

REGISTRO DE TRABAJO DE GRADO

Fecha	31	05	2021
-------	----	----	------

DATOS DEL ESTUDIANTE

NOMBRES: JORGE ANDRES						APELLIDOS: IBÁÑEZ HUERTAS	
TIPO DE IDENTIFICACIÓN:	T.I.		C.C.	X	C.E.	NÚMERO: 1001045370	
CORREO INSTITUCIONAL: jibanezh@ucentral.edu.co						TELÉFONO: 3045560527	

AVAL DEL DOCENTE DIRECTOR

NOMBRES: CARLOS ISAAC ZAINEA					DEPARTAMENTO: MATEMÁTICAS
CORREO INSTITUCIONAL: czaineam@ucentral.edu.co					TELÉFONO-EXT: 3239868 - 3457

COMPONENTES

1. TÍTULO DEL TRABAJO DE GRADO O PASANTIA:

Diseño de un algoritmo en redes neuronales para predecir el comportamiento de una enfermedad, en entornos simulados con autómatas celulares

2. INTRODUCCIÓN Y JUSTIFICACIÓN (máximo 1500 palabras):

Las matemáticas han sido pieza fundamental en el estudio epidemiológico, el cual busca predecir el comportamiento de una enfermedad para mitigar su impacto, tomando medidas de control y prevención. Un punto de partida es el modelo descrito por Kermack y McKendrick en 1927, también conocido como modelo SIR, en el que se establecen relaciones entre tres estados de una enfermedad (Susceptible-Infectado-Recuperado) y se implementan los conceptos de tasa de contagio y de recuperación [11]. Desde entonces se han desarrollado múltiples variaciones sobre el modelo SIR, con el objetivo de analizar diferentes tipos de enfermedades de una manera más precisa, considerando por ejemplo diferentes estados, tasas de natalidad y mortalidad, entre otros [10].

Por otra parte y debido a los avances tecnológicos de las últimas décadas, se han desarrollado modelos y simulaciones que permiten analizar características que no eran posibles con los modelos anteriormente mencionados. Por ejemplo, patrones de movilidad [1, 4], disminución en la cantidad de contagios debido a un aislamiento preventivo [6], contagios de individuo a individuo [12] e interacciones en masa [2, 9], la mayoría realizadas con una fuerte influencia de las redes neuronales y complejas, apoyadas fuertemente sobre la teoría de grafos.

La inexistencia de un algoritmo predictivo que tenga en cuenta las interacciones más cercadas de cada individuo, es un problema que se ha evidenciado principalmente, a causa de la naturaleza de las bases de datos disponibles. A pesar del fuerte acercamiento de los modelos epidemiológicos a las dinámicas sociales con la llegada del coronavirus, se consideran interacciones a escalas que podrían ser mayores a las que se contagia la enfermedad, si bien existen modelos que analizan dinámicas de persona a persona, sus capacidades predictivas resultan limitadas por la naturaleza con la que se generan las bases de datos, como ocurre en [2, 9].

Teniendo en cuenta la aplicabilidad de los autómatas celulares y las cualidades predictivas de los modelos en redes neuronales, se pretende simular el comportamiento de una enfermedad y desarrollar un algoritmo que permita realizar pronósticos sobre su comportamiento, para responder a la pregunta: ¿Qué impacto tienen las relaciones sociales cercanas, en la propagación de una enfermedad?

Este trabajo permitirá comprender de manera más certera, como las interacciones de las personas afectan la propagación de alguna enfermedad y podrá establecer características para construir las bases de datos, que pueden dar lugar a análisis predictivos más realistas. Así mismo, se abre la puerta a que se propongan diferentes algoritmos con el fin de obtener diferentes resultados óptimos que alimenten el estudio epidemiológico.

3. OBJETIVO GENERAL:

Diseñar un algoritmo en redes neuronales que permita realizar pronósticos en el comportamiento de una enfermedad simulada con autómatas celulares, teniendo en cuenta aspectos topológicos que describan las interacciones entre individuos.

4. OBJETIVOS ESPECÍFICOS:

1. Simular la propagación de una enfermedad por medio de un algoritmo basado en autómatas celulares.
2. Desarrollar herramientas que permitan analizar el comportamiento de la enfermedad, en entornos topológicos diferentes para los autómatas.
3. Diseñar un algoritmo en redes neuronales que permita realizar pronósticos sobre el comportamiento de la enfermedad simulada.
4. Validar el comportamiento del algoritmo propuesto, en ambientes topológicos diferentes.

5. ANTECEDENTES (máximo 2000 palabras):

Los modelos epidemiológicos compartimentales clásicos, descritos por Kermack y McKendrick, han sido fuente de numerosos estudios y simulaciones relacionadas con el estudio epidemiológico. Como caso particular, encontramos el modelo propuesto por Fuentes y Kuperman en [3], el cual permite estudiar la evolución de un núcleo de infectividad en varias condiciones y para dos tipos de enfermedades epidemiológicamente distintas, en autómatas celulares. A diferencia de los modelos clásicos en ecuaciones diferenciales [10], proponen un conjunto de reglas de evolución para determinar el estado de un individuo en función de su propio estado y el de sus vecinos en pasos de tiempo anteriores, considerando un radio de "interacción".

En trabajos como el de Mikler-Venkatachalam-Abbas en [8], los autómatas son interpretados como regiones, de modo que se puedan tener en cuenta factores geográficos, demográficos, de medio ambiente y patrones de migración, usando un autómata celular estocástico global (GSCA). Algo comparable se ve en [14], en donde H. White, Martín del Rey y G. Rodríguez implementan la misma noción de autómata espacial, para analizar entre otras cosas el proceso de vacunación para una enfermedad modelada con un enfoque determinista.

Un enfoque diferente se ve en [7], en donde Marr y Hutt se enfatizan en la relación entre topología y teoría de grafos, para analizar dinámicas de redes biológicas. Mostrando que la distribución de grados puede estar relacionada con las propiedades dinámicas de la red, además de mencionar posibles mejoras, basadas en ideas de topología de redes para analizar diferentes grados de entropía.

Encontramos también aplicaciones entre la teoría de redes y la epidemiología. Moo Kim realiza un análisis interesante para la propagación de enfermedades en redes complejas, puntualmente en la red de Barabasi-Albert, presente en diversas redes reales [5]. Realiza además, una implementación de los modelos compartimentales clásicos en redes regulares y complejas.

Desplazándonos un poco a la teoría de redes neuronales y los procesos de machine-learning, encontramos trabajos como el de Schmidt en [13], en donde implementa una red neuronal artificial, basada en modelos compartimentales para predecir la propagación del virus de la influenza, para posteriormente realizar una comparación con un modelo de regresión lineal múltiple.

Otra aplicación interesante es el modelo desarrollado en [1], el cual se centra en la predicción de los ILI a largo plazo, usando datos con un rango de tiempo limitado a 20 semanas. Teniendo presente la pérdida de información temporal, exploran un modelo de propagación en grafos con representaciones espaciales. Usan además, las redes neuronales recurrentes para capturar las dependencias secuenciales en los datos de las series temporales locales y diseñan un módulo de convolución temporal para extraer automáticamente las dependencias temporales cortas y largas de los datos de series temporales de múltiples localizaciones.

La reciente aparición del coronavirus a finales del 2019, ha incentivado estudios que proponen diversos enfoques basados en redes neuronales convolucionales de grafos (GCNs), memorias de corto plazo (LSTMs) y modelos de aprendizaje de transferencia basados en el meta-aprendizaje (MAML) [2, 4]. Mientras que en [4] proponen un modelo basado en machine-learning para estimar los parámetros de un modelo epidemiológico que se ajusta a tendencias observadas sobre el movimiento de individuos, en [2] proponen un enfoque de fusión entre redes neuronales basadas en grafos y modelos epidemiológicos para combinar directamente los datos de movilidad diádica y conectividad derivadas de Facebook con información estructural y espacial de distritos alemanes.

Panagopoulos, Nikolentzos y Michalis Vazirgiannis realizan un trabajo comparable en [9], en donde se centran en predecir el comportamiento de la enfermedad basándose en los datos de movilidad de masa, y consideran la circulación dentro de una región como un indicador de interacción. Exponen además, la inexistencia de una base de datos que considere las posibles interacciones entre individuos y sus implicaciones.

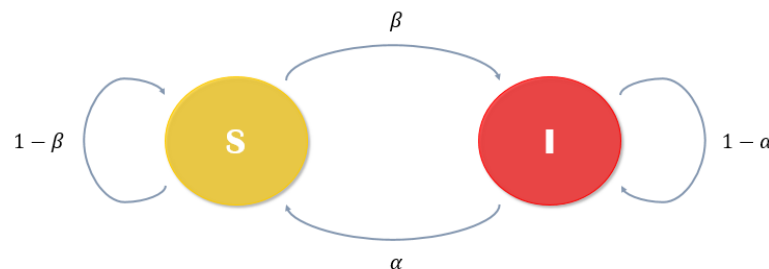
Después de revisar diferentes desarrollos en el campo de la epidemiología, se ratifica la importancia de un análisis predictivo que tenga en cuenta las relaciones o interacciones más cercanas, de modo que se establezcan estrategias de prevención y contención mucho más certeras.

6. MARCO CONCEPTUAL:

Modelos compartimentales

El modelo SIS

El modelo *SIS* describe la interacción entre dos estados, susceptible (S) e infectado (I), el paso de un estado a otro viene determinado por las constantes α y β , las cuales representan las tasas de recuperación e infección respectivamente.



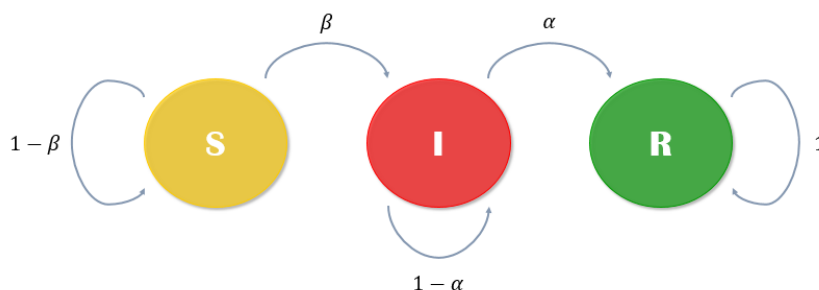
De donde se obtiene el siguiente sistema de ecuaciones diferenciales:

$$\begin{cases} S' = \alpha I - \beta SI \\ I' = \beta SI - \alpha I \end{cases} \quad (1)$$

Para modelar el comportamiento de los individuos mediante el sistema de ecuaciones diferenciales descrito en 1, se consideran los contactos entre las poblaciones de susceptibles e infectados a partir del producto de las variables S e I . El término αI indica la probabilidad de que un individuo se recupere mientras que el término βSI indica la probabilidad de que un individuo susceptible se infecte al de tener contacto con un individuo infectado.

El modelo SIR

A diferencia del modelo SIS, el modelo SIR describe las interacciones entre los estados susceptible (S), infectado (I) y recuperado (R). Es posible que un individuo susceptible adquiera el virus y que posteriormente pueda recuperarse, sin embargo en este modelo no es posible que un individuo que no ha tenido la enfermedad en algún momento se cure de está, para comprender de una mejor manera el funcionamiento del modelo *SIR* observemos el siguiente diagrama:



Al igual que en el caso del modelo *SIS*, las constantes α y β representan las tasas de recuperación e infección respectivamente; una vez dicho esto podemos describir el modelo *SIR* en ecuaciones diferenciales:

$$\begin{cases} S' = -\beta SI \\ I' = \beta SI - \alpha I \\ R' = \alpha I \end{cases} \quad (2)$$

Tal y como ocurría en el modelo *SIS*, los contactos entre poblaciones se van a representar con el producto de variables y los factores αI y βSI serán equivalentes respectivamente, a la recuperación de un porcentaje α de individuos infectados y al paso al estado de infección por parte de un porcentaje β de individuos susceptibles, luego de tener contacto con infectados.

Autómatas celulares

Son un modelo matemático capaz de describir el comportamiento de diferentes sistemas dinámicos, están compuestos por un conjunto de agentes (también llamados celdas, células, individuos o píxeles) que toman algún valor o estado.

El *estado* de cada individuo es alterado en mediciones discretas de tiempo, usualmente esta alteración del estado de la célula depende del comportamiento sus individuos cercanos también llamados *vecinos*, la regla que describe la relación entre el conjunto de estados, el agente y sus vecinos está determinada por una expresión matemática la cual se conoce la *regla de evolución local*.

Para definir las reglas de evolución local de manera que se tengan en cuenta las tasas de infección y de recuperación, junto con el estado de cada vecino se usarán las reglas semi-totalísticas, las cuales consideran el estado de la célula central para posteriormente asignar un elemento del conjunto de estados a la suma de los valores de los elementos que forman la vecindad, estos valores determinan el comportamiento de todas las vecindades cuya suma de valores de sus elementos corresponda a un mismo valor.

Cada individuo puede tomar un único valor del conjunto de estados que se modelan con el autómata celular, en nuestro caso será uno de dos estados para el modelo SIS (o tres en el caso del modelo SIR). Es importante resaltar la complejidad que puede llegar a tener un autómata celular, debido al comportamiento en la vecindad de cada celda.

En el caso de sistemas 2-dimensionales encontramos una gran cantidad de vecindades, entre las más conocidas encontramos la *vecindad de Von Neumann* la cual considera a la célula central y a los individuos ubicados a la izquierda, derecha, arriba y abajo y la *vecindad de Moore* la cual añade los individuos diagonales a la vecindad de Von Neumann. Para los intereses del proyecto, se decidió trabajar con vecindades de Moore.



Vecindad de Von Neumann



Vecindad de Moore

Topología

Dado que uno de los objetivos del proyecto es trabajar con diferentes topologías definidas por un sistema de vecindades locales, debemos tener presentes las siguientes definiciones:

Definición: Sea X un conjunto no vacío y τ una colección de subconjuntos de X . Diremos que τ es una *topología* sobre X si:

1. \emptyset y X están en τ .
2. La unión de elementos de cualquier sub colección de τ es un elemento en τ .
3. Toda intersección finita de elementos de τ está en τ .

Diremos que la dupla (X, τ) es un *espacio topológico* si τ es una topología sobre X , adicionalmente conoceremos a los elementos de τ como conjuntos abiertos.

Definición: Sea (X, τ) un espacio topológico y $Y \subset X$. La colección $\tau' = \{A \cap Y | A \in \tau\}$ es una topología sobre Y , la cual conoceremos como topología relativa de Y . Con esta topología, Y se denomina subespacio de X .

Definición: Sea (X, τ) un espacio topológico y $x \in X$. Diremos que un conjunto V es una vecindad de x si existe $A \in \tau$ tal que $x \in A \subset V \subset X$.

Denotaremos a la familia de vecindades de x por $\mathcal{V}(x)$.

Proposición: Sea (X, τ) un espacio topológico, entonces para cada x se tiene lo siguiente:

1. Si $U \in \mathcal{V}(x)$, entonces $x \in U$.
2. Si $U, V \in \mathcal{V}(x)$, entonces $U \cap V \in \mathcal{V}(x)$.
3. Si $U \in \mathcal{V}(x)$ y $u \in U$, entonces $\mathcal{V}(u) \subset U$.
4. Para cada $U \in \mathcal{V}(x)$ existe $V \in \mathcal{V}(x)$ tal que $U \in \mathcal{V}(y)$ para cada $y \in V$.

Definición: Definimos la *base* \mathcal{B} de la topología de un conjunto X , como la colección de subconjuntos de X tales que:

1. Para cada $x \in X$, existe por lo menos un elemento $B \in \mathcal{B}$ que contiene a x .
2. Si $x \in B_1 \cap B_2$, para $B_1, B_2 \in \mathcal{B}$, entonces existe $B_3 \in \mathcal{B}$ tal que $x \in B_3$ y $B_3 \subset B_1 \cap B_2$.

Los elementos B de \mathcal{B} serán conocidos como elementos básicos.

Definición: Sea \mathcal{B} una base para una topología sobre X , un subconjunto U de X se dice que es *abierto* de X , si para cada $x \in U$, existe $B \in \mathcal{B}$ tal que $x \in B$ y $B \subset U$. La colección de subconjuntos U con esta propiedad, se define topología generada por \mathcal{B} .

Definición: Sean (X, τ) un espacio topológico y x un elemento en X . Diremos que una colección $\mathcal{B}(x) \subset \mathcal{V}(x)$ es una *base de vecindades de x* en (X, τ) si para cada $V \in \mathcal{V}(x)$ existe $B \in \mathcal{B}(x)$ tal que $B \subset V$.

Proposición: Sea (X, τ) un espacio topológico, supongamos que para cada $x \in X$ hemos elegido una base de vecindades $\mathcal{B}(x)$, entonces

1. Si $V_1, V_2 \in \mathcal{B}(x)$, entonces existe $V_3 \in \mathcal{B}(x)$ tal que $V_3 \subset V_1 \cap V_2$.
2. Para cada $V \in \mathcal{B}(x)$ se puede escoger $V_0 \in \mathcal{B}(x)$ de modo que para $y \in V_0$, existe $W \in \mathcal{B}(y)$ tal que $W \subset V$.

Definición: Definimos la *frontera* de un conjunto $A \subset X$, como el conjunto de puntos $x \in X$ tales que para todo $V \in \mathcal{V}(x)$ se cumple que $V \cap A \neq \emptyset$ y $V \cap (X \setminus A) \neq \emptyset$.

7. METODOLOGÍA:

Este proyecto estará dividido en cuatro fases principales, que se articulan entre sí y con los objetivos específicos del proyecto. Las fases se ejecutarán en forma secuencial y cada una de ellas generará una serie de módulos concretos que servirán de insumo a las fases subsecuentes. A continuación, se describe brevemente cada una de las fases metodológicas junto con las etapas que se desarrollarán en cada fase:

Etapas	Fases Metodológicas
1. Revisión de modelos compartimentales	Fase 1: Modelos basados en autómatas celulares para la simulación de una enfermedad
2. Diseño de reglas probabilísticas para el comportamiento de la enfermedad en autómatas celulares	
3. Implementación de reglas de comportamiento	
4. Simulación del comportamiento de la enfermedad	
5. Obtención de datos para el algoritmo en redes neuronales	Fase 2: Análisis sobre la simulación de la enfermedad y comparación con modelos compartimentales clásicos
6. Análisis del impacto de la condición de frontera	
7. Comparación con modelos compartimentales clásicos	
8. Revisión de métodos de clasificación en redes neuronales	Fase 3: Diseño e implementación del algoritmo en redes neuronales sobre datos simulados.
9. Definición de métricas para el algoritmo en redes neuronales	
10. Diseño e implementación del algoritmo en redes neuronales sobre los datos simulados	
11. Validación del comportamiento del algoritmo con los datos previamente simulados	Fase 4: Validación del algoritmo en diferentes entornos topológicos
12. Implementación de herramienta para aplicar cambios de topología	
13. Cambios de topología sobre la simulación en autómatas celulares	
14. Validación del algoritmo en redes neuronales con datos obtenidos de diferentes cambios de topología	

Fase 1: Modelos basados en autómatas celulares para la simulación de una enfermedad.

La primera fase del proyecto consistirá en simular el comportamiento de la enfermedad sobre un autómata celular. Para alcanzar este objetivo, será necesario comprender las dinámicas que describen los modelos SIS, SIR y algunas de sus variaciones. Esto permitirá definir las reglas probabilísticas que en conjunto, describirán la propagación de la enfermedad de modo que se tengan en cuenta las interacciones de cada célula.

Inicialmente se trabajará sobre una red uniforme, en donde se consideren las vecindades de Moore o las de Von Neumann.

Fase 2: Análisis sobre la simulación de la enfermedad y comparación con modelos compartimentales clásicos.

Una vez se llegue a un entorno de simulación estable, se realizarán modificaciones en la condición de frontera para posteriormente analizar su impacto en la propagación de la enfermedad.

A continuación se realizarán comparaciones entre la simulación y los modelos compartimentales clásicos. Esto permitirá validar el comportamiento simulado, de acuerdo con la manera en la que se diseñen las reglas de comportamiento en la fase 1.

Fase 3: Diseño e implementación del algoritmo en redes neuronales sobre datos simulados.

Durante esta fase del proyecto, se realizará inicialmente un ejercicio investigativo en modelos de clasificación. Esto permitirá comprender la manera en la que se definen las métricas, que posteriormente serán monitoreadas por el algoritmo en redes neuronales. Tan pronto como se definan las métricas, se dará inicio en paralelo, al diseño e implementación del algoritmo en redes neuronales.

Fase 4: Validación del algoritmo en diferentes entornos topológicos.

Daremos inicio a la fase final del proyecto, con la validación del algoritmo desarrollado en la fase anterior. Las pruebas realizadas durante esta validación, se elaborarán sobre los datos obtenidos de las simulaciones realizadas en la primera fase.

Posteriormente se realizarán las modificaciones sobre la topología con la que se simuló la enfermedad en la primera fase. Para esto, será necesario implementar una herramienta que permita definir y modificar la topología sobre la cual se desarrollan las simulaciones.

Una vez se realice dicha implementación, se desarrollarán las validaciones finales sobre el algoritmo en redes neuronales, con lo que se espera poner a prueba la efectividad del algoritmo propuesto.

8. RECURSOS:

Se espera que el proyecto este desarrollado en su totalidad sobre Python 3.8.

Puntualmente para el desarrollo de la simulación, se usarán las librerías **numpy** para cálculos sobre arreglos matriciales; **matplotlib**, **cv2** y **seaborn** para visualizaciones y algunas características; **random** y **scipy.stats** para valores aleatorios. Mientras que el desarrollo del algoritmo en redes neuronales se usarán las librerías **Keras**, **TensorFlow** y **PyTorch**.

9. CRONOGRAMA DE ACTIVIDADES:

Los tiempos estimados para la ejecución del proyecto se presentan en la siguiente tabla:

Actividades a realizar	Semanas de ejecución de cada actividad															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Revisión de modelos compartimentales	■															
Diseño de reglas probabilísticas		■														
Implementación de reglas de comportamiento		■	■													
Simulación del comportamiento de la enfermedad			■	■												
Obtención de datos para el algoritmo en redes neuronales				■												
Análisis del impacto de la condición de frontera					■											
Comparación con modelos compartimentales clásicos						■										
Revisión de métodos de clasificación en redes neuronales							■									
Definición de métricas para el algoritmo en redes neuronales								■								
Diseño e implementación del algoritmo en redes neuronales sobre datos simulados									■	■	■					
Validación del comportamiento del algoritmo con datos simulados												■				
Implementación de herramienta para aplicar cambios de topología													■			
Cambios de topología sobre la simulación en autómatas celulares														■		
Validación del algoritmo en redes neuronales con datos obtenidos de los cambios de topología															■	■
Redacción del documento final del trabajo de grado		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

10. BIBLIOGRAFÍA:

- [1] Songgaojun Deng y col. "Cola-GNN: Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction". En: *arXiv* (2019).
- [2] Cornelius Fritza, Emilio Dorigattia y David Rugamera. "Combining Graph Neural Networks and Spatio-temporal Disease Models to Predict COVID-19 Cases in Germany". En: *arXiv* (2021).
- [3] M.A. Fuentes¹ y M.N. Kuperman. "Cellular automata and epidemiological models with spatial dependence". En: *ELSEVIER* (1999).
- [4] Valerio La Gatta y col. "An Epidemiological Neural Network exploiting Dynamic Graph Structured Data applied to the COVID-19 outbreak". En: *ELSEVIER* (2021).
- [5] Kyung Moo Kim. "Algunos modelos epidemiológicos en grafos completos y redes complejas". Tesis de mtría. Universidad de Sao Paulo, 2019.
- [6] Ilias N. Lymperopoulos. "stayhome to contain Covid-19: Neuro-SIR – Neurodynamical epidemic modeling of infection patterns in social networks". En: *ELSEVIER* (2021).
- [7] Carsten Marr y Marc-Thorsten Hutt. "Cellular Automata on Graphs: Topological Properties of ER Graphs Evolved towards Low-Entropy Dynamics". En: *entropy* (2012).
- [8] Armin R. Mikler, Sangeeta Venkatachalam y Kaja Abbas. "Modeling infectious diseases using global stochastic cellular automata". En: *World Scientific Publishing Company* (2005).
- [9] George Panagopoulos, Giannis Nikolentzos y Michalis Vazirgiannis. "Transfer Graph Neural Networks for Pandemic Forecasting". En: *arXiv* (2021).
- [10] Diego de Pereda Sebastián. "Modelización matemática de la difusión de una epidemia de peste porcina entre granjas". Tesis de mtría. Universidad Complutense de Madrid, 2010.
- [11] Mario Castro Ponce, Manuel de León y Antonio Gómez Corral. "Las matemáticas contra la malaria y el modelo SIR". En: *Matemáticas y sus fronteras* (2020).
- [12] Kao RR. "Networks and Models with Heterogeneous Population Structure in Epidemiology". En: *Network Science*, 2010. Cap. 4.

- [13] Carina Schmidt. *Prediction of the Influenza Virus Propagation by using different Epidemiological and Machine Learning Models*. 2019.
- [14] S. Hoya White, A. Martín del Rey y G. Rodríguez Sánchez. "Modeling epidemics using cellular automata". En: *ELSEVIER* (2006).

11. FIRMAS

FIRMA DEL ESTUDIANTE:	FIRMA DEL DIRECTOR O TUTOR:

12. DATOS DE TRÁMITE COMITÉ DE INVESTIGACIÓN DEL PROGRAMA (Espacio para diligenciar por el Comité del Programa)

No.CONSECUTIVO	
No. ACTA	
FECHA	