

NUEVAS FRONTERAS EN CIENCIAS SOCIALES COMPUTACIONALES

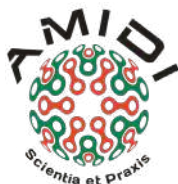
**Análisis del discurso,
la opinión y la
dinámica social**



**Antonio Aguilera
Ontiveros**
(Coordinador)

NUEVAS FRONTERAS EN CIENCIAS SOCIALES COMPUTACIONALES:

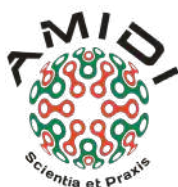
**Análisis del discurso,
la opinión y la dinámica
social**



NUEVAS FRONTERAS EN CIENCIAS SOCIALES COMPUTACIONALES:

**Análisis del discurso,
la opinión y la dinámica
social**

Antonio Aguilera Ontiveros
(Coordinador)



Este libro fue sometido a un proceso de dictamen por pares doble ciego, de acuerdo con las normas establecidas por el Comité Editorial de la editorial Academia Mexicana de Investigación y Docencia en Innovación (AMIDI).



Esta obra se encuentra bajo la licencia Atribución-No Comercial-Sin Derivadas 4.0 (CC BY-NC-ND 4.0), de Creative Commons. Usted puede descargar esta obra y distribuir en cualquier medio o formato dando crédito a los autores, pero no se permite su uso comercial ni la generación de obras derivadas.

Primera edición, 2025

© D.R. 2025, Academia Mexicana de Investigación
y Docencia en Innovación (AMIDI)
Av. Paseo de los Virreyes 920, Col. Virreyes Residencial
C.P. 45110, Zapopan, Jalisco, México
direccion@amidi.mx

Ilustración de portada: Simon Lee (licencia gratuita)
Sitio: unsplash.com



ISBN: 978-607-26875-6-1

Hecho en México
Made in Mexico

Contenido

Introducción	7
Antonio Aguilera Ontiveros	
Capítulo 1. Algoritmo de autorreflexión de agentes sociales para la formación de grupos de opinión a través de un modelo de dinámica de opinión de acuerdo relativo	13
Norma Leticia Abrica Jacinto y Evguenii Kurmyshev	
Capítulo 2. Identificando ecuaciones parsimoniosas que gobiernan la dinámica de modelos basados en agentes. Una primera aproximación	43
Jorge Zazueta Gutiérrez	
Capítulo 3. Diseminación cultural con aumento de presencia de medios de comunicación en el modelo de Axelrod	53
Sadan Josué López Morales, Ricardo Armando Gonzalez Silva y Mario Ignacio Gonzalez Silva	
Capítulo 4. Análisis de tópicos en comentarios de YouTube durante la campaña presidencial de México 2024: un estudio con LDA	75
Antonio Aguilera Ontiveros y Adela Martínez Galván	

Capítulo 5. Herramientas y técnicas computarizadas al servicio de la construcción de <i>corpus</i> y <i>datasets</i> . Desafíos metodológicos del análisis de discursos mediatizados en plataformas con base en Internet.....	121
Natalia Raimondo Anselmino, Irene Gindin, Alejandro Sartorio, Franco Cellone, Verónica Aráoz, Francisco J. Alomar	
Capítulo 6. Factores psicosociales en la dinámica de la calidad de vida. Análisis Cualitativo Comparado con Conjuntos Difusos (fs/QCA).....	149
Gabriela Sánchez Mondragón y Anthony Pérez Balcázar	

Introducción

Antonio Aguilera Ontiveros¹

En las últimas tres décadas, el campo de las ciencias sociales ha experimentado una profunda transformación gracias a la incorporación de métodos computacionales y simulaciones basadas en agentes (ABM, por sus siglas en inglés) (cf. Epstein, 1999; Banks, 2002; Macy y Willer, 2002; Heath *et al.* 2009), técnicas de minería de texto basadas en *Machine Learning* (cf. Hopkins y King, 2010; Grimmer y Stewart, 2013) y enfoques de indagación causalística basados en teoría de conjuntos comparativos (cf. Ragin, 2006; Fiss, 2011). Estos desarrollos permiten no solo explorar dinámicas colectivas a gran escala, sino también analizar discursos mediáticos y construir marcos analíticos que integran datos cuantitativos y cualitativos de manera inédita. Tal sinergia metodológica abre una nueva frontera para comprender fenómenos sociales complejos, desde la formación de opinión pública hasta la calidad de vida en contextos nacionales e internacionales. Todo lo anterior forma parte del repertorio de lo que se denomina las ciencias sociales computacionales (cf. Conte *et al.*, 2012).

Es de nuestro interés indagar en cómo, desde una perspectiva latinoamericana, podemos contribuir al conocimiento en nuestros países y nuestra lengua materna, el español, a que las ciencias sociales computacionales se difundan como una metodología robusta y concreta, y que pueda contribuir a la comprensión de nuestras realidades tanto sociales, económicas como políticas. Es por ello que presentamos este

1. Programa de Estudios Políticos e Internacionales, El Colegio de San Luis, A. C. ORCID: <https://orcid.org/0000-0003-1548-9956>. Correo electrónico: antonio.aguilera@colsan.edu.mx

libro a la comunidad científica, social e interdisciplinaria con el objeto de poner nuestro granito de arroz a la mejor comprensión de los fenómenos sociopolíticos de nuestra región. El volumen se estructura en torno a tres ejes temáticos que combinan teoría, método y aplicación:

1. Modelado y simulación de dinámicas sociales

La primera sección explora la generación de patrones macrosociales a partir de reglas microfundamentadas. Bonabeau (2002) ya señalaba que los modelos basados en agentes permiten simular sistemas humanos heterogéneos y revelar propiedades emergentes no evidentes a partir de análisis agregados. En este eje se incluyen los siguientes capítulos:

- Capítulo 1: un algoritmo de autorreflexión de agentes para la formación de grupos de opinión mediante un modelo de “acuerdo relativo”, que extiende los clásicos esquemas de interacción de Deffuant y Weisbuch.
- Capítulo 2: la identificación de ecuaciones parsimoniosas que aproximan la dinámica de ABM, aportando herramientas de reducción paramétrica.
- Capítulo 3: la adaptación del modelo de Axelrod (1997) para incorporar la influencia creciente de medios masivos, explicando cómo transmisiones culturales focalizadas modulan la polarización social.

2. Análisis de discurso y opinión pública digital

El auge de las plataformas en línea ha generado volúmenes masivos de texto que requieren métodos robustos de análisis temático y de sentimientos. La combinación de Latent Dirichlet Allocation (LDA) (Blei, Ng, y Jordan, 2003) junto con técnicas estadísticas —especialmente regresión logística— ofrece un marco para relacionar tópicos discursivos con actitudes ciudadanas. En este eje se despliegan los siguientes capítulos:

- Capítulo 4: un estudio de tópicos y sentimientos en comentarios de YouTube durante la campaña presidencial de México 2024,

que ilustra la potencialidad de LDA para mapear narrativas políticas y su asociación con el apoyo electoral.

- Capítulo 5: un compendio de herramientas computarizadas para la construcción de *corpus* y conjuntos de datos, incluyendo flujos de trabajo reproducibles para la recolección, limpieza y anotación de discursos mediáticos.

3. Métodos comparativos y calidad de vida

Finalmente, la discusión se amplía a nivel macro mediante técnicas conjunto-teóricas que combinan condiciones cualitativas y cuantitativas. El fs/QCA (fuzzy-set Qualitative Comparative Analysis) permite identificar configuraciones de factores psicosociales asociadas a variaciones en la calidad de vida entre países (Dusa, 2010). En este apartado está el capítulo:

Capítulo 6: aplica fs/QCA a un estudio comparado en 38 países, analizando cómo dimensiones como la cohesión social, la percepción de corrupción y el bienestar subjetivo configuran trayectorias de desarrollo humano.

Contribución y perspectivas

Este libro reúne perspectivas *complementarias* que, al dialogar, enriquecen la comprensión de fenómenos sociales contemporáneos. Por un lado, los modelos basados en agentes nos muestran cómo interacciones locales generan cambios globales; por otro, el análisis de texto digital revela las narrativas y emociones que movilizan a las sociedades. Finalmente, los métodos comparativos aportan claridad sobre las condiciones necesarias y suficientes para explicar la calidad de vida en diferentes contextos nacionales.

Al integrar estos tres ejes, aspiramos a ofrecer un *marco metodológico integral* para investigadores y estudiantes de ciencias sociales computacionales, así como para responsables de políticas públicas interesadas en evidencias cuantitativas y cualitativas de alto rigor. Además, los capítulos incluyen *implementaciones prácticas* (en Python/Mesa,

entornos de análisis de texto en Python y *software* de QCA gratuito) que facilitan la reproducibilidad y adaptación a nuevos casos de estudio.

Esperamos que esta obra inspire nuevas investigaciones interdisciplinarias —por ejemplo, la combinación de ABM con datos provenientes de redes sociales en tiempo real— y fomente la formación de *científicos sociales digitales* capaces de utilizar herramientas avanzadas para abordar los retos del siglo XXI: polarización política, difusión de información, cohesión cultural y bienestar social.

Referencias

- Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2), 203-226. <https://doi.org/10.1177/0022002797041002001>
- Banks, S. C. (2002). Agent-Based Modeling: A Revolution? *Proceedings of the National Academy of Sciences of the United States of America*, 99(suppl 3), 7199-7200. <https://doi.org/10.1073/pnas.072081299>
- Blei, D. M., Ng, A. Y., y Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3), 7280-7287. <https://doi.org/10.1073/pnas.082080899>
- Conte, R., Gilbert, N., Bonelli, G., y Ciampaglia, G. L. (2012). Manifesto of computational social science. *EPJ Data Science*, 1(1), 1-9. <https://doi.org/10.1140/epjds4>
- Dusa, A. (2010). QCA: A package for Qualitative Comparative Analysis in R. *Journal of Statistical Software*, 36(4), 1-13. <https://doi.org/10.18637/jss.v036.i04>
- Epstein, J. M. (1999). Agent Based Computational Models and Generative Social Science. *Complexity*, 4(5), 41-60. [https://doi.org/10.1002/\(SICI\)1099-0526\(199905/06\)4:5<41::AID-CPLX9>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0526(199905/06)4:5<41::AID-CPLX9>3.0.CO;2-F)
- Fiss, P. C. (2011). Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. *Academy of Management Journal*, 54(2), 393-420. <https://doi.org/10.5465/amj.2011.60263120>
- Grimmer, J., y Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267-297. <https://doi.org/10.1093/pan/mps028>

- Heath, B., Hill, R., y Ciarallo, F. (2009). A Survey of Agent-Based Modeling Practices (January). *Journal of Artificial Societies and Social Simulation*, 12(2), 2. <https://www.jasss.org/12/4/9.html>.
- Hopkins, D. J., y King, G. (2010). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, 54(1), 229-247. <https://tinyurl.com/y6plmzlg>
- Macy, M. W., y Willer, R. (2002). From Factors to Actors: Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, 28, 143-166. <https://doi.org/10.1146/annurev.soc.28.110601.141117>
- Ragin, C. C. (2006). Set Relations in Social Research: Evaluating Their Consistency and Coverage. *Political Analysis*, 14(3), 291-310. <https://doi.org/10.1093/pan/mpj019>

Capítulo 1

Algoritmo de autorreflexión de agentes sociales para la formación de grupos de opinión a través de un modelo de dinámica de opinión de acuerdo relativo

Norma Leticia Abrica Jacinto¹
Evguenii Kurmyshev²

1. Introducción

La psicología social busca comprender cómo los pensamientos y acciones de los individuos son influenciados por los grupos a los que pertenecen. Un aspecto clave de esta influencia es la opinión, ya que a partir de ella los individuos toman decisiones que pueden afectar el comportamiento del grupo en su conjunto. Por tanto, influir en la opinión de los miembros de un grupo puede modificar sus decisiones y, en consecuencia, su comportamiento social.

Desde las matemáticas y la computación se han desarrollado modelos para estudiar la dinámica de la formación de opinión en grupos sociales. Estos modelos consideran factores psicológicos y socio-

1. Instituto de Matemáticas y Actuaría, Universidad del Mar. ORCID: <https://orcid.org/0000-0003-0676-1276>. Correo electrónico: norma-abrica@aulavirtual.umar.mx
2. Centro Universitario de los Lagos, Universidad de Guadalajara. ORCID: <https://orcid.org/0000-0003-2832-5918>. Correo electrónico: evguenii.kourmychev@academicos.udg.mx

lógicos, como la estructura de la sociedad, la conciencia sobre un tema y la influencia de líderes de opinión o medios de comunicación. Las matemáticas permiten traducir estos comportamientos en ecuaciones y algoritmos que ayudan a predecir si un grupo llegará a un consenso o se dividirá en subgrupos con opiniones diferentes.

Los agentes dentro de un grupo social interactúan entre sí y con su entorno de manera dinámica, cambiando su opinión con base en experiencias pasadas y la información que reciben. Estas interacciones no son lineales ni predecibles de manera simple, ya que la formación de opinión es un proceso complejo. No es posible analizar el comportamiento de un agente de manera aislada, pues su evolución depende de la interacción con los demás y del contexto en que se desarrolla.

A diferencia de los sistemas físicos o biológicos, los fenómenos sociales no se rigen por leyes universales y estables, sino por patrones de comportamiento que cambian con el tiempo y que suelen manifestarse de manera probabilística (Castellano, Fortunato y Loreto, 2009). Esta variabilidad introduce un grado de incertidumbre mayor en el análisis social, dificultando la predicción del comportamiento colectivo a partir de las decisiones individuales (Miller y Page, 2007). En consecuencia, la formación de opiniones no puede entenderse simplemente como la suma de acciones individuales, ya que emergen fenómenos colectivos no lineales y a menudo inesperados (Helbing, 2012).

Para abordar esta complejidad, los modelos de dinámica de opinión conciben la sociedad como un sistema dinámico compuesto por agentes interdependientes cuyas decisiones y creencias se ven influenciadas por su entorno social, sus interacciones y características individuales. Estos modelos permiten analizar cómo surgen y evolucionan grupos de opinión, considerando factores como la homofilia, la influencia de líderes de opinión, la topología de las redes sociales y los sesgos cognitivos (Lorenz, 2007; Flache *et al.*, 2017).

A continuación se presenta un resumen de los elementos fundamentales y características principales de los modelos de dinámica de opinión.

1.1. Modelos de dinámica de opinión

Desde la década de 1990, el modelado matemático de la opinión ha cobrado relevancia como herramienta clave para comprender los procesos de formación, difusión y transformación de las opiniones en contextos sociales. Diversos modelos han sido desarrollados para estudiar la dinámica de la opinión pública, abordando fenómenos como la propagación de rumores, el surgimiento de consensos, la persistencia de opiniones minoritarias, la polarización, la toma de decisiones colectivas, el extremismo ideológico y la competencia entre partidos políticos (Castellano, Fortunato y Loreto, 2009; Flache *et al.*, 2017). Estos enfoques permiten representar las interacciones sociales con agentes que actualizan sus opiniones en función de reglas determinadas, facilitando el análisis de comportamientos emergentes en poblaciones complejas.

La dinámica de la opinión suele considerarse un campo de la física interdisciplinaria que investiga y analiza cómo se forman, evolucionan y se propagan las opiniones y creencias de los individuos en la sociedad.

En términos generales, son cuatro los elementos básicos que definen a un modelo de dinámica de opinión:

- *El espacio de opinión.* Definido como el conjunto de valores que toma la variable del sistema: la opinión.
- *La estructura social.* Representada a través de un grafo $G = (V, E)$ donde V , el conjunto de vértices, representa los agentes sociales (individuos) y E , el conjunto de aristas, representa la comunicación entre pares de agentes.
- *Dinámica de interacción.* Define la forma y secuencia de la interacción de los agentes y las condiciones para la actualización de la variable de estado, en caso de ser así. Existen dos tipos: a) por pares de agentes, b) por agentes vecinos. En el primer caso, la dinámica de interacción puede ser unidireccional (solo un agente es influenciado por el otro) o bidireccional (ambos agentes son influenciados). En el segundo, cada agente es influenciado por el grupo de agentes vecinos a él.
- *La regla de actualización de la variable del sistema.* Son básicamente las ecuaciones del modelo que describen el cambio de la variable de estado de cada agente.

Los primeros modelos de dinámica de opinión fueron definidos en un espacio de opinión binario, con lo cual solo se podía simular y predecir la evolución de una opinión mayoritaria o minoritaria en la sociedad. Por ejemplo, el modelo del votante (Clifford y Sudbury, 1973; Holley y Liggett, 1975) o el modelo de la regla de la mayoría (Galam, 2002). En el modelo del votante los agentes están colocados en una cuadrícula, donde cada celda representa un agente. Cada agente puede tener la opinión -1 o $+1$, es decir, el estado de los agentes es ± 1 . La dinámica de interacción es por agentes vecinos, interactuando solo con los vecinos adyacentes. La regla de actualización del modelo es simple: en cada paso de tiempo se forman parejas de agentes (celdas) adyacentes; si ambos agentes comparten el mismo estado, nada sucede; de lo contrario, uno adopta el estado del otro. En el modelo de la regla, la mayoría, los agentes también están representados por celdas de una cuadrícula y el estado de la variable opinión es ± 1 . La dinámica de interacción es diferente al modelo anterior; aquí los agentes pueden comunicarse con cualquier otro agente de la sociedad. En cada paso de tiempo se forman grupos de agentes, los cuales pueden variar de un paso de tiempo a otro. La actualización de la opinión de los agentes se da en función de la opinión de la mayoría del grupo al que pertenecen.

Posteriormente aparecieron los primeros modelos de espacio de opinión continuo, entre los que destaca el modelo de Deffuant-Weisbuch (Deffuant *et al.*, 2000; Weisbuch *et al.*, 2002, 2003) y el Hegselmann-Krause (Krause, 2000; Hegselmann y Krause, 2002), ambos pertenecientes a la categoría de modelos de confianza acotada (BC, *bounded confidence* en inglés, nosotros los denotaremos como CA, *confianza acotada*). Tales modelos se basan en el supuesto de *homofilia*, es decir, la tendencia de los individuos a asociarse y formar vínculos sociales con aquellos que comparten características o rasgos similares como edad, género, etnia, nivel educativo, creencias, opiniones. Por ejemplo, en el modelo de Deffuant-Weisbuch cada agente solo puede interactuar con aquellos con los que están conectados (a través de la estructura social). Luego, la dinámica de interacción se da por parejas, de forma bidireccional en cada paso de tiempo. Cada agente tiene un valor asignado en el espacio de opinión continuo $[0,1]$, y además poseen un nivel de confianza $\epsilon > 0$. Si la diferencia de opinión entre ambos agentes es menor que o igual al nivel de confianza, entonces sus opiniones se aproximan al promedio de sus opiniones; en caso

contrario, no se actualiza la opinión de ambos agentes. Es aquí donde el supuesto de homofilia toma lugar: los agentes interaccionan y se dejan influir por aquellos que son similares entre sí.

Por otro lado, el modelo de Hegselmann-Krause (Krause, 2000; Hegselmann y Krause, 2002) se basa en el mismo supuesto de homofilia con la diferencia en la dinámica de interacción. En este caso, en cada paso de tiempo los agentes interactúan con aquellos agentes vecinos que queden “dentro de su nivel de confianza”, es decir, de todos aquellos agentes para los cuales su diferencia de opinión es menor o igual que el umbral del agente. Como resultado de la dinámica, los agentes en un grupo alcanzarán el mismo valor de opinión, la cual será el promedio de las opiniones anteriores de todos los agentes del grupo. Puede consultar las referencias anteriores y el trabajo de Abrica y Aguilera (2024) donde los autores muestran algunas simulaciones de ambos modelos en Netlogo.

Con la finalidad de acercar a los modelos de dinámica de opinión a situaciones más realistas, se definió otra categoría: los modelos de acuerdo relativo (AR). Inicialmente, estos modelos surgieron como una extensión de los modelos de CA con el propósito de capturar interacciones sociales más matizadas. A diferencia de los modelos de CA, los modelos de AR permiten grados parciales de influencia, basados en la cercanía entre las opiniones de los agentes. Por ejemplo, el trabajo de Deffuant *et al.* (2002) propone un modelo que permite modelar la transición de opiniones moderadas a opiniones extremas, en un espacio de opinión continuo $[-1, 1]$. Esto permite una interpretación muy aproximada a diversos contextos, tales como condiciones políticas, económicas o sociales prevalecientes durante el auge del extremismo. Aquí, el extremismo son aquellos agentes que presentan opiniones de valor -1 o $+1$, o muy cercanos a ellos. Esto lo hacen definiendo una nueva variable, llamada incertidumbre, la cual está relacionada con la opinión de cada agente. Este nuevo enfoque reconoce que los agentes no cambian su postura de manera absoluta, sino que ajustan su opinión en función de cuán similares son a las de los demás. Así, los modelos de acuerdo relativo ofrecen una herramienta más flexible y realista para estudiar procesos de consenso, polarización o fragmentación dentro de una población.

El modelo de concordia y antagonismo parcial (C/PA), propuesto por Kurmyshev *et al.* (2011), es una extensión del modelo de acuerdo

relativo, ya que incorpora, además de la variable de opinión y la incertidumbre, el perfil psicológico de los agentes y la tolerancia media inicial de la sociedad. En la Sección 2.1 se presentará una descripción más detallada de este modelo.

1.2. Modelos basados en agentes

Para estudiar los modelos de dinámica de opinión se emplean diversos métodos computacionales, dado que estos modelos describen la evolución del estado de opinión de múltiples agentes en interacción. Matemáticamente, suelen representarse como sistemas dinámicos acoplados, que pueden formularse mediante ecuaciones en tiempo discreto o, en algunos casos, mediante ecuaciones diferenciales. En la mayoría de los enfoques, se utilizan ecuaciones en tiempo discreto para modelar la evolución paso a paso del sistema. Para analizar el comportamiento futuro del sistema y su sensibilidad ante cambios en los parámetros, se recurre a herramientas como la simulación computacional y los métodos numéricos. Estas herramientas permiten explorar fenómenos colectivos emergentes, como el consenso, la polarización o la fragmentación de opiniones.

Una metodología destacada para este propósito es el Modelado Basado en Agentes (MBA). Esta técnica representa a cada individuo como una entidad autónoma —un agente— que posee características particulares, reglas de comportamiento y capacidad de decisión. Según Rodríguez-Zoya *et al.* (2015), el MBA permite representar explícitamente tres elementos fundamentales en las ciencias sociales: los agentes, el ambiente y las reglas de comportamiento. Además, facilita la incorporación de estructuras de interacción entre agentes y su entorno.

Un modelo de opinión implementado mediante MBA inicia con una población de N agentes y una estructura social representada por un grafo $G = (N, L)$ donde N es el conjunto de nodos (agentes) y L el conjunto de aristas en el grafo (enlaces de comunicación). Esta estructura social, o red, puede representarse mediante la *matriz de adyacencia* $A = (a_{ij})_{N \times N}$ donde $a_{ij} = 1$ si el agente i se comunica con el agente j , y $a_{ij} = 0$ en caso contrario. En general, se considera un grafo no dirigido, es decir, $a_{ij} = a_{ji}$, aunque es posible modelar interacciones dirigidas si el fenómeno lo requiere.

Otra representación es la *matriz de aristas* E , donde para cada $l \in L$ existen agentes i y j con $e_{1l} = i$, $e_{2l} = j$ y $i < j$. Es decir, que existen agentes i, j relacionados en la sociedad y que la matriz E está ordenada (figura 1). Esta notación resulta conveniente al implementar algoritmos de simulación y visualización de redes.

Figura 1. Representación de la matriz de aristas

$$E = \begin{pmatrix} e_{11} & e_{12} & e_{13} & \cdots & e_{1L} \\ e_{21} & e_{22} & e_{23} & \cdots & e_{2L} \end{pmatrix}$$

La construcción de un modelo MBA implica varios pasos fundamentales:

- Definición de características y variables de estado: En los modelos de opinión, las variables principales suelen ser la opinión y el grado de incertidumbre de cada agente. Adicionalmente, pueden incorporarse atributos jerárquicos u organizacionales.
- Establecimiento de condiciones de interacción: Por ejemplo, en el modelo de Deffuant, dos agentes solo interactúan si sus opiniones están dentro de un umbral de proximidad.
- Aplicación de reglas de actualización: Se actualizan las variables de estado de los agentes conforme a las reglas definidas, que pueden ser determinísticas o estocásticas.

Para fomentar la replicabilidad y claridad metodológica en el diseño de los modelos basados en agentes, se diseñó una herramienta llamada *ODD protocol* (Overview, Design concepts, Details); en español: descripción general, conceptos de diseño y detalles (Grimm *et al.*, 2006, 2010). Este protocolo es un formato estándar para describir los modelos basados en agentes, facilitar su escritura, lectura y replicabilidad del modelo, sin ser sumamente técnico. Las descripciones de modelos con ODD incluyen ecuaciones y algoritmos cortos, pero se basan en texto escrito y están destinadas a ser leídas de una forma fácil. Son independientes del *hardware* y el *software* utilizados para implementar el modelo. La figura 2 describe los elementos del ODD, en cada una de sus categorías. Cada una de ellas tiene un propósito diferente: dar una visión general del modelo, explicar cómo se utilizaron los conceptos

de diseño importantes para los MBA y explicar todos los detalles de la “maquinaria” del modelo. Para más detalle del ODD se puede consultar Grimm *et al.* (2006, 2010, 2020).

Figura 2. Estructura de las categorías del protocolo ODD

Descripción general (O)	1. Propósito y patrones 2. Entidades, variables de estado y escalas 3. Visión general y programación del proceso	
Diseño de conceptos (D)	4. Diseño de Conceptos	Principio básicos Emergencia Adaptación Objetivos Aprendizaje Predicción Detección Interacción Estocasticidad Colectivos Observación
Detalles (D)	5. Inicialización 6. Datos de entrada 7. Submodelos	

El protocolo ODD de este trabajo se encuentra en el *Apéndice: Protocolo ODD del modelo C/PA con el proceso de autorreflexión*.

1.3. Justificación del trabajo y estructura social utilizada: red de mundo pequeño

En general, los modelos de dinámica de opinión suponen interacciones reactivas, es decir, los agentes modifican su estado únicamente como respuesta a las interacciones sociales con otros agentes. Estos modelos no suelen incorporar mecanismos que permitan cambios internos autónomos en la opinión del individuo, es decir, procesos mentales independientes de la interacción social directa.

En este contexto, proponemos incorporar un mecanismo de autorreflexión en un modelo de dinámica de opinión basado en el principio de acuerdo relativo. La autorreflexión se define aquí como la capacidad del agente de evaluar su propio estado y modificarlo internamente, sin necesidad de interactuar con su entorno. Este mecanismo permite estudiar la posibilidad de cambios endógenos, más cercanos a procesos cognitivos, lo cual representa una extensión importante a los modelos tradicionales.

Para evaluar la influencia de este mecanismo, se ha optado por implementar el modelo sobre una red de mundo pequeño. Este tipo de red, formalizado por Watts y Strogatz (1998), se caracteriza por una alta agrupación local y caminos cortos entre pares de nodos. Las redes de mundo pequeño reflejan con fidelidad diversas estructuras sociales reales, como redes de colaboración científica, conexiones neuronales o vínculos de amistad. Su estructura permite una propagación eficiente de información al tiempo que conserva núcleos locales de interacción intensiva.

La elección de esta topología tiene como objetivo minimizar sesgos estructurales y garantizar un equilibrio entre conectividad global y local, de modo que se pueda observar con mayor claridad el efecto de la autorreflexión sobre la dinámica colectiva del sistema. Al tratarse de una red con propiedades relativamente homogéneas en cuanto al grado medio, se espera que los resultados obtenidos reflejen de forma más precisa el impacto del proceso interno propuesto, sin verse distorsionados por efectos topológicos extremos.

En resumen, este trabajo explora la incorporación de procesos individuales de reflexión en modelos de dinámica de opinión, bajo una estructura social realista y bien estudiada, como lo es la red de mundo pequeño, con el fin de ampliar la comprensión de los mecanismos que contribuyen a la formación y transformación de opiniones en contextos sociales.

2. Autorreflexión en un MDO de AR

El modelo de Concordia y Antagonismo Parcial (C/PA), propuesto por Kurmyshev *et al.* (2011), es una extensión del modelo de acuerdo

relativo de Deffuant-Weisbuch *et al.* (2002), ya que incorpora, además de la opinión y la incertidumbre de los agentes, el perfil psicológico de los agentes y la tolerancia media inicial de la sociedad.

En el trabajo de Kurmyshev y Abrica-Jacinto (2022), los autores mostraron que la variable de incertidumbre, tal como se define en los modelos de acuerdo relativo, no estaba adecuadamente descrita. A través de diversas simulaciones del modelo C/PA, argumentaron que dicha variable representa más bien un proceso interno de los agentes, por lo que debe entenderse como una característica psicológica, y no como una variable de naturaleza social.

En consecuencia, propusieron renombrarla como *tolerancia*. A partir de esta reinterpretación, y en coherencia con dicho enfoque, en lo que sigue del presente trabajo se utilizará este término. Así, las variables de estado de los agentes serán la opinión y la tolerancia de opinión.

La autorreflexión se define aquí como la capacidad del agente de evaluar su propio estado y modificarlo internamente, sin necesidad de interactuar con su entorno. En este sentido, solo podrá evaluar su opinión mediante su tolerancia de opinión. Es decir, este proceso interno de reflexionar está en función de una variable psicológica.

2.1. El modelo C/PA

El modelo C/PA contempla sociedades formadas por agentes de dos tipos psicológicos: agentes de concordia (agentes tipo C) y agentes de acuerdo con antagonismo parcial (agentes tipo PA). Para caracterizar a los N -agentes de la sociedad, se hace una partición en dos subconjuntos con $N_C = pN$ igual al número de agentes de concordia, y $N_{PA} = (1 - p)N$ el número de agentes de antagonismo parcial, con $p \in [0, 1]$ un parámetro del modelo C/PA. Tales agentes son llamados C-agentes y PA-agentes, respectivamente.

En cada instante discreto de tiempo t , el estado del agente i se describe mediante dos variables: opinión $x_i(t) \in [-1, 1]$ y tolerancia de opinión $u_i(t) \in [0.05, 1]$. La dinámica de interacción en el modelo C/PA es estocástica y por pares de agentes. En cada instante t , se seleccionan al azar M aristas de la red (de la matriz E). Cada arista conecta un par de agentes (i, j) , y de forma aleatoria a uno de ellos, digamos i , se le designa como *agente activo* (influyente), y al otro agente, j , como

agente pasivo (influenciado). De este modo, en caso de cumplir con las condiciones del modelo, el agente i influirá en las variables de estado del agente j , mientras que las variables de estado del agente i quedan sin cambio.

Cada agente i posee su opinión $x_i(t)$ y su tolerancia de opinión $u_i(t)$. Con esto, es posible definir el *segmento de opinión* como $[x_i(t) - u_i(t), x_i(t) + u_i(t)]$, este describe la opinión del agente y qué tan amplia es su tolerancia al cambio de opinión. Para cada par seleccionado (i, j) , se define $h_{ij}(t)$ como el *traslape de opinión*, el cual cuantifica la longitud de la intersección de los segmentos de opinión $s_i(t)$ y $s_j(t)$. La Ecuación 1 muestra el valor del traslape de opinión.

$$h_{ij}(t) = \min\{x_i(t) + u_i(t), x_j(t) + u_j(t)\} - \max\{x_i(t) - u_i(t), x_j(t) - u_j(t)\} \quad (\text{Ec. 1})$$

Cuando $h_{ij}(t) > 0$, las variables de estado del agente pasivo se actualizan de acuerdo con las Ecuaciones 2a y 2b.

$$x_j(t + 1) = x_j(t) + \mu_1 \cdot ar_{ij}^{C,PA}(t) \cdot [x_i(t) - x_j(t)], \quad (\text{Ec. 2a})$$

$$u_j(t + 1) = u_j(t) + \mu_1 \cdot ar_{ij}^{C,PA}(t) \cdot [u_i(t) - u_j(t)], \quad (\text{Ec. 2b})$$

donde $\mu_{1,2} \in [0, 0.5]$ son parámetros de convergencia (Kurmyshev y Abrica-Jacinto, 2022), y $ar_{ij}^{C,PA}(t)$ es el acuerdo relativo del agente pasivo j según su perfil psicológico, C-agente o PA-agente. Las Ecuaciones 3a y 3b describen los valores de los acuerdos:

$$ar_{ij}^C(t) = \frac{h_{ij}(t)}{u_i(t)}, \quad (\text{Ec. 3a})$$

$$ar_{ij}^{PA}(t) = \frac{h_{ij}(t)}{2 u_i(t)} \left[\frac{h_{ij}(t)}{u_i(t)} - 1 \right]. \quad (\text{Ec. 3b})$$

Cuando $h_{ij}(t) \leq 0$, no hay cambios en las variables de estado. En el modelo C/PA, la tendencia de la opinión de los C-agentes se aproxima a la opinión del agente con el que interactúa, comportamiento similar al presentado en el modelo de Deffuant-Weisbuch (DW) (Deffuant *et al.*, 2002). Por otro lado, la dinámica de los PA-agentes puede ser repulsiva o atractiva, lo que depende del valor del acuerdo relativo ar^{PA} .

w2.2. Inserción de autorreflexión

No todos los agentes de la sociedad tendrán la capacidad de autorreflexionar y modificar su opinión debido a este proceso. Denominaremos como *agentes con potencial reflexivo* a aquellos agentes que pueden actualizar su opinión a partir de un proceso interno de reflexión. Y de esos agentes, llamaremos *agentes reflexivamente activos* a aquellos que sí llevan a cabo el proceso de autorreflexión y, por tanto, modifican su estado de opinión, en cada paso de tiempo. Esta distinción no implica diferencias en capacidades cognitivas, sino únicamente en el mecanismo de actualización.

De este modo, la *autorreflexión* se define como el proceso mediante el cual agentes con potencial reflexivo actualizan su opinión en función de su tolerancia de opinión para luego realizar la dinámica de actualización del modelo C/PA. Para esto, sea $r \in [0,1]$ la proporción de agentes con potencial reflexivo, por ejemplo, $\{i_1, i_2, \dots, i_r\}$ representa el conjunto de agentes con potencial reflexivo. De este conjunto de agentes, sea $\rho \in [0,1]$ la proporción de agentes reflexivamente actualizados, es decir, $\{j_1, j_2, \dots, j_\rho\} \subseteq \{i_1, i_2, \dots, i_r\}$ son los agentes que cambian su estado de opinión debido al proceso de autorreflexión. Entonces, el número $\frac{\rho}{r} \leq 1$ define la intensidad del proceso de autorreflexión. Cuando $r = 0$, la sociedad está compuesta únicamente por agentes reactivos (modelo C/PA original); en el caso opuesto, $r = 1$, todos los agentes de la sociedad son agentes con potencial reflexivo. Por definición, si $r = 0, \rho = 0$ o ambos, se considera que la intensidad es cero.

Ahora, en cada paso de tiempo t , se tiene la opinión $x_i(t)$ y tolerancia de opinión $u_i(t)$ de todos los agentes i s de la sociedad. Para los agentes $\{j_1, j_2, \dots, j_\rho\}_t$, la modificación de la opinión $\hat{x}_j(t)$ por la autorreflexión se realiza mediante la Ecuación 4:

$$\hat{x}_j(t) = x_j(t) + \delta_j(t), \quad (\text{Ec. 4})$$

donde $\delta_j(t) \in [-u_j(t), u_j(t)]$ es un valor elegido aleatoriamente dentro del segmento de tolerancia. Para todos los agentes $j \notin \{j_1, j_2, \dots, j_\rho\}_t$ se tiene $\delta_j(t) = 0$; es decir, la opinión no se modifica por este mecanismo. Se asume que el cambio de opinión por autorreflexión no altera la tolerancia de opinión de los agentes. Con este nuevo valor de opinión

$\hat{x}_j(t)$, se definen los segmentos de opinión, los traslapes de opinión, acuerdos relativos y las ecuaciones de actualización, como se muestra en las Ecuaciones 5a y 5b.

$$\hat{x}_j(t+1) = \hat{x}_j(t) + \mu_1 \cdot \widehat{ar_{ij}^{C,PA}}(t) \cdot [x_i(t) - x_j(t)], \quad (\text{Ec. 5a})$$

$$u_j(t+1) = u_j(t) + \mu_2 \cdot \widehat{ar_{ij}^{C,PA}}(t) \cdot [u_i(t) - u_j(t)], \quad (\text{Ec. 5b})$$

donde,

$$\widehat{ar_{ij}^C}(t) = \frac{\widehat{h_{ij}}(t)}{u_i(t)},$$

$$\widehat{ar_{ij}^{PA}}(t) = \frac{\widehat{h_{ij}}(t)}{2u_i(t)} \left[\frac{\widehat{h_{ij}}(t)}{u_i(t)} - 1 \right],$$

$$\widehat{h_{ij}}(t) = \min\{\widehat{x_i}(t) + u_i(t), \widehat{x_j}(t) + u_j(t)\} - \max\{\widehat{x_i}(t) - u_i(t), \widehat{x_j}(t) - u_j(t)\}.$$

Es importante señalar que este proceso autorreflexivo no distingue entre los tipos psicológicos C o PA. Finalmente, la fase reactiva de actualización de opiniones (que afecta tanto a agentes reflexivos como reactivos) se lleva a cabo de forma convencional (Kurmyshev *et al.*, 2011; Kurmyshev y Abrica-Jacinto, 2022).

2.3. Algoritmo de la dinámica autorreflexiva y reactiva en el modelo C/PA

A continuación se describe la dinámica autorreflexiva y reactiva del modelo C/PA implementada en el algoritmo computacional, para obtener la evolución temporal de la opinión y tolerancia de opinión de los agentes en una red dada.

ENTRADAS DEL MODELO

- Número total de agentes: N .
- Estructura social: grafo representado por la matriz de aristas E .

- Perfil psicológico: $p \in [0,1]$, representado como un vector binario unidimensional y ordenado ($1 = \text{C-agente}$; $0 = \text{PA-agente}$), con pN de C-agentes y $(1 - p)N$ de AP-agentes.
- Parámetros de convergencia de opinión y tolerancia de opinión: $\mu_1 = 0.5$ y $\mu_2 = 0.05$ (Kurmyshev y Abrica-Jacinto, 2022).
- Tolerancia media inicial: $U \in [0.2,1]$
- Proporción r de agentes con potencial reflexivo: $r \cdot N$
- Proporción p de agentes reflexivamente activos: $p \cdot r \cdot N$
- Número total de pasos de tiempo: T

PROCESO DEL MODELO

- Inicialización: Se asignan opiniones aleatorias $x_i(0) \in [-1,1]$ y tolerancias de opinión $u_i(0) \in [U - 0.15, U + 0.15]$ a todos los agentes i s de la sociedad.

Se eligen aleatoriamente los agentes con potencial reflexivo $\{i_1, i_2, \dots, i_r\}$.

En cada paso de tiempo t :

Se eligen aleatoriamente los agentes reflexivamente activos $\{j_1, \dots, j_p\} \subseteq \{i_1, i_2, \dots, i_r\}$.

Se realiza el cambio de opinión $\hat{x}_{ij}(t)$ (Ec. 4).

Se seleccionan N -aristas de la red de forma aleatoria, para cada una de ellas uno de los nodos será el agente pasivo y el otro el agente activo.

Se calcula el valor de los traslapes de opinión y los acuerdos relativos.

La opinión y tolerancia de opinión del agente pasivo se actualiza siguiendo las Ecuaciones 5a y 5b, si $\widehat{h}_{ij}(t) > 0$.

Salida del modelo:

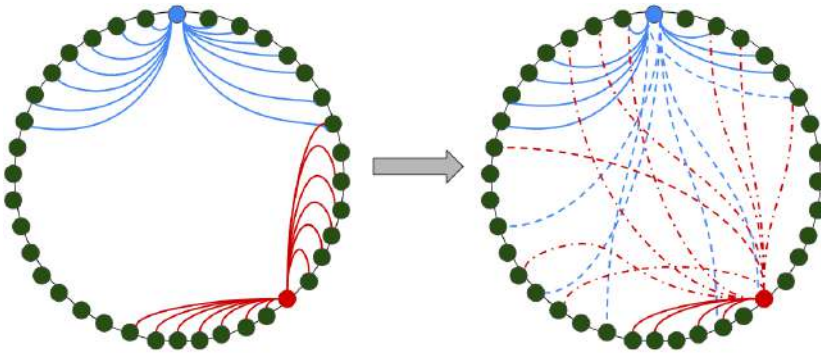
Matriz de opinión y matriz de tolerancia de opinión, cada una de tamaño $T \times N$.

3. Resultados de las simulaciones

3.1. Diseño de experimentos

La estructura social se simula en una red con $N = 1000$ agentes y estructura de mundo pequeño, siguiendo el algoritmo de Watts-Strogatz (1998) teniendo la red $SW(\beta, \kappa)$ con $\beta = 0.25$ la probabilidad de reconexión de las aristas y $\kappa = 7$ el número de agentes a los que se conecta cada nodo a su izquierda y derecha. Con esto, el grado promedio de la red es $\kappa = 14$. Por ejemplo, la figura 3 muestra el acomodo en círculo de varios nodos, y cómo dos de ellos pueden ser redireccionados.

Figura 3. Ejemplo del algoritmo de Watts-Strogatz para redes de mundo pequeño



Se realizan experimentos en sociedades mixtas mediante el perfil psicológico p , considerando los valores de $p = 0, 0.3, 0.7, 1$, definido así para estructuras sociales fijas caracterizadas por p y $SW(\beta, \kappa)$. Estas configuraciones se mantienen fijas durante las simulaciones de dinámica de opinión para diferentes valores de los parámetros (U, r, p) . Se exploran diferentes valores para la terna (U, r, p) con el objetivo de comprender cómo la tolerancia de opinión y los agentes reflexivos

influyen en la formación de grupos de opinión en sociedades definidas por p y $SW(\beta, \kappa)$.

Para mostrar los resultados de las simulaciones en la formación de grupos de opinión se realizó el promedio de 35 experimentos, para cada valor de (p, U) , elaborando *matrices de histogramas de opinión* mediante una paleta de colores, es decir, son arreglos de la forma $X(p, r, \rho) = (f(u, x))_{13 \times 51}$ donde $f(u, x)$ es el histograma de opinión en el último paso de tiempo T , en una sociedad con parámetros (p, U, r, ρ) . Este histograma se obtiene de la matriz de opinión. En este caso el número total de pasos de tiempo es $T = 1000$ el número de bins para el histograma de opinión en el espacio $[-1, 1]$ es 51, con 13 valores de U en el intervalo $[0.2, 0.8]$. Cabe señalar que $X(p, r, \rho)(u, :)$ denotará el u -ésimo renglón de la matriz de histograma de opinión, siendo cada entrada de ese vector-renglón el respectivo histograma de opinión en el tiempo T .

Con el objetivo de identificar los patrones de influencia del proceso de autorreflexión, la proporción r de agentes con potencial reflexivo y la proporción ρ de agentes reflexivamente activos varían de 0 a 1 en incrementos de 0.1, resultando en 10 valores para cada uno de ellos. En total, se estudian $10 \times 10 = 100$ escenarios distintos para cada combinación de valores de (p, U) . Nuevamente, se cuenta con un número grande de resultados y para analizar la influencia del proceso de autorreflexión en la sociedad hemos definido una métrica llamada *distancia de opinión* $E(p, U)$ para observar la diferencia entre las matrices de histogramas de opinión y la matriz de referencia, según el valor de la tolerancia media inicial U y la composición de la sociedad p . Esta matriz se construye a partir de la matriz de referencia $X(p, 0, 0)$, de la siguiente manera:

Considere la matriz $X(p, r, \rho)$, para cada valor de r y ρ .

Recordemos que $X(p, r, \rho)(u, :)$ denota el u -ésimo renglón de la matriz $X(p, r, \rho)$. Entonces definimos la distancia entre las distribuciones de opinión, para (p, U, r, ρ) dados como

$$e(r, \rho|U) = \sum_x |X(p, r, \rho)(u, :) - X(p, 0, 0)(U, :)|. \quad (\text{Ec. 6})$$

Estos valores definen $E(p, U)$ como la *matriz de distancia entre los histogramas de opinión*. Notemos que el orden de esta matriz representa el orden de los valores de (r, ρ) respectivamente. Es decir, que el primer

renglón de $e(r, \rho|U)$ es para el primer valor de $r = 0.1$, y las columnas corresponden a los valores $\rho = 0.1, 0.2, \dots, 1$; en ese orden. El resto de los renglones, sigue el orden de $r = 0.2, \dots, 1$.

Observación. Los valores posibles de las entradas de la matriz de distancia de opinión $E(p, U)$ están acotados en el intervalo $[0, 2N]$. El valor mínimo 0 se alcanza cuando no hay diferencia entre el histograma $X(p, r, \rho)(U, :)$ y $X(p, 0, 0)(U, :)$, mientras que el valor máximo $2N$ ocurre cuando los histogramas son completamente disjuntos en su distribución.

Por ejemplo, supongamos que $X(p, 0, 0)$ y $X(p, r, \rho)$ son como se muestran en la figura 4, con tres valores de U , dos valores de r y ρ , y cuatro valores de opinión, para una sociedad de $N = 10$ agentes. Notemos que la suma por renglón de las matrices de histogramas de opinión debe ser $N = 10$. Además, supongamos que $r_1 < r_2$, $\rho_1 < \rho_2$ y que $U_1 < U_2 < U_3$. Las figuras 5 y 6 muestran el resultado para cada $E(p, U)$ aplicando la fórmula de la Ecuación 6.

Figura 4. Ejemplo de matrices de histogramas de opinión para una sociedad de $N = 10$ agentes, tres valores de tolerancia media inicial U , dos valores de r y ρ , y cuatro bins para el histograma de opinión

$$X(p, 0, 0) = \begin{pmatrix} 2 & 2 & 4 & 2 \\ 3 & 1 & 1 & 5 \\ 2 & 3 & 3 & 2 \end{pmatrix}$$

$$X(p, r_1, \rho_1) = \begin{pmatrix} 0 & 1 & 5 & 4 \\ 0 & 1 & 4 & 5 \\ 1 & 0 & 9 & 0 \end{pmatrix} \quad X(p, r_1, \rho_2) = \begin{pmatrix} 2 & 0 & 1 & 7 \\ 1 & 8 & 1 & 0 \\ 3 & 1 & 3 & 3 \end{pmatrix}$$

$$X(p, r_2, \rho_1) = \begin{pmatrix} 1 & 5 & 0 & 4 \\ 1 & 5 & 1 & 3 \\ 2 & 1 & 5 & 2 \end{pmatrix} \quad X(p, r_2, \rho_2) = \begin{pmatrix} 0 & 0 & 6 & 4 \\ 0 & 2 & 1 & 7 \\ 2 & 1 & 5 & 2 \end{pmatrix}$$

Figura 5. Matriz de distancia de opinión $E(p, U_1)$ de la figura 4

$$E(p, U_1) = \begin{pmatrix} |2-0| + |2-1| + |4-5| + |2-4| & |2-2| + |2-0| + |4-1| + |2-7| \\ |2-1| + |2-5| + |4-0| + |2-4| & |2-0| + |2-0| + |4-6| + |2-4| \end{pmatrix}$$

$$E(p, U_1) = \begin{pmatrix} 6 & 10 \\ 10 & 8 \end{pmatrix}$$

Figura 6. Matrices de distancia de opinión $E(p, U_s)$, con $s = 2, 3$ de la figura 4

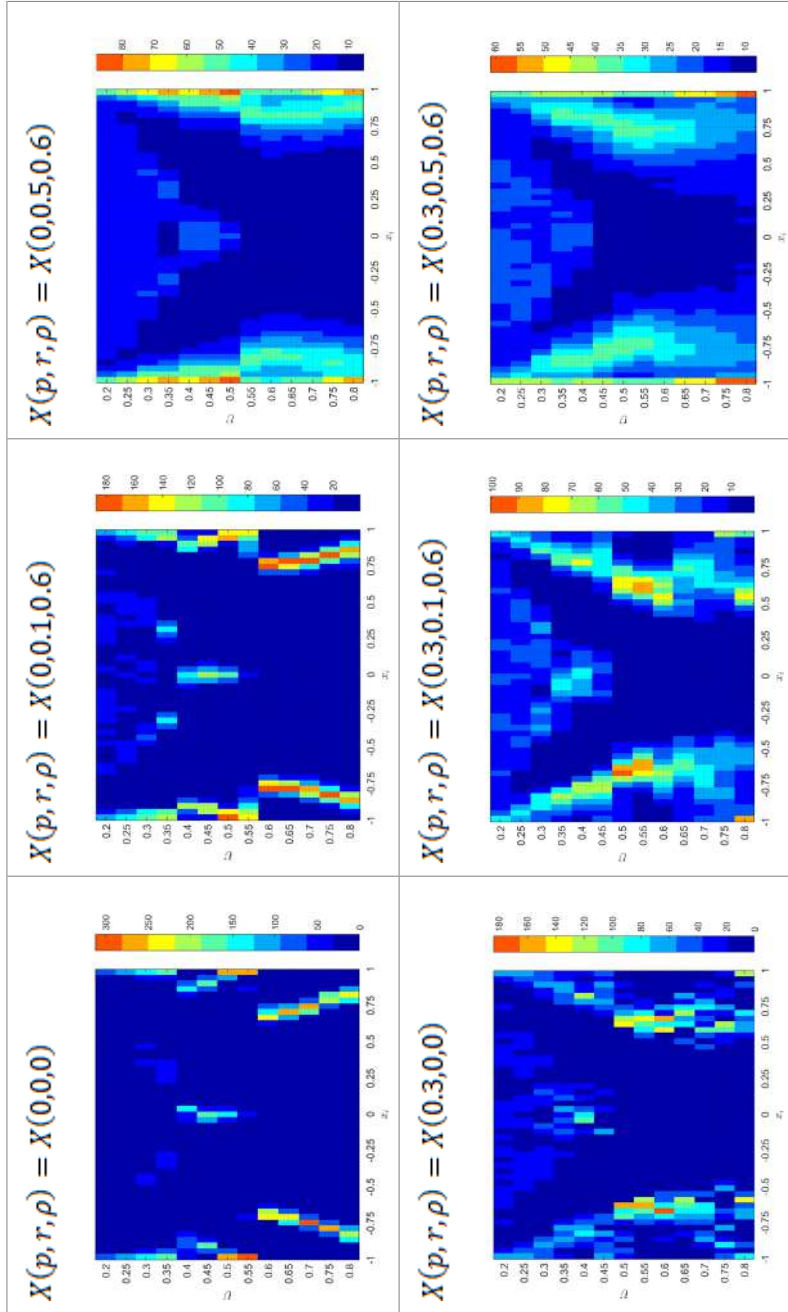
$$E(p, U_2) = \begin{pmatrix} 6 & 14 \\ 8 & 6 \end{pmatrix} \quad E(p, U_3) = \begin{pmatrix} 12 & 4 \\ 4 & 4 \end{pmatrix}$$

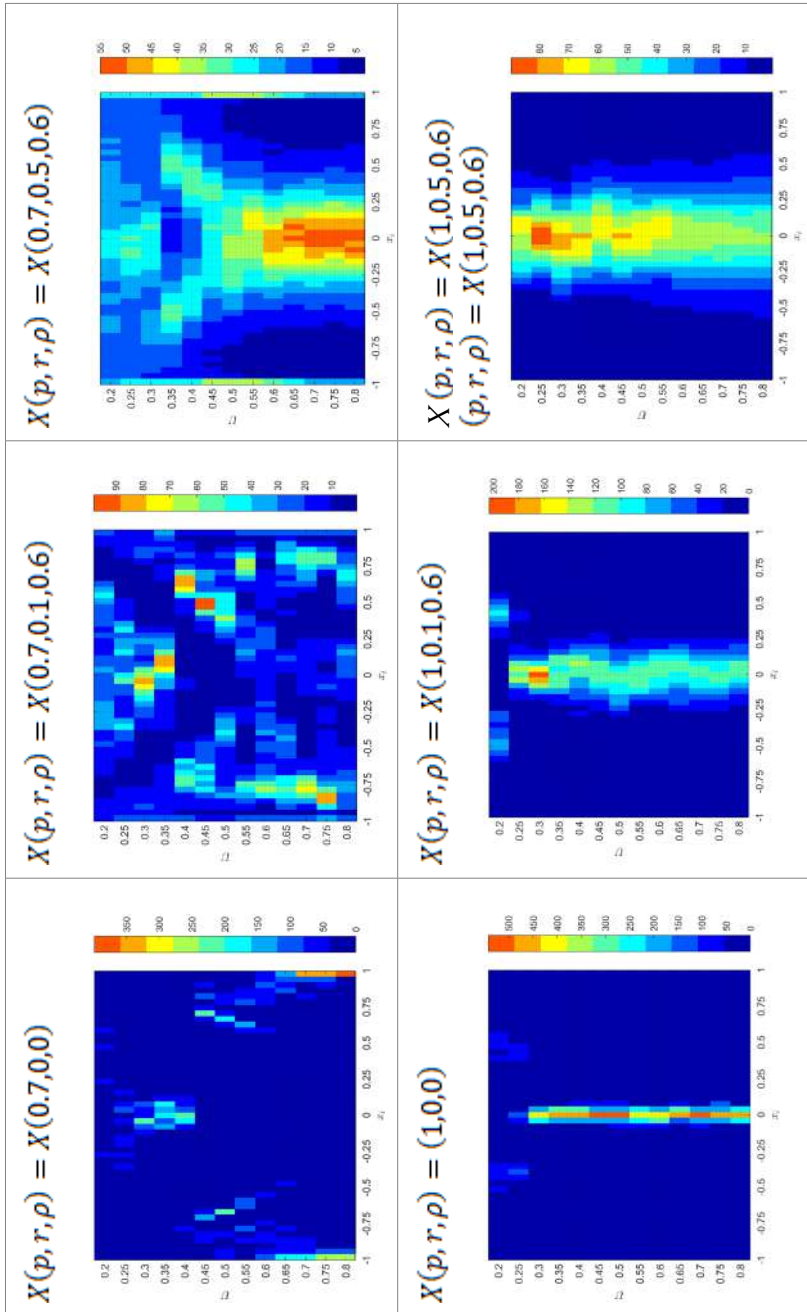
3.2. Análisis de los resultados

Tomamos como *matriz de referencia* a $X(p, 0, 0)$, correspondiente al caso en el que no se incluye el proceso de autorreflexión. En este escenario, las configuraciones finales de la distribución de opinión están determinadas exclusivamente por la dinámica reactiva del modelo C/PA. Esta matriz permite identificar de manera clara los efectos específicos atribuibles a la inclusión del mecanismo de autorreflexión en las distintas simulaciones comparativas.

La figura 7 muestra las matrices de histogramas de opinión $X(p, r, p) = (f(u, x))_{13 \times 51}$ para diferentes valores de (p, r, p) . Es evidente el efecto que tiene la presencia de agentes potencialmente reflexivos—controlado por el parámetro r — sobre la configuración final de las opiniones. Es posible observar cómo se alejan progresivamente de la matriz de referencia conforme aumenta el valor de r . Esta diferencia es particularmente evidente cuando $p = 0.7$. Por ejemplo, la matriz $X(0.7, 0.5, 0.6)$ muestra un patrón distinto—incluso en su topología—respecto a $X(0.7, 0, 0)$, destacándose la formación de un grupo centrista de opinión para valores de tolerancia media inicial en el intervalo $U \in [0.6, 0.85]$. Además, la presencia de agentes potencialmente reflexivos (r) en la sociedad favorece la aparición de grupos extremistas bien definidos en valores pequeños de U , especialmente en contextos donde predominan los PA-agentes. Esto puede apreciarse al comparar las matrices $X(0, 0.5, 0, 0.6)$ y $X(0.3, 0.5, 0.6)$ con sus respectivas matrices de referencia.

Figura 7. Matrices de histogramas de opinión $X(p, r, \rho) = (f(u, x))_{13 \times 51}$. En cada figura se especifica el valor de la terna (p, r, ρ)





Las figuras 8 y 9 presentan las matrices de distancia entre los histogramas de opinión $E(p,U)$ en sociedades mixtas con $p = 0,0,3,0,7,1$, y diferentes valores de la tolerancia media inicial U . El eje horizontal representa la proporción de agentes potencialmente reflexivos (r). El eje vertical representa la proporción de agentes reflexivamente activos (ρ). En general, se observa que el incremento en la tolerancia inicial media U genera modificaciones significativas en la matriz de distancia de opinión, incluso cuando la proporción de agentes potencialmente reflexivos (r) es baja (ver las gráficas de las figuras, por columna).

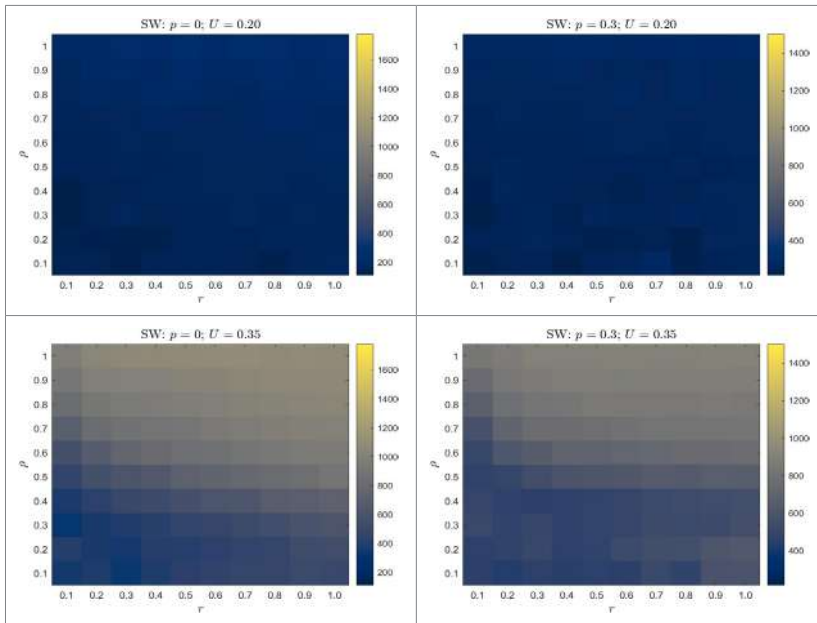
Por ejemplo, en la sociedad de PA-agentes (columna izquierda, figura 8) y con baja tolerancia ($U = 0.2$), los valores de la matriz $E(p,U)$ son uniformemente bajos, lo que sugiere una dispersión débil y poca respuesta al proceso de autorreflexión. A medida que aumenta el valor de la tolerancia de opinión U , se observa la presencia de zonas más amarillas que van emergiendo paulatinamente en la parte superior de cada subfigura. Esto puede indicar que una mayor tolerancia de opinión permite que los PA-agentes respondan de forma más activa al proceso de autorreflexión.

En el caso de sociedades mixtas con 30% de C-agentes ($p = 0.3$), columna derecha de la figura 8), el impacto del proceso de autorreflexión comienza a ser más evidente a partir de $U = 0.55$, en comparación con su contraparte puramente PA-agentes. A medida que la tolerancia U aumenta, el patrón en los mapas de colores tiende a hacerse más uniforme, lo que sugiere una estabilización del sistema causada por el proceso de autorreflexión.

La figura 9 muestra las matrices de distancia de opinión en sociedades donde predominan los C-agentes: $p = 0.7$ (columna izquierda), $p = 1$ (columna derecha). A diferencia de lo observado con valores bajos de p , aquí desde $U = 0.2$ surgen cambios notorios: para $p = 0.7$ el sistema no parece ser perturbado significativamente por el proceso de autorreflexión; para $p = 1$ el efecto del proceso de autorreflexión es muy marcado con la presencia de una franja de transición, lo que sugiere un posible umbral en los valores de r y ρ a partir del cual el proceso de autorreflexión se vuelve dominante sobre la dinámica reactiva del modelo C/PA. En general, se observa que el incremento simultáneo de la proporción de agentes con potencial reflexivo (r) y la proporción de agentes reflexivamente activos (ρ) conducen a un aumento significativo en la distancia de opinión, lo que sugiere un

mayor cambio en la distribución de la opinión debido al proceso de autorreflexión aun en sociedades que tienden al consenso ($p = 1$). Estos resultados sugieren que una proporción del 70% de C-agentes pueden ser suficientes para inducir un comportamiento colectivo similar al de una sociedad completamente de C-agentes, siempre que la tolerancia inicial sea suficientemente alta, y en este caso el proceso de autorreflexión tiene mayor presencia.

Figura 8. Matrices de distancia entre los histogramas de opinión en sociedades mixtas $p = 0, 0.3$, y varios valores de la tolerancia media inicial $U: E(p, U)$



Capítulo 1. Algoritmo de autorreflexión de agentes sociales para la formación de grupos de opinión a través de un modelo de dinámica de opinión de acuerdo relativo

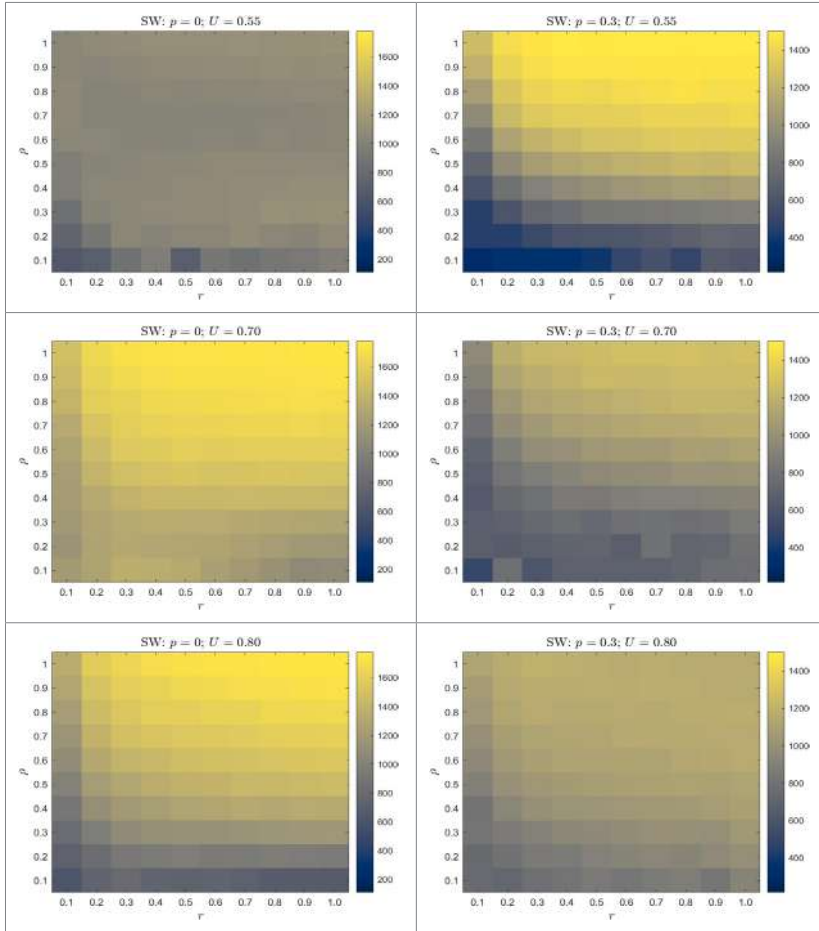
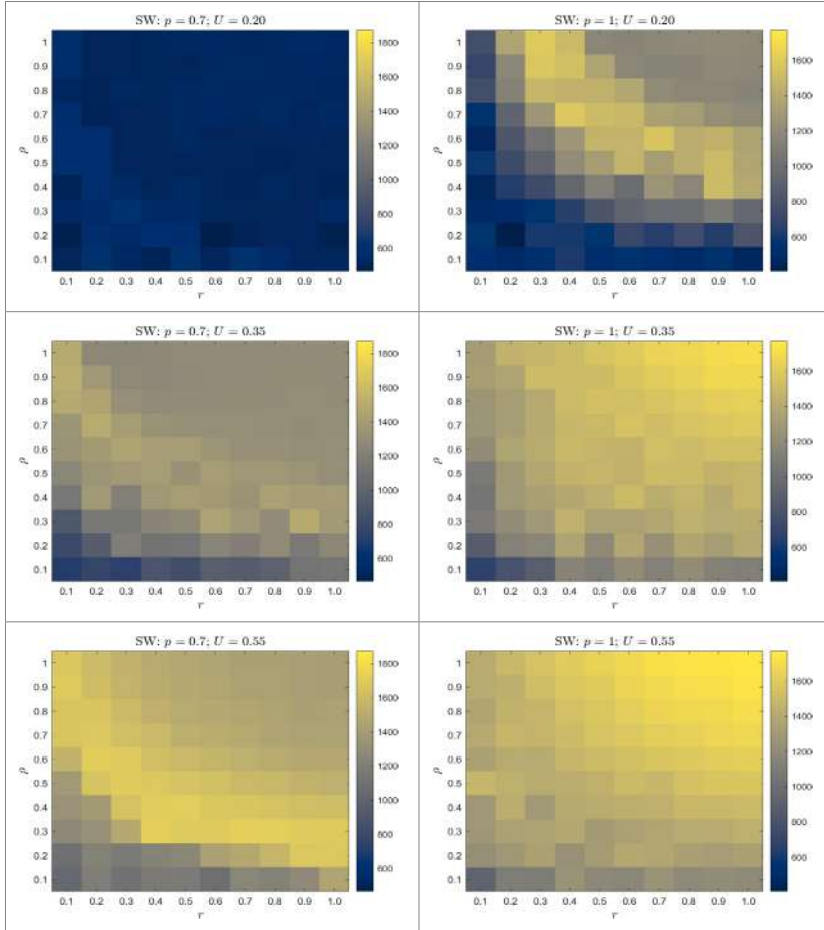
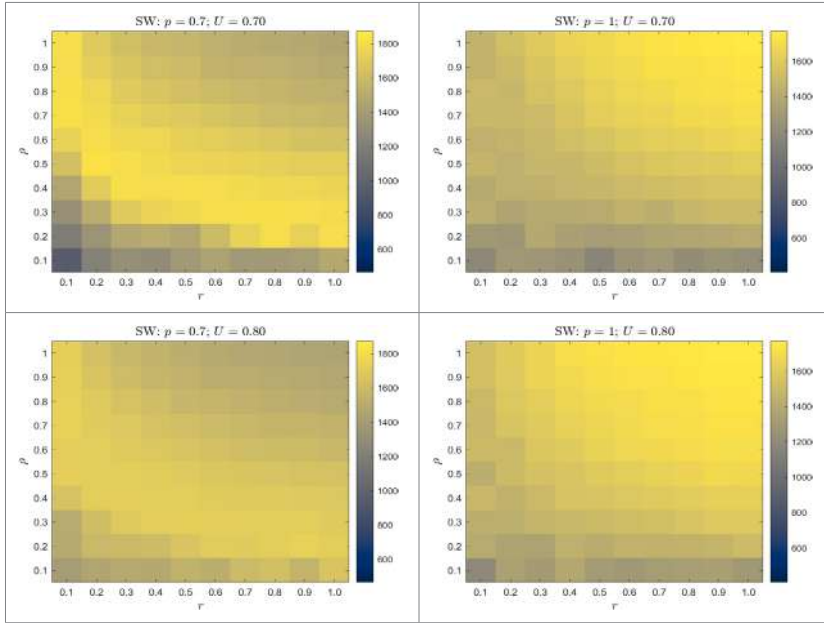


Figura 9. Matrices de distancia entre los histogramas de opinión en sociedades mixtas $p = 0.7, 1$, y varios valores de la tolerancia media inicial $U: E(p, U)$





4. Conclusiones

Los resultados presentados permiten identificar patrones relevantes sobre el impacto del proceso de autorreflexión en la dinámica de opinión, bajo diferentes configuraciones de la tolerancia media inicial U , la composición psicológica p , y las proporciones de agentes potencialmente reflexivos r y reflexivamente activos ρ .

De manera general, se observa que el aumento en la tolerancia media inicial U provoca cambios significativos en la matriz de distancia de opinión $E(p, U)$. Este efecto se acentúa en sociedades con predominancia de C-agentes, lo que sugiere que la tolerancia inicial puede actuar como un catalizador que amplifica la influencia del proceso autorreflexivo.

En el caso de una sociedad de PA-agentes exclusivamente, la dinámica es poco sensible al proceso de autorreflexión cuando la tolerancia inicial media es baja ($U = 0.2$). Sin embargo, conforme aumenta U , comienzan a surgir patrones de mayor intensidad en la matriz $E(p, U)$,

indicando que los PA-agentes pueden responder al proceso autorreflexivo bajo condiciones de mayor tolerancia.

Al introducir en el sistema una proporción moderada de C-agentes ($p = 0.3$), el efecto del proceso de autorreflexión es perceptible a partir de $U = 0.55$, mostrando un contraste con la sociedad exclusivamente de PA-agentes. A medida que la tolerancia aumenta, los patrones de color se tornan más uniformes, lo que indica una estabilización de la distancia de opinión inducida por el proceso de autorreflexión.

Este patrón se intensifica en configuraciones con mayor proporción de C-agentes, y en menor valor de tolerancia. En sociedades $p = 0.7$ y $U = 0.2$, el efecto del proceso de autorreflexión es bajo, mientras que para $p = 1$ es muy notable, lo que sugiere un umbral a partir del cual el proceso de autorreflexión domina la dinámica reactiva del modelo C/PA. Es decir, que a partir de dichos umbrales la dinámica del sistema deja de estar dominada por las interacciones reactivas tradicionales del modelo C/PA y pasa a ser regida por el mecanismo de autorreflexión. Además, cabe mencionar que los cambios importantes también aparecen con la presencia de agentes reflexivamente activos (es decir, valores altos ρ).

Apéndice. Protocolo ODD del modelo C/PA con el proceso de autorreflexión

Con el objetivo de la reproducibilidad del modelo, se describe el protocolo ODD del modelo C/PA con el proceso de autorreflexión.

Descripción general

Propósito: estudiar la dinámica de opinión en sociedades mixtas compuestas por agentes de dos perfiles psicológicos (concordia y antagonismo parcial), considerando tanto interacciones reactivas (basadas en traslape de opinión) como procesos de autorreflexión. Se analiza cómo diferentes proporciones de agentes reflexivos (con capacidad de reflexionar) y reflexivamente activos (que efectivamente actualizan

su opinión por reflexión) afectan la evolución de la distribución de opiniones en la sociedad.

Entidades, variables de estado y escalas. Los agentes poseen una opinión, tolerancia de opinión, perfil psicológico, capacidad de reflexionar y el estado de reflexivamente activo. Los agentes están conectados mediante una red de mundo pequeño. Los parámetros globales son el número de agentes, la proporción de agentes de concordia, la proporción de agentes potencialmente reflexivos, la proporción de agentes reflexivamente activos, la tolerancia media inicial, los parámetros de convergencia de opinión y tolerancia (μ_1, μ_2).

Proceso temporal y espacial. El modelo evoluciona en pasos discretos $t = 1, 2, \dots, T$. El espacio social está representado por la red de conexiones entre agentes.

Diseño de conceptos

- Teoría básica. El modelo se basa en el enfoque de modelos de dinámica de acuerdo relativo y la extensión al modelo C/PA, el cual introduce perfiles psicológicos. Se incorpora el proceso de autorreflexión para modelar cambios internos de opinión independientes de las interacciones sociales.
- Emergencia. Se observa la formación de consenso, polarización, centrismo o extrema dispersión en dependencia de la configuración inicial de parámetros. La interacción entre procesos reactivos y reflexivos produce patrones no triviales en la matriz de distancia de opinión.
- Adaptación. Los agentes modifican su opinión y tolerancia de opinión según la dinámica reactiva y el proceso de autorreflexión.
- Objetivos. No hay objetivos individuales definidos más allá de responder al entorno social o interno.
- Interacciones. Son binarias, unidireccionales, y ocurren a través de aristas seleccionadas aleatoriamente en la red. Un agente influye a otro si existe traslape de opinión.
- Estocasticidad. En la selección de aristas, en la asignación de agente pasivo-reactivo, en la asignación de agentes potencialmente reflexivos y reflexivamente activos.

- Recolección de datos. Matrices de distribución de opinión y tolerancia y la matriz de distancia de opinión.

Detalles

Descrita en la Sección 2.3. Algoritmo de la dinámica autorreflexiva y reactiva en el modelo C/PA.

Agradecimientos

Los autores expresan su agradecimiento a la Universidad del Mar por facilitar la estancia de investigación del Dr. Kurmyshev en el Campus Huatulco, cuya participación fue decisiva en el desarrollo de este trabajo.

Referencias

- Abrica-Jacinto, N. L., y Aguilera Ontiveros, A. (2024). Influencia de la opinión pública en un modelo de rebelión. En A. Aguilera Ontiveros y N. L. Abrica-Jacinto (Coords.), *Temas selectos para las ciencias sociales computacionales: Contribuciones desde América Latina*. Comunidad Editora Latinoamericana.
- Castellano, C., Fortunato, S., y Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2), 591-646. <https://doi.org/10.1103/RevModPhys.81.591>
- Deffuant, G., Amblard, F., Weisbuch, G., y Faure, T. (2002). How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5.
- Deffuant, G., Neu, D., Amblard, F., y Weisbuch, G. (2001). Mixing beliefs among interacting agents. En *Advances in Complex Systems*, p. 11.
- Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., y Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4), 2. <https://doi.org/10.18564/jasss.3521>
- Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J. U., Jørgensen,

- C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., y DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198, 115-126. DOI:10.1016/j.ecolmodel.2006.04.0231
- Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., y Railsback, S. F. (2010). The ODD protocol: a review and first update. *Ecological Modelling*, 221(23), 2760-2768. DOI:10.1016/j.ecolmodel.2010.08.0191
- Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., DeAngelis, D. L., ..., y Ayllón, D. (2020). The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. *Journal of Artificial Societies and Social Simulation*, 23(2).
- Hegselmann, R., Krause, U. et al. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5.
- Helbing, D. (2012). *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*. Springer.
- Kurmyshev, E., y Abrica-Jacinto, N. L. (2022). The effect of agents' psychology and social environment on the opinion formation: C/PA relative agreement model in sw and sf societies. *Chaos Theory and Applications*, 4, 212-225.
- Kurmyshev, E., Juárez, H. A., y González-Silva, R. A. (2011). Dynamics of bounded confidence opinion in heterogeneous social networks: Concord against partial antagonism. *Physica A: Statistical Mechanics and its Applications*, 390, 2945-2955.
- Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12), 1819-1838. <https://doi.org/10.1142/S0129183107011789>
- Miller, J. H., y Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press.
- Rodríguez-Zoya, L. G., y Roggero, P. (2015). Modelos basados en agentes: aportes epistemológicos y teóricos para la investigación social. *Revista Mexicana de Ciencias Políticas y Sociales*, 60(225), 227-261.
- Wang, C. (2022). Opinion dynamics with higher-order bounded confidence. *Entropy*, 24, 1300.
- Watts, D. J., y Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393, 440-442.

Capítulo 2

Identificando ecuaciones parsimoniosas que gobiernan la dinámica de modelos basados en agentes. Una primera aproximación

Jorge Zazueta Gutiérrez¹

Introducción

Los modelos basados en agentes (MBA) son ampliamente utilizados en las ciencias sociales tanto para replicar y estudiar fenómenos reales complejos como para explorar propiedades emergentes de la interacción entre agentes artificiales (Elsenbroich y Polhill, 2023; Giardini *et al.*, 2024). Paquetes especializados como NetLogo (Wilensky, 1999) o Mesa (Masad y Kazil, 2015) equipan al investigador con una herramienta razonablemente intuitiva para desarrollar modelos detallados de simulación y analizar resultados.

A pesar de estas ventajas, los MBA sufren de dos problemas principales. Son difíciles de interpretar y de generalizar. En contraste, los modelos basados en ecuaciones diferenciales (ED) tienden a ser más interpretables y generalizables a cambio de abstraer el minucioso detalle de los MBA (Nardini *et al.*, 2021). Esta abstracción corresponde a una reducción de dimensiones, expresada como un sistema acoplado de

1. Facultad de Economía. Universidad Autónoma de San Luis Potosí. ORCID: <https://orcid.org/0000-0001-7138-2753>. Correo electrónico: jorge.zazueta@uaslp.mx.

ecuaciones diferenciales, que permite utilizar técnicas estándar de sistemas dinámicos para estudiar su comportamiento y hacer predicciones.

En respuesta a esta necesidad, se han desarrollado metodologías para extraer modelos simbólicos ya sea a partir de datos observados o simulados (North *et al.*, 2022; Ghadami y Epureanu, 2022; Velázquez-Sánchez *et al.*, 2024; Wang *et al.*, 2016). En particular, el método de *Identificación Dispersa de Dinámicas No-lineales* (SINDy, por sus siglas en inglés) introducido por Brunton *et al.* (2016) ofrece un proceso intuitivo y computacionalmente accesible para trabajar con MBA.

En este trabajo utilizamos como materia prima un modelo de presa-depredador basado en agentes desarrollado por Niemann *et al.* (2021) y exploramos posibles ecuaciones regidoras del sistema.

Objetivos

Este trabajo tiene dos objetivos principales:

- Desarrollar una metodología simple para la identificación de sistemas de ecuaciones diferenciales parsimoniosas que representen la dinámica generada por un modelo de agentes.
- Presentar el estudio de caso de un modelo de presa-depredador.

Metodología

La figura 1 describe esquemáticamente la metodología propuesta, que se centra en la aplicación del método de identificación dispersa de sistemas no-lineales (SINDy,² por sus siglas en inglés). A continuación describimos sus elementos individuales.

Generación de un MBA

Asumimos que existe un modelo basado en agentes que genera una serie de tiempo describiendo el comportamiento del fenómeno a estudiar.³ Un elemento común en este tipo de modelado es la introducción

2. Sparse Identification of Nonlinear Dynamics (SINDy).

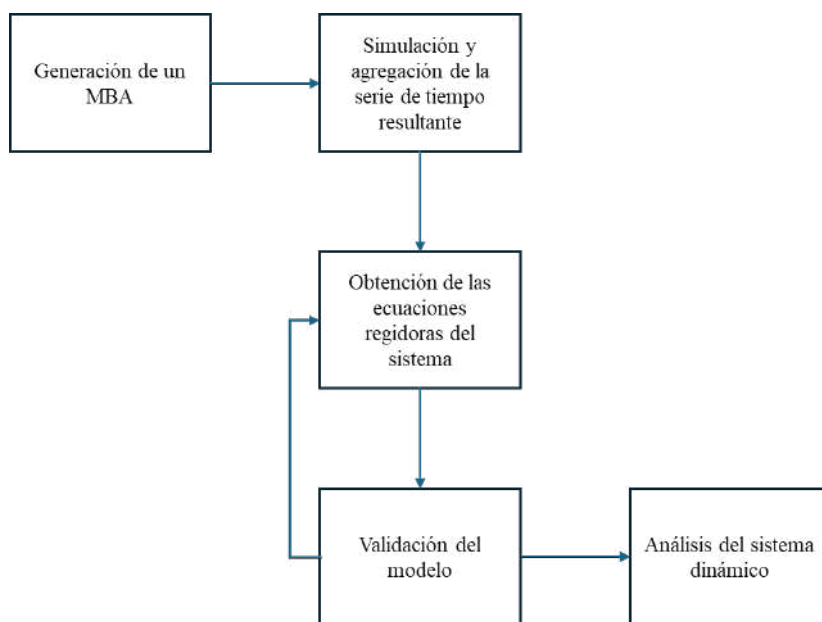
3. Es claro que los modelos basados en agentes (MBA) no solamente generan series de tiempo. Un aspecto importante y común incluye distribuciones espaciales en diferentes topologías,

de aleatoriedad en las reglas de comportamiento de los agentes, por lo que la serie de tiempo resultante es típicamente ruidosa (ver figura 2).

Simulación y agregación de series de tiempo resultantes

Una manera común de reducir el ruido generado por la estocasticidad del modelo es generando un número de simulaciones con parámetros idénticos, pero con semillas distintas generadas también aleatoriamente para *suavizar* el efecto del ruido estadístico y mejorar la identificación de las ecuaciones regidoras.

Figura 1. Metodología sugerida



La figura 3 muestra este efecto, al promediar 500 trayectorias del mismo modelo, variando únicamente la semilla del generador de

sin embargo, en este trabajo nos restringimos al caso de las series de tiempo.

números aleatorios. La figura es claramente más suave, aunque sugiere una acumulación de ruido a través del tiempo.

Figura 2. Simulando una sola trayectoria

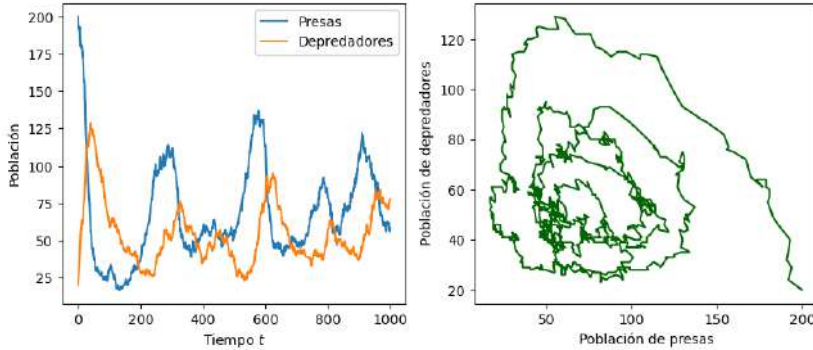
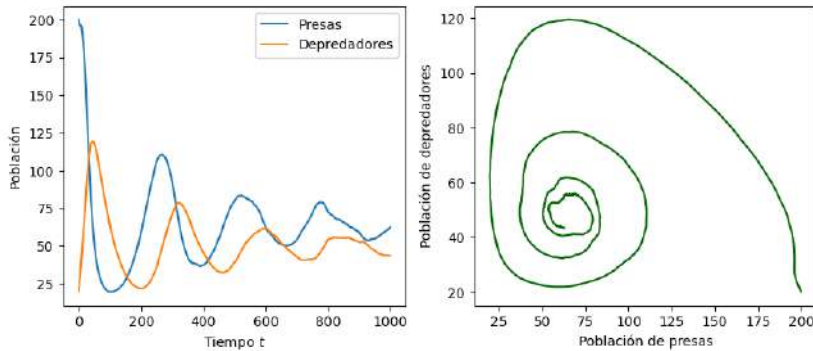


Figura 3. Promediando 500 trayectorias



Obtención de las ecuaciones regidoras del sistema

A partir de los datos suavizados, podemos estimar el sistema dinámico que gobierna el comportamiento del sistema utilizando el método SINDy que describimos brevemente a continuación.

SINDy fue propuesto originalmente por Brunton *et al.* (2016) han refinado y enriquecido el método desde entonces (Fasel *et al.*, 2022; Messenger y Bortz, 2021; Wentz y Doostan, 2023).

Uno de los principales supuestos de SINDY es que las ecuaciones regidoras de un sistema dinámico son típicamente dispersas. Esto es, tienden a depender de pocas variables. Ejemplos de esta observación abundan en física y biología. Si bien las ciencias sociales no cuentan aún con una historia de modelado disperso tan rica como otras disciplinas, en este trabajo asumimos que es razonable esperar un comportamiento similar.

Consideremos el sistema dinámico de la forma

$$\dot{x}(t) = f(x(t)), \quad (1)$$

donde $\dot{x}(t) \in \mathbb{R}^n$. Para determinar la forma explícita de $f(x)$ muestreemos la trayectoria de $x(t)$, que en nuestro caso proviene de la simulación de un MBA, y construimos dos matrices.

$$X = \begin{pmatrix} x_1(t_1) & \cdots & x_n(t_1) \\ \vdots & \ddots & \vdots \\ x_1(t_n) & \cdots & x_n(t_n) \end{pmatrix} \text{ and } \dot{X} = \begin{pmatrix} \dot{x}_1(t_1) & \cdots & \dot{x}_n(t_1) \\ \vdots & \ddots & \vdots \\ \dot{x}_1(t_n) & \cdots & \dot{x}_n(t_n) \end{pmatrix}$$

Donde \dot{X} se deriva numéricamente. A continuación, construimos la matriz $\Theta(X)$ de funciones no-lineales candidatas de las columnas de X :

$$\Theta(X) = \begin{pmatrix} | & | & | & \cdots \\ \theta_1(X) & \theta_2(X) & \theta_3(X) & \cdots \\ | & | & | & \end{pmatrix}$$

De este modo, podemos representar las posibles formas de la ecuación 1 como

$$\dot{X} = \Theta(X)\Xi$$

Donde $\Xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ es un vector de coeficientes *dispersos* que determinan cuáles de las funciones de $\Theta(X)$ están activas, lo que implica regularizar el problema de regresión utilizando, por ejemplo, LASSO o STLSQ.

Una vez determinada Ξ , los renglones de la ecuación 1 se construyen de la siguiente manera:

$$\dot{x}_k = f_k(x) = \Theta(x^T)\xi_k$$

Donde $\Theta(x^T)$ es un vector de funciones simbólicas, mientras que $\Theta(X)$ es una matriz numérica.

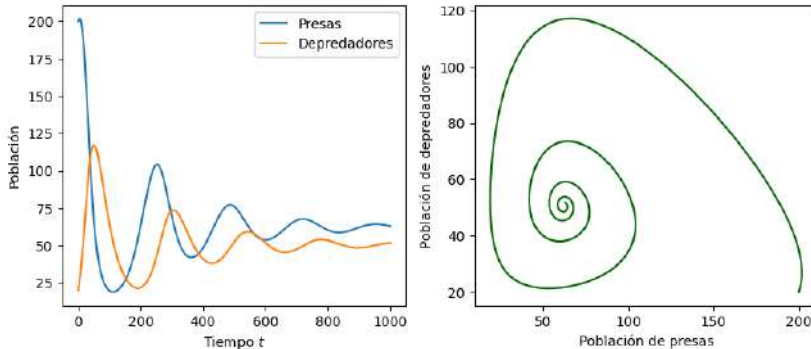
Resultados preliminares

Aplicando el modelo descrito en la sección anterior al modelo basado en agentes de una dinámica de presa-depredador obtenemos el siguiente sistema de ecuaciones diferenciales ordinarias no-lineales acopladas:

$$\begin{cases} \dot{p}(t) = 0.035 s(t) - 0.002d(t) - 0.001s(t)d(t) \\ \dot{d}(t) = 0.002s(t) - 0.025w(t) \end{cases} \quad (2)$$

Simulando el sistema (2) observamos que la dinámica replica las características cualitativas del modelo original.

Figura 4. Simulación del sistema obtenido a través de SINDy



Discusión

La dinámica del modelo analítico obtenido refleja razonablemente el comportamiento del sistema original, al menos de manera cualitativa. Es interesante observar la acumulación de ruido en el tiempo en el sistema original. Aparentemente, esto es un fenómeno común en los MBA debido a su naturaleza estocástica. La acumulación de ruido en la simulación afecta el desempeño de SINDy dado que la estimación numérica de las derivadas introduce ruido adicional, exacerbando el problema de detección. Para atenuar este efecto, utilizamos métodos numéricos de diferenciación suavizada y optimizamos el problema disperso utilizando ensambles. Tanto el trabajo numérico como las gráficas fueron realizados en Python con las librerías Numpy, Matplotlib, Mesa y PysINDy.

Conclusiones

Este trabajo introduce el potencial de los métodos de aprendizaje automático científico como herramientas de análisis de modelos basados en agentes que generan series de tiempo, al representar dichas dinámicas de manera analítica. En particular, se abre el camino a la exploración directa del efecto de diferentes condiciones iniciales y la identificación y clasificación de puntos fijos. Un problema abierto tanto en la identificación de sistemas dinámicos como en el modelado basado en agentes es la acumulación de ruido en el sistema a través del tiempo.

Referencias

- Brunton, S. L., Proctor, J. L., y Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15), 3932-3937. <https://doi.org/10.1073/pnas.1517384113>
- Elsenbroich, C., y Polhill, J. G. (2023). Agent-based modelling as a method for prediction in complex social systems. *International Journal of Social*

- Research Methodology*, 26(2), 133-142. <https://doi.org/10.1080/13645579.2023.2152007>
- Fasel, U., Kutz, J. N., Brunton, B. W., y Brunton, S. L. (2022). Ensemble-SINDY: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 478(2260), 20210904. <https://doi.org/10.1098/rspa.2021.0904>
- Ghadami, A., y Epureanu, B. I. (2022). Data-driven prediction in dynamical systems: Recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2229), 20210213. <https://doi.org/10.1098/rsta.2021.0213>
- Giardini, F., Borit, M., Verhagen, H., y Wijermans, N. (2024). Modeling Realistic Human Behavior in Disasters. A Rapid Literature Review of Agent-Based Models Reviews. En C. Elsenbroich y H. Verhagen (Eds.), *Advances in Social Simulation* (pp. 151-162). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-57785-7_13
- Masad, D., y Kazil, J. (2015). Mesa: An Agent-Based Modeling Framework. *Proceedings of the 14th Python in Science Conference*, 51-58. Proceedings of the 14th Python in Science Conference. <https://doi.org/10.25080/Majora-7b98e3ed-009>
- Messenger, D. A., y Bortz, D. M. (2021). Weak SINDY: Galerkin-Based Data-Driven Model Selection. *Multiscale Modeling & Simulation*, 19(3), 1474-1497. <https://doi.org/10.1137/20M1343166>
- Nardini, J. T., Baker, R. E., Simpson, M. J., y Flores, K. B. (2021). Learning differential equation models from stochastic agent-based model simulations. *Journal of the Royal Society Interface*, 18(176), 20200987. <https://doi.org/10.1098/rsif.2020.0987>
- Niemann, J.-H., Klus, S., y Schütte, C. (2021). Data-driven model reduction of agent-based systems using the Koopman generator. *PLOS ONE*, 16(5), e0250970. <https://doi.org/10.1371/journal.pone.0250970>
- North, J. S., Wikle, C. K., y Schliep, E. M. (2022). *A Review of Data-Driven Discovery for Dynamic Systems* (arXiv:2210.10663). arXiv. <http://arxiv.org/abs/2210.10663>
- Velázquez-Sánchez, R. D., Escobedo-Alva, J. O., Peña-García, R., Tapia-Herrera, R., y Meda-Campaña, J. A. (2024). Identification of High-Order Nonlinear Coupled Systems Using a Data-Driven Approach. *Applied Sciences*, 14(9), Article 9. <https://doi.org/10.3390/app14093864>
- Wang, W.-X., Lai, Y.-C., y Grebogi, C. (2016). Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports*, 644, 1-76. <https://doi.org/10.1016/j.physrep.2016.06.004>

- Wentz, J., y Doostan, A. (2023). Derivative-based SINDy (DSINDy): Addressing the challenge of discovering governing equations from noisy data. *Computer Methods in Applied Mechanics and Engineering*, 413, 116096. <https://doi.org/10.1016/j.cma.2023.116096>
- Wilensky, U. (1999). *NetLogo*. <https://ccl.northwestern.edu/netlogo/>

Capítulo 3

Diseminación cultural con aumento de presencia de medios de comunicación en el modelo de Axelrod

Sadan Josué López Morales¹

Ricardo Armando Gonzalez Silva²

Mario Ignacio Gonzalez Silva³

Antecedentes

La diseminación cultural es el proceso de propagación y adopción de ideas, creencias, habilidades, valores y comportamientos entre los miembros de una sociedad. El primer modelo de simulación para este proceso es el AM propuesto por Axelrod (1997).

El AM emplea modelos basados en agentes (ABM, por sus siglas en inglés) para representar a los individuos en una sociedad. Los ABM son marcos computacionales utilizados para simular las interacciones de agentes individuales que representan entidades tales como personas,

1. Centro Universitario de Los Lagos (Universidad de Guadalajara). Correo electrónico: sadan.lopez8247@alumnos.udg.mx. ORCID: <https://orcid.org/0009-0002-2582-8683>
2. Centro Universitario de Los Lagos (Universidad de Guadalajara). Correo electrónico: ricardo.gsilva@academicos.udg.mx. ORCID: <https://orcid.org/0000-0002-7448-3178>
3. Centro Universitario de Los Lagos (Universidad de Guadalajara). Correo electrónico: mario.gonzalezs@academicos.udg.mx. ORCID: <https://orcid.org/0000-0002-1238-8909>

organizaciones o, incluso, organismos biológicos dentro de un entorno (Railsback y Grimm, 2019).

Cada agente en estos modelos está diseñado con características, reglas y comportamientos específicos, lo que le permite actuar de manera autónoma e interactuar con otros agentes según protocolos predefinidos. En el AM, la cultura de un agente consiste en atributos tales como creencias, valores y comportamientos.

El modelo simula interacciones entre estos agentes bajo dos supuestos clave:

- Es más probable que los individuos interactúen con otros que posean atributos culturales similares.
- Las interacciones tienden a aumentar el número de atributos compartidos, lo que refuerza la similitud cultural y aumenta la probabilidad de futuras interacciones.

Varias extensiones del AM exploran factores que influyen en la diseminación cultural, como los efectos del ruido y el rol de la topología de la red en el orden cultural (Klemm *et al.*, 2003). Otros estudios investigan los efectos homogeneizadores de la retroalimentación global, demostrando que una retroalimentación fuerte acelera la monocultura, mientras que una retroalimentación más débil mantiene una diversidad limitada (Shibanai *et al.*, 2001). Las estrategias de propaganda en el AM también han sido exploradas, destacando la importancia del momento y la intensidad para lograr una influencia efectiva (Carletti *et al.*, 2006).

Existen investigaciones sobre el modelo AM que abordan los mecanismos que sostienen la diversidad cultural, enfatizando cómo la confianza limitada y los rasgos métricos afectan la estabilidad del sistema (Flache y Macy, 2006). Otra aplicación del modelo AM se encuentra en la dinámica de opiniones, con experimentos que analizan cómo la tolerancia y la conectividad también forman grupos sólidos (Jacomier, 2005).

Además, la coevolución de grupos culturales y redes es discutida en Centola *et al.* (2007), subrayando cómo la dinámica de redes en forma de retícula determina la estabilidad del multiculturalismo o la convergencia hacia una sociedad completamente homogénea. El doble rol de los medios de comunicación masivos se estudia en González-Avella *et al.* (2005), donde una fuerte influencia mediática promueve la fragmentación cultural, mientras que una influencia débil impulsa la

homogeneización cultural. Además, Rodríguez *et al.* revelan cómo el uso estratégico de los medios puede manipular la percepción pública para fomentar tanto la resistencia como la convergencia cultural.

Los estudios sobre difusión cultural han evolucionado significativamente desde la introducción del AM. Numerosas investigaciones han refinado y ampliado el marco teórico para abordar diversos fenómenos sociales complejos. Una extensión clave examinó el impacto de los medios de comunicación masivos mediante la creación de un tipo de superagente que interactúa con todos los agentes del sistema, demostrando que los medios pueden impulsar la convergencia cultural en sociedades diversas o promover la diversidad cultural en contextos moderadamente diversos debido a la interacción entre los medios y las influencias locales (Rodríguez y Moreno, 2010).

Otras investigaciones han explorado el rol de la tolerancia cultural. Gracia-Lázaro *et al.* (2011) demostraron que los rasgos tolerantes poseen una ventaja evolutiva debido a su capacidad para facilitar la superposición cultural y mejorar las interacciones sociales. De manera similar, en Crokidakis (2012) se destacó cómo la influencia de los medios de comunicación puede suprimir fenómenos críticos y acelerar la formación de consensos en la dinámica de opinión.

Un estudio sobre el AM aplicado a interacciones interculturales reveló cómo la influencia recíproca de los medios entre distintos grupos sociales puede llevar a una diversidad estructurada, equilibrando los efectos homogenizantes globales con la heterogeneidad local (González-Avella *et al.*, 2012). En otro estudio, se identificaron estados ordenados similares a los estados quimera, demostrando que una población puede alcanzar un orden coherente mientras otras permanecen desordenadas (González-Avella *et al.*, 2014).

Investigaciones adicionales analizaron las estructuras jerárquicas de los rasgos culturales, revelando que una alta ultrametricidad y una variación significativa en los rasgos culturales son esenciales para preservar la diversidad en sistemas complejos (Stivala *et al.*, 2014). También se estudiaron las dinámicas de difusión de la innovación, mostrando que diferentes topologías de red influyen en la propagación de innovaciones, siendo las redes aleatorias altamente conectadas las que producen curvas de adopción en forma de S, características de los sistemas del mundo real (Tilles y Fontanari, 2015).

La interacción entre diversidad cultural y cooperación fue investigada por Stivala *et al.* (2016), demostrando que los estados multiculturales fomentan la cooperación, mientras que los estados monoculturales la inhiben. Hernández *et al.* (2018) examinaron mecanismos para preservar la diversidad cultural, desarrollando comunidades culturales robustas que evitan estados monoculturales mientras mantienen grupos fragmentados pero cohesionados.

El impacto y la evolución de las redes sociales en la difusión cultural fueron explorados por Raducha y Gubiec (2017), quienes demostraron cómo diferentes mecanismos de reconfiguración influyen en la topología de la red y en las configuraciones culturales. Además, se analizaron las dinámicas de los grupos minoritarios en sociedades globalizadas, revelando cómo ciertas condiciones estructurales permiten que estos grupos prosperen y coexistan a pesar de las tendencias dominantes de los medios de comunicación masivos (Cosenza *et al.*, 2020).

En 2021, se estudió la polarización en el modelo AM, revelando efectos disruptivos en la dinámica cultural y rasgos polarizantes que inhiben las transiciones tradicionales entre estados monoculturales y multiculturales, con énfasis en las divisiones ideológicas y las estructuras de red (Gracia-Lázaro *et al.*, 2021). En 2024, se investigó la competencia entre múltiples influencias mediáticas, descubriendo cómo los medios más débiles pueden dominar bajo ciertas condiciones y cómo pueden emerger estados ordenados alternativos a partir de las interacciones entre agentes, proporcionando información integral sobre el equilibrio entre interacciones locales, globales y estructurales (Alvarez-Llamoza *et al.*, 2024).

En nuestra investigación replicamos el modelo de difusión cultural descrito en Rodríguez *et al.* (2009), centrándonos en los índices cuantitativos de influencia directa e indirecta de los medios de comunicación masivos. Se introdujo un nuevo enfoque analítico, específicamente identificando configuraciones donde emergen estados coexistentes en el modelo (una idea equivalente a la de Campos Cantón, 2025).

Este nuevo enfoque emplea ciencia de datos, presentando un análisis gráfico, descriptivo y categorizado de los diferentes porcentajes de información de campo dentro de la red en retícula. Los resultados numéricos describen las relaciones entre variables dependientes e independientes, destacando los rangos de valores en los que la diversidad cultural fomenta la aparición de estados coexistentes.

Modelo

Características y grupos de los agentes

En este modelo, los agentes representan individuos con características o rasgos culturales distintos. Un vector de rasgos culturales define a cada agente, y su interacción está influenciada por el grado de similitud con otros agentes en el sistema.

Sea:

$$[q] = \{0, 1, \dots, q - 1\}, q \in \mathbb{N}$$

La estructura y agrupación de los agentes se describen de la siguiente manera:

Características de los agentes:

Cada agente i posee un vector de rasgos nominal de dimensión F , denotado por:

$$\vec{\sigma}_i = (\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{iF}) \in [q]^F$$

donde σ_{if} es el valor del rasgo f para el agente i , este puede tomar cualquier valor dentro del conjunto finito $[q]$, es decir, $\sigma_{if} \in [q]$. Estos rasgos representan la diversidad de posibles configuraciones culturales en el modelo y pueden simbolizar diversos aspectos de la cultura, como creencias, preferencias o comportamientos sociales.

Campo vectorial externo (VF) y condiciones externas:

Para incorporar la influencia de factores externos en los agentes, el modelo introduce un campo vectorial externo (VF), que representa una fuente constante de influencia cultural análoga a los medios de comunicación masivos (M). Este VF actúa sobre todos los agentes con el objetivo de inducir la adopción de ciertos rasgos culturales dentro del sistema (ecuación 6). Se caracteriza por

$$\vec{\sigma}_M = (\sigma_{M1}, \sigma_{M2}, \dots, \sigma_{MF})$$

y está diseñado para compartir al menos un rasgo con cada agente de la red, asegurando una conexión mínima y una interacción con todos los individuos (González-Avella *et al.*, 2005).

Grupos de agentes:

Según la similitud de características, los agentes pueden agruparse en dos categorías principales:

- Grupo A: Consiste en agentes con alta similitud cultural, lo que aumenta su probabilidad de interacción y fomenta la homogeneización. La interacción frecuente entre los agentes del Grupo A facilita la transmisión de rasgos culturales a través de redes locales de influencia. Este grupo se caracteriza por una alta probabilidad de interacción con agentes culturalmente similares, convirtiéndolos en puentes de influencia cultural hacia otros agentes dentro del sistema.
- Grupo B: Este grupo está compuesto por agentes que poseen pocos o ningún rasgo en común con otros agentes en su vecindario. La baja similitud en los rasgos reduce la probabilidad de interacción y adopción de nuevas características culturales, permitiendo a los agentes del Grupo B mantener una identidad cultural distinta o formar subculturas aisladas dentro del sistema. Para estos agentes, la probabilidad de interacción directa con otros es baja. Sin embargo, si están ubicados cerca de agentes del Grupo A, pueden adoptar ciertos rasgos culturales a través de una interacción indirecta, mediada por las interacciones de los agentes del Grupo A con otros.

Entorno y condiciones externas

El modelo propuesto simula una sociedad estructurada en una retícula G_{LL} (de dimensión $L \times L$), donde cada sitio en la cuadrícula representa un agente con características culturales definidas. Este entorno espacial permite que los agentes interactúen principalmente con sus cuatro vecinos más cercanos, promoviendo la difusión de rasgos culturales basada en la similitud de características.

El efecto del ν_f sobre los agentes se cuantifica y regula mediante dos parámetros fundamentales:

RASGOS EFECTIVOS ()

Este parámetro representa una cantidad adicional de rasgos que el ν_F comparte de manera efectiva con cada agente. Estos rasgos efectivos actúan como elementos universales (como el lenguaje o símbolos ampliamente comprendidos) que facilitan la interacción y la influencia del ν_F , incluso cuando no hay coincidencia en los rasgos culturales nominales de un agente. Así, aumenta la probabilidad de interacción entre el ν_F y los agentes, permitiendo que todos los agentes tengan la posibilidad de ser influenciados.

VALOR DE CONFIANZA (C)

Este parámetro mide la credibilidad o aceptación de la información proveniente del ν_F . La confianza (C) se incorpora como un factor de probabilidad adicional (explicado en la sección de actualización de rasgos culturales) que determina la probabilidad de que un agente adopte directamente un rasgo del ν_F . Valores altos de C aumentan la influencia del ν_F y promueven una mayor homogeneización cultural dentro del sistema. En contraste, valores bajos reducen su impacto, permitiendo la coexistencia de diversas configuraciones culturales.

Dinámica y actualización de agentes

El modelo define las reglas de interacción que moldean la evolución cultural a lo largo del tiempo. Estas reglas dictan cómo los agentes modifican sus rasgos culturales en función de las interacciones con sus vecinos y las influencias externas. Se basan en la similitud cultural y en la influencia de las interacciones locales con otros agentes, así como en la influencia externa representada por el ν_F . A continuación se detallan los principales procesos internos y el mecanismo de actualización de los agentes.

Probabilidad nominal de interacción

La interacción entre dos agentes i y j está condicionada por el grado de superposición o similitud entre sus rasgos culturales. Esta similitud se mide a través del traslape $l(i,j)$, definido como el número de rasgos compartidos entre ambos agentes. Matemáticamente, se expresa como:

$$l(i,j) = \sum_{f=1}^F \delta_{\sigma_{if}, \sigma_{jf}} \quad (1)$$

donde δ es la función de Kronecker, que toma el valor 1 si $\sigma_{if} = \sigma_{jf}$ y 0 en caso contrario. La probabilidad nominal de interacción entre los agentes i y j está dada por:

$$p(i,j) = \frac{l(i,j)}{F} \quad (2)$$

Esto implica que la probabilidad de que dos agentes interactúen e influyan mutuamente aumenta con la cantidad de rasgos culturales compartidos.

Probabilidad extendida de interacción

Para modelar la interacción entre los agentes y el campo vectorial externo (\mathbf{vF}), así como la probabilidad de que los agentes adopten rasgos externos, se define una probabilidad extendida de interacción. Esta probabilidad incorpora tanto los rasgos culturales nominales de los agentes como los rasgos efectivos compartidos. Se expresa como:

$$p(i, M) = \frac{l(i, M) + \epsilon}{F + \epsilon} = \frac{l(i, M)/F + \epsilon/F}{1 + \epsilon/F} \quad (3)$$

Se cumple que: $p(i, M) \geq l(i, M)/F$. La igualdad se da si $\epsilon = 0$ o si $l(i, M) = F$.

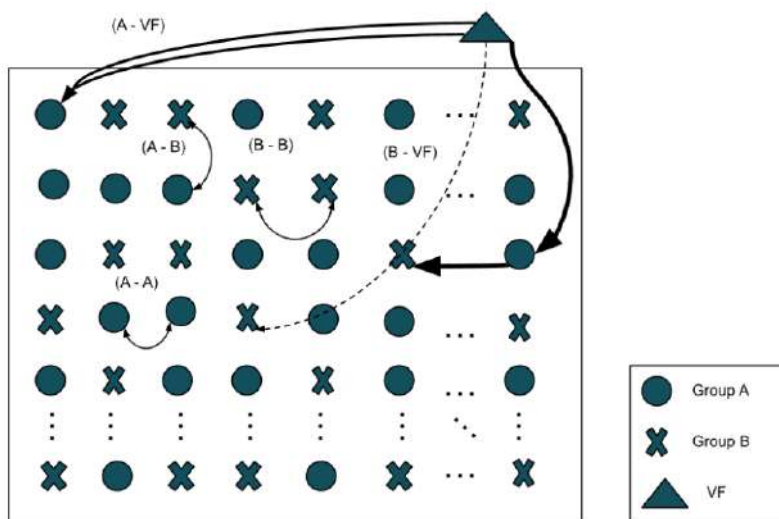
Esta probabilidad extendida $p(i, M)$ regula la interacción efectiva entre cada agente y el \mathbf{vF} , aumentando la posibilidad de que los agentes adopten rasgos del \mathbf{vF} incluso cuando la superposición $l(i, M)$ es baja. El parámetro ϵ amplía la influencia del \mathbf{vF} sobre el sistema (Rodríguez *et al.*, 2009).

Interacción entre grupos de agentes

Como se mencionó anteriormente, los agentes se dividen en dos grupos principales según su relación con el VF:

- Grupo A: Incluye agentes que comparten al menos un rasgo con el VF. Estos agentes pueden interactuar directamente con el VF y recibir su influencia de manera directa.
- Grupo B: Son agentes que no comparten ningún rasgo con el VF y, por tanto, no pueden interactuar directamente con él. Sin embargo, pueden verse influenciados indirectamente a través de sus interacciones con agentes del Grupo A.

Figura 1. Interacción entre grupos de agentes



Nota: Esta es una representación de agentes e interacciones dinámicas en la red de malla con VF incluida (Rodríguez *et al.*, 2009). Las diferentes líneas representan los tipos de interacción. Las líneas dobles representan la interacción entre los agentes del Grupo A con el VF, las líneas delgadas representan interacciones dentro del Grupo A y dentro del Grupo B. La interacción indirecta del VF hacia los agentes del Grupo B, mediada por los agentes del Grupo A, se representa con líneas gruesas. Las líneas discontinuas representan la interacción directa entre el VF y los agentes del Grupo B.

Las interacciones en el modelo se pueden clasificar de la siguiente manera:

- *Interacciones entre agentes del Grupo A y el VF ($A-VF$):* Los agentes del Grupo A, al tener características similares al VF, pueden interactuar directamente con él y adoptar sus rasgos. Estas interacciones son unidireccionales y sincrónicas, es decir, todas las actualizaciones ocurren en el mismo paso de tiempo, permitiendo que el VF influya directamente en el Grupo A, difundiéndose sus características culturales. Esta relación es fundamental, ya que convierte a los agentes del Grupo A en un canal a través del cual los rasgos del VF pueden potencialmente extenderse por toda la sociedad.
- *Interacciones entre agentes del Grupo A ($A-A$):* Los agentes del Grupo A interactúan con frecuencia entre sí debido a sus similitudes culturales, lo que facilita la propagación de los rasgos adoptados del VF dentro del grupo. Estas interacciones son unidireccionales y sincrónicas, consolidando y fortaleciendo la influencia del VF en el Grupo A, ya que los rasgos adoptados por un agente del VF pueden compartirse rápidamente con otros agentes del mismo grupo. De este modo, los agentes del Grupo A tienden a formar grupos homogéneos que están culturalmente alineados con el VF.
- *Interacciones entre agentes del Grupo B ($B-B$):* Los agentes del Grupo B, que no comparten rasgos con el VF, interactúan entre sí basándose en sus propias características culturales. Estas interacciones son unidireccionales y sincrónicas, reforzando su identidad separada, ya que estos agentes no están directamente expuestos a los rasgos del VF y, por tanto, son menos susceptibles a su influencia. Esta dinámica conduce a la formación de subculturas dentro de la red.
- *Interacciones entre agentes del Grupo A y del Grupo B ($A-B$):* Aunque los agentes del Grupo B no interactúan directamente con el VF, pueden verse influenciados indirectamente a través de sus interacciones con los agentes del Grupo A. Cuando un agente del Grupo A que ya ha adoptado un rasgo del VF interactúa con un agente del Grupo B, existe la posibilidad de que este último adopte dicho rasgo. Esta interacción es unidireccional y sincrónica, ya que los rasgos pueden potencialmente compartirse en ambas direcciones. De esta manera, el Grupo A actúa como un puente, permitiendo que la influencia del VF se propague al Grupo B, aunque de manera indirecta y con menor frecuencia. Esta cadena de interacciones

(A-VF seguido de A-B) permite que los rasgos del VF afecten todo el sistema, incluso alcanzando a los agentes culturalmente más distantes del Grupo B.

- *Interacción entre el Grupo B y el VF (B-VF)*: No hay interacción directa entre el VF y los agentes del Grupo B debido a la ausencia de rasgos compartidos. En el modelo, esta falta de interacción directa significa que los agentes del Grupo B están “aislados” de la influencia directa del VF. Sin embargo, el modelo incorpora rasgos efectivos (ϵ) para simular una conexión mínima y unidireccional que el VF mantiene con todos los agentes, incluidos aquellos del Grupo B.

Actualización de características culturales

La dinámica de actualización de los agentes se lleva a cabo mediante un proceso iterativo, siguiendo los siguientes pasos:

- *Selección del agente activo*: En cada iteración, se selecciona aleatoriamente un agente i para que actúe como el agente *activo* en la interacción.
- *Selección del agente de interacción*: Para cada agente activo, se selecciona aleatoriamente un vecino cercano (uno de los cuatro vecinos más próximos) o el VF para la fase de interacción.
- *Determinación de la superposición $I(i, s)$* : Se calcula la superposición, donde $s = j$ para un agente vecino o $s = M$ para el VF. Cuando $s = M$, el agente i interactúa con el VF utilizando la probabilidad extendida (3). Por otro lado, si $s = j$ y $0 < I(i, j) < F$, los agentes i y j interactúan con la probabilidad nominal (2). En ambos casos, si $I(i, s) < F$, la interacción es necesaria; de lo contrario, cuando $I(i, s) = F$, la interacción no es necesaria porque ya comparten todos los rasgos culturales.
- *Dinámica de copia de rasgos*: Si la interacción es válida, el agente activo i procede a copiar un rasgo del agente j con cierta probabilidad, de la siguiente manera: se elige aleatoriamente una posición de rasgo f , tal que $\sigma_{if} \neq \sigma_{jf}$. Si el agente de interacción es un vecino j , el agente i copia el rasgo σ_{jf} ; si el agente de interacción es el VF, el agente i copia el rasgo σ_{Mf} del VF. Cualquiera de estos casos ocurre con una probabilidad p .

La probabilidad p' de adoptar el rasgo copiado depende del valor de confianza C . Esta probabilidad p' regula si la característica seleccionada es copiada o no, según la siguiente regla:

$$p' = p'(i, s) = \begin{cases} C, & \text{if } \sigma_{sf} = \sigma_{Mf} \text{ or } \sigma_{jf} = \sigma_{Mf}. \\ 1 - C, & \text{if } \sigma_{if} = \sigma_{Mf}. \\ 1, & \text{if } \sigma_{if} \neq \sigma_{Mf} \neq \sigma_{jf}. \end{cases} \quad (4)$$

Probabilidad de copia entre agentes: La probabilidad total de que el agente activo i copie un rasgo del agente de interacción s (donde s puede ser un vecino j o el VF) está dada por:

$$P(i, s) = \begin{cases} \frac{1-B}{4} \cdot p(i, j) \cdot p'; & s = j. \\ B \cdot p(i, M) \cdot p'; & s = M. \end{cases} \quad (5)$$

Donde B representa la probabilidad de elegir interacción con el VF, esta toma valores desde 0.2 hasta 0.5, al aumentar el valor de B la probabilidad restante se divide entre los cuatro vecinos del agente i , y p' es la probabilidad adicional de copiar o descartar el rasgo en función del valor de confianza C .

Dado que G_{LL} puede considerarse un subconjunto de \mathbb{R}^{F+1} y $[q] \subset \mathbb{R}^{F+1}$, los medios masivos M , definidos como un campo vectorial VF, tienen la siguiente forma:

$$VF : G_{LL} \rightarrow [q]^F, VF(i) = \sigma'_i \quad (6)$$

Donde σ'_i es similar a σ_i pero con la posición f modificada según las reglas de la dinámica de copia de rasgos cuando el agente i interactúa con el VF.

Estados absorbentes

Cuando la dinámica de evolución del sistema se detiene o no experimenta más cambios, se dice que el modelo ha alcanzado un estado absorbente. Los estados absorbentes del modelo pueden ser:

- *Estados monoculturales*: Representan una clase de estados absorbentes caracterizados por una homogeneidad cultural completa dentro de la red. En estos estados, $l(i,s) = F$.
- *Estados n -culturales*: Este estado absorbente está organizado en múltiples regiones culturales homogéneas, distintas entre sí. En estos estados $l(i,s) = 0$ entre agentes de diferentes regiones culturales y $l(i,s) = F$ entre agentes de la misma región cultural.

Regiones culturales y la región cultural más grande (S_{max})

Un camino W en una red de malla es una secuencia ordenada de vértices que comienza en un vértice inicial y termina en un vértice final, de manera que cada par consecutivo de vértices en la secuencia está conectado por una arista.

Para una característica cultural σ , definimos un w_σ - camino como un conjunto de nodos (agentes) tal que $\sigma_i = \sigma$ para cada agente (nodo) en un camino w . También definimos una σ región cultural como:

$$R_\sigma = \{i: \sigma_i = \sigma \forall i, j \sigma_i = \sigma = \sigma_j \text{ existe un } w_\sigma \text{ - camino}\}$$

Finalmente, definimos S_{max} como el tamaño de la región cultural más grande. De la forma:

$$S_{max} = \max_\sigma \{|R_\sigma|\}$$

El parámetro S_{max} se utiliza para cuantificar la estructura cultural final del sistema mediante la evaluación de la homogeneidad o fragmentación de la red. Este valor corresponde al número de agentes que pertenecen al dominio cultural más grande, es decir, al conjunto más grande de agentes que comparten exactamente el mismo vector de rasgos culturales. Además, en el contexto de los estados absorbentes, podemos definir:

$S_{max} \approx L^2$. Estado monocultural.

$S_{max} \ll L^2$. Estado multicultural.

El análisis de S_{max} en función de la diversidad cultural inicial q y otros parámetros del modelo (como la confianza C y las características efectivas ϵ) permite estudiar la transición entre los estados monocultural y multicultural.

Porcentaje de información ν_F (ρ)

El porcentaje de información ν_F (ρ) mide cuánto de la información externa ν_F ha sido adoptada por los agentes al final de la dinámica. Este parámetro se define como la superposición promedio entre los rasgos culturales de los agentes y la ν_F , expresado como un porcentaje:

$$\rho = 100 \times \frac{1}{L^2 F} \sum_{i=1}^{L^2} \sum_{f=1}^F \delta_{\sigma_{if}, \sigma_{Mf}} \quad (7)$$

El análisis de ρ evalúa la efectividad de la ν_F en la homogeneización cultural del modelo y complementa métricas como S_{max} y el número de culturas para caracterizar la dinámica global del modelo.

Las secciones anteriores corresponden a López Morales *et al.* (2025), con la excepción de la ecuación (5), la cual ha sido modificada en esta investigación. Este ajuste resulta relevante, ya que introduce un incremento en la presencia de los medios de comunicación dentro de las interacciones. Es decir, la probabilidad de que el ν_F sea seleccionado como agente de interacción es mayor en comparación con la de cualquier vecino, a diferencia del enfoque adoptado en López Morales *et al.* (2025), donde dicha probabilidad se distribuye de manera homogénea entre los agentes y el ν_F .

Diseño experimental

El diseño experimental se llevó a cabo con los siguientes parámetros y características:

Hay L^2 agentes, donde $L = 30$.

Cada agente i tiene un vector de características nominales $\overline{\sigma}_i$ con un valor fijo de $F = 4$ rasgos, y cada uno toma valores en $[q]$.

Se incluyen tres parámetros: B , C y ϵ .

El parámetro B toma valores de 0.2 a 0.5 con pasos de 0.1, este maximiza la presencia de los medios de comunicación ante el agente i .

El parámetro ϵ toma los valores de 0.01, 0.1, 0.50 y 1, maximizando la cantidad de información.

El parámetro de confianza C varía de 0 a 0.45 en incrementos de 0.05 y representa la confianza en la información de los medios de comunicación masivos.

El parámetro de simulación $[q]$ toma valores de 1 a 39.

La simulación se dividió en tres experimentos. En cada experimento B , ϵ y C toman el rango completo de valores, y $[q]$ se divide en tres intervalos, de la siguiente manera:

Experimento 1: $[q] \in \{1, \dots, 13\}$

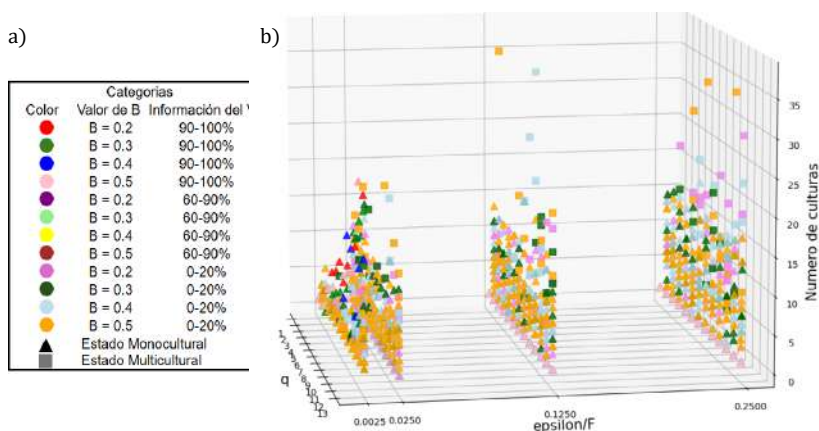
Experimento 2: $[q] \in \{14, \dots, 26\}$

Experimento 3: $[q] \in \{27, \dots, 39\}$

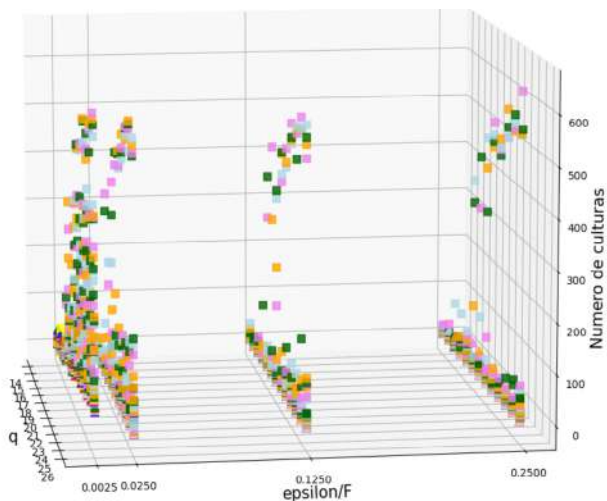
Una vez finalizados los experimentos, cada uno generó una base de datos. La base de datos completa contiene aproximadamente 225 millones de datos (2.33 GB).

Análisis y resultados

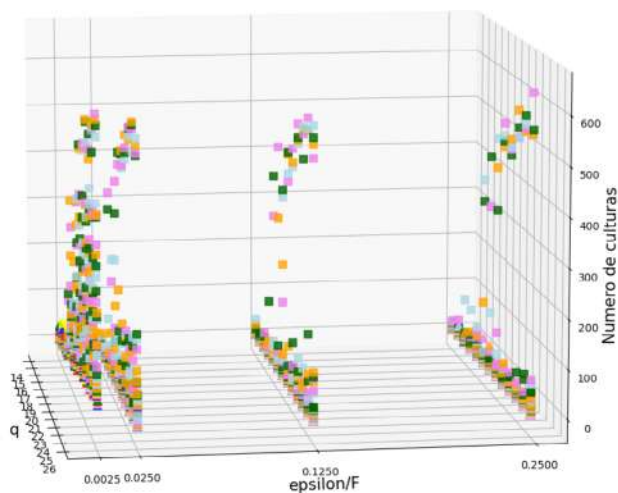
Figura 2. Cambios en el número de culturas



d)



c)



Nota: Este conjunto de gráficos presenta la variación en el número de culturas en función de los valores de ϵ/F . a) Se muestran las distintas categorías de clasificación de puntos según los valores del parámetro B, así como los porcentajes de información correspondiente al ϵ/F dentro de la red. b), c) y d) exponen la evolución del número de culturas obtenida en los experimentos 1, 2 y 3, respectivamente.

EXPERIMENTO 1

En este experimento se pueden apreciar claramente monoculturas, la escala nos permite ver que tenemos diversos triángulos de colores alrededor de la altura cinco, a partir de ahí se van generando más culturas, según el grado de presencia del campo vectorial y el aumento de los valores de q . Existen algunos valores atípicos, alrededor de 35 culturas, para casos aunque la presencia del campo llega a 0.5, en este caso, podríamos decir que esas culturas tienen muy poca información del campo vectorial, esto casi siempre pasa para valores de q entre 6 y 13.

El patrón, en general, muestra que el mayor número de culturas es a lo más 15, aunque hay unos casos especiales donde llegan a 35. Los casos donde son menores que cinco culturas se consideran como monocultura.

Resaltamos que en múltiples casos se tiene una monocultura para valores de q menores que cinco y para diversos valores del rango de presencia del campo vectorial.

EXPERIMENTO 2

Esta imagen con una escala de 0 a 600 muestra la diversidad cultural que existe para valores pequeños de ϵ/F , se tiene una diversidad más grande que para valores mayores pero, en general, el patrón es que se tienen muchas culturas. A casos muy especiales de valores altos de la presencia del campo vectorial, y también de confianza cuando existe monocultura. La imagen en otra perspectiva mostraría una serie de triángulos en la parte inferior. Podemos resaltar en estos patrones que si la presencia del campo vectorial es igual a 0.3 se tiene multicultura.

En este caso, donde q tiene valores entre 18 y 26, la multiculturalidad predomina, no importa qué valores se tengan de la presencia del campo. En general, esa multiculturalidad surge con porcentajes de información del campo pequeña menores que 60 %.

EXPERIMENTO 3

En este gráfico de simulación tenemos una escala de 0 a 700, podemos ver los patrones en los cuales, nuevamente, para valores de ϵ/F pequeños se tiene una gran diversidad cultural para cualquier valor

de q ; además, para valores mayores de ϵ/F tenemos multiculturalidad muy grande, alrededor de 600 tipos de culturas, esto también sucede para valores donde la presencia del campo puede tomar casi cualquier valor a partir de 0.3. Otro de los patrones de estos datos es que no importa cuál sea la presencia del campo y se tienen alrededor de 100 culturas diferentes.

En este rango de valores de q de 27 a 39, la multiculturalidad es total aun para valores de porcentaje de información de más del 90 %.

Discusión

Nuestro modelo, en comparación con el propuesto en Rodríguez *et al.* (2009), tiene un tiempo de convergencia cultural más rápida a la adopción o repulsión de la información de los medios de comunicación, de igual manera.

En los cambios de valores de ϵ y B se logra apreciar cómo se generan mayores estados de multiculturalidad en comparación con Rodríguez *et al.* (2009) (ya que se aumenta gradualmente la presencia de los medios de comunicación en la red de individuos se genera un mayor porcentaje de información, pero, a su vez, este genera una mayor multiculturalidad).

Conclusiones

La inclusión de un campo vectorial externo que representa los efectos de los medios masivos de comunicación, así como el aumento de la presencia de este dentro de la red, permiten entender cómo estas fuerzas pueden influenciar una cultura homogénea o multicultural dentro de una sociedad.

Otra de las cosas que también podemos resaltar en los patrones de comportamiento de este modelo es la formación de monoculturas o de pocas culturas sin información del campo vectorial.

Referencias

- Alvarez-Llamoza, O., Cosenza, M. G., Gonzalez-Avella, J. C., Suarez, M. A., Tucci, K., y Valverde, P. (2024). Mass media competition and alternative ordering in social dynamics. *Physical Review E*, 110(2). <https://doi.org/10.1103/physreve.110.024311>
- Axelrod, R. (1997). The dissemination of culture. *Journal of Conflict Resolution*, 41(2), 203-226. <https://doi.org/10.1177/0022002797041002001>
- Campos Cantón, E. (2025). Orbits of families of discrete dynamical systems evolving in the natural numbers. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 35(1). <https://doi.org/10.1063/5.0233348>
- Carletti, T., Fanelli, D., Grolli, S., y Guarino, A. (2006). How to make an efficient propaganda. *Europhysics Letters (EPL)*, 74(2), 222-228. <https://doi.org/10.1209/epl/i2005-10536-9>
- Centola, D., González-Avella, J. C., Eguíluz, V. M., y San Miguel, M. (2007). Homophily, cultural drift, and the co-evolution of cultural groups. *Journal of Conflict Resolution*, 51(6), 905-929. <https://doi.org/10.1177/0022002707307632>
- Cosenza, M. G., Gavidia, M. E., y González-Avella, J. C. (2020). Against mass media trends: Minority growth in cultural globalization. *PLOS ONE*, 15(4), e0230923. <https://doi.org/10.1371/journal.pone.0230923>
- Crokidakis, N. (2012). Effects of mass media on opinion spreading in the Sznaid sociophysics model. *Physica A: Statistical Mechanics and Its Applications*, 391(4), 1729-1734. <https://doi.org/10.1016/j.physa.2011.11.038>
- Flache, A., y Macy, M. W. (2006). What sustains cultural diversity and what undermines it? Axelrod and beyond. arXiv preprint physics/0604201.
- González-Avella, J. C., Cosenza, M. G., y San Miguel, M. (2012). A model for cross-cultural reciprocal interactions through mass media. *PLOS ONE*, 7(12), e51035. <https://doi.org/10.1371/journal.pone.0051035>
- (2014). Localized coherence in two interacting populations of social agents. *Physica A: Statistical Mechanics and Its Applications*, 399, 24-30. <https://doi.org/10.1016/j.physa.2013.12.035>
- González-Avella, J. C., Cosenza, M. G., y Tucci, K. (2005). Nonequilibrium transition induced by mass media in a model for social influence. *Physical Review E*, 72(6). <https://doi.org/10.1103/physreve.72.065102>
- Gracia-Lázaro, C., Brigatti, E., Hernández, A. R., y Moreno, Y. (2021). Polarization inhibits the phase transition of Axelrod's model. *Physical Review E*, 103(6). <https://doi.org/10.1103/physreve.103.062306>

- Gracia-Lázaro, C., Floría, L. M., y Moreno, Y. (2011). Selective advantage of tolerant cultural traits in the Axelrod-Schelling model. *Physical Review E*, 83(5). <https://doi.org/10.1103/physreve.83.056103>
- Hernández, A. R., Gracia-Lázaro, C., Brigatti, E., y Moreno, Y. (2018). Robustness of cultural communities in an open-ended Axelrod's model. *Physica A: Statistical Mechanics and Its Applications*, 509, 492-500. <https://doi.org/10.1016/j.physa.2018.06.023>
- Jacobmeier, D. (2005). Multidimensional Consensus Model on a Barabási-Albert Network. *International Journal of Modern Physics C*, 16(04), 633-646. <https://doi.org/10.1142/s0129183105007388>
- Klemm, K., Eguíluz, V. M., Toral, R., y Miguel, M. S. (2003). Global culture: A noise-induced transition in finite systems. *Physical Review E*, 67(4). <https://doi.org/10.1103/physreve.67.045101>
- (2003). Nonequilibrium transitions in complex networks: A model of social interaction. *Physical Review E*, 67(2). <https://doi.org/10.1103/physreve.67.026120>
- Lopez Morales, S. J., González Silva, R. A., y González Silva, M. (2025). Emergence of multistability in the cultural dissemination of the Axelrod model with mass media. *Chaos Theory and Applications*, 7(1), 78-86. <https://doi.org/10.51537/chaos.1603833>
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431-441. <https://doi.org/10.1137/0111030>
- Raducha, T., y Gubiec, T. (2017). Coevolving complex networks in the model of social interactions. *Physica A: Statistical Mechanics and Its Applications*, 471, 427-435. <https://doi.org/10.1016/j.physa.2016.12.079>
- Railsback, S. F., y Grimm, V. (2019). *Agent-Based and individual-based modeling: A practical introduction, second edition*. Princeton University Press.
- Rodríguez, A. H., del Castillo-Mussot, M., y Vázquez, G. J. (2009). Induced monoculture in Axelrod model with clever mass media. *International Journal of Modern Physics C*, 20(08), 1233-1245. <https://doi.org/10.1142/s012918310901431x>
- Rodríguez, A. H., y Moreno, Y. (2010). Effects of mass media action on the Axelrod model with social influence. *Physical Review E*, 82(1). <https://doi.org/10.1103/physreve.82.016111>
- Shibanai, Y., Yasuno, S., e Ishiguro, I. (2001). Effects of global information feedback on diversity. *Journal of Conflict Resolution*, 45(1), 80-96. <https://doi.org/10.1177/0022002701045001004>
- Stivala, A., Kashima, Y., y Kirley, M. (2016). Culture and cooperation in a spatial public goods game. *Physical Review E*, 94(3). <https://doi.org/10.1103/physreve.94.032303>

- Stivala, A., Robins, G., Kashima, Y., y Kirley, M. (2014). Ultrametric distribution of culture vectors in an extended Axelrod model of cultural dissemination. *Scientific Reports*, 4(1). <https://doi.org/10.1038/srep04870>
- Tilles, P. F. C., y Fontanari, J. F. (2015). Diffusion of innovations in Axelrod's model. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(11), P11026. <https://doi.org/10.1088/1742-5468/2015/11/p11026>

Capítulo 4

Análisis de tópicos en comentarios de YouTube durante la campaña presidencial de México 2024: un estudio con LDA

Antonio Aguilera Ontiveros¹

Adela Martínez Galván²

Introducción

En el contexto de la ciencia política contemporánea, el análisis de datos provenientes de las redes sociales se ha convertido en una herramienta esencial para comprender las dinámicas de opinión pública y las estrategias de comunicación política que realizan tanto gobiernos como partidos políticos. Si bien, se busca que sean principalmente herramientas para predecir resultados electorales, como muestra el trabajo de Skoric *et al.* (2020), donde se realizó un metaanálisis de 74 estudios publicados donde se examinó la capacidad predictiva de diferentes enfoques basados en redes sociales, como el análisis de sentimientos y las características estructurales de las redes, se llegó a la conclusión de que estos enfoques no son muy fiables para

1. Programa de Estudios Políticos e Internacionales, El Colegio de San Luis, A. C. Correo electrónico: antonio.aguilera@colsan.edu.mx. ORCID: <https://orcid.org/0000-0003-1548-9956>
2. Becaria de proyecto, Programa de Estudios Políticos e Internacionales, El Colegio de San Luis, A. C. Correo electrónico: adela.martinez@colsan.edu.mx. ORCID: <https://orcid.org/0009-0003-4999-7593>

esta tarea predictiva. No obstante, este tipo de estudios ofrecen una fuente valiosa de información para entender la opinión pública y el comportamiento político en la era digital. El estudio posevento de los comentarios vertidos en una campaña electoral puede dar luz sobre diferentes asuntos, por ejemplo: a) la evaluación de la eficacia de los mensajes políticos; b) la detección de tópicos emergentes entre la audiencia; c) la identificación de líderes de opinión; d) la comprensión de la polarización y los sentimientos del electorado; e) el análisis del impacto de la desinformación (Permadi y Putri, 2022).

Por otra parte, la campaña presidencial de México en 2024 generó un volumen significativo de comentarios en plataformas digitales como X, Facebook y YouTube, ofreciendo una oportunidad única para explorar las percepciones y reacciones de la ciudadanía ante los diferentes discursos vertidos tanto por los candidatos a la presidencia como por los líderes de opinión a favor de una u otra candidata o candidato. El estudio de estos comentarios se hace relevante en el sentido de que nos permiten medir la aceptación a los proyectos de nación que cada candidata y candidato esgrimía. Asimismo, nos permite observar el nivel del debate ya no de los candidatos, sino de la ciudadanía, pudiendo observar a aquellos ciudadanos partidarios de uno u otro partido y los argumentos que vertían para posicionar su opinión. Así, los debates presidenciales fueron un escaparate donde se pudieron observar las emociones, los prejuicios y las simpatías y fobias hacia cada una de las candidatas y el único candidato.

Este capítulo presenta un análisis de tópicos basado en el algoritmo *Latent Dirichlet Allocation* (LDA, por sus siglas en inglés) (Blei *et al.*, 2003) aplicado a más de 9,500 comentarios extraídos de YouTube con referencia a los tres debates previos a las elecciones de 2024. Además, se realizó un análisis de sentimientos a los comentarios. El objetivo general es analizar la percepción pública y la dinámica de la opinión en torno a los candidatos presidenciales de México 2024 a través del estudio de los comentarios en YouTube de dos fuentes periodísticas: CNN en Español e Imagen Noticias, con el fin de comprender los patrones de la distribución temática de las discusiones electorales por parte de la audiencia. Los objetivos específicos son los siguientes:

- Identificar los principales tópicos presentes en los comentarios de YouTube sobre los candidatos presidenciales de México 2024 durante cada debate presidencial.

- Evaluar los comentarios hacia cada candidato.
- Analizar la existencia de sesgos positivos y sesgos negativos hacia los tres candidatos.

Metodología

Los datos fueron recolectados usando técnicas de *web scraping* codificadas en Python. Los comentarios que se vertieron durante los debates presidenciales, que fueron tres, se recolectaron de las siguientes fuentes noticiosas: CNN en Español e Imagen Noticias. No se usó la fuente del INE (Instituto Nacional Electoral) porque esta ya no contiene los comentarios y solo aparece el debate en sí. Se recolectó la información en febrero de 2025. Se analizaron 708 comentarios para el primer debate, 2,275 comentarios para el segundo debate y 6,621 para el tercer debate, un total de 9,604 comentarios. Estos comentarios constituyeron el *corpus* del análisis de texto.

Para el *web scraping* se usó primero *yt-dlp*, un código en Python derivado del proyecto *youtube-dl*. Está disponible en código abierto en el siguiente GitHub: <https://github.com/yt-dlp/yt-dlp>. Para instalarlo solo se debe tener Python en nuestro sistema operativo y luego tipear `pip install yt-dlp`. Una vez que se instala se procede a bajar los comentarios usando el siguiente código: `yt-dlp --skip-download --write-comments --output "comentarios.json" https://www.youtube.com/watch?v=gGXTaJNuZcw`. Esto crea un *dataframe* llamado "comentarios.json" en formato `json`³ del video en YouTube identificado con la cadena `gGXTaJNuZcw` que corresponde al video del primer debate presidencial transmitido por CNN en Español. El comando `--skip-download` hace que no se baje el video y el comando `--write-comments` baja los comentarios.

El preprocesamiento de datos incluyó la eliminación de ruido, como pueden ser los emoticones, cadenas de caracteres sin sentido,

3. Un archivo *JSON* (*JavaScript Object Notation*) es un formato de texto ligero y estructurado para el intercambio de datos. Se basa en la sintaxis de objetos de JavaScript, pero es independiente del lenguaje de programación, lo que lo hace ampliamente utilizado para almacenar y transmitir información en aplicaciones web, bases de datos y API.

etc. También se realizó la tokenización, que consiste en el proceso de dividir un texto en unidades más pequeñas llamadas *tokens*. Un *token* puede ser una palabra, un carácter o una oración, dependiendo del nivel de granularidad requerido para el análisis. Después se procedió a hacer la lematización, que es el proceso de reducir una palabra a su forma base o lema, manteniendo su significado lingüístico. A diferencia de la *stemming*, que solo recorta palabras según reglas heurísticas, la lematización utiliza diccionarios morfológicos y reglas lingüísticas para obtener una forma gramatical válida. Para la identificación de los tópicos se aplicó el algoritmo LDA (Blei *et al.* 2003). El análisis de temas o tópicos se caracteriza por lo siguiente:

- *Extracción de tópicos latentes.* Para ello se identifican grupos de palabras que tienden a aparecer juntas en distintos documentos, lo que permite inferir de qué trata el texto sin necesidad de clasificación manual previa.
- *Reducción de la dimensionalidad.* Esto implica que se representa un gran volumen de datos textuales en un número reducido de temas, facilitando con ello su interpretación y análisis.
- *Uso de métodos probabilísticos o basados en matrices.* Los métodos más utilizados, como el NMF (Non-Negative Matrix Factorization) y el LDA (Latent Dirichlet Allocation), usan métodos basados en la factorización de matrices y modelos probabilísticos para realizar la tarea de encontrar los tópicos de un texto.

En este texto usaremos el método basado en modelos probabilísticos llamado LDA. A continuación se explica en qué consiste este método.

El LDA (Latent Dirichlet Allocation) fue introducido por primera vez por Blei *et al.* (2003) y se ha convertido en una de las técnicas de modelado de temas más usadas. Este modelo se ha utilizado en muchas aplicaciones de aprendizaje automático y recuperación de información. Por ejemplo, Wei y Croft (2006) aplicaron este modelo a la recuperación de información basada en modelos lingüísticos y lo compararon con la indexación semántica latente probabilística y la recuperación basada en conglomerados. Griffiths y Steyvers (2004) utilizaron LDA para capturar de temas científicos en una colección de documentos. Igualmente, este modelo fue utilizado por Rosen-Zvi *et al.* (2010) para extraer modelos autor-tema. En su modelo, además de las estadísticas de palabras dentro de los documentos, se utilizó la

información sobre los autores de cada documento y cada autor se representó mediante una distribución de probabilidad sobre temas.

El LDA es un modelo generativo, probabilístico y jerárquico bayesiano que induce temas a partir de una colección de documentos en tres pasos:

- Cada documento de la colección se distribuye en tópicos que se muestrean para ese documento basándose en una distribución de Dirichlet.
- Cada palabra del documento se asocia con un único tema basado en esta distribución Dirichlet.
- Cada tema se representa como una distribución multinomial sobre las palabras asignadas al tema muestreado.

La distribución de temas sobre documentos (θ) y de palabras sobre temas (z) son las variables latentes del modelo (Siklos *et al.* 2018). Entonces, tenemos que:

$$\theta \sim \text{Dirichlet}(\alpha)$$

Una distribución Dirichlet es una distribución sobre distribuciones; en concreto, es una prior conjugada para la distribución multinomial. En estadística bayesiana, una prior conjugada es una elección especial de la distribución *a priori* que tiene una relación matemática conveniente con la distribución de probabilidad verosimilitud. Esta relación permite que la distribución *a posteriori* resultante sea de la misma familia que la distribución *a priori*, simplificando así los cálculos.

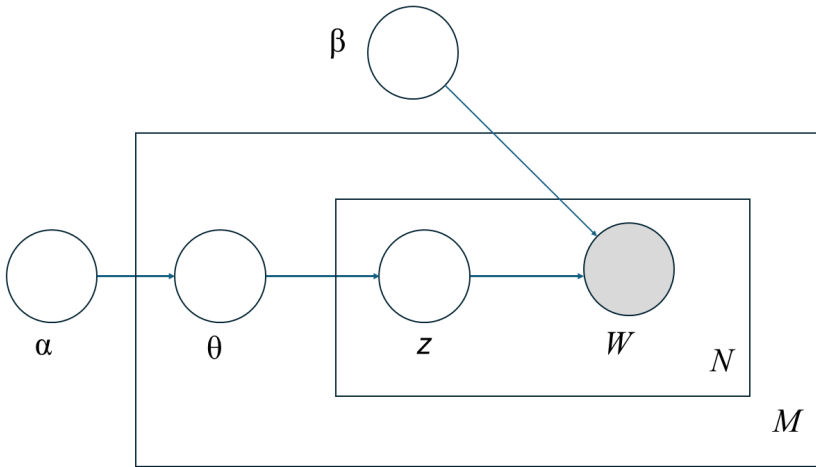
Para cada una de las N palabras en un documento d , w_{dn} , la asignación de una palabra a un tópico (z_{dn}) es escogida de la distribución multinomial de los tópicos en dicho documento (θ_d). O sea:

$$z_n \sim \text{Multinomial}(\theta)$$

El LDA sigue un proceso de tres capas (figura 2.1). La primera capa son parámetros que se fijan a nivel de *corpus*, lo que significa que se supone que se muestrean una vez en el proceso de generación de un *corpus*. El primero de estos hiperparámetros es α , que define la distribución Dirichlet a partir de la cual se muestrea la distribución de temas sobre documentos (θ). El segundo hiperparámetro es δ , que define otra

distribución de Dirichlet a partir de la cual se muestrea la distribución de palabras sobre temas (β). Más concretamente, β es la distribución de palabras sobre temas, una matriz $k \times V$ donde los elementos representan la probabilidad de que la palabra ($w_i \in \{1, \dots, V\}$) pertenezca a un determinado tema z_n .

Figura 1. Proceso del LDA



Fuente: Tomado de Blei *et al.* (2003).

La segunda capa son parámetros exclusivos de cada documento: θ_d es la distribución de temas sobre los documentos, muestreada una vez por documento d_i , donde ($d_i \in \{1, \dots, M\}$). La tercera capa son parámetros únicos para cada palabra: z_{dn} y w_{dn} se muestrean una vez por cada palabra en cada documento w_{dn} , donde ($w_d \in \{1, \dots, N_d\}$).

La distribución posterior de la variable latente $p(\theta, z | w, \alpha, \beta)$ no puede calcularse de forma manejable. Existen algunos métodos para derivar esta distribución a partir de los datos observados, w_{dn} (Blei *et al.*, 2003; Griffiths y Steyvers, 2004; Hoffman *et al.*, 2013; Grün y Hornik, 2011). Para este trabajo se usó la implementación estándar empleada en la biblioteca Python Gensim, la cual utiliza una aproximación basada en el método de Muestreo de Gibbs Colapsado. El muestreo de Gibbs es un algoritmo MCMC (Monte Carlo Markov Chain) que permite muestrear desde una distribución de probabili-

dad multivariante, siendo más sencillo muestrear desde la distribución condicional que desde la distribución conjunta (Gelfand, 2000). Esta secuencia puede emplearse para aproximar la distribución conjunta, como al generar un histograma de dicha distribución; para aproximar la distribución marginal de una de las variables o de algún subconjunto de ellas, como parámetros desconocidos o variables latentes; o para calcular una integral, como el valor esperado de una variable. Generalmente, algunas variables corresponden a observaciones con valores conocidos, por lo que no es necesario muestrearlas.

Todos los procesos anteriores fueron implementados en Python, los códigos para hacer esto se pueden bajar desde nuestro GitHub:

El análisis de sentimientos, también conocido como análisis de opiniones o minería de opiniones, es una técnica en el campo del procesamiento de lenguaje natural (NLP, por sus siglas en inglés) que se utiliza para determinar y evaluar la actitud, emoción o sentimiento expresado en un texto o un conjunto de datos de texto. El análisis de sentimientos es un proceso automatizado que implica el uso de algoritmos y técnicas de procesamiento de lenguaje natural para evaluar la polaridad emocional o actitud expresada en un texto. Esta polaridad puede ser positiva, negativa o neutral. El objetivo es comprender cómo se siente o qué piensa una persona o un grupo de personas sobre un tema, producto, servicio, evento u entidad específica a partir de la información textual disponible.

El modelo que utilizamos se basa en la arquitectura de BERT (*Bidirectional Encoder Representations from Transformers*). El modelo en particular:

Está entrenado para análisis de sentimientos en múltiples idiomas, incluyendo español.

Clasifica el texto en una escala ordinal (una a cinco estrellas) para representar niveles de sentimiento desde muy negativo hasta muy positivo. Así, un texto con solo una estrella lo podemos interpretar como muy negativo, mientras que un texto con tres estrellas sería un texto neutro.

Resultados, análisis LDA, primer debate presidencial

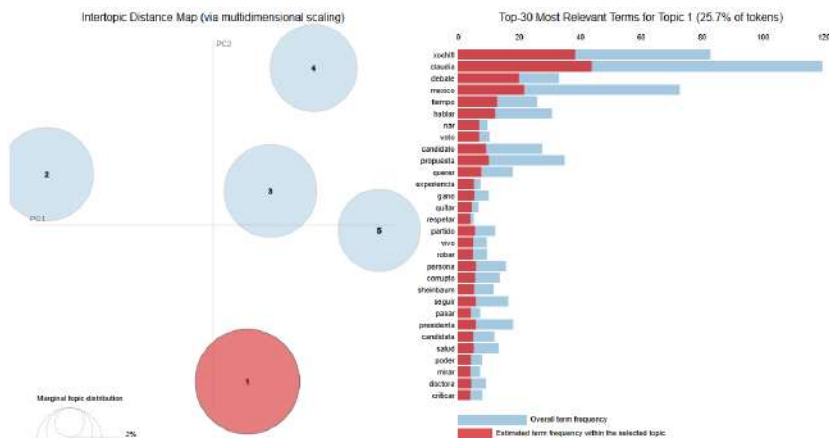
El primer debate presidencial de México 2024 se llevó a cabo el 7 de abril a las 20:00 horas, marcando un hito en el proceso electoral. Este evento fue transmitido a través de diversas plataformas digitales, destacando la cobertura de CNN en Español e Imagen Noticias, cuyos espacios sirvieron como fuentes para recopilar los comentarios y reacciones del público. En el debate participaron los tres candidatos principales a la presidencia del país:

- Claudia Sheinbaum, candidata por la coalición integrada por el Movimiento de Regeneración Nacional (Morena), el Partido del Trabajo (PT) y el Partido Verde Ecologista de México (PVEM).
- Xóchitl Gálvez, candidata por la coalición conformada por el partido Revolucionario Institucional (PRI), el Partido de la Revolución Democrática (PRD) y el Partido Acción Nacional (PAN).
- Jorge Álvarez Máynez, candidato por el partido Movimiento Ciudadano.

Durante el debate, los candidatos presentaron sus propuestas y confrontaron sus posturas en torno a temas clave de la agenda nacional, generando un espacio de discusión que captó la atención de millones de ciudadanos.

Como parte del análisis de la conversación digital generada en torno a este evento, se recopilaron y analizaron los comentarios publicados en YouTube mediante técnicas de *web scraping*, utilizando la herramienta yt-dlp. A partir de estos datos, se aplicó el modelo LDA para identificar los principales temas abordados por los usuarios en sus interacciones.

Figura 2. Tópico 1, primer debate presidencial



Fuente: Elaboración propia.

En la figura 2 se visualiza con detalle el Tópico 1, el cual representa el 25.7 % total de *tokens* analizados, este tópico fue identificado como uno de los ejes más relevantes en la discusión pública derivada del debate, reflejando las preocupaciones e intereses predominantes entre la audiencia. A continuación se presentan los términos más relevantes y su posible significado dentro del contexto del debate:

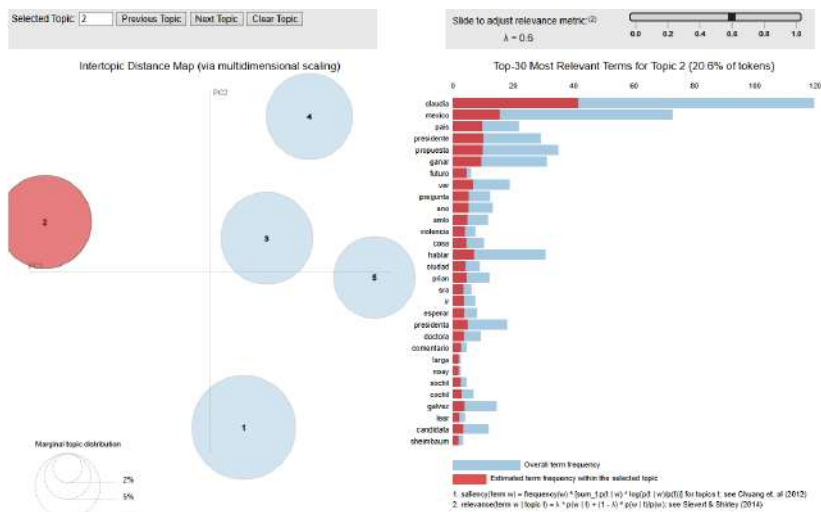
- “Xóchitl”, “Claudia”, “Máynez”: La mención concurrente de los tres candidatos sugiere que este tópico está centrado en la evaluación de su desempeño en el debate, se destaca la presencia de Xóchitl Gálvez y Claudia Sheinbaum, lo que podría indicar que la confrontación entre ambas fue uno de los ejes centrales de la discusión. Este enfoque refleja el interés de los usuarios en analizar y comparar las posturas, argumentos y estrategias de las dos candidatas, quienes, al parecer, generaron mayor interacción y comentarios en la plataforma digital.
- “Votar”, “Presidencia”, “México”: La presencia de términos asociados con el proceso electoral y la representación política indica que este tópico agrupa discusiones en torno a la importancia del voto y la elección presidencial como un evento de relevancia nacional.

- “Honestidad”, “Corrupción”, “Mentira”: Estos términos sugieren que una parte de la discusión estuvo orientada a la credibilidad de los candidatos y su historial político. Su uso en los comentarios podría reflejar tanto las acusaciones como defensas en torno a la integridad de las participaciones en el debate.
- “Propuesta”, “Gobierno”, “Pueblo”: La inclusión de estos términos indica que parte del debate en los comentarios se enfocó en la evaluación de las plataformas políticas y programas de gobierno expuestos por los candidatos, así como en su impacto potencial en la ciudadanía.
- “Manipulación”, “Medios”: La aparición de estos términos sugiere que dentro del Tópico 1 también se discutieron percepciones sobre el rol de los medios de comunicación en la cobertura del debate y la posible existencia de sesgos mediáticos.

El Tópico 1 se encuentra directamente asociado a la evaluación del desempeño de los candidatos presidenciales durante el debate, como lo demuestra la recurrencia de sus nombres propios y de términos vinculados a valores políticos y percepciones ciudadanas. La identificación de palabras con una marcada carga subjetiva, tales como “honesto”, “mentira” y “corrupción”, engloba opiniones polarizadas, en las que los comentarios reflejan tanto adhesión como rechazo hacia las figuras políticas, centrándose en una valoración de carácter moral y ético.

Estudios previos han señalado que los debates electorales suelen generar altos niveles de polarización en redes sociales, lo que se ve amplificado por la naturaleza algorítmica de estas plataformas, que favorecen la difusión de contenido altamente emotivo (Allcott y Gentzkow, 2017; Vosoughi *et al.*, 2018). Este fenómeno es relevante para comprender cómo las percepciones de los candidatos pueden ser influenciadas no solo por sus propias intervenciones en el debate, sino también por la narrativa construida en los espacios digitales.

Figura 3. Tópico 2, primer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

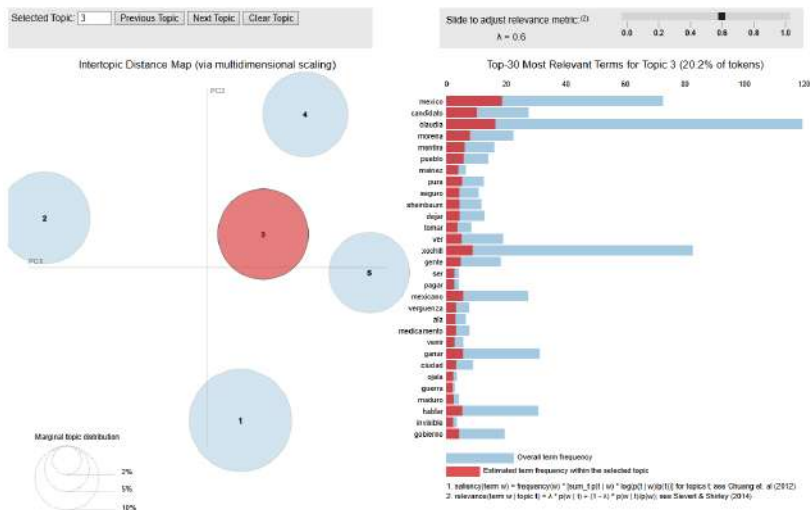
A continuación, el análisis del Tópico 2 (figura 3). Este tópico representado en color rojo abarca una porción significativa del discurso, correspondiente al 20.6 % del total de las palabras analizadas, lo que evidencia su relevancia dentro de la estructura temática examinada. No obstante, su proximidad con otros tópicos, como el Tópico 1 (azul), podría indicar la existencia de cierta relación conceptual o temática entre ambos, aunque el Tópico 2 mantiene características únicas que lo definen. A partir del análisis realizado se identifica que este tópico se enfoca principalmente en términos como:

- “Claudia”: La candidata Sheinbaum es el término más frecuente en este tópico, lo que sugiere que gran parte de la conversación estuvo enfocada en su desempeño durante el debate.
- “México”, “País”: La inclusión de estos términos indica que la discusión trascendió el debate en sí mismo y abordó cuestiones sobre el futuro del país y el impacto de las elecciones en la sociedad.
- “Presidente”, “Presidenta”: La presencia de estos términos refuerza la centralidad del debate electoral en la discusión, con énfasis en la figura presidencial y el liderazgo de las candidatas.

- “Propuesta”, “Ganar”, “Futuro”: Estos términos sugieren que una parte de la conversación giró en torno a las propuestas presentadas por los candidatos y su viabilidad, así como a las especulaciones sobre los posibles resultados electorales.
- “AMLO”: La referencia al presidente Andrés Manuel López Obrador indica que su figura estuvo presente en el debate y en la percepción de los usuarios sobre la continuidad o el cambio en el próximo Gobierno.
- “Violencia”: La mención de este término sugiere que la seguridad pública fue un tema central en las discusiones dentro de este tópico.
- “PRIAN”: La inclusión de este término, utilizado para referencias a la alianza entre el PRI y el PAN, sugiere que una parte de la conversación abordó críticas o comparaciones entre los partidos políticos y sus estrategias electorales.

Este conjunto de términos sugiere que el Tópico 2 está fuertemente vinculado con la figura de Claudia Sheinbaum y su rol en la contienda presidencial. Más allá de centrarse exclusivamente en su participación en el debate, este tópico también abarca cuestiones relacionadas con la dirección política del país, el legado del actual Gobierno y la percepción de continuidad o cambio en el próximo sexenio. La convergencia de términos como “presidenta”, “México”, “propuesta” y “violencia” indica que la discusión en torno a la candidata no solo giró en torno a su desempeño en el debate, sino que también incluyó evaluaciones sobre propuestas, el estado del país y las precipitaciones centrales del electorado.

Figura 4. Tópico 3, primer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

La figura 4 refleja la visualización del Tópico 3, resaltado en rojo, representando el 20.2 % de los *tokens*. La presencia de términos como “México”, “candidato”, “mentira”, “guerra” y “gobierno” sugiere de los discursos políticos, la gobernabilidad y el rol del Estado en la resolución de problemáticas nacionales. Los términos predominantes del Tópico 3 incluyen:

- “México”, “Gobierno”, “Ciudad”: Los comentarios dentro de este tópico se centran en cuestiones nacionales y en la evaluación del rol del Estado en la gestión de problemas de interés público.
- “Candidato”, “Claudia”, “Xóchilt”, “Máñez”: La mención recurrente de los tres principales contendientes a la presidencia sugiere una comparación entre sus posturas y su credibilidad como figuras políticas.
- “Mentira”, “Pura”, “Seguro”: Estos términos apuntan a una discusión sobre la veracidad de las declaraciones de los candidatos. Es posible que los comentarios hayan cuestionado la autenticidad de ciertos discursos o propuestas presentadas durante el debate.
- “Pueblo”, “Gente”, “mexicano”: Asimismo, también incluye referencias a la relación entre el Gobierno y la ciudadanía, posible-

mente en torno a temas de representación política y demandas sociales.

- “Guerra”, “Maduro”, “Vergüenza”: Los anteriores términos están relacionados con comparaciones entre la situación política de México y la de otros países, o con preocupaciones sobre conflictos internos y la estabilidad democrática.
- “Medicamento”, “Pagar”, “Invisible”: Dichos conceptos abordan problemáticas fundamentales en la salud pública, centrándose en la disponibilidad de medicamentos y los mecanismos de financiamiento del país.

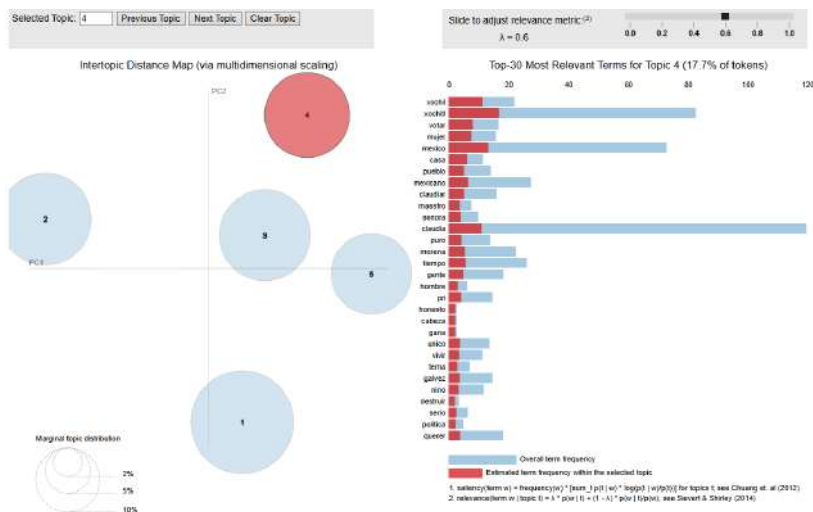
El Tópico 3 se estructura en tres dimensiones clave: credibilidad y percepción de los candidatos, gobernabilidad y confianza en el Estado, y crisis y comparaciones internacionales.

En la dimensión de credibilidad y percepción de los candidatos, términos como *mentira*, *pura* y *seguro*, junto con los nombres de los principales candidatos, indican que el foco de la discusión fue la prominencia de la realidad de los discursos políticos. Los comentarios de los usuarios analizaron promesas de campaña y declaraciones emitidas durante el debate, cuestionando la autenticidad y la factibilidad, evidenciando una polarización con opiniones divididas, entre quienes defendían la integridad de ciertos candidatos y quienes los acusaban de manipular la información o engañar al electorado.

En cambio, en la dimensión de gobernabilidad y confianza en el Estado, términos como *“México”*, *“gobierno”* y *“ciudad”* sugieren que la conversación se centró en la evaluación de la administración pública y el rol del Estado en la resolución de problemáticas nacionales. En los comentarios se discutió la capacidad de los candidatos para gestionar áreas clave como seguridad, economía y servicios públicos, comparando antecedentes y propuestas.

Por otra parte, en la dimensión de crisis y comparaciones internacionales persistieron términos como *“guerra”*, *“Maduro”* y *“vergüenza”*, estableciendo comparaciones entre la situación política de México y la de otros países, particularmente Venezuela, reflejando preocupaciones sobre posibles crisis económicas o autoritarismo; asimismo, la gestión gubernamental.

Figura 5. Tópico 4, primer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

Del mismo modo, en la figura 5 se representa el 17.7 % del total de *tokens* analizados, su composición sugiere fuerte presencia de términos relacionados con las principales candidatas, el acto de votar y conceptos ligados a la representación política y ciudadana.

El Tópico 4 concentra términos que evidencian el enfoque de la conversación, entre los términos más representativos destacan:

- “Votar”, “Mujer”: Sugieren que la participación ciudadana y el liderazgo femenino fueron temas destacados, reflejando un interés particular en la elección de mujeres para cargos políticos y su impacto en la representación de género.
- “Pueblo”, “Mexicano”: La recurrencia de estos términos indica una dimensión nacionalista en la discusión, vinculada a temas de identidad nacional, soberanía y la representación de los intereses del país en el ámbito político.
- “PRI”, “Morena”: Al hacer referencia a los partidos políticos, establece una comparación entre los proyectos políticos que representan los candidatos, evidenciando un contraste entre sus ideologías y propuestas de gobierno.

- “Honesto”, “Serio”, “Gana”: El uso de dichas palabras refleja la percepción pública sobre la ética, transparencia y capacidad de liderazgo de los aspirantes, cualidades personales que inciden en su viabilidad como candidatos.

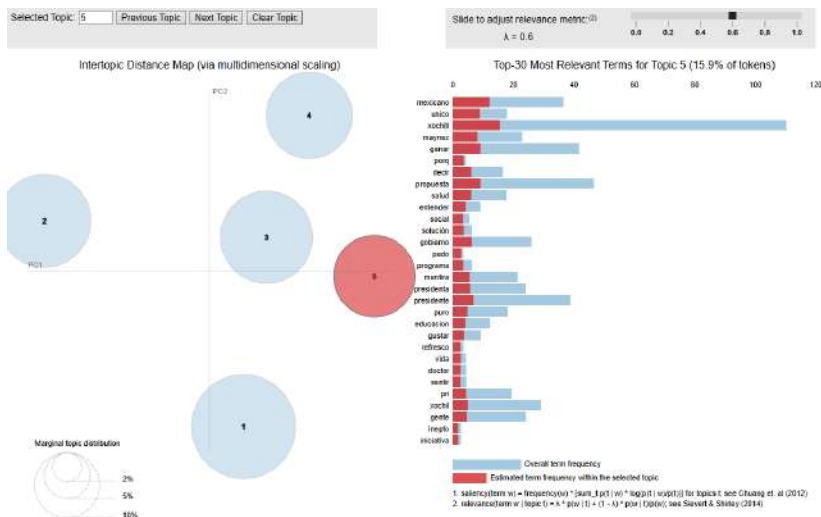
Uno de los aspectos más destacados ha sido la representación femenina en la política, evidenciada por la presencia de Xóchitl Gálvez y Claudia Sheinbaum. Este tema abarcó tanto evaluaciones positivas del progreso en la igualdad de género, como cuestionamientos sobre si su postulación representa un cambio fundamental o simplemente una continuación de la política tradicional.

El análisis revela dos ejes fundamentales: la participación ciudadana y la evaluación de los actores políticos. La recurrencia de términos como “votar”, “pueblo” y “política” evidencia la relevancia de la movilización democrática, aunque también refleja inquietudes sobre la transparencia del proceso.

Asimismo, la mención de los partidos políticos PRI y Morena, junto con la valoración de atributos como honestidad, seriedad y capacidad, demuestra el interés por comparar proyectos políticos y evaluar la idoneidad de los candidatos. Este escrutinio generó una marcada polarización, reflejando la intensidad de la confrontación ideológica.

La polarización del discurso se manifestó en tres narrativas principales: apoyo a Sheinbaum, resaltando su preparación y continuidad con el proyecto de la 4T; apoyo a Xóchitl Gálvez, destacando su perfil de cambio; y críticas a ambas candidatas, cuestionando su viabilidad. Este patrón es consistente con estudios sobre la influencia de las redes sociales en la polarización política (Tucker *et al.*, 2018).

Figura 6. Tópico 5, primer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

El Tópico 5 (figura 6) comprende el 15.9 % del total de los *tokens*. Los hallazgos indican que la conversación en torno a este tópico se enfocó en el examen de propuestas de gobierno, programas sociales y la valoración de la factibilidad de las soluciones propuestas por los candidatos.

Los términos con mayor representatividad son los siguientes:

- “Xóchitl”, “Máynez”: Indican que un segmento significativo de los comentarios de los usuarios se enfocó en la evaluación comparativa de las propuestas formuladas por estos candidatos.
- “Propuesta”, “Programa”, “Solución”: Subraya la importancia otorgada a la discusión de las estrategias presentadas durante el debate y a la valoración de su viabilidad.
- “Salud”, “Educación”, “Gobierno”: Señala que los comentarios abordaron sectores fundamentales de la agenda pública, reflejando las preocupaciones prioritarias de la ciudadanía.
- “Mentira”, “Inepto”, “PRI”: Representa una manifestación de escepticismo o crítica hacia ciertos planteamientos o figuras políticas presentes en el debate.

- “Presidente”, “Presidenta”: Sugiere que los comentarios trascendieron de la mera discusión de propuestas, extendiéndose a la evaluación de la idoneidad de los candidatos para ejercer el liderazgo del país.

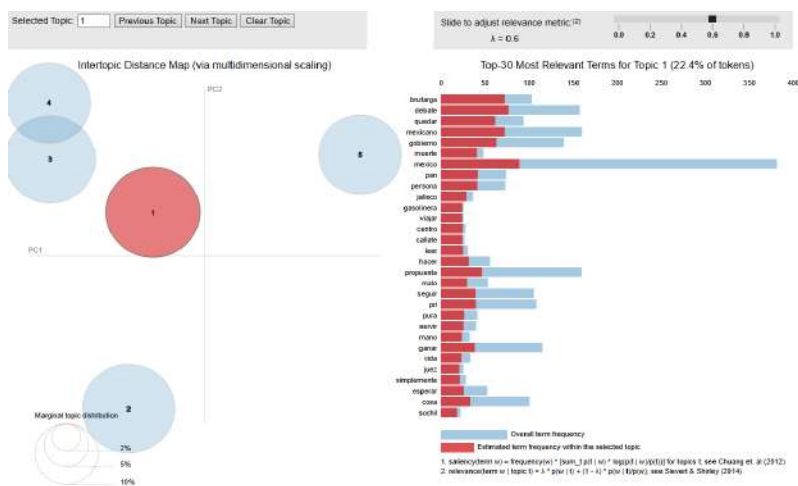
En el contexto de este análisis, destaca un enfoque predominante en la evaluación de las propuestas y programas de gobierno presentados por los candidatos. Se observa una clara división entre los usuarios que percibieron las propuestas como soluciones concretas y quienes las consideran discursos carentes de sustento o realismo.

Por otro lado, también se abordaron temas estructurales, reflejando reflexiones sobre la necesidad de mejorar sectores clave, así como comparaciones entre las estrategias propuestas por los candidatos y su impacto potencial en la sociedad. La discusión en torno a políticas públicas constituye un pilar esencial de debate democrático, permitiendo a los ciudadanos evaluar las plataformas de los candidatos y su capacidad para abordar los desafíos del país (Dunn, 2017).

No obstante, una parte del debate estuvo marcada por el escepticismo político, reflejado en el uso de las palabras como “mentira”, “inepto” y “PRI”; estas expresiones interpretan críticas hacia los candidatos o hacia la clase de política general, evidenciando una visión respecto a las promesas de campaña.

Resultados, análisis LDA, segundo debate presidencial

Figura 7. Tópico 1, segundo debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

El panorama político mexicano actual se caracteriza por una marcada polarización y un intenso debate sobre el futuro del país, especialmente en áreas críticas como la inseguridad, la economía y la gobernabilidad (Casar, 2021). La contienda electoral de 2024 se ha visto significativamente influenciada por el desempeño del gobierno de Andrés Manuel López Obrador, el legado de los partidos tradicionales y la búsqueda de alternativas por parte de la oposición. En este contexto, los debates presidenciales adquieren una relevancia crucial, ofreciendo a los ciudadanos la oportunidad de evaluar el perfil, las propuestas y la capacidad de respuesta de los candidatos ante los desafíos más apremiantes de la nación (Lawson, 2019).

El segundo debate presidencial de México 2024, llevado a cabo el 28 de abril de 2024, representó un momento clave en la contienda electoral rumbo a las elecciones del 2 de junio. De igual forma, como en el anterior debate, tuvo la participación de los tres candidatos a la presidencia.

Bajo el lema “México, prosperidad y bienestar”, el debate abordó temas de seguridad, economía, desarrollo social y política exterior, ejes estratégicos que han sido objeto de discusión en la campaña. Uno de los aspectos más destacados fue la intensidad de los intercambios directos entre los candidatos, particularmente Sheinbaum y Gálvez, realizando acusaciones sobre la gestión del país, el manejo de la seguridad pública y la viabilidad de los proyectos económicos.

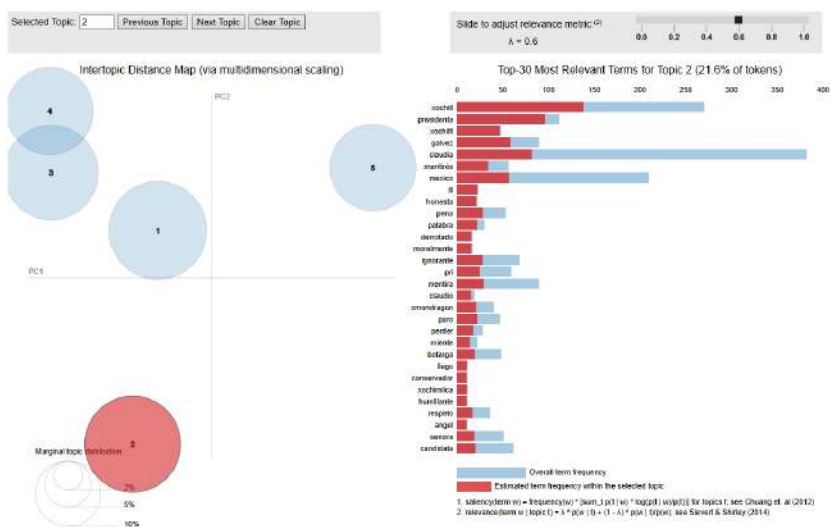
El análisis de la figura 7 revela que el Tópico 1, correspondiente al segundo debate presidencial, representa el 22.4 % del total de *tokens* analizados. Este tópico se caracteriza por la alta frecuencia de términos como “muerte”, “gobierno”, “México” y “debate”, entre otros. Los términos específicos que conforman este tópico incluyen los siguientes:

- “Muerte”, “México”, “Gobierno”: De esto se deduce que los comentarios del debate se enfocaron en la crisis de inseguridad que vive el país y en cómo el Gobierno ha anejado esta problemática.
- “PAN”, “PRI”: Cuando aparecen PAN y PRI es por que existieron comparaciones entre los gobiernos anteriores (de esos partidos) y el actual.
- “Jalisco”, “Gasolinera”, “Viajar”: En el periodo de campañas electorales, es de conocimiento público que Jalisco enfrenta desafíos significativos en materia de seguridad pública. Por tanto, la aparición de este término podría estar vinculada a incidentes relevantes ocurridos en la entidad, generando un debate sobre la efectividad de las políticas de seguridad implementadas.
- “Malo”, “Mentira”, “Esperar”: Estas palabras reflejan un tono crítico hacia el discurso y el desempeño de los candidatos, denotan escepticismo o descontento hacia las afirmaciones de los aspirantes.

El Tópico 1 se centra en el análisis de la crisis de seguridad y gobernabilidad que aqueja al país, evidenciando una marcada preocupación ciudadana. Los comentarios analizados abordan tanto las acciones implementadas por la administración actual como las soluciones propuestas por los aspirantes a la presidencia. Asimismo, se incluye una comparación entre administraciones anteriores, lo que refleja un intento de contextualizar la problemática y evaluar la efectividad de las estrategias implementadas,

Los tópicos precedentes revelan la crisis política y de inseguridad que persiste en México. En los comentarios se observa que los ciudadanos de Jalisco experimentan una crisis de delincuencia, lo cual incita a realizar una comparación entre el Gobierno actual y los candidatos postulados. Los ciudadanos analizan las estrategias propuestas con el objetivo de discernir si estas ofrecen soluciones viables o si la situación continuará sin cambios significativos.

Figura 8. Tópico 2, segundo debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

A diferencia del tópico anterior, el Tópico 2 (figura 8) del segundo debate presidencial se caracterizó por un enfoque en la exposición y defensa de propuestas en áreas cruciales como la economía, la seguridad y el bienestar social. Asimismo, se observó un énfasis en el cuestionamiento de la honestidad de los candidatos.

Este cambio de enfoque refleja la dinámica propia de los debates presidenciales, donde los candidatos buscan contrastar sus plataformas y persuadir al electorado. Según el Instituto Nacional Electoral (INE), el segundo debate presidencial de México 2024 se estructuró para

fomentar un contraste de ideas y la participación ciudadana directa, incorporando preguntas videograbadas por ciudadanos de todo el país.

El tópico se enfoca en la percepción de Xóchitl Gálvez dentro del debate y su rol en la contienda electoral, algunos de los términos más representativos incluyen:

- “Xóchitl”, “Gálvez”, “presidenta”, “Candidata”: Indican que una parte significativa de la discusión giró en torno a su desempeño en el debate y la posibilidad de que se convierta en la primera presidenta de México.
- “Claudia”, “PRI”, “Mentira”: Sugiere una comparación entre ambas candidatas, así como críticas y acusaciones relacionadas con su trayectoria política.
- “Honesta”, “Ignorante”, “Moralmente”, “Derrotado”: Reflejan la polarización en la percepción de Gálvez, con algunos comentarios que resaltan su integridad y otros que la descalifican.
- “Conservador”, “Respetar”, “Humillación”: Muestra una relación con el rol de la candidata dentro del bloque opositor y la narrativa sobre su estrategia política.
- “Perder”, “Ganar”, “Mentir”: Indican que el debate generó posturas encontradas sobre quién tuvo mejor desempeño y quién fue objeto de mayores críticas.

El análisis del Tópico 2 revela una discusión polarizada en torno a la figura de Xóchitl Gálvez, quien aparece como la principal protagonista del debate. El debate se analiza desde tres dimensiones: credibilidad y honestidad, confrontación con Claudia Sheinbaum y posicionamiento de Gálvez dentro de la oposición.

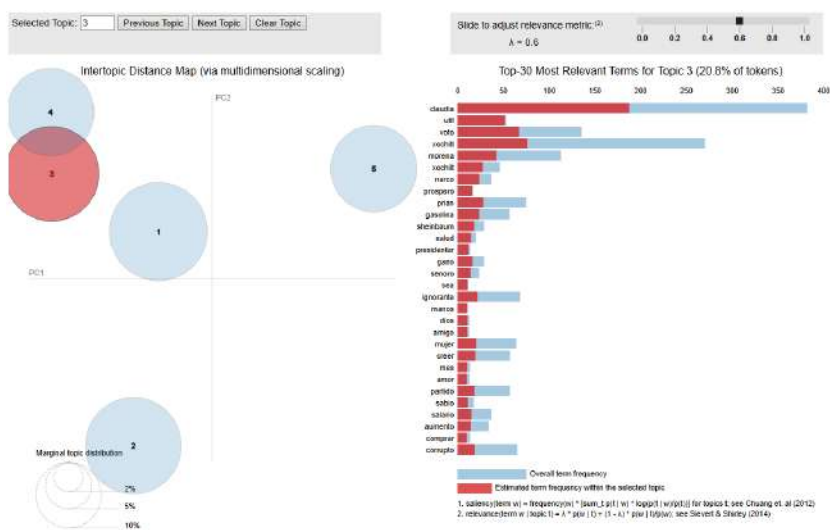
El primer eje de la conversación gira en torno a la credibilidad y honestidad de la candidata, con términos como “honesta”, “mentira”, “ignorante”, los cuales sugieren una evaluación dividida. Mientras algunos comentarios resaltan su intensidad y compromiso con la ciudadanía, otros la acusan de falta de preparación o de manipular su discurso para ganar simpatizantes. Esta tensión se acentúa en los debates digitales, donde la identidad política de los usuarios influye en la interpretación de su desempeño.

El segundo aspecto clave del análisis es la comparación directa con Claudia Sheinbaum, evidenciada por la presencia de términos como “Claudia”, “PRI” y “mentir”. En este caso, la discusión parece centrarse

en contrastar los estilos de liderazgo, experiencia y afiliación políticas de ambas candidatas. Parte de la conversación incluye críticas cruzadas sobre el rol en la administración pública y la veracidad de sus declaraciones durante el debate.

Finalmente, el tercer eje del tópico aborda el posicionamiento de Gálvez dentro de la oposición y su rol en la contienda electoral. La mención de términos como “conservador”, “humillación” y “perder” sugiere que durante parte del debate se analizó si su estrategia política fue efectiva o si quedó debilitada frente a sus contrincantes. Esta narrativa fue impulsada tanto por partidarios de Morena como por críticos internos que consideran que su candidatura no ha logrado consolidar un bloque opositor fuerte.

Figura 9. Tópico 3, segundo debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

A partir de la evidencia del Tópico 3 (figura 9) se estructura en una evaluación centrada en la candidata Claudia Sheinbaum, el cual representa el 20.8 % del total de los *tokens* analizados, examinando tanto el impacto de su postulación como las diversas posturas expresadas en

relación con su desempeño y viabilidad como aspirante a la presidencia, como se puede observar en los siguientes términos:

- “Claudia”, “Sheinbaum”, “Morena”: Reflejan que una parte significativa de la conversación giró en torno a la candidata y su desempeño en el debate, dándole popularidad entre los votantes.
- “Voto”, “Presidenta”, “Ganó”: Indican que el tema del sufragio y la viabilidad de su victoria electoral fueron discutidos ampliamente.
- “PRIAN”, “Narco”, “Corrupto”: Sugiere que hubo una discusión polarizada sobre el oficialismo y la oposición. Los términos “narco” y “corrupto” apuntan a acusaciones de vínculos con el crimen organizado y prácticas de corrupción, respectivamente, lo que refleja una preocupación por la transparencia y la ética en la política.
- “Salud”, “Gasolina”, “Salario”: Muestra que algunos comentarios abordaron políticas públicas y preocupaciones económicas en el contexto electoral.
- “Ignorante”, “Amigo”, “Sabio”: Hace referencia a la percepción de los candidatos en cuanto a su conocimiento y preparación para gobernar.

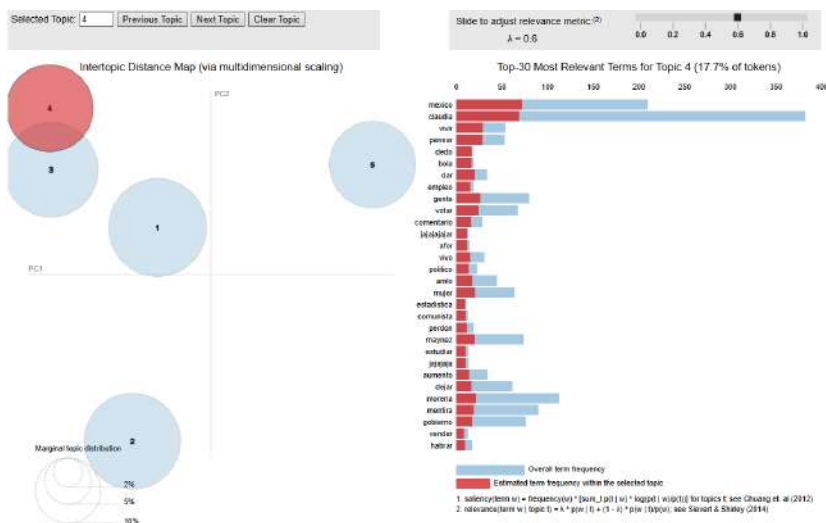
Claramente, el Tópico 3 refleja una discusión marcada por la evaluación de Claudia Sheinbaum, las implementaciones de su posible triunfo electoral y el contraste entre Morena y la oposición. Se analizan tres dimensiones principales: viabilidad de Sheinbaum como presidenta, percepción de los partidos políticos, y el debate sobre las políticas públicas.

El primer eje de la conversación gira en torno a Claudia Sheinbaum y su candidatura, con términos como “voto”, “presidenta” y “ganó”, lo que indica que los usuarios analizaron su potencial para alcanzar la victoria. Mientras algunos comentarios resaltaron su preparación y alineación con el proyecto de gobierno actual, otros expresaron críticas hacia su liderazgo y cuestionaron su capacidad para dirigir el país.

Un segundo aspecto fue la comparación entre Morena y la oposición, reflejando que existe una confrontación, donde algunos participantes defendieron a Sheinbaum como símbolo de la continuidad de la transformación política iniciada en 2018, mientras que otros la percibieron como parte de un sistema que no ha logrado resolver los problemas estructurales del país.

Por último, el tercer eje del tópico se enfocó en temas económicos y sociales, por tanto, se abordaron temas como el impacto de las políticas actuales en la vida cotidiana de la población, generando un contraste entre las posturas gubernamentales y las demandas ciudadanas en áreas sensibles y prioritarias.

Figura 10. Tópico 4, segundo debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

En la figura 10, el Tópico 4, que representa el 17.7 % total de *tokens* analizados, está orientado a la percepción del país y la relevancia del proceso electoral, entre los términos más predominantes se encuentran:

- “Vivir”, “Pensar”, “Empleo”: Indican que parte de la conversación estuvo enfocada en expectativas sobre la calidad de vida y oportunidades en el país.
- “Votar”, “Gente”, “Político”: Muestra que la participación ciudadana y la influencia de los políticos en la sociedad fueron temas abordados.
- “AMLO”, “Morena”, “Gobierno”: Sugiere que el debate incluyó referencias al rol del actual presidente y la continuidad del partido oficialista.

- “Mentira”, “Comunista”, “Perdón”: Indica que la discusión también tuvo un tono crítico y polarizado, con cuestionamientos sobre el modelo político y económico del país.

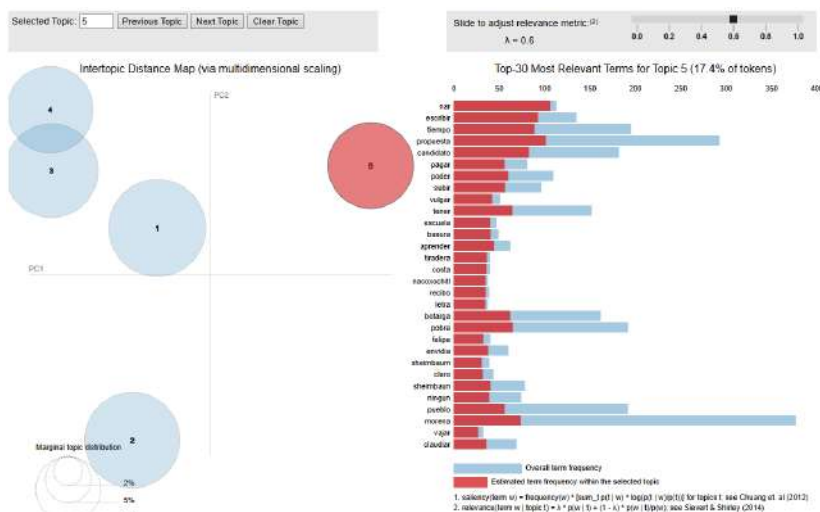
El Tópico 4 revela una conversación marcada por la incertidumbre sobre el futuro del país y el rol que desempeñan las elecciones en la vida de los ciudadanos. A partir del análisis de los comentarios, se pueden identificar tres dimensiones clave en el debate: la percepción sobre la situación de México y su calidad de vida, la importancia de la participación electoral y el rol de los políticos, así como las críticas y la polarización ideológica en torno a los proyectos de nación en disputa.

La primera dimensión se relaciona con las expectativas sobre el rumbo del país y la calidad de vida, evidenciada en términos como “vivir”, “empleado” y “pensar”, esto es un reflejo de preocupaciones sobre las oportunidades económicas y sociales, evaluando la capacidad de los candidatos para proponer soluciones viables a estos problemas.

El segundo eje de la discusión gira en torno a la participación ciudadana y el rol de los políticos, destacan términos como “votar”, “gente”, “político”. Esto sugiere que una parte del debate se enfocó en la relevancia del sufragio y en cómo la ciudadanía percibe a la clase política en su capacidad para generar cambios.

Por último, los comentarios también estuvieron atravesados por una fuerte carga ideológica y crítica política, reflejada en palabras como “mentira”, “comunista” y “perdón”, indicando que los comentarios se dividieron entre quienes respaldan la gestión actual y quienes la consideran perjudicial.

Figura 11. Tópico 5, segundo debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

Por último, el Tópico 5 (figura 11) representa el 17.4 % del total de *tokens* analizados, este tópico giró en torno a la viabilidad de los planteamientos de los aspirantes, así como las deudas y críticas sobre sus capacidades para cumplir con sus compromisos.

Los términos más significativos incluyen:

- “Propuesta”, “Candidato”, “Poder”: Indican que el foco de la conversación estuvo en la evaluación de los compromisos y promesas realizadas durante el debate.
- “Pagar”, “Subir”, “Tener”: En los comentarios se discutió la factibilidad económica de las iniciativas presentadas, particularmente en términos de costos e impacto en la ciudadanía.
- “Basura”, “Vulgar”, “Botarga”: Reflejan una parte de la discusión que tuvo un tono crítico y despectivo hacia algunos de los candidatos o sus discursos.
- “Morena”, “Sheinbaum”, “Claudia”: Sugiere que una parte de los comentarios se enfocó en la candidata oficialista, evaluando su desempeño y la viabilidad de sus propuestas.

- “Escribir”, “Tiempo”, “Aprender”: Podrían estar vinculados a debates sobre la preparación de los candidatos o la claridad con la que expusieron sus ideas.

El Tópico 5 muestra una discusión centrada en la viabilidad de las propuestas presentadas por los candidatos, así como en la percepción sobre su capacidad para implementarlas. Este análisis se puede dividir en tres dimensiones principales: evaluación de las propuestas y su factibilidad económica, críticas y escepticismo sobre los discursos, y el rol de Morena y Claudia Sheinbaum en el debate.

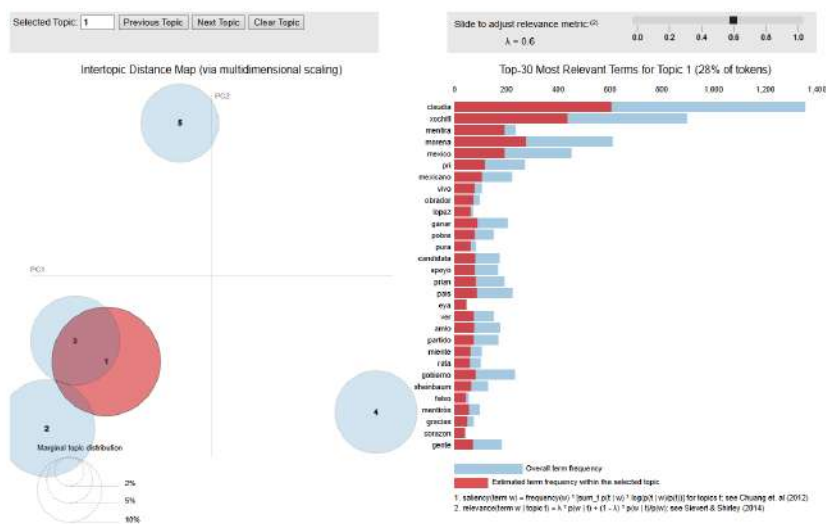
El primer eje de la discusión se relaciona con la evaluación de las propuestas y su viabilidad económica, destacada por términos como “pagar”, “subir” y “tener”. Esto sugiere que los comentarios estuvieron marcados por el análisis de costos y el impacto de los compromisos adquiridos por los candidatos, con posturas que van desde el respaldo hasta la incredulidad ante la posibilidad de cumplirlos.

En segundo lugar, la conversación también estuvo influenciada por un tono crítico y de desconfianza hacia los discursos de los candidatos, reflejado en términos como “basura”, “vulgar” y “botarga”. Estos términos sugieren que parte del debate digital estuvo cargado de descalificaciones hacia ciertos aspirantes, lo que denota una reacción emocional ante su desempeño en el debate.

Por último, una parte de los comentarios se enfocó en el rol de Morena y Claudia Sheinbaum en la contienda, con términos como “Sheinbaum”, “Claudia” y “Morena”. Esto sugiere que la conversación incluyó análisis sobre la viabilidad de su proyecto político y su desempeño en el debate, así como comparaciones con sus adversarios.

Resultados, análisis LDA, tercer debate presidencial

Figura 12. Tópico 1, tercer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

El tercer debate presidencial de México 2024 representó el último gran enfrentamiento entre los candidatos Claudia Sheinbaum, Xóchitl Gálvez y Jorge Álvarez Máynez, antes de la jornada electoral. En este evento, los aspirantes reafirmaron sus posturas y buscaron consolidar el respaldo de sus seguidores, mientras que la audiencia evaluó su desempeño en redes sociales, con especial énfasis en YouTube, donde se generaron debates sobre la viabilidad de sus propuestas y su capacidad de liderazgo.

En la figura 12 destaca el Tópico 1, el cual concentra el 28 % total de los *tokens* analizados. La presencia de términos como “Claudia”, “Xóchitl”, “mentira”, “Morena” y “México” sugiere que el debate digital se centró en el enfrentamiento entre las dos principales candidatas y en la evaluación del impacto de sus declaraciones y antecedentes políticos.

Los términos más representativos incluyen:

- “Claudia”, “Xóchitl”: Indican que la conversación giró principalmente en torno a ambas candidatas, sus intervenciones y su impacto en la contienda.
- “Mentira”, “Falso”, “Miente”: Sugieren que una parte considerable de la discusión estuvo orientada a cuestionar la veracidad de sus declaraciones.
- “Morena”, “PRI”, “PRIAN”: Indican que la conversación también incluyó referencias a los partidos políticos y sus antecedentes en la administración del país.
- “México”, “Mexicano”, “País”: Sugiere que los comentarios abordaron las implicaciones de esta elección para el futuro del país.
- “Obrador”, “AMLO”, “López”: Muestra que la figura del actual presidente sigue siendo un referente clave en la contienda electoral.
- “Apoyo”, “Ganar”, “Candidata”: Indican que una parte del debate digital estuvo enfocada en la viabilidad de cada candidata para ganar la elección.

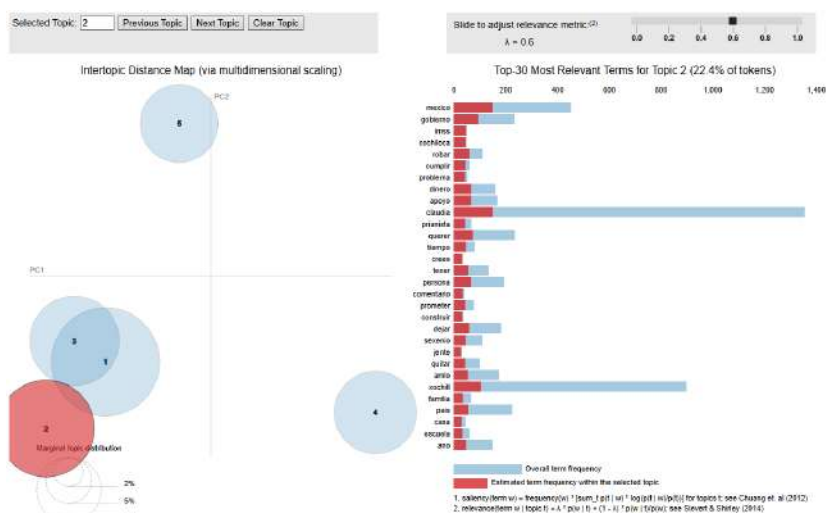
El Tópico 1 refleja una discusión fuertemente polarizada en la que los usuarios debatieron sobre la credibilidad de las candidatas, el impacto de sus declaraciones y la relación de cada una con los partidos políticos que representan. Esta conversación puede dividirse en tres dimensiones principales: confianza en las candidatas y veracidad de sus declaraciones, afiliación partidista y desempeño de los gobiernos anteriores, y evaluación de la contienda electoral y posibilidades de triunfo.

El primer eje del debate se centró en la confianza en las candidatas y la veracidad de sus declaraciones, reflejado en términos como “mentira”, “falso” y “miente”. La discusión giró en torno a si las afirmaciones de Sheinbaum y Gálvez eran genuinas o si estaban distorsionadas por intereses políticos. Este patrón es común en campañas electorales altamente competitivas, donde la desinformación y la percepción de manipulación del discurso pueden influir en la opinión pública.

En segundo lugar, la afiliación partidista y la evaluación del desempeño de administraciones pasadas desempeñaron un rol clave en la conversación. La mención de términos como Morena, PRI y PRIAN sugiere que los comentarios abordaron comparaciones entre los gobiernos previos y la administración actual. En este contexto, las críticas se dirigieron tanto a la continuidad del proyecto de Morena como al regreso de los partidos tradicionales al poder.

Por último, el debate también incluyó una evaluación de la contienda electoral y la percepción sobre el posible triunfo de cada candidata, con términos como ganar, apoyo y candidata. Este aspecto muestra que los usuarios no solo debatieron sobre el desempeño en el debate, sino que también especularon sobre quién tiene mayores posibilidades de obtener la victoria en las urnas.

Figura 13. Tópico 2, tercer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

En la figura 13, el Tópico 2, que representa el 22.4 % del total de *tokens* analizados, muestra la relevancia de términos como “México”, “gobierno”, “IMSS” y “dinero”; esto sugiere que la conversación giró en torno a la gestión de recursos públicos, el sistema de salud y las preocupaciones sobre la administración gubernamental.

Entre los términos más representativos destacan:

- “México”, “Gobierno”, “IMSS”: Indican que una parte significativa del debate digital giró en torno a la gestión del Estado en sectores clave como la salud pública.

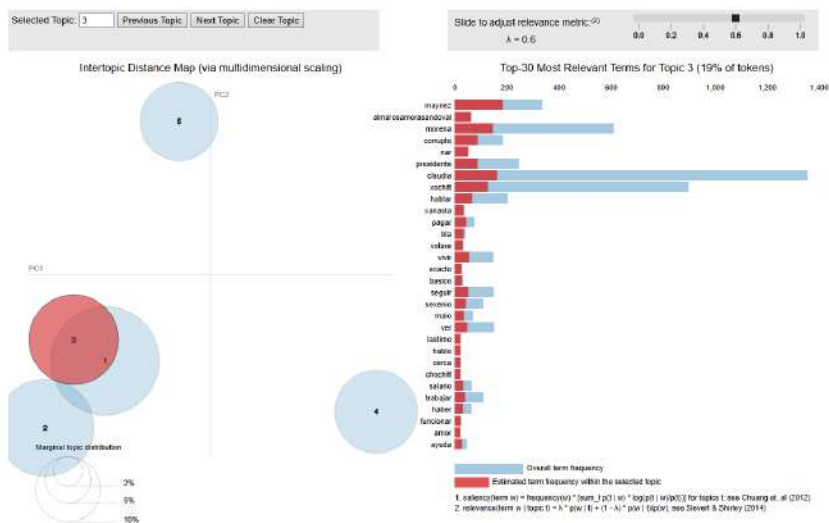
- “Robar”, “Cochinoca”, “Cumplir”: Sugiere que parte de los comentarios abordaron acusaciones de corrupción o dudas sobre la capacidad de los candidatos para cumplir sus compromisos.
- “Dinero”, “Apoyo”, “Problema”: Reflejan inquietudes sobre el manejo de los recursos públicos y la eficiencia en su distribución.
- “Claudia”, “Xóchitl”: Indican que los comentarios estuvieron divididos entre quienes apoyan o critican a las principales candidatas en su visión sobre la administración pública.
- “Familia”, “Escuela”, “Casa”: Sugiere que la conversación incluyó reflexiones sobre el impacto de las políticas gubernamentales en la vida cotidiana de los ciudadanos.

El primer eje del debate se enfoca en las críticas hacia la administración pública y la gestión de recursos, representado en términos como “robar”, “dinero” y “cumplir”. La conversación sugiere que una parte de los usuarios expresó escepticismo sobre la eficiencia del Gobierno en la distribución del presupuesto, mientras que otros respaldaron su desempeño y defendieron su continuidad.

Un segundo aspecto relevante del tópico aborda las preocupaciones sobre el acceso y la calidad de los servicios públicos, con menciones a “IMSS”, “familia” y “escuela”. Esto indica que la discusión incluyó reflexiones sobre el impacto de las políticas gubernamentales en aspectos como la educación, la salud y el bienestar social.

Finalmente, el debate también se caracterizó por la confianza o desconfianza en las candidatas principales, reflejado en términos como “Claudia”, “Xóchitl” y “apoyo”. Esto sugiere que los comentarios contrastaron la capacidad de cada candidata para atender las problemáticas mencionadas, reforzando la polarización entre quienes respaldan la continuidad del Gobierno actual y quienes buscan un cambio de administración.

Figura 14. Tópico 3, tercer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

En la figura 14 destaca en rojo el Tópico 3, el cual representa el 19 % del total de *tokens* analizados. La presencia de términos como “Claudia”, “Xóchitl”, “presidente” y “corrupto” indica que esta conversación giró en torno a la disputa entre las principales candidatas y el debate sobre la honestidad y viabilidad de sus proyectos políticos.

Algunos de los términos más representativos incluyen:

- “Claudia”, “Xóchitl”, “Maynez”: Indican que el debate se centró en la evaluación del desempeño de los tres aspirantes.
- “Presidente”, “Votar”, “Vivir”: Reflejan la importancia de la decisión electoral y las implicaciones de elegir a un líder adecuado.
- “Morena”, “Corrupción”, “Narco”: Sugiere que parte del debate estuvo marcado por cuestionamientos a los partidos políticos y acusaciones sobre su gestión.
- “Canasta”, “Salario”, “Trabajar”: Indica que la conversación incluyó referencias a temas económicos y políticas sociales.
- “Seguir”, “Malo”, “Ver”: Podrían estar vinculados a la evaluación del desempeño de los candidatos en el debate y la interpretación de sus intervenciones.

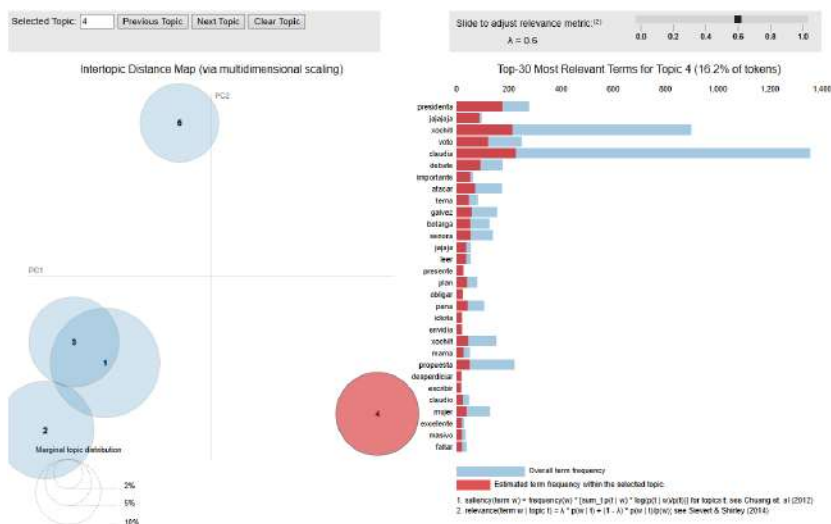
El Tópico 3 refleja una conversación marcada por la evaluación del liderazgo de los candidatos, el impacto de sus propuestas en la economía y la percepción de transparencia en la contienda electoral. Se pueden identificar tres dimensiones principales dentro del debate: credibilidad y confianza en los candidatos, impacto de la elección en la economía y bienestar social, y acusaciones de corrupción y cuestionamientos a los partidos políticos.

El primer eje de la discusión se centra en la credibilidad y confianza en los candidatos, reflejada en términos como “Claudia”, “Xóchitl”, “Maynez” y “presidente”. Esto sugiere que los comentarios estuvieron divididos entre quienes respaldan a las candidatas principales y quienes consideran que su desempeño no fue convincente. Asimismo, se debatió sobre cuál de los aspirantes tenía mayores posibilidades de ganar.

En segundo lugar, el debate incluyó referencias al impacto de la elección en la economía y el bienestar social, evidenciado en términos como “canasta”, “salario” y “trabajar”. Estos comentarios pudieron haberse enfocado en el análisis de las propuestas de los candidatos en materia económica, así como en la viabilidad de sus planteamientos para mejorar las condiciones laborales y salariales de los ciudadanos.

Por último, el tópico también estuvo marcado por acusaciones de corrupción y cuestionamientos a los partidos políticos, con menciones a “Morena”, “corrupción” y “narco”. Esto indica que una parte significativa de la discusión giró en torno a la relación de los candidatos con los partidos que representan y las críticas hacia su historial político.

Figura 15. Tópico 4, tercer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

En la figura 15, el mapa de distancia intertópica resalta en rojo el Tópico 4, que representa el 16.2 % del total de *tokens* analizados. La presencia de términos como “presidenta”, “voto”, “debate” y “propuesta” sugiere que este tópico se enfocó en la percepción de los candidatos en términos de liderazgo, desempeño en el debate y credibilidad de sus compromisos de campaña.

Entre los términos más representativos destacan:

- “Presidenta”, “Xóchitl”, “Claudia”: Indican que la conversación giró en torno a la percepción de las candidatas y la posibilidad de que México tenga su primera mujer presidenta.
- “Voto”, “Debate”, “Importante”: Sugiere que los comentarios incluyeron reflexiones sobre la relevancia del evento y su impacto en la intención de voto.
- “Propuesta”, “Desperdiciar”, “Plan”: Refleja el análisis de las promesas de campaña y la viabilidad de los proyectos planteados.
- “Gálvez”, “Botarga”, “Señora”: Indica que una parte de la discusión estuvo marcada por ataques y descalificaciones a las candidatas, lo que denota un tono de confrontación.

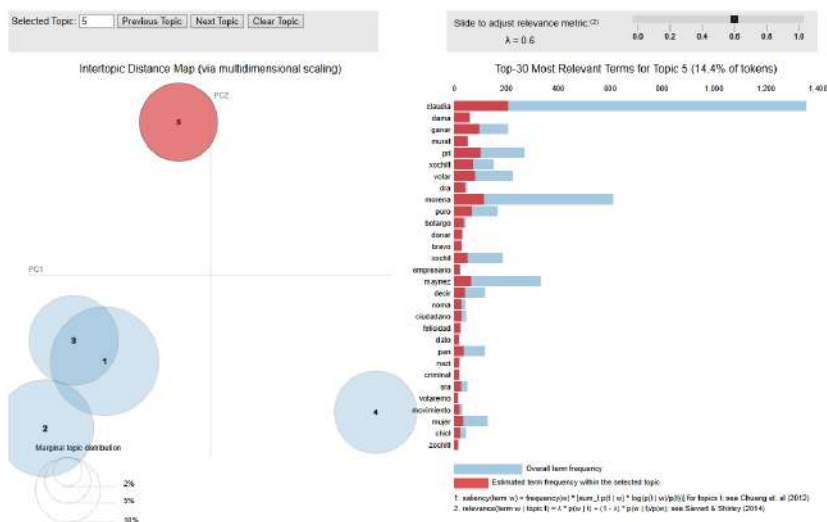
- “Escribir”, “Faltar”, “Excelente”: Puede sugerir que algunos comentarios evaluaron la claridad y efectividad del mensaje de las candidatas en el debate.

El primer eje del debate gira en torno a la posibilidad de que México tenga su primera presidenta, reflejado en términos como “presidenta”, “Xóchitl” y “Claudia”. Los comentarios parecen haber abordado tanto el respaldo como el escepticismo sobre la capacidad de las candidatas para liderar el país, con posturas divididas entre quienes consideran que su género es irrelevante frente a su experiencia y quienes lo ven como un factor simbólico en la elección.

En segundo lugar, el debate incluyó una evaluación de las propuestas y su impacto en la campaña electoral, con términos como “propuesta”, “plan” y “desperdiciar”. Esto sugiere que los usuarios discutieron la viabilidad de los proyectos presentados y si estos representan soluciones concretas a los problemas del país o si son únicamente promesas de campaña.

Por último, la conversación también estuvo marcada por críticas y descalificaciones, reflejadas en términos como “botarga”, “señora” e “idiota”. Esto indica que parte de la discusión adoptó un tono de confrontación, con ataques directos a las candidatas, lo que refuerza la polarización de la contienda.

Figura 16. Tópico 5, tercer debate presidencial, $\lambda = 0.6$



Fuente: Elaboración propia.

En la figura 16, el Tópico 5, que representa el 14.4 % del total de *tokens* analizados, muestra la preeminencia de términos como “Claudia”, “votar”, “PRI” y “Morena”, lo cual sugiere que esta conversación giró en torno a la afiliación partidista de los candidatos y la percepción de sus trayectorias políticas.

Entre los términos más representativos destacan:

- “Claudia”, “Xóchitl”, “Máynez”: Indican que la conversación giró en torno a la evaluación de los tres candidatos y sus antecedentes políticos.
- “PRI”, “PAN”, “Morena”: Reflejan el énfasis en la afiliación partidista y el contraste entre los proyectos políticos en competencia.
- “Votar”, “Movimiento”, “Volver”: Sugiere que los comentarios incluyeron llamados al voto y reflexiones sobre el impacto de la elección.
- “Ganar”, “Empresario”, “Murat”: Indica que parte de la discusión se enfocó en los respaldos políticos y los grupos de poder detrás de cada candidato.

- “Criminal”, “Nazi”, “Bravo”: Puede reflejar el tono crítico y polarizado con el que algunos usuarios debatieron sobre las figuras políticas y sus trayectorias.

El Tópico 5 evidencia una conversación centrada en la identidad política de los candidatos, el impacto de sus partidos en la elección y la percepción de los ciudadanos sobre su viabilidad como líderes. Se pueden identificar tres dimensiones principales dentro del debate: afiliación partidista y lealtad electoral, discusión sobre el impacto del voto y la movilización ciudadana, y críticas y descalificaciones en torno a los aspirantes.

El primer eje del debate gira en torno a la afiliación partidista de los candidatos y la lealtad de los votantes, evidenciado en términos como “PRI”, “PAN” y Morena. Esto sugiere que una parte de la discusión estuvo marcada por la comparación entre los partidos y sus proyectos de nación, con argumentos tanto a favor como en contra de la continuidad de Morena o el regreso de los partidos tradicionales al poder.

En segundo lugar, la conversación incluyó un debate sobre la importancia del voto y la movilización ciudadana, con términos como “votar”, “movimiento” y “volver”. Esto indica que los comentarios reflejaron llamados a la participación electoral y reflexiones sobre el impacto que la decisión de los votantes tendrá en el futuro del país.

Finalmente, el tópico también estuvo marcado por críticas y descalificaciones hacia los candidatos y sus alianzas políticas, reflejado en términos como “criminal”, “nazi” y “empresario”. Esto sugiere que una parte del debate digital se centró en ataques directos a los aspirantes, lo que refuerza el tono polarizado de la contienda.

Discusión

A continuación, se presenta una discusión que sintetiza el análisis de los comentarios de YouTube en torno a los tres debates previos a las elecciones de 2024 en México, enfatizando tanto sus similitudes como sus diferencias, y proponiendo temas derivados para cada uno de los tópicos identificados en cada debate.

Similitudes entre los debates

A pesar de las particularidades de cada encuentro, se observan varios elementos en común:

- *Polarización y evaluación de candidatos:* Los tres debates se caracterizan por una discusión intensa sobre la credibilidad y el desempeño de los candidatos, donde los términos asociados a la honestidad, integridad y veracidad (por ejemplo, “mentira”, “honesto”, “corrupción”) aparecen de manera recurrente.
- *Comparaciones entre propuestas y partidos:* Tanto en el primer como en el segundo y tercer debate se evidencia una tendencia a contrastar las propuestas de gobierno y a realizar comparaciones entre los partidos políticos (Morena, PRI, PAN) y sus antecedentes, reflejando una preocupación por la continuidad o el cambio en la administración pública.
- *Impacto en la percepción ciudadana:* Los comentarios ponen de relieve la relevancia del voto y la movilización democrática, tema que atraviesa los debates, en tanto la audiencia evalúa la factibilidad de los programas sociales y las estrategias electorales.

Diferencias entre los debates

Aunque comparten elementos comunes, cada debate presenta énfasis distintos en función de la coyuntura y de los discursos de los candidatos:

- *Primer debate (7 de abril de 2024):*
 - Enfoque en la evaluación comparativa de desempeño:* Se destaca la confrontación directa entre los tres candidatos, con una atención particular a la comparación entre Claudia Sheinbaum y Xóchitl Gálvez.
 - Discusión sobre integridad y manipulación mediática:* La presencia de términos como “honestidad”, “corrupción” y “medios” refleja la preocupación por la credibilidad de los discursos y la posible manipulación en la cobertura mediática.

- Segundo debate (28 de abril de 2024):
Énfasis en la crisis de seguridad y el contexto regional: Los términos relacionados con “muerte”, “Jalisco” y “gasolinera” sugieren que la discusión se centró en la problemática de la inseguridad, haciendo énfasis en problemas locales que impactan en la percepción nacional.
Foco en la evaluación de Xóchitl Gálvez: Se destaca la polarización en torno a su figura, generando debates sobre su honestidad y capacidad de liderazgo en contraste con Claudia Sheinbaum.
- Tercer debate (último encuentro antes de las elecciones):
Intensificación de la confrontación sobre veracidad y afiliación partidista: La discusión se centra en la credibilidad de las declaraciones, la relación de los candidatos con sus partidos (Morena, PRI, PAN) y la posibilidad de una presidencia femenina.
Evaluación final de propuestas y movilización electoral: Se observan debates sobre la capacidad de los candidatos para gestionar los recursos públicos y responder a las expectativas de la ciudadanía, lo que se traduce en términos relacionados con “IMSS”, “dinero” y “votar”.

Temáticas y tópicos por debate

Primer debate:

Tópico 1: Evaluación del desempeño y la credibilidad

- *Temas derivados:*
 - Evaluación comparativa de desempeño (presencia de nombres y valoración ética).
 - Críticas y defensas sobre la integridad de los candidatos.
 - Discusión sobre el rol y la posible manipulación mediática.

Tópico 2: Centrarse en Claudia Sheinbaum y la continuidad política

- *Temas derivados:*
 - Potencial continuidad del proyecto político de la 4T.
 - Análisis de propuestas y seguridad pública.
 - Comparaciones implícitas con la oposición.

Tópico 3: Gobernabilidad y comparaciones internacionales

- *Temas derivados:*
 - Evaluación del rol del Estado en la gestión de problemas nacionales.
 - Debate sobre la veracidad de los discursos.
 - Referencias a crisis comparativas a escala internacional.

Tópico 4: Participación ciudadana y liderazgo femenino

- *Temas derivados:*
 - Discusión sobre el valor del voto y la representación política.
 - Análisis del impacto de tener candidatas femeninas.
 - Evaluación de la ética y capacidad de liderazgo.

Tópico 5: Propuestas de gobierno y programas sociales

- *Temas derivados:*
 - Valoración de las propuestas presentadas y su factibilidad.
 - Evaluación de estrategias de políticas públicas.
 - Críticas sobre la ejecución de los programas sociales.

Segundo debate:

Tópico 1: Crisis de seguridad y gobernabilidad

- *Temas derivados:*
 - Evaluación de la gestión de la seguridad pública en contextos regionales.
 - Comparación entre administraciones pasadas y presentes.
 - Impacto de la inseguridad en la percepción ciudadana.

Tópico 2: La figura de Xóchitl Gálvez en el debate

- *Temas derivados:*
 - Evaluación de la credibilidad y honestidad de la candidata.
 - Comparación con Claudia Sheinbaum en términos de liderazgo.
 - Debate sobre la efectividad de la estrategia de oposición.

Tópico 3: Evaluación de Claudia Sheinbaum y las políticas públicas

- *Temas derivados:*
 - Discusión sobre la viabilidad electoral y el liderazgo.

- Análisis comparativo entre la continuidad del oficialismo y la oposición.
- Evaluación de propuestas en seguridad y bienestar.

Tópico 4: Perspectivas sobre la situación nacional y la participación electoral

- *Temas derivados:*

- Reflexiones sobre la calidad de vida y las oportunidades.
- Importancia del voto y la participación ciudadana.
- Críticas a modelos políticos y polarización ideológica.

Tópico 5: Factibilidad de propuestas y críticas a los discursos

- *Temas derivados:*

- Evaluación de la factibilidad económica y social de las propuestas.
- Críticas sobre la preparación y claridad en los discursos.
- Debate sobre la capacidad de los candidatos para cumplir compromisos.

Tercer debate:

Tópico 1: Confrontación directa y credibilidad de las declaraciones

- *Temas derivados:*

- Análisis de la veracidad de las afirmaciones de las candidatas.
- Comparación del estilo comunicativo y la retórica de cada candidato.
- Debate sobre la influencia de antecedentes políticos en la percepción actual.

Tópico 2: Gestión de recursos públicos y servicios esenciales

- *Temas derivados:*

- Evaluación de la eficiencia en la administración de recursos (ej. imss, dinero público).
- Discusión sobre la calidad de los servicios públicos y su impacto social.
- Críticas sobre prácticas de corrupción y mal manejo de recursos.

Tópico 3: Evaluación del liderazgo y las propuestas económicas

- *Temas derivados:*
 - Análisis del impacto de las propuestas en la economía y el bienestar social.
 - Debate sobre la viabilidad de proyectos y el compromiso electoral.
 - Acusaciones y defensas en torno a la corrupción.

Tópico 4: Liderazgo femenino y propuestas de campaña

- *Temas derivados:*
 - Discusión sobre la posibilidad de una presidencia femenina.
 - Evaluación de la efectividad y claridad de las propuestas presentadas.
 - Análisis de la movilización ciudadana en respuesta a las propuestas.

Tópico 5: Afiliación partidista y movilización electoral

- *Temas derivados:*
 - Evaluación de la lealtad electoral y respaldo de cada partido (Morena, pri, pan).
 - Discusión sobre el impacto de las alianzas políticas y los respaldos de grupos de poder.
 - Críticas sobre la identidad política y llamados a la participación.

Conclusiones

El análisis de los tres debates evidencia que, si bien existe una base común en términos de evaluación de candidatos, discusión sobre la gobernabilidad y la polarización mediática, cada debate presenta énfasis específicos. El primer debate se centra en la evaluación comparativa y en la discusión sobre la integridad de los discursos, el segundo debate destaca la crisis de seguridad y la evaluación diferencial de las candidatas, mientras que el tercer debate intensifica la confrontación sobre la veracidad, la gestión de recursos públicos y la polarización partidista. Esta diversidad temática refleja la complejidad del escenario político mexicano de 2024 y la importancia de analizar, desde una

perspectiva sociopolítica, tanto los discursos emitidos en vivo como las interpretaciones y reacciones que estos generan en el espacio digital.

Esta discusión, al mismo tiempo que propone temas para profundizar en cada tópico, ofrece un panorama integral sobre cómo la opinión pública articula sus percepciones en función de las dinámicas propias de cada debate, permitiendo comprender mejor la intersección entre comunicación política, polarización y movilización ciudadana en contextos electorales contemporáneos.

Referencias

- Allcott, H., y Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-236. DOI:10.1257/jep.31.2.211
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Fallin Hunzaker, M. B., ... y Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221. DOI:10.1073/pnas.1804840115
- Blei, D. M., Ng, A. Y., y Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Casar, M. A. (2021). México: la polarización como estrategia. *Revista de Ciencia Política (Santiago)*, 41(2), 481-507.
- Dunn, W. N. (2017). *Public policy analysis*. Routledge.
- Gelfand, A. E. (2000). Gibbs Sampling. *Journal of the American Statistical Association*, 95(452), 1300-1304. <https://doi.org/10.2307/2669775>
- Griffiths, T. L., y Steyvers, M. (2004). Finding scientific topics. *Proc Nat Acad Sci USA*, 101, 5228-5235. <https://www.pnas.org/doi/epdf/10.1073/pnas.0307752101>
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., y Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. *Science*, 363(6425), 374-378. DOI:10.1126/science.aau2706
- Grün, B., y Hornik, K. (2011). Topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1-30.
- Hoffman, M. D., Blei, D. M., Wang, C., y Paisley, J. (2013). Stochastic Variational Inference. *Journal of Machine Learning Research*, 14, 1303-1347.
- Lawson, C. H. (2019). *Mexico's 2018 elections: AMLO's landslide and the transformation of Mexican democracy*. Routledge.

- Permadi, D., y Putri, T. E. (2022). Social Media Analytics for Policy Studies. En S. Nair, y N. Varma (Eds.), *Emerging Pedagogies for Policy Education*. Palgrave Macmillan, Singapur. https://doi.org/10.1007/978-981-16-5864-8_7
- Rosen-Zvi, M., Chemudugunta, C., Griffiths, T., Smyth, P., y Steyvers, M. (2010). Learning author-topic models from text corpora. *ACM Trans Inform Syst*, 28(1), 1-38. <https://doi.org/10.1145/1658377.1658381>
- Siklos, P. L., St. Amand, S., y Wajda, J. (2018). Appendix A: Latent Dirichlet Allocation. En *The Evolving Scope and Content of Central Bank Speeches* (pp. 20-20). Centre for International Governance Innovation. <http://www.jstor.org/stable/resrep51882.12>
- Skoric, M. M., Liu, J., y Jaidka, K. (2020). Electoral and Public Opinion Forecasts with Social Media Data: A Meta-Analysis. *Information*, 11(4), 187. DOI: 10.3390/info11040187
- Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., ... y Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political Science Research and Methods*, 6(2), 215-242. DOI:10.1017/psrm.201735
- Vosoughi, S., Roy, D., y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. DOI:10.1126/science.aap9559
- Wei, X., y Croft, W. B. (2006). LDA-based document models for ad-hoc retrieval. *Proceedings of the international ACM SIGIR Conference on Research and Development in Information Retrieval*. Nueva York: ACM, pp. 178-185. <https://doi.org/10.1145/1148170.1148204>

Capítulo 5

Herramientas y técnicas computarizadas al servicio de la construcción de *corpus* y *datasets*. Desafíos metodológicos del análisis de discursos mediatizados en plataformas con base en Internet

Natalia Raimondo Anselmino¹

Irene Gindin²

Alejandro Sartorio³

Franco Cellone⁴

Verónica Aráoz⁵

Francisco J. Alomar⁶

1. Centro de Investigaciones en Mediatizaciones (Universidad Nacional de Rosario). Consejo Nacional de Investigaciones Científicas y Técnicas (Conicet). Correo electrónico: nraimondo@conicet.gov.ar. ORCID: <https://orcid.org/0000-0001-7217-8754>
2. Centro de Investigaciones en Mediatizaciones (Universidad Nacional de Rosario). Consejo Nacional de Investigaciones Científicas y Técnicas (Conicet). Correo electrónico: iregindin@gmail.com. ORCID: <https://orcid.org/0000-0002-3080-0207>
3. Centro de Altos Estudios en Tecnología Informática (Universidad Abierta Interamericana). Correo electrónico: Alejandro.Sartorio@uai.edu.ar. ORCID: <https://orcid.org/0000-0002-0735-1810>
4. Centro de Estudios Avanzados (Universidad Nacional de Córdoba). Correo electrónico: francellone@mi.unc.edu.ar. ORCID: <https://orcid.org/0009-0009-3948-5248>
5. Centro de Conocimiento, Formación e Investigación en Estudios Sociales (Universidad Nacional de Villa María). Consejo Nacional de Investigaciones Científicas y Técnicas (Conicet). Correo electrónico: varaoz@unvm.edu.ar. ORCID: <https://orcid.org/0000-0003-0500-699X>
6. Facultad de Ciencias Exactas, Ingeniería y Agrimensura (Universidad Nacional de Rosario). Correo electrónico: franciscojalomar@gmail.com. ORCID: <https://orcid.org/0009-0001-3010-7533>

Introducción

Estas páginas tienen como finalidad compartir la experiencia de aplicación (aún en curso) de una estrategia metodológica híbrida desplegada por un equipo multidisciplinario en pos de la producción de conocimiento respecto a la plataformización (van Dijck, Poell y de Waal, 2018) de los discursos sobre lo público común.⁷ Partimos de considerar que, en nuestras sociedades contemporáneas, los discursos sobre lo público común deben entenderse en inevitable conexión con la organización de la experiencia colectiva, la praxis política y las tensiones entre disputa y consenso que atraviesan la esfera pública de la época (Raimondo Anselmino, Reviglio y Diviani, 2015). Durante gran parte del siglo xx, dicha esfera pública se encontró predominantemente nucleada en torno a la relación entre el espacio público tradicional (el de la plaza o la calle) y el sistema de medios con modalidad *broadcasting*. Sin embargo, el incremento de la complejidad inherente al estadio actual de la mediatización (Verón, 2013) nos sitúa en una coyuntura que Fernández (2018) define como *ecosistema post-broadcasting* y Carlón (2020) sintetiza con la denominación de *sociedad hipermediatizada*. En ambos casos, se advierte la consolidación de un sistema de medios con base en la sinergia entre Internet y la telefonía móvil, y con modalidad *networking*, que otorga a individuos y colectivos sociales la posibilidad de constituirse en enunciadores con alcance masivo y repercusión relativamente equivalente a la que antes era exclusiva de medios como la prensa, la radio o la televisión (Fratlicelli, 2023: 40). En cualquier caso, un escenario en el cual, como señaló

7. El proyecto Discursos sobre lo público-común plataformizados: caracterización interdisciplinaria de las estrategias enunciativas empleadas en plataformas mediáticas contemporáneas y sus flujos de sentido" (código 80020220700094UR) se encuentra acreditado ante el Programa para la Investigación Universitaria Argentina (2023-2026). Es ejecutado en el marco de la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Rosario (UNR), bajo la dirección de Natalia Raimondo Anselmino e Irene Gindin, con radicación en el Centro de Investigaciones en Mediatizaciones (CIM) de la UNR. Integran, asimismo, el proyecto las y los siguientes colegas: Pablo Torres (asesor); Cecilia Echecopar, Guillermo Leale, Alejandro Sartorio, María Beatriz Abrach, María Clara Musante y Dolores Sollberger (en calidad de investigadores); Franco Cellone, Verónica Aráoz, Emmanuel Pérez Zamora, Adrián Pino y Ezequiel Aguilera Matute (como becarios/as); Natalia Cioitti, Tomás Milano, Francisco Alomar y Daniela Sánchez (como auxiliares de investigación).

Cingolani (2013), se advierte “un acceso a lo público cada vez menos generalizable” (p. 111) y, por tanto, fragmentado, con consecuencias para las democracias mediante.

Es en este contexto que se considera relevante analizar las estrategias enunciativas (Raimondo Anselmino, 2011) que distinguen a los discursos sobre lo público-común cuando son publicados en distintas plataformas mediáticas (Fernández, 2018), así como los flujos de sentido (Carlón, 2020) que dichas estrategias propician. Dadas las dificultades que supone un estudio sociosemiótico multiplataforma, esta meta se enfrenta circunscribiendo la investigación a un caso único de carácter instrumental (Marradi, Archenti y Piovani, 2018).

El dominio empírico de este estudio comprende, por tanto, un conjunto voluminoso y diverso de discursos circulantes a través de distintas plataformas (portales de prensa *online*, y cuentas de Instagram, X, Facebook y WhatsApp) en torno a un conjunto de hechos violentos suscitados en la ciudad de Rosario (Provincia de Santa Fe, Argentina) entre el 2 de marzo y el 6 de mayo de 2024, que fueron rápidamente calificados (por actores políticos, institucionales y mediáticos) como *narcoterrorismo*⁸.

La secuencia de acontecimientos a los que refieren los discursos comprendidos en el caso se encuentra desarrollada en Gindin y Raimondo Anselmino (2024), aunque puede sintetizarse que se trató de una escalada de hechos delictivos caracterizados por utilizar la fuerza organizada de bandas criminales de la ciudad como mecanismo de coerción dirigido, explícitamente, al poder político provincial. Hechos que no solo desencadenaron cuatro asesinatos, sino que generaron una situación de alarma tan generalizada en la población que paralizaron el funcionamiento de múltiples instituciones (escuelas, centros de salud, transporte público, comercios, etcétera) durante varios días y requirieron la intervención del Gobierno nacional, las fuerzas de seguridad federales y los distintos poderes del Estado, con consecuencias que llegan hasta la actualidad y repercuten, incluso, por fuera de las fronteras locales.

8. Una revisión de las implicancias de este término, en tanto construcción conceptual y teórica alrededor de fenómenos ligados con el narcotráfico y el terrorismo, puede leerse en Mantilla-Valbuena (2008).

A los fines de esta publicación colectiva, nos limitaremos a exponer las definiciones de orden metodológico que propiciaron la construcción de los *corpus* de esta investigación, así como el arduo proceso de generación de *sets* de datos asociados a los discursos que es de nuestro interés analizar, con todas las dificultades que esta instancia supone en el caso de estudios de enfoque híbrido como el nuestro (Gindin y Busso, 2018; Kataishi y Milia, 2024). Esto ha sido posible gracias a la colaboración interdisciplinaria e interinstitucional de saberes provenientes de las ciencias sociales (como la sociología, la psicología, la semiótica, la comunicación social, la lingüística o la antropología) y otros derivados de las ciencias computacionales (que, a su vez, se encuentran en estrecha relación con técnicas y procedimientos de la matemática, la estadística o la informática). Dicho esto, a continuación se explicitan los objetivos y el abordaje empírico de nuestra investigación, y se presentan los distintos procedimientos técnicos llevados a cabo, así como las potencialidades y limitaciones que hemos encontrado en el camino hasta aquí recorrido para producir, desde el sur latinoamericano, conocimiento respecto a la dimensión significativa de un fenómeno social complejo.

Definición de objetivos y estrategia metodológica

En función de caracterizar las variaciones en las estrategias enunciativas empleadas en discursos (de enunciadores socioindividuales, institucionales o mediáticos) sobre el caso publicado en diversas plataformas, y sus flujos de sentido, esta investigación se propuso los siguientes objetivos específicos: 1. Reconocer operaciones enunciativas recurrentes que sean infrecuentes fuera de las plataformas mediáticas en cuestión; 2. Describir el funcionamiento discursivo que asumen esas operaciones enunciativas; 3. Identificar los modos de empleo y el tipo de función que desempeñan los recursos paratextuales y paralingüísticos (entendidos en los términos propuestos en Raimondo Anselmino *et al.*, 2019); 4. Analizar las lógicas de construcción de los destinatarios manifiestas, atendiendo especialmente a la conformación de identidades colectivas en oposición a la configuración de coenunciadores socioindividuales; 5. Analizar el modo en que se despliega la polémica; 6. Tipologizar los

modos de composición de las unidades discursivas, identificando rasgos que manifiesten la performatividad de cada plataforma mediática en cuestión; y 7. Reconstruir y caracterizar los flujos de sentido promovidos por las estrategias enunciativas presentes en los discursos analizados.

Para responder a los objetivos propuestos partimos de un abordaje definido como *semiodata* (Raimondo Anselmino, 2022), en tanto se aplica una estrategia metodológica combinada que integra interdisciplinariamente el estudio empírico de la puesta en discurso desde un punto de vista sociosemiótico (Verón, 1998) con el empleo de herramientas y técnicas computacionales diversas, y el análisis univariante y multivariante de datos y metadatos (Aldas y Uriel, 2017). Además, para la ejecución del objetivo particular 7, utilizamos el modelo teórico-metodológico propuesto por Carlón (2020) en lo que atañe a la distinción entre flujos de sentido con modo descendente y ascendente (esto es, de los medios tradicionales a las plataformas mediáticas con modalidad *networking*, o viceversa) así como entre flujos de adentro hacia fuera o de fuera hacia dentro (tomando en consideración actores sociales diversos, así como colectivos e instituciones).

Como en parte se adelantó, nuestro dominio empírico comprende un paquete voluminoso de textualidades digitales que (junto con sus datos y metadatos asociados), y a diferencia de lo que sucede en investigaciones semióticas tradicionales usualmente concentradas en el análisis de un único *corpus*, aquí se encuentran compendiados en un total de cinco *datasets*. Esto es así porque, como bien anticipan Kataishi y Milia (2024), “a mayor tamaño y complejidad del *corpus*, mayores serán las dificultades de construir un único *dataset* en el que convivan variables estructurales, derivadas de los metadatos, y variables relacionadas con el texto a analizar” (p. 286). Siendo así, un primer *dataset* contiene registros de notas publicadas en seis portales de prensa locales y nacionales (lacapital, rosario3, elciudadanoweb, por un lado; lanacion, infobae y clarin, por el otro),⁹ así como los comentarios de usuarios-lectores a estas (esto exclusivamente para el caso de los periód-

9. Los portales lacapital.com.ar, elciudadanoweb.com, lanacion.com.ar y clarin.com son ediciones *online* producidas desde las redacciones de periódicos impresos (*La Capital*, *El Ciudadano & La Región*, *La Nación* y *Clarín*, respectivamente). En cambio, rosario3.com es un sitio web producido por el tres TV, canal de televisión abierta con alcance regional, e infobae.com es un diario nativo digital con alcance nacional.

dicos locales). Un segundo *dataset* recoge publicaciones en Facebook, y un tercero, en Instagram, en ambos casos realizados tanto por las cuentas oficiales de los medios anteriormente mencionados como por parte de distintos funcionarios públicos locales, provinciales y nacionales que tuvieron vinculación con el caso en cuestión. En cuarto lugar, otro *dataset* combina publicaciones en las cuentas de los portales de noticias y de los funcionarios en X. Y, por último, una quinta colección aglutina intervenciones¹⁰ (de texto, imágenes, audio y video) compartidas por la población local a través de WhatsApp.

Entre todas esas unidades discursivas ya compendiadas, las únicas recuperadas completamente de modo artesanal (es decir, sin mediar para su recolección el empleo de herramientas y técnicas informáticas) fueron las de WhatsApp. Dicha colección se efectuó aplicando la técnica de muestreo no probabilístico conocida como estrategia de bola de nieve (Martín-Crespo Blanco y Salamanca Castro, 2007), mediante la cual se obtienen progresivamente las unidades de análisis, partiendo de los contactos facilitados por informantes con quienes tiene vínculo el investigador. De esta manera, este *corpus* en particular comprende un total de 221 intervenciones que, al momento de escribir este informe, se encuentran en proceso de clasificación empleando una matriz con variables cualitativas definidas *ad hoc*. No desconocemos que este paquete, en particular, es un archivo acotado, heterogéneo y fragmentario de lo que circuló a través de la plataforma de mensajería instantánea durante el periodo bajo estudio, sobre todo debido que (a diferencia de lo que sucede con las otras plataformas consideradas) lo compartido en WhatsApp no tiene estatuto público. Aun así, se decidió contemplar esta plataforma en el relevamiento de nuestra investigación para tener al menos un pequeño atisbo exploratorio de qué sucede en el nivel *underground* del sistema hipermediático (Carlón, 2020).

En cambio, para la captura y preprocesamiento de los discursos publicados en portales periódicos, así como en Facebook, Instagram

10. Se recupera aquí una categoría propuesta por la Escuela de Ginebra para aludir a las unidades que se saben parte de una conversación (Kerbrat-Orecchioni, 1996). Aunque en nuestro caso no será posible recuperar el sistema de intercambios discursivos (Fernández, 2018) en el cual se enmarca cada una de esas intervenciones, nos proponemos, aun así, hacer de estas un análisis enunciativo, en tanto en las piezas recolectadas es posible observar, al menos, la definición de una escena enunciativa (Fratelli, 2023) y un tipo particular de propuesta de interacción.

y X (y sus respectivos datos y metadatos asociados) se emplearon herramientas y técnicas informáticas que automatizaron (o, en algunos casos, semiautomatizaron) parte de estos procedimientos, tal como se explicará en los siguientes apartados.

Recolección y preprocesamiento de notas publicadas en portales de información general

Para construir el primero de los *datasets* arriba mencionados se empleó una infraestructura metodológica y tecnológica avanzada desarrollada en el marco del Centro de Altos Estudios en Tecnología Informática (CAETI) de la Universidad Abierta Interamericana, en su sede Rosario. Esta fue diseñada para extraer noticias de portales web y crear un *dataset* enriquecido adecuado para su análisis mediante inteligencia artificial generativa (IAG), utilizando modelos de lenguaje de gran escala (LLM) altamente sobreentrenados. Esta plataforma permitió la recuperación de una parte de la información digital requerida, al tiempo que también ofrece funcionalidades para construir *embeddings* detallados, generar reportes comprensivos, y visualizar resultados a través de tableros interactivos. Los procedimientos implementados facilitan la ejecución de análisis complejos y la integración de LLM, permitiendo interoperabilidad y construcción de *pipelines* personalizados adaptados a diversas necesidades analíticas.

Para comprender en profundidad el funcionamiento de esta herramienta y su valor como componente técnico central del abordaje metodológico, es necesario detallar los pasos operativos que componen su lógica de uso. Su puesta en marcha implicó una secuencia de acciones técnicas articuladas, desde su instalación y configuración inicial hasta la recolección automatizada de datos, su análisis mediante técnicas de recuperación aumentada (RAG) y la posterior inserción de resultados en módulos especializados. A través de este procedimiento se garantizó la trazabilidad del flujo de datos, la replicabilidad del proceso investigativo y la posibilidad de escalar el sistema hacia otros dominios temáticos.

Complementariamente, desde CAETI se desarrolló un conjunto de módulos funcionales que amplían las capacidades de la herramienta

más allá de la simple extracción de contenidos. Estas funcionalidades comprenden desde la configuración dinámica de fuentes de datos, la parametrización de reglas de búsqueda, el etiquetado semántico automatizado y el análisis con técnicas de procesamiento de lenguaje natural (NLP), hasta la visualización avanzada en distintos formatos y la gestión de usuarios y permisos. Este ecosistema de servicios convierte a la herramienta en una solución integral, orientada tanto a la sistematización de información como a su explotación analítica en investigaciones sociales.

Finalmente, para dar solidez técnica al sistema, se optó por una arquitectura modular que articula todos los componentes involucrados en el proceso: desde los portales de noticias como fuentes de entrada, pasando por las librerías de extracción y los planificadores de tareas, hasta los módulos desarrollados sobre el *framework* Odoo y la integración con modelos de lenguaje tipo RAG. Esta arquitectura permite visualizar con claridad cómo se distribuyen las responsabilidades técnicas.

Proceso técnico de implementación

La recolección de notas periodísticas en portales constituye uno de los primeros eslabones del proceso automatizado de construcción del *corpus* analizado. Este procedimiento fue viabilizado a través del desarrollo e implementación de esta herramienta de *software* construida sobre el *framework* Odoo, que permite articular capacidades de gestión, automatización, visualización y análisis de datos en un entorno modular, flexible y ampliable. La puesta en marcha de la herramienta contempló una secuencia de fases que garantiza su operatividad de forma integral, a saber:

Paso 1. Instalación de la herramienta: El sistema fue instalado como un módulo *add-on* personalizado dentro de una instancia de Odoo, con soporte para los módulos base de Contactos, Mensajería y Tableros. La arquitectura fue diseñada para funcionar en servidores locales o en entornos *cloud*, según los requerimientos del proyecto.

Paso 2. Configuración de los componentes: Se establecieron perfiles de usuarios, permisos diferenciados y criterios de acceso a los distintos modelos de datos. La modularidad del sistema permitió

una configuración específica para los fines investigativos, orientada al análisis de medios digitales.

Paso 3. Puesta en servicio: Mediante interfaces administrativas, se habilitó la carga de portales a ser monitoreados. A su vez, se configuraron tareas automatizadas según criterios de periodicidad y prioridad de actualización.

Paso 4. Acceso a portales: Se implementaron técnicas de *web scraping* del tipo *spider*, que permiten explorar la estructura de los portales seleccionados, recuperando publicaciones en tiempo real o bajo programación periódica. La prioridad de cada fuente puede definirse mediante escalas preestablecidas (alta, media, baja, urgente), lo cual incide en la frecuencia de recolección.

Paso 5. Extracción de datos: El sistema recupera elementos clave de cada publicación (título, bajada, cuerpo del texto, URL) de forma estructurada. Las reglas de extracción son definidas mediante expresiones regulares y operadores lógicos, que permiten orientar el proceso hacia determinados términos o patrones lingüísticos.

Paso 6. Aplicación del proceso RAG: Una vez extraída, la información es procesada mediante técnicas de RAG, combinando recuperación contextualizada y generación asistida por LLM. Esto permite enriquecer el texto con metadatos, análisis de sentimiento, etiquetas semánticas y nubes de palabras.

Paso 7. Obtención y visualización de resultados: La información es almacenada, etiquetada y organizada en función de las reglas de clasificación configuradas. Posteriormente, es accesible a través de formularios especializados que permiten navegar, filtrar, editar o exportar los datos según distintos criterios.

Paso 8. Inserción de los resultados en módulos especializados: El sistema cuenta con módulos de visualización gráfica y tabular que presentan los resultados según estándares de sistemas de gestión de información. Además, incluye opciones de exportación en formatos .xls y .csv, facilitando la interoperabilidad con otras herramientas de análisis.

Servicios y funcionalidades de la herramienta

La aplicación desarrollada no se limita a la automatización del *scraping*, sino que también constituye un entorno completo para la gestión y análisis de información extraída de portales web. Entre sus principales funcionalidades se destacan: registro, edición y configuración de portales web a ser monitoreados; gestión de reglas de búsqueda avanzadas con uso de expresiones regulares; control de frecuencia de monitoreo basado en prioridades; visualización de contenido enriquecido con metadatos; representación de resultados en formato de tablas dimensionales y gráficos; exportación e importación en formatos interoperables (.xls, .csv); búsqueda, filtrado y agrupamiento de resultados mediante *dashboards* personalizados; configuración de perfiles de usuario y control de accesos; sistema de mensajería interna para usuarios de la aplicación; integración con un módulo de gestión de contenido (CMS) para publicación web de resultados.

De manera que esta infraestructura técnica desarrollada por CAETI responde a una necesidad creciente en las ciencias sociales computacionales: disponer de herramientas capaces de articular la recolección automatizada de datos con su análisis, ofreciendo una solución escalable y reproducible.

Descripción de la arquitectura de la aplicación

La arquitectura general de la herramienta comprende seis componentes principales, organizados bajo un enfoque modular y escalable, lo cual garantiza la sostenibilidad técnica del sistema, así como su adaptabilidad futura en entornos de investigación de datos abiertos.

Componente 1 - Portales de noticias: constituye la fuente primaria de información. Se trata de un conjunto de URL correspondientes a medios de comunicación digitales, previamente registrados en el sistema por usuarios administradores. Cada portal se configura con metadatos básicos (nombre, categoría, URL base) y con una prioridad de monitoreo (alta, media, baja, urgente) que impacta directamente en la frecuencia con la cual será accedido por el

sistema. Esta colección de portales alimenta continuamente al sistema y es el punto de partida para el despliegue automatizado de tareas de extracción.

Componente 2 - Librerías computacionales (Python): agrupa todas las librerías del lenguaje Python utilizadas para habilitar la conectividad, extracción, parseo y estructuración de datos web. Entre ellas se incluyen librerías para el acceso a URL, como *requests* o *http.client*, que permiten establecer conexiones HTTP seguras; herramientas específicas para *scraping*, como BeautifulSoup, lxml o Selenium, que permiten recorrer estructuras DOM y localizar elementos relevantes mediante selectores CSS o etiquetas HTML; así como librerías auxiliares destinadas a la gestión de excepciones, registro de *logs*, conversión de caracteres y tratamiento de texto. Todas estas bibliotecas son orquestadas mediante *scripts* propios que se ejecutan a través de planificadores internos del sistema (ver Componente 3). Su implementación garantiza la ejecución confiable de los pasos de extracción de información y constituyen una práctica recomendada en desarrollos orientados al análisis automatizado de medios digitales.

Componente 3 - Módulos de tareas planificadas: representa la infraestructura de automatización y ejecución periódica de procesos. A través de la utilización del sistema de acciones planificadas de Odoo, se crean tareas programadas que disparan los *scripts* de *scraping* en función de los criterios definidos para cada portal (frecuencia, horarios, condiciones). Cada tarea es parametrizable, puede activarse o desactivarse según calendario, y permite asegurar la continuidad en la alimentación del sistema con nuevos datos, lo cual responde al requerimiento de actualización constante en el tratamiento de fenómenos de alta dinámica discursiva. Este componente da soporte técnico al paso 3 del proceso técnico y activa los procesos definidos en el paso 4, descritos anteriormente.

Componente 4 - Odoo Framework: Odoo actúa como entorno de ejecución y *framework* de desarrollo, proporcionando servicios de base de datos (PostgreSQL), capa de servidor (Python) y *frontend web* (Qweb, XML, JS). La aplicación fue construida sobre la versión Community Edition, lo que garantiza su disponibilidad como *software* libre y su adecuación a entornos de investigación académica y de innovación abierta. Odoo provee el contexto organizativo para

integrar de manera nativa los módulos específicos desarrollados para esta solución, incluyendo *dashboards*, formularios, permisos y flujos de trabajo colaborativo.

Componente 5 - Módulo *Addon* de la aplicación: constituye el núcleo funcional del sistema. Se trata de un módulo personalizado (*add-on*) construido sobre el *framework* Odoo, que concentra las funcionalidades ya descritas. Este módulo materializa todos los pasos definidos en el proceso técnico, desde la configuración inicial (pasos 1 y 2), hasta la inserción y visualización de resultados (pasos 7 y 8), consolidando la lógica de ciclo completo del sistema.

Componente 6 - Módulo *RAG*: Finalmente, se integra un módulo específico para la aplicación de modelos de *RAG*, el cual opera como una capa semántica de análisis sobre los datos recolectados. Este componente permite: ejecutar procesos de clasificación temática y etiquetado automático; aplicar análisis de sentimiento a los textos recogidos; generar nubes de palabras y resúmenes asistidos por IA; relacionar los contenidos recolectados con repositorios internos o externos de conocimiento. Además, complementa y enriquece el proceso de obtención de resultados y aporta valor interpretativo a los datos mediante técnicas de NLP.

Procedimientos de *scraping* complementario y elaboración integral del *dataset*

Ahora bien, una limitación encontrada en la aplicación, para nuestra investigación, del *framework* desarrollado por CAETI radicó en la imposibilidad de extraer notas publicadas con cierta anterioridad a la puesta en funcionamiento de la herramienta, por lo que una parte importante de las unidades de análisis requeridas quedaron en principio inaccesibles. Por lo cual, y en función de resolver esta dificultad, se decidió implementar como complemento diversas técnicas de *scraping* con algoritmos adaptados *ad hoc* a cada portal, destacando el uso de las bibliotecas *Selenium* y *Newspaper3k*. Mediante *Selenium* se automatizó la navegación web y se simuló la interacción con las páginas como un usuario humano, mientras que *Newspaper3k* facilitó la extracción de texto relevante en las notas publicadas (como el título, bajada y cuerpo

del texto), entre otras cuestiones que son factibles de hacer con dicha librería. En este punto, cuando no fue posible recuperar las noticias desde los portales originales, se abordó el problema de dos maneras: a) accediendo a *Wayback Machine*¹¹ y, b) mediante búsquedas automatizadas en Google. Además, en esta instancia del trabajo investigativo se obtuvieron los comentarios a las noticias que otros usuarios-lectores realizaron en los portales de alcance local a través de Facebook y que estaban disponibles mediante incrustaciones *iframes*.¹²

La selección de las notas se presentó como un problema, dado que la temática que engloba al conjunto es demasiado específica. Si la búsqueda a desarrollar hubiera sido, por ejemplo, sobre secciones temáticas (política, deporte, economía, etcétera), habría resultado más sencillo diseñar un modelo de IA y NLP (supervisado) para seleccionar las notas, ya que es factible encontrar numerosos y extensos *sets* de datos etiquetados para entrenar y testear un modelo tal. Pero este no fue el caso.

Como se describe en la introducción de este capítulo, el caso es particular y, además, su recorte temporal resulta acotado. Tal es así, que hay noticias que se vinculan al narcotráfico en la ciudad de Rosario durante este periodo, involucran asesinatos, amenazas y, sin embargo, no son pertinentes al estudio. Además, la situación de emergencia que se vivenció en la ciudad fue tan transversal, que una noticia que sirviera al estudio del caso podría hallarse en cualquier sección temática de un portal.

Dado este escenario, se diseñó la siguiente estrategia de recolección suplementaria: a) Definimos sintagmas nominales probables de ser mencionados en los textos;¹³ b) Si ninguno de los sintagmas

11. Archivo digital que permite recuperar versiones anteriores de sitios web: <https://web.archive.org/>.

12. Los *inline frame* de una página web son elementos *HTML* que permiten insertar otro documento *HTML* dentro de la misma página.

13. Estos sintagmas fueron introducidos en la variable *sintagmas_clasificadores* = ["amenaza", "armadas", "armado", "armados", "asesinada", "asesinadas", "asesinado", "asesinados", "asesinato", "asesinatos", "ataque", "ataques", "baleada", "baleadas", "baleado", "baleados", "bullrich", "ciudadano", "civil", "cococcioni", "colectivero", "colectiveros", "colectivo", "colectivos", "conmoción", "crimen", "crímenes", "cárcel", "cárceles", "ejército", "estaciones", "estación", "fuerzas", "gendarmería", "javkin", "mafia", "mafioso", "milei", "narco", "narcos", "narcoterrorismo", "narcotráfico", "ola", "paquete", "penitenciario", "petri", "playero", "playeros", "policial", "policía", "policías", "pullaro", "rosario", "sangre", "seguridad",

aparecía en el texto (título, bajada, cuerpo del texto), ni tampoco en los resúmenes automáticos generados por *Newspaper3k* de cada nota, el texto sería descartado; c) Para los periódicos de alcance nacional (lanacion, clarin e infobae) ponderamos la aparición de la palabra “rosario” en la selección de la nota, criterio no aplicable para los portales locales (cuyas publicaciones aluden a la ciudad de distintos modos y no siempre la llaman por su nombre).

Puede advertirse que planteamos un *threshold* útil para rechazar las notas, pero no así para aceptarlas. Por ello, se efectuó un acercamiento exploratorio a lo recolectado para permitir el filtrado y la selección que, de esta manera, se hicieron de forma manual, esto es, sin aplicar procesamiento automático. Esta instancia de lectura convencional de las notas inicialmente relevadas hizo posible, además, identificar unidades informativas que no fueron captadas por la herramienta de CAETI ni por los códigos diseñados *ad hoc* para la recolección, pero que al ser detectadas pudieron luego procesarse y agregarse a este set de datos.

Una vez obtenidas las notas pertinentes para el caso, se recolectaron los comentarios a estas realizados vía Facebook, circunscribiéndolos a los publicados en portales locales de la ciudad de Rosario (lacapital, elciudadanoweb y rosario3). Dichos comentarios fueron recolectados de manera desatendida, es decir, se cargaron en el *dataset* bajo la pauta de estar relacionados a la nota. Lamentablemente, en el proceso se perdió información relevante según el tipo de análisis que se planifique, como la estructura dialogal de los comentarios o metadatos relacionados (*likes*, *replies* y tiempo transcurrido desde la publicación).

Por otra parte, al recolectar automáticamente las notas y comentarios obtuvimos información no deseada. Por ejemplo, en los comentarios de Facebook, “levantamos” publicidades (mayormente en inglés) o, desde el cuerpo del texto de los artículos y de modo variable, comentarios de la plataforma X, publicidades e hipervínculos a otras noticias. Ante la necesidad de limpiar el *set* de datos se procedió a detectar primero las estructuras repetidas que debían extraerse, explorando manualmente el *dataset*. Una vez identificadas estas, se diseñaron algo-

“servicio”, “sicario”, “sicarios”, “tachero”, “tacheros”, “taxi”, “taxis”, “taxista”, “taxistas”, “terror”, “terrorismo”, “violencia”].

ritmos que mediante expresiones regulares las sustrajeran del *corpus*.¹⁴ En las publicidades de Facebook se utilizó la biblioteca *Langdetect*, con la cual se identificaron y eliminaron las publicidades en inglés.

Vale aclarar que, debido a las particularidades de los portales, iteramos el flujo de trabajo descrito anteriormente por cada uno de ellos. Sin embargo, dada la heterogeneidad de la fuente de los datos (o sea, de los portales) y los distintos *approaches* adoptados en su recolección (por un lado la herramienta de CAETI y, por otro, los códigos generados especialmente para esta investigación), debimos normalizar los datos, garantizar su integridad (por ejemplo, quitar registros repetidos) y concatenarlos en un solo *dataset*. Dicho *set* contiene 596 registros (es decir, notas relacionadas con el caso de estudio) y siete columnas, las cuales son: 1. medio: contiene el nombre del portal fuente; 2. url: contiene el enlace a la nota original; 3. fecha: contiene la fecha de publicación en formato YYYY-MM-DD; 4. título: contiene el título original de la nota; 5. bajada: contiene el fragmento de texto publicado por el portal como bajada; 6. cuerpo_texto: contiene el cuerpo de texto principal de la nota y; 7. comentarios_fb: contiene en una lista los comentarios extraídos de la sección de Facebook de cada nota de medios locales, y en algunos registros esta columna contiene valores nulos (NaN).

Construcción de corpus y datasets de publicaciones en Facebook e Instagram vía Meta

Con el objetivo de construir *corpus* y *datasets* de publicaciones sobre el caso en cuentas de Facebook e Instagram, se procuró obtener acceso formal (mediante solicitudes gestionadas a través del Consorcio Interuniversitario para la Investigación Política y Social, ICPSR, de la Universidad de Michigan, y *Social Science One*, del Instituto de Ciencias Sociales Cuantitativas de Harvard) a dos herramientas y a un conjunto de datos de *Meta Platforms*. Se trata, por un lado, de la biblioteca de contenido

14. Por ejemplo, en los textos de lacapital aparecen a menudo conexiones a otras noticias que comienzan con la expresión >> Leer más: , seguida de un hipervínculo y que se encuentra entre saltos de línea (\n). Entonces, la lógica a seguir para este sitio web fue quitar del *dataset* todo lo que comience con >> Leer más: , situado entre saltos de líneas.

(*Meta Content Library*) y de la API (*Application Programming Interface*) de la biblioteca de contenido y, por otro, del conjunto de enlaces compartidos en Facebook (*URL Shares*).¹⁵ A partir de la autorización obtenida para el trabajo con estos recursos, se llevó a cabo la recolección del segundo y tercer *dataset* de nuestra investigación.

La biblioteca de contenido y la API están alojadas en el entorno virtual de *Social Media Archive* (SOMAR) del ICPSR. Estas herramientas permiten realizar consultas automatizadas sobre el archivo de contenido público de Facebook (desde sus inicios) e Instagram¹⁶ y extraer información detallada sobre páginas, grupos, eventos y perfiles públicos, ya sean personales o comerciales, con más de 25,000 seguidores o con la insignia de cuenta verificada. Asimismo, facilita la identificación de la autoría de las publicaciones, el número de reacciones y las veces que se ha compartido un contenido.

La biblioteca de contenido, por su parte, opera como un entorno de acceso controlado basado en la web, mientras que la API habilita a realizar consultas programáticas utilizando lenguajes como R o Python. La primera, además, facilita la ejecución de búsquedas, con un límite de 100,000 registros, aplicando filtros predeterminados y posibilitando la descarga de datos en formato .csv (*Comma-Separated Values*), con un máximo de diez archivos por día. En cambio, la API proporciona un volumen de datos considerablemente mayor y, principalmente, permite extraer comentarios públicos, función no disponible en la biblioteca.

Hasta el momento de escribir este capítulo, hemos trabajado con el entorno virtual de acceso controlado de la biblioteca de contenido. Esta herramienta cuenta con la opción de crear listas de productores (*Producer lists*), colecciones personalizables de perfiles o páginas en las que se pueden realizar búsquedas, extraer la información en .csv y trabajar de manera colaborativa entre los investigadores que cuentan con autorización de acceso. Para organizar mejor el análisis, se generaron conjuntos diferenciados de las cuentas a analizar según plataforma: por un lado, las de Facebook y, por otro, las de Instagram. De esta manera, se clasificaron según el tipo de fuente, generando grupos espe-

15. <https://socialscience.one/rfps>

16. Actualmente, también permite el acceso a contenido de Threads, aunque dicha plataforma no es de interés para nuestro estudio.

cíficos para los portales locales de noticias (La Capital, El Ciudadano y La Región y Rosario³), los portales nacionales (Infobae, La Nación, Clarín), así como los perfiles de un conjunto de funcionarios públicos. Sobre estos últimos, vale aclarar, a escala local la indagación se circunscribió al intendente de la ciudad de Rosario, Pablo Javkin, al gobernador de la provincia de Santa Fe, Maximiliano Pullaro, y al ministro de Seguridad y Justicia de la provincia de Santa Fe, Pablo Cococcioni. Por su parte, como referentes de la escena política nacional vinculados al caso, se consideraron las publicaciones en las cuentas del presidente de la nación, Javier Milei, de la ministra de Seguridad, Patricia Bullrich, y del ministro de Defensa, Luis Petri.

Una vez agrupadas las cuentas a analizar en distintas listas de productores diferencias por plataforma, se realizó la búsqueda de publicaciones asociadas al caso de estudio. Los filtros aplicados incluyeron el recorte temporal (del 2 de marzo al 6 de mayo de 2024) y operadores booleanos para la identificación, dentro del contenido de la publicación, de los siguientes sintagmas: narco, droga, penitenciario, arma, terror, crimen, asesina, balea, amenaza, violencia, crímenes, ejército, sicario y fuerzas. Tras completar la búsqueda por cada lista, se obtuvieron 3,892 registros (120 de los perfiles de figuras políticas en Facebook y 123 de Instagram; 1,515 de las cuentas de los portales de noticias locales en Facebook y 597 de Instagram; y, por último, 1,188 de los portales de noticias nacionales en Facebook y 349 de Instagram) y se exportaron en documentos .csv.¹⁷ Posteriormente, para llevar a cabo un acercamiento inicial a los primeros datos capturados, se procedió a trabajar con cada archivo descargado. En primer lugar, se unificó el formato de fecha y hora en GMT (*Greenwich Mean Time*) ya que se encontraban en distintos sistemas de coordenadas: Unix y *Universal Time Coordinated*, UTC. Además, se eliminaron algunas variables que no resultaban significativas y, asimismo, se incorporaron otras *ad hoc*. Luego, se filtraron y eliminaron, manualmente, aquellas publicaciones que no estaban vinculadas con el caso de estudio. Por último, se inte-

17. Es importante señalar que estos paquetes de registros permanecen en el entorno virtual por treinta días y luego expiran, al igual que los enlaces que llevan al contenido multimedia que posee cada publicación. Por tanto, se vuelve fundamental para la investigación implementar estrategias de resguardo y respaldo de la información en un soporte autónomo y estable por fuera del entorno web de la herramienta.

graron las publicaciones de todas las cuentas en dos bases de datos organizadas según plataforma: una de Instagram con 461 registros y otra de Facebook con 985 unidades.

Con los dos *datasets* completos, los próximos pasos consisten en, por un lado, realizar análisis descriptivos y comparativos de las entradas de cada plataforma; esto es, verificar, por ejemplo, si los contenidos compartidos en Instagram y en Facebook son los mismos por cada una de las cuentas estudiadas. Por otro lado, se procederá a comparar los registros obtenidos de los periódicos locales y nacionales con la búsqueda realizada mediante *web scraping* y desarrollada en el apartado anterior de este capítulo.

Asimismo, y como parte de los beneficios que posibilita el acceso formal a Meta, se recabarán mediante la API los comentarios correspondientes a las publicaciones relevadas. Es menester recalcar que, por la política de privacidad de Meta, los comentarios que se podrán obtener pertenecen, únicamente, a aquellas cuentas públicas.

Por otro lado, a través de *Social Science One*, del Instituto de Ciencias Sociales Cuantitativas de Harvard, parte del equipo obtuvo el permiso para trabajar con el conjunto de enlaces compartidos más de 100 veces públicamente en Facebook, conocido como *URL Shares*. Aunque aún no se ha iniciado el análisis de estos datos, entendemos que nos permitirán identificar los enlaces relacionados con narcoterrorismo que circularon en Facebook durante el periodo temporal delimitado por nuestra investigación.

Como puede imaginarse, el uso de estas herramientas informáticas conlleva tanto potencialidades como limitaciones. Entre los aspectos favorables, se destaca la posibilidad de obtener publicaciones de manera retrospectiva, así como de hacernos del material multimedia y los comentarios de los usuarios de las distintas cuentas públicas, trabajando de manera automatizada en la recolección y análisis de datos. Además, las plataformas provistas por Meta permiten la colaboración dentro del equipo de investigación y cuentan con asistencia ágil por parte de las instituciones que regulan su acceso. Sin embargo, se observa que el proceso de solicitud ha sido altamente burocrático, con requisitos como la radicación del equipo en una misma institución académica y la necesidad de avales legales, lo cual inevitablemente ralentiza la dinámica de trabajo y posterga la labor concreta con las herramientas informáticas en cuestión. Otro factor que puede repre-

sentar una barrera adicional, es que la documentación, las actualizaciones y las capacitaciones impartidas por las organizaciones intermedias han sido diseñadas exclusivamente en inglés. Finalmente, una de las limitaciones más relevantes que hemos registrado por el momento radica en la imposibilidad de difundir y publicar los datos obtenidos, así como las visualizaciones derivadas de ellos, sin contar con el permiso expreso y la autorización de Meta. Estas restricciones entran, en parte, en tensión con los principios de la ciencia abierta y el acceso libre a los datos, generando un dilema entre la necesidad de compartir información o hallazgos para avanzar de modo comunitario en la producción de conocimiento científico y la protección de determinados intereses comerciales o legales de las plataformas en cuestión.

Captura mixta y sistematización de *corpus* y *datasets* de publicaciones en X

Acceder a los datos suministrados por X representa ciertamente algunos desafíos importantes. Twitter fue una de las primeras plataformas en implementar una política de datos abiertos a la comunidad científica, poniendo a disposición una API con la cual acceder a determinados datos en caso de justificar esa recolección en el marco de una investigación respaldada. Incluso, Twitter mantuvo dicha posibilidad, aún después del escándalo de *Cambridge Analytica* que empujó al entonces Facebook (ahora Meta) a cerrar el ingreso a su API. No obstante, esta lógica se vio alterada radicalmente a partir de la adquisición de la plataforma por parte del empresario Elon Musk, quien en 2023 cambió el nombre de Twitter por X y suspendió la recolección de datos a todas las cuentas autorizadas para investigación, que se vieron totalmente restringidas para descargar datos. En este marco, y ante la necesidad de construir un *corpus* con publicaciones en X, se optó por mixturar técnicas de recolección de datos tradicionales con otras computarizadas.

Recolección automatizada de tweets

En primer lugar, se utilizó *Twscrape*, una librería de *scraping web* desarrollada por terceros en Python.¹⁸ Esta permitió efectuar la extracción desatendida de *tweets* (y sortear las mencionadas limitaciones de acceso), empleando como parámetros tanto la identificación de determinadas cuentas como el periodo temporal de estudio.

Twscrape no usa la API oficial de X; en lugar de eso, efectúa un *web scraping* avanzado, simulando el comportamiento de un cliente web y haciendo ingeniería inversa sobre los *endpoints* internos de la API *GraphQL* que la plataforma reserva para su propio uso web. Estos *endpoints* tienen la ventaja de que, al ser parte de una API *GraphQL*, sirven para consultas complejas (más flexible que REST). Se obtuvo así un *dataset* inicial con 21,675 publicaciones, colección de ensayo que comprendió los *tweets* publicados por cuentas¹⁹ que consideramos relevantes para nuestro estudio. Luego, se aplicó sobre ese conjunto un filtro de pertinencia temática (tabla 1) diseñado sobre dos criterios: i) identificación de sintagmas clave y, ii) asignación de pesos diferenciales según su capacidad discriminante,²⁰ definidos a partir de expresiones significativas del periodo observado. Para cada publicación se sumaron los pesos de los sintagmas presentes; aquellas con puntaje mayor o igual a 3 se consideraron pertinentes. Con el fin de reducir falsos positivos, se incorporaron pesos negativos (-3) para expresiones frecuentes que desviaban el foco del fenómeno. Este procedimiento

18. <https://github.com/vladkens/twscrape>

19. De acuerdo con la documentación de *Twscrape*, los métodos de cronología de usuario (`user_tweets` y `user_tweets_and_replies`) recuperan como máximo ~3,200 tweets por cuenta. Por eso en nuestro estudio no utilizamos esos métodos, sino búsquedas parametrizadas por cuenta y ventana temporal (con `since:/until:`) en la pestaña Latest, elevando el parámetro limit para cubrir íntegramente el periodo de referencia. De este modo, para cada cuenta de interés (@maxipullaro, @pablojavkin, @PatoBullrich, @JMLEi, @luispetri, @PabloCocoX @Rosario3, @elCiudadanoWeb, @laCapital) se recolectaron exclusivamente las publicaciones comprendidas entre el 2 de marzo y el 6 de mayo de 2024, y no los últimos 3,200 de su cronología.

20. Este se encuentra asociado a la medida en que se estimó que la sola presencia del sintagma resulta suficiente (o insuficiente) para considerar pertinente una publicación. De modo que expresiones con alta especificidad temática, capaces por sí solas de indicar relevancia, recibieron un peso de 4; otras, con menor capacidad de discriminación, fueron ponderadas con valores de 3, 2 o 1.

automático se complementó con una revisión manual, destinada a validar la pertinencia y corregir sesgos del umbral, que incluyó: a) la evaluación de casos fronterizos y, b) la supervisión específica de *tweets* con menos de tres palabras. El proceso produjo un conjunto final de 350 publicaciones.

Tabla 1. Criterio de relevancia temática aplicado como filtro

Peso asignado	Sintagmas clave y su peso en el análisis
	Sintagmas
4	raul justino pereyra, raul pereyra, david joel pereyra, narcosario, narcotraficorosario, narcoestado, chucky monedita, lindor alvarado, claudio morocho, claudio mansilla, morocho mansilla
3	narcoterrorismo, movi taxi
2	narco, narcotrafico, sicario, narcocriminal, recompensa, amenaza, taxista, taxi, tachero
1	rosario, pullaro, droga, crisis, bullrich, ataque, armado, terrorismo, terror, crimen, maxipullaro, javkin, pablojavkin_x, petri, servicio penitenciario, imputabilidad, violencia, ola de violencia (incluye oladeviolencia, olaviolenta)
-3	hezbollah, israel, paritarias

Fuente: Elaboración propia.

Recolección artesanal de tweets

Durante el proceso de recolección de datos hubo un conjunto de publicaciones que no pudo ser obtenido de manera automatizada, debido a una falla en la librería *Twscrape*. Al operar sin claves oficiales de API, *Twscrape* depende de *endpoints* que pueden modificarse en cualquier momento y sobre los cuales los desarrolladores de la librería no tienen control. Como consecuencia, su funcionamiento puede verse interrumpido sin previo aviso, lo que efectivamente ocurrió durante el proceso que estamos describiendo. Las publicaciones afec-

tadas correspondían a cuentas de especial interés para nuestro estudio: @clarincom, @LANACION y @infobae.

Ante la imposibilidad de resolver dicho error técnico en el tiempo disponible y dada la falta de recursos para explorar alternativas automatizadas más complejas, se optó por desarrollar en este punto una estrategia de recolección artesanal de los datos. Para ello, se diseñó un esquema de procedimientos para aprovechar la sintaxis de búsqueda avanzada de la propia plataforma X. Esta decisión permitió evitar la operación más rudimentaria del *scrolling* manual (esto es, navegar publicación por publicación en cada perfil requerido), y facilitó la ejecución de consultas precisas que cumplieran con un grupo de principios (tabla 2) orientados a la eficiencia y organización del *corpus*. A saber:

- *Filtrado progresivo*: Cada término de búsqueda adiciona una capa de exclusión de términos previamente consultados para evitar la superposición y asegurar que los *tweets* recopilados estén claramente segmentados por tema.
- *Especificidad de fuente*: Las consultas están diseñadas para recolectar *tweets* específicos de cuentas de interés (Clarín, La Nación, Infobae).
- *Temporalidad*: Todas las consultas están limitadas al rango de fechas buscado, lo que sugiere un interés por estudiar estos temas en un contexto temporalmente relevante y limitado.
- *Incremento de la especificidad*: A medida que avanzamos en la búsqueda, cada término nuevo excluye todos los términos anteriores para evitar la redundancia en la captura del dato.

Tabla 2. Ejemplos de *Querys* utilizados para la búsqueda artesanal en X²¹

Esquema de búsqueda progresiva en X para la recolección manual de publicaciones relevantes			
Palabras a buscar	Query en X - Clarín	Query en X - La Nación	Query en X - Infobae
Rosario	from:@clarincom "Rosario" since:2024-03-01 until:2024-05-07	from:@LANACION "Rosario" since:2024-03-01 until:2024-05-07	from:@infobae "Rosario" since:2024-03-01 until:2024-05-07
ataque	from:@clarincom "ataque" -"Rosario" since:2024-03-01 until:2024-05-07	from:@LANACION "ataque" -"Rosario" since:2024-03-01 until:2024-05-07	from:@infobae "ataque" -"Rosario" since:2024-03-01 until:2024-05-07
armado	from:@clarincom "armado" -"Rosario" -"ataque" since:2024-03-01 until:2024-05-07	from:@LANACION "armado" -"Rosario" -"ataque" since:2024-03-01 until:2024-05-07	from:@infobae "armado" -"Rosario" -"ataque" since:2024-03-01 until:2024-05-07
narcoterrorismo	from:@clarincom "narcoterrorismo" -"Rosario" -"ataque" -"armado" since:2024-03-01 until:2024-05-07	from:@LANACION "narcoterrorismo" -"Rosario" -"ataque" -"armado" since:2024-03-01 until:2024-05-07	from:@infobae "narcoterrorismo" -"Rosario" -"ataque" -"armado" since:2024-03-01 until:2024-05-07
terrorismo	from:@clarincom "terrorismo" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" since:2024-03-01 until:2024-05-07	from:@LANACION "terrorismo" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" since:2024-03-01 until:2024-05-07	from:@infobae "terrorismo" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" since:2024-03-01 until:2024-05-07

21. El presente cuadro es un ejemplo parcial que forma parte de otra tabla más amplia que reúne la totalidad de los *querys* utilizados durante la etapa de recolección de datos.

Esquema de búsqueda progresiva en X para la recolección manual de publicaciones relevantes			
Palabras a buscar	Query en X - Clarín	Query en X - La Nación	Query en X - Infobae
terror	from:@clarincom "terror" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" since:2024-03-01 until:2024-05-07	from:@LANACION "terror" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" since:2024-03-01 until:2024-05-07	from:@infobae "terror" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" since:2024-03-01 until:2024-05-07
narcotráfico	from:@clarincom "narcotráfico" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" -"terror" since:2024-03-01 until:2024-05-07	from:@LANACION "narcotráfico" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" -"terror" since:2024-03-01 until:2024-05-07	from:@infobae "narcotráfico" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" -"terror" since:2024-03-01 until:2024-05-07
narco	from:@clarincom "narco" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" -"terror" -"narcotráfico" since:2024-03-01 until:2024-05-07	from:@LANACION "narco" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" -"terror" -"narcotráfico" since:2024-03-01 until:2024-05-07	from:@infobae "narco" -"Rosario" -"ataque" -"armado" -"narcoterrorismo" -"terrorismo" -"terror" -"narcotráfico" since:2024-03-01 until:2024-05-07

Fuente: Elaboración propia.

Este enfoque, si bien más laborioso que una recolección automatizada, ofreció un mayor control cualitativo sobre el contenido recuperado, lo cual resultó crucial dada la necesidad de precisión en la construcción del *corpus* analítico. Además, representa un ejemplo de cómo el conocimiento de las estructuras técnicas de las plataformas puede compensar, parcialmente, la falta de herramientas y técnicas informáticas mediante soluciones más accesibles en términos de recursos. Gracias a este procedimiento, se logró agregar a la base otras 117 publicaciones adicionales, por lo que el *dataset* final por X comprende un total de 467 registros.

Algunas reflexiones de cierre

Compartimos con Kataishi y Milia (2024) que “los procesos de digitalización afectan qué y cómo investigamos” (p. 270), y los estudios de corte semiótico (o sociosemiótico) no tienen por qué ser una excepción a ello. Como se deduce de lo presentado a lo largo de estas páginas, nuestra investigación está aún lejos de concluir, y tan solo hemos finalizado la instancia de construcción y preprocesamiento de *corpus* y *datasets*. Por ello, lo que compartimos no son hallazgos ni resultados de nuestra observación, sino la experiencia adquirida hasta el momento en tren de aplicar una estrategia metodológica combinada (denominada como semidata) que permita optimizar el estudio empírico de la producción de sentido mediatizada en plataformas con base en Internet. Embarcados en esa decisión, nos proponemos integrar subsidiariamente las técnicas y métodos computacionales (ya sea de base estadística, matemática o informática) al análisis sociosemiótico, lo cual deriva, en cierta medida, en la experimentación de procedimientos que hagan viable el manejo de paquetes textuales de cierta magnitud.

Los *sets* de datos compendiados permiten contar con acceso a los textos datificados (además de sistematizar los enlaces a los discursos online) y eso tiene al menos dos implicancias directas en nuestra investigación. En primer lugar, disminuye las “chances” de perder con el tiempo información relevante para la investigación, algo muy frecuente cuando se estudia el presente en transición (Raimondo Anselmino, 2013; Gindin y Busso, 2018). En segundo lugar, propicia la combinación de diferentes niveles de observación que hacen asequible el estudio intermedial y el trabajo con diversas fuentes de información disponibles en Internet y distintos tipos de datos.

Al mismo tiempo, esta posición que hemos adoptado remodela, en parte, las dinámicas y flujos de trabajo en el campo de la investigación social actual, así como demanda el desarrollo de una cultura de la investigación interdisciplinaria. En el contexto actual, atravesado por una política de desmantelamiento de las capacidades científicas estatales en Argentina, resulta imposible soslayar las dificultades materiales que enfrentamos quienes investigamos en el país. Aun así, insistimos en sostener formas de producción de conocimiento colaborativas y

situadas, que dan cuenta no solo de los límites, sino también de la potencia crítica del trabajo académico en nuestra región.

Referencias

- Aldas, J., y Uriel, E. (2017). *Análisis multivariante aplicado con R*. Paraninfo.
- Carlón, M. (2020). *Circulación del sentido y construcción de colectivos en una sociedad hipermediatizada*. Nueva Editorial Universitaria.
- Fernández, J. (2018). *Plataformas mediáticas. Elementos de análisis y diseño de nuevas experiencias*. La Crujía.
- Fratlicelli, D. (2023). *El humor hipermediático. Una nueva era de la mediatización reidera*. Teseo.
- Gindin, I., y Busso, M. (2018). Investigaciones en comunicación en tiempos de *big data*: sobre metodologías y temporalidades en el abordaje de redes sociales. *adComunica. Revista Científica de Estrategias, Tendencias e Innovación en Comunicación*, (15), 25-43. <https://www.raco.cat/index.php/adComunica/article/view/343017>
- Gindin, I., y Raimondo Anselmino, N. (2024). Mediatizaciones del discurso sobre lo público-común platformizado: estudio interdisciplinario del “narcoterrorismo” en la ciudad de Rosario (Argentina). En I. Gindin y M. C. Reviglio, *La mediatización insomne. 40 años de sueños y pesadillas democráticas* (pp. 78-96). UNR Editora. <https://cim.unr.edu.ar/publicaciones/1/libros/181/la-mediatizacion-insomne-40-anos-de-suenos-y-pesadillas-democraticas>
- Kataishi, R., y Milia, M. (2024). El análisis semántico-estadístico como estrategia de abordaje metodológico: reflexiones sobre su pertinencia en el estudio de problemáticas latinoamericanas. En J. M. Natera y D. Suárez (Comps.), *Métodos para el análisis de los procesos de ciencia, tecnología e innovación: herramientas para el estudio del desarrollo de América Latina. Métodos mixtos y emergentes*, vol. 3. Ediciones UNGS.
- Kerbrat-Orecchioni, K. (1996). *La conversation*. Seuil. Traducción al español: Preiti, M. Cátedra de Lengua Española II, Facultad de Humanidades y Artes, UNR.
- Mantilla-Valbuena, S. (2008). Más allá del discurso hegemónico: narcotráfico, terrorismo y narcoterrorismo en la era del miedo y la inseguridad global. *Papel Político*, 13(1), 227-259.
- Marradi, A., Archenti, N., y Piovani, J. I. (2018). *Manual de metodología de las ciencias sociales*. Siglo XXI.

- Martín-Crespo Blanco, C., y Salamanca Castro, A. (2007). El muestreo en la investigación cualitativa. *NURE Investigación: Revista Científica de Enfermería*, (27).
- Montes del Castillo, Á., y Montes Martínez, A. (2014). Guía para proyectos de investigación. *Universitas*, 12(20), 91-126.
- Raimondo Anselmino, N. (2011). O ocaso do modelo intencional: a noção de “estratégia discursiva” sob o olhar sócio-semiótico. *Semeiosis: Semiótica e Transdisciplinaridade em Revista*, 1(2). <https://semeiosis.com.br/issues?issue=YEvsnsst2sPcL29M5KWG&article=EaGHosEF8L5DVptfXaWC>
- (2013). Un análisis sociosemiótico de la prensa *online*: investigar el presente en transición. En N. Raimondo Anselmino y M. C. Reviglio (Eds.), *Territorios de comunicación. Recorridos de investigación para abordar un campo heterogéneo*. Quipus, Ciespal.
- (25 de julio de 2022). *Semiodata: una estrategia combinada para el estudio de la producción de sentido en el estadio actual de la mediatización*. Conferencia realizada en el marco del Seminario “Arqueologías de la mediatización. Tiempos, espacios y tecnologías del mundo actual”. <https://www.youtube.com/watch?v=7rP8k6YpHvo>
- Raimondo Anselmino, N., Cardoso, A. L., Rostagno, J., y Sambrana, A. (2019). Recursos paratextuales y paralingüísticos en las *fanpages* de los periódicos argentinos *Clarín* y *La Nación*. Atributos del discurso de la prensa en las redes. *Perspectivas de la Comunicación*, 2(2), 245-280. <https://www.perspectivasdelacomunicacion.cl/index.php/perspectivas/article/view/2028>
- Raimondo Anselmino, N., Reviglio, M. C., y Diviani, R. (2015). Esfera pública y redes sociales en Internet: ¿Qué es lo nuevo en Facebook? *Revista Mediterránea de Comunicación*, 7(1).
- van Dijck, J., Poell, T., y de Waal, M. (2018). *The Platform Society. Public values in a connective world*. Oxford University Press.
- Verón, E. (1998). *La semiosis social. Fragmentos de una teoría de la discursividad*. Gedisa.
- (2013). *La semiosis social, 2. Ideas, momentos, interpretantes*. Paidós.

Capítulo 6

Factores psicosociales en la dinámica de la calidad de vida. Análisis Cualitativo Comparado con Conjuntos Difusos (fs/QCA)

Gabriela Sánchez Mondragón¹

Anthony Pérez Balcázar²

Introducción

La calidad de vida es un concepto que ha trascendido lo que significa vivir bien, ya que no solo se mide por la posesión de bienes materiales o acceso a recursos; además, involucra una experiencia profunda que integra tanto las condiciones objetivas como las percepciones subjetivas de bienestar (Cummins y Cahill, 2000). Esto representa que las variables objetivas como la salud, la educación, la seguridad y la prosperidad económica influyen de manera tangible en la forma en la que vive un ser humano, pero es fundamental considerar la dimensión interna (procesos cognitivos y afectivos) que contribuyen en la percepción de sentirse pleno, conectado y satisfecho con las condiciones en las que se vive. Así, este concepto adquiere

1. Universidad Intercultural del Estado de México (Plantel Tepetlixpa). Correo electrónico: gabriela.sanchez@uiem.edu.mx. orcid: <https://orcid.org/0000-0001-5732-0070>

2. Universidad Intercultural del Estado de México (Plantel Tepetlixpa). Correo electrónico: anthony.perez@uiem.edu.mx. orcid: <https://orcid.org/0000-0002-5845-7834>

una naturaleza compleja y dinámica, pues depende no solo de lo que poseemos, sino también de cómo lo vivimos y lo interpretamos.

Desde sus orígenes en la economía del bienestar posterior a la Segunda Guerra Mundial (Tonon, 2010; Salas y Garzón, 2013), la calidad de vida como sinónimo de bienestar estaba representada por el crecimiento económico, el consumo y la acumulación de bienes; pero hasta los años 80, el término se consolida como un concepto multidimensional que integra dimensiones objetivas como ingresos, salud, educación, vivienda y seguridad y subjetivas como la satisfacción personal, la felicidad y el sentido de vida.

Urzúa (2012), así como Salas y Garzón (2013), convergen en que uno de los primeros modelos conceptuales en proponer una visión integradora de la calidad de vida fue el propuesto por Felce y Perry en 1995, al especificar que la calidad de vida se integraba por dimensiones como la satisfacción personal y las condiciones de vida.

En este contexto, el análisis de la calidad de vida constituye un campo multidisciplinario que articula enfoques psicológicos, sociales, filosóficos y políticos, que explora las interacciones entre las estructuras que configuran la experiencia humana. Las definiciones convergen en su naturaleza multidimensional integrada por factores que promueven en el individuo un estado de satisfacción general derivado de la realización de sus potencialidades o percepción de estas (Ardila, 2003; Urzúa y Caqueo-Urizar, 2012; Nicolás *et al.*, 2022).

La satisfacción de las necesidades personales es fundamental para el desarrollo de las esferas vitales. Por tanto, la calidad de vida se configura como un proceso de evaluación que está condicionado por factores personales y políticos, sociales, culturales, ambientales y económicos. Es decir, no basta con que una persona tenga o perciba como suficientes los recursos de que dispone o a que puede acceder, sino que representa un asunto de las políticas sociales, lo que le adhiere componentes políticos, gubernamentales, económicos y culturales.

Distintos autores han realizado aportaciones destacando la multidimensionalidad del concepto en cuestión. Palomba (2002) refirió que está vinculado estrechamente a las políticas sociales, el cual abarca tanto condiciones de vida objetivas como un alto nivel de bienestar subjetivo. Distingue dos dimensiones centrales: una evaluación objetiva del nivel de vida mediante indicadores concretos y la percepción subjetiva del individuo, comúnmente asociada al bienestar (*wellbeing*).

Esto representa que la satisfacción personal no se limita al plano individual, sino que además está mediatizada por las condiciones sociales y políticas que estructuran las posibilidades de desarrollo humano.

Desde esta perspectiva, la calidad de vida integra aspectos materiales y no materiales, objetivos y subjetivos, individuales y colectivos. En cuanto a los recursos materiales, se refiere a elementos como el ingreso, la posición en el mercado laboral, el estado de salud y el nivel educativo; mientras que los factores ambientales consideran el acceso a servicios, la seguridad, la infraestructura urbana y la tecnología, todos ellos determinantes del bienestar. En cuanto a la dimensión subjetiva, aspectos como los vínculos sociales, el apoyo comunitario, la participación en organizaciones y el uso del tiempo libre desempeñan un rol fundamental en la percepción de una vida satisfactoria. Es evidente que desde el enfoque expuesto por Palomba (2002) las políticas gubernamentales son un componente estructural en la configuración de la calidad de vida, por lo que en su perspectiva se resalta la acción del Estado y el diseño de políticas públicas para el bienestar personal y colectivo.

Cabe señalar que las políticas públicas, entendidas como respuestas estatales frente a problemas sociales, deben contemplar no solo los indicadores cuantitativos tradicionales —que miden condiciones estructurales como salud, educación o vivienda—, sino también indicadores cualitativos que reflejen las experiencias y valoraciones de los ciudadanos (Tonon, 2010). Los indicadores exclusivamente objetivos, al generalizar datos sin atender a los contextos históricos y socioculturales, corren el riesgo de distorsionar la realidad o invisibilizar desigualdades. Por ello, se requiere incorporar la percepción de los individuos que integran comunidades, con el objetivo de garantizar decisiones más pertinentes, justas y eficaces, capaces de atender tanto las condiciones externas de vida, como al bienestar percibido por las personas.

Por otra parte, Schalock y Verdugo (2012) identificaron ocho dimensiones donde destacan el bienestar emocional, relaciones interpersonales, bienestar material, desarrollo personal, bienestar físico, autodeterminación, inclusión social y derechos como los principales componentes de la calidad de vida. Su aporte estuvo enfocado principalmente en comprender la calidad de vida con miras a crear sistemas de medición válidos.

Urzúa y Caqueo-Urizar (2012) mencionaron que la calidad de vida está relacionada con las condiciones en las que vive una persona

y que son objetivamente medibles, como la salud física, condiciones de vida en general, relaciones sociales, actividades funcionales u ocupación; mientras que la dimensión subjetiva está asociada a elementos subjetivo-valorativos que se manifiestan en la satisfacción con la vida, satisfacción personal (en algunos casos percibido como felicidad).

Estos antecedentes conceptuales ponen de manifiesto que la calidad de vida es un concepto multidimensional, dinámico y relacional; que articula tanto elementos objetivos (acceso a salud, educación, seguridad, ingresos, etcétera) como elementos subjetivos (percepción de bienestar, satisfacción vital y realización personal). Además, subrayan la influencia de factores estructurales (políticas públicas, equidad social, inclusión) en la construcción del bienestar.

Estas definiciones coinciden en que el bienestar individual depende tanto de las condiciones materiales disponibles que se concentran como *condiciones de vida*, como de la valoración personal de dichas condiciones. En consecuencia, la calidad de vida se configura como una construcción compleja, situada en la intersección entre lo individual y lo colectivo. Por ello, es necesario identificar cuáles son los factores suficientes o mínimos necesarios para que una persona pueda percibir bienestar desde una perspectiva psicosocial (Urzúa, Cortés y Pino, 2011; Varela, 2017).

La relevancia de conocer estos elementos es amplia, para este caso en particular, llama la atención el indagar en las condiciones de necesidad y/o de suficiencia en la calidad de vida de la población en 38 países con el propósito de detectar cuáles son las condiciones que propician calidad de vida actualmente en dichos países, y con ello poder indagar sobre algunos factores que podrían ser replicados a través de la propuesta de políticas públicas.

Para ello, se utilizó el Análisis Cualitativo Comparado con Conjuntos Difusos (fs/QCA) como herramienta de análisis, identificando, así, tres tipos de soluciones de concurrencia; parsimoniosa, compleja e intermedia. El objetivo no solo fue aplicar un análisis descriptivo-comparativo, sino también una comparación que permitiera realizar inferencias científicas más rigurosas, confiables y válidas sobre la forma en la que interactúan indicadores objetivos y subjetivos asociados a la calidad de vida. Esto es, de una estrategia de análisis de causalidad coyuntural entre casos similares y diferentes, contruidos bajo tipologías y normas lógicas.

Así bien, el análisis en los países seleccionados sostuvo que la calidad de vida de una población se debe a factores causales altamente correlacionados asociados a componentes mentales y sociales, es decir, de factores psicosociales. Por ello, el trabajo se centró en dos hipótesis plausibles:

- H1: *Los derechos universales, la salud y la seguridad son condiciones necesarias, pero no suficientes para la calidad de vida de una población.*
- H2: *El desarrollo humano, la felicidad y los ingresos económicos son condiciones suficientes para incrementar la calidad de vida de una población.*

Metodología

Es importante mencionar que la base metodológica que aquí se describe no profundiza en la base matemática de análisis. Se describen cualitativamente las bases principales que permiten construir dicho análisis y los elementos que estructuran la discusión de los resultados.

Así bien, el Análisis Cualitativo Comparado (QCA, por sus siglas en inglés) es una herramienta que permite generar inferencias científicas con mayor precisión y validez, esto es, del análisis y explicación de sucesos causales multivariados. QCA permite comparar una gran cantidad de casos (países, personas, grupos, etcétera) cuyos resultados (similitudes y diferencias) son la construcción de soluciones de concurrencia (parsimoniosa, compleja e intermedia) basados en dos condiciones principales: condiciones de necesidad y condiciones de suficiencia.

Para una solución de concurrencia (μ), es indispensable que la condición α esté presente, en caso contrario no es posible tal resultado, esto es:

$$\alpha \Rightarrow \mu$$

donde α es la condición de necesidad y μ es la solución de concurrencia.

Para el caso de las condiciones de suficiencia, basta con que α , β y γ estén presentes para la solución de concurrencia μ , es decir:

$$\alpha * \beta * \gamma \Rightarrow \mu$$

donde, α , β , γ son condiciones y μ es la solución de concurrencia.

Dentro de las ciencias sociales, el uso de QCA ha permitido realizar análisis con variables cuyos valores numéricos permiten calcular grados de pertenencias a ciertos conjuntos [A, B, o C] de casos. Regularmente, en las ciencias sociales estos valores numéricos son puntuaciones parciales o bien difusos, esto es, de puntuaciones con valores decimales que pueden estar, por ejemplo, entre [0] y [1], permitiendo que las variables se relacionen entre sí, expresados en términos de suficiencia o necesidad, ya sea de manera causal o descriptiva (cfr. Schneider, 2012, pp. 2-4). En este sentido, QCA es una herramienta sumamente fuerte cuando se relacionan fenómenos sociales de carácter causal, revelando patrones de asociación entre casos.

Los casos pueden ser catalogados como: personas, países, ciudades, o bien algún elemento susceptible de comparar. Entonces, tenemos tres componentes importantes para poder realizar un análisis difuso con el uso de QCA: casos, variables y puntuaciones numéricas. En relación con este trabajo, los casos son países de la Organización para la Cooperación y el Desarrollo Económico (OCDE), cuyas variables están estrechamente relacionadas con un proceso psicosocial que en su condición de causalidad mejora la calidad de vida de la población de un país. Para este análisis, se utilizan:

Valores numéricos tomados de los índices³ *Global Social Progress Index*, *The Better Policies for Better Lives* de la OCDE, índice de calidad de vida, índice de desarrollo humano y *World Happiness Report* (2023), y datos del PIB per cápita del Banco Mundial, los cuales fungen como factores o condiciones psicosociales para mejorar la calidad de vida.

El método directo de calibración de RAGIN para la generación de membresía de pertenencia o no pertenencia a dicho subconjunto.⁴

Se presenta un histograma denominado *XY-plot* que muestra la correlación entre los factores psicosociales y su condición de necesidad para la calidad de vida, es una condición necesaria cuando su consistencia se encuentre por encima de 0.85, esto es, que entre más

3. Se usaron los índices actualizados a septiembre de 2024.

4. Para profundizar sobre el método de calibración véase: Ragin, C. (2008). *Resigning Social Inquiry. Fuzzy Set and Beyond*. The University of Chicago Press.

cercano a [1] mayor es su consistencia, en caso contrario la consistencia es no-observada. La consistencia mide el grado en el cual una relación de necesidad o suficiencia entre una condición causal (o combinación de condiciones) cumpla un resultado con un conjunto de datos dado (Legewie, 2013, p. 11).

Posteriormente, mediante la calibración de Ragin se construye una tabla de verdad. Esta permite observar las condiciones presentes o ausentes en cada uno de los países analizados, su valor de presencia se define según su nivel de consistencia, en este caso por encima de 0.89. El resultado es la formación de subconjuntos de países que contienen condiciones presentes o ausentes, dando como resultado patrones causales de suficiencia, donde [0] denota la ausencia de la variable o factor psicosocial y [1] la presencia de dicho factor psicosocial.

Finalmente, a través de suposiciones lógicas, o bien, de correlaciones lógicas, se construyen las soluciones de concurrencia donde el símbolo [*] es el ancla del valor lógico “AND” o “y”; [v] es el valor lógico “OR” o “O”; [~] es la negación, es decir, la ausencia de la condición. En este sentido, la meta es identificar condiciones o combinaciones de necesidad o de suficiencia para μ resultado.

QCA, al tratarse de valores entre [0] y [1], utiliza minimización booleana para reducir las expresiones y llegar a soluciones inteligibles, es decir, identificar más y más combinaciones generales de condiciones suficientes para los resultados que permanecen lógicamente verdaderas.

Variables analizadas

Se ha mencionado que las variables que se analizaron provienen de distintos índices. En la tabla 1 se muestran cuáles fueron y cómo estuvieron integradas.

Tabla 1. Variables consideradas para el análisis

Nombre	Abreviatura	Datos que contempla	Fuente
Calidad de vida	CAL_VID	Incluye la estabilidad económica y política, el sistema legal y los derechos civiles, la disponibilidad y calidad de los servicios de salud (incluyendo la esperanza de vida), la seguridad y la paz global, el clima, los costes e ingresos asociados a la vida diaria, así como la popularidad o aceptación social.	Datos Mundial (2024)
PIB per cápita	PIB ´P	División del producto bruto interno total de un país entre su población, lo cual permite medir el ingreso promedio por habitante y ofrece una aproximación al nivel económico general de un país.	Banco Mundial (2024)
Desarrollo humano	IDH	Indicadores como la esperanza de vida al nacer, los años esperados y promedio de escolaridad, así como la renta nacional bruta per cápita ajustada por paridad de poder adquisitivo (PPP) correspondiente al año 2021. Estos indicadores reflejan el bienestar material y las oportunidades educativas disponibles para la población.	Eustat (2024)
Progreso social	PROG-SOC	Múltiples indicadores que abarcan fundamentos de bienestar, necesidades humanas básicas y oportunidades. Entre estos se encuentran el analfabetismo, la cobertura y calidad de la educación, el acceso a tecnologías como Internet, celulares y computadoras, la calidad del medio ambiente urbano, tasas de mortalidad infantil y materna, y aspectos relacionados con seguridad, violencia, participación política y derechos civiles.	Social Progress Imperative (2024)

Nombre	Abreviatura	Datos que contempla	Fuente
Índice de felicidad	IFB	Integra aspectos económicos y sociales, considerando no solo los ingresos (PIB per cápita), sino también factores subjetivos como el apoyo social (tener una red de apoyo), la percepción de salud física y mental (esperanza de vida saludable y autopercepción de salud), el sentido de libertad para tomar decisiones importantes en la vida, la generosidad y la percepción de ausencia de corrupción.	Helliwell <i>et al.</i> (2023)
Apoyo social	AP_SOC	Se refiere a la evaluación de la red de apoyo que tiene una persona, es decir, la disponibilidad de alguien con quien contar en momentos de necesidad. Este componente es fundamental para el bienestar emocional y la resiliencia ante situaciones adversas.	OCDE (2024)
Autorreporte de salud	AUTO-SAL	Consiste en la percepción que tienen las personas acerca de su propio estado de salud. Este indicador subjetivo complementa las mediciones objetivas, ya que refleja cómo los individuos valoran su bienestar físico y mental en su vida diaria.	OCDE (2024)
Nivel educativo	NIV_EDU	Se mide a partir del promedio de avance escolar, lo cual indica el grado de progreso y logro educativo alcanzado por la población. Este factor es crucial para el desarrollo personal y social, y está relacionado con mejores oportunidades laborales y calidad de vida.	OCDE (2024)

Nombre	Abreviatura	Datos que contempla	Fuente
Satisfacción con la vida	SAT_VIDA	Evalúa la percepción personal de diversos aspectos como la salud, la educación, los ingresos, la satisfacción personal y las condiciones sociales en las que vive el individuo. Este índice busca captar el bienestar subjetivo y el sentido de realización que experimentan las personas en su entorno.	OCDE (2024)

Fuente: Elaboración propia a partir de *Global Social Progress Index, PIB per cápita*, “Índice para una vida mejor”, *World Happiness Report 2023*, “Índice de desarrollo humano por indicadores según países” e “Índice de calidad de vida”.

Resultados

Condiciones de necesidad con fs/QCA para la calidad de vida

Para el análisis de las condiciones de necesidad se consideró el umbral cuyo valor fuera igual o mayor a 0.85, lo que representó que cualquiera que cumpliera esa característica se consideraría como necesaria para la calidad de vida en los países analizados. Conforme a este criterio, en la tabla 2 se muestran los valores obtenidos para cada condición. Se observa que la satisfacción con la vida (0.863) y el progreso social (0.855) son las condiciones que contribuyen a crear mayor calidad de vida; aunque el desarrollo humano, la felicidad, y las demás por debajo son importantes, al parecer para estos países no son tan necesarias.

El análisis de este resultado refleja lo que teóricamente se ha revisado sobre la calidad de vida, que tanto el factor subjetivo como la *satisfacción por la vida* representada por el bienestar subjetivo y el sentido de realización que experimentan las personas en la totalidad de su entorno en aspectos como la salud, la educación, los ingresos, la satisfacción personal y las condiciones sociales en las que vive el individuo (OCDE, 2024) es el principal factor necesario para la calidad de vida, seguido por el *progreso social* compuesto por múltiples indicadores

que abarcan fundamentos de bienestar, necesidades humanas básicas y oportunidades (Social Progress Imperative, 2024).

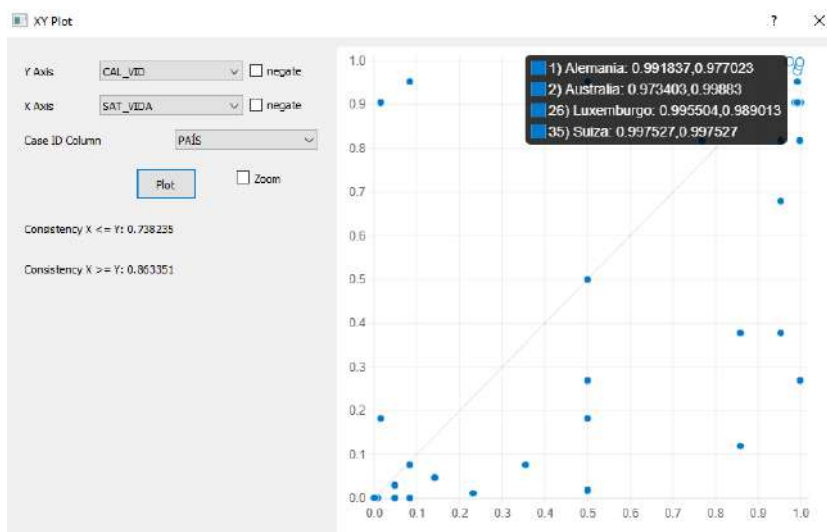
Tabla 2. Análisis de condiciones necesarias para la calidad de vida

	Consistencia	Promedio
SAT_VIDA	0.863351	0.738239
PROG-SOC	0.855446	0.820954
IDH	0.819346	0.799016
IFB	0.798023	0.751999
NIV_EDU	0.688177	0.621413
AUTO_SAL	0.672594	0.612276
AP_SOC	0.668566	0.666712
PIB´P	0.637732	0.841248

Fuente: Elaboración propia.

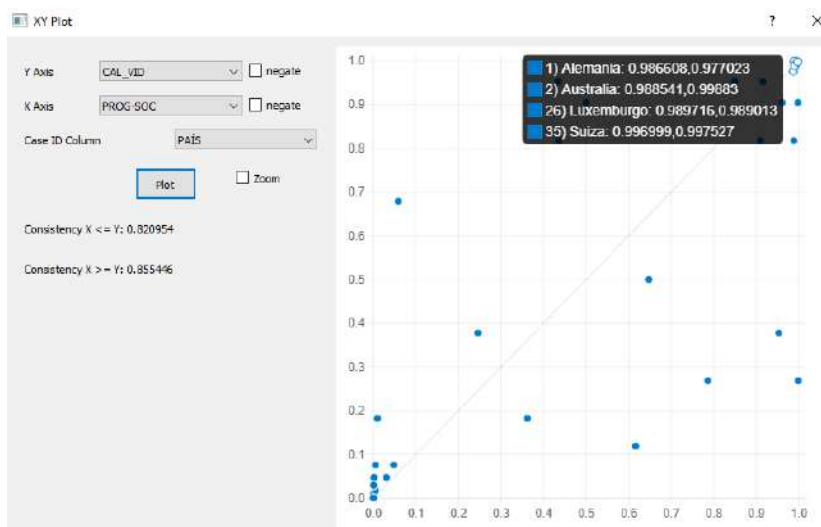
En la figura 1 se muestran los países donde la satisfacción con la vida es una condición necesaria para la calidad de vida. Estos se ven representados por los puntos azules sobre y bajo la diagonal en la región superior derecha. Los más altos fueron Suiza, Luxemburgo (0.995), Alemania y Australia; en contraste, los países más bajos fueron Colombia y Turquía.

Figura 1. XYPLOT de la satisfacción con la vida como condición necesaria para la calidad de vida



Por otra parte, en la figura 2 se muestran los países donde el progreso social desempeñó una condición necesaria para la calidad de vida. En este caso, de igual manera se muestra que los más altos puntajes los tienen los mismos países con ligeras variaciones: Suiza, Luxemburgo, Australia y Alemania. En el caso de los más bajos, fueron Costa Rica, Colombia, Brasil, México, Eslovaquia, Hungría y Turquía.

Figura 2. XYPLOT del progreso social como condición necesaria para la calidad de vida



Condiciones de suficiencia con fs/QCA

En este análisis de datos, una vez realizada la calibración y obtenidos los valores booleanos de presencia o ausencia de las condiciones causales tomando como referencia que [1] representa la condición causal presente y [0] la ausente, se construyó una tabla de verdad, la cual permitió visualizar las condiciones que impactan en la calidad de vida bajo tres soluciones: compleja, intermedia y parsimoniosa, las cuales fueron idénticas.

Nótese en la tabla 3 una serie de subconjuntos con minimización booleana. La primera fila muestra el subconjunto de países con condiciones similares; de la fila 2 a la 8, las condiciones (presentes o ausentes); la fila 9, la solución booleana de calidad de vida; y, finalmente, en la última fila la tasa de consistencia que determina la calidad de vida, que en este caso se encuentra por una tasa de consistencia por arriba de 0.88.

Nótese que países como *Irlanda, Noruega, Países Bajos, Suecia y Suiza* son países que tienen presente cada una de las condiciones que

incentivan la mejora de la calidad de vida. En contraste con países del continente americano —aun cuando hay dos europeos—, los cuales no muestran la presencia de condiciones que mejoren la calidad de vida.

Es importante mencionar que en *Finlandia*, aun cuando no existe una autopercepción personal de su propio estado de salud, no es una condición que incide en la mejora de la calidad de vida, como lo es la presencia del producto interno bruto per cápita (PIB_P), el índice de desarrollo humano (IDH), el progreso social (PROG-SOC), el índice de felicidad bruta (IFB), el apoyo social (AP_SOC), el nivel educativo (NIV_EDU), la satisfacción de vida (SAT_VID) y la calidad de vida (CAL_VID).

Por otro lado, en países como Luxemburgo y Austria la ausencia del apoyo social no es una condición de importancia que afecte negativamente al incremento de la calidad de vida.

Tabla 3. Minimización booleana de la calidad de vida

País	PIB_P	IDH	PROG-SOC	IFB	AP_SOC	AUTO_SAL	NIV_EDU	SAT_VID	CAL_VID	CONSISTENCIA
Irlanda, Noruega, Países Bajos, Suecia y Suiza	1	1	1	1	1	1	1	1	1	0.895707
Brasil, Chile, México y Turquía	0	0	0	0	0	0	0	0	0	0.271587
Colombia, Costa Rica y Grecia	0	0	0	0	0	0	0	0	0	0.327481
Hungría y Polonia	0	0	0	1	0	0	1	0	0	0.348295
Letonia	0	0	0	0	0	0	1	0	0	0.409808
Lituania	0	0	0	1	0	0	1	0	0	0.526624
Luxemburgo	1	1	1	1	0	1	0	1	1	0.916226
Islandia	1	1	1	1	1	1	0	1	0	0.85366
Francia	0	0	0	1	1	0	1	1	1	0.907284
Finlandia	1	1	1	1	1	0	1	1	1	0.991871
Austria	1	1	1	1	0	1	1	1	1	0.921391
Estados Unidos	1	1	0	1	1	1	1	1	0	0.832165
Nueva Zelanda	0	1	1	1	1	1	1	1	1	0.904245

Fuente: Elaboración propia. *Nota:* La tabla 3 muestra las distintas correlaciones de la minimización booleana de 36 países respecto a la calidad de vida.

Por otra parte, las cuatro combinaciones o caminos causales (tabla 4) representan diferentes configuraciones suficientes para la ocurrencia del resultado de la calidad de vida (Mejía-Trejo, 2021). Para el primer caso, con nivel de consistencia de 0.93, $PIB^*P^*IDH^*PROG_SOC^*IFB^*\sim AP_SOC^*AUTO_SAL^*SAT_VIDA$, se encontró que si hay condiciones adecuadas respecto al PIB per cápita, al desarrollo humano, al progreso social y a la felicidad, aunado a la ocurrencia de una adecuada valoración de salud y satisfacción con la vida, son suficientes para que haya calidad de vida, aunque no haya apoyo social. En otras palabras, la calidad de vida se puede lograr incluso sin apoyo social, si el entorno económico y social es suficientemente fuerte y satisfactorio, pero también es importante la presencia de las variables subjetivas como la percepción de salud, la satisfacción con la vida y la felicidad. En esta solución los casos representativos fueron Luxemburgo y Austria.

Tabla 4. Solución parsimoniosa

Solución parsimoniosa	Cobertura bruta	Cobertura única	Consistencia
$PIB^*P^*IDH^*PROG_SOC^*IFB^*\sim AP_SOC^*AUTO_SAL^*SAT_VIDA$	0.218193	0.0804366	0.932399
$\sim PIB^*P^*\sim IDH^*\sim PROG_SOC^*IFB^*AP_SOC^*\sim AUTO_SAL^*NIV_EDU^*SAT_VIDA$	0.0756617	0.0281952	0.907284
$PIB^*P^*IDH^*PROG_SOC^*IFB^*AP_SOC^*\sim AUTO_SAL^*NIV_EDU^*SAT_VIDA$	0.153853	0.0392075	0.991871
$\sim PIB^*P^*IDH^*PROG_SOC^*IFB^*AP_SOC^*AUTO_SAL^*NIV_EDU^*SAT_VIDA$	0.200195	0.0492797	0.904245

Fuente: Elaboración propia. *Nota:* La consistencia de la solución parsimoniosa fue de 0.928944 y la cobertura de la solución fue de 0.378236.

En cuanto al segundo camino causal:

$$\sim PIB^*P^*\sim IDH^*\sim PROG_SOC^*IFB^*AP_SOC^*\sim AUTO_SAL^*NIV_EDU^*SAT_VIDA$$

Se reflejó que con que haya felicidad, apoyo social, nivel educativo y satisfacción con la vida, puede ser una cuestión suficiente para que exista calidad de vida, aunque no haya un alto PIB per cápita, ni desa-

rollo humano, ni autoevaluación de salud alta. Este resultado mostró alta consistencia (0.90) y como caso representativo fue Francia.

En el tercer camino causal, se identificó una combinación causal con un alto nivel de consistencia (0.99) $PIB^*P^*IDH^*PROG\ SOC^*IFB^*AP_SOC^*\sim AUTO_SAL^*NIV_EDU^*SAT_VIDA$, lo que respalda su carácter de condición suficiente para la presencia de una alta calidad de vida. Esta combinación se compuso por un alto PIB per cápita, elevado nivel de desarrollo humano, progreso social, altos niveles de felicidad, apoyo social, nivel educativo y satisfacción con la vida, incluso en ausencia de una alta autoevaluación del estado de salud. Finlandia emerge como el caso empíricamente más representativo de esta configuración.

El cuarto camino causal mostró una consistencia alta (0.90) $\sim PIB^*P^*IDH^*PROG\ SOC^*IFB^*AP_SOC^*AUTO_SAL^*NIV_EDU^*SAT_VIDA$, lo que indica que no es imprescindible tener un alto PIB para alcanzar calidad de vida, si se cuenta con capital humano y social fuerte representado por el resto de las variables analizadas. El caso representativo de esta solución fue Nueva Zelanda.

A partir de estos resultados, se observaron ciertas semejanzas en los distintos caminos causales, como la presencia de variables subjetivas en la mayoría de las configuraciones (en tres de las cuatro soluciones estuvieron presentes la satisfacción con la vida y la felicidad). Estas variables se mantienen, aunque exista la ausencia de factores estructurales (por ejemplo, PIB per cápita bajo). Por otra parte, la condición de autoevaluación de salud aparece ausente en tres de los cuatro caminos, lo cual es interesante porque sugiere que la percepción de salud no es determinante siempre que otros factores estructurales estén presentes.

Estas configuraciones permiten concluir que la calidad de vida es un fenómeno multidimensional, al que se puede acceder mediante diferentes combinaciones de condiciones. Algunos caminos priorizan lo estructural, otros lo subjetivo, y algunos logran una integración de ambos dominios, como fue el caso en cada camino causal. Esta flexibilidad es coherente con la naturaleza equifinal de los fenómenos sociales (Fiss, 2011), y en particular de la calidad de vida, como se expuso. Sin embargo, es claro que sí existen condiciones que deben estar presentes para que exista calidad de vida.

La diversidad de caminos en este análisis aporta evidencia empírica a favor de un enfoque multidimensional en la formulación de políticas públicas. Las configuraciones que excluyen al PIB per cápita como con-

dición presente resultan particularmente relevantes para países, ya que son los que mayor desarrollo tienen, dejando fuera del análisis a países de Latinoamérica (tabla 5). Finalmente, la cobertura moderada de la solución indica que, si bien estas combinaciones explican una parte significativa del fenómeno, existen otros factores no contemplados en este análisis que también inciden en la calidad de vida (Medina *et al.*, 2017). En futuras investigaciones podrían incorporarse dimensiones culturales, institucionales o ambientales para enriquecer el modelo.

Tabla 5. Países representativos por camino causal

Miembros en términos de:	País
$PIB^+P^+IDH^+PROGSOC^+IFB^+AP_SOC^+AUTO_SAL^+SAT_VIDA:$	Luxemburgo, Austria
$\sim PIB^+P^+\sim IDH^+\sim PROGSOC^+IFB^+AP_SOC^+\sim AUTO_SAL^+NIV_EDU^+SAT_VIDA:$	Francia
$PIB^+P^+IDH^+PROGSOC^+IFB^+AP_SOC^+\sim AUTO_SAL^+NIV_EDU^+SAT_VIDA:$	Finlandia
$\sim PIB^+P^+IDH^+PROGSOC^+IFB^+AP_SOC^+AUTO_SAL^+NIV_EDU^+SAT_VIDA:$	Nueva Zelanda

Fuente: Elaboración propia.

Conclusiones

La calidad de vida es un constructo dinámico y complejo que cambia conforme a elementos individuales derivados de la cognición y percepción subjetiva, y a elementos culturales, históricos, sociales, entre otros. En el presente estudio, a partir del Análisis Cualitativo Comparado con Conjuntos Difusos (fs/QCA) se indagó en la relación de los indicadores psicosociales objetivos y subjetivos de la calidad de vida en 38 países de la OCDE.

Los resultados obtenidos mediante el análisis fs/QCA permitieron identificar configuraciones causales múltiples que explican la calidad de vida en los países estudiados. Una de las principales contribuciones del estudio radica en evidenciar que existen caminos alternativos para alcanzar altos niveles de calidad de vida, como se observó con los resultados de los caminos causales (Fiss, 2011). Este hallazgo es consistente con investigaciones previas que destacan la naturaleza mul-

tidimensional del bienestar a través de la multiplicidad de definiciones, en la que debe considerarse la gama de factores internos y externos, así como objetivos y subjetivos (Veenhoven, 2000). Estos hallazgos no permitieron confirmar las hipótesis planteadas, no obstante, amplían las posibilidades de interpretación sobre los factores necesarios y suficientes para la calidad de vida en los países de la muestra analizada.

En tres de las cuatro soluciones encontradas, las variables subjetivas como la satisfacción con la vida y la felicidad estuvieron presentes como condiciones suficientes. Esto refuerza la idea de que los factores psicosociales no son meramente complementarios, sino centrales en la experiencia del bienestar. La relevancia del bienestar subjetivo, incluso cuando faltan factores estructurales como el PIB per cápita o el desarrollo humano, sugiere que la percepción individual del entorno puede modular fuertemente la calidad de vida. No obstante, no son excluyentes los factores estructurales (objetivos), ya que, con independencia de las percepciones individuales o colectivas, es parte de los derechos humanos contar con las condiciones de vida (factores estructurales) adecuadas para asegurar la satisfacción de las necesidades, más allá de la percepción de la existencia de estas.

Fue relevante que el PIB per cápita no representara una condición necesaria ni suficiente en todos los casos. Esto coincide con estudios que advierten sus límites como predictor del bienestar general, pues el PIB no fue creado como índice de bienestar social; aunque tiene un impacto en el bienestar económico que permite el acceso a distintos recursos, no tiene una incidencia directa en la calidad de vida (Kahneman y Deaton, 2010; Govea, 2018). Países como Francia y Nueva Zelanda logran altos niveles de calidad de vida sin cumplir todas las condiciones estructurales, pero con fuertes dimensiones psicosociales.

A modo de conclusión, los hallazgos invitan a reorientar las políticas hacia enfoques integrales del bienestar que no se limiten al crecimiento económico, pues, como lo plantea Veenhoven (2000), el bienestar no está únicamente relacionado con los factores estructurales. Por ello, se sugiere:

- *Invertir en programas que fortalezcan el bienestar subjetivo, incluyendo salud mental, educación emocional y sentido de propósito.* Todas estas esferas de la salud mental están, a su vez, alineadas con los Objetivos de Desarrollo Sostenible (ODS) como un aspecto prioritario para el desarrollo en la agenda de la ONU. Es menester mencionar

que, al mismo tiempo, el concepto de calidad de vida permite la aproximación de las personas al desempeño en diferentes aspectos de su vida y comprender la perspectiva sobre su propio bienestar (Ornelas y Ruíz, 2017), lo que incrementa su relevancia.

- *Promover entornos sociales que aumenten la satisfacción con la vida, el apoyo social y las oportunidades personales.* El contexto ambiental rico en estimulación y oportunidades es relevante para los seres humanos en todas las etapas de la vida. Pensar en espacios donde los niños, adolescentes, adultos y ancianos puedan colaborar y aprender representa un escenario necesario. Estos escenarios podrían abordarse desde una perspectiva de desarrollo tomando como parámetro la clasificación de Bronfenbrenner y Morris (2006), es decir, desde mejores condiciones de vivienda, contextos escolares o recreativos, condiciones laborales, vecindarios con parques, etcétera.
- *Evaluar el éxito de las políticas públicas no solo con indicadores económicos, sino también con indicadores de bienestar multidimensional.* Para ello se han realizado aproximaciones, no obstante, se requiere de nuevas estrategias para operacionalizar y llevar a la práctica las diferentes propuestas analizadas, incluyendo qué se entiende por bienestar (OCDE, 2022).

Toda investigación es perfectible; en este estudio, si bien la cobertura total de la solución parsimoniosa fue moderada (0.37), los caminos causales identificados explican una parte sustancial del fenómeno, lo que pone de manifiesto la necesidad de explorar otras variables que podrían incrementar el poder para explicar las condiciones de necesidad y suficiencia. Asimismo, la muestra se limita a países de la OCDE, lo que restringe la generalización a otros países y, dentro de este grupo, a los menos desarrollados.

Se propone que futuras investigaciones incorporen dimensiones adicionales tales como factores ambientales, institucionales o culturales con el fin de ampliar la capacidad explicativa del modelo y captar con mayor precisión la complejidad del fenómeno de la calidad de vida. Asimismo, la inclusión de una muestra más diversa que contemple países latinoamericanos permitiría explorar configuraciones causales propias de contextos con menor grado de desarrollo estructural, contribuyendo así a una comprensión más global y equitativa del bienestar.

Con relación a la metodología, si bien el análisis con conjuntos difusos (fs/QCA) no permite realizar inferencias probabilísticas en sentido clásico, su valor radica en la identificación de múltiples configuraciones causales suficientes, lo que resulta particularmente útil en fenómenos sociales de naturaleza compleja, no lineal y equifinal. Esta característica metodológica constituye una fortaleza para representar de manera realista las distintas trayectorias hacia un mismo resultado (Fiss, 2011). El fs/QCA no busca sustituir otros enfoques cuantitativos, sino complementarlos mediante una lógica de análisis que privilegia la diversidad, la conjunción de condiciones y la contextualización de los hallazgos.

Referencias

- Ardila, R. (2003). Calidad de vida: una definición integradora. *Revista Latinoamericana de Psicología*, 35(2), 161-164.
- Banco Mundial (2024). PIB per cápita (USD actuales). <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
- Bronfenbrenner, U., y Morris, P. A. (2006). The bioecological model of human development. En R. M. Lerner (Ed.), *Handbook of child psychology* (6.ª ed., vol. 1, pp. 793-828). John Wiley & Sons.
- Datos Mundial (2024). Comparación de calidad de vida por país. <https://www.datosmundial.com/calidad-de-vida.php>
- Eustat (2024). Índice de desarrollo humano por indicadores según países 2024. Eustat-Instituto Vasco de Estadística. https://www.eustat.eus/elementos/ele0013500/ti_indice-de-desarrollo-humano-por-indicadores-segun-paises-2019/tbl0013566_c.html
- Felce, D., y Perry, J. (1995). Quality of life: Its definition and measurement. *Research in Developmental Disabilities*, 16, 51-74.
- Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54(2), 393-420.
- Govea, A. M. (2018). Evolución e impacto del PIB y el IDH en un mundo desigual. *Revista Vinculando*, 16(1).
- Helliwell, J. F., Layard, R., Sachs, J. D., Aknin, L. B., De Neve, J.-E., y Wang, S. (Eds.). (2023). *World Happiness Report 2023* (11.ª ed.). Sustainable Development Solutions Network. <https://worldhappiness.report/ed/2023/#appendices-and-data>

- Kahneman, D., y Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38), 16489-16493.
- Legewie, N. (2013). *An Introduction to Applied Data Analysis with Qualitative Comparative Analysis*. Deutsches Institut für Wirtschaftsforschung.
- Medina, I., Álamos-Concha, P., Castillo Ortiz, P. J., y Rihoux, B. (2017). *Análisis cualitativo comparado (QCA)* (vol. 56). CIS-Centro de Investigaciones Sociológicas.
- Mejía-Trejo, J. (2021). *Análisis Cualitativo Comparativo (fsQCA) y su relación con la innovación: Discusión e interpretación de resultados (tomo II)*. Universidad de Guadalajara.
- Nicolás, S. C. M., Alejandro, A. H. S., Gabriel, R. C. E., y Katherine, R. A. C. (2022). Calidad de vida: el camino de la objetividad a la subjetividad en población general y grupos como: niños y jóvenes, personas con discapacidad y adultos mayores. *Revista Médica Vozandes*, 33(1), 61.
- Organización para la Cooperación y el Desarrollo Económicos (OCDE) (2022). *¿Cómo va la vida en América Latina? Medición del bienestar para la formulación de políticas públicas*. OECD Publishing. <https://doi.org/10.1787/7f6a948f-es>
- (2024). Índice para una vida mejor. <https://www.oecdbetterlifeindex.org/es/>
- Ornelas, A. R., y Ruíz, A. M. (2017). Salud mental y calidad de vida: Su relación en los grupos etarios. *PSIENCIA. Revista Latinoamericana de Ciencia Psicológica*, 9(2), 1-16.
- Palomba R. (2002). *Calidad de Vida: Conceptos y medidas. Documentos del Taller sobre calidad de vida y redes de apoyo de las personas adultas mayores*. Centro Latinoamericano y Caribeño de Demografía (CELADE) / División de Población, Comisión Económica para América Latina y el Caribe (CEPAL). Fondo de Población de las Naciones Unidas, julio de 2002.
- Ragin, C. (2008). *Resigning social Inquiry. Fuzzy Set and Beyond*. The University of Chicago Press.
- Salas, C., y Garzón, M. O. (2013). La noción de calidad de vida y su medición. *Revista CES Salud Pública*, 4(1), 36-46.
- Schneider, C., y Wagemand C. (2012). *Set Theoretic Methods for the Social Sciences. A guide to Qualitative Comparative Analysis*. Cambridge.
- Social Progress Imperative (2024). *Global Social Progress Index*. <https://www.socialprogress.org/social-progress-index>
- Tonon, G. (2010). La utilización de indicadores de calidad de vida para la decisión de políticas públicas. *Polis. Revista Latinoamericana*, (26).
- Urzúa, A., y Caqueo-Úrizar, A. (2012). Calidad de vida: Una revisión teórica del concepto. *Terapia Psicológica*, 30(1), 61-71.

- Urzúa, A., Pavlov, R., Cortés, R., y Pino, V. (2011). Factores psicosociales relacionados con la calidad de vida en salud en pacientes hemodializados. *Terapia Psicológica*, 29(1), 135-140.
- Varela, L. F. (2016). Salud y calidad de vida en el adulto mayor. *Revista Peruana de Medicina Experimental y Salud Pública*, 33, 199-201.
- Veenhoven, R. (2000). The four qualities of life: Ordering concepts and measures of the good life. *Journal of Happiness Studies*, 1(1), 1-39.

**Gestión tributaria y competitividad
Una visión sistémica desde los negocios**

© 2025, Academia Mexicana de Investigación y Docencia en Innovación (AMIDI).

Distribución digital:

Academia Mexicana de Investigación y Docencia en Innovación (AMIDI).
Responsable del registro DOI, la gestión de metadatos y la publicación en **AMIDI.Biblioteca**.



Maquetación y diseño:

Talleres gráficos de **Ediciones de la Noche**
Madero #687, Zona Centro Guadalajara, Jalisco, México.



Hecho y editado en México / Made and edited in Mexico
Se terminó de editar en Noviembre de 2025.

Edición digital con posibilidad de impresión bajo demanda.
Las copias impresas se generan individualmente a solicitud del lector;
por lo tanto, **no existe un tiraje fijo.**

Coordinado por Antonio Aguilera Ontiveros, quien además participa como autor de un capítulo, *Nuevas fronteras en ciencias sociales computacionales: análisis del discurso, la opinión y la dinámica social* constituye un aporte innovador al estudio de fenómenos sociales mediante métodos computacionales. La obra integra enfoques como la simulación basada en agentes, la minería de texto y el análisis comparativo, aplicados al examen de procesos de formación de opiniones, diseminación cultural y comportamiento en plataformas digitales.

Con una estructura en seis capítulos, el libro combina teoría y práctica al explorar desde la construcción de corpus hasta el análisis de tópicos en campañas electorales recientes, mostrando cómo la interacción entre ciencias sociales y herramientas digitales permite comprender problemáticas contemporáneas como la polarización, la cohesión social y la calidad de vida.

Por su carácter interdisciplinario y el liderazgo de su coordinador, el texto se convierte en una referencia clave para investigadores, docentes y estudiantes interesados en aplicar enfoques computacionales al análisis social, aportando nuevos marcos metodológicos y perspectivas críticas que enriquecen el campo de las ciencias sociales en el contexto latinoamericano.

ISBN 978-607-26875-6-1



9 786072 687561



Scientia et Praxis

AMIDI
Academia Mexicana
de Investigación y Docencia
en Innovación

Zapopan, Jal. a 20 de Septiembre de 2025

Dictamen de Obra. AMIDI.DO.20250920

Los miembros del equipo editorial de la Academia Mexicana de Investigación y Docencia en Innovación (**AMIDI**), **RENIECYT-SECIHTI 2200092**, ver:

<https://www.amidibiblioteca.amidi.mx/index.php/AB/about/editorialTeam>

se reunieron para atender la invitación a dictaminar el libro:

Nuevas fronteras en ciencias computacionales: análisis del discurso, la opinión y la dinámica social

Siendo los siguientes participantes de la misma:

Nombre	Rol
Aguilera Ontiveros, Antonio	Coordinador editorial
Abrica Jacinto, Norma Leticia	Autor
Evguenii, Kurmyshev	Autor
Zazueta Gutiérrez, Jorge	Autor
López Morales, Sadan Josué	Autor
Gonzalez Silva, Ricardo Armando	Autor
Gonzalez Silva, Mario Ignacio	Autor
Aguilera Ontiveros, Antonio	Autor
Martínez Galván, Adela	Autor
Raimondo Anselmino, Natalia	Autor
Gindín, Irene	Autor
Sartorio, Alejandro	Autor
Cellone, Franco	Autor
Araoz, Verónica	Autor
Alomar, Francisco J.	Autor
Sánchez Mondragón, Gabriela	Autor
Pérez Balcázar, Anthony	Autor

Dicho documento fue sometido al proceso de evaluación por pares doble ciego, de acuerdo a la política de la editorial, para su dictaminación de aceptación, ver:

<https://www.amidibiblioteca.amidi.mx/index.php/AB/procesodeevaluacionporparesen ciego>

Los miembros del equipo editorial se reúnen con el curador principal del repositorio digital para convocar:

1. Que el comité científico, de forma colegiada, revise los contenidos y proponga a los pares evaluadores que colaboran dentro del comité de redacción, tomando en cuenta su especialidad, pertinencia, argumentos, enfoque de los capítulos al tema central del libro, entre otros.
2. Se invita a los pares evaluadores a participar, formalizando su colaboración.
3. Se envía así, el formato de evaluación para inicio del proceso de evaluación doble ciego a los evaluadores elegidos de la mencionada obra.
4. El comité científico recibe las evaluaciones de los pares evaluadores e informa a el/los autor/(es) los resultados a fin de que se atiendan las observaciones, el requerimiento de reducción de similitudes, y recomendaciones de mejora a la obra.
5. La obra evaluada, consta de:

Contenido
Contenido
Introducción
Capítulo 1. Algoritmo de autorreflexión de agentes sociales para la formación de grupos de opinión a través de un modelo de dinámica de opinión de acuerdo relativo Norma Leticia Abrica Jacinto Evguenii Kurmyshev
Capítulo 2. Identificando ecuaciones parsimoniosas que gobiernan la dinámica de modelos basados en agentes . Una primera aproximación Jorge Zazueta Gutiérrez
Capítulo 3. Diseminación cultural con aumento de presencia de medios de comunicación en el modelo de Axelrod Sadan Josué López Morales Ricardo Armando Gonzalez Silva y Mario Ignacio Gonzalez Silva
Capítulo 4. Análisis de tópicos en comentarios de YouTube durante la campaña presidencial de México 2024: un estudio con Ida Antonio Aguilera Ontiveros Adela Martínez Galván
Capítulo 5. Herramientas y técnicas computarizadas al servicio de la construcción de corpus y datasets . Desafíos metodológicos del análisis de discursos mediatizados en plataformas con base en Internet Natalia Raimondo Anselmino, Irene Gindin, Alejandro Sartorio, Franco Cellone, Verónica Aráoz

Francisco J. Alomar

Capítulo 6. Factores psicosociales en la dinámica de la calidad de vida . Análisis Cualitativo Comparado con Conjuntos Difusos (fs/qca)

Gabriela Sánchez Mondragón y Anthony Pérez Balcázar

6. Una vez emitidas las observaciones, el requerimiento de reducción de similitudes, y recomendaciones de mejora a la obra por los evaluadores y todas ellas resueltas por el/los autor/(es), el resultado resalta que el contenido del libro:
 - a. Reúne los elementos teóricos actualizados y prácticos desglosados en cada uno de sus capítulos.
 - b. Los capítulos contenidos en la obra, muestran claridad en el dominio del tema, congruencia con el título central del libro, y una estructura consistente
 - c. Se concluye finalmente, que la obra dictaminada, puede fungir como libro de texto principal o de apoyo tanto para estudiantes de licenciatura como de posgrados, así como público en general interesado.
7. De esta forma, el resultado del dictamen de aceptación de la obra fue:

FAVORABLE PARA SU PUBLICACIÓN

Sirva la presente para los fines que a los Interesados convengan.

Atentamente



Mtro. Rodrigo Mejía-Mancilla
Curador AMIDI.Biblioteca
AMIDI