

UNIVERSIDAD DE SALAMANCA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA



VNiVERSiDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

TRABAJO DE FIN DE GRADO

MODELOS ESTOCÁSTICOS
PARA SIMULAR LA
PROPAGACIÓN DEL
MARKETING VIRAL EN UNA
RED SOCIAL

Autor: Pablo Díaz Castillo

Tutor: Angel Martín del Rey

Grado en Estadística

Curso académico 2022-23

UNIVERSIDAD DE SALAMANCA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADÍSTICA



VNiVERSiDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

TRABAJO DE FIN DE GRADO

MODELOS ESTOCÁSTICOS
PARA SIMULAR LA
PROPAGACIÓN DEL
MARKETING VIRAL EN UNA
RED SOCIAL

Firmado por:

Una firma manuscrita en tinta oscura, que parece ser la del autor, Pablo Díaz Castillo. La firma es fluida y estilizada, con una gran 'P' inicial.

Autor: Pablo Díaz Castillo

Tutor: Ángel Martín del Rey

Grado en Estadística

Curso académico 2022-23



Certificado de los tutores TFG Grado en Estadística

D. Ángel Martín del Rey, profesor del Departamento de Matemática Aplicada de la Universidad de Salamanca.

HACEN CONSTAR:

Que el trabajo titulado “*Modelos estocásticos para simular la propagación del marketing viral en una red social*”, que se presenta, ha sido realizado por D. Pablo Díaz Castillo, con DNI 72202490Q y constituye la memoria del trabajo realizado para la superación de la asignatura Trabajo de Fin de Grado en Estadística en esta Universidad.

Salamanca, a 4 de julio de 2023.

Fdo.: Ángel Martín del Rey

Índice general

1. Marketing Viral	19
2. Estado del arte	25
2.1. Epidemiología Matemática	25
2.2. Modelo SIR de Kermack-McKendrick	26
2.3. Modelo SIS	31
2.4. Revisión de modelos de propagación del Marketing viral	32
3. Modelos estocásticos de propagación del marketing viral	37
3.1. Cadenas de Markov en tiempo discreto.	37
3.1.1. Clasificación de estados	39
3.1.2. Comportamiento asintótico	40
3.2. Modelos de propagación del marketing viral	44
3.3. Modelos estocásticos SIS	46
3.3.1. Comportamiento asintótico	49
3.4. Modelos estocásticos SIR	50
3.4.1. Comportamiento asintótico	52
3.5. Pseudocódigo y ejemplo numérico	52
3.5.1. Modelos CMTD SIS	52
3.5.2. Modelos CMTD SIR	54
4. Conclusiones	57

Índice de figuras

2.1. Diagrama de flujo del modelo SIR.	27
2.2. Evolución del modelo SIR con $N = 100$ y condiciones iniciales $S(0) = 99, I(0) = 1$	30
2.3. Diagrama de flujo del modelo SIS.	31
2.4. Diagrama del modelo que muestra todas las transiciones posibles de un estado a otro.	35
3.1. Diagrama del modelo de propagación SIS estocástico	49
3.2. Simulaciones de un modelo de propagación CMTD SIS con tres trayectorias estocásticas usando los parámetros $N = 100, I(0) = 5, \beta = 1, \gamma = 0.5$ y $\Delta t = 0.01$	54
3.3. Simulaciones de un modelo de propagación CMTD SIR con tres trayectorias estocásticas usando los parámetros $N = 100, I(0) = 5, \beta = 1, \gamma = 0.5$ y $\Delta t = 0.01$	56

Resumen

Aunque en la literatura existen modelos que intentan reproducir el proceso de difusión del marketing viral, la mayoría de estos son deterministas y de naturaleza global. El presente trabajo de Fin de Grado tiene como fin diseñar un modelo estocástico de naturaleza individual para simular la propagación del marketing viral en una red social. Para lograr este objetivo, hemos realizado una revisión bibliográfica de los modelos compartimentales deterministas clásicos (SIS y SIR) y de los modelos estocásticos propuestos en la literatura sobre la propagación viral. A partir de esta revisión, hemos creado el modelo basándonos en los modelos deterministas pero aplicando cadenas de Markov en tiempo discreto en vez de ecuaciones diferenciales y a su vez, haciendo uso de los modelos de propagación viral propuestos, hemos diseñado el comportamiento de los compartimentos y la dinámica del modelo.

Palabras clave: marketing viral, epidemiología matemática, modelos compartimentales, cadenas de Markov, mensaje, redes sociales.

Abstract

Although there are models in the literature that attempt to reproduce the viral marketing diffusion process, most of them are deterministic and global. The aim of the present work is to design a stochastic model of individual nature to simulate the spread of viral marketing in a social network. To achieve this goal, we have conducted a literature review of classical deterministic compartmental models (SIS and SIR) and proposed stochastic models of viral propagation. From this review, we have created the model based on the deterministic models but applying discrete-time Markov chains instead of differential equations and using the proposed viral propagation models we have designed the behavior of the compartments and the dynamics of the model.

Keywords: viral marketing, mathematical epidemiology, compartmental models, Markov chains, message, social networks.

Introducción

En la era digital actual, las redes sociales se han convertido en una parte fundamental de la sociedad, transformando la manera en que las personas interactúan, se comunican, comparten información y consumen contenido en línea. El uso de smartphones y otros dispositivos móviles se ha vuelto indispensable, por ello las redes sociales se han convertido en una herramienta omnipresente y accesible para millones de personas en todo el mundo.

Las redes sociales ofrecen a los usuarios la posibilidad de conectarse con amigos, familiares y conocidos, así como de conocer a nuevas personas con intereses y opiniones similares. Además, han cambiado la forma en que las personas consumen información, permitiendo a los usuarios seguir noticias, eventos y tendencias en tiempo real.

Aparte de sus efectos en la comunicación, las redes sociales también han tenido un impacto en la salud mental y emocional de los usuarios. Las redes sociales pueden ser adictivas y consumir mucho tiempo, y la comparación constante con los demás puede afectar negativamente la autoestima y la percepción de la vida de uno.

A pesar de estas preocupaciones, las redes sociales continúan evolucionando y su influencia en la sociedad y cultura sigue creciendo. Como tal, el estudio de estas se ha convertido en un área importante de investigación en campos como la sociología, la psicología y el marketing digital, entre otros. Es por ello por lo que el marketing digital ha emergido como una forma altamente efectiva de llegar a los clientes. Las empresas ahora pueden aprovechar las redes sociales para promocionar sus productos y servicios, interactuar con los clientes, y obtener información valiosa sobre los intereses y las necesidades de los consumidores.

Dentro de este ámbito existe un fenómeno denominado ‘Marketing viral’, que ha despertado gran interés en el área de la investigación. Su estrategia se basa en la creación de contenidos que sean atractivos y que se propaguen rápidamente por los usuarios a través de las redes sociales, con el propósito de aumentar la visibilidad de una marca

o producto de manera económica y efectiva.

El objetivo del marketing viral es lograr una amplia difusión del mensaje con un mínimo de inversión publicitaria. Para ello, se crea contenido que resulte atractivo, innovador y divertido, con el fin de que los usuarios sientan la necesidad de compartirlo con sus contactos. Dado que varios estudios señalan claramente que la herramienta del boca-boca en internet es la clave del marketing en el mundo actual, la importancia del análisis cualitativo para comprender la dinámica del marketing viral cada vez es más innegable. Así pues, la simulación de la propagación del marketing viral se ha convertido en una herramienta valiosa para los investigadores y profesionales del sector, ya que permite entender mejor los procesos de difusión del contenido en línea y desarrollar estrategias efectivas para maximizar su impacto.

Es aquí donde entra la epidemiología matemática, disciplina que combina los principios de la epidemiología, la estadística y las matemáticas para comprender y predecir la propagación de los agentes biológicos (virus, bacterias, etc.) que causan enfermedades en las poblaciones. Utilizando modelos matemáticos, la epidemiología matemática busca analizar y cuantificar la dinámica de las enfermedades infecciosas, así como evaluar la efectividad de estrategias de control y prevención.

La dinámica de propagación de las campañas de marketing viral es muy similar a la de las enfermedades infecciosas, ya que tienen una cualidad contagiosa y se propagan a través de la interacción social. Es por ello que nuestro trabajo está basado en modelos de la epidemiología matemática pero aplicándolos al marketing viral.

El trabajo a realizar consiste en diseñar un modelo matemático de naturaleza estocástica para simular la propagación del marketing viral. Para poder realizar dicho estudio se utilizan modelos matemáticos basados en diferentes herramientas matemáticas: ecuaciones diferenciales ordinarias, modelos basados en agentes y autómatas celulares, cadenas de Markov, etc. Siendo este último el implementado en este trabajo de Fin de Grado.

Para el diseño de dichos modelos se suele considerar que el tamaño de la población es constante y que hay diferentes tipos de compartimentos en los que se dividen los individuos, de forma análoga a como se hace en los estudios epidemiológicos de enfermedades infecciosas. Algunos tipos de compartimentos que se han considerado son: ignorantes (I), difusores (D), inertes (S), etc. A partir de los compartimentos se pueden considerar diferentes modelos según sus dinámicas: SIR ; susceptible, infectado y recuperado (ignorantes, difusores e inertes), SIS ; susceptible e infectado (ignorantes y

difusores), etc.

Atendiendo a distintas características matemáticas que tiene el modelo podemos clasificarlo de diferentes maneras entre las que destacan las descritas en [13]:

- Modelos estocásticos o determinísticos: Los modelos estocásticos son aquellos que tienen en cuenta los factores aleatorios que influyen en la dinámica del fenómeno. Por otro lado están los deterministas, que no tienen en cuenta dichos factores debido a que no son muy relevantes en el problema o se conoce la evolución de todas las propiedades características del fenómeno.
- Modelos globales o individuales: Los modelos globales son aquellos que se construyen a partir de características globales de fenómeno (obtenemos soluciones globales de un problema). Opuestamente están los modelos individuales los cuales consideran las características individuales de cada uno de los actores del fenómeno (obtenemos soluciones individuales y globales del problema).

En este trabajo nos centraremos fundamentalmente en diseñar un modelo de difusión de naturaleza individual y basado en autómatas celulares estocásticos usando cadenas de Markov. Para ello, haremos una revisión bibliográfica de los modelos ya propuestos dentro de la literatura científica con el inconveniente de que existen muy pocos modelos ya desarrollados y la gran mayoría de ellos de naturaleza global. Es, por tanto, necesario el desarrollo e implementación computacional de un modelo de naturaleza individual para simular la propagación del marketing viral.

Ahora una vez definidos los términos principales de este trabajo, veremos su estructura:

En el primer capítulo explicaremos en que consiste el marketing viral, sus fundamentos básicos y analizaremos los actos y características que intervienen en un proceso de difusión.

Posteriormente, en el segundo capítulo realizaremos una revisión de la teoría necesaria para entender los modelos creados, describiremos el modelo SIR de Kermack-McKendrick y el modelo SIS, modelos basados en ecuaciones diferenciales ordinarias y en los que nos basaremos para realizar el modelo posteriormente y finalmente, haremos una revisión de los modelos estocásticos referidos a la propagación del marketing viral. El tercer capítulo estará enfocado en el diseño de los modelos estocásticos para la propagación del marketing viral, específicamente los modelos SIR y SIS. Previo a esto, se realizará una introducción a las cadenas de Markov en tiempo discreto, que sirven de

base para nuestros modelos. El capítulo concluirá con la presentación de un ejemplo numérico que ilustrará los dos modelos propuestos.

En nuestro capítulo final explicaremos las conclusiones extraídas acerca del tema tratado.

Por último, se ofrecerá la bibliografía utilizada para la realización de este trabajo, en la que encontraremos diversos libros, artículos, y páginas web.

Objetivos

El objetivo principal de este trabajo es diseñar un modelo estocástico de naturaleza individual y basado en cadenas de Markov en tiempo discreto para simular la propagación del marketing viral en una red social.

Objetivos secundarios

- Realizar una revisión bibliográfica acerca del marketing viral y los factores que influyen en la propagación.
- Realizar una revisión bibliográfica de la epidemiología matemática y los modelos compartimentales deterministas.
- Realizar una revisión bibliográfica de los modelos estocásticos propuestos para simular la diseminación del marketing viral a través de una red social.
- Indicar las ventajas y desventajas de los modelos estocásticos frente a los deterministas.

Capítulo 1

Marketing Viral

En esta primera sección, abordaremos el concepto de marketing viral y analizaremos cómo se propaga, así como los elementos y características que intervienen en este proceso de difusión.

Los consumidores se resisten cada vez más a las formas tradicionales de publicidad, como los anuncios en la televisión o en la prensa, por lo que los profesionales han recurrido a estrategias alternativas, como el marketing viral. El marketing viral es una estrategia de difusión de información de un producto, servicio o empresa que se basa en la propagación de contenido de manera rápida y masiva a través de diferentes medios como las redes sociales, el correo electrónico, blogs u otros medios digitales.

El objetivo del marketing viral es generar una gran cantidad de interacciones con el contenido (compartir, comentar, dar me gusta) y así lograr una amplia difusión del mensaje con un mínimo de inversión publicitaria. Para ello, se crea contenido que resulte atractivo, innovador y divertido para el público objetivo, con el fin de que los usuarios sientan la necesidad de compartirlo con sus contactos.

El marketing viral se propaga a través de una red de contactos existentes. Predominan las redes sociales, que son el medio más apto para la difusión rápida y sencilla de contenidos. Estas, proporcionan una audiencia masiva, facilidad para compartir, interacción con la audiencia y medición de los datos en tiempo real. Dentro de las redes sociales, según Grace Parazzoli [19], existen cuatro tipos de estrategias de viralidad: Estrategia de atracción, de contenido, ‘boca-boca’ y de recomendación.

La primera de estas consiste en atraer a los usuarios a utilizar un producto o servicio mediante la mejora de su experiencia y el incremento de los beneficios asociados. Un ejemplo claro de esta estrategia son las redes sociales, como Facebook, Instagram o

Twitter, que ofrecen una plataforma para conectar con amigos, compartir contenido y obtener entretenimiento. A medida que se incrementa el número de personas que utilizan estas redes sociales, aumentan los beneficios del uso de la plataforma. Otro ejemplo de esta estrategia son los programas de comunicación y colaboración, como Zoom, Dropbox o Spotify. Al utilizar estos servicios, los usuarios experimentan una mejora en su productividad, eficacia o entretenimiento, lo que aumenta su motivación para seguir utilizando el producto y, potencialmente, compartirlo con otros.

La estrategia de contenido es el tipo de viralidad más reconocible. Esto es cuando el contenido, en particular, publicado en las redes sociales se hace viral. En el contexto de internet y de las redes sociales, los clientes o las empresas pueden compartir sus experiencias y opiniones con todo el mundo o con toda su red de contactos o seguidores. En estos casos, la valoración de los productos y comerciantes influye en la probabilidad de que se compre un artículo [17]. Hoy en día, los sistemas de recomendación permiten a los usuarios puntuar las opiniones de otros o puntuar directamente a otros críticos para formar implícitamente una red de críticos de críticos de confianza que puede coincidir muy poco con el círculo social real de una persona.

La estrategia boca-boca, se basa en el principio de que las personas confían más en las recomendaciones de amigos y familiares que en la publicidad tradicional. Es por esto, que el contenido generado por los usuarios (UGC) a menudo tiene un impacto mucho mayor que el contenido promocional creado por las empresas, ya que es percibido como más auténtico y confiable. La naturaleza humana está más interesada en lo que compra un amigo que en lo que compra una persona anónima y es más probable que confíe en su opinión y se deje influir más por sus acciones. Como es de suponer, nuestros amigos también conocen nuestras necesidades y gustos y pueden hacernos recomendaciones adecuadas. Según una encuesta de Lucid Marketing, el 68 % de los encuestados consultó a sus amigos y familiares antes de comprar un producto, más de la mitad de los que utilizaron motores de búsqueda para encontrar información sobre este [5]. Frenzen y Nakamoto [8] estudiaron a un grupo de personas y descubrieron que cuanto más fuerte es el riesgo moral que presenta la información, más fuertes deben ser los vínculos para fomentar la propagación de la información.

Por último, estaría el marketing de recomendación, el cual es el único tipo no orgánico ya que incentiva la transmisión del mensaje o de la información a través de recompensas como descuentos en tiendas o diversas promociones.

Además las características de las redes de comunicación influyen en las pautas de

compra de sus miembros. Por ejemplo, la probabilidad de que un individuo compre un producto aumenta inicialmente a medida que recibe más recomendaciones sobre él, pero llega un punto en el que se alcanza un punto de saturación. Curiosamente a medida que se envían más recomendaciones entre las dos mismas personas, disminuye la probabilidad de que sean tenidas en cuenta [17]. Por consiguiente, el marketing viral se centra mucho en la psicología de compartir, esto es, las razones que llevan a las personas compartir contenido. Entre ellas destacan, la identidad social, las emociones, el entretenimiento y la utilidad [19].

Aunque crear una campaña publicitaria viral es una forma rentable y rápida de hacer correr la voz, en la redes sociales de hoy en día, tan activas, existe una enorme posibilidad de que una campaña publicitaria sea increíblemente efímera. Es por ello importante tener en cuenta que no todos los intentos de marketing viral tienen éxito, no existe una fórmula que haga que los usuarios empiecen a compartir algo de forma masiva, y a menudo se desconoce cuál es la causa exacta hasta que interviene la viralidad. Sin embargo, en todas las campañas existen una serie de características esenciales para incentivar la viralización y así tener mejores posibilidades de éxito.

A continuación se describen algunos de los actos y características que intervienen en el proceso de difusión (véase [3] y [21]):

- **Identificación del público objetivo:** Para que el marketing viral tenga éxito es fundamental identificar y entender al público objetivo, para así adaptar el contenido y la estrategia a sus intereses y necesidades.
- **Creación de contenido atractivo:** El contenido que se comparte debe ser atractivo y llamativo, para que así las personas se sientan motivadas a compartirlo con sus contactos. Este contenido suele requerir de una estrategia visual, que guíe a los clientes potenciales a que entiendan su marca mediante el uso de imágenes y vídeos. Una campaña debe contar una historia y esa historia se cuenta mejor utilizando elementos visuales que resuenen con su audiencia.
- **Estrategia de incentivos:** Ofrecer incentivos o recompensas a las personas que comparten el contenido es una estrategia común en el marketing viral. Estos incentivos pueden ser descuentos, promociones, participación en un sorteo, etc.
- **Elección de la plataforma:** Es importante elegir el medio a través el cual vas a difundir el contenido, ya que no todas las plataformas son igualmente efectivas

para todas las audiencias. En cuanto a las redes sociales, el contenido visual como imágenes y vídeos tiende a ser más efectivo en plataformas como Instagram y TikTok, mientras que los artículos y publicaciones largas son más efectivos en plataformas como Facebook y LinkedIn.

- **Compartir en el momento adecuado:** Tener en cuenta la fecha y la hora en la que se comparte el contenido es esencial para maximizar el alcance. Aprovechar las grandes fiestas, como la navidad, y los grandes acontecimientos, como la Super Bowl o Eurovision, puede aumentar la visibilidad y propiciar la propagación.
- **Influencers:** Los influencers son personas con una gran cantidad de seguidores en las redes sociales y suelen ser una herramienta efectiva para propagar el marketing viral. Estos comparten el contenido con su audiencia y aumentan el alcance del mensaje.

Una vez descritos los fundamentos básicos del marketing viral e identificados los principales factores y características que intervienen en el proceso de difusión. Procedemos a explicar como vamos a abordar el problema, para ello haremos uso de la modelización y epidemiología matemática.

La mayor parte de la investigación previa sobre el flujo de información e influencia a través de las redes se ha realizado en el contexto de la epidemiología y la propagación de enfermedades. Los modelos clásicos de propagación de enfermedades se basan en las etapas de una enfermedad de un huésped: una persona es primero susceptible a una enfermedad, luego, si se expone a un contacto infeccioso, puede infectarse y, por tanto, ser infecciosa. Tras el cese de la enfermedad, la persona se recupera o se elimina. La persona es inmune durante algún tiempo. La inmunidad puede desaparecer y la persona vuelve a ser susceptible. Así, el modelo SIR (susceptible/infectado/recuperado) modela enfermedades en las que una persona recuperada nunca vuelve a ser susceptible, mientras que el modelo SIS, (susceptible/infectado/susceptible) modela una población en la que el huésped recuperado puede volver a ser susceptible. Estos son los modelos en los que nos basaremos para realizar el modelo estocástico, y los profundizaremos en el siguiente capítulo. En nuestro contexto, los individuos susceptibles serán aquellos que desconozcan el mensaje o la información objeto de estudio. Los individuos infecciosos serán los difusores, individuos que son conocedores del mensaje o de la información y son capaces de propagarlo y compartirlo a otros individuos. Por último, los individuos

recuperados serán individuos inertes, individuos conocedores del mensaje pero que no forman parte de la propagación por diversas razones como el olvido o el aburrimiento.

En el siguiente capítulo, llevaremos a cabo una revisión del estado del arte de los modelos previamente desarrollados y realizaremos una breve introducción a la epidemiología matemática, el campo en el que se fundamentan estos modelos. Además, presentaremos el modelo de Kermack-McKendrick, un pilar fundamental de la epidemiología matemática moderna.

Capítulo 2

Estado del arte

En este capítulo se hará un resumen de la teoría necesaria para entender los modelos creados en este trabajo de fin de grado así como sus antecedentes.

2.1. Epidemiología Matemática

La Epidemiología Matemática es la disciplina científica que se encarga del diseño y análisis de modelos matemáticos que simulan la propagación de agentes biológicos (virus, bacterias, etc). Esta rama de las matemáticas y la estadística es esencial para comprender la dinámica de las enfermedades infecciosas y la eficacia de las medidas de control y prevención.

La modelización matemática de la propagación de agentes biológicos combina diferentes herramientas y técnicas de las matemáticas, la estadística y la informática, que permiten desarrollar e implementar computacionalmente modelos para simular la propagación de enfermedades infecciosas y su posterior control.

La epidemiología matemática se ha utilizado para abordar una amplia variedad de enfermedades, desde la peste o la viruela hasta enfermedades emergentes como el Covid-19. Esta última, ha provocado un gran crecimiento en esta disciplina en los últimos años, debido a la necesidad de intentar entender como se comportaba la pandemia ocasionada por dicha enfermedad. Sin embargo, en la actualidad la epidemiología matemática ya no tiene como único objeto de estudio los agentes de naturaleza biológica sino que también tiene en cuenta otros como: El código malicioso (malware), las noticias falsas (fake news) o el marketing viral. Es por ello, que el modelo a realizar en este trabajo esta basado en la Epidemiología Matemática.

El primer ejemplo de utilización de la modelización epidémica apareció en 1760

cuando D. Bernoulli [6] propuso un modelo para el estudio de la propagación de la viruela basado en un sistema de ecuaciones diferenciales ordinarias. No obstante, el estudio y desarrollo de estos modelos no se inició hasta principios del siglo XX. En 1906, W.H Hamer [12] propuso un modelo matemático discreto para estudiar la propagación del sarampión. Este modelo fue el primero que sugirió el principio de acción de masas, que establece que la incidencia de una epidemia es proporcional al producto de la densidad de individuos susceptibles por la densidad de individuos infecciosos. Por otro lado, destacar el modelo matemático de R.Ross [20] en 1911, el cual intenta predecir el comportamiento de una epidemia de malaria.

Se puede afirmar que la Epidemiología Matemática moderna nace en 1927 con el trabajo de W.O. Kermack y A.G. McKendrick [16], puesto que a partir del paradigma establecido se han basado la gran mayoría de modelos propuestos desde entonces.

El trabajo clásico de Kermack y McKendrick sentó las bases de la modelización compartimental para comprender la propagación de enfermedades. Estos compartimentos varían en función del tipo de infección que se quiera modelizar. Cada individuo es clasificado en compartimentos según su estado de salud en relación con el desarrollo de la enfermedad. En los casos generales, un miembro de la población cae en uno de estos tres compartimentos: Susceptible, infeccioso o recuperado. Dependiendo de la enfermedad, pueden ser necesarios distintos compartimentos. En el caso de la varicela, después de recuperarse de la infección, es poco probable que una persona vuelva a infectarse y, por lo tanto, el compartimento de recuperado es necesario. Sin embargo, en infecciones como la gripe se puede volver a entrar en contacto y no hay inmunidad permanente. Así que la gripe puede modelarse sin compartimento recuperado.

A continuación, explicaremos el funcionamiento y formulación de los modelos SIR de Kermack-McKendrick y el modelo SIS, ya que son los dos modelos en los que nos vamos a basar para desarrollar los modelos estocásticos basados en cadenas de Markov. Posteriormente dichos modelos se explicarán particularizándolos al caso de propagación del marketing viral en una red social.

2.2. Modelo SIR de Kermack-McKendrick

El modelo Kermack-McKendrick (véase [16]) es un modelo matemático basado en ecuaciones diferenciales ordinarias que se utilizó para la simulación de la propagación de la peste bubónica acaecida en Londres desde 1665 a 1666 y que se saldó con un 20 %

de la población.

Su característica fundamental es que se trata del primer modelo compartimental en el que la población es dividida en tres clases diferentes: individuos susceptibles, individuos infecciosos e individuos recuperados o muertos [9].

- Susceptibles: Aquellos individuos que se encuentran sanos y pueden ser infectados por el agente biológico.
- Infecciosos: Aquellos individuos que han sido infectados por el agente biológico y son capaces de transmitirlo.
- Recuperados o muertos: Aquellos individuos infectados que se han recuperado de la infección o que han muerto como consecuencia de la misma.

El modelo de Kermack-Mckendrick es un modelo SIR, el cual describe la transmisión de enfermedades infecciosas en la que los individuos recuperados ya no pueden volver a infectarse. Una persona susceptible se infecta con la enfermedad y más tarde se recupera de ella. En este modelo existe inmunidad permanente y los individuos recuperados permanecen permanentemente en el compartimento recuperado. La Figura 2.1 ilustra la relación entre las tres compartimentos.

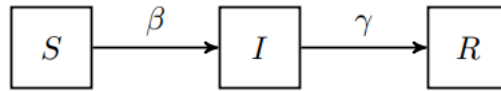


Figura 2.1: Diagrama de flujo del modelo SIR.

La dinámica del modelo de Kermack-McKendrick viene regida por el siguiente sistema de ecuaciones diferenciales ordinarias:

$$S'(t) = -\beta \cdot S(t) \cdot I(t), \quad (2.1)$$

$$I'(t) = \beta \cdot S(t) \cdot I(t) - \gamma \cdot I(t), \quad (2.2)$$

$$R'(t) = \gamma \cdot I(t), \quad (2.3)$$

$$S(0) > 0, I(0) > 0, R(0) \geq 0,$$

donde $S(t)$, $I(t)$ y $R(t)$ representan el número de individuos susceptibles, infecciosos y recuperados en el instante de tiempo t , respectivamente. Las constantes β y γ son positivas, siendo β la tasa de transmisión, que podemos definir grosso modo como la tasa, por unidad de tiempo, de que un individuo susceptible se contagie al estar en

contacto con un individuo infectado y γ , la tasa de recuperación, (tasa, por unidad de tiempo, de que un individuo infectado se recupere).

La ecuación (2.1) indica que la variación del número de susceptibles depende del número de contactos entre los individuos susceptibles y los individuos infectados y de la tasa de transmisión. La ecuación (2.2) indica que la variación del número de infectados es la diferencia entre los nuevos infectados y los que se han recuperado. La ecuación (2.3) indica que la variación del número de recuperados es proporcional al número de infectados.

El tamaño total de la población satisface que $S(t) + I(t) + R(t) = N$ y consecuentemente es constante $\frac{dN}{dt} = 0$, suponemos que no hay ni nacimientos ni muertes naturales. Asimismo, la infección sólo puede transmitirse horizontalmente, es decir, la transmisión es de persona a persona y no de padres a hijos.

La dinámica del modelo viene determinada por el *número reproductivo básico*, que está definido como:

$$R_0 = \frac{\beta \cdot N}{\gamma}. \quad (2.4)$$

R_0 es el número esperado de contagios causados por un único individuo infeccioso durante su periodo infeccioso en una población enteramente susceptible. De manera que si R_0 es mayor que 1, es decir, si la tasa de transmisión β es mayor que la tasa de recuperación γ habrá epidemia, mientras que si R_0 es menor que 1 ($\beta < \gamma$), no se producirá epidemia, en el sentido de que el número de individuos infecciosos no aumentará.

Una de las principales aportaciones del modelo de Kermack-McKendrick fue el enunciado del Teorema Umbral, el cual permite determinar cuándo una enfermedad se convierte en epidémica: esto sucede cuando el número de individuos inicialmente susceptibles es mayor que un determinado valor umbral. Grosso modo dicho teorema dice lo siguiente [9]:

- Si $S(0) < \frac{\gamma}{\beta}$ entonces el número de individuos infectados decrecerá progresivamente y el brote desaparecerá (no se producirá epidemia).
- Si $S(0) > \frac{\gamma}{\beta}$ entonces el número de individuos infectados crecerá (produciéndose una epidemia).

Para analizar mejor el comportamiento de las soluciones de las ecuaciones (2.1) - (2.3), realizamos un estudio cualitativo del modelo, con la ayuda de [18].

Empezando por los susceptibles, sabemos que $S(t)$ es monótona decreciente, puesto que β es positivo, podemos deducir que $S'(t) < 0$ (véase la ecuación (2.1)). Si un individuo se infecta, este adquiere inmunidad, por lo que no regresa al compartimento de los susceptibles y además está acotada inferiormente $S(t) \geq 0$, debido a que el número de individuos susceptibles no puede ser negativo. Ahora, si tomamos la primera (2.1) y la tercera ecuación (2.3) de nuestro sistema *SIR* podemos obtener la ecuación

$$\frac{S'(t)}{R'(t)} = -\frac{\beta S(t)}{\gamma},$$

entonces $S(t)$ satisface lo siguiente:

$$S(t) = S(0) \cdot e^{-\frac{\beta}{\gamma} \cdot R(t)}. \quad (2.5)$$

Ahora, como $R \leq N$ podemos observar que $S(t) \geq S(0) \cdot e^{-\frac{\beta \cdot N}{\gamma}}$ y por tanto $S(t)$ convergerá hacia un valor constante positivo ($S(\infty) > 0$):

$$S(\infty) = \lim_{t \rightarrow \infty} S(t) = S(0) \cdot e^{-\frac{\beta}{\gamma} \cdot R(\infty)}.$$

Esto quiere decir que algunos individuos nunca sufren la enfermedad, ya que todavía hay susceptibles disponibles cuando la enfermedad se detiene. También esto implica que la enfermedad desaparece por falta de infecciosos y no por falta de susceptibles.

Pasando al análisis de los infecciosos. De la ecuación (2.2) se deduce lo siguiente:

$I'(t) > 0$ si y solo si $\beta S(t) - \gamma > 0$ y esto se cumple si y solo si $\frac{\beta S}{\gamma} > 1$.

De manera similar se concluye que $I'(t) > 0$ si y solo si $\frac{\beta S(t)}{\gamma} > 1$.

Como hemos visto que S decrece para todo t concluimos que si en un principio $\frac{\beta S}{\gamma} > 1$ (I crece), luego de un tiempo $\frac{\beta S}{\gamma} < 1$ y por tanto I decrecería. Por tanto, como $R_0 > \frac{\beta S}{\gamma}$, ya que $N > S$ entonces

- Si $R_0 \leq 1$, entonces $I(t)$ es monótona decreciente ($I'(t) < 0$) y:

$$I(\infty) = \lim_{t \rightarrow \infty} I(t) = 0$$

- Si $R_0 > 1$, entonces $I(t)$ crece hasta alcanzar su valor máximo. Dicho valor sería:

$$I_{max} = N - \frac{\gamma}{\beta} - \frac{\gamma}{\beta} \cdot \log\left(\frac{\beta \cdot S(0)}{\gamma}\right),$$

este valor máximo viene determinado por el *valor umbral* ($\frac{\gamma}{\beta}$). Como hemos explicado anteriormente, en el momento en el que el número de susceptibles se

reduzca hasta ser inferior que el *valor umbral*, el número de individuos infectados decrecerá hasta desaparecer.

$$I(\infty) = \lim_{t \rightarrow \infty} I(t) = 0.$$

Por tanto, si $R_0 > 1$, el número de infecciosos aumenta hasta I_{max} y luego empieza a disminuir hasta desaparecer.

Por último, los recuperados $R(t)$ tienen una función creciente (véase la ecuación (2.3)), ya que a medida que los individuos se van recuperando, estos adquieren la inmunidad a la enfermedad. Además está acotada superiormente, debido a que el número de recuperados no puede superar al número de la población total. Por tanto, $R(t)$ convergerá hacia un valor constante positivo:

$$R(\infty) = \lim_{t \rightarrow \infty} R(t).$$

Una vez analizado cualitativamente las tres ecuaciones de nuestro modelo SIR , podemos concluir que el modelo SIR posee un único punto de equilibrio, denominado *punto de equilibrio libre de infección* $E_0 = (S(\infty), 0, R(\infty))$. El modelo solo alcanza el equilibrio en el momento que el número de individuos infectados desaparece por completo.

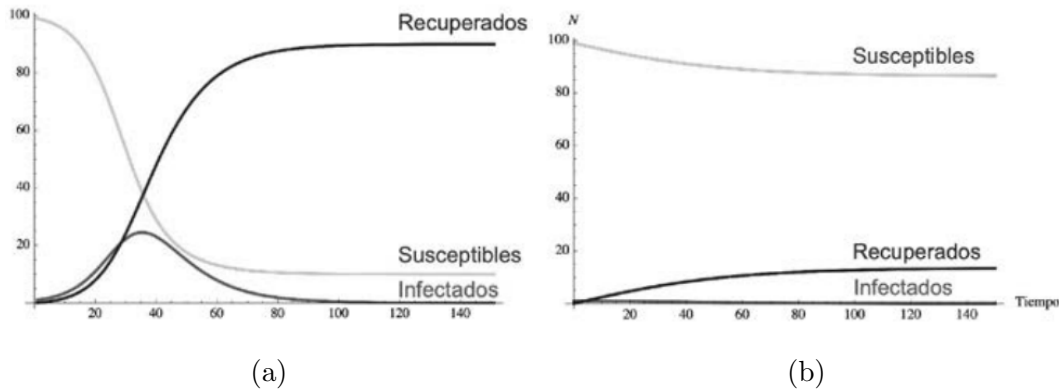


Figura 2.2: Evolución del modelo SIR con $N = 100$ y condiciones iniciales $S(0) = 99, I(0) = 1$.

En la Figura 2.2 observamos un ejemplo de la evolución de las diferentes clases de individuos. En el primer apartado consideramos $\beta = 0,25, \gamma = 0,1$ con lo que $\beta > \gamma$ ($R_0 > 1$), se producirá epidemia. En el siguiente apartado tenemos $\beta = 0,15$ y $\gamma = 0,15$ por lo que no aumentará los individuos infectados ya que $R_0 = 1$.

2.3. Modelo SIS

El modelo epidemiológico SIS (véase [7]) divide la población en dos categorías: Susceptibles (S) e infectados (I). A diferencia del modelo SIR, en este modelo no se considera la recuperación permanente de los individuos infectados, ya que si un individuo susceptible, tras un contacto exitoso con un individuo infeccioso, se infecta y se vuelve infeccioso, este no desarrolla inmunidad a la enfermedad y tras la recuperación vuelve a la clase susceptible pudiendo ser reinfectado. La Figura 2.3 indica la relación entre los dos compartimentos.

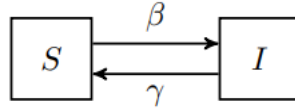


Figura 2.3: Diagrama de flujo del modelo SIS.

Al igual que en el modelo SIR, suponemos que el tamaño de la población es constante $\frac{dN}{dt} = 0$, no consideramos ni nacimientos ni muertes durante ese tiempo. A su vez, las infecciones no pueden transmitirse verticalmente (la infección no puede pasar de la madre al feto).

El modelo SIS puede ser formulado como un sistema de dos ecuaciones diferenciales.

$$S'(t) = -\beta \cdot S(t) \cdot I(t) + \gamma \cdot I(t), \quad (2.6)$$

$$I'(t) = \beta \cdot S(t) \cdot I(t) - \gamma \cdot I(t). \quad (2.7)$$

$$S(0) > 0, I(0) > 0.$$

$S(t)$ y $I(t)$ representan el número de individuos susceptibles e infecciosos en el instante de tiempo t . Del mismo modo que en el modelo SIR, estas ecuaciones vienen definidas por la tasa de contagio, $\beta > 0$ y por la tasa de recuperación, $\gamma > 0$.

Puesto que el tamaño de la población N es fijo, podemos reducir el sistema a una dimensión con la sustitución $S(t) = N - I(t)$, resultando la ecuación diferencial ordinaria siguiente:

$$I'(t) = (\beta \cdot N - \gamma) \cdot I(t) \cdot \left(1 - \frac{I(t)}{N - \frac{\gamma}{\beta}}\right) \quad (2.8)$$

El número reproductivo básico R_0 del modelo SIS está definido con la misma ecuación (2.4) que en el modelo SIR. Sin embargo, el comportamiento en este caso es diferente. Al igual que en el modelo SIR, si $R_0 \leq 1$ la enfermedad se extingue, en cambio, si $R_0 > 1$ la enfermedad permanece en la población como enfermedad endémica.

A partir del número reproductivo básico y haciendo uso de [10] y [18], analizamos el comportamiento cualitativo del modelo *SIS*:

- Si $R_0 < 1$, entonces $\beta N - \gamma < 0$ y $N - \frac{\gamma}{\beta} < 0$, por tanto $I'(t) < 0$ (véase ecuación (3.6)). Esto indica que $I(t)$ es decreciente y $S(t)$ creciente hasta que la población sea enteramente susceptible, teniendo el *punto de equilibrio libre de enfermedad*

$$\lim_{t \rightarrow \infty} (S(t), I(t)) = (N, 0).$$

- Si $R_0 > 1$, entonces $\beta N - \gamma > 0$ y $N - \frac{\gamma}{\beta} > 0$, por tanto $I'(t) > 0$ (véase ecuación (3.6)). $I(t)$ es creciente y $S(t)$ decreciente hasta alcanzar el punto de equilibrio

$$\lim_{t \rightarrow \infty} (S(t), I(t)) = \left(\frac{N}{R_0}, N \cdot \left(1 - \frac{1}{R_0} \right) \right), \quad (2.9)$$

denominado *punto de equilibrio endémico*.

A diferencia del modelo *SIR* que solo posee el *punto de equilibrio libre de enfermedad*, el modelo *SIS* también posee el *punto de equilibrio endémico*. Esto significa que en el modelo *SIR* cuando $t \rightarrow \infty$, la enfermedad acaba desapareciendo ($I(\infty) = 0$), en cambio en el modelo *SIS*, cuando $t \rightarrow \infty$, puede que desaparezca la enfermedad ($I(\infty) = 0$) o puede que esta permanezca endémica (véase (2.9)).

2.4. Revisión de modelos de propagación del Marketing viral

En este subapartado haremos una revisión de los modelos estocásticos de propagación del marketing viral presentes en la literatura científica.

Primero usaremos el artículo [9] para definir lo que son los modelos estocásticos de carácter individual basados en autómatas celulares (cadenas de Markov) y posteriormente procederemos a la revisión de los modelos ya desarrollados en la literatura científica. En este caso, nos hemos centrado en los trabajos realizados por Valerie Isham et al., [15] y Saumik Bhattacharya et al., [4].

Además de los modelos basados en ecuaciones diferenciales descritos en los anteriores apartados, existen otros enfoques que buscan superar algunas de sus limitaciones. Entre ellos, destacan los modelos de carácter individual y naturaleza estocástica, como los basados en autómatas celulares.

Los autómatas celulares son modelos de computación simples pero efectivos para simular sistemas complejos de manera eficiente. Están formados por un número finito de unidades de memoria denominadas células que se encuentran conectadas entre sí siguiendo un determinado patrón, de tal manera que en cada instante de tiempo cada célula está en un estado entre un número finito de ellos. Este estado va cambiando con el paso discreto del tiempo de acuerdo a una regla de transición local cuyas variables son los estados en el instante anterior de la propia célula y sus vecinas. En nuestro caso, haremos uso de las cadenas de Markov finitas que son el método más habitual en este tipo de modelos. Antes de pasar a la formulación de nuestro modelo, procedemos a realizar una revisión sobre los artículos ya mencionados, para así obtener una idea previa sobre este tipo de modelos enfocados en el ámbito del marketing viral y los rumores.

Primero de todo recalcar que los modelos de propagación de marketing viral, a diferencia de los modelos de propagación de enfermedades que intentan comprender los mecanismos por los que se propaga una infección y así intentar tomar medidas para su control y para evitar su aparición o propagación, tienen como objetivo encontrar la forma más eficaz de difundir rápidamente una información o rumor a través de una red informática.

Isham et al., [15] divide la población, en el contexto de rumor o marketing viral, en ignorantes, propagadores y sofocadores, con notación X, Y, Z respectivamente, que corresponden a los susceptibles, infecciosos y recuperados del caso epidémico, también supone que la población es cerrada y de tamaño N . La tasa de infección β , tiene el mismo comportamiento que en los modelos deterministas, consistiendo en ignorantes que se convierten en difusores con un tasa $\beta Y(t)/N$. Sin embargo, la tasa de recuperación, en este caso denominada tasa de sofocamiento, tiene un comportamiento diferente ya que propone dos formas diferentes de pasar del compartimento de los difusores al de los sofocados (el individuo decide de dejar de propagar el rumor). La primera de ellas es si un difusor entra en contacto con otro propagador o sofocador, es decir si entra en contacto con una persona que ya sabe el mensaje o la información. Para esto Isham et al., asume una forma de mezcla homogénea para el contacto, y supone que en tal contacto el difusor es sofocado con una probabilidad p , independiente entre contactos. La segunda forma consiste en que los propagadores se convierten espontáneamente (sin ninguna forma de contacto) en sofocadores a un ritmo constante, δ ; denominado 'olvido'. Así, suponemos que la tasa total a la que un propagador se convierte en un

sofocador es $p\beta\{Y(t) + Z(t)\}/N + \delta$. Con respecto a las formas de sofocamiento se puntualiza que el olvido es un efecto de primer orden, mientras que el contacto es de segundo. Esto se debe a que al principio de un rumor, casi todos los contactos se producen entre ignorantes y propagadores, de modo que el contacto no puede actuar por sí mismo como freno a la propagación del rumor, mientras que el olvido puede dominar la propagación si δ es suficientemente grande.

Al igual que con los modelos epidémicos, las propiedades de interés se centran en las condiciones en las que se propagará un rumor y, si lo hace, el número de los que finalmente lo oirán.

Lo comentado hasta ahora está bajo el supuesto de que la población esta homogéneamente mezclada, en la que los contactos entre todos los pares son igualmente probables, para así poder comparar el modelo pertinente con los modelos deterministas anteriores. Sin embargo, la estructura de contactos que propone Isham et al., viene dada por una red aleatoria. La estructura de esta red es fija y no evoluciona con el tiempo. Las diferencias con el caso de una población homogéneamente mezclada son que un nodo propagador sólo puede transmitir el rumor o ser sofocado por un nodo vecino, según especifique la red, y que ahora los propagadores contactan con cada vecino a una velocidad λ , de modo que la velocidad total a la que los nodos contactan con sus vecinos varía de un nodo a otro. Si todos los nodos tienen el mismo grado (k), entonces la correspondencia con la población de mezcla homogénea es que $\beta = \lambda \cdot k$.

Isham et al., supone que hay n_k nodos de grado k y que de ellos X_k, Y_k, Z_k son respectivamente ignorantes, propagadores y sofocadores (omitiendo la dependencia del tiempo) y denota la función de correlación grado - grado de la red por

$$p_{jk} = P(\text{nodo vecino tiene grado } k \mid \text{nodo índice tiene grado } j).$$

Siendo $\lambda k X_k$, la tasa total a la que los ignorantes de grado k son contactados por sus vecinos, la tasa total de transiciones de (X_k, Y_k) a $(X_k - 1, Y_k + 1)$ se define como

$$\lambda k X_k \sum_j p_{kj} Y_j / n_j.$$

Realizando una aproximación similar con la tasa de sofocamiento, tenemos que la tasa total de transiciones de (X_k, Y_k) a $(X_k, Y_k - 1)$ es

$$Y_k \left\{ \delta + \lambda p k \sum_j p_{kj} [n_j - X_j] / n_j \right\}.$$

Pasamos ahora a la revisión del trabajo de Saumik Bhattacharya et al., [4], en el que se plantea una división de la población por compartimentos mutuamente excluyentes muy similar a la del trabajo de Isham et al., en la que tan solo cambia el nombre de estos, denominándolos: inconscientes, difusores e inertes. La tasa de contacto, en este caso, entre los inconscientes y difusores, tiene el mismo comportamiento, con una probabilidad p . En cambio, la tasa de sofocamiento tiene ciertas diferencias, ya que Bhattacharya et al., no considera el contacto entre un difusor y otro difusor una razón para entrar en la clase inerte. En este caso, solo considera efectos como el olvido, la pérdida de interés o el aburrimiento, razones de peso que llevan a los difusores pasar al compartimento inerte. Estos efectos se combinan en el parámetro σ del modelo.

La principal novedad de este trabajo, es que consideraron posibilidades de que el individuo inerte recuperase su interés en volver a difundir el mensaje o la información. Según ellos, esto es debido a que muchos de los difusores llega un momento en el que se olvidan del mensaje o de la información, por lo que ciertos recordatorios estratégicos pueden influir en ellos para que recuperen de inmediato su estado activo y empiecen a contribuir de nuevo en la propagación de la campaña. Todas estas posibles transiciones del estado inerte al estado difusor fueron incluidas en el modelo permitiendo una retroalimentación de la clase difusora con una tasa de α . En cuanto a la población, es un modelo que considera dinámica poblacional, es decir que introduce nacimientos y fallecimientos. Tanto la tasa de natalidad como la tasa de mortalidad se mantienen iguales a μ , para poder mantener así un tamaño de población fijo. Para una dinámica de marketing viral concreta, el nacimiento y la muerte pueden verse como eventos en los que los individuos se unen o abandonan una plataforma social concreta en la que se desarrolla la campaña.

El esquema del modelo se ilustra en la Figura (2.4), la clase inconsciente denotada por U , la clase difusora por B y la clase inerte por I .

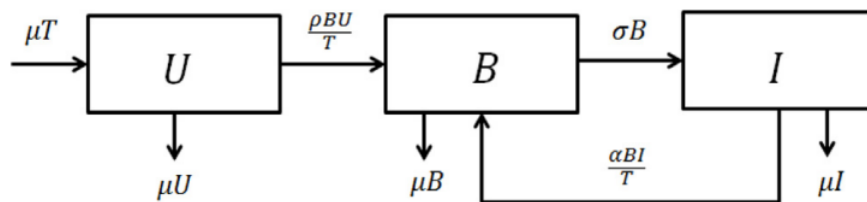


Figura 2.4: Diagrama del modelo que muestra todas las transiciones posibles de un estado a otro.

Capítulo 3

Modelos estocásticos de propagación del marketing viral

En este capítulo, utilizaremos cadenas de Markov de tiempo de discreto para formular los modelos estocásticos *SIR* y *SIS*, enfocándolos en la propagación del marketing viral.

3.1. Cadenas de Markov en tiempo discreto.

En este subapartado hablaremos de los fundamentos básicos de las cadenas de Markov en tiempo discreto basándonos en [11] y [1].

Los modelos epidemiológicos deterministas *SIR* y *SIS* pueden interpretarse mediante procesos estocásticos.

Definición 1. Un **proceso estocástico** es una sucesión de observaciones $x_1, x_2, \dots$ que no se pueden predecir exactamente pero se pueden especificar las probabilidades para los distintos valores posibles en cualquier instante de tiempo.

A diferencia de los modelos deterministas explicados en el capítulo anterior, los modelos estocásticos pueden manejar la aleatoriedad y, por tanto, dar resultados diferentes por cada entrada específica.

Denotamos por E al conjunto de estados en los que puede estar el sistema S en cada instante.

$$E = \{0, 1, 2, 3, \dots, N\} \text{ siendo } N \geq 2.$$

En la modelización de la propagación del marketing viral, el conjunto de estados esta formado por el número de individuos en cada uno de los compartimentos. En el

caso del modelo SIS, el conjunto de estados es el número de individuos en cada una de las dos clases: ignorante y difusora. Para los modelos SIR, el espacio de estados es el número de individuos para cada una de las tres clases (ignorantes, difusores y sofocados).

El conjunto T es el conjunto de tiempos en los que se puede observar el sistema. Un proceso estocástico modela cómo cambia una variable aleatoria con el tiempo. Cuando T es numerable y uniformemente espaciado, el proceso estocástico es un proceso de tiempo discreto.

Los procesos estocásticos en la propagación del marketing o en la transmisión de enfermedades se modelan mejor utilizando cadenas de Markov.

Definición 2. Una *cadena de Markov* es un modelo matemático dinámico y estocástico que permite describir la evolución de un sistema a lo largo del tiempo utilizando la probabilidad, siendo un tipo de proceso estocástico en que la probabilidad de que ocurra un suceso depende solamente del suceso inmediatamente anterior.

Es decir, sólo necesitamos información sobre el estado actual para poder predecir el futuro, y cualquier información adicional sobre el estado pasado no importa. Este hecho se denomina *propiedad de Markov*, también nombrada a veces como la propiedad de ‘falta de memoria’.

Entonces $\{X(t) : t \in T\}$ son un conjunto de variables aleatorias discretas que denotan el estado E de un sistema S en el tiempo $t \in T = 0, \Delta t, 2\Delta t, \dots$

Un proceso estocástico es una cadena de Markov en tiempo discreto (CMTD) si satisface la siguiente ecuación:

$$P[X(t + \Delta t) = j | X(t) = i] = p_{j \leftarrow i}(\Delta t),$$

que indica la probabilidad de que el proceso transite del estado i en el tiempo t al estado j en el tiempo $t + \Delta t$. Satisfaciendo así la *propiedad de Markov* ya que el proceso en cualquier momento $t + \Delta t$ depende sólo del estado del pasado inmediato en el tiempo t .

Estas probabilidades se denominan *probabilidades de transición* y suponemos que no cambian con el tiempo, es decir, el proceso es homogéneo en el tiempo y por tanto independiente de t .

Por ende, una cadena de Markov es un proceso aleatorio homogéneo en el tiempo (las probabilidades de transición no dependen de t), de parámetro discreto y es finito por ser el conjunto de estados de la cadena E , un conjunto finito.

Definición 3. Se llama **matriz de transición**, a la matriz estocástica M , cuyas columnas representan las probabilidades de transición del proceso de un estado dado a cualquiera de los estados de la cadena en una etapa.

Así, el elemento de la columna i de la fila j representa la probabilidad de tránsito del estado i al estado j en cualquier etapa. Dado que los estados pueden tomar valores de 0 a N , M es una matriz $(N + 1) \times (N + 1)$.

$$M = \begin{bmatrix} p_{0 \leftarrow 0} & p_{0 \leftarrow 1} & \cdots & p_{0 \leftarrow N} \\ p_{1 \leftarrow 0} & p_{1 \leftarrow 1} & \cdots & p_{1 \leftarrow N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N \leftarrow 0} & p_{N \leftarrow 1} & \cdots & p_{N \leftarrow N} \end{bmatrix}. \quad (3.1)$$

Al ser una matriz estocástica, las columnas de M forman una distribución de probabilidad y suman 1, denominándose así matriz estocástica izquierda.

3.1.1. Clasificación de estados

Los estados de una cadena de Markov se clasifican en función de sus propiedades. En general, una cadena de Markov puede tener dos tipos de estados: transitorios y recurrentes, aun así existen otros muchos términos a definir como los mencionados en [22].

Definición 4. En una cadena de Markov, un estado j se dice **alcanzable** desde un estado i si $p_{ij} > 0$ para algún $n \geq 0$. Esto significa que empezando desde el estado i , es posible (con probabilidad positiva) entrar en el estado j en un número finito de transiciones. Se denota $i \rightarrow j$.

Definición 5. Se dice que los estados i y j se comunican si cada estado es alcanzable desde el otro. Es decir, $i \rightarrow j$ y $j \rightarrow i$. Se denota por $i \leftrightarrow j$.

Definición 6. A dos estados que se comunican se dice que están en la misma clase. Una cadena de Markov es **irreducible** si todos los estados pertenecen a la misma clase, esto es, todos se comunican entre sí.

Definición 7. Una clase de comunicación C se dice **cerrada** si para cualquier par de estados $i \in C, j \notin C$, se tiene que $p_{ij} = 0$. Es decir, desde cualquier estado dentro de C no se puede alcanzar ningún estado de fuera de C .

Definición 8. Para cualquier estado i en una cadena de Markov, sea f_i la probabilidad de que empezando en el estado i , el proceso vuelva a pasar alguna vez por el estado i . El estado i se dice **recurrente** si $f_i = 1$ y **transitorio** si $f_i < 1$.

Definición 9. Un estado i recurrente se dice **absorbente** si $p_{ii} = 1$, por tanto es imposible abandonarlo.

Una cadena de Markov queda determinada si se conocen los siguiente datos:

1. Su espacio de estados, el conjunto finito $E = 0, 1, 2, \dots, N$ (siendo $N \geq 2$)
2. La matriz estocástica $M = [p_{j \leftarrow i}]_{1 \leq i, j \leq N}$, matriz de probabilidades de transición en una etapa.
3. El vector de probabilidad inicial $P(0) = [p_1(0), p_2(0), \dots, p_N(0)]^t$ siendo $p_i(0) = p(x_0 = i)$ probabilidad de que el proceso se encuentre en el estado i , para $i = 1, 2, \dots, N$, en el instante inicial.

Definición 10. Definimos $p_{j \leftarrow i}^n$ como la probabilidad de que un proceso en el estado i se encuentre en el estado j después de n pasos.

Proposición 1. La matriz de probabilidades de transición de n pasos M^n es $\overbrace{P \cdot P \cdots P}^{n \text{ veces}}$.

Con estos datos se puede deducir la ley de probabilidad para cualquier etapa n . La ley de probabilidad de la etapa n -ésima es el producto de la potencia n -ésima de M por el vector de probabilidades iniciales. Esto significa que conocido el vector de probabilidades iniciales, $P(0)$, la evolución de la cadena sólo depende de la matriz de transición M .

Llamamos $P(n)$ al vector de probabilidad en la etapa n -ésima. Si $n = 1$, entonces: $P(1) = M \cdot P(0)$. Por tanto, la ley de probabilidad de la etapa n -ésima es:

$$P(n) = M^n \cdot P(0).$$

3.1.2. Comportamiento asintótico

Nuestro objetivo es diseñar un modelo estocástico para simular la propagación del marketing viral haciendo uso de cadenas de Markov, es decir, queremos ver como evoluciona la propagación de un mensaje a lo largo del tiempo. Por ello, es importante estudiar el comportamiento de la cadena a medida que el tiempo tiende a infinito,

es decir, analizar lo que sucede con la cadena a largo plazo. Esto se denomina, comportamiento asintótico de una cadena de Markov. Para analizar el comportamiento asintótico, tenemos que estudiar los valores propios de la matriz de transición de una cadena de Markov, es decir, los valores propios de una matriz estocástica, siendo esta, una matriz cuadrada cuyos elementos son no negativos y tal que la suma de los elementos de cada columna es igual a 1.

Las matrices estocásticas presentan algunas particularidades con respecto a los valores propios como [1]:

- Toda matriz estocástica tiene el valor propio $\lambda = 1$.
- Si λ_i es un valor propio de una matriz estocástica, el módulo de λ_i es $|\lambda_i| \leq 1$.

Sabemos que la forma canónica de Jordan asociada a una matriz $M_{r \times r}$ es la matriz J diagonal:

$$J = P^{-1} \cdot M \cdot P = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_r \end{bmatrix}, \quad (3.2)$$

siendo P una matriz regular que tiene por columnas r autovectores linealmente independientes asociados respectivamente a los autovalores λ .

Si no es posible encontrar r autovectores linealmente independientes para una matriz dada $M_{r \times r}$, entonces el teorema de Jordan asegura que siempre se podrá encontrar una matriz regular P tal que

$$J = P^{-1} \cdot M \cdot P = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_r \end{bmatrix} \quad (3.3)$$

siendo las cajas

$$J_i = \begin{bmatrix} \lambda_i & 0 & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & 0 & \cdots & 0 & 0 \\ 0 & 1 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 0 \\ 0 & 0 & 0 & \cdots & 1 & \lambda_i \end{bmatrix}, 1 \leq i \leq r, \quad (3.4)$$

de orden s o inferior (s es el orden de multiplicidad del autovalor λ_i), teniendo en la diagonal principal todos los elementos iguales a λ_i , y la paralela inferior a la diagonal principal formada por unos. En este caso, la forma canónica de Jordan, J , es diagonal por cajas.

Para analizar el comportamiento asintótico de la ley de probabilidad de la etapa n -ésima $P(n)$, es necesario realizar el estudio del $\lim_{n \rightarrow \infty} P(n) = \lim_{n \rightarrow \infty} M^n \cdot P(0)$.

Como $J = P^{-1} \cdot M \cdot P$, será el $\lim_{n \rightarrow \infty} P(n) = \lim_{n \rightarrow \infty} M^n \cdot P(0) = \lim_{n \rightarrow \infty} (P \cdot J^n \cdot P^{-1}) \cdot P(0)$.

El análisis y los resultados de este límite dependen del tipo de matriz de transición de una cadena de Markov que estemos analizando.

Cadenas completamente ergódicas

Cuando la matriz de transición, M , es irreducible y primitiva, entonces $\lambda = 1$ es el único autovalor de módulo máximo. Si los autovalores de M son $\lambda_1, \lambda_2, \dots, \lambda_r$, sólo hay uno de módulo máximo: $\lambda_1 = 1 > |\lambda_2| \geq \dots \geq |\lambda_r|$. Como consecuencia:

$$\lim_{n \rightarrow \infty} J^n = \lim_{n \rightarrow \infty} \begin{bmatrix} 1^n & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_r \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

y

$$\begin{aligned} \lim_{n \rightarrow \infty} M^n &= P \cdot \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \cdot P^{-1} \\ &= \begin{bmatrix} P^1 & P^2 & \dots & P^r \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & \dots & 1 \\ & F_2 & & \\ & \vdots & & \\ & F_r & & \end{bmatrix} \\ &= \begin{bmatrix} P^1 & P^2 & \dots & P^r \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} P^1 & P^1 & \dots & P^1 \end{bmatrix} \end{aligned}$$

es una matriz con todas las columnas iguales al vector P^1 .

Por tanto, el $\lim_{n \rightarrow \infty} P(n)$ es un vector de probabilidad P^1 con todas las componentes estrictamente positivas, que además es autovector asociado al autovalor 1, es decir, el $\lim_{n \rightarrow \infty} P(n)$ es un vector P^1 de probabilidades estacionario, siendo además un vector de estado permanente. El vector P^1 representa las probabilidades de que el sistema esté en cada uno de los estados a la larga. La probabilidad de encontrarse en un estado después de un número n suficientemente grande de etapas es independiente de su estado inicial $P(0)$, ya que:

$$\lim_{n \rightarrow \infty} P(n) = \lim_{n \rightarrow \infty} M^n \cdot P(0) = \begin{bmatrix} P^1 & P^1 & \dots & P^1 \end{bmatrix} \cdot \begin{bmatrix} p_1(0) \\ p_2(0) \\ \dots \\ p_r(0) \end{bmatrix} = P^1$$

esto significa que la cadena consta de una sola clase final ergódica y no posee estados transitorios.

Cadenas simplemente ergódicas

Si la matriz M de transición de la cadena es reducible pero tiene un solo autovalor de módulo máximo, $\lambda = 1$, el autovector P^1 , asociado a este autovalor, tiene alguna componente nula. Esto indica que la cadena consta de una clase final de estados ergódicos y de t clases de estados transitorios, siendo $1 \leq t \leq m$ componentes nulas de P^1 .

Por tanto, al igual que en el caso anterior existe un vector de estado permanente que coincide con la única distribución estacionaria P^1 de la cadena, que a su vez cuando el número de etapas tiende a infinito no depende del vector de estado inicial $P(0)$. Sin embargo, este vector de estado permanente contiene componentes iguales a cero, las correspondientes a los estados transitorios porque en el momento en que se deja un estado transitorio ya no se puede volver a él.

Cadenas periódicas

Si $M_{r \times r}$ es una matriz irreducible pero no primitiva, tiene $m > 1$ autovalores de módulo 1, estos son las m raíces enteras de la unidad, siendo m divisor del número de autovalores distintos de cero de la matriz M .

Si $M_{r \times r}$ es una matriz de una cadena de Markov que es no singular e irreducible de orden r primo, sólo puede tener uno o r autovalores de módulo 1, porque r tiene que

ser divisor del número de autovalores distintos de cero de la matriz M que, por ser no singular, es r .

En estos dos casos, la matriz de transición de la cadena tiene el autovalor 1 simple pero posee otros m autovalores de módulo 1. Por tanto, la cadena es periódica de periodo m , es decir, la cadena tiene una distribución estacionaria que viene dada por el autovector P^1 asociado al autovalor dominante, pero no es una distribución de estado permanente debido a que no existe el $\lim_{n \rightarrow \infty} P(n) = \lim_{n \rightarrow \infty} M^n \cdot P(0)$ por no existir el $\lim_{n \rightarrow \infty} M^n$. La evolución a la larga depende de las condiciones iniciales.

Cadenas múltiples

Si la matriz de transición M de una cadena de Markov es una matriz reducible con $k > 1$ autovalores de módulo máximo, entonces contiene el autovalor dominante $\lambda = 1$ múltiple de orden de multiplicidad k , la cadena es múltiple y consta de k clases finales, y puede tener o no alguna clase de estados transitorios, pero a la larga quedará atrapada en una de las clases finales. La evolución en el interior de una de estas clases finales se puede hacer utilizando la submatriz de la matriz de transición correspondiente a los estados de esa clase final. A esta submatriz le corresponde el autovalor dominante $\lambda = 1$ simple.

Una clase final será ergódica si $\lambda = 1$ es el único autovalor de módulo 1 de la submatriz correspondiente y será periódica de periodo m si existen $m > 1$ autovalores de módulo 1. Si una de las clases finales consta de un solo elemento, esa clase es absorbente.

La evolución a la larga de la cadena depende de la distribución de probabilidad inicial. Una vez que la cadena entra en una de las clases finales ya es imposible que salga de ella.

3.2. Modelos de propagación del marketing viral

Para el desarrollo de nuestro modelo hemos decidido compartimentar la población en tres compartimentos al igual que los modelos mencionados durante el trabajo. En este caso, hemos asociado a cada individuo a uno de los tres estados mutuamente excluyentes: Ignorante (I), Difusor (D) e Inerte (S). A diferencia de los modelos de propagación de enfermedades, en estos modelos tenemos que tener en cuenta diferentes factores para comprender la dinámica de la población y entender la psicología del

consumidor hacia las campañas de marketing viral. Entre estos factores destacan la cuantía de las recompensas, las tendencias recientes y la asociación con una marca, razones que motivan a la gente a participar activamente en una campaña viral. Aunque hay otros factores, como la seguridad, el olvido o experiencias pasadas adversas, que provocan que la gente no participe en la campaña, después de haber conocido la oferta.

El esquema del modelo consta de la clase ignorante, denotada por I , formada por aquellos individuos que todavía no han recibido el mensaje. Estos serían el equivalente a la clase susceptible, ya que, pueden recibir el mensaje publicitario con ofertas de marketing y decidir participar en la campaña. La clase difusora D está formada por individuos que conocen el mensaje y tienen el potencial de enviarlo a la población, pudiendo participar en la campaña, difundiendo y transmitiendo el mensaje a toda la población recomendándolo a través de sus contactos sociales. Por último, está la clase inerte S , formada por aquellos individuos que pertenecían a la clase de difusores, pero que no comparte el mensaje en la actualidad. Hemos considerado dos formas principales de entrar en la clase inerte: Primero aquella en que las personas que solían difundir el mensaje viral, pierden interés en seguir participando en la propagación y por tanto, pasan a la clase inerte. Esta situación puede ser debida a la baja relación beneficio-esfuerzo, al aburrimiento o el sentirse dudosos con la seguridad. Por otro lado, están los individuos que conociendo el mensaje tenían la opción de compartir el mensaje pero que por diversas razones como el olvido, la distracción o el desinteresamiento, deciden no compartir el mensaje ni una sola vez.

Suponemos que β es la tasa a la que un difusor entra en contacto con un individuo de la clase inconsciente, y comparte el mensaje viral, creando un nuevo propagador potencial y la tasa γ es la combinación de los efectos que provocan el paso de la clase difusora a la clase sofocada. Para nuestro modelo solo hemos considerado los factores mencionados para el paso de difusor a inerte, y no hemos tenido en cuenta el contacto entre dos personas conocedoras del mensaje, ya que como mencionó Isham et al. en su trabajo [15], esta cuestión se considera de segundo orden. También mencionar que una vez el individuo pertenece a la clase inerte, este no puede volver a propagar el mensaje. Así pues, no hemos considerado la posibilidad de que los individuos inertes vuelvan a difundir el mensaje o la información volviendo al compartimento de los difusores, cuestión mencionada por Bhattacharya et al., en [4].

Una vez definido el modelo estocástico de propagación del marketing viral que vamos a llevar a cabo, procedemos a desarrollar los modelos SIR y SIS del mismo.

3.3. Modelos estocásticos SIS

Para formular los modelos epidémicos SIS estocásticos utilizando cadenas de Markov de tiempo discreto, seguimos el enfoque de Allen [2] y de Ige [14], pero particularizándolo a la propagación del marketing viral.

Sean $I(t)$ y $D(t)$ variables aleatorias discretas que cuentan el número de personas ignorantes y difusoras en cada momento $t \in T = \{0, \Delta t, 2\Delta t, \dots\}$ con $I(t), D(t) \in \{0, 1, 2, \dots, N\}$.

Suponemos que la población es constante durante todo el tiempo observado, es decir, en cualquier momento t , $N = I(t) + D(t)$, por tanto

$$\frac{dN}{dt} = \frac{dI(t)}{dt} + \frac{dD(t)}{dt} = 0.$$

Como ya vimos en el modelo determinista, podemos reducir el modelo SIS a una sola ecuación ya que al sólo haber dos compartimentos bastaría con utilizar el valor de la clase difusora para estimar el valor de la clase inconsciente. Por lo que nos quedaría el proceso estocástico $\{D(t)\}_{t=0}^{\infty}$, con tan solo la variable aleatoria del número de difusores y con la función de probabilidad asociada,

$$p_d(t) = \text{Prob}\{D(t) = d\}$$

donde

$$\sum_{d=0}^N p_d(t) = 1.$$

Para reducir el número de transiciones en el tiempo Δt , suponemos que el paso del tiempo se elige lo suficientemente pequeño como para que el número de individuos propagadores cambie como máximo en uno durante el intervalo de tiempo Δt . Es decir, para $D(t) = d$ y un paso de tiempo de Δt , sólo puede ocurrir una de las siguientes situaciones:

$$d \rightarrow d + 1, \quad d \rightarrow d - 1, \quad d \rightarrow d.$$

Por cada cambio de tiempo, una persona puede pasar a ser difusora del mensaje, puede sofocarse y decidir dejar de participar en la propagación y pasar a ser ignorante o puede que no se produzca ningún cambio en el número de personas difusoras en la población.

Para un modelo de propagación de marketing viral SIS simple, consideramos sólo la estructura de contacto entre las poblaciones ignorantes y difusoras para la transmisión del mensaje. Definimos $\beta > 0$ y $\gamma > 0$ como las tasas de compartir el mensaje y sofocamiento, respectivamente. Suponemos que existe una mezcla homogénea de las

dos poblaciones y que la estructura de contactos depende de la frecuencia. El hecho de que la estructura de contactos depende de la frecuencia implica que la tasa de compartir el mensaje β no cambia con el tamaño de la población. Esto significa que β permanece constante incluso cuando el número de personas difusoras aumenta. Por lo tanto, el cambio de la clase ignorante en el paso temporal t viene dado por:

$$\frac{dI(t)}{dt} = -\frac{\beta \cdot I(t) \cdot D(t)}{N} + \gamma D(t).$$

La probabilidad de pasar de d a $d + 1$ es

$$p_{d+1 \leftarrow d}(\Delta t) = \frac{\beta d(N-d)}{N} \Delta t,$$

donde $i = N - d$, debido al tamaño fijo de la población. El número de individuos que se sofocan en cualquier momento t viene dado por $\gamma D(t)$. La probabilidad de pasar de d a $d - 1$ es

$$p_{d-1 \leftarrow d}(\Delta t) = \gamma d \Delta t.$$

La suma de las probabilidades de todas las transiciones posibles tiene que dar 1. Por lo tanto, la probabilidad de que el número de difusores permanezca invariable después de un paso de tiempo es

$$p_{d \leftarrow d}(\Delta t) = 1 - \left[\frac{\beta d(N-d)}{N} + \gamma d \right] \Delta t.$$

Las probabilidades de transición del modelo de propagación de marketing viral CMTD SIS vienen dadas por

$$p_{j \leftarrow d}(\Delta t) = \begin{cases} \frac{\beta d(N-d)}{N} \Delta t, & \text{si } j = d + 1 \\ \gamma d \Delta t, & \text{si } j = d - 1 \\ 1 - \left[\frac{\beta d(N-d)}{N} + \gamma d \right] \Delta t, & \text{si } j = d \\ 0, & \text{en otro caso.} \end{cases}$$

Para simplificar el modelo denotamos las probabilidades de transición del nacimiento de un nuevo difusor y el sofocamiento de un difusor como $b(d)\Delta t$ y $g(d)\Delta t$, respectivamente. Por lo tanto, el modelo SIS de propagación de marketing viral puede representarse

como

$$p_{j \leftarrow i}(\Delta t) = \begin{cases} b(d)\Delta t, & \text{si } j = d + 1 \\ g(d)\Delta t, & \text{si } j = d - 1 \\ 1 - [b(d) + g(d)]\Delta t, & \text{si } j = d \\ 0, & \text{en otro caso.} \end{cases}$$

Para garantizar que estas probabilidades de transición de encuentran en el intervalo $[0, 1]$, el paso del tiempo Δt debe elegirse lo suficientemente pequeño como para que

$$\max_{d \in \{1, 2, \dots, N\}} \{[b(d) + g(d)]\Delta t\} \leq 1.$$

Formamos la matriz de transición M , con los estados ordenados de 0 a N . Por ejemplo, el elemento (1,2) de la matriz de transición es la probabilidad de transición del estado 2 al estado 1, es decir, sería la probabilidad de que uno de los dos difusores presentes en la población se sofoque y pase a ser ignorante, quedando tan solo un difusor en la población.

Utilizando las probabilidades de transición, la matriz M , viene dada como sigue:

$$\begin{bmatrix} 1 & g(1)\Delta t & 0 & \dots & 0 & 0 \\ 0 & 1 - [b(1) + g(1)]\Delta t & g(2)\Delta t & \dots & 0 & 0 \\ 0 & b(1)\Delta t & 1 - [b(2) + g(2)]\Delta t & \dots & 0 & 0 \\ 0 & 0 & b(2)\Delta t & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & g(N-1)\Delta t & 0 \\ 0 & 0 & 0 & \dots & 1 - [b(N-1) + g(N-1)]\Delta t & g(N)\Delta t \\ 0 & 0 & 0 & \dots & b(N-1)\Delta t & 1 - g(N)\Delta t \end{bmatrix} \quad (3.5)$$

El proceso de propagación de marketing viral CMTD SIS $\{D(T)\}_{t=0}^{\infty}$ estaría completamente formulado, por lo que dado un vector de probabilidades iniciales $p(0)$, podríamos obtener la ley de probabilidad para cualquier etapa ya que

$$p(t + \Delta t) = M^{n+1} \cdot p(0), \quad (3.6)$$

donde $t = n \cdot \Delta t$.

3.3.1. Comportamiento asintótico

Para estudiar como evoluciona el modelo con el tiempo, analizamos la matriz de transición M (3.5) de la cadena de Markov SIS. Los estados de la matriz se clasifican según su conectividad en un gráfico dirigido. El gráfico correspondiente a la cadena de Markov SIS se ilustra en la Figura 3.1, donde $d = 0, 1, \dots, N$ son los estados difusores.

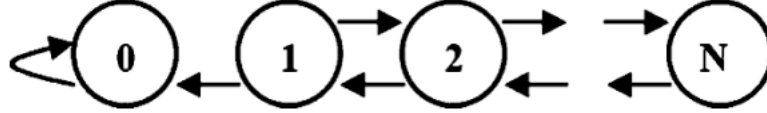


Figura 3.1: Diagrama del modelo de propagación SIS estocástico

Los estados $0, 1, \dots, N$ pueden dividirse en dos conjuntos formados por el estado recurrente, $\{0\}$, y los estados transitorios $\{1, \dots, N\}$. El estado $d = 0$ es un estado absorbente, ya que es imposible la transición de $d = 0$ a otro estado $d > 0$, debido a que no es posible que se propague el mensaje en una población sin ningún difusor, por tanto el conjunto 0 está cerrado. En la clase con estados $d > 0$, la probabilidad de transición entre dos estados cualesquiera es positiva, por lo tanto, son estados transitorios, ya que es posible que vuelvan al cabo de un tiempo, pero no está garantizado. Aun así, el estado conjunto no es cerrado porque $p_{0,1}(\Delta t) > 0$.

Para los estados transitorios se puede demostrar que los elementos de la matriz de transición tienen la siguiente propiedad: Sea $M^n = p_{i,j}^{(n)}$, donde $p_{i,j}^{(n)}$ es el elemento (i, j) de la n -ésima potencia de la matriz de transición, M^n , entonces

$$\lim_{n \rightarrow \infty} p_{i,j}^{(n)} = 0$$

para cualquier estado j y cualquier estado transitorio i . Por tanto, el $\lim_{n \rightarrow \infty} M^n$, nos resultará en una matriz estocástica formada tan solo por el estado absorbente.

$$\lim_{n \rightarrow \infty} M^n = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

De la ecuación 3.6 y de la teoría de las cadenas de Markov se deduce que

$$\lim_{t \rightarrow \infty} p(t) = (1, 0, \dots, 0)^T,$$

donde $t = n\Delta t$.

Este resultado implica que la población en el modelo de propagación de marketing viral *CMTD SIS* se aproxima a un punto de equilibrio sin difusores (la probabilidad de absorción es uno) independientemente de la magnitud del número básico. A medida que el tiempo avance, el mensaje que se está propagando perderá fuerza y los individuos irán dejando de participar en la propagación y por tanto, pasarán de ser difusores a ser ignorantes hasta el punto en que la población no tenga ningún difusor, lo que implica que el mensaje no se propague más. Dado que este resultado es asintótico, la tasa de convergencia al equilibrio libre de difusores puede ser muy lenta. El tiempo medio hasta que se alcanza este punto de equilibrio (absorción) depende de las condiciones iniciales y de los valores de los parámetros, pero puede ser extremadamente largo.

3.4. Modelos estocásticos SIR

Para la formulación de los modelos SIR de propagación viral, seguimos usando el enfoque de Allen [2] y de O.Ige [14].

Para los modelos SIR, hay tres compartimentos y se requiere de las variables aleatorias $I(t)$, $D(t)$ y $S(t)$. Como en el modelo SIS anterior, $I(t)$, $D(t)$ cuentan el número de individuos ignorantes y difusores, en el tiempo $t \in T = \{0, \Delta t, 2\Delta t, \dots\}$ con $I(t), P(t) \in \{0, 1, 2, \dots, N\}$. La variable aleatoria adicional $S(t)$ cuenta el número de individuos inertes en la población en el momento $t \in T$, también con $S(t) \in \{0, 1, 2, \dots, N\}$. Los inertes son aquellos individuos que anteriormente eran difusores, pero que por diversos factores deciden dejan de difundir el mensaje. Al igual que en el modelo SIS, seguimos con la hipótesis de que la población permanece constante en cualquier momento t .

$$\frac{dN}{dt} = \frac{dI(t)}{dt} + \frac{dD(t)}{dt} + \frac{dS(t)}{dt} = 0.$$

Para reducir el sistema y así facilitar la interpretación. Utilizamos las clases ignorante y difusora para calcular la clase inerte, $S(t) = N - I(t) - D(t)$. Por tanto, nos queda el proceso bivalente $\{I(t), D(t)\}_{t=0}^{\infty}$, con la función de probabilidad conjunta

$$p_{(i,d)}(t) = \text{Prob}\{I(t) = i, D(t) = d\}.$$

Posteriormente, elegimos un paso de tiempo lo suficientemente pequeño, para poder suponer que en cada paso de tiempo ocurre como máximo una transición. Para un paso de tiempo Δt , solo puede ocurrir:

$$(i, d) \leftarrow (i - 1, d + 1), (i, d) \leftarrow (i, d - 1), (i, d) \leftarrow (i, d).$$

Es posible que un individuo de la población pase a ser difusor (se reduce en uno el número de ignorantes y aumenta en uno el número de difusores), que un difusor no quiera propagar más el mensaje y pase a ser inerte (se reduce en uno el número de difusores) o que no se produzca ningún cambio.

Como ya vimos en el modelo SIR determinista, el número de ignorantes (susceptibles) ira disminuyendo con el tiempo. Por lo que la probabilidad de que aparezca un nuevo difusor es

$$p_{(i-1,d+1) \leftarrow (i,d)}(\Delta t) = \frac{\beta \cdot i \cdot d}{N} \Delta t.$$

Los individuos del compartimento difusor también transitan constantemente al compartimento inerte. La probabilidad de pasar a ser inerte es

$$p_{(i,d-1) \leftarrow (i,d)}(\Delta t) = \gamma \cdot d \Delta t.$$

Dado que la población es constante, la probabilidad de que el número de difusores permanezca sin cambio es

$$p_{(i,d) \leftarrow (i,d)}(\Delta t) = 1 - \left[\frac{\beta i d}{N} + \gamma d \right] \Delta t.$$

Por tanto, las probabilidades de transición del modelo SIR de propagación viral son

$$p_{(i+k,d+j) \leftarrow (i,d)}(\Delta t) = \begin{cases} \frac{\beta i d}{N} \Delta t, & \text{si } (k, j) = (-1, 1) \\ \gamma d \Delta t, & \text{si } (k, j) = (0, -1) \\ 1 - \left[\frac{\beta i d}{N} + \gamma d \right] \Delta t, & \text{si } (k, j) = (0, 0) \\ 0, & \text{en otro caso.} \end{cases}$$

Para simplificar el modelo, usamos el mismo método de simplificación que en el modelo SIS.

$$p_{(i+k,d+j) \leftarrow (i,d)}(\Delta t) = \begin{cases} b(d) \Delta t, & \text{si } (k, j) = (-1, 1) \\ g(d) \Delta t, & \text{si } (k, j) = (0, -1) \\ 1 - [b(d) + g(d)] \Delta t, & \text{si } (k, j) = (0, 0) \\ 0, & \text{en otro caso.} \end{cases}$$

Como ahora los estados son pares ordenados, la matriz de transición es más compleja y su forma depende de cómo estén ordenados los estados (i, d) .

3.4.1. Comportamiento asintótico

Al igual que en el modelo SIS, analizamos la matriz de transición de la cadena de Markov SIR para ver como evoluciona el modelo a medida que pasa el tiempo. En este caso el gráfico dirigido asociado a la cadena se encuentra en una red bidimensional. Es fácil demostrar que el estado $(N, 0)$ es absorbente ($p_{(N,0),(N,0)}(\Delta t) = 1$) y que todos los demás estados son transitorios. Así, asintóticamente, todas las trayectorias de la muestra acaban siendo absorbidas por el estado libre de difusores $(N, 0)$, es decir que a medida que avanza el tiempo, los difusores dejan de compartir el mensaje hasta el punto en el que no queda ningún difusor y por tanto el mensaje deja de propagarse, tal y como pasaba con el modelo SIS.

3.5. Pseudocódigo y ejemplo numérico

En esta sección, procedemos a explicar el pseudocódigo que vamos a utilizar para posteriormente realizar un ejemplo numérico de los dos modelos estocásticos de propagación del marketing viral utilizando el enfoque de O.Ige [14].

3.5.1. Modelos CMTD SIS

Siempre que se cumplan todos los parámetros y condiciones iniciales propuestas, podemos realizar un gráfico del modelo de propagación viral *SIS* comparando los cambios en el número de individuos difusores frente al tiempo.

El algoritmo del código utilizado es el siguiente:

1. Inicializar los valores de β, γ , el tamaño de la población N , el paso del tiempo Δt y la duración del brote t_{end} .
2. Crear las matrices $I(t)$ y $D(t)$. Establezca $t = 0$. Sea $D(0)$ el número inicial de individuos difusores y $I(0) = N - D(0)$.
3. Para t desde 0 a t_{end} en un incremento de Δt , mientras t sea menor que t_{end} :
 - a) Calcule la probabilidad de aparición de un nuevo difusor (p_1) utilizando $(\frac{\beta I(t) D(t)}{N}) \Delta t$.

- b) Calcule la probabilidad de retorno al compartimento ignorante (p_2) utilizando $\gamma D(t)\Delta t$.
 - c) Calcule la probabilidad de que no cambie el número de difusores utilizando $1 - (p_1 + p_2)$.
 - d) Seleccione un número aleatorio u de una distribución uniforme de $(0,1)$.
 - e) Compare u con p_1, p_2 y $1 - (p_1 + p_2)$.
 - f) Si $0 < u \leq p_1$, disminuye el número de individuos ignorantes en 1 y aumenta el número de difusores en 1. Por tanto, $I(t + \Delta t) = I(t) - 1$ y $D(t + \Delta t) = D(t) + 1$.
 - g) Si $p_1 < u \leq p_1 + p_2$, aumenta en 1 el número de individuos ignorantes y disminuye en 1 el número de difusores. Por tanto, $I(t + \Delta t) = I(t) + 1$ y $D(t + \Delta t) = D(t) - 1$.
 - h) En caso contrario, $p_1 + p_2 < u < 1$. En este caso, el número de ignorantes y difusores sigue siendo el mismo. Por tanto, $I(t + \Delta t) = I(t)$ y $D(t + \Delta t) = D(t)$.
4. Traza la gráfica de los tiempos t frente al número de difusores en cada tiempo, $D(t)$.

En el algoritmo, simulamos una variable aleatoria discreta u para determinar cual de los sucesos se ha producido. En general, para n sucesos, dividimos el intervalo $(0,1)$ en n partes, según la probabilidad de cada uno de los sucesos individuales $p_1, p_2, \dots, p_n$. La variable aleatoria u tiene la misma distribución que p_i .

Procedemos a la simulación de un ejemplo de modelo *CMTD SIS* de propagación viral. Supongamos que se intentamos propagar una campaña publicitaria en una comunidad formada por 100 personas, de las cuales 5 son personas difusoras, es decir, personas conocedoras del mensaje o información capaces de propagarlo.

La tasa de contacto entre un individuo ignorante y difusor β es 1, la tasa de paso de difusor a ignorante es de 0,5, y la fracción de tiempo entre eventos consecutivos es de 0,05. Para tres simulaciones, la gráfica del tiempo frente al número de difusores esta representada en la Figura 3.2, siguiendo el pseudocódigo presentado en esta sección.

En este ejemplo, el número de reproducción básico R_0 es 2. En dos de las tres simulaciones de propagación viral (trayectorias verde y roja), la población alcanza un equilibrio endémico con alrededor de unas 50 personas conocedoras del mensaje. Es

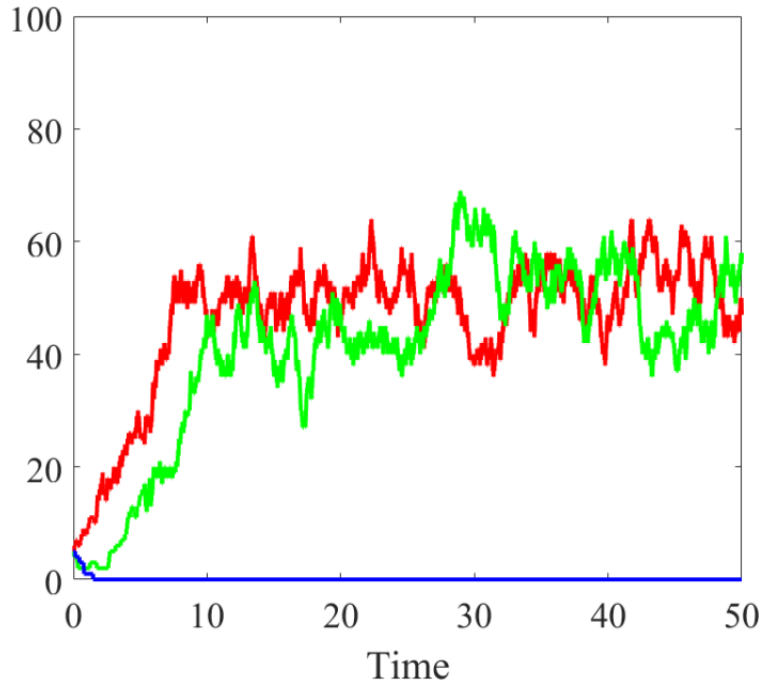


Figura 3.2: Simulaciones de un modelo de propagación CMTD SIS con tres trayectorias estocásticas usando los parámetros $N = 100$, $I(0) = 5$, $\beta = 1$, $\gamma = 0.5$ y $\Delta t = 0.01$.

decir, que en estas dos simulaciones, alrededor de la mitad de la población le llegó el mensaje y formó parte de la propagación. Por tanto, estas dos campañas publicitarias tuvieron éxito. Sin embargo, la trayectoria azul muestra que la propagación se extingue muy rápidamente, fracasando así la campaña publicitaria. Aún teniendo $R_0 > 1$, el número de difusores decreció muy rápidamente. Esto es debido a la naturaleza estocástica del modelo.

3.5.2. Modelos CMTD SIR

El algoritmo de un código de modelo CMTD SIR es el siguiente:

1. Inicializar los valores de β, γ , el tamaño de población N , el paso de tiempo Δt y la duración del brote t_{end} .
2. Crear las matrices $I(t)$, $D(t)$ y $S(t)$. Establezca $t = 0$ y $S(0) = 0$. Sea $I(0)$ el número inicial de individuos difusores y $I(0) = N - D(0)$.
3. Para t desde 0 a t_{end} en un incremento de Δt , mientras t sea menor que t_{end} :
 - a) Calcule la probabilidad de aparición de un nuevo difusor (p_1) utilizando $(\frac{\beta I(t)D(t)}{N})\Delta t$.

- b) Calcule la probabilidad de retorno al compartimento ignorante (p_2) utilizando $\gamma D(t)\Delta t$.
- c) Calcule la probabilidad de que no cambie el número de difusores utilizando $1 - (p_1 + p_2)$.
- d) Seleccione un número aleatorio u de una distribución uniforme de $(0,1)$.
- e) Compare u con p_1, p_2 y $1 - (p_1 + p_2)$.
- f) Si $0 < u \leq p_1$, disminuye el número de individuos ignorantes en 1 y aumenta el número de difusores en 1. Por tanto, tenemos

$$I(t + \Delta t) = I(t) - 1, \quad D(t + \Delta t) = D(t) + 1 \quad \text{y} \quad S(t + \Delta t) = S(t).$$

- g) Si $p_1 < u \leq p_1 + p_2$, disminuye en 1 el número de individuos difusores y aumenta en 1 el número de inertes. Por tanto tenemos

$$I(t + \Delta t) = I(t), \quad D(t + \Delta t) = D(t) - 1 \quad \text{y} \quad S(t + \Delta t) = S(t) + 1.$$

- h) En caso contrario, $p_1 + p_2 < u < 1$. En este caso, el número de ignorantes, difusores e inertes sigue siendo el mismo. Por tanto,

$$I(t + \Delta t) = I(t), \quad D(t + \Delta t) = D(t) \quad \text{y} \quad S(t + \Delta t) = S(t).$$

- 4. Represente gráficamente los tiempos t frente al número de difusores en cada tiempo, $D(t)$.

Procedemos a simular un ejemplo de modelo de propagación CMTD SIR utilizando el algoritmo que acabamos de describir.

Vamos a suponer los mismos parámetros que en el ejemplo del modelo SIS anterior. Una población de 100 personas, de las cuales 5 de ellas son individuos difusores inicialmente. La tasa de contacto entre ignorantes y difusores β es de 1, la tasa de sofocamiento γ es de 0.5 y el tiempo entre eventos consecutivos es 0.01. Para tres simulaciones, el gráfico del tiempo frente al número de difusores se muestra en la Figura 3.3.

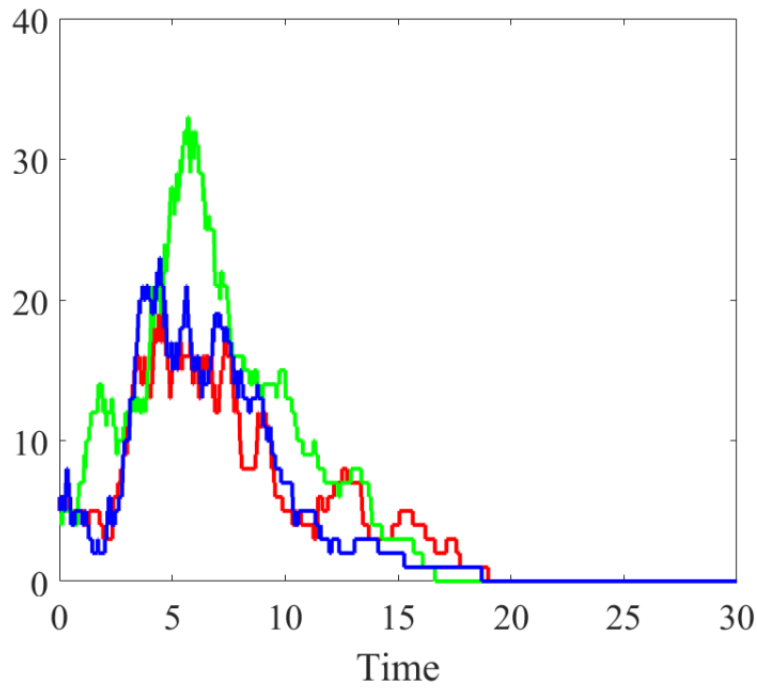


Figura 3.3: Simulaciones de un modelo de propagación CMTD SIR con tres trayectorias estocásticas usando los parámetros $N = 100$, $I(0) = 5$, $\beta = 1$, $\gamma = 0.5$ y $\Delta t = 0.01$.

Observamos que la trayectoria de las tres simulaciones muestra que, tras alcanzar un número máximo de difusores, los individuos se van convirtiendo en inertes gradualmente, hasta el punto de no quedar difusores. El pico de difusores en la simulación mas alta (verde), superó el valor 30, pero se quedo por debajo del número de difusores obtenido por el modelo *SIS*.

Capítulo 4

Conclusiones

El trabajo de Fin de Grado desarrollado ha consistido en el diseño de un modelo estocástico de naturaleza individual y basado en cadenas de Markov en tiempo de discreto, para simular la propagación del marketing viral en una red social. Para alcanzar este objetivo hemos realizado una revisión bibliográfica, primero de los modelos compartimentales deterministas SIS y SIR de Kermack-McKendrick y posteriormente sobre los modelos estocásticos de propagación del marketing viral propuestos en la literatura científica.

Finalmente, hemos presentado el diseño de dos modelos distintos basados en cadenas de Markov de tiempo discreto. El primer modelo sigue un comportamiento tipo SIS, mientras que el segundo modelo sigue un comportamiento tipo SIR. Al realizar un ejemplo numérico con ambos modelos, hemos observado diferentes patrones en la propagación del marketing viral en la población. En el modelo tipo SIS, de las tres simulaciones obtenidas, en dos ellas el mensaje permanecía endémico en la población. Esto indica que, un vez el mensaje se hace viral, este permanece en la población durante mucho tiempo. Como hemos observado en el comportamiento asintótico de este tipo de modelos, la población tiende a converger hacia un punto de equilibrio libre de difusores. Sin embargo, en algunos casos, este proceso puede ser extremadamente largo, como sucede en nuestro caso. Esto contrasta con lo obtenido en el modelo tipo SIR, en el que introducimos el compartimento formado por individuos inertes. En este modelo, una vez que el mensaje se vuelve viral, experimenta un rápido crecimiento hasta alcanzar un punto máximo. A partir de ese punto, el número de difusores comienza a disminuir rápidamente hasta que desaparecen por completo, llegando al punto de equilibrio sin difusores.

Por tanto, hemos diseñado dos modelos que muestran dos tipos de comportamientos diferentes en cuanto al proceso de difusión del marketing viral. Un primer modelo en el que el una vez se hace viral el mensaje, este se mantiene presente en la población durante bastante tiempo y un segundo modelo en el que el mensaje se hace viral rápidamente pero este llega un punto en el que se satura y desaparece de la población. En comparación con los modelos deterministas, podemos observar que los modelos estocásticos presentados siempre tienden hacia una población libre de difusores. En cambio, el modelo determinista SIS, contiene un punto de equilibrio en el que el mensaje puede permanecer endémico. En el ámbito de la difusión de mensajes y el marketing viral, los modelos estocásticos se asemejan más a la realidad que los modelos deterministas. Si se ejecuta una campaña publicitaria exitosa o si un mensaje se recibe favorablemente por la población, es probable que permanezca en esa población por un tiempo prolongado. Sin embargo, en la era de las redes sociales y la sobreabundancia de información a la que estamos expuestos, es común que dicho mensaje termine siendo olvidado o saturado.

Para finalizar, cabe destacar que los modelos epidemiológicos referidos a la propagación del marketing viral carecen de una definición clara acerca de su dinámica y comportamiento, ya que no existe un consenso en la literatura con respecto a su definición. Esto contrasta con los modelos clásicos referidos a las enfermedades infecciosas, en los que dependiendo de la enfermedad a tratar se considera un tipo de modelo compartimental u otro.

Bibliografía

- [1] Eme Aka. Cadenas de Markov Procesos Estocásticos. In *Procesos Estocásticos*, pages 103–127.
- [2] Linda J. S. Allen. An Introduction to Stochastic Epidemic Models. In J. M. Morel, F. Takens, B. Teissier, Fred Brauer, Pauline Van Den Driessche, and Jianhong Wu, editors, *Mathematical Epidemiology*, volume 1945, pages 81–130. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. Series Title: Lecture Notes in Mathematics.
- [3] Kristen Baker. The Ultimate Guide to Viral Campaigns.
- [4] Saumik Bhattacharya, Kumar Gaurav, and Sayantari Ghosh. Viral marketing on social networks: An epidemiological perspective. *Physica A: Statistical Mechanics and its Applications*, 525:478–490, July 2019.
- [5] Kevin Burke. As consumer attitudes shift, so must marketing strategies. In *Proc. 12th ACM Conf. Electron. Commerce (EC)*, pages 157–166, 2003.
- [6] José Antonio Camúñez Ruiz, Jesús Basulto Santos, and Francisco Javier Ortega Irizo. La memoria de Daniel Bernoulli sobre la inoculación contra la viruela (1760): Un problema de decisión bajo incertidumbre. In *La memoria de Daniel Bernoulli sobre la inoculación contra la viruela (1760): Un problema de decisión bajo incertidumbre*, pages 47–60. Universidad de Huelva, 2007.
- [7] Carlos Castillo-Chávez and Abdul-Aziz Yakubu. *Intraspecific Competition, Dispersal and Disease Dynamics in Discrete-Time Patchy Environments*, page 22 p. 01 2000.
- [8] Jonathan Frenzen and Kent Nakamoto. Structure, Cooperation, and the Flow of Market Information. *Journal of Consumer Research*, 20(3):360–375, December 1993.

- [9] María José Fresnadillo Martínez, Enrique García Sánchez, Enrique García Merino, Ángel María Martín del Rey, and José Elías García Sánchez. Modelización matemática de la propagación de enfermedades infecciosas: de dónde venimos y hacia dónde vamos. *Revista Española de Quimioterapia*, 26(2):81–91, 2013.
- [10] Laura García Rovira. Modelos matemáticos compartimentales en epidemiología.
- [11] María Teresa González Manteiga. Cadenas de Markov finitas. In *Modelos matemáticos discretos en las ciencias de la naturaleza: teoría y problemas*, pages 87–147. Díaz de Santos, 2002.
- [12] William Heaton Hamer. *The milroy lectures on epidemic diseases in england: The evidence of variability and of persistency of type; delivered before the royal college of physicians of london, march 1st, 6th, and 8th, 1906*. Bedford Press, 1906.
- [13] Jose Diamantino Hernández Guillén. *Modelización Matemática de la propagación de malware: Un nuevo enfoque basado en la seguridad de la información*. PhD thesis, USAL, 2020.
- [14] Oluwatobiloba Ige. Markov Chain Epidemic Models and Parameter Estimation. *Theses, Dissertations and Capstones*, January 2020.
- [15] Valerie Isham, Simon Harden, and Maziar Nekovee. Stochastic epidemics and rumours on finite random networks. *Physica A: Statistical Mechanics and its Applications*, 389(3):561–576, February 2010.
- [16] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, January 1997.
- [17] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1):5, May 2007.
- [18] Luis Miguel Marcos Mejía. Estudio cualitativo y dinámico del modelo matemático SIR planteado por Kermack y Mckendrick. *Repositorio de Tesis - UNMSM*, 2021. Accepted: 2021-11-09T22:30:49Z Publisher: Universidad Nacional Mayor de San Marcos.

- [19] Grace Parazzoli. The Complete Guide to Organic Viral Marketing, March 2022.
- [20] Ronald Ross. Some Quantitative Studies in Epidemiology. *Nature*, 87(2188):466–467, October 1911.
- [21] Isidora Subercaseaux. Marketing viral: qué es y cuáles son sus ventajas - InboundCycle, September 2022.
- [22] Rosario Susi García. *Cadenas de Markov en tiempo discreto*. 2008.