

Tesis de grado

Mitigación y propagación de noticias falsas en redes sociales

Autor: Bruno Russo

Directora: Inés Caridi

`inescaridi@gmail.com`

Co-Director: Pablo Balenzuela

`balen@df.uba.ar`

Primer Cuatrimestre 2025

Índice general

1. Introducción	3
2. Marco teórico	5
2.1. La analogía epidemiológica	5
2.2. Definiciones: tendencia, rumor, desinformación y noticia falsa	6
2.3. Grafos: algunas definiciones	6
3. Datos	8
3.1. Origen y recolección de datos	8
3.2. Estructura de un <i>tweet</i>	8
3.3. Preprocesamiento y selección de subconjuntos de datos	9
3.4. Construcción de la red de <i>retweets</i> :	14
4. Modelos epidemiológicos	16
4.1. Modelo SI (Susceptible-Infectado)	16
4.2. Modelos con un tercer estado: SIR y SIZ	17
4.2.1. Modelo SIR (Susceptible-Infectado-Recuperado)	17
4.2.2. Modelo SIZ (Susceptible-Infectado-Escéptico)	17
4.3. Modelos con estado de latencia: SEI y SEIZ	18
4.3.1. Modelo SEI (Susceptible-Expuesto-Infectado)	18
4.3.2. El modelo SEIZ (Susceptible-Expuesto-Infectado-Escéptico)	18
5. Metodología	20
5.1. Ajuste y estimación de parámetros de los modelos epidemiológicos	20
5.2. Métricas de evaluación del ajuste	21
5.3. Análisis topológico de la red de <i>retweets</i>	22
5.3.1. Métricas de centralidad	22
5.4. Implementación computacional	23
6. Resultados y análisis	24
6.0.1. Caso de estudio 1: SputnikV	24
6.0.2. Caso de estudio 2: AstraZeneca	26
6.0.3. Caso de estudio 3: Vacunatorio VIP	28
6.1. Análisis de los resultados de los modelos epidemiológicos	30
6.1.1. Análisis del rendimiento cuantitativo	30
6.1.2. Análisis del ajuste cualitativo	30
6.1.3. Interpretación del rendimiento superior del modelo SEIZ sin ρ	31
6.2. Resultados de la red de <i>retweets</i>	31
6.2.1. Distribución de grados	31

6.2.2.	Análisis de centralidad de intermediación	33
6.2.3.	Comparación entre métricas de la red dos a dos	33
6.2.4.	Identificación de nodos clave	36
6.3.	Integración de la dinámica temporal y la estructura de red	40
7.	Conclusiones y próximos pasos	41
7.1.	Trabajos futuros y próximos pasos	41
A.	Análisis de sensibilidad al parámetro N: metodología y justificación	43
A.1.	Resultados de parámetros por caso de estudio	43
A.1.1.	Análisis de parámetros de tasa de contacto (dependientes de N): . .	51
	Bibliografía	51

Capítulo 1

Introducción

La consolidación de las redes sociales como plataformas primarias para el consumo de información ha transformado radicalmente el ecosistema mediático. Plataformas como *Twitter* (actual *X*), *Facebook* e *Instagram* permiten una diseminación de contenidos a una velocidad y escala sin precedentes, democratizando la creación y distribución de noticias. Sin embargo, esta misma virtud, caracterizada por bajas barreras de entrada y algoritmos que favorecen a la viralización de contenido, ha creado un terreno fértil para la propagación de desinformación. Este fenómeno, definido como la difusión de información falsa o engañosa, representa uno de los desafíos más complejos de la era digital, ya que tiene la capacidad de erosionar la confianza pública, polarizar el debate social e impactar significativamente en procesos políticos, económicos y sociales. [1, 2].

La crisis sanitaria global provocada por la pandemia de COVID-19 sirvió como un catalizador para una **infodemia** a escala mundial [3]. En este contexto, la desinformación sobre el origen del virus, los tratamientos y, de manera crucial, la seguridad y eficacia de las vacunas, se propagó masivamente, amenazando los esfuerzos de salud pública y generando focos de escepticismo y rechazo a la vacunación.

Si bien la mitigación de noticias falsas es la motivación última, este trabajo de tesis se enfoca en un estudio fundamental sobre la dinámica de propagación de tendencias en redes sociales. Se parte de la premisa de que los mecanismos de viralización de un rumor, una noticia falsa o incluso un evento verídico de alto impacto comparten características de propagación comunes que pueden ser estudiadas y modeladas. Si bien en los datos trabajados no encontramos noticias falsas que hayamos podido constatar que efectivamente lo eran, nos limitamos a estudiar tendencias ya que, las noticias falsas, podrían pensarse como un subconjunto dentro de las tendencias.

Este trabajo de tesis aborda el problema de la propagación de tendencias en redes sociales desde una perspectiva cuantitativa, aplicando un marco conceptual tomado de la epidemiología. La premisa fundamental es que la difusión de una idea, rumor o noticia falsa en una red social puede ser modelada como un proceso de contagio, análogo a la propagación de una enfermedad infecciosa en una población. Los usuarios de la red pueden ser clasificados en distintos estados (por ejemplo, susceptibles de creer la desinformación, infectados que la propagan activamente, o escépticos que la rechazan), y la transición entre estos estados puede ser descrita mediante sistemas de ecuaciones diferenciales.

El objetivo principal de esta investigación fue analizar la dinámica de propagación de información y tendencias relacionada con las vacunas contra el COVID-19 en *Twitter*, utilizando un conjunto de datos recolectados entre el 29 de diciembre de 2020 y mediados de abril de 2021 asociados a la vacunación y a vacunas contra el COVID-19. Se exploraron

y compararon varios modelos epidemiológicos —desde el simple modelo SI (suceptible-infectado) hasta el más complejo SEIZ (Susceptible-Expuesto-Infectado-Escéptico)- para determinar cuál de ellos se ajustó mejor a los patrones empíricos observados. A través del ajuste de estos modelos a los datos reales, se buscó estimar parámetros clave que gobiernan la dinámica de difusión, como la tasa de contagio o la tasa de escepticismo.

Capítulo 2

Marco teórico

En este capítulo se establecen las bases teóricas que fundamentan el presente estudio. Se comienza por introducir la analogía epidemiológica, que constituye el pilar central de nuestra metodología, para luego realizar una distinción conceptual clave entre los términos de rumor, desinformación y noticia falsa y cerrar el capítulo dando algunas definiciones sobre grafos en relación a nuestro caso de estudio.

2.1. La analogía epidemiológica

La similitud entre la propagación de un virus biológico y la difusión de información, tendencia o de un rumor es sorprendentemente fuerte, y ha sido notada en la literatura desde hace décadas [4]. Esta analogía permite importar un vasto conjunto de herramientas matemáticas desarrolladas en el campo de la epidemiología para modelar y analizar la difusión de contenido en redes sociales.

Para entender la idea central del modelo, se definen los siguientes elementos:

- **Población:** El conjunto de todos los usuarios de la red social de nuestro conjunto de datos.
- **Individuos:** Los usuarios de la red social.
- **Contagio:** Un usuario “infectado” (que se suma a la tendencia) transmite el “virus” (contenido relacionado con la tendencia) a un usuario “susceptible” (que aún no la ha visto) a través de un *tweet*, un *retweet*, un mensaje o una cita. En este trabajo, un usuario “infectado” es aquel que publica un *tweet* bajo algunos de los *hashtags* observados.
- **Estados de los individuos:** Al igual que en una epidemia, los usuarios pueden transitar por diferentes estados. Los modelos más básicos consideran los estados de **Susceptible (S)** e **Infectado (I)**. Modelos más sofisticados, como los que se usaron en esta tesis, introducen estados adicionales como **Expuesto (E)** (ha visto la información pero aún no la comparte), **Recuperado (R)** (ha compartido la información pero ya no lo hace y es inmune) o **Escéptico (Z)** (en caso de ver la información, tiene la capacidad de identificarla como falsa y por eso no compartirla).

Esta perspectiva permite pasar de un análisis puramente descriptivo a un modelado predictivo y cuantitativo y que permita hacer predicciones. Al formular la dinámica de transición entre estados mediante ecuaciones diferenciales, es posible simular la evolución

temporal de la tendencia, predecir su alcance y, fundamentalmente, evaluar el impacto potencial de intervenciones para frenarla, en caso de tratarse de desinformación. [5, 6].

2.2. Definiciones: tendencia, rumor, desinformación y noticia falsa

Aunque a menudo se usan indistintamente, es útil distinguir entre los conceptos de tendendia, rumor, desinformación (*misinformation*) y desinformación intencionada (*disinformation*).

- **Tendencia (Trend):** En el contexto de las redes sociales, se refiere a un tema, evento o narrativa que experimenta un rápido y masivo aumento de visibilidad y discusión en un período de tiempo acotado. Se caracteriza por un alto volumen de publicaciones e interacciones por parte de un gran número de usuarios, generando un pico de actividad temporal bien definido. Fundamentalmente, una tendencia es agnóstica a la veracidad del contenido, pudiendo originarse a partir de noticias reales, rumores o campañas de desinformación. Para este trabajo, las tendencias se identifican y cuantifican a través del uso de *hashtags* específicos.
- **Rumor:** Es una historia o afirmación de interés público cuya veracidad no está verificada en el momento de su circulación [7]. Los rumores pueden terminar siendo verdaderos o falsos.
- **Desinformación (Misinformation):** Se refiere a información falsa que es compartida por personas que no necesariamente saben que es falsa. La propagación es a menudo no intencionada, basada en la creencia errónea de que la información es verídica [8].
- **Desinformación Intencionada (Disinformation):** Es información falsa creada y difundida con el propósito deliberado de engañar. Este tipo de contenido suele tener motivaciones políticas, ideológicas o económicas.

En el contexto de este estudio, que analizó la dinámica de propagación de tendencias, la distinción sobre la veracidad o intencionalidad del contenido es secundaria al patrón de difusión. Los modelos epidemiológicos aquí empleados son agnósticos a la naturaleza del “virus” informativo, ya que han demostrado ajustarse bien tanto a la propagación de noticias falsas como a tendencias sobre tópicos de cualquier naturaleza. Por ello, el enfoque del trabajo es modelar brotes de información o tendencias (representadas por el uso masivo de *hashtags*), un fenómeno que abarca tanto noticias verificadas como rumores y desinformación. Se utilizarán a lo largo de este trabajo términos como **desinformación** o **tendencia** para referirse al contenido que se propaga, centrándose en su dinámica más que en su veracidad [9].

2.3. Grafos: algunas definiciones

El análisis de la red de usuarios que interactúan, permite observar otro enfoque del análisis; mientras que el analogía epidemiológica nos puede proporcionar información valiosa a una escala “macro” de la propagación, la red de interacción, permite ver propiedades estructurales que podríamos considerar un análisis a nivel “micro” de cómo se propaga

la tendencia. Los grafos son estructuras perfectas para representar redes de interacción. En este trabajo utilizamos este objeto matemático para representar una red de *retweets* relacionados al "Vacunatorio VIP", donde los usuarios estarán vinculados si existe esta interacción entre ambos. En la sección 3.4 se profundiza sobre el armado. A continuación, algunas definiciones pertinentes con su respectiva interpretación con el caso de estudio.

- **Grafo:** Un grafo $G = (V, E)$ es una estructura matemática que consiste en un conjunto de *nodos* V y un conjunto de *aristas* E , donde cada arista conecta un par de nodos. Es decir la *arista* $(u, v) = (v, u)$. En nuestro caso de estudio, los *nodos* representan usuarios de *Twitter* y las *aristas* representan una interacción entre usuarios. En conjunto conforman la red, es decir, el grafo.
- **Dígrafo (Grafo Dirigido):** Un dígrafo $D = (V, E)$ es un grafo donde el conjunto de *aristas* E está compuesto por pares ordenados de *nodos*. Cada *arista* (u, v) representa una conexión con una dirección específica desde el *nodo* u hacia el *nodo* v . A diferencia del grafo definido anteriormente, $(u, v) \neq (v, u)$. En la red de *retweets*, el sentido en el que trazamos la dirección de la *arista* define la interpretación de la información que la red provee.
- **Grado de Entrada (In-degree):** En un dígrafo, el grado de entrada de un *nodo* v es el número total de *aristas* que apuntan hacia él. Formalmente, es la cardinalidad del conjunto $\{u \in V \mid (u, v) \in E\}$.
- **Grado de Salida (Out-degree):** En un dígrafo, el grado de salida de un *nodo* v es el número total de *aristas* que se originan en él. Formalmente, es la cardinalidad del conjunto $\{u \in V \mid (v, u) \in E\}$.
- **Intermediación (Betweenness Centrality):** Es una medida de la centralidad de un *nodo* dentro de un grafo. La intermediación de un vértice v se calcula como la suma de la fracción de los caminos más cortos entre todos los pares de vértices (s, t) que pasan a través de v . Un vértice con alta intermediación actúa como un *punto* importante en el flujo de información de la red.

Capítulo 3

Datos

El pilar de cualquier estudio de modelado empírico es el conjunto de datos sobre el cual se construye y valida. En este capítulo, se describe en profundidad el origen, la estructura y el proceso de curación del corpus de *tweets* utilizado en esta tesis.

3.1. Origen y recolección de datos

El conjunto de datos utilizado en esta investigación fue recolectado originalmente con otros fines durante la pandemia de COVID-19, abarcando un período clave de la campaña de vacunación en Argentina y el mundo, aproximadamente entre diciembre de 2020 y abril de 2021. La recolección se realizó utilizando la API de *Twitter*, que permite acceder a un flujo de datos históricos y en tiempo real.

Se utilizaron filtros de palabras clave para capturar tweets relevantes. Las consultas de búsqueda incluían términos generales como **vacuna**, **vacunación**, **vacunarse**, así como los nombres específicos de las vacunas que eran prominentes en el debate público en ese momento, tales como **Sputnik V**, **AstraZeneca**, **Sinopharm** y **Pfizer**. El resultado fue un conjunto de datos masivo de aproximadamente 58 millones de tweets en idioma español. Este corpus, por su temática y temporalidad, representa un gran recurso para estudiar la propagación de narrativas y tendencias, tanto verídicas como falsas, sobre la vacunación.

3.2. Estructura de un *tweet*

Para comprender el proceso de análisis, es fundamental entender la estructura de los datos. Cada tweet recuperado a través de la API de *Twitter* es un objeto en formato JSON (JavaScript Object Notation), que contiene una rica variedad de metadatos. A continuación se describe la estructura y las claves más relevantes para este estudio.

Un *tweet* puede ser una publicación original, una respuesta (*reply*), una cita (*quote*) o un *retweet*. Esta naturaleza anidada es crucial, ya que un *retweet* o una cita contienen, a su vez, el objeto completo del *tweet* original.

```
1 {  
2   "created_at": "Tue Apr 06 15:24:15 +0000 2021",  
3   "id_str": "1379459288880562176",  
4   "full_text": "Cuidado con la vacuna Sputnik V, dicen que causa efectos  
5     secundarios graves. #sputnikv #peligro",  
6   "entities": {
```

```

6   "hashtags": [
7     { "text": "sputnikv", "indices": [78, 87] },
8     { "text": "peligro", "indices": [88, 96] }
9   ],
10  "user_mentions": [],
11  "urls": []
12 },
13 "user": {
14   "id_str": "123456789",
15   "screen_name": "usuario_antivacunas",
16   "followers_count": 150,
17   "friends_count": 300
18 },
19 "retweeted_status": {
20   // Objeto completo del tweet original que fue retuiteado
21   // ...
22 },
23 "quoted_status": {
24   // Objeto completo del tweet que fue citado
25   // ...
26 },
27 "in_reply_to_status_id_str": "1379459000000000000"
28 }

```

Listing 3.1: Estructura simplificada de un objeto Tweet en formato JSON.

Las claves más importantes para nuestro análisis fueron:

- **id_str**: El identificador único del *tweet*. Es la clave primaria para cada registro.
- **created_at**: La fecha y hora exactas de creación del tweet (timestamp). Este campo es fundamental para construir las series temporales de propagación.
- **full_text**: El contenido textual completo del *tweet* (hasta 280 caracteres). Este es el campo que puede contener la desinformación.
- **entities**: Un objeto que contiene arrays de entidades extraídas del texto, como **hashtags**, **user_mentions** (menciones a otros usuarios) y **urls** (enlaces). Los *hashtags* son particularmente útiles para rastrear conversaciones temáticas.
- **user**: Un objeto que contiene información sobre el autor del *tweet*, como su **id_str**, su nombre de usuario (**screen_name**) y métricas de popularidad como **followers_count** (número de seguidores).
- **retweeted_status** (Objeto completo del tweet original que fue retuiteado) y **quoted_status** (Objeto completo del tweet que fue citado): Si un *tweet* es un *retweet* o una cita, estos campos contendrán el objeto completo del *tweet* original. Esto permite rastrear la difusión de una pieza de contenido específica a través de la red.

3.3. Preprocesamiento y selección de subconjuntos de datos

El conjunto de datos crudo de 58 millones de *tweets* es demasiado grande para un análisis directo. Por lo tanto, se llevó a cabo un riguroso proceso de preprocesamiento y selección de *tweets* de interés para estudiar la propagación de información.

1. **Análisis exploratorio de *hashtags*:** El primer paso fue realizar un conteo de la frecuencia de todos los *hashtags* presentes en el conjunto de datos. Esto permitió identificar los temas de conversación más prominentes y detectar *hashtags* asociados a temas de alta repercusión, los cuales a menudo generan picos de actividad viral.
2. **Selección de casos de estudio:** Basado en el análisis anterior y en el conocimiento del contexto sociopolítico de la época, se seleccionaron varios *hashtags* de interés para un análisis más profundo. Entre ellos se encuentran:
 - **#sputnikv:** Para analizar las narrativas en torno a la vacuna rusa, que fue objeto de intenso debate.
 - **#astrazeneca:** Para estudiar las conversaciones sobre los casos de trombosis, que generaron una crisis de confianza.
 - Un conglomerado de *hashtags* relacionados con el evento del “Vacunatorio VIP” en Argentina (por ejemplo, **#vacunatoriovip**, **#vacunasvip**). Este evento, al estar basado en un hecho real —la vacunación irregular durante la campaña contra la COVID-19— se convirtió en un punto focal de debate público y participación masiva en redes sociales [10, 11, 12]. Además, generó narrativas polarizadas, rumores y una cascada de reacciones digitales bajo el hashtag **#vacunasvip**, demostrando que los modelos de propagación son aplicables a cualquier brote de información de alto impacto, no solo a noticias categóricamente falsas [12].
 - Además, para la temática del “Vacunatorio Vip”, se extendió el conjunto de datos más allá de los *hashtags*; dicho conjunto de datos es el resultado de la búsqueda: [“vacunatorio vip”, “vacunatoriovip”, “vakunatorio”, “vacuna vip”, “vacunación vip”, “vacunados vip”, “gines”, “lista de vacunados”, “escándalo vacunación”, “escandalo vacunas”, “turno vip”] con el fin de extender un poco más el análisis. Dicho conjunto de datos, se utilizó para el armado de la red de *retweets*, que será definida en la sección 3.4.
3. **Análisis de la dinámica temporal:** Para cada *hashtag* o conjunto de *hashtags* de interés, se generó una serie temporal de su uso a lo largo de todo el período del conjunto de datos (diciembre 2020 - abril 2021). Esto se hizo contando el número de *tweets* por día (o por rango de tiempo arbitrario) que contenían el **hashtag**.

El objetivo de este análisis fue identificar picos de actividad muy pronunciados, que suelen corresponder al estallido de un evento noticioso, un rumor o una campaña de desinformación. Estos picos representan brotes epidémicos de información, ideales para ser modelados.

Como ejemplo, en la Figura 3.1 se muestra la actividad temporal para un conjunto de *hashtags* relacionados con el **Vacunatorio VIP** en Argentina. Aunque cada *hashtag* tiene su propia dinámica, se observa claramente un pico masivo y sincronizado a finales de febrero de 2021, indicando un evento de gran interés público.

Para simplificar el análisis y capturar el impacto total del evento, se agregaron las series temporales de todos los *hashtags* relacionados. El resultado, visible en la Figura 3.2, es una única serie temporal que muestra de forma inequívoca un brote de información concentrado principalmente entre el 19 y el 22 de febrero de 2021. Esta ventana temporal, caracterizada por un rápido crecimiento y posterior saturación, se selecciona como uno de los principales casos de estudio para el ajuste de modelos.

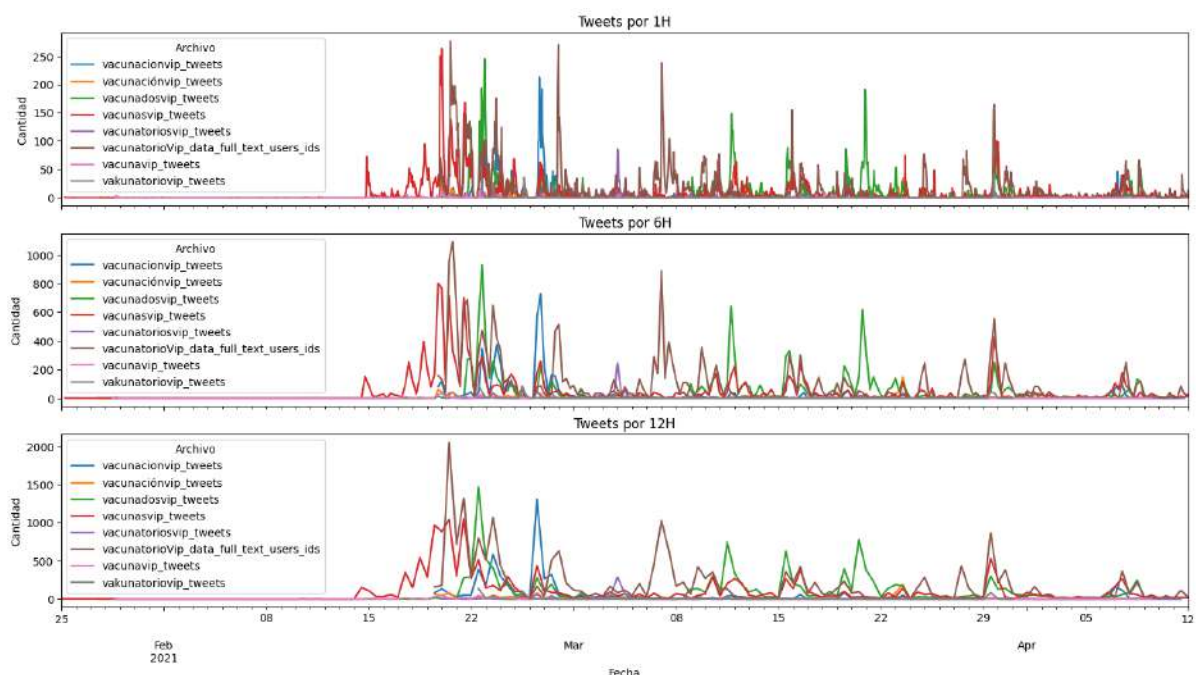


Figura 3.1: Frecuencia de tweets agrupados por 1 hora (1H), 6 horas (6H) y 12 horas (12H) para varios hashtags relacionados con el **Vacunatorio VIP**. Se observa un pico de actividad sincronizado y masivo alrededor de finales de febrero de 2021.

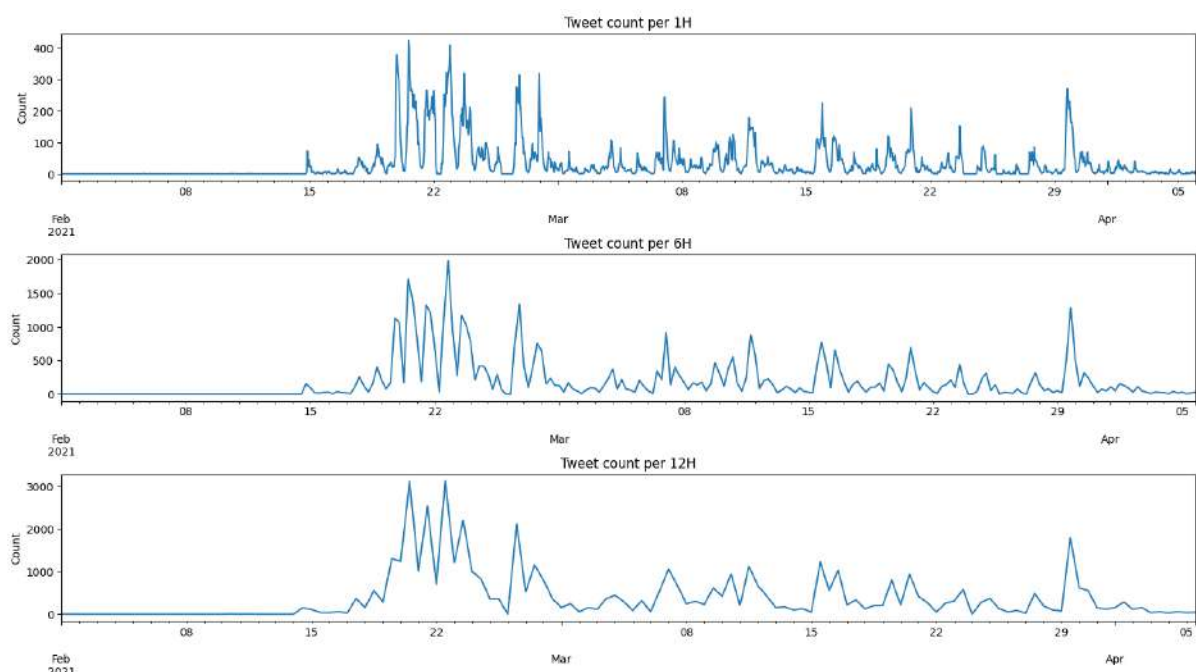


Figura 3.2: Serie temporal agregada para el conjunto de *hashtags* del **Vacunatorio VIP**. Al sumar las contribuciones, el pico de actividad se vuelve inequívoco, concentrándose principalmente entre el 19 y el 22 de febrero de 2021. Esta ventana temporal se selecciona como un caso de estudio.

Este método se aplicó a otros temas de interés. La Figura 3.3 muestra la dinámica temporal para el *hashtag* #sputnikv. Aquí también se pueden identificar múltiples picos de interés, como el de principios de febrero y otro muy pronunciado a finales de marzo. Cada uno de estos picos representa un evento de difusión distinto y podría constituir un caso de estudio independiente.

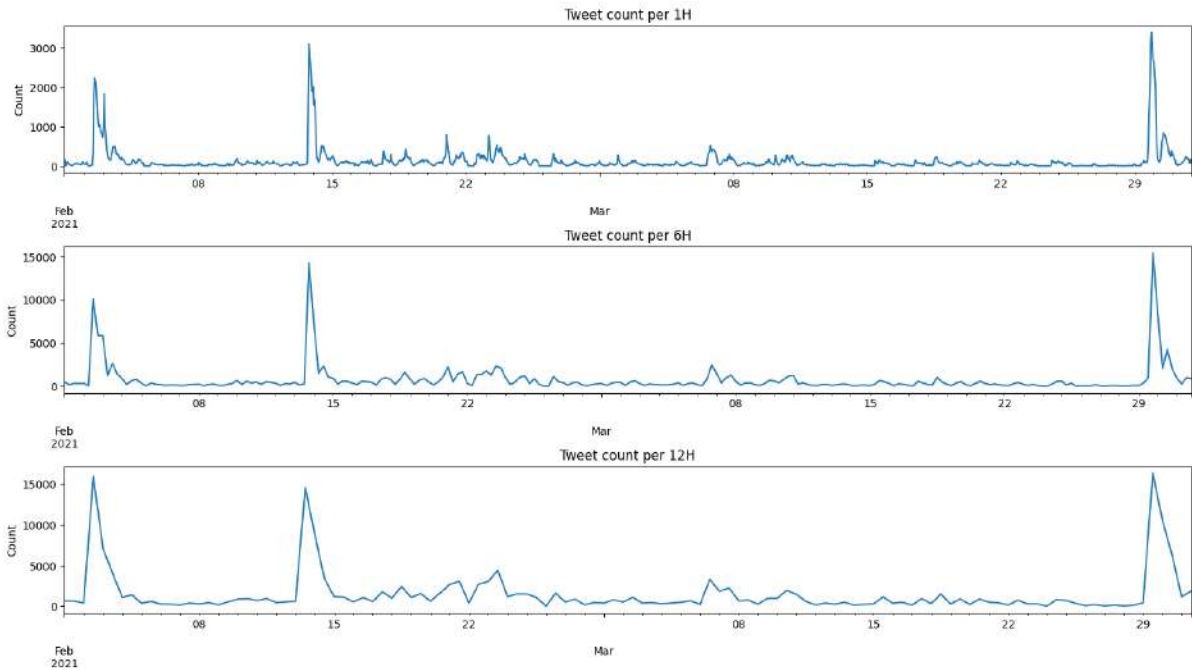


Figura 3.3: Ejemplo de la dinámica temporal para el *hashtag* #sputnikv. Se pueden identificar múltiples picos de interés, cada uno de los cuales podría constituir un caso de estudio independiente para el modelado.

Contexto en torno a los picos principales:

- **2–5 de febrero de 2021:** Se publicaron los primeros resultados del ensayo clínico de fase III de **SputnikV**, mostrando una eficacia del 91,6 —informado por medios como Capital 21 el 2 de febrero— y se comenzaron a aplicar dosis en centros de salud, incluso en espacios como el estadio River Plate el 2 de febrero -[13, 14].
- **13–15 de febrero de 2021:** Varias jurisdicciones informaron incidentes con dosis de **SputnikV** (por ejemplo, desaparición de 40 dosis en Comodoro Rivadavia el 24 de febrero), y la ANMAT actualizó información sobre su seguridad y uso en mayores de 60 años el 21 de febrero [15, 16].
- **29–31 de marzo de 2021:** Llegaron a Argentina nuevos lotes de **SputnikV** desde Rusia y otros países reportaron aprobación del uso de emergencia, generando aumentos en la conversación pública hacia esos días [17, 18].

De manera similar, la Figura 3.4 muestra la dinámica para los *hashtags* relacionados con AstraZeneca. Se identifica claramente un brote de información masivo entre el 15 y 22 de marzo de 2021, coincidiendo con la cobertura mediática internacional sobre los posibles efectos adversos de la vacuna, lo que lo convierte en un excelente caso de estudio sobre una crisis de confianza.

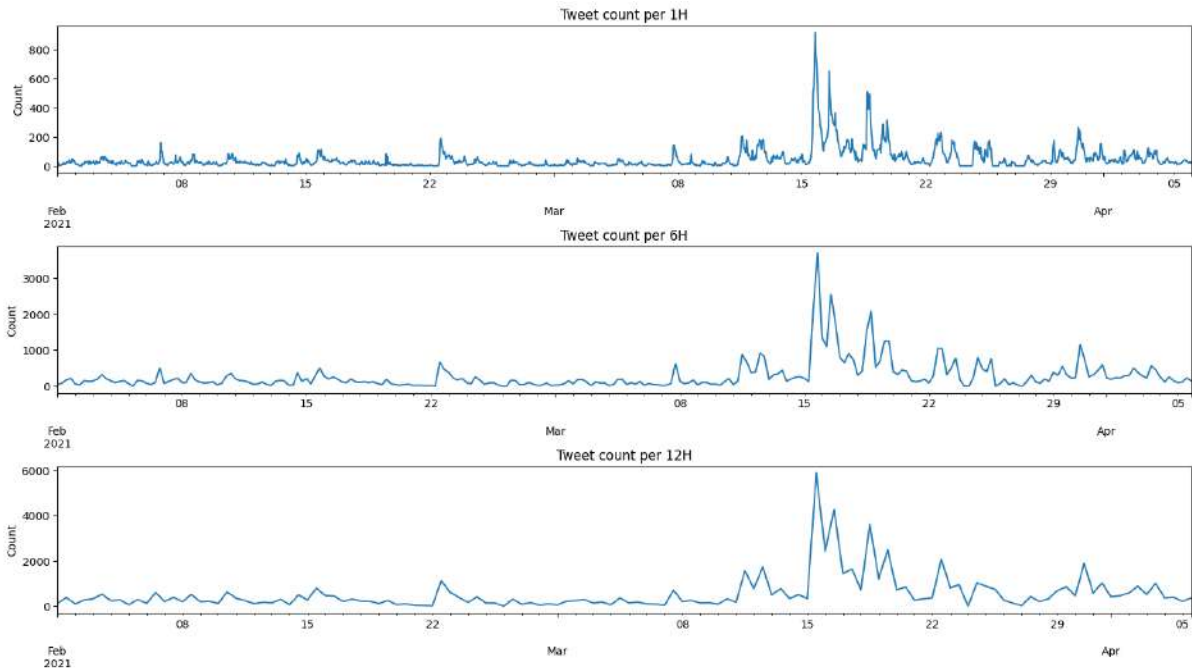


Figura 3.4: Dinámica temporal para el hashtag #astrazeneca. Se observa un pico de interés muy pronunciado alrededor del 15 de marzo de 2021, coincidiendo con las noticias sobre casos de trombosis.

Contexto del pico (15–22 de marzo de 2021):

- El 15 de marzo varios países europeos como Francia, Alemania, Italia, España y otros suspendieron temporalmente el uso de AstraZeneca por “precaución” tras informes de casos muy raros de trombosis y trombocitopenia, aunque la OMS y la EMA reafirmaron que la vacuna seguía siendo beneficiosa [19, 20].
- El 16 de marzo la EMA señaló que era “plausible” un vínculo inmunológico con los casos de coágulos cerebrales, aunque reiteró que el balance riesgo-beneficio seguía siendo favorable y alentó a continuar con la administración vigilante [21, 22].
- Para el 19 de marzo, el subcomité de seguridad de la OMS también confirmó que se habían registrado casos raros de tromboembolias asociadas, pero sin evidencia de aumento global de riesgo y recomendó continuar con la vacunación mediante una monitorización activa [23].

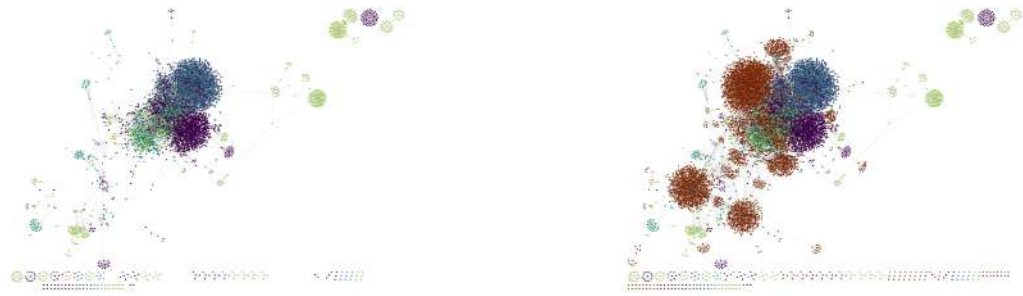
4. **Creación de curvas acumuladas:** Una vez identificado un pico de actividad (por ejemplo, un período de 2 a 3 días donde el uso de un hashtag se disparó), se procedió a crear una serie temporal para esa ventana específica. Se filtraron todos los *tweets* dentro de esa ventana temporal que contenían el *hashtag* de interés. Luego, se calculó el **número de usuarios únicos** que participaron de la tendencia en intervalos de 15 minutos desde el inicio del pico. El resultado es una curva empírica de propagación que representa cómo la “infección” (el uso del *hashtag*) se extendió a través de la población de usuarios a lo largo del tiempo. Esta curva acumulada es el dato empírico sobre la evolución de la cantidad de usuarios en estado infectado en el tiempo que se utilizará para ajustar y validar los modelos epidemiológicos descritos en el siguiente capítulo.

3.4. Construcción de la red de *retweets*:

Para analizar la dinámica de difusión de la información durante el período de interés, se construyó una red de *retweets*. Esta red fue concebida como un **grafo dirigido** para modelar el flujo de información entre los participantes de la conversación en *Twitter*. Como ha sido mencionado previamente, el análisis de la red de *retweets* puede proveernos de herramientas para complementar el análisis macroscópico del modelo epidemilógico.

- **Datos utilizados para la red:** del subconjunto de datos mencionado al final de la sección 2, se seleccionaron exclusivamente aquellos usuarios que fueron retuiteados por otros y aquellos que realizaron un *retweet*. Esta identificación se basó en la presencia de los metadatos que distinguen a un *retweet* de un tuit original (también detallado en la sección 3.2), permitiendo así aislar las acciones de redifusión de contenido.
- **Nodos (V):** El conjunto de nodos de la red está constituido por usuarios de *Twitter*, en específico aquellos usuarios que tuitearon o retuitearon en relación al "Vacunatorio VIP". Cada usuario único, ya sea como autor de un mensaje original o como difusor del mismo, se representa como un nodo en el grafo.
- **Aristas (E):** Las aristas simbolizan la acción de retuitear y, por ende, el flujo de información. Para cada *retweet*, se trazó una arista dirigida desde el nodo del autor original del tuit hacia el nodo del usuario que lo retuiteó.
- **Grado de salida:** Por la construcción de la red, el grado de salida representa las veces que un usuario fue retuiteado por otros.
- **Grado de entrada:** Por la construcción de la red, el grado de entrada representa las veces que un usuario retuiteó a otros.

De esta forma, una arista $u \rightarrow v$ en la red significa que “el usuario v retuiteó al usuario u al menos una vez”, modelando así la propagación de un mensaje a través de la plataforma. Este enfoque permite identificar a los actores centrales en la difusión de información y visualizar los patrones de comunicación. La red total de este conjunto de datos consta de aproximadamente 54 mil nodos y 130 mil aristas. Aprovechando nuevamente la metadata de los *tweets*, en este caso la fecha de creación, podemos ver la evolución de la red de *retweets* en el tiempo. A su vez, podemos ver momentos específicos de forma cronológica; por ejemplo, ver cuando la red tiene 5000, 10000 usuarios hasta la red completa contenida en el conjunto de datos.



(a) Visualización de la red de *retweets* con 5000 usuarios. (b) Visualización de la red de *retweets* con 10000 usuarios.

Figura 3.5: Comparación de la red en dos momentos específicos. Cada color representa el paso en el cual los nodos aparecen en la red. Cada paso representa hasta la cantidad de usuarios que había en la red hasta ese momento. El paso es arbitrario y para el armado de este gráfico fueron cuatro; se muestran los últimos dos.

Capítulo 4

Modelos epidemiológicos

Este capítulo presenta los modelos matemáticos que se utilizaron para analizar la dinámica de propagación de información. Se comenzó con los modelos compartimentales más simples y que asumen menos estados y transiciones entre estados, y se avanzó progresivamente hacia el modelo SEIZ, que es el más completo y el enfoque central de esta tesis. Todos los modelos se describen mediante sistemas de Ecuaciones Diferenciales Ordinarias (EDOs).

4.1. Modelo SI (Susceptible-Infectado)

El modelo SI es el modelo epidemiológico más fundamental. Divide a la población total N en dos compartimentos:

- **Susceptibles (S):** Usuarios que aún no han visto la tendencia y son vulnerables a ser infectados o influenciados.
- **Infectados (I):** Usuarios que han visto la tendencia y la están propagando activamente (por ejemplo, han tuiteado al respecto).

La dinámica asume que un individuo susceptible se infecta a una tasa β al entrar en contacto con un individuo infectado. Una vez infectado, un individuo permanece en ese estado para siempre. Cabe destacar que, la interpretación de un infectado, no distingue si el usuario publica una única vez o varias. Esto es importante a tener en cuenta ya que, alguien que publica una única vez podría considerarse un “recuperado”.

El sistema de EDOs que describe este modelo es:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N}\end{aligned}$$

donde $N = S + I$ es la población total, que se mantiene constante.

- β : Tasa de contacto efectivo o tasa de “contagio”. Representa la probabilidad de que un contacto entre un susceptible y un infectado resulte en una nueva infección.

Limitación: La principal debilidad del modelo SI es que no contempla ninguna forma de recuperación o escepticismo. Predice que, con el tiempo, toda la población susceptible se infectará, lo cual no es realista para la propagación de información, donde el interés decae o los usuarios se vuelven escépticos.

4.2. Modelos con un tercer estado: SIR y SIZ

4.2.1. Modelo SIR (Susceptible-Infectado-Recuperado)

El modelo SIR añade un tercer compartimento:

- **Recuperados (R):** Usuarios que estuvieron infectados pero que ya no propagan la tendencia y se han vuelto inmunes. Si bien no es el compartimento que mejor modela la dinámica, lo estudiamos por ser una variante.

Los individuos pasan de I a R a una tasa γ .

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

- γ : Tasa de recuperación. Su inversa, $1/\gamma$, es el período infeccioso promedio.

Este modelo es más realista que el SI, ya que predice un pico de infecciones seguido de un declive, como se observa en muchas epidemias reales y en picos de atención en redes sociales pero sigue sin ser de los más acordes para el estudio de propagación de tendencias.

4.2.2. Modelo SIZ (Susceptible-Infectado-Escéptico)

El modelo SIZ es una variante del SIR, conceptualmente más adecuada para la propagación de información. Reemplaza el compartimento de Recuperados (R) por el de Escépticos (Z), representando una forma de resistencia o abandono de la tendencia.

- **Escépticos (Z):** Usuarios que han estado expuestos a la tendencia —por ejemplo, al ver un tuit que lo contiene— pero deciden no reaccionar públicamente al mismo. No lo retuitean, no lo comentan, ni generan nuevas publicaciones al respecto. Este compartimento representa a aquellos individuos que, ya sea por escepticismo activo hacia la veracidad de la información, por indiferencia, o por decisión consciente de no participar en su difusión, se mantienen inactivos ante el contenido. Es importante aclarar que pertenecer al compartimento Z no implica necesariamente tener una postura crítica o informada: simplemente significa no contribuir activamente a la propagación del contenido [6].

En su formulación análoga al SIR, la dinámica asume que los individuos Infectados se convierten en Escépticos a una tasa constante ϵ .

Las ecuaciones que describen el modelo son:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} - \frac{bSZ}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N} + \epsilon \frac{SZ}{N} \\ \frac{dZ}{dt} &= \frac{bSZ}{N}\end{aligned}$$

El nuevo parámetro es:

- b : Tasa de contacto entre Susceptibles (S) y Escépticos (Z)

4.3. Modelos con estado de latencia: SEI y SEIZ

4.3.1. Modelo SEI (Susceptible-Expuesto-Infectado)

El modelo SEI introduce un período de latencia, representado por el compartimento “Expuesto”.

- **Expuestos (E):** Usuarios que han visto la tendencia (han sido expuestos al “virus”) pero aún no la están propagando. Están en un período de incubación, durante el cual pueden estar evaluando la información o decidiendo si compartirla o no.

Los susceptibles se convierten en expuestos a una tasa β , cuando se encuentran con un infectado, y los expuestos se convierten en infectados a una tasa ϵ .

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dE}{dt} &= \frac{\beta SI}{N} - \epsilon E \\ \frac{dI}{dt} &= \epsilon E\end{aligned}$$

- ϵ : Tasa de progresión de expuesto a infeccioso. Su inversa, $1/\epsilon$, es el período de latencia promedio.

El estado Expuesto es muy relevante para modelar la desinformación, ya que captura el retraso entre el consumo de una noticia y la decisión de compartirla.

4.3.2. El modelo SEIZ (Susceptible-Expuesto-Infectado-Escéptico)

Finalmente, el modelo SEIZ es el más completo para nuestro propósito, ya que combina tanto un período de latencia (E) como un estado de resistencia activa (Z). Este modelo fue propuesto y validado para la difusión de noticias y rumores en *Twitter* [5, 6] y es el que mejor captura la complejidad del comportamiento del usuario en el contexto que nos interesa.

Los cuatro compartimentos son:

- **S (Susceptible):** Usuario que no ha visto la tendencia.
- **E (Expuesto):** Usuario que ha visto la tendencia pero está en un período de deliberación antes de decidir si la comparte o no.
- **I (Infectado):** Usuario que está compartiendo activamente la tendencia.
- **Z (Escéptico):** Usuario que ha visto la tendencia y ha decidido no compartirla, volviéndose inmune.

La dinámica es más rica. Un usuario Susceptible (S), al entrar en contacto con un Infectado (I), puede tomar dos caminos:

1. Convertirse en **Infectado (I)** directamente con una probabilidad p .
2. Convertirse en **Expuesto (E)** con una probabilidad $1 - p$, entrando en un estado de latencia.

Además, un usuario Susceptible (S) puede encontrarse con la tendencia y volverse directamente **Escéptico (Z)** o **Expuesto (E)**. El modelo completo, adaptado de [6], se describe con el siguiente sistema de EDOs:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} - \frac{bSZ}{N} \\ \frac{dE}{dt} &= (1-p)\frac{\beta SI}{N} + (1-l)\frac{bSZ}{N} - \rho E \frac{I}{N} - \epsilon E \\ \frac{dI}{dt} &= p\frac{\beta SI}{N} + \rho E \frac{I}{N} + \epsilon E \\ \frac{dZ}{dt} &= l\frac{bSZ}{N}\end{aligned}$$

Cuadro 4.1: Definición de los parámetros del modelo SEIZ.

Parámetro	Definición en el contexto de <i>Twitter</i>
β	Tasa de contacto entre Susceptibles (S) e Infectados (I).
b	Tasa de contacto entre Susceptibles (S) y Escépticos (Z).
p	Probabilidad de que un S se vuelva I directamente tras contacto con I.
$1-p$	Probabilidad de que un S se vuelva E tras contacto con I.
l	Probabilidad de que un S se vuelva Z tras contacto con Z.
$1-l$	Probabilidad de que un S se vuelva E tras contacto con Z.
ϵ	Tasa de transición espontánea de E a I (incubación/deliberación).
ρ	Tasa de contacto entre Expuestos (E) e Infectados (I) que acelera la infección.

Este modelo es poderoso, aunque incluye más parámetros a ajustar, porque permite bifurcaciones complejas en el comportamiento del usuario, capturando la posibilidad de creer inmediatamente, dudar, o ser escéptico desde el principio. La tarea en la siguiente fase será ajustar este modelo a los datos empíricos.

Modelos simplificados derivados del SEIZ. Para explorar la influencia de las distintas dinámicas internas del modelo y facilitar su interpretación, también se definieron dos variantes simplificadas:

- **SEIZ sin ρ :** Se elimina la contribución de la interacción entre Expuestos e Infectados en la transición a I , es decir, se anula el término $\rho E \frac{I}{N}$. Esto permite estudiar una versión donde los usuarios deliberan de forma independiente, sin influencia adicional de sus pares.
- **SEIZ sin ϵ :** Se elimina la transición espontánea de Expuestos a Infectados (ϵE), quedando únicamente la transición por contacto con Infectados. Esta variante asume que la deliberación solo se resuelve al ver a otros usuarios compartiendo activamente.

Estas variantes permiten analizar cómo cambia la dinámica de propagación al suprimir mecanismos específicos de contagio o activación, lo que puede facilitar la interpretación de los modelos.

Capítulo 5

Metodología

Este capítulo detalla el procedimiento metodológico que se siguió para conectar el marco teórico y los modelos descritos en el Capítulo 4 con los datos empíricos presentados en el Capítulo 3. El objetivo es corroborar qué modelo se ajusta mejor a cada caso de estudio y, en última instancia, utilizar los parámetros de los modelos epidemiológicos que mejor describen la propagación observada de información y, adicionalmente, caracterizar la estructura de la red de interacciones.

5.1. Ajuste y estimación de parámetros de los modelos epidemiológicos

El núcleo de la metodología consiste en un problema de optimización: encontrar el conjunto de parámetros para un modelo dado (por ejemplo, el modelo SEIZ) que minimice la discrepancia entre la curva de propagación simulada por el modelo y la curva de propagación empírica obtenida de los datos de *Twitter*.

La curva empírica es la serie temporal del número acumulado de usuarios únicos que publicaron en *Twitter* con el **hashtag** de interés, medida en intervalos de 15 minutos, como se describió en la Sección 3.3. La curva simulada se obtiene resolviendo numéricamente el sistema de Ecuaciones Diferenciales Ordinarias (EDOs) del modelo para un conjunto específico de valores de sus parámetros (por ejemplo, $\beta, p, \epsilon, \dots$).

El procedimiento es el siguiente:

1. **Definir una función de Costo (o error):** Esta función mide la diferencia entre la curva real y la simulada. Para el compartimento de Infectados $I(t)$, que se corresponde con los usuarios que tuitean, el SSE se define como:

$$SSE(\theta) = \sum_{t=t_0}^{t_f} (I_{real}(t) - I_{sim}(t, \theta))^2$$

donde $I_{real}(t)$ es el número acumulado de tweets en el tiempo t , $I_{sim}(t, \theta)$ es el número de infectados predicho por el modelo en el tiempo t con el vector de parámetros θ , y la sumatoria se realiza sobre todos los puntos temporales de la ventana de observación.

El vector de parámetros θ está compuesto por las constantes del modelo epidemiológico correspondiente (por ejemplo, $\theta = [\beta, \gamma]$ en el modelo SIR o $\theta = [\beta, b, p, l, \epsilon, \rho]$ en el modelo SEIZ), incluyendo también las condiciones iniciales cuando se optimizan (como S_0, E_0, I_0, Z_0).

La Tabla 5.1 resume los parámetros incluidos en θ para cada modelo utilizado en este estudio.

Cuadro 5.1: Parámetros incluidos en el vector θ para cada modelo epidemiológico

Modelo	θ	Significado de los parámetros
SI	β	Tasa de contacto entre susceptibles e infectados.
SIR	β, γ	β : tasa de infección; γ : tasa de recuperación.
SEI	β, ϵ	ϵ : tasa de progresión de expuesto a infectado.
SIZ	β, b	b : tasa de contacto entre susceptibles y escépticos.
SEIZ	$\beta, b, p, l, \epsilon, \rho$	β : contacto S-I; b : contacto S-Z; p : probabilidad de transición directa de S a I; l : probabilidad de transición de S a Z; ϵ : tasa de incubación (E→I); ρ : tasa de activación por exposición repetida.

2. **Optimización numérica:** Se utilizaron algoritmos de optimización para encontrar el vector de parámetros θ^* que minimiza la función de costo. Dado que los modelos son no lineales, se emplearon técnicas de optimización por mínimos cuadrados no lineales utilizando las bibliotecas científicas **scipy** y **sklearn** de Python.

5.2. Métricas de evaluación del ajuste

Una vez que se encuentran los parámetros óptimos θ^* , es necesario evaluar cuantitativamente qué tan bueno es el ajuste del modelo a los datos. Para ello, se utilizaron las siguientes métricas:

- **Error cuadrático medio (MSE):** Es la media de los errores al cuadrado. Es fácil de interpretar y penaliza más los errores grandes.

$$MSE = \frac{1}{n} \sum_{t=t_0}^{t_f} (I_{real}(t) - I_{sim}(t, \theta^*))^2$$

- **Coefficiente de determinación (R^2):** Mide la proporción de la varianza en los datos reales que es predecible a partir del modelo. Un valor de R^2 cercano a 1 indica un ajuste casi perfecto, mientras que un valor cercano a 0 indica que el modelo no explica la variabilidad de los datos.

$$R^2 = 1 - \frac{\sum (I_{real}(t) - I_{sim}(t, \theta^*))^2}{\sum (I_{real}(t) - \bar{I}_{real})^2}$$

donde $\bar{I}_{real}$ es la media de los datos reales.

- **Error relativo en norma-2:** Como se utiliza en [5], esta métrica normaliza el error por la magnitud de los datos, lo que facilita la comparación entre diferentes casos de estudio que pueden tener escalas muy distintas.

$$\text{Error Relativo} = \frac{\|I_{real} - I_{sim}\|_2}{\|I_{real}\|_2}$$

donde $\|\cdot\|_2$ es: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$

Se ajustaron los diferentes modelos (SI, SIR, SEI, SEIZ) a los mismos datos y se compararon estas métricas para determinar, de manera cuantitativa, cuál de ellos ofrece la mejor descripción de la dinámica de propagación.

5.3. Análisis topológico de la red de *retweets*

Además del análisis temporal mediante modelos epidemiológicos, se estudió la estructura topológica de la red de interacciones subyacente a la difusión del *hashtag*. Este análisis permite caracterizar los roles de los diferentes usuarios y la estructura general de la red, complementando la visión macroscópica de las EDOs.

5.3.1. Métricas de centralidad

Para cuantificar la importancia y el rol de cada nodo dentro de la red, se calculó la centralidad de intermediación. Esta métrica nos ayuda a entender cómo la posición de un usuario en la red se relaciona con su papel en la propagación de la información.

- **Centralidad de intermediación (betweenness centrality):** Mide cuántas veces un nodo actúa como intermediario en los caminos más cortos entre pares de nodos en la red. Formalmente, para un nodo v , su centralidad de intermediación se define como:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

donde:

- σ_{st} es el número total de caminos más cortos entre los nodos s y t .
- $\sigma_{st}(v)$ es el número de esos caminos que pasan por el nodo v .

Un nodo con alta centralidad de intermediación actúa como un **punto** entre distintas partes de la red. Es crucial para facilitar la difusión de información a gran escala, ya que conecta grupos que de otro modo estarían aislados.

En la implementación, se utilizó la función `networkx.betweenness centrality()`, que emplea un algoritmo basado en el método de **Brandes (2001)** para computar esta métrica de manera eficiente. Este algoritmo tiene una complejidad de:

$$\mathcal{O}(nm)$$

La función permite normalizar los resultados para que los valores estén entre 0 y 1, y puede operar sobre grafos dirigidos y no dirigidos, lo cual es de suma utilidad en este análisis ya que la red trabajada es dirigida.

Adicionalmente, se estudiaron las correlaciones entre esta métrica con las distribuciones de grado para cada nodo (e.g., centralidad de intermediación vs. grado de entrada/salida) para investigar la relación entre la actividad, la influencia y el rol de intermediación de los usuarios.

5.4. Implementación computacional

La implementación de toda la metodología se realizó íntegramente en el lenguaje de programación **Python**, haciendo uso de un conjunto de librerías científicas especializadas.

- **Manejo de datos:** Se utilizó la librería **Pandas** para cargar, limpiar y manipular tanto las series temporales para los modelos epidemiológicos como los datos de interacciones para la construcción de la red.
- **Resolución de EDOs:** Para resolver numéricamente los sistemas de ecuaciones diferenciales de los modelos epidemiológicos, se empleó la función `odeint` del módulo `scipy.integrate`.
- **Optimización:** Para encontrar los parámetros óptimos de los modelos, se utilizó la función `curve_fit` del módulo `scipy.optimize`.
- **Análisis de redes:** Para la construcción de la red dirigida y el cálculo de las métricas topológicas (grados de entrada/salida, centralidad de intermediación), se utilizó la librería **NetworkX**.
- **Visualización:** Las librerías **Matplotlib** y **Seaborn** se usaron para generar todos los gráficos, incluyendo las comparaciones de curvas ajustadas, los histogramas de distribuciones y los diagramas de dispersión de las métricas de red. Además, se utilizó el entorno de redes *Cytoscape* para graficar las redes de *retweets*.

El proceso computacional para cada caso de estudio consistió en construir un *script* que integra estos componentes para cargar los datos, realizar el ajuste de modelos, construir y analizar la red de interacciones, y finalmente reportar y visualizar los resultados de ambos enfoques.

¹La notación Big-O, o Gran O, es una notación que se utiliza para describir el comportamiento de un algoritmo en su peor escenario. En lugar de medir el tiempo exacto de ejecución, que depende del hardware, Big-O describe cómo escala el rendimiento o el uso de memoria del algoritmo a medida que aumenta el tamaño de la entrada.

Capítulo 6

Resultados y análisis

En este capítulo se presentan los resultados obtenidos al aplicar la metodología descrita para cada uno de los tres casos de estudio: La vacuna rusa SputnikV, la vacuna AstraZeneca y el caso de Argentina bautizado como "Vacunatorio VIP". El análisis se enfoca en la robustez de los modelos aplicados, comparando las métricas obtenidas para cada uno de ellos a los tres casos de estudio analizados en este trabajo. Por último, se realiza un análisis de la red de *retweets* para el caso de estudio 3 ("Vacunatorio VIP") junto con la integración de los resultados de ambos análisis.

6.0.1. Caso de estudio 1: SputnikV

Ajuste de curvas

La Figura 6.1 muestra el ajuste visual de las curvas generadas por cada modelo frente a la curva de tweets acumulados reales. Este gráfico permite una evaluación cualitativa de cómo cada modelo captura la forma y la temporalidad de la difusión de la información.

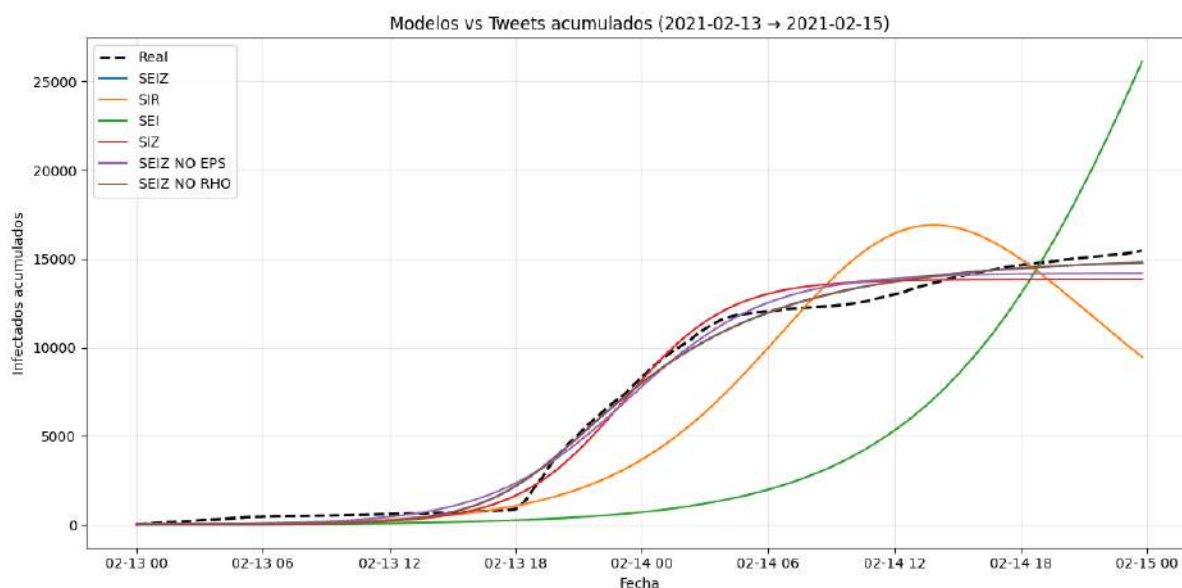
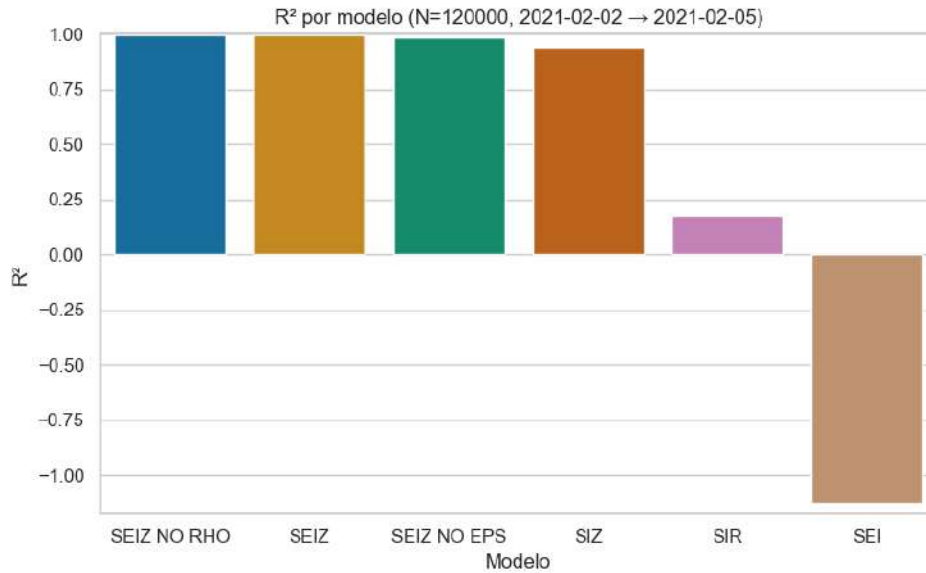


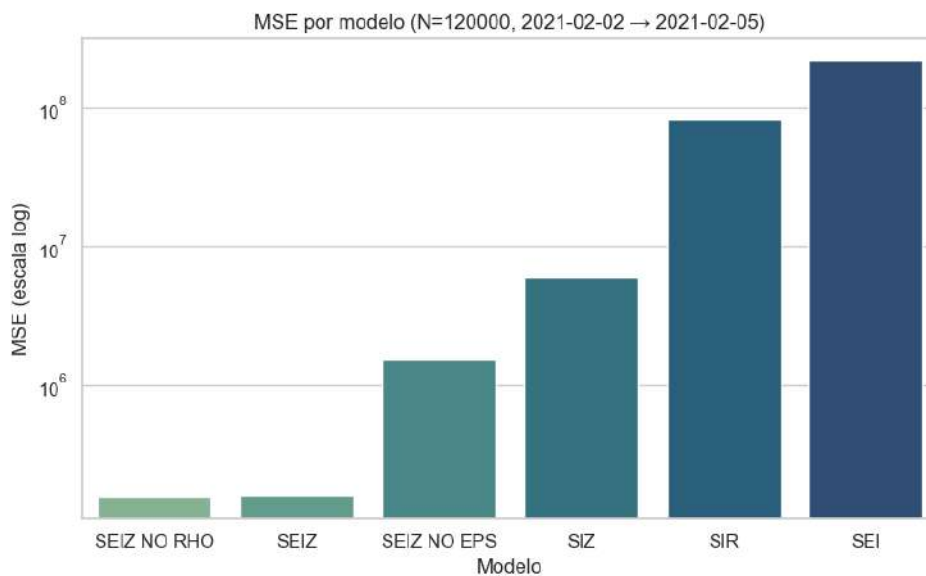
Figura 6.1: Comparación de las curvas de infectados acumulados de cada modelo con los datos reales para el caso SputnikV (2021-02-13 a 2021-02-15).

Métricas de ajuste

Además del análisis de estabilidad, se evaluó el rendimiento de cada modelo utilizando las métricas de Coeficiente de Determinación (R^2) y Error Cuadrático Medio (MSE). Los resultados se muestran en la Figura 6.2.



(a) R^2 por modelo.



(b) MSE por modelo (escala logarítmica).

Figura 6.2: Métricas de ajuste para los modelos en el caso de estudio SputnikV enmarcadas en la ventana del 2 al 5 de febrero de 2021.

6.0.2. Caso de estudio 2: AstraZeneca

Ajuste de curvas

El ajuste cualitativo de los modelos a los datos de la controversia de AstraZeneca se visualiza en la Figura 6.3.

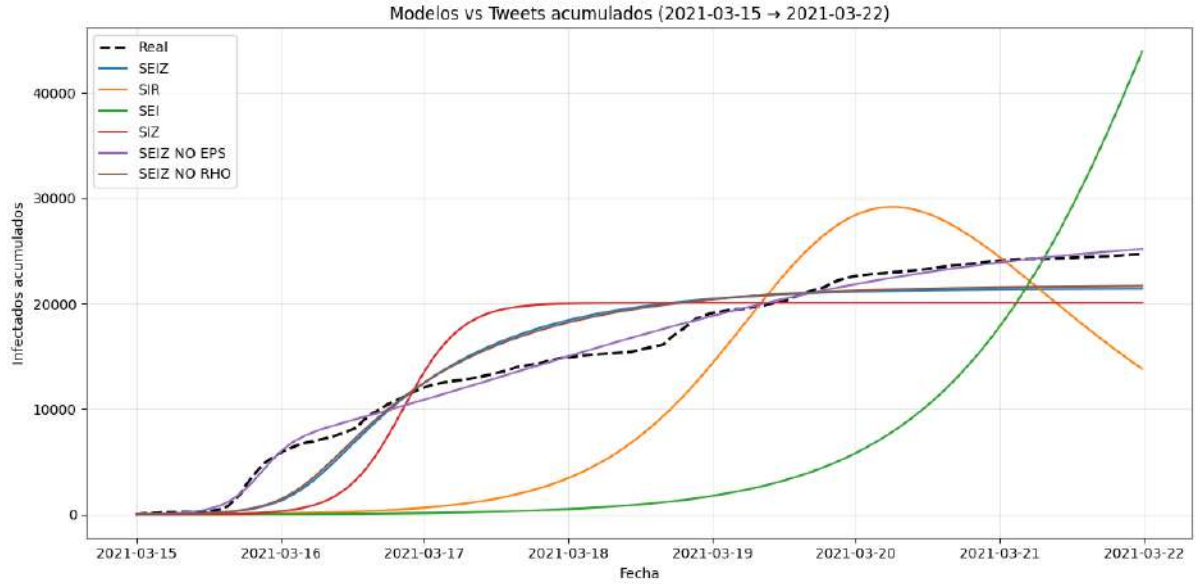
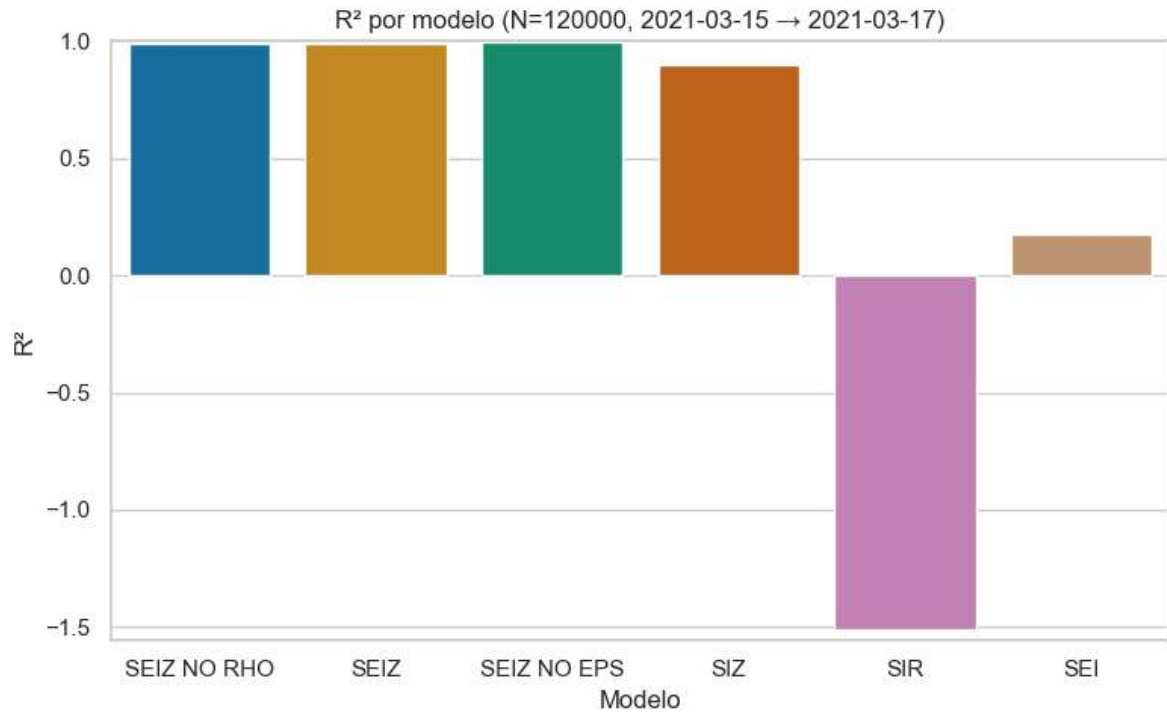


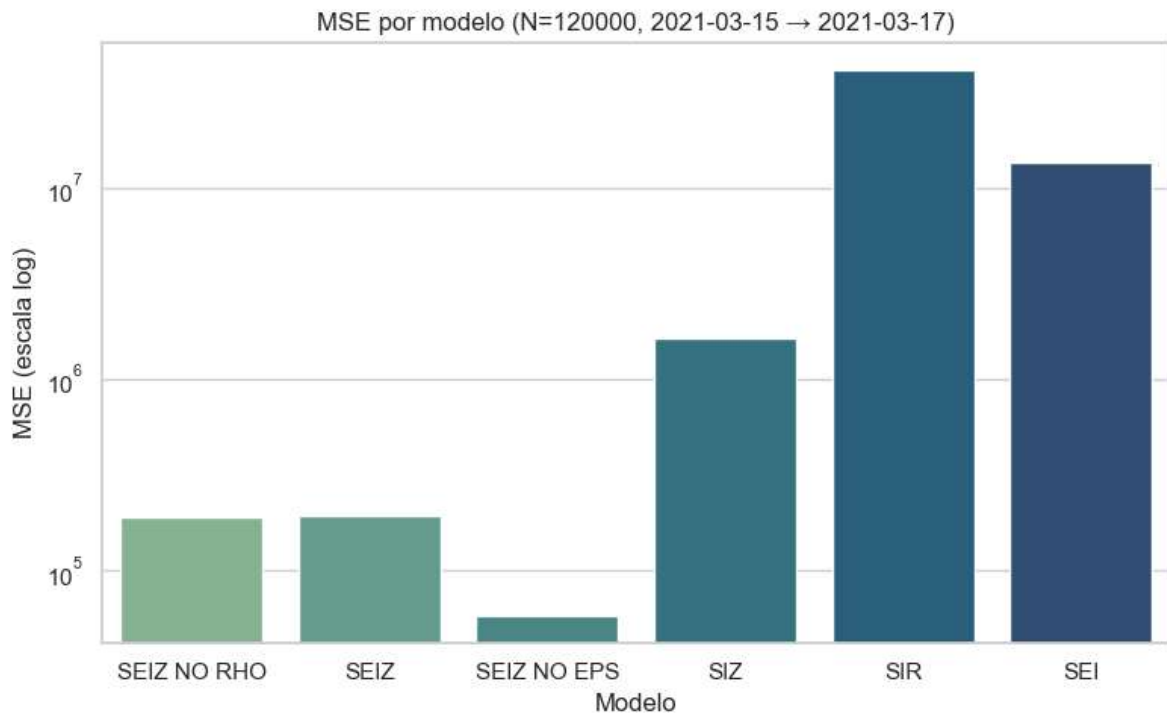
Figura 6.3: Comparación de las curvas de “infectados” acumulados de cada modelo contra los datos reales para el caso AstraZeneca (2021-03-15 a 2021-03-22).

Métricas de ajuste

Las métricas de R^2 y MSE para el caso de AstraZeneca se presentan en la Figura 6.4, complementando el análisis de estabilidad de los parámetros.



(a) R^2 por modelo.



(b) MSE por modelo (escala logarítmica).

Figura 6.4: Métricas de ajuste para los modelos en el caso de estudio AstraZeneca enmarcadas en la ventana del 15 al 17 de marzo de 2021

6.0.3. Caso de estudio 3: Vacunatorio VIP

Ajuste de curvas (Vacunatorio VIP)

La Figura 6.5 ilustra el desafío que este conjunto de datos presentó para todos los modelos, mostrando el ajuste visual de cada uno.

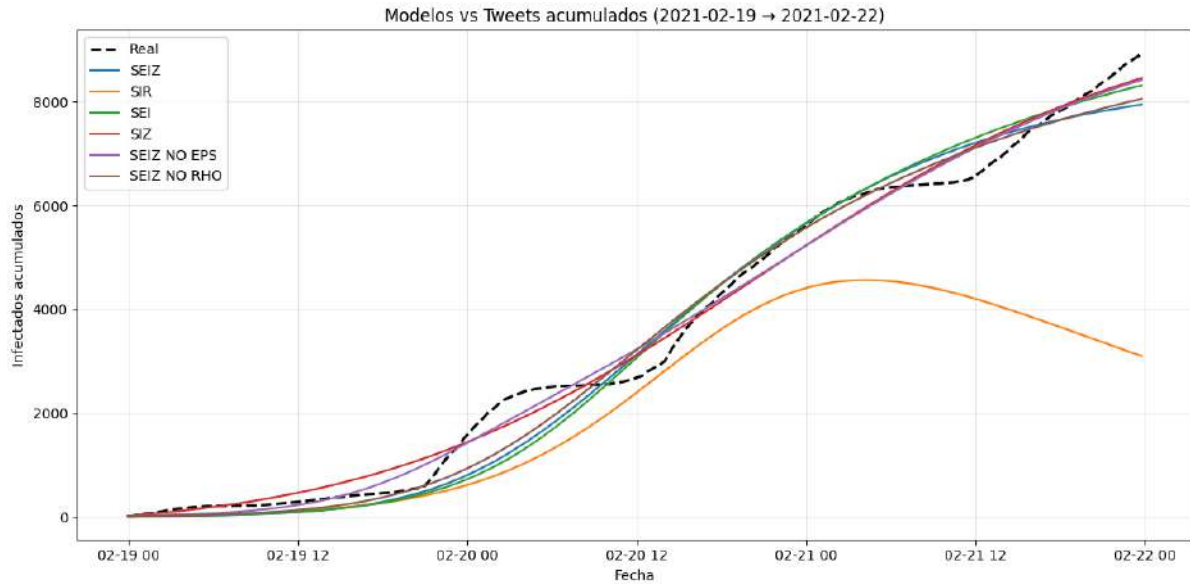
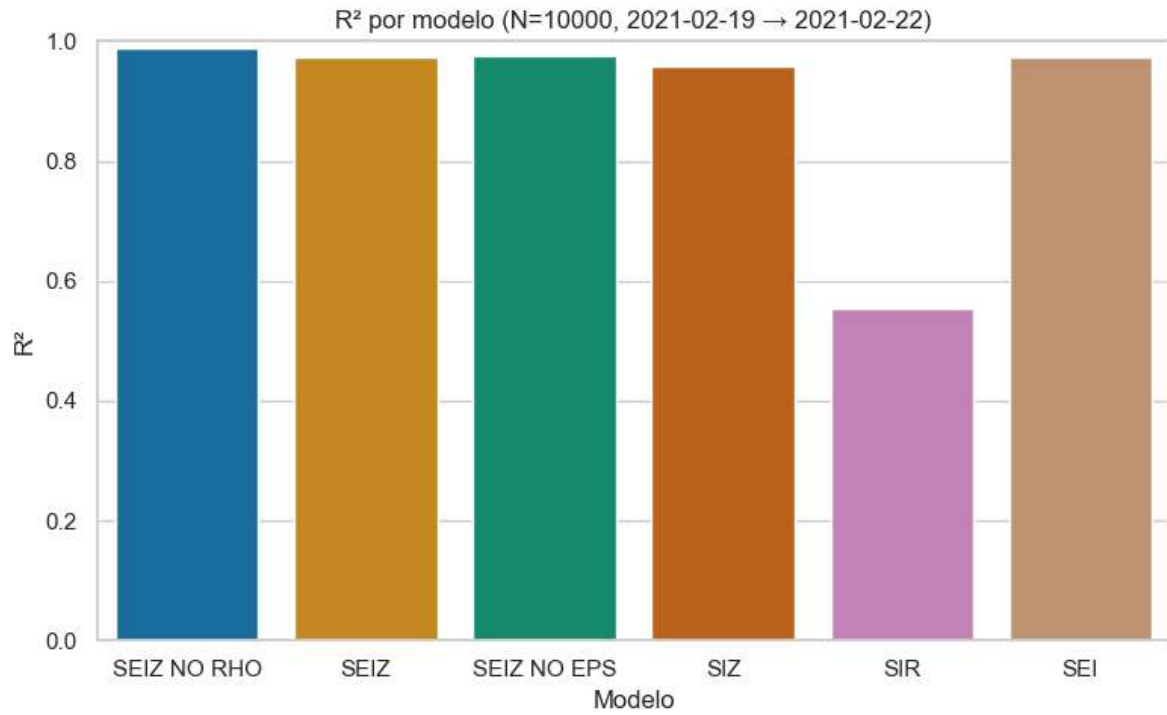


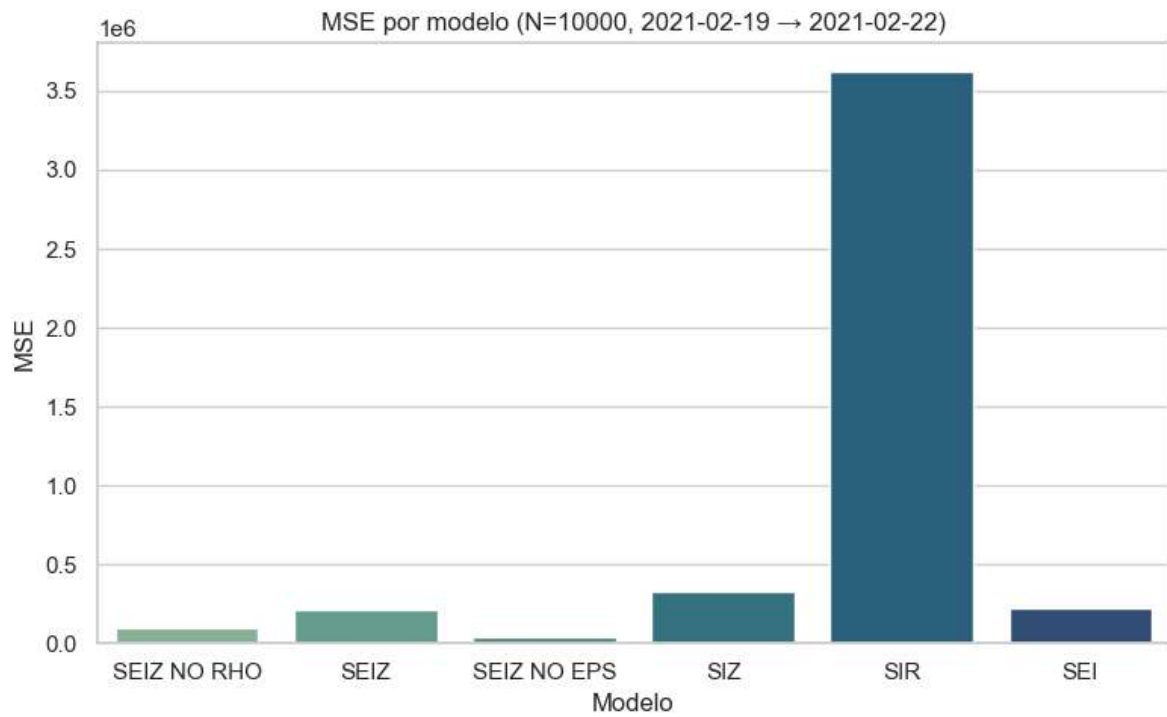
Figura 6.5: Comparación de las curvas de “infectados” acumulados de cada modelo contra los datos reales para el caso Vacunatorio VIP (2021-02-19 a 2021-02-22).

Métricas de ajuste (Vacunatorio VIP)

Para el caso del Vacunatorio VIP, las métricas de ajuste se observan en la Figura 6.6, proporcionando una visión cuantitativa del rendimiento de cada modelo.



(a) R^2 por modelo.



(b) MSE por modelo.

Figura 6.6: Métricas de ajuste para los modelos en el caso de estudio Vacunatorio VIP enmarcadas en la ventana del 19 al 22 de febrero de 2021.

6.1. Análisis de los resultados de los modelos epidemiológicos

6.1.1. Análisis del rendimiento cuantitativo

Las métricas de ajuste (R^2 y MSE), presentadas en las Figuras 6.2, 6.4 y 6.6, ofrecen una visión cuantitativa clara del rendimiento de cada modelo.

- **Dominio consistente del modelo SEIZ:** En los tres casos de estudio, el modelo **SEIZ** y sus variantes superan de manera contundente a los demás. Sus valores de R^2 son consistentemente los más cercanos a 1.0, como se observa en las subfiguras 6.2a, 6.4a y 6.6a.
- **Fracaso de los modelos simples:** El modelo **SEI** muestra un rendimiento extremadamente pobre, llegando a tener un R^2 negativo en el caso de SputnikV (Figura 6.2a). Un R^2 negativo significa que el modelo se ajusta peor a los datos que una simple línea horizontal, lo que equivale a un fracaso en la modelización. El modelo **SIR**, aunque mejor que el **SEI**, es consistentemente superado por la familia **SEIZ**.
- **Análisis del Error (MSE):** Los gráficos de **MSE** (como en la Figura 6.2b) confirman estos hallazgos. El MSE de los modelos **SEIZ** es órdenes de magnitud más bajo que el de los modelos **SIR** y **SEI**.

6.1.2. Análisis del ajuste cualitativo

Los gráficos que comparan las curvas de los modelos con los datos reales (Figuras 6.1, 6.3 y 6.5) permiten un análisis visual que complementa las métricas.

- **Captura de la curva empírica:** Los modelos de la familia **SEIZ** logran replicar con gran fidelidad la cascada de información empírica. Capturan bien la fase inicial de crecimiento lento, la fase de aceleración exponencial y, crucialmente, la fase de saturación final donde la tasa de nuevos “infectados” disminuye.
- **Incapacidad de los modelos simples para modelar la saturación:** El modelo **SIR** (por ejemplo, la curva verde en la Figura 6.1) y el **SEI** fallan visiblemente en capturar la dinámica completa. A menudo, o bien subestiman el crecimiento inicial o no logran modelar correctamente la fase de saturación, desviándose significativamente de la curva de datos reales en las etapas posteriores.
- **Importancia de los estados “Expuesto” (E) y “Escéptico” (Z):** El éxito del modelo **SEIZ** sugiere que la inclusión de un estado de latencia (**Expuesto**) y un estado de resistencia o abandono (**Escéptico**) es fundamental para modelar con precisión la difusión de información en redes sociales. El estado **E** captura el retraso entre ver la información y compartirla, mientras que el estado **Z** explica por qué no toda la población susceptible termina “infectada”, reflejando la saturación del interés o el escepticismo activo.

6.1.3. Interpretación del rendimiento superior del modelo SEIZ sin ρ

Un patrón recurrente en los tres casos de estudio es el rendimiento sistemáticamente superior de la variante del modelo **SEIZ sin ρ** . Como se observa en las métricas de ajuste (por ejemplo, en la Figura 6.2, el MSE de **SEIZ NO RHO** es el más bajo), esta versión simplificada supera consistentemente al modelo SEIZ completo y a la variante sin ε .

Para interpretar mejor esto, recordemos qué representa el parámetro ρ . Este parámetro modela un mecanismo de contagio secundario: la tasa de contacto entre un individuo en estado **Expuesto (E)** con un individuo en estado **Infectado (I)**.

Al anular este parámetro ($\rho = 0$), el modelo **SEIZ NO RHO** asume que la única vía para que un individuo expuesto se convierta en infectado es a través de una tasa de transición espontánea ε , que representa un proceso de deliberación interna e independiente. Esta simplificación parece ajustarse de manera óptima a la naturaleza de los datos empíricos utilizados.

Por lo tanto, el modelo, no puede capturar el matiz de si un usuario fue “presionado” a tuitear por la exposición repetida. Simplemente registra el resultado final: el primer tweet. El modelo **SEIZ NO RHO**, al eliminar precisamente ese mecanismo de transición por contacto con un individuo en estado **Infectado (I)**, parecería estar capturando mejor la dinámica.

6.2. Resultados de la red de *retweets*

6.2.1. Distribución de grados

Las distribuciones de grado de entrada y salida (Figuras 6.7 y 6.8) revelan la topología de la red. Ambos histogramas, presentados en escala logarítmica en el eje de frecuencia, muestran una distribución de cola pesada. Esto indica que la mayoría de los usuarios tienen una actividad muy baja (pocos *retweets* hechos o recibidos), mientras que un número reducido de nodos concentra una cantidad desproporcionadamente alta de conexiones. Este patrón es característico de las redes libres de escala, comunes en los sistemas sociales.

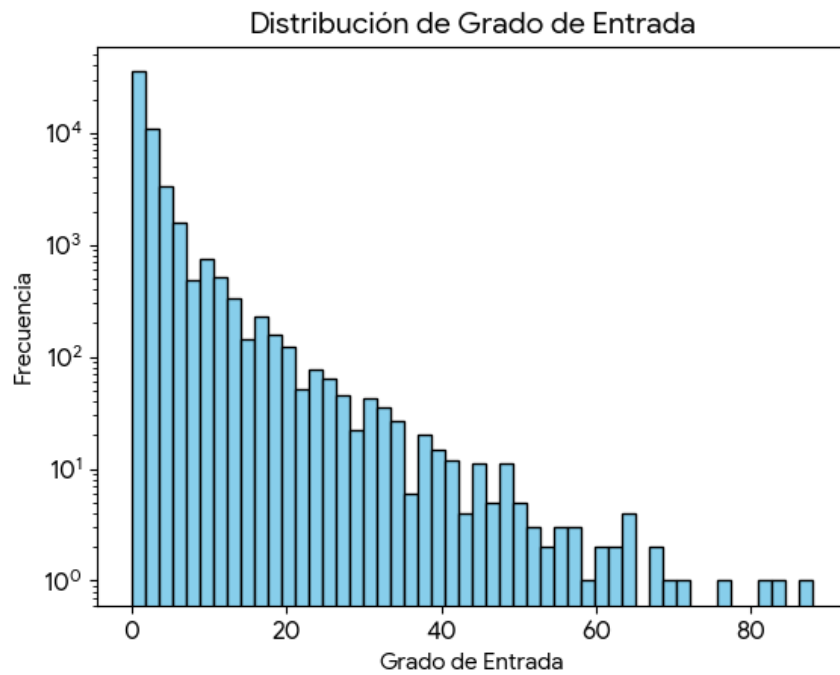


Figura 6.7: Distribución de grado de entrada (in-degree) para la red completa. El grado de entrada representa la cantidad de veces que un usuario fue retuiteado por otros.

La mayoría de los usuarios son retuiteados pocas veces, mientras que unos pocos nodos alcanzan altos niveles de popularidad.

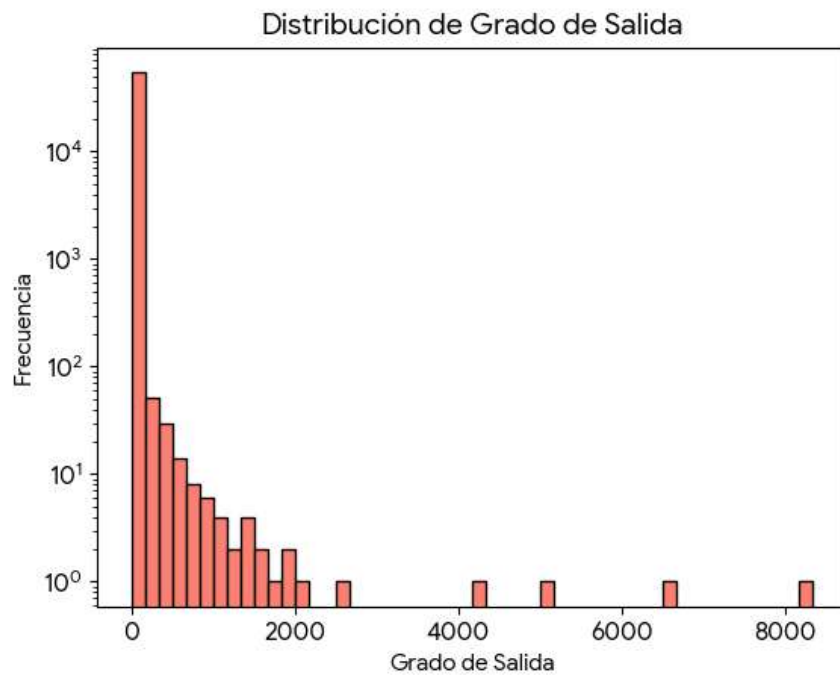


Figura 6.8: Distribución de grado de salida (out-degree) para la red completa. El grado de salida representa la cantidad de veces que un usuario retuiteó a otros.

La mayoría de los usuarios realizan pocos *retweets*, mientras que un pequeño grupo de usuarios es extremadamente activo en la difusión. Recordamos que la red está construida a partir de una extensión del conjunto de datos del "Vacunatorio VIP".

6.2.2. Análisis de centralidad de intermediación

La distribución de la centralidad de intermediación (Figura 6.9) también sigue un patrón de cola pesada, similar a las distribuciones de grado. La gran mayoría de los nodos tiene una centralidad cercana a cero, lo que significa que no actúan como puentes en la red. Por el contrario, un grupo muy selecto de nodos posee una alta centralidad, indicando que son fundamentales para el flujo de información entre diferentes clústeres de la red.

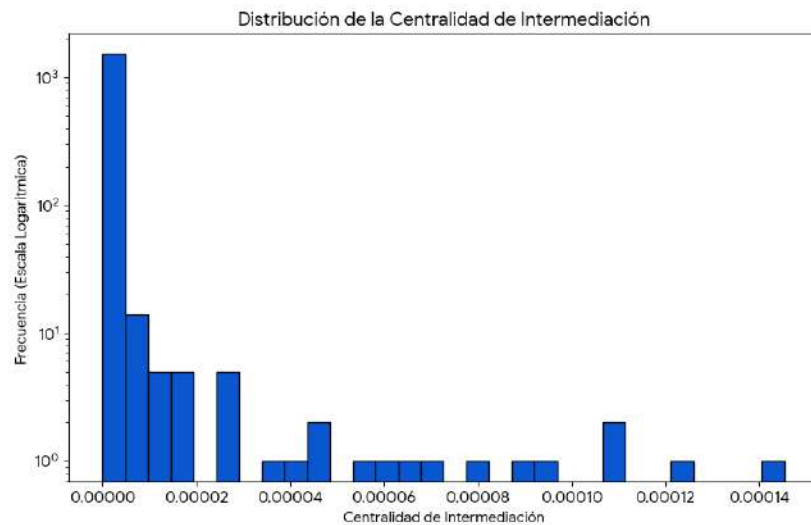


Figura 6.9: Distribución de la centralidad de intermediación en la red completa de *retweets* para el caso de estudio "Vacunatorio VIP"

6.2.3. Comparación entre métricas de la red dos a dos

Para entender mejor los roles de los usuarios, se compararon los valores de las distintas propiedades de cada usuario en la red completa.

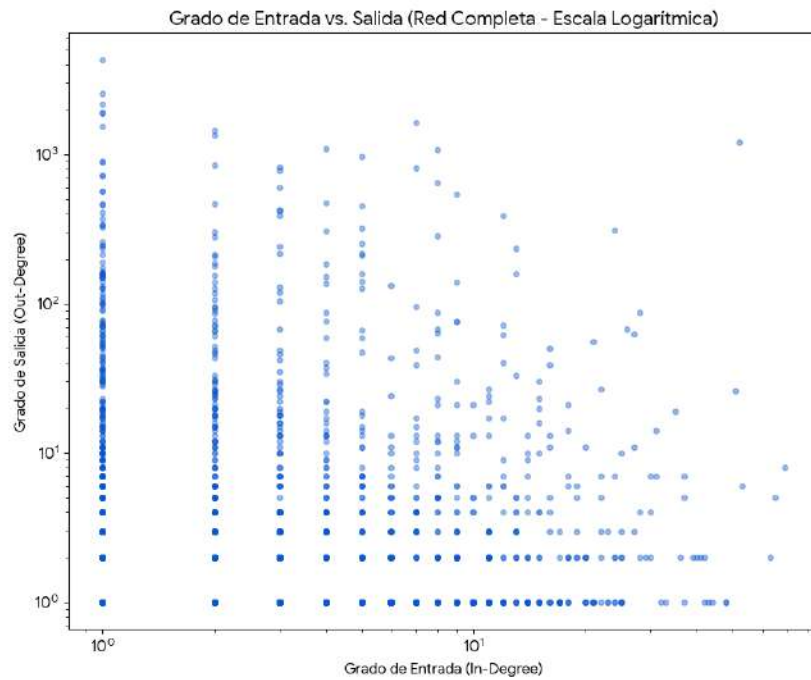


Figura 6.10: Comparación entre grado de entrada y grado de salida en la red completa de *retweets* para el caso de estudio "Vacunatorio VIP". Cada punto representa un usuario y está ubicado en la grilla según su grado de entrada y su grado de salida

No se observa una correlación directa, indicando que ser popular (alto in-degree) no implica necesariamente ser un gran difusor (alto out-degree) y viceversa.

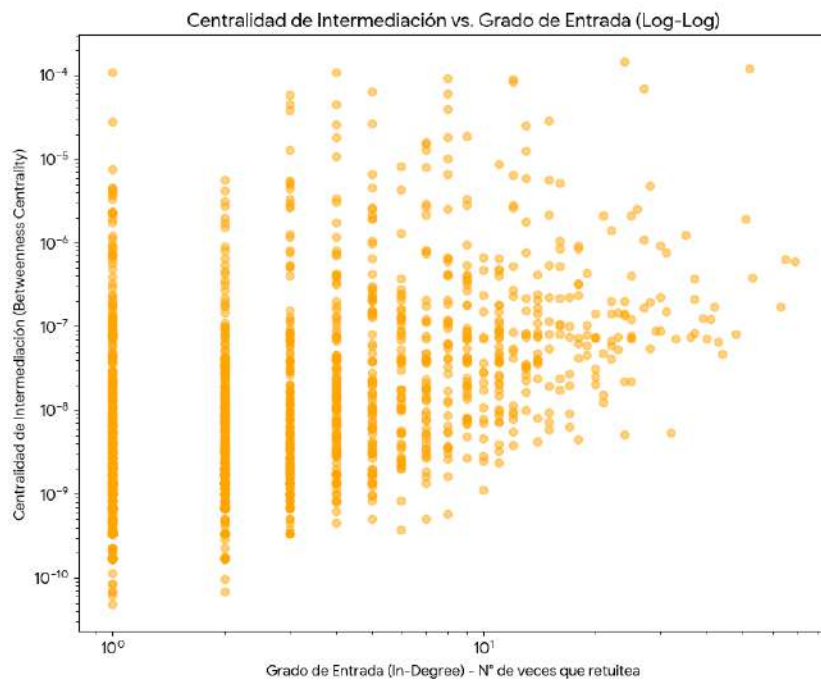


Figura 6.11: Centralidad de intermediación vs. grado de entrada en la red completa de *retweets* para el caso de estudio "Vacunatorio VIP"

Aunque hay nodos con alto grado de entrada y alta intermediación, muchos nodos influyentes (alto in-degree) no son puentes importantes en la red.

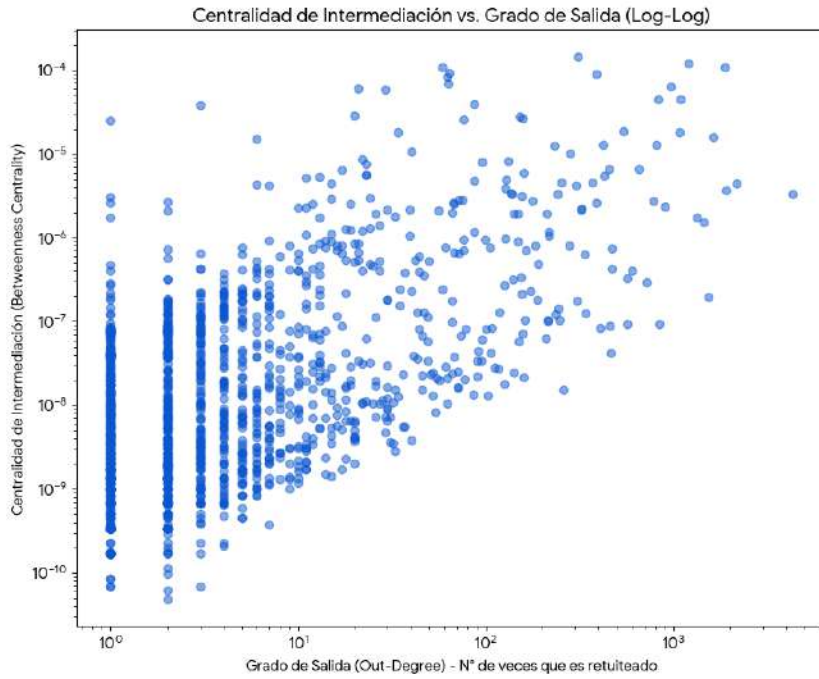


Figura 6.12: Centralidad de intermediación vs. grado de salida en la red completa de *retweets* para el caso de estudio "Vacunatorio VIP"

Similarmente, ser un gran difusor (alto out-degree) no garantiza una alta centralidad de intermediación.

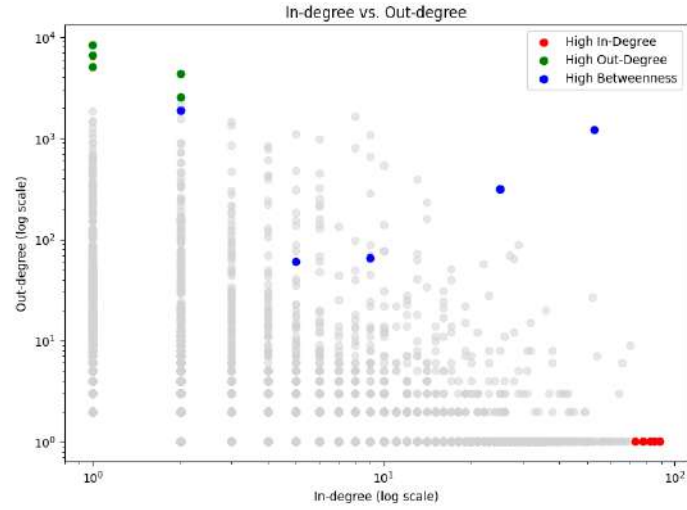
6.2.4. Identificación de nodos clave

Finalmente, se identificaron los nodos con los valores más altos en cada una de las tres métricas para visualizar sus roles distintivos dentro de la red (Figuras 6.13a, 6.13b y 6.13c).

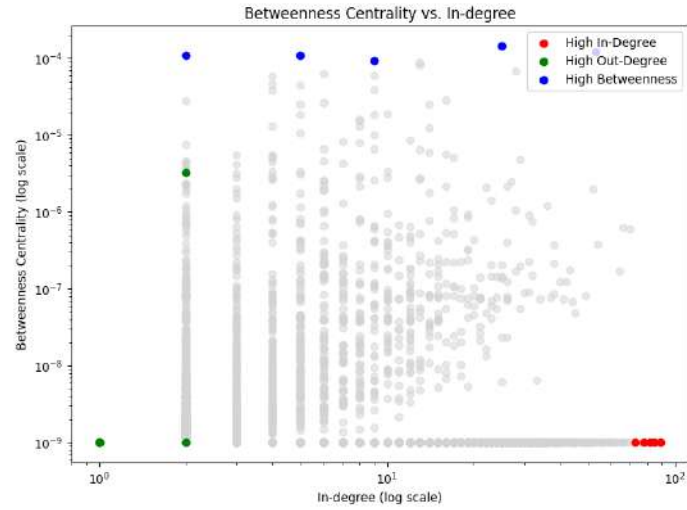
Métrica	ID del nodo	Valor
In-Degree (Propagadores)	4442645356	88
	909358892	84
	343636885	81
	566854299	77
	290671142	72
Out-Degree (Originadores)	54414081	8344
	33989170	6621
	24900072	5102
	159210352	4323
	862841604482224132	2550
Betweenness (Puentes)	3321734585	0.000145
	2564582235	0.000122
	2365779397	0.000110
	97094716	0.000109
	170505141	0.000093

Cuadro 6.1: Top 5 nodos según las métricas de centralidad.

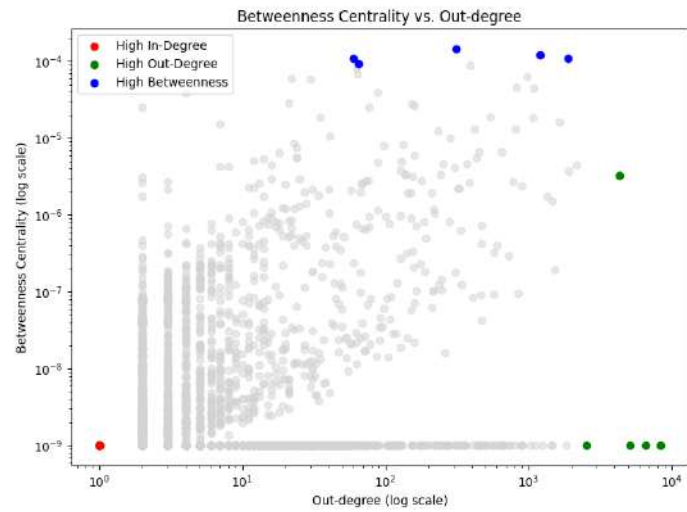
- **Nodos de alto in-degree (Rojo):** Son los *propagadores* más activos. Como se observa en la Figura 6.13a, tienen un grado de salida muy bajo, lo que significa que son fuentes de propagación de contenido.
- **Nodos de alto out-degree (Verde):** Son los *originadores* con más influencia. Tienen un grado de entrada relativamente bajo, lo que indica que su rol principal es generar contenido propagable.
- **Nodos de alta intermediación (Azul):** Son los *puentes*. Ocupan posiciones estructurales únicas. No siempre tienen el grado de entrada o salida más alto, pero son esenciales para conectar grupos de usuarios que de otro modo estarían aislados.



(a) Grado de salida vs. Grado de entrada con nodos clave resaltados. Se distingue claramente entre los mas retuiteadores (rojo) y los mas retuiteados (verde). Con los nodos de mayor centralidad con una distribución menos homogénea (azul)



(b) Centralidad de intermediación vs. Grado de entrada con nodos clave resaltados. Los nodos de mayor centralidad (azul), los nodos que más fueron retuiteados (verde) y los nodos que más retuitearon (rojo).



(c) Centralidad de intermediación vs. Grado de salida con nodos clave resaltados. Los nodos de mayor centralidad (azul), los nodos que más fueron retuiteados (verde) y los nodos que más retuitearon (rojo).

6.3. Integración de la dinámica temporal y la estructura de red

Finalmente, para conectar la estructura estática de la red con su evolución temporal, se superpuso la aparición de los nodos clave (identificados en la Tabla ??) sobre la curva de crecimiento de usuarios únicos acumulados. La Figura 6.14 compara la curva de datos reales (negro) con el ajuste de un modelo epidemiológico SEIZ (azul), que modela la propagación de información.

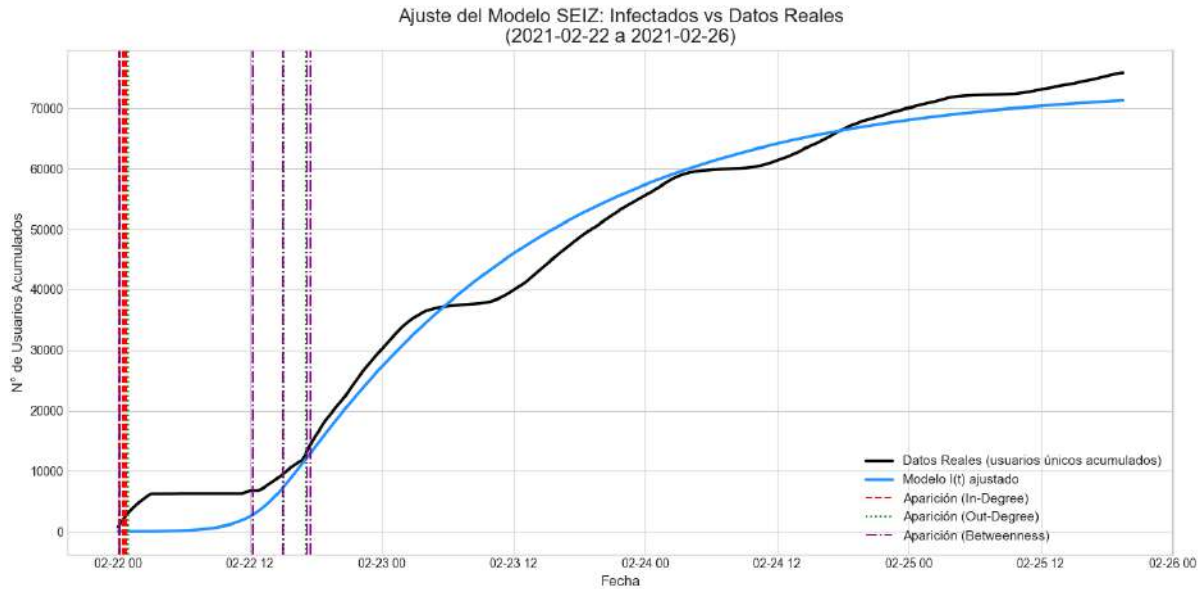


Figura 6.14: Ajuste del modelo SEIZ frente a los datos reales de usuarios acumulados. Las líneas verticales indican el momento exacto en que aparecieron por primera vez los nodos clave identificados: propagadores (rojo), originadores (verde) y puentes (morado).

El análisis de esta figura revela el rol temporal distintivo de cada tipo de nodo en la cascada de información:

- **Fase de inicio (propagadores y esparcidores):** Las líneas verticales rojas (alto in-degree) y verdes (alto out-degree) se concentran masivamente en las primeras horas del período analizado (22 de febrero). Esto sugiere que los *propagadores* y los mayores *originadores* son fundamentales para **iniciar** la difusión. Actúan como las “semillas” que introducen y dan el primer impulso a la información en la red. Sin embargo, su aparición por sí sola no desata la fase de crecimiento exponencial.
- **Fase de aceleración (puentes):** Las líneas verticales moradas, que marcan la aparición de los nodos con alta centralidad de intermediación (*puentes*), aparecen más tarde. Notablemente, su aparición coincide con el punto de inflexión de la curva de datos reales, justo cuando la difusión abandona el estancamiento inicial y entra en una fase de **crecimiento rápido y viral**.

Capítulo 7

Conclusiones y próximos pasos

Este trabajo se propuso analizar la dinámica de propagación de tendencias en *Twitter* mediante una doble aproximación: el modelado temporal con herramientas epidemiológicas y el análisis estructural de la red de interacciones. Los resultados obtenidos no solo validaron la superioridad de ciertos modelos, sino que también revelaron la riqueza que emerge al cruzar ambos enfoques.

Uno de los hallazgos más significativos fue la robustez y consistencia del modelo SEIZ. Su capacidad para predecir la curva de infectados se mantuvo notablemente estable ante variaciones del parámetro N , logrando un coeficiente de determinación (R^2) consistentemente cercano a 1. Este resultado sugiere que la fortaleza del modelo radica en su compartimento de escépticos (Z). Más que representar únicamente a usuarios que rechazan activamente la información, este compartimento parece capturar a la vasta mayoría de la población que nunca llega a interactuar con la tendencia. De esta manera, el modelo SEIZ refleja de forma implícita la naturaleza algorítmica de *Twitter*, donde el contenido solo alcanza a una fracción del total de usuarios, evitando así la predicción irreal de una “infección” total. El estado Z actúa como un sumidero poblacional que modela de forma efectiva los límites de la viralización.

Por otro lado, el análisis de la red de *retweets* del caso “Vacunatorio VIP” permitió enriquecer la interpretación de la curva de usuarios acumulados. Se observó que los nodos con mayor centralidad de intermediación aparecieron en un punto de inflexión clave del proceso de viralización. Este hallazgo es fundamental, ya que expone una de las limitaciones inherentes de los modelos epidemiológicos clásicos: su supuesto de homogeneidad poblacional. La evidencia de la red demuestra que no todos los individuos son iguales; ciertos nodos actúan como catalizadores críticos. Esto abre la puerta a un ida y vuelta, donde la estructura de la red informa al modelo temporal y viceversa.

Es destacable también que, si bien la red se construyó únicamente a partir de interacciones de *retweet*, sus propiedades estructurales demostraron tener una fuerte influencia predictiva sobre la dinámica global. La curva de usuarios acumulados incluye a todos los que tuitearon, no solo a los que participaron en la red de *retweets*. El hecho de que los nodos clave de esta red expliquen momentos críticos de la curva general valida el uso de la red de *retweets* como un estimador del flujo de influencia en *Twitter*.

7.1. Trabajos futuros y próximos pasos

A partir de los resultados y las limitaciones identificadas, se proponen las siguientes líneas de investigación futura:

- **Modelos epidemiológicos avanzados:** Explorar modelos más complejos que incorporen una mayor granularidad, como el modelo “Cognition-Driven SEIZ” [24]. Este tipo de modelos distingue entre las diferentes formas de interacción en *Twitter* (como *retweets*, citas o respuestas), asignando diferentes probabilidades y dinámicas a cada una, lo que podría capturar de forma más fiel la complejidad del comportamiento del usuario.
- **Integración de parámetros en modelos de red:** Utilizar los parámetros de salida de los modelos epidemiológicos (como las tasas de contagio o escepticismo) para construir redes de interacción ponderadas. En lugar de tratar todas las conexiones como iguales, los pesos de las aristas podrían estar modulados por estos parámetros, creando así un modelo de red que refleje la dinámica temporal observada.
- **Colaboración interdisciplinaria:** Como punto final, este trabajo subraya la necesidad imperativa de abordar estos problemas desde una perspectiva interdisciplinaria. La colaboración con expertos en Sociología, Lingüística y Neurociencia Cognitiva es esencial para comprender no solo el “qué” y el “cómo” de la propagación, sino también el “por qué” subyacente al comportamiento humano en las redes sociales.

Apéndice A

Análisis de sensibilidad al parámetro N : metodología y justificación

La evaluación de los modelos no se limita a su capacidad de ajuste (medida por métricas como el MSE o el R^2), sino que debe complementarse con un análisis de sensibilidad estructural. En este caso, nos enfocamos en el parámetro N , que representa el tamaño total de la población susceptible —es decir, la cantidad total de usuarios que podrían haber estado expuestos a la desinformación—. Este valor no se observa directamente en los datos, ya que solo se dispone de una muestra parcial (los tweets recolectados), por lo que se trata de un parámetro desconocido que debe ser asumido.

A.1. Resultados de parámetros por caso de estudio

Visualización de estabilidad de parámetros frente a variaciones de N (SputnikV)

A continuación se presentan los gráficos de estabilidad de parámetros para cada modelo ajustado al conjunto de datos de SputnikV. Es importante observar que este análisis se realiza haciendo variaciones al parámetro desconocido N .

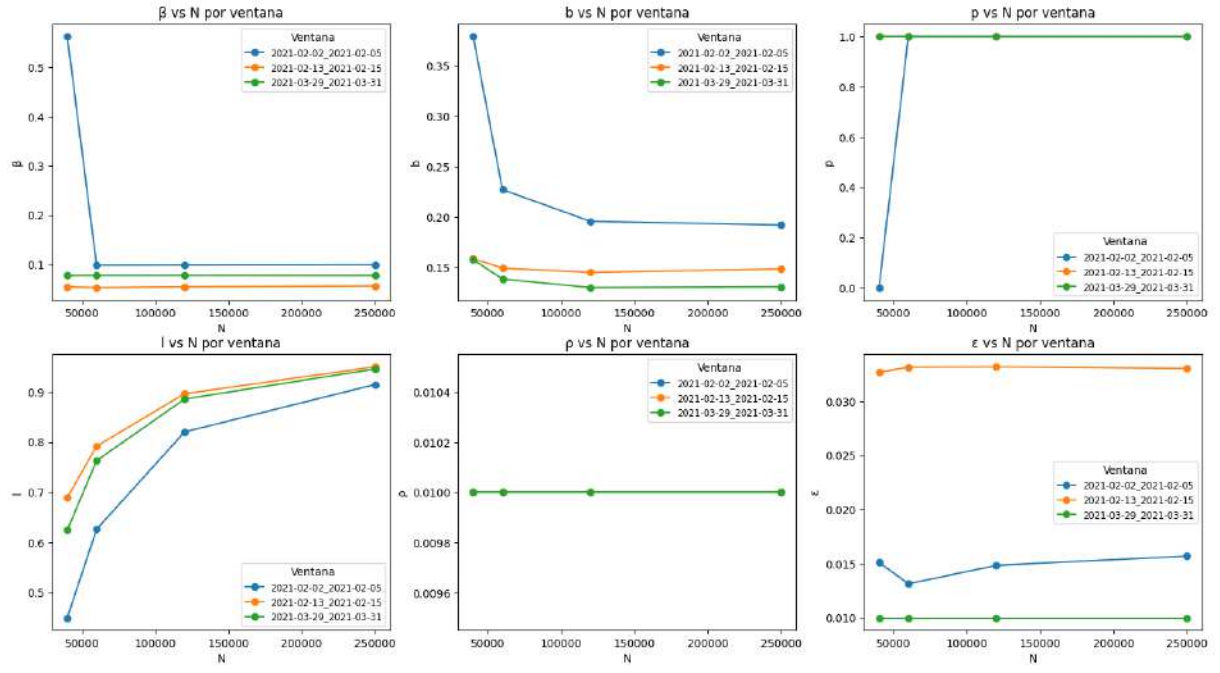


Figura A.1: Estabilidad de parámetros para el modelo SEIZ frente a variaciones de N (SputnikV).

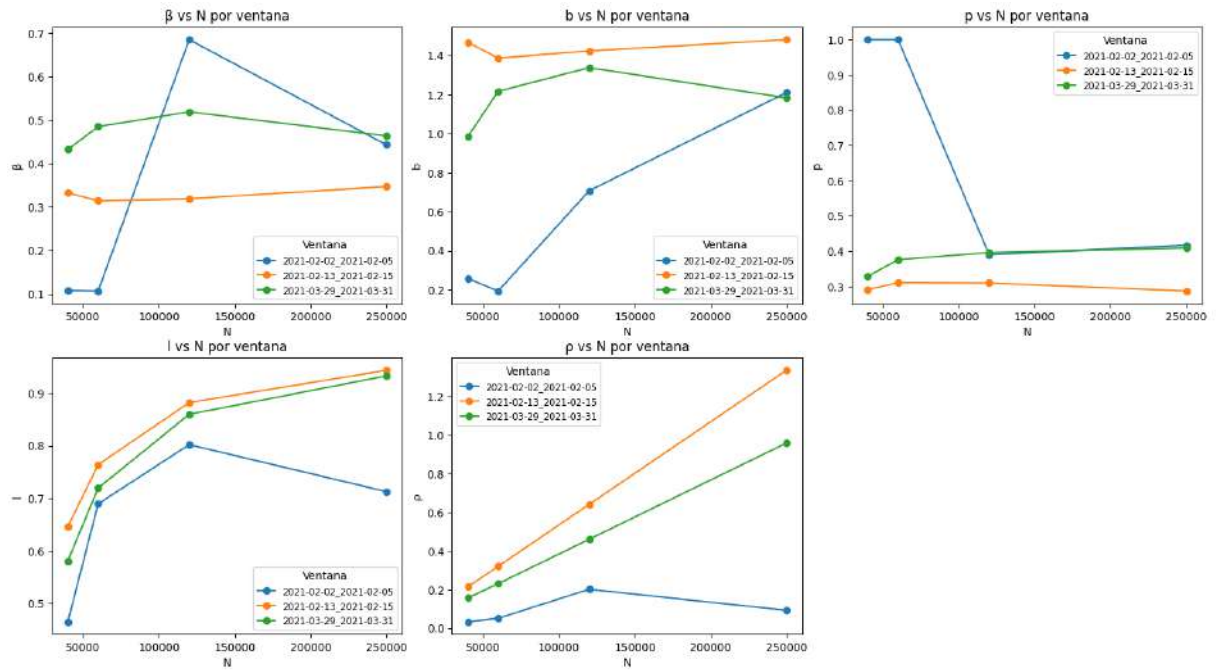


Figura A.2: Estabilidad de parámetros para el modelo SEIZ NO EPS frente a variaciones de N (SputnikV).

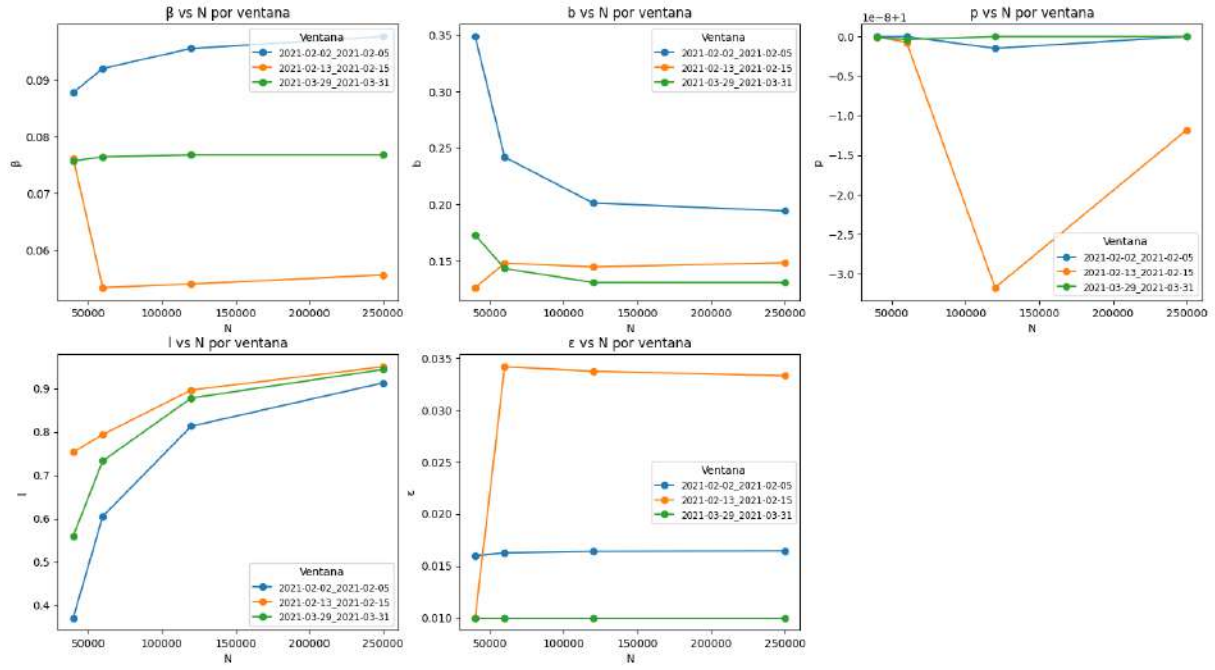


Figura A.3: Estabilidad de parámetros para el modelo SEIZ NO RHO frente a variaciones de N (SputnikV).

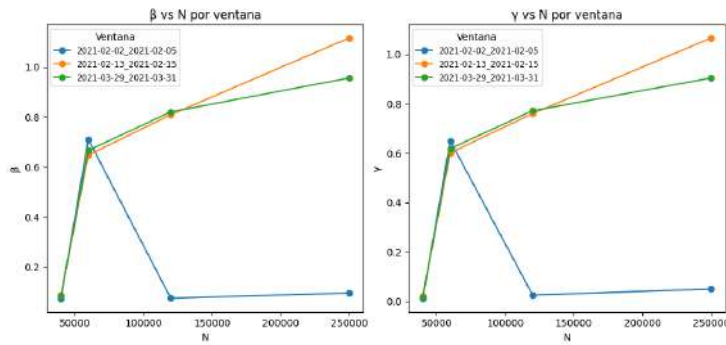


Figura A.4: Estabilidad de parámetros para el modelo SIR frente a variaciones de N (SputnikV).

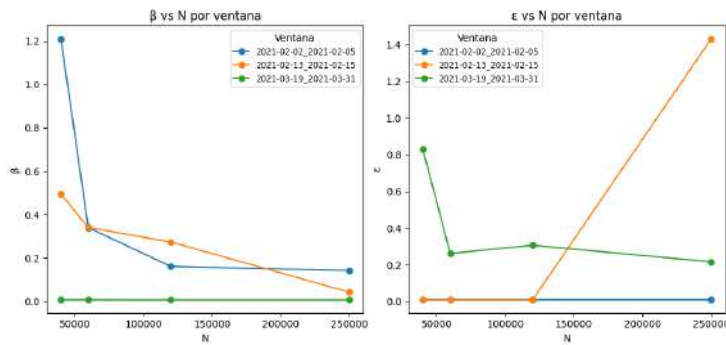


Figura A.5: Estabilidad de parámetros para el modelo SEI frente a variaciones de N (SputnikV).

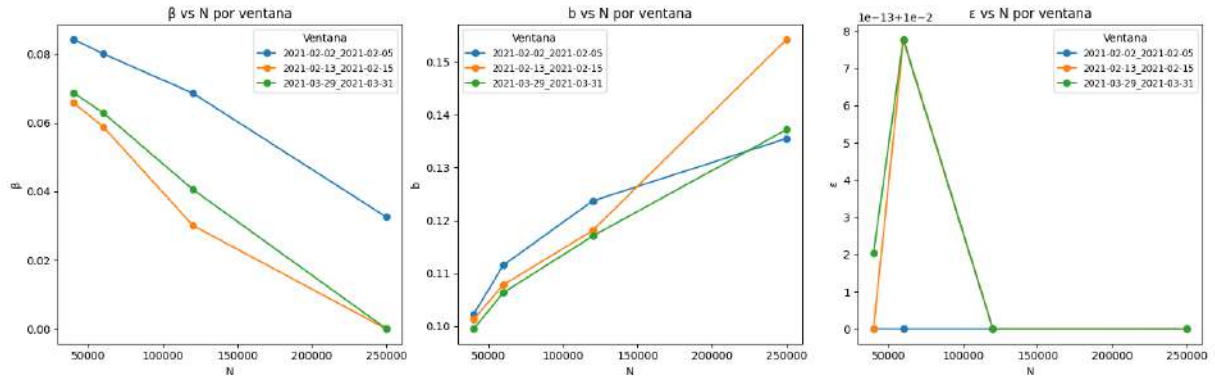


Figura A.6: Estabilidad de parámetros para el modelo SIZ frente a variaciones de N (SputnikV).

Visualización de estabilidad de parámetros frente a variaciones de N (Astra-Zeneca)

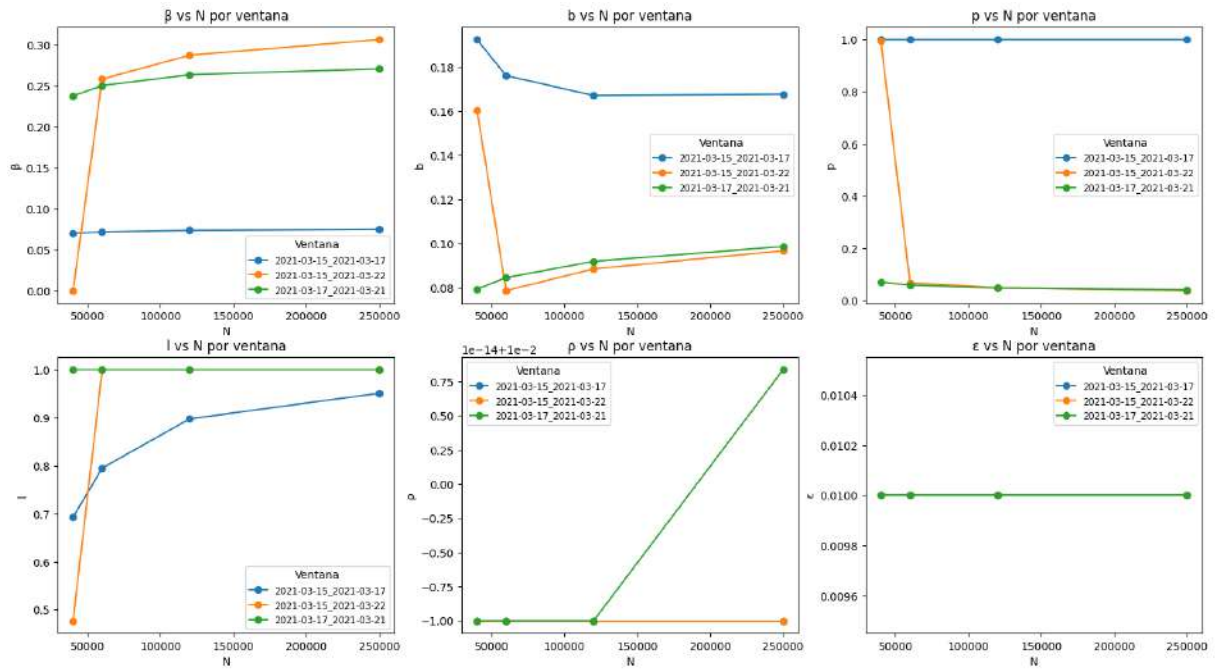


Figura A.7: Estabilidad de parámetros para el modelo SEIZ frente a variaciones de N (Astra-Zeneca).

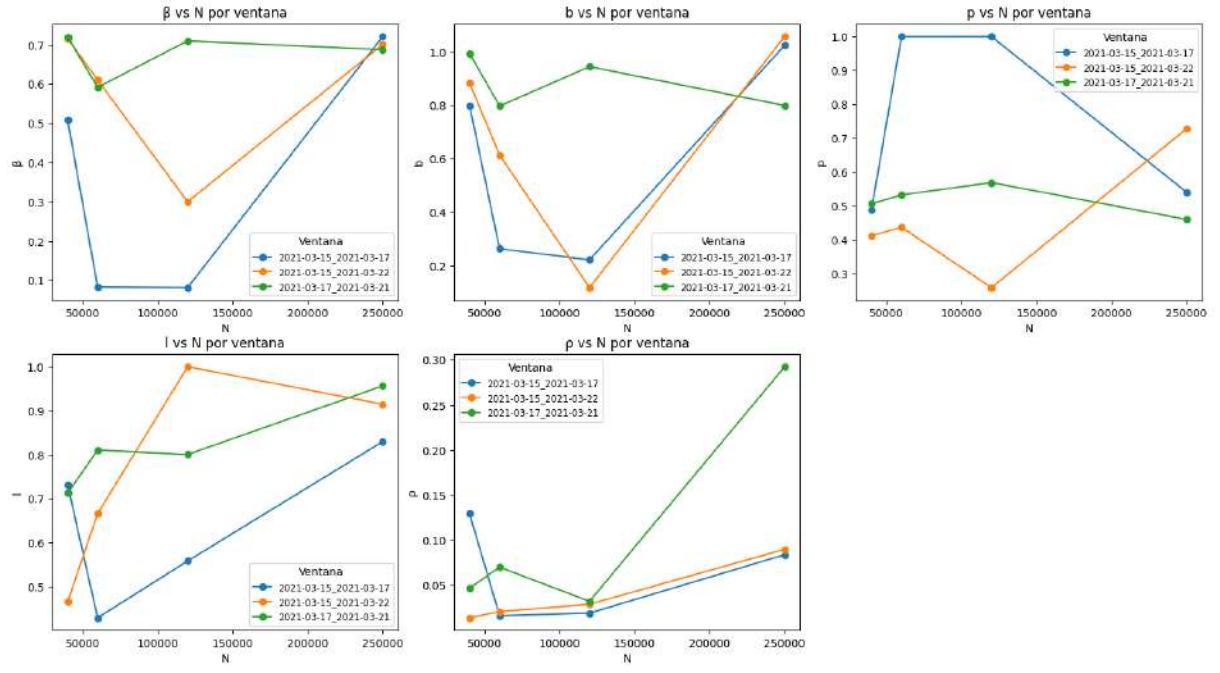


Figura A.8: Estabilidad de parámetros para el modelo SEIZ NO EPS frente a variaciones de N (AstraZeneca).

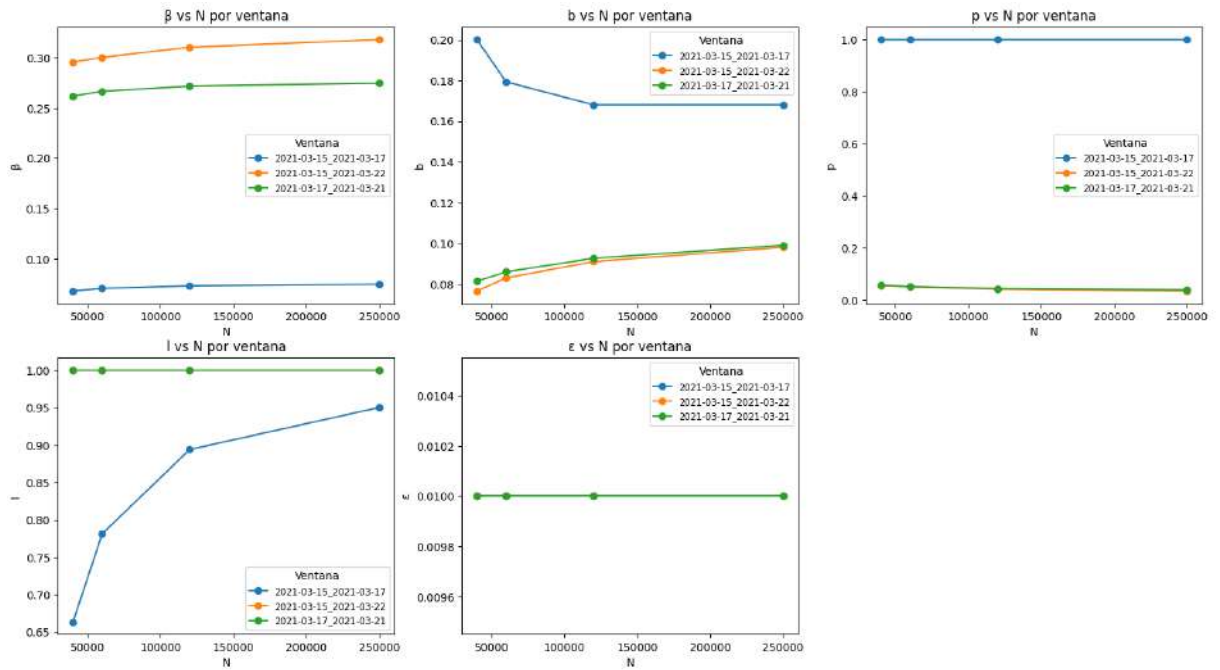


Figura A.9: Estabilidad de parámetros para el modelo SEIZ NO RHO frente a variaciones de N (AstraZeneca).

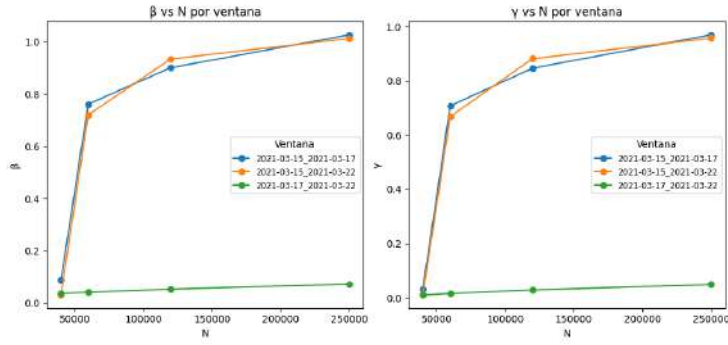


Figura A.10: Estabilidad de parámetros para el modelo SIR frente a variaciones de N (Astra-Zeneca).

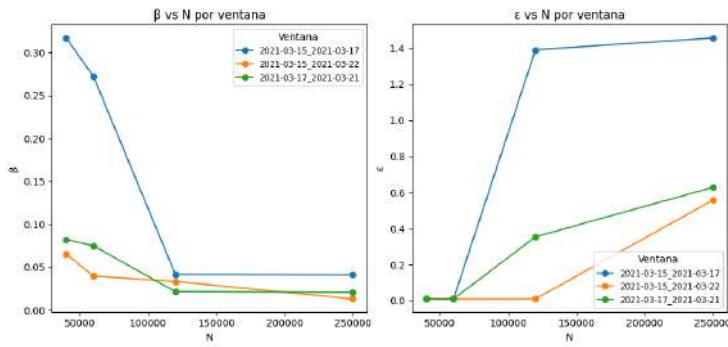


Figura A.11: Estabilidad de parámetros para el modelo SEI frente a variaciones de N (Astra-Zeneca).

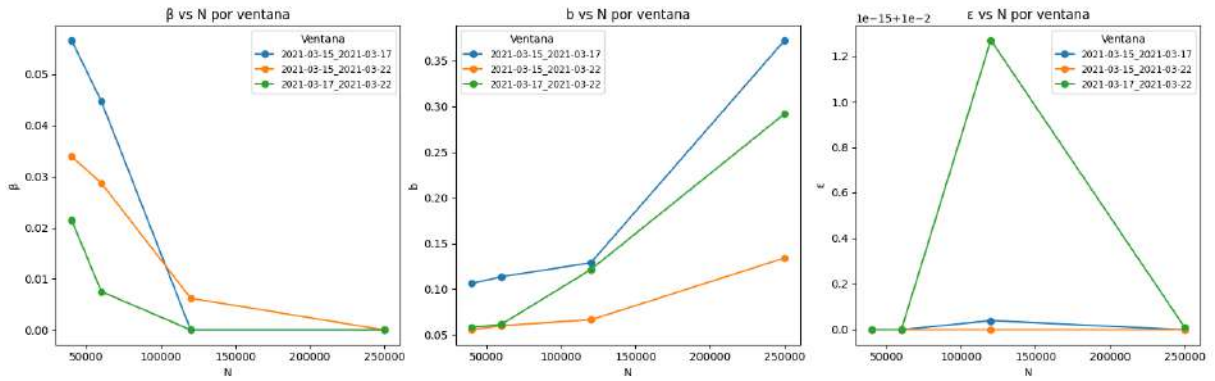


Figura A.12: Estabilidad de parámetros para el modelo SIZ frente a variaciones de N (Astra-Zeneca).

Visualización de estabilidad de parámetros frente a variaciones de N (Vacunatorio VIP)

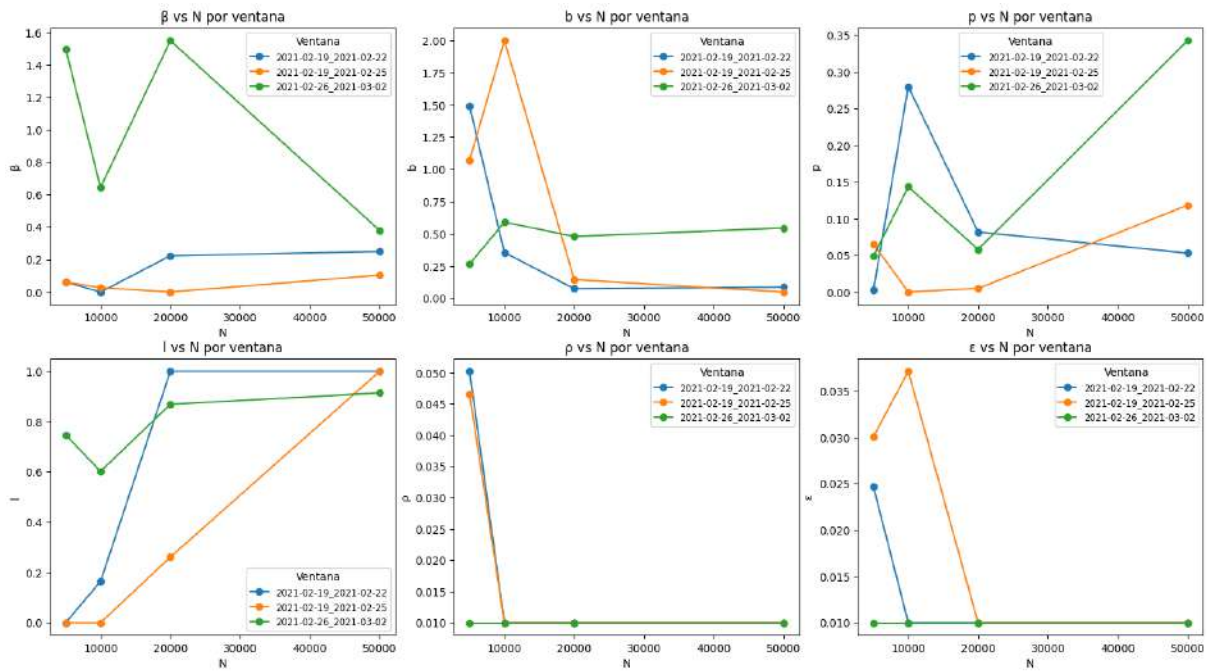


Figura A.13: Estabilidad de parámetros para el modelo SEIZ (Vacunatorio VIP).

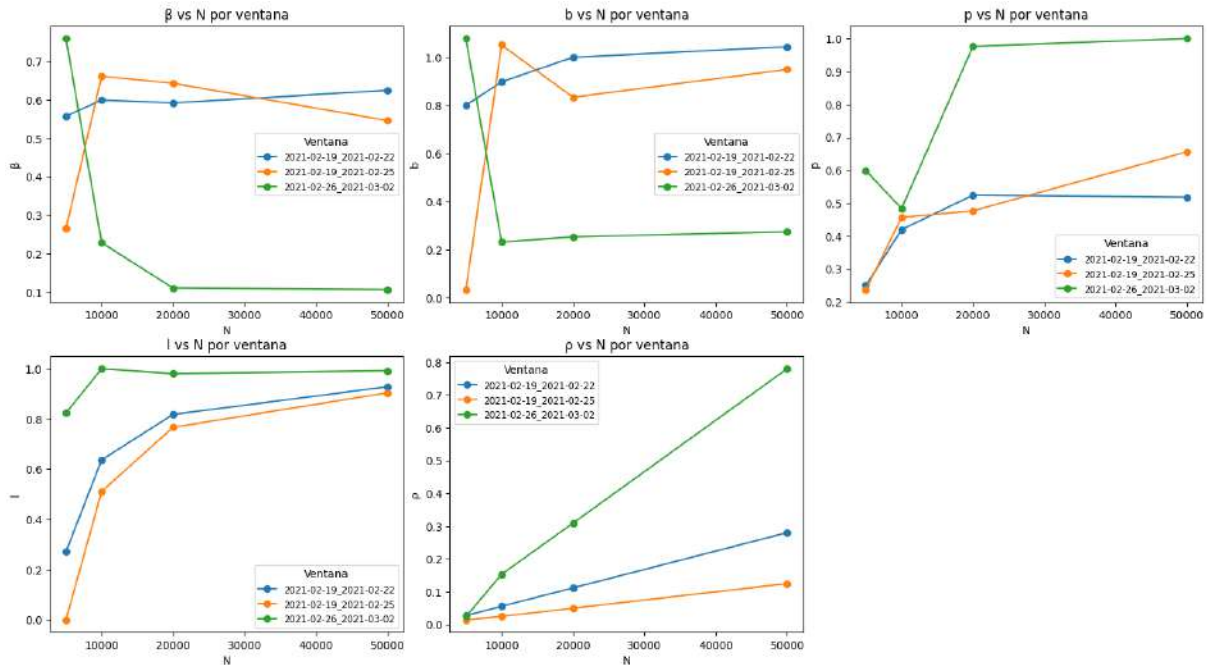


Figura A.14: Estabilidad de parámetros para el modelo SEIZ NO EPS frente a variaciones de N (Vacunatorio VIP).

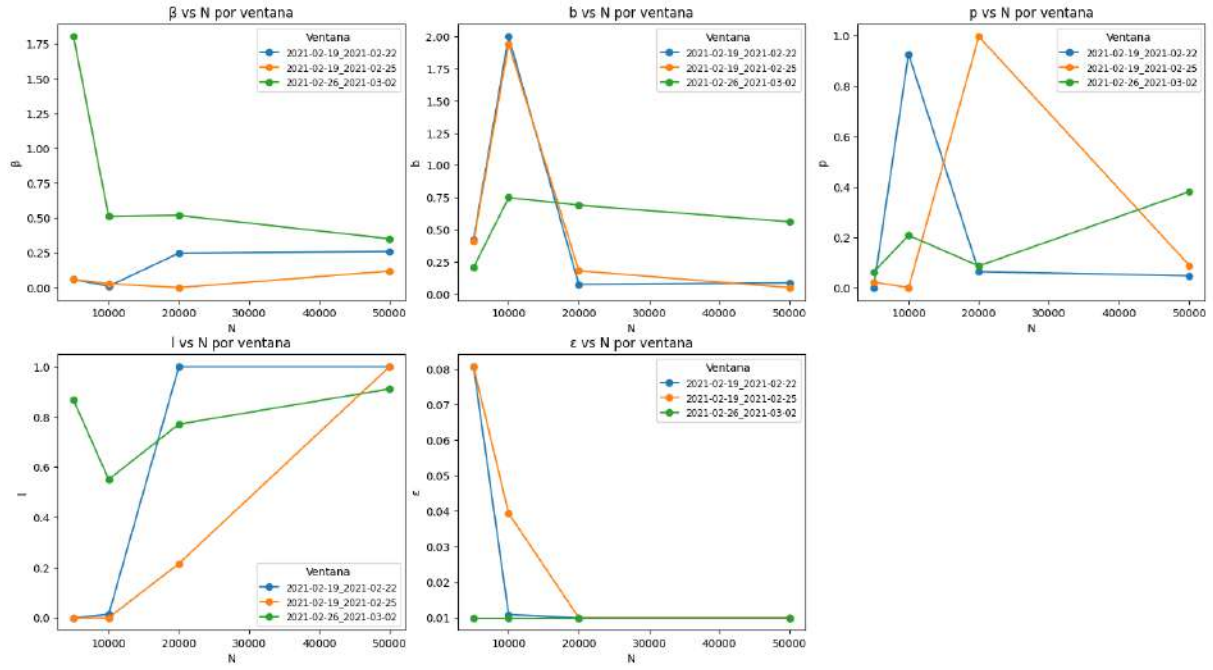


Figura A.15: Estabilidad de parámetros para el modelo SEIZ NO RHO frente a variaciones de N (Vacunatorio VIP).

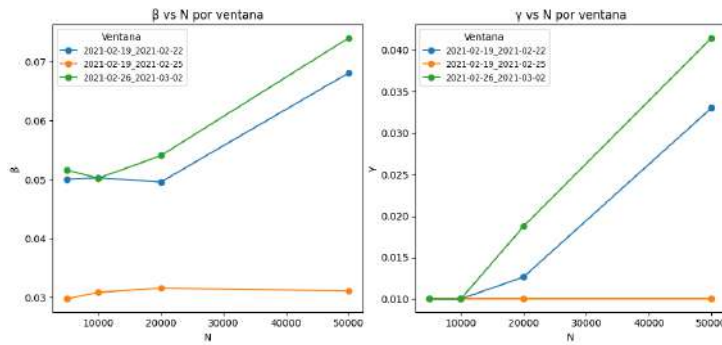


Figura A.16: Estabilidad de parámetros para el modelo SIR (Vacunatorio VIP).

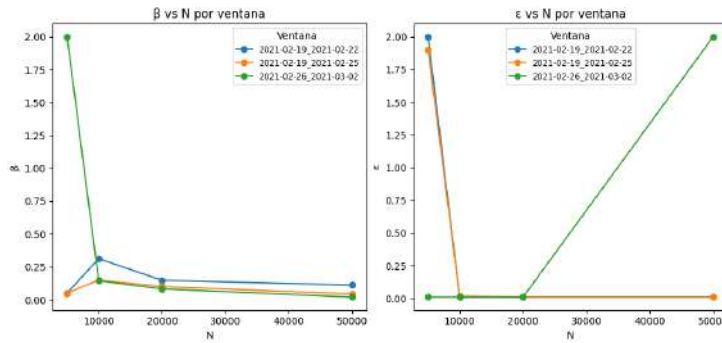


Figura A.17: Estabilidad de parámetros para el modelo SEI frente a variaciones de N (Vacunatorio VIP).

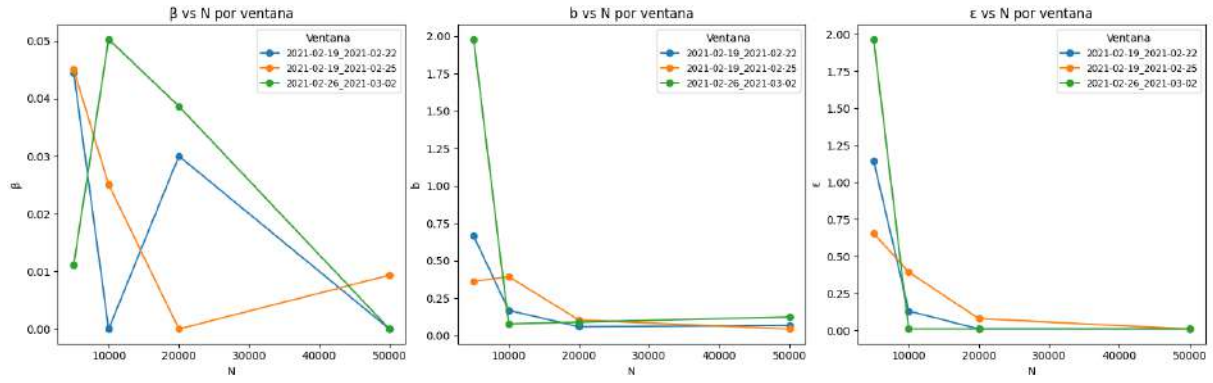


Figura A.18: Estabilidad de parámetros para el modelo SIZ frente a variaciones de N (Vacunatorio VIP).

A.1.1. Análisis de parámetros de tasa de contacto (dependientes de N):

Las tasas como β están ligadas a N por la formulación del modelo ($\beta \cdot S \cdot I/N$). Para replicar el mismo número de infecciones reales con un N mayor, el modelo debe reducir β . Esta tendencia decreciente es visible en la Figura A.1 (β vs N) y no representa un fallo, sino un ajuste lógico del modelo a diferentes escalas.

Sin embargo, se podría esperar que, al aumentar N , la probabilidad de volverse escéptico (l) o la tasa de contacto con escépticos (b) aumentaran para desviar a más individuos de la ruta de infección y así mantener la curva de infectados ajustada a los datos empíricos.

No obstante, los resultados muestran comportamientos erráticos y no parece haber un patrón claro y universal en todos los casos de estudio. Por ejemplo, mientras que la tasa de contacto con escépticos (b) muestra una tendencia ligeramente creciente para el caso SputnikV (Figura A.1), en el caso de AstraZeneca (Figura A.7) su comportamiento es más volátil y no sigue un patrón claro. De manera similar, para el evento del "Vacunatorio VIP" (Figura A.13), la tasa b muestra una fuerte caída inicial antes de estabilizarse, un comportamiento distinto a los otros dos casos.

Esta falta de un patrón consistente podría atribuirse a la alta sensibilidad del modelo al contexto específico de cada evento. La **elección de la ventana de tiempo** para el análisis es probablemente un factor crucial: una ventana corta podría no capturar completamente la fase de saturación, llevando al algoritmo de optimización a encontrar equilibrios de parámetros diferentes a los que encontraría en una ventana más larga. Esto sugiere que, si bien el modelo SEIZ es estructuralmente superior, la interpretación de sus parámetros se dificulta al tratarse de un modelo más complejo que otros modelos epidemiológicos estudiados.

Bibliografía

- [1] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018.
- [2] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236, May 2017.
- [3] World Health Organization. Infodemic management: a key component of the covid-19 global response, 2020.
- [4] D. J. Daley and D. G. Kendall. Epidemics and rumours. *Nature*, 204(4963):1118–1118, 1964.
- [5] Feng Jin, Evan Dougherty, Prateek Saraf, Yuxin Cao, and Naren Ramakrishnan. Epidemiological modeling of news and rumors on twitter. pages 1–9, 2013.
- [6] Maryam Maleki, Esther Mead, Mohammad Arani, and Nitin Agarwal. Using an epidemiological model to study the spread of misinformation during the black lives matter movement. Technical report, University of Arkansas at Little Rock, 2021. Preprint.
- [7] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36, 2017.
- [8] Liang Wu, Fred Morstatter, Kathleen M Carley, and Huan Liu. Misinformation in social media: A survey. *ACM SIGKDD Explorations Newsletter*, 21(2):80–90, 2019.
- [9] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.
- [10] vacunatoriovip. aproximación al estudio de los escándalos de corrupción en twitter. *Scielo*, 2024.
- [11] Editor de Latimes Español. Miles de argentinos protestan por privilegios de vacunación. Latimes (edición en español), 2021.
- [12] Deutsche Welle. Dimite ministro de salud de argentina por escándalo ”vacunas vip”. DW, 2021.
- [13] Capital 21 Web. Sputnik v muestra eficacia del 91.6 Capital 21, Feb 2021.
- [14] Associated Press. Argentina vacuna contra el covid-19 en estadios de fútbol. AP via Latimes (edición en español), Feb 2021.

- [15] France24. Con sobresaltos y polémicas avanza la vacunación contra el covid-19. France24 (español), Feb 2021.
- [16] ANMAT (Argentina). Preguntas frecuentes sobre la vacuna contra covid-19. Sitio oficial Gobierno Argentina, Feb 2021.
- [17] France24. En medio de la crisis sanitaria, venezuela recibe nuevo cargamento de sputnik v. France24 (español), Mar 2021.
- [18] Wikipedia. Sputnik v covid-19 vaccine. Wikipedia, 2025.
- [19] Oliver Holmes. Which european states have paused astrazeneca jabs due to clotting concerns? The Guardian, Mar 2021.
- [20] Al Jazeera. Which countries have stopped using astrazeneca’s covid vaccine? Al Jazeera, Mar 2021.
- [21] Allyson Chiu. Does the astrazeneca vaccine cause blood clots? Washington Post, Mar 2021. Consultado 24 jun 2025.
- [22] European Medicines Agency. Astrazeneca’s covid-19 vaccine: Ema finds possible link to very rare cases of unusual blood clots. EMA Press Release, Mar 2021.
- [23] PAHO / OMS. Declaración del subcomité sobre seguridad de vacunas astrazeneca. PAHO, Mar 2021.
- [24] Ece Çiğdem Mutlu, Amirarsalan Rajabi, and Ivan Garibay. Cd-seiz: Cognition-driven seiz compartmental model for the prediction of information cascades on twitter. *arXiv:2008.12723v1 [cs.SI]*, August 2020. A PREPRINT August 31, 2020.