendimiento con dos procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.009,33	1.008,88	-	-	-	-	-	-
2ª	1.005,95	1.005,68	-	-	-	-	-	-
3ª	1.010,99	1.010,71	-	-	-	-	-	-
<b>Tiempo medio</b>	1.008,76	1.008,42	-	-	-	-	-	-
<b>Desviación estándar</b>	2,57	2,55	-	-	-	-	-	-

Tabla 11. Tiempo total de ejecución para el supuesto 1 con dos procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	987,62	987,62	-	-	-	-	-	-
2ª	984,57	984,57	-	-	-	-	-	-
3ª	989,14	989,13	-	-	-	-	-	-
<b>Tiempo medio</b>	987,11	987,11	-	-	-	-	-	-
<b>Desviación estándar</b>	2,33	2,32	-	-	-	-	-	-

Tabla 12. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con dos procesos.



## Rendimiento con cuatro procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	960,62	961,26	959,19	959,19	-	-	-	-
2ª	967,32	967,06	966,62	966,62	-	-	-	-
3ª	971,34	971,18	969,41	969,40	-	-	-	-
<b>Tiempo medio</b>	966,43	966,50	965,07	965,07	-	-	-	-
<b>Desviación estándar</b>	5,42	4,98	5,28	5,28	-	-	-	-

Tabla 13. Tiempo total de ejecución para el supuesto 1 con cuatro procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	940,90	940,86	940,86	940,86	-	-	-	-
2ª	946,30	946,33	946,30	946,30	-	-	-	-
3ª	951,01	951,02	950,98	950,98	-	-	-	-
<b>Tiempo medio</b>	946,07	946,07	946,05	946,05	-	-	-	-
<b>Desviación estándar</b>	5,06	5,08	5,06	5,06	-	-	-	-

Tabla 14. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con cuatro procesos.

## Rendimiento con ocho procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	736,53	736,02	734,74	736,23	734,45	736,23	736,24	735,55
2ª	743,97	743,69	743,68	743,69	743,68	743,69	743,69	743,69
3ª	738,02	737,75	737,74	737,75	737,75	737,75	736,45	737,50
<b>Tiempo medio</b>	739,51	739,15	738,72	739,22	738,63	739,22	738,79	738,91
<b>Desviación estándar</b>	3,94	4,02	4,55	3,94	4,68	3,94	4,24	4,25

Tabla 15. Tiempo total de ejecución para el supuesto 1 con ocho procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	719,02	719,02	719,02	719,02	719,05	719,06	719,03	719,02
2ª	726,16	726,15	726,16	726,16	726,18	726,17	726,15	726,15
3ª	720,37	720,37	737,74	720,37	720,41	720,41	720,37	720,37
<b>Tiempo medio</b>	721,85	721,85	727,64	721,85	721,88	721,88	721,85	721,85
<b>Desviación estándar</b>	3,79	3,79	9,45	3,79	3,79	3,78	3,78	3,79

Tabla 16. Tiempo de ejecución de la rutina de propagación para el supuesto 1 con ocho procesos.

## Evaluación de rendimiento para el supuesto 2

### Rendimiento con un proceso

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.915,09	-	-	-	-	-	-	-
2ª	1.930,32	-	-	-	-	-	-	-
3ª	1.913,03	-	-	-	-	-	-	-
<b>Tiempo medio</b>	1.919,48	-	-	-	-	-	-	-
<b>Desviación estándar</b>	9,44	-	-	-	-	-	-	-

Tabla 17. Tiempo total de ejecución para el supuesto 2 con un proceso.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.856,78	-	-	-	-	-	-	-
2ª	1.872,09	-	-	-	-	-	-	-
3ª	1.854,69	-	-	-	-	-	-	-
<b>Tiempo medio</b>	1.861,19	-	-	-	-	-	-	-
<b>Desviación estándar</b>	9,50	-	-	-	-	-	-	-

Tabla 18. Tiempo de ejecución de la rutina de propagación para el supuesto 2 con un proceso.

## Rendimiento con dos procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.123,35	1.121,34	-	-	-	-	-	-
2ª	1.133,41	1.133,12	-	-	-	-	-	-
3ª	1.117,58	1.117,29	-	-	-	-	-	-
<b>Tiempo medio</b>	1.124,78	1.123,92	-	-	-	-	-	-
<b>Desviación estándar</b>	8,01	8,22	-	-	-	-	-	-

Tabla 19. Tiempo total de ejecución para el supuesto 2 con dos procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.088,22	1.088,19	-	-	-	-	-	-
2ª	1.099,23	1.099,22	-	-	-	-	-	-
3ª	1.083,24	1.083,24	-	-	-	-	-	-
<b>Tiempo medio</b>	1.090,23	1.090,22	-	-	-	-	-	-
<b>Desviación estándar</b>	8,18	8,18	-	-	-	-	-	-

Tabla 20. Tiempo de ejecución de la rutina de propagación para el supuesto 2 con dos procesos.

## Rendimiento con cuatro procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.090,53	1.090,36	1.088,43	1.088,49	-	-	-	-
2ª	1.062,16	1.061,90	1.060,01	1.061,42	-	-	-	-
3ª	1.082,39	1.082,31	1.081,26	1.080,40	-	-	-	-
<b>Tiempo medio</b>	1.078,36	1.078,19	1.076,57	1.076,77	-	-	-	-
<b>Desviación estándar</b>	14,61	14,67	14,78	13,90	-	-	-	-

Tabla 21. Tiempo total de ejecución para el supuesto 2 con cuatro procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.057,66	1.057,68	1.057,66	1.057,66	-	-	-	-
2ª	1.029,20	1.029,22	1.029,18	1.029,18	-	-	-	-
3ª	1.049,83	1.049,80	1.049,77	1.049,77	-	-	-	-
<b>Tiempo medio</b>	1.045,56	1.045,57	1.045,54	1.045,54	-	-	-	-
<b>Desviación estándar</b>	14,70	14,69	14,70	14,70	-	-	-	-

Tabla 22. Tiempo de ejecución de la rutina de propagación para el supuesto 2 con cuatro procesos.

## Rendimiento con ocho procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	793,95	793,40	793,67	793,34	793,64	793,68	792,07	793,56
2ª	792,02	791,75	791,74	791,74	790,47	791,74	791,74	791,74
3ª	790,21	789,92	789,91	788,01	789,92	789,93	789,92	789,92
<b>Tiempo medio</b>	792,06	791,69	791,77	791,03	791,34	791,78	791,24	791,74
<b>Desviación estándar</b>	1,87	1,74	1,88	2,74	2,01	1,88	1,16	1,82

Tabla 23. Tiempo total de ejecución para el supuesto 2 con ocho procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	766,48	766,48	766,52	766,48	766,48	766,53	766,49	766,53
2ª	764,65	764,64	764,64	764,64	764,67	764,71	764,64	764,64
3ª	762,33	762,21	762,20	762,21	762,20	762,27	762,20	762,21
<b>Tiempo medio</b>	764,49	764,44	764,45	764,44	764,45	764,50	764,44	764,46
<b>Desviación estándar</b>	2,08	2,14	2,17	2,14	2,15	2,14	2,15	2,17

Tabla 24. Tiempo de ejecución de la rutina de propagación para el supuesto 2 con ocho procesos.

## Evaluación de rendimiento para el supuesto 3

### Rendimiento con un proceso

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.964,38	-	-	-	-	-	-	-
2ª	2.052,97	-	-	-	-	-	-	-
3ª	2.019,87	-	-	-	-	-	-	-
<b>Tiempo medio</b>	2.012,41	-	-	-	-	-	-	-
<b>Desviación estándar</b>	44,76	-	-	-	-	-	-	-

Tabla 25. Tiempo total de ejecución para el supuesto 3 con un proceso.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.930,16	-	-	-	-	-	-	-
2ª	2.018,98	-	-	-	-	-	-	-
3ª	1.985,56	-	-	-	-	-	-	-
<b>Tiempo medio</b>	1.978,23	-	-	-	-	-	-	-
<b>Desviación estándar</b>	44,86	-	-	-	-	-	-	-

Tabla 26. Tiempo de ejecución de la rutina de propagación para el supuesto 3 con un proceso.

## Rendimiento con dos procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1 <sup>a</sup>	1.334,46	1.334,18	-	-	-	-	-	-
2 <sup>a</sup>	1.363,77	1.363,46	-	-	-	-	-	-
3 <sup>a</sup>	1.311,93	1.311,64	-	-	-	-	-	-
<b>Tiempo medio</b>	1.336,72	1.336,43	-	-	-	-	-	-
<b>Desviación estándar</b>	25,99	25,98	-	-	-	-	-	-

Tabla 27. Tiempo total de ejecución para el supuesto 3 con dos procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1 <sup>a</sup>	1.312,86	1.312,84	-	-	-	-	-	-
2 <sup>a</sup>	1.341,78	1.341,75	-	-	-	-	-	-
3 <sup>a</sup>	1.290,07	1.209,07	-	-	-	-	-	-
<b>Tiempo medio</b>	1.314,90	1.287,89	-	-	-	-	-	-
<b>Desviación estándar</b>	25,92	69,77	-	-	-	-	-	-

Tabla 28. Tiempo de ejecución de la rutina de propagación para el supuesto 3 con dos procesos.



## Rendimiento con cuatro procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.290,77	1.290,52	1.289,28	1.289,28	-	-	-	-
2ª	1.206,80	1.206,55	1.204,66	1.204,66	-	-	-	-
3ª	1.243,26	1.243,00	1.241,19	1.241,20	-	-	-	-
<b>Tiempo medio</b>	1.246,94	1.246,69	1.245,04	1.245,05	-	-	-	-
<b>Desviación estándar</b>	42,11	42,11	42,44	42,44	-	-	-	-

Tabla 29. Tiempo total de ejecución para el supuesto 3 con cuatro procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	1.269,63	1.269,63	1.269,61	1.269,61	-	-	-	-
2ª	1.186,16	1.186,17	1.186,13	1.186,13	-	-	-	-
3ª	1.222,39	1.222,38	1.222,35	1.222,35	-	-	-	-
<b>Tiempo medio</b>	1.226,06	1.226,06	1.226,03	1.226,03	-	-	-	-
<b>Desviación estándar</b>	41,86	41,85	41,86	41,86	-	-	-	-

Tabla 30. Tiempo de ejecución de la rutina de propagación para el supuesto 3 con cuatro procesos.

### Rendimiento con ocho procesos

- Tiempo total de ejecución (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	823,06	824,26	824,77	823,05	824,44	824,78	824,78	823,42
2ª	861,32	861,04	861,02	861,03	860,98	861,03	861,03	861,03
3ª	794,79	794,50	794,50	794,50	794,50	794,50	794,50	794,50
<b>Tiempo medio</b>	826,39	826,60	826,76	826,19	826,64	826,77	826,77	826,32
<b>Desviación estándar</b>	33,39	33,33	33,30	33,38	33,29	33,31	33,31	33,36

Tabla 31. Tiempo total de ejecución para el supuesto 3 con ocho procesos.

- Tiempo de ejecución de la rutina de propagación (medido en segundos)

Número de ejecución	Número de proceso							
	0	1	2	3	4	5	6	7
1ª	807,48	807,46	807,54	807,49	807,46	807,52	807,46	807,46
2ª	843,80	843,79	843,79	843,80	843,80	843,87	843,81	843,85
3ª	777,20	777,17	777,18	777,17	777,19	777,23	777,19	777,19
<b>Tiempo medio</b>	809,49	809,47	809,50	809,49	809,48	809,54	809,49	809,50
<b>Desviación estándar</b>	33,35	33,36	33,35	33,36	33,35	33,36	33,36	33,38

Tabla 32. Tiempo de ejecución de la rutina de propagación para el supuesto 3 con ocho procesos.

## Análisis de speedup, eficiencia y escalabilidad

### Speedup y eficiencia para el supuesto 1

- Análisis de tiempo total de ejecución

Número de procesos	Tiempo secuencial	Tiempo paralelo	Speedup	Eficiencia
1	1.694,11	1.694,11	1,00	1,00
2	1.694,11	1.008,76	1,68	0,84
4	1.694,11	966,50	1,75	0,44
8	1.694,11	739,51	2,29	0,29

Tabla 33. Tabla de análisis de speedup y eficiencia del supuesto 1.

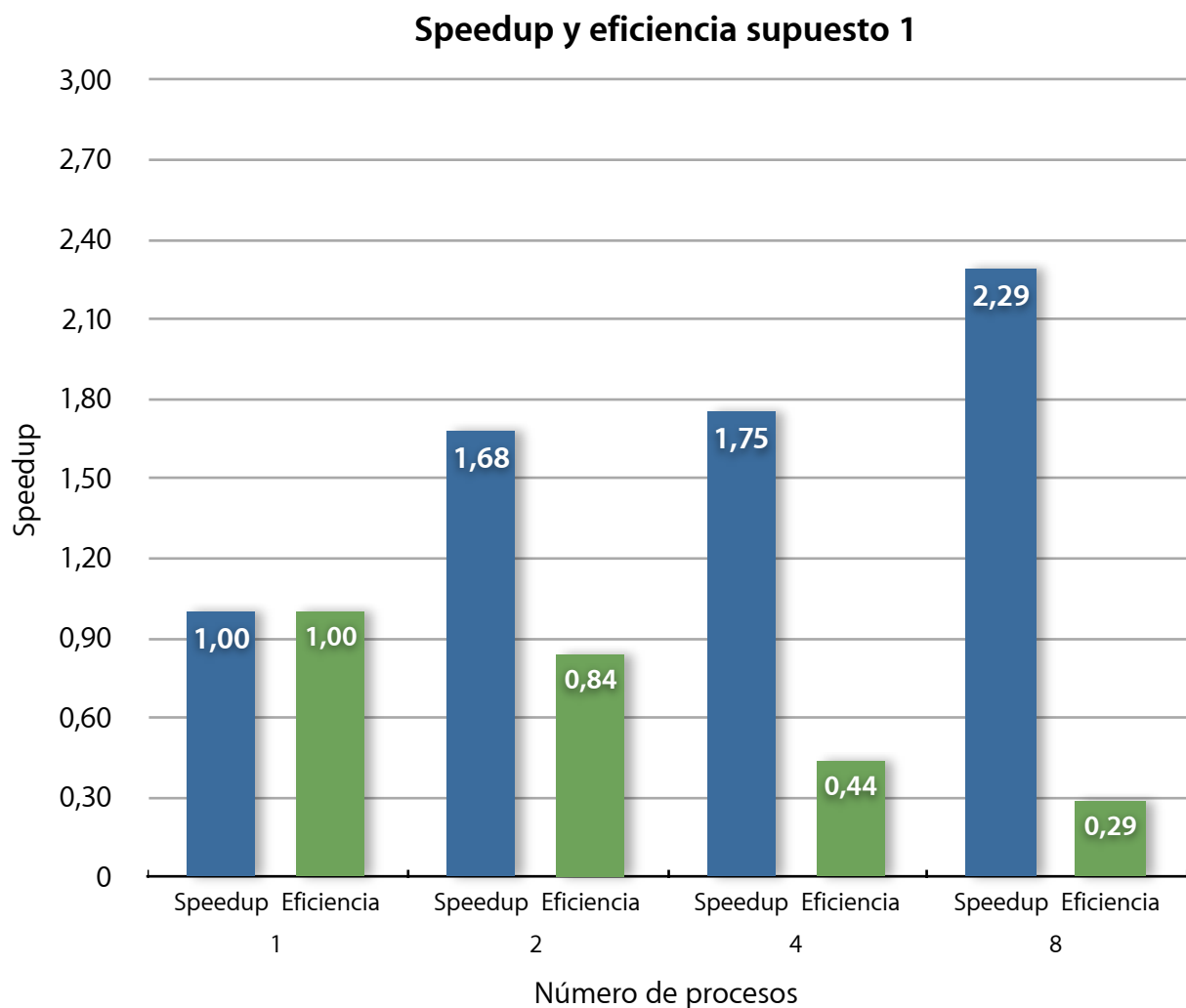


Ilustración 53. Análisis de speedup para el supuesto 1.

- Análisis de tiempo de ejecución de la rutina de propagación

Número de procesos	Tiempo secuencial	Tiempo paralelo
1	1.660,13	1.660,13
2	1.660,13	987,11
4	1.660,13	946,07
8	1.660,13	727,64

Tabla 34. Tabla de análisis de speedup y eficiencia de la rutina de propagación del supuesto 1.

Analizando las medidas y gráficas de tiempos, aceleración y eficiencia se observa que para la ejecución del supuesto 1 la aplicación no se comporta de una forma escalable. Sí se observa una mejora en la aceleración de la ejecución según se incrementa el número de procesos, pero en cambio la eficiencia disminuye en la misma medida.

## Speedup y eficiencia para el supuesto 2

- Análisis de tiempo total de ejecución

Número de procesos	Tiempo secuencial	Tiempo paralelo	Speedup	Eficiencia
1	1.919,48	1.919,48	1,00	1,00
2	1.919,48	1.124,78	1,71	0,85
4	1.919,48	1.078,36	1,78	0,44
8	1.919,48	792,06	2,42	0,30

Tabla 35. Tabla de análisis de speedup y eficiencia del supuesto 2.

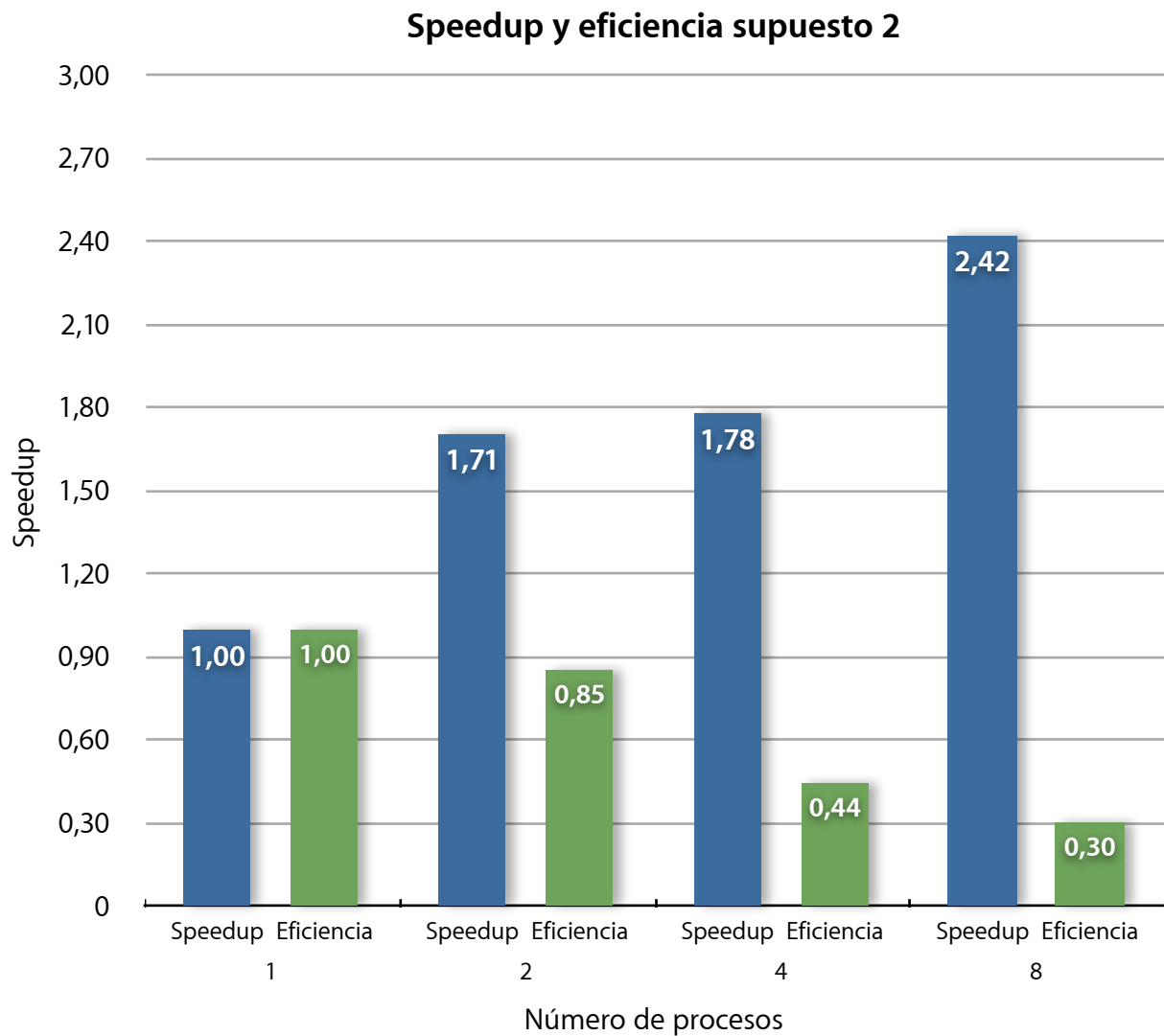


Ilustración 54. Análisis de speedup para el supuesto 2.

- Análisis de tiempo de ejecución de la rutina de propagación

Número de procesos	Tiempo secuencial	Tiempo paralelo
1	1.861,19	1.861,19
2	1.861,19	1.090,23
4	1.861,19	1.045,57
8	1.861,19	764,50

Tabla 36. Tabla de análisis de speedup y eficiencia de la rutina de propagación del supuesto 2.

Para el caso del supuesto 2 la aplicación, al igual que para el primer supuesto, no se comporta de una forma escalable. Sí se observa una mejora en la aceleración de la ejecución según se incrementa el número de procesos, pero igualmente la eficiencia disminuye en la misma medida.

### Speedup y eficiencia para el supuesto 3

- Análisis de tiempo total de ejecución

Número de procesos	Tiempo secuencial	Tiempo paralelo	Speedup	Eficiencia
1	2.012,41	2.012,41	1,00	1,00
2	2.012,41	1.336,72	1,51	0,75
4	2.012,41	1.246,94	1,61	0,40
8	2.012,41	826,77	2,43	0,30

Tabla 37. Tabla de análisis de speedup y eficiencia del supuesto 3.

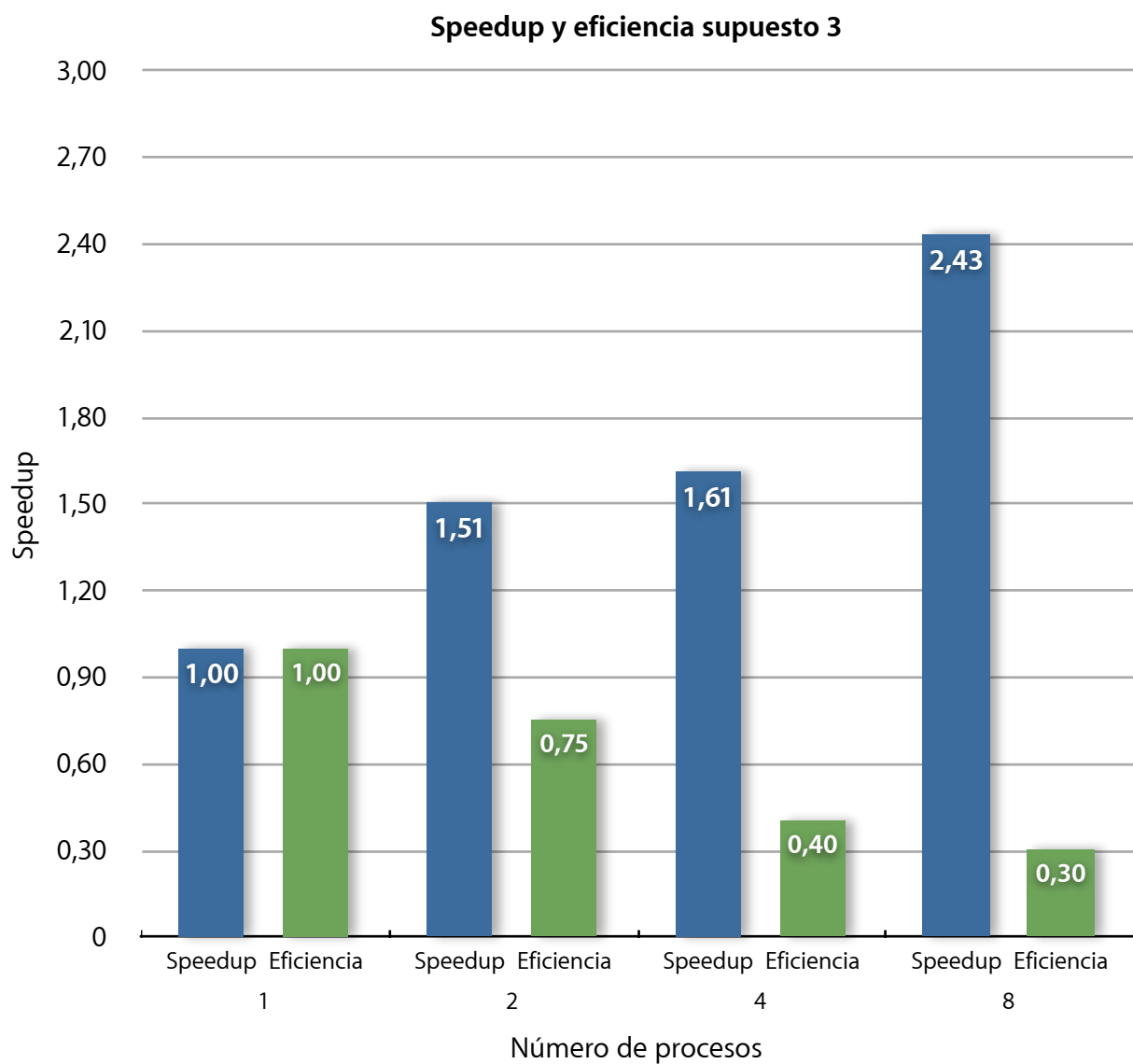


Ilustración 55. Análisis de speedup para el supuesto 3.

- Análisis de tiempo de ejecución de la rutina de propagación

Número de procesos	Tiempo secuencial	Tiempo paralelo
1	1.978,23	1.978,23
2	1.978,23	1.314,90
4	1.978,23	1.226,06
8	1.978,23	849,54

Tabla 38. Tabla de análisis de speedup y eficiencia de la rutina de propagación del supuesto 3.

Como en los supuestos anteriores la aplicación tampoco comporta de una forma escalable para la ejecución del supuesto 3. También se observa una mejora en la aceleración de la ejecución según se incrementa el número de procesos, pero del mismo modo la eficiencia disminuye en la misma medida.



# Conclusiones y trabajo futuro

*En este capítulo se realiza una comparativa entre los resultados obtenidos frente a los objetivos planteados de forma inicial de la que se extraen las conclusiones sobre el proyecto así como la perspectiva de trabajo futuro.*

## Conclusiones

Las conclusiones que se muestran a continuación surgen como consecuencia de los objetivos planteados al inicio del desarrollo del proyecto, mostrando el grado de éxito alcanzado por el mismo. Posteriormente se muestran una serie de reflexiones a modo de experiencia personal sobre los objetivos, desarrollo y conclusiones del proyecto.

- Se ha conseguido modelar la propagación de la gripe A (H1N1) en un entorno de población urbana, lo que supone que se ha alcanzado el objetivo principal planteado.
- Se ha conseguido modelar matemáticamente el colectivo social de una población, dando forma a individuos, colectivos y conexiones entre ellos.
- Se ha conseguido modelar el comportamiento de una enfermedad contagiosa y simular su propagación en un entorno urbano.
- Se ha conseguido desarrollar una herramienta distribuida que permita simular la propagación de una enfermedad contagiosa en entornos urbanos.
- Se ha conseguido evaluar la simulación de la propagación en la herramienta distribuida del virus de la gripe A (H1N1) sobre la población de Leganés.
- La herramienta construida es flexible, ya que se adapta a distintos modelos de sociedad y a diferentes enfermedades. En el caso de estudio se ha planteado la simulación de la propagación del virus de la gripe A (H1N1) en el municipio madrileño de Leganés, aunque existe la posibilidad de modelar diferentes virus y poblaciones.
- Tras los datos obtenidos en la evaluación de rendimiento se ha valorado que la herramienta construida no se comporta de forma escalable, al menos para el conjunto de pruebas evaluado. Se han obtenido mejoras en rendimiento de aceleración, pero no en cuanto a eficiencia.
- Con el objetivo de construir una herramienta de la que poder obtener datos reales se han modelado los principales parámetros que influyen en la propagación de un virus en una población, y en particular la gripe A (H1N1), apoyándonos en documentos oficiales y médicos. Así mismo se ha recurrido a datos reales sobre la población de Leganés extraídos del [Ayto de Leganés, 2010] para modelar los individuos junto con los datos sobre interacciones personales reales que se han extraído de [Enron Corporation] y [Facebook].

- 
- La validación final del simulador (contrastando los valores obtenidos con los reales) forma parte del trabajo futuro, quedando fuera del alcance de este proyecto. Sin embargo, los resultados obtenidos muestran un comportamiento similar al de la enfermedad: un periodo inicial de fuerte expansión y contagio, seguido de un periodo prolongado de transmisión en el que el número de infectados disminuye paulatinamente hasta desaparecer.

## Trabajo futuro

En esta sección se enumeran diferentes propuestas como trabajo futuro para continuar el desarrollo del proyecto:

- Implementación de la utilización de grafos dinámicos que evolucionen con el tiempo, de forma que se generen y destruyan conexiones entre los individuos de la población en tiempo real.
- Dependencia de contagio por el virus en función de la edad del individuo. Se propone la implementación de una función que determine una nueva probabilidad en función de la edad del individuo transmisor y receptor del virus.
- Validación del modelo con casos reales. Para realizar esto es necesario acceder a estadísticas detalladas de infecciones (datos que actualmente no disponemos). Empleando esta información sería posible validar el modelo propuesto así como afinar los parámetros empleados para dotarlo de una mayor precisión.
- Optimización del tiempo de ejecución de la herramienta con el objetivo de obtener un comportamiento escalable en la misma.

# Presupuesto

*En este apartado se realiza un análisis detallado del coste presupuestado para el presente proyecto. En primer lugar se realiza un análisis del cálculo del coste desglosado según las diferentes categorías y a continuación se presenta el presupuesto definitivo alcanzado.*

## Cálculo de costes

### Descripción del proyecto

**Autor:** Gonzalo Martín Cruz.

**Departamento:** Arquitectura de Computadores, Comunicaciones y Sistemas.

**Título:** Simulación de la transmisión de una enfermedad contagiosa en entornos urbanos.

**Duración:** La fecha de inicio del proyecto se establece el **7 de Julio de 2009** y la fecha de finalización prevista se emplaza al día **6 de Julio de 2010**. El tiempo total de realización del proyecto que se sitúa entre estas dos fechas tiene una duración de **450 horas**.

### Costes de personal

Durante el período de realización ha existido una única persona implicada, el autor del mismo, que ha desarrollado las diversas funciones necesarias. A continuación se ilustra el desglose del coste de personal considerando en el coste el gasto por IRPF y Seguridad Social:

Fase	Coste/Hora	Total horas	Total coste
Análisis	€ 63,00	20	€ 1.260,00
Diseño	€ 63,00	30	€ 1.890,00
Codificación	€ 50,00	250	€ 12.500,00
Pruebas	€ 25,00	100	€ 2.500,00
Documentación	€ 50,00	50	€ 2.500,00
<b>TOTAL</b>		<b>450</b>	<b>€ 20.650,00</b>

Tabla 39. Costes de personal.

## Costes de equipos y tecnología

Para la realización del proyecto ha sido necesario un equipo durante la duración total, además de 16 equipos necesarios para la realización de pruebas y evaluación de la aplicación durante 3 meses. En base a lo descrito en [RD1777, 2004] se establece un plazo de amortización de 8 años para los elementos de tratamiento de información y además se incluye el coste de la conexión ADSL mensual:

Concepto	Unidades	Total coste
Equipo principal	1	€ 137,38
Equipos pruebas	8	€ 249,50
ADSL	11	€ 329,45
<b>TOTAL</b>		<b>€ 716,33</b>

## Costes de software

Tabla 40. Costes de equipos y tecnología.

El coste del Sistema Operativo del equipo principal de realización del proyecto está incluido en la anterior tabla de costes de equipos y tecnologías, ya que se incluyó su venta de forma conjunta. El coste del Sistema Operativo de los equipos de pruebas es una distribución *Linux* gratuita y se incluye el coste de la suite ofimática *iWork 09*. El coste del resto de software se incluye a continuación, estimados en base al [RD1777, 2004] en el que se establece un plazo de amortización de 6 años para los programas informáticos:

Concepto	Unidades	Total coste
S.O. equipos de pruebas	16	€ 0,00
iWork 09	1	€ 1,65
<b>TOTAL</b>		<b>€ 1,65</b>

Tabla 41. Costes de software.

## Costes de material fungible

Se asignan los siguientes costes de material fungible para papel y otros gastos no contemplados en los conceptos anteriores:

Concepto	Total coste
Papel	€ 20,00
Otros gastos	€ 75,00
<b>TOTAL</b>	<b>€ 95,00</b>

Tabla 42. Costes de material fungible.

## Presupuesto

Se establece un 7% sobre el total dedicado al riesgo del proyecto y un 15% de beneficio:

Concepto	Total coste
Costes de personal	€ 20.650,00
Costes de equipos y tecnología	€ 716,33
Costes de software	€ 1,65
Costes de material fungible	€ 95,00
Riesgo (7%)	€ 1.502,41
Beneficio (15%)	€ 3.219,45
<b>TOTAL</b>	<b>€ 26.184,84</b>

Tabla 43. Presupuesto total.

Por lo tanto el coste total presupuesto para la realización del proyecto asciende a **VEINTISÉIS MIL CIENTO OCHENTA Y CUATRO CON OCHENTA Y CUATRO CÉNTIMOS sin contabilizar el Impuesto sobre Valor Añadido (I.V.A.):**

**26.184,84 € (SIN I.V.A.)**

# Anexo I: Manual de usuario

## Contenido

El manual de usuario contiene la información necesaria para llevar a cabo la instalación, configuración, ejecución y, opcionalmente, compilación de la herramienta de simulación distribuida desarrollada. El último apartado corresponde a una rápida guía de resolución de problemas relacionados con la ejecución.

## Requisitos hardware y software

Los requisitos hardware y software se presentan tanto para ejecución como para compilación, ya que se ofrece al usuario la posibilidad de generar el ejecutable para distintas plataformas llevando a cabo la compilación de los archivos fuente.

### Requisitos hardware de compilación

Los requisitos hardware mínimos descritos para llevar a cabo la compilación de la herramienta son:

- Procesador: Intel Pentium IV a 3,0Ghz. o similar.
- Memoria RAM: 1GB.
- Espacio libre en disco duro: 30MB.

### Requisitos software de compilación

Los requisitos software necesarios para llevar a cabo la compilación de la herramienta son:

- Sistema Operativo: Linux. La herramienta está construida en SO Linux y lenguaje C, por lo que es posible llevar a cabo la compilación en otras plataformas siempre que se disponga de las librerías necesarias para tal fin.
- Librería GSL: [GSL] es la GNU Scientific Library para C y C++.
- Librería XML: [LibXml].
- Librería MPI: [MPICH2].
- Librería CSparse: [Davis 2006].

## Requisitos hardware de ejecución

Los requisitos hardware mínimos descritos para llevar a cabo la ejecución de la herramienta de forma distribuida son los siguientes:

- Procesador: Intel Pentium IV a 3,0Ghz. o similar.
- Memoria RAM: 1GB.
- Espacio libre en disco duro: se recomienda un mínimo de 2GB variable en función del tamaño de la población y del tiempo de la simulación.

## Requisitos software de ejecución

Los requisitos software necesarios para llevar a cabo la compilación de la herramienta son:

- Sistema Operativo: el mismo en el que se haya llevado a cabo la compilación de los archivos fuente.
- Sistema gestor de colas de trabajo del clúster: PBS.

## Compilación e instalación de la aplicación

La instalación de la aplicación se realiza copiando la estructura y archivos contenidos en el CD sobre la ruta deseada en el disco duro o, si se desea realizar una ejecución distribuida, en el espacio de usuario del clúster. Posteriormente será necesario generar los diferentes ficheros sobre la ruta especificada:

- en `/data` se debe situar `config.xml`, fichero de configuración general que contiene los parámetros de ejecución en formato XML.
- en `/data` se debe situar `config_grafos.xml`, fichero de configuración que contiene los parámetros de generación de los colectivos de la población en formato XML.
- en `/data/graphs/social` se deben situar los grafos origen necesarios para la ejecución.
- en `/data/graphs/pop` se debe situar `grafos.dat`, fichero que contiene la definición en colectivos de la población de la simulación. Para generar dicho fichero se incluye la herramienta `gengrafo`, a la que proporcionado un número de procesos y un tamaño de población como entrada genera el fichero `grafos.dat` correspondiente.
- en `/data/graphs/acq` se deben situar los grafos de configuración de los grados de conexión entre colectivos necesarios para la ejecución.
- en `/include` se debe situar los archivos correspondientes a la librería `Csparse` [Davis 2006].

La compilación de la aplicación se realizará ejecutando el archivo `make.sh` provisto para su ejecución en consola de comandos.



## Guía rápida sobre el uso y funcionamiento de la aplicación

Para realizar una ejecución distribuida sobre el clúster PBS únicamente se deberá crear un script para lanzar el trabajo con el número de nodos y procesos deseado. Un aspecto a tener en cuenta es que el número de procesos lanzados debe coincidir con el número de procesos de la construcción del fichero `grafos.dat` con la herramienta `gengrafo`.

Una vez finalizada la ejecución se generarán una serie de ficheros de salida:

- `age.dat`: fichero que contiene la edad de cada uno de los individuos de la población que interviene en la simulación.
- `condition.dat`: fichero que contiene la condición final de cada uno de los individuos de la población que interviene en la simulación.
- `occupation.dat`: fichero que contiene la ocupación de cada uno de los individuos de la población que interviene en la simulación.
- `v_sanos.dat`: fichero que contiene el número de individuos sanos en cada instante en saltos de tiempo especificados en el parámetro `test` del fichero de configuración `config.xml`.
- `v_inc.dat`: fichero que contiene el número de individuos incubando en cada instante en saltos de tiempo especificados en el parámetro `test` del fichero de configuración `config.xml`.
- `v_inf.dat`: fichero que contiene el número de individuos infectados en cada instante en saltos de tiempo especificados en el parámetro `test` del fichero de configuración `config.xml`.
- `v_inm.dat`: fichero que contiene el número de individuos inmunes en cada instante en saltos de tiempo especificados en el parámetro `test` del fichero de configuración `config.xml`.
- `v_hos.dat`: fichero que contiene el número de individuos hospitalizados en cada instante en saltos de tiempo especificados en el parámetro `test` del fichero de configuración `config.xml`.
- `v_fall.dat`: fichero que contiene el número de individuos fallecidos en cada instante en saltos de tiempo especificados en el parámetro `test` del fichero de configuración `config.xml`.
- `v_time.dat`: fichero que contiene el número el instante de tiempo en minutos en saltos de tiempo especificados en el parámetro `test` del fichero `config.xml`.
- `condition_N.dat`: fichero que contiene la condición de cada uno de los individuos de la población en el instante de tiempo `N`, el cuál es el valor correspondiente al parámetro `tfich` del fichero de configuración `config.xml`.
- `adjac_N.dat`: fichero que contiene la matriz de adyacencia del proceso `N` en formato `CSC`.

Habrás notado que al tratarse de una herramienta distribuida cada uno de los procesos generará su propio fichero correspondiente a su porción de matriz de adyacencia, por lo que al finalizar la ejecución será necesario compactar las matrices de todos los procesos en ejecución para obtener el grafo completo. Para realizar dicha tarea se provee la herramienta `matrix_sum`, a la que únicamente se ha de proporcionar como entrada el número de procesos en ejecución y al finalizar generará tres archivos correspondientes al grafo en formato de representación coordinado:

- `scoldata.dat`: fichero que contiene los índices de columna correspondientes a las entradas de la matriz de adyacencia.
- `srowdata.dat`: fichero que contiene los índices de fila correspondientes a las entradas de la matriz de adyacencia.
- `svaldata.dat`: fichero que contiene los valores de las entradas de la matriz de adyacencia.

## Resolución de problemas

Los principales problemas que pueden surgir durante la ejecución de la aplicación están relacionados con la gestión de memoria de los procesos. A continuación se describen los problemas más comunes y su modo de solucionarlos:

- Aparece el error: "epidemiology\_mpi: epidemiology.c:685: GenerarGrupos: Assertion ``*offset < limite_offset'` failed".

Este error se produce por falta de memoria para generar el grafo, como consecuencia de una valoración por defecto del número medio de conocidos por persona. Para solucionarlo se debe incrementar el valor del parámetro `avgn` del fichero de configuración "config.xml".

- Aparece el error: "epidemiology\_mpi: epidemiology\_mpi.c:170: main: Assertion ``(array_val = (double*) malloc (limite_offset * sizeof (double))) != ((void *)0)'` failed".

Este error se produce por falta de memoria del sistema, como consecuencia de una valoración por exceso del número medio de conocidos por persona. Para solucionarlo se debe disminuir el valor del parámetro `avgn` del fichero de configuración "config.xml".

El resto de problemas más comunes están relacionados con la gestión del clúster y del gestor de trabajos, que quedan ambos fuera del ámbito de la aplicación.

# Referencias

## [ARCOS, 2009]

Grupo de investigación de Arquitectura de Computadores, Comunicaciones y Sistemas (ARCOS): *Material docente de la asignatura Arquitectura de Computadores II*. 2009. Recurso electrónico disponible en:

<http://www.arcos.inf.uc3m.es>

## [Argonne]

Argonne National Laboratory, Mathematics and Computer Science Division. Recurso electrónico disponible en:

<http://www.mcs.anl.gov/index.php>

## [Ayto de Leganés, 2010]

Ayuntamiento de Leganés: *"Estadísticas de la población de Leganés"*. 2010. Recurso electrónico disponible en:

<http://www.leganes.org/leganes/Estadisticas/Inicio.html>

## [CDC, 2009]

Centers for Disease Control and Prevention (CDC): *"Seasonal Influenza"*. 2009. Recurso electrónico disponible en:

<http://www.cdc.gov/flu/weekly/>

## [Culler et al, 1999]

David E. Culler, Jaswinder Pal Singh con Anoop Gupta: *"Parallel Computer Architecture. A hardware/software approach"*. Morgan Kaufmann Publishers, Inc. 1999.

## [Davis, 2006]

Timothy A. Davis: *"A Concise Sparse matrix package"*. 2006. Recurso electrónico disponible en:

[http://people.sc.fsu.edu/~burkardt/c\\_src/csparse/csparse.html](http://people.sc.fsu.edu/~burkardt/c_src/csparse/csparse.html)

### **[Enron Corporation]**

Enron Corporation. Recurso electrónico disponible en:

<http://www.enron.com>

### **[Facebook]**

Facebook. Recurso electrónico disponible en:

<http://es-es.facebook.com/>

### **[GoogleTrends]**

Google Inc.: Google Trends, herramienta de estadísticas sobre el historial de búsquedas en el sistema Google. Recurso electrónico disponible en:

<http://www.google.es/trends>

### **[GSL]**

GSL: GNU Scientific Library, a numerical library for C and C++. Recurso electrónico disponible en:

<http://www.gnu.org/software/gsl/>

### **[INE, 2001-2009]**

Instituto Nacional de Estadística (INE): "*Clasificación de la población de 16 y más años según su relación con la actividad económica*". 2001-2009. Recurso electrónico disponible en:

<http://www.ine.es>

### **[Kasukabe]**

Clúster PBS de la Universidad Carlos III de Madrid. Recurso electrónico disponible en:

<html://kasukabe.lab.inf.uc3m.es>

### **[Leyland]**

Tony Leyland, B.Ed. M.Sc., Department of Biomedical Physiology and Kinesiology at Simon Fraser University, Canadá: "*Infectious Diseases. Chapter 13 Immunity and Infection*". Recurso electrónico disponible en:

<http://www.sfu.ca/~leyland>

### **[LibXML]**

XML C parser and toolkit. 1999. Recurso electrónico disponible en:

<http://xmlsoft.org/>

### [Ministerio de Sanidad y Política Social del Gobierno de España - Gripe A]

Ministerio de Sanidad y Política Social del Gobierno de España: "Gripe A (H1N1)". Recurso electrónico disponible en:

<http://www.informaciongripe.es/>

### [MPIForum]

Official MPI standards documents, errata, and archives of the MPI Forum. Recurso electrónico disponible en:

<http://www.mpi-forum.org/>

### [MPICH2]

MPICH2, a high-performance and widely portable implementation of Message Passing Interface (MPI) standard. Recurso electrónico disponible en:

<http://www.mcs.anl.gov/research/projects/mpich2/index.php>

### [RD1777, 2004]

Ministerio de Economía y Hacienda de España: "Real Decreto 1777/2004, de 30 de Julio, por el que se aprueba el Reglamento del Impuesto de Sociedades. Anexo: Tablas de Coeficientes de Amortización". 2004. Recurso electrónico disponible en:

[http://www.boe.es/g/es/bases\\_datos/doc.php](http://www.boe.es/g/es/bases_datos/doc.php)

### [Saad, 2003]

Yousef Saad: "Iterative methods for sparse linear systems". Society for Industrial and Applied Mathematics. 2003.

### [WSDH]

Washington State Department of Health: "Public health surveillance. Notifiable Conditions". Recurso electrónico disponible en:

<http://www.doh.wa.gov/>