POLARIZACIÓN EN REDES SOCIALES AYUDA A QUE LOS *INFLUENCERS* TENGAN MÁS INFLUENCIA: ANÁLISIS Y DOS ESTRATEGIAS DE INOCULACIÓN

POLARIZATION IN SOCIAL MEDIA ASSISTS INFLUENCERS TO BECOME MORE INFLUENTIAL: ANALYSIS AND TWO INOCULATION STRATEGIES

Ivan Garibay[1] , Alexander V. Mantzaris[2], Amirarsalan Rajabi[1] & Cameron E . Taylor[3]

RECEPCIÓN: 26 DE AGOSTO DEL 2019;
ACEPTACIÓN: 23 DE NOVIEMBRE DEL 2019

ABSTRACT
This work explores simulations of polarized discussions from a general and theoretical premise. Specifically the question of whether a plausible avenue exists for a subgroup in an online social network to find a disagreement beneficial and what that benefit could be. A methodological framework is proposed which represents key factors that drives social media engagement including the iterative accumulation of influence and the dynamics for the asymmetric treatment of messages during a disagreement. It is shown that prior to a polarization event a trend towards a more uniform distribution of relative influence is achieved which is then reversed by the polarization event. The reasons for this reversal are discussed and how it has a plausible analogue in real world systems. A pair of inoculation strategies are proposed which aim at returning the trend towards uniform influence across users while refraining from violating user privacy (by remaining topic agnostic) and from user removal operations.

Key words:

RESUMEN
Este trabajo explora simulaciones de debates polarizados desde una premisa general y teórica. Específicamente, trata sobre la existencia de una vía verosímil para un subgrupo en una red social en línea para encontrar un desacuerdo beneficioso y cuál podría ser ese beneficio. Se propone un marco metodológico que representa los factores clave que impulsan la participación en las redes sociales, incluida la acumulación iterativa de influencia y la dinámica para el tratamiento asimétrico de mensajes durante un desacuerdo. Se muestra que, antes de un evento de polarización, se logra una tendencia hacia una distribución más uniforme de relativa influencia, lo que entonces se invierte por el evento de polarización. Se debaten las razones de esta reversión y cómo tiene un análogo verosímil en los sistemas del mundo real. Además, se propone un par de estrategias de inoculación, cuyo objetivo es devolver la tendencia hacia una influencia uniforme entre los usuarios, mientras que se abstiene de violar la privacidad del usuario (por mantener el tema agnóstico) y de las operaciones de eliminación de usuarios.

Palabras clave:

1 University of Central Florida, Department of Industrial Engineering and Management Systems, Orlando, 32816, USA. email: igaribay@ucf.edu.
2 University of Central Florida, Department of Statistics and Data Science, Orlando, 32816, USA.
3 University of Central Florida, Department of Computer Science, Orlando, 32816, USA.

El tema de la polarización en las sociedades es un problema de gran preocupación y que dispone de una plataforma diferente en la que puede desarrollarse, debido a la disponibilidad de las redes sociales en línea. Los diferenciadores clave son el bajo costo de comunicación con grandes grupos de personas a largas distancias y, aunque a menudo pasado por alto, el historial de contenido popular que puede estar en exhibición pública [1], [2]. Este último punto es muy importante para los usuarios que dedican una cantidad considerable de su tiempo, ya que son indicadores del rango de influencia («vistas», «re-tweets», «me gusta», etc.). El esfuerzo realizado para obtener influencia puede ser apoyado a través de la *cooperación* de los destinatarios del contenido cuando repropagan mensajes (re-compartiendo o «re-tuiteando»). Se puede esperar que esto produzca comportamientos agregados complejos no triviales [3], [4]. Por su parte, este trabajo investiga cómo la polarización (desacuerdos) puede beneficiar a los usuarios que buscan aumentar su influencia relativa. Basado en el enfoque de modelado que produce los efectos donde la polarización puede beneficiar a un grupo particular de usuarios, se exploran estrategias de inoculación que revierten los efectos inducidos a los estados de prepolarización.

Los temas para la polarización pueden provenir de una amplia gama de diferentes áreas de la sociedad, tales como del discurso político [5], [6], donde recientemente se han observado aumentos en los Estados Unidos [7]-[9]. Otras fuentes de polarización pueden surgir de temas como la desigualdad de riqueza [10]-[12], que ha sido un desafío para una gran parte de la historia, pero que ahora puede estudiarse como un tema de participación en línea [13]. Una exploración de la anatomía del debate en torno a los temas puede llevarse a cabo con varias herramientas modernas, tales como la asignación de Dirichlet latente [14], en la que pueden identificarse factores clave entre las categorías de contenido y sus etiquetas ideológicas. Existen, asimismo, varios enfoques para resumir el texto de diferentes grupos para que puedan examinarse las características clave que los diferencian [15].

No está claro si esa información proporciona alguna dirección sobre cómo puede reducirse una discusión polarizada o, incluso, qué debe «reducirse» en primer lugar. Existe literatura sobre los enfoques que pueden adoptarse para reducir la polarización en las redes sociales [16]-[20]; sin embargo, hay desafíos que aparecen comúnmente. Estos están relacionados con que los enfoques requieren interpretación humana en cuanto a qué direcciones tomar al comprender la motivación de la audiencia [21], que el contenido es visible para la inspección [22] o que los usuarios particulares pueden ser eliminados de la participación en la red [22]. El documento [23] discute cómo el debate grupal público puede proporcionar un medio para reducir la polarización, pero esto puede no tener lugar cuando lo hacen representantes de facciones opuestas. El autor también analiza los incentivos que tienen los representantes y cómo se puede utilizar el discurso público para motivos privados al influir en las opiniones de los demás.

Nuestro enfoque para abordar los incentivos no supone que el contenido de un mensaje pueda examinarse para adherirse a la defensa de la privacidad [24] o depender de una etiqueta en su estado. También, explora métodos en los que la polarización se puede reducir e invertir sin la necesidad de eliminar nodos. Hay varias razones por las que se busca este enfoque y una clave es que ningún lado particular del argumento es visto como un objetivo por las autoridades «extralimitadoras». Se supone que los usuarios de las redes sociales pueden explorar estos incentivos para beneficiarse de la polarización, incluso si la única característica que notan es que su puntaje de influencia cambia con el tiempo. Aunque las plataformas de discusión en línea difieren en sus mecanismos para compartir contenido, se propone una ecuación estocástica de intercambio de mensajes, en la que la mayoría de las plataformas compartirán algún aspecto de esa dinámica básica. Los intercambios de mensajes en plataformas de redes sociales en línea, donde los usuarios están expuestos a un conjunto de contenido y su intercambio, que pueden considerarse equivalentes al popular mecanismo de retweet, son, en efecto, una amplificación de contenido por parte del autor original en el que existen varias características de crédito de autoría [25].

Este es un aspecto clave que puede atraer a los usuarios, ya que sus mensajes no necesitan estar basados en plataformas que tengan una asociación con grandes canales directos con altas barreras de entrada, como los canales convencionales de difusión de contenido (televisión, diarios o editoriales). En [26], se describe esta baja inversión a alta influencia que puede buscarse en muchas plataformas modernas. El alcance potencial de los perfiles independientes en estas redes, para que su contenido (y el reconocimiento de autoría) se propague a millones, si no más, a partir de una mezcla de bordes dirigidos y propagaciones indirectas (efecto cascada), es lucrativo, por decir lo menos.

Esto se ha visto, en la práctica, desde una perspectiva comercial durante el *Super Bowl* de 2013, en el que hubo una falla de energía que resultó en que los asistentes usaran sus teléfonos para iluminar el estadio con aplicaciones de redes sociales involucradas. Antes del evento, la cuenta Twitter de Pepsi estaba en tendencia con la mayor influencia, aunque luego se redujo debido al apagón. Las cuentas de Audi y, en particular, las galletas Oreo, lograron proporcionar la creación de contenido en tiempo real con referencias al corte de energía. Esto dio lugar a que sus cuentas dominaran la discusión posterior [27]. Oportunidades como esta proporcionan una inversión lucrativa desde la perspectiva de los estrategas de marketing para encontrar oportunidades de bajo costo con las que se puedan lograr grandes impactos en la audiencia.

La sección 4 de la metodología describe el desarrollo del marco sobre el cual el modelo de miembros de redes sociales se comunica e intercambia mensajes. De los diversos modelos de comunicación propuestos, el trabajo de *Dynamic Communicators* [28] se elige para ampliarlo, ya que ofrece un mecanismo explicativo intuitivamente razonable sobre cómo los usuarios explotan y utilizan las capacidades de intercambio ad-hoc en las plataformas sociales en línea para obtener más influencia que las capacidades de transmisión directa proporcionan. Asimismo, se cuantifica el efecto del intercambio de contenido que produce «caminatas temporales» creadas a través de propagaciones de mensajes por medio de instantáneas de red ordenadas por tiempo. La naturaleza dinámica de las caminatas temporales se explora en [25] y [29], aunque también se aplica a grandes contextos comerciales de publicidad [30]. Estos estudios respaldan que los usuarios efectivos en las redes sociales no dependen exclusivamente de grandes anchos de banda (muchos mensajes directos) para aplicar su influencia en otros nodos, sino que dependen de intermediarios para propagar su contenido entre adyacencias temporales. Es decir, el efecto en el que los usuarios producen números bajos de mensajes se propaga a un gran número de usuarios indirectamente a través de intermediarios o retransmisores. Una característica clave del desarrollo propuesto es que el parámetro de influencia subyacente [31] debe cambiar como resultado de éxitos o fracasos cuando el contenido se difunde con éxito a través de intermediarios. Se considera que, con el tiempo, los éxitos deberían producir un efecto de arrastre que sea visible como una medida relativa entre otros miembros en el éxito la red de influencia.

Dado un modelo que representa un marco en el que los usuarios pueden explotar el mecanismo o dinámica del contenido compartido, dentro de ese marco se construye la dinámica de cómo los usuarios pueden comportarse hipotéticamente durante un desacuerdo, lo que crea el efecto acuñado de «polarización», como se presenta en la ecuación 3. Una característica clave de los experimentos realizados en la sección «Resultados» es qué cambios afectan la experiencia de comunicación de la comunidad y cómo se altera el flujo que debería existir en ella. Se supone que un «campo de juego más nivelado» es un signo positivo en una reunión social saludable en cualquier plataforma. Nuestros resultados muestran que, antes de la polarización, las redes exhibían una tendencia hacia un aumento en la homogeneidad con respecto a la capacidad de difundir contenido a través de receptores proxy. Esto puede interpretarse como resultado de un tipo de mayor familiaridad entre los usuarios a lo largo de los puntos de tiempo y una evaluación más uniforme del valor del contenido.

La fase de polarización, cuya dinámica está representada por las ecuaciones 3 y 4, invierte esta tendencia mostrando un favor en el contenido originado en los miembros de la red que tuvieron un

puntaje inicial de influencia de alto rango. Esto permite una recaptura del control del diálogo a través de una característica de la dinámica que es similar a la «intimidación» de los intercambios de mensajes durante un evento de polarización. Las subsecciones 4.1 y 4.2 describen las estrategias de inoculación que aliviarán este efecto, como se muestra en la sección «Resultados». Las estrategias de inoculación propuestas no suponen acceso al contenido ni requieren la eliminación de cuentas que se consideran catalizadoras del contenido polarizado y tienen incentivos de la fase polarizada. La dinámica que produce una polarización está de acuerdo con [32], que muestra cómo se puede observar la formación de diferentes opiniones opuestas en las simulaciones y en los eventos del mundo real.

El aspecto clave de la metodología que permite un examen de la difusión de contenido ad-hoc es que todos los usuarios comparten la misma velocidad a la que se produce el contenido y la cantidad de personas a las que llega. Sin embargo, las probabilidades de una mayor propagación mediada por estos eventos de contactos están reguladas por un parámetro de «influencia» (que se muestra en la ecuación 1). Se permite que este parámetro aumente con el número de respuestas exitosas (un indicador similar al recuento de retuits) y se introduce como se describe en la ecuación 2. Durante la simulación, la dinámica se modifica para simular las ecuaciones de un evento de polarización (como es el caso de la 3 y la 4) y se toma como una fase diferente a lo largo de una traza de simulación que representa el tiempo. Los efectos de la polarización producida se mitigan y revierten mediante la introducción de cualquiera de las dos estrategias de inoculación diferentes descritas en las subsecciones 4.1 y 4.2, que son independientes del contenido y que preservan la privacidad.

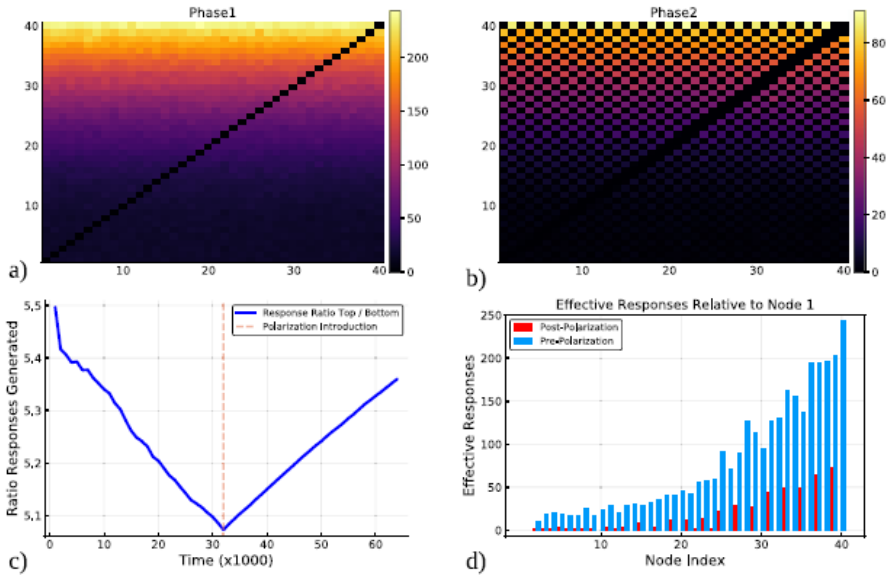

Figura 1. Cómo cambian los puntajes de influencia antes y después de la polarización. La gráfica (a) muestra los recuentos acumulados de las respuestas exitosas entre los nodos que envían mensajes sin polarización; la gráfica (b), la misma información durante la fase de comunicación polarizada (los pares y los impares tienen un desacuerdo). Cada celda representa el número de mensajes exitosos enviados por el usuario i al usuario j. El gráfico (c) muestra el valor de la influencia relativa de la mitad superior de los nodos con respecto a la mitad inferior durante la simulación en la que se introduce un evento de polarización en la línea vertical discontinua. La gráfica (d) muestra la cantidad total de mensajes exitosos enviados por cada usuario al influencer de menor rango (antes y después de la introducción de la polarización). Elaboración propia.

RESULTADOS

Los resultados de las simulaciones ejecutadas utilizando la metodología propuesta en la sección 4 se presentan aquí, donde el modelo se utiliza para examinar los efectos y las posibles soluciones para la comunicación en las redes sociales, que se ha polarizado. Una característica clave de las simulaciones son los cambios en el número relativo de respuestas exitosas que tiene un grupo de nodos con respecto a otro, como resultado de la introducción de una fase de polarización. La distribución de los valores de los parámetros de influencia asignados a partir de una distribución cúbica a lo largo de los números de nodos l_n: $0 < l_1 \leq l_2 \leq \ldots \leq l_N$. Se supone que los nodos producen contenido a una velocidad igual que no va en contra de las asignaciones de influencia, ya que la mayoría de las plataformas no restringen ni regulan la difusión del contenido del usuario en función de las clasificaciones de influencia.

La acumulación de influencia se basa en la cantidad de veces que otro nodo ha propagado el contenido de un nodo regido por una ecuación de paso de mensaje estocástico. La figura 1 muestra una simulación con dos fases de comunicación, sin polarización y luego con polarización. Se examinan los efectos sobre los grupos agregados, así como para los nodos individuales en la red de 40 usuarios sintéticos (lo que se debate más abajo). La figura 2 muestra los resultados de una simulación con tres fases de un intercambio simulado: dinámica no polarizada, polarizada y polarizada continuamente superpuesta con la de la *estrategia de inoculación 1*. La figura 3 también muestra tres fases del intercambio simulado: dinámica no polarizada, polarizada y polarizada continua, esta vez superpuesta con la *estrategia de inoculación 2*. La característica clave de la estrategia de inoculación es que se reduce la disparidad entre los usuarios para que su contenido se propague a través de intermedios. El material suplementario analiza cómo se utilizaron los datos de trazas para calcular las mediciones presentadas en los gráficos.

El número de nodos elegidos es 40, de acuerdo con el número de tamaños de comunidad Dunbar, según el cual podemos esperar que dicho grupo pueda enviar mensajes a todos los demás miembros de manera directa y uniforme [33], [34]. Este número también se elige en algunas de las investigaciones relacionadas con el discurso público sobre debates políticos, como [35], que estudia 40 cuentas de blogueros que se consideraron en una *A-List* en 2004 antes de las elecciones presidenciales de Estados Unidos. Otro análisis de blogs con orientación política opuesta en [36] también usó 40 blogs que estaban interconectados. Cambiar el número de nodos en las simulaciones no cambia las características de los resultados observados.

Por su parte, la figura 1 muestra un conjunto de cuatro tramas que describen los efectos que tiene una discusión no polarizada sobre la capacidad de los nodos para iniciar una respuesta de difusión de contenido, en comparación con el efecto que una discusión polarizada puede tener sobre esa capacidad. El panel (a) es un mapa de calor donde cada celda es el número acumulado de «propagaciones de mensajes exitosas» s_n (descrito, en detalle, en los materiales suplementarios; el número de veces que un nodo *i* tuvo una ventaja sobre otro nodo *j* que posteriormente produjo contenido). A medida que los nodos producen e intentan difundir contenido, la simulación registra la transferencia exitosa de contenido s_n entre pares de bordes *(i, j)* y presenta esa acumulación en el mapa de calor. Esta difusión de información, mediante la creación de enlaces, se cuenta en cada entrada y es proporcional al esquema de color de la leyenda. Los nodos de números de ID más bajos comenzaron con puntajes de *influencia* inferiores, que también muestran recuentos más bajos sobre la capacidad de instigar la propagación de mensajes en otros nodos por los bordes dirigidos que se crearon originalmente de acuerdo con una tasa basal uniforme. Los nodos con valores de ID más altos que comenzaron la simulación con un puntaje de influencia mayor tienen un mayor éxito acumulado en la producción de respuestas de propagación. El panel (b) muestra el mapa de calor equivalente del panel (a), pero derivado de la fase cuando existe la polarización introducida en la red.

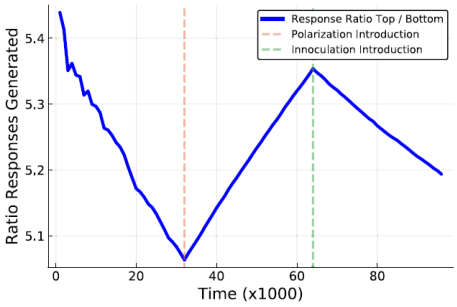

Figura 2. Cómo la estrategia de inoculación 1 puede reducir el efecto de polarización. La gráfica muestra los cambios en los valores de la influencia percibida que usan los nodos al decidir si propagar un mensaje enviado por otro nodo en tres fases de una discusión: la discusión no polarizada, la discusión polarizada y la discusión polarizada con la estrategia de inoculación 1 aplicada. Se introduce un evento de polarización en la primera línea vertical discontinua y, en la segunda línea vertical discontinua, se aplica la estrategia de inoculación 1. Elaboración propia.

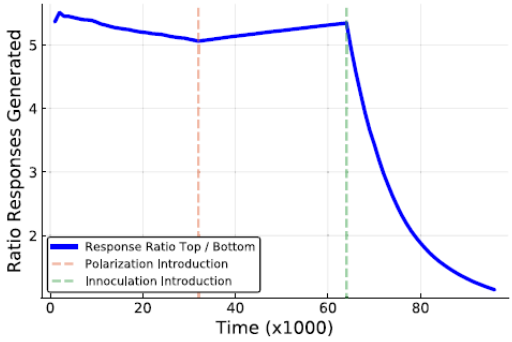

Figura 3. Cómo la estrategia de inoculación 2 puede reducir el efecto de polarización. La gráfica muestra los cambios en los valores de la influencia percibida de los nodos durante 3 fases de una discusión: intercambio no polarizado, intercambio polarizado y el intercambio polarizado que tiene aplicada la estrategia de inoculación 2. Esta disminución en la tercera fase muestra la capacidad de esta estrategia de inoculación para devolver la red a una discusión con una distribución más uniforme de las capacidades de difusión entre los participantes.

Hay un efecto de entrelazado («tablero de verificación»), debido a la probabilidad reducida que tiene un nodo de compartir su contenido entre miembros no alineados. Las acumulaciones generales también se reducen en la fase polarizada como resultado de situaciones en las que un nodo recibe mensajes de ambos lados del desacuerdo («mensajes mixtos» o «intimidación» si el contenido no alineado es de gran influencia). El panel c muestra la relación de las propagaciones de mensajes exitosas producidas entre la mitad superior de los identificadores de nodo influyentes ($n \in \{21, 40\}$) contra la mitad inferior de los ids de nodo ($n \in \{1, 40\}$). Durante el curso de la simulación, los valores de influencia del nodo pueden cambiar en función del éxito derivado de las propagaciones de mensajes que elevan su klout percibido (ecuación 2). En el punto donde hay una línea vertical discontinua, es el momento en que la fase de polarización se introduce en la simulación de intercambio de mensajes y se puede ver un cambio en la relación de tendencia para la relación de éxitos de mensajes de una tendencia

decreciente a una creciente. Esto representa que los nodos iniciales con ID de nodo más altos tenían una tasa relativa decreciente antes de la introducción de la polarización, que luego se recupera debido a la polarización. El panel d muestra el número relativo de respuestas producidas entre cada ID de nodo y el nodo 1 para las fases no polarizadas y polarizadas. El punto clave aquí es que, después de la introducción de la polarización, la proporción de los mensajes exitosos que los ID de nodo de alto rango pueden propagar al nodo 1 en relación con la de los nodos de bajo rango aumenta con respecto a la misma relación antes de la polarización.

Las características clave de los resultados de la figura 1 son que, durante la simulación en el panel c, una disminución en las tasas relativas de éxito de propagación de mensajes puede verse como una tendencia para la fase de prepolarización. Esto significa que, con el tiempo, los nodos que comienzan con valores de influencia más pequeños pueden reducir su distancia efectiva. Lo anterior puede deberse a la mecánica del algoritmo de propagación de mensajes estocásticos, que permite que los nodos con baja influencia envíen un mensaje a otro nodo durante la misma iteración de tiempo que coincide con la recepción de un mensaje desde un ID de nodo altamente clasificado. Este efecto representa un tipo de «*piggy-backing*» que puede ocurrir con el reempaquetado o la replicación de contenido (*retweeting* manteniendo o eliminando el crédito original «*re-tweet* sombra»). A medida que el número de propagaciones de mensajes aumenta con el tiempo, la diferencia inicial se vuelve menos importante y, por lo tanto, dentro de este marco, podemos postular que las discusiones no polarizadas producen una discusión más uniforme entre todos los miembros. Esto cambia durante la fase de polarización, en la que los nodos que comienzan con una capacidad relativamente más alta de diseminar contenido logran recuperar su poder de corretaje original sobre la difusión de este. Ello se puede atribuir al efecto de que las contribuciones de diferentes lados del desacuerdo cancelarán sus contribuciones de puntaje de influencia como una forma de intimidación o incertidumbre que neutraliza la probabilidad de disparar (ecuación 4). Lo que se puede ver en el panel d es que el nodo 1, que cuenta con el valor de influencia más bajo para comenzar, tiene una barrera mayor para diseminar el contenido a todos los nodos en la fase posterior a la polarización. El resultado es que este nodo tiene poco potencial para producir una respuesta visible en otros nodos y menos posibilidades de amplificar sus mensajes de contenido. Esta reinstauración del estado inicial del efecto de influencia, por lo tanto, puede usarse como una instalación para los nodos que perdieron su influencia *Klout*, a través de una dilución de la necesidad en su dominio de la red, para recuperar la influencia instigando o apoyando el desacuerdo polarizado. Esto, entonces, puede restablecer su contenido como una mayor proporción de la difusión.

La figura 2 muestra una gráfica equivalente producida en la figura 1 (c), en el contexto de tres fases de la dinámica del debate. Las dos primeras fases son las mismas que en esa figura, en la que la primera fase presenta comunicación entre los nodos sin ningún desacuerdo y la segunda fase es aquella en que se produce el evento de polarización y la influencia percibida es directamente contraproducente para difundir contenido a los usuarios al otro lado del argumento. En la tercera fase, se introduce la estrategia de inoculación 1 (subsección 4.1) para reducir los efectos de la polarización. El efecto de la polarización es particularmente la acumulación de un subconjunto de la red que tiene un mayor éxito en su capacidad de propagar contenido a través de intermediarios. Esta estrategia de inoculación, denominada simplemente «estrategia de inoculación 1», tiene como objetivo reintroducir la tendencia hacia una relación de influencia más uniforme entre los extremos superior e inferior de los influyentes clasificados. Los resultados de la aplicación de la estrategia de inoculación 1 muestran que la inoculación puede reintroducir una influencia uniforme en los nodos, como era la tendencia anterior a la polarización, lo que brinda más posibilidades de que los nodos en el extremo inferior de la distribución de puntaje de influencia inicial difundan el contenido. Como resultado, existe una mayor defensa contra los intereses aislados de un pequeño subconjunto de nodos.

Los resultados de la aplicación de la estrategia de inoculación 1, descritos en la subsección 4.1 y definidos en la ecuación 5, muestran una inversión de la disparidad de influencia producida por el período de polarización anterior. Aunque los nodos que están clasificados en la mitad superior del rango de influencia, estos todavía participan en el intercambio de contenido polarizado. Asimismo, los de menor rango pueden disminuir unilateralmente la disparidad. Esto no requiere de ninguna subcategorización de contenido específica que no sea la aparente neutralidad percibida por los mensajes enviados a otros nodos.

La figura 3 muestra los resultados de aplicar la estrategia de inoculación 2 después de un evento de polarización. La simulación es análoga a la de la figura 2, en la que, antes de la primera línea discontinua vertical, hay una fase de comunicación no polarizada e intercambios polarizados, respectivamente, entre las dos líneas verticales. La tercera fase de la comunicación sucede cuando se aplica esta segunda estrategia de inoculación (subsección 4.2). Esta estrategia de inoculación 2 tiene como objetivo colocar comunicadores en niveles donde su comunicación se realiza solo entre otros usuarios de valores de influencia similares. Al igual que con la figura anterior, la tendencia del sesgo acumulado para que los mensajes se propaguen cuando se originan a partir de identificadores de nodo superiores se invierte en una dirección que admite una relación de difusión de mensaje más uniforme. Podemos ver que la tasa de disminución en el sesgo de la distribución es mayor en este enfoque. Aunque estos resultados demuestran una superioridad de la tasa de mejora, esto requiere una participación de la administración de la plataforma, en comparación con una campaña hacia aquellos con pocos incentivos para continuar un movimiento de polarización, ya que su retorno es negativo en términos de influencia. El componente clave de la estrategia de inoculación 2 es que existe una separación escalonada de grupos de influencia, que no requiere del aislamiento de los usuarios, debido a la agrupación de contenido. La separación se orienta a restringir el contenido de los usuarios, lo que puede producir un gran contenido intimidante etiquetado debido a las clasificaciones de alta influencia en lugar de cualquier otro aspecto. Se puede esperar que los usuarios con un puntaje de influencia más bajo no encuentren que su voz sea sustancial al participar en un hilo de discusión dirigido por un usuario con más seguidores y una historia más larga de encontrar una audiencia de seguidores, cuando sus opiniones están en desacuerdo.

Los resultados de la figura 2 y la figura 3 muestran que dos estrategias de inoculación que pueden parecer diferentes permiten restablecer la tendencia hacia la difusión uniforme del contenido. Dado que el dominio de la difusión de contenido del panorama actual de las redes sociales en línea es un activo potencialmente valioso en ciertas manos, puede ser mal utilizado. Los resultados de las simulaciones de las diseminaciones de contenido polarizado muestran cómo los *influencers* que pierden el control de la diseminación de contenido en una red pueden adaptar o crear instancias de la polarización como un medio para recuperar el *klout* relativo derivado en fases anteriores. Incluso si el concepto no es conocido por esos *influencers*, los cambios en sus tasas de difusión efectivas serán monitoreados atentamente y se puede anticipar un esfuerzo por cambiar las estrategias. En tal situación, se puede adoptar un enfoque más «agresivo» de los *influencers* altamente calificados para evitar que sus seguidores difundan contenido de otros. Se puede esperar que el enfoque de diseminar contenido polarizado para mantener la influencia sea explorado y probado. De esa manera, se crean perspectivas negativas sobre el contenido de la competencia de los *influencers* y sus ideas pueden llegar a los partidarios, que pueden instigar una forma simple de dividir grupos.

DEBATE

Los desarrollos en el área de las redes sociales en línea han traído un sello distintivo en la forma en que los humanos pueden comunicarse y hay muchas formas en que sus comportamientos pueden cambiar

como resultado de una comunicación rápida y barata en todo el mundo con contenido audiovisual de alto ancho de banda. Mantener los componentes positivos de las interacciones en estas plataformas digitales debería ser una alta prioridad, dada la gran población que interactúa a través de ellas en la mayoría de los países desarrollados y en desarrollo. Debido a que los usuarios pueden usar las plataformas en un esfuerzo por establecer su influencia, gracias a las características de compartir de nuevo un mensaje, un pequeño grupo puede dominar el discurso. Permitir que más usuarios tengan una contribución igual es un componente clave para un diálogo abierto que debería ser apoyado. Por lo tanto, en las situaciones donde una distribución uniforme está en peligro, se han explorado estrategias para mitigar los efectos en este trabajo. El modelo explora cómo un conjunto de dinámicas establecidas para el intercambio de mensajes en las redes sociales en línea, junto con un posible intercambio de incentivos para la difusión de contenido, puede proporcionar una estructura de recompensa para los *influencers* para introducir la polarización. Las simulaciones muestran cómo la polarización puede permitir que estas personas, si poseen un alto rango, mantengan su dominio social y obtengan una clasificación relativa de los seguidores alineados, lo que, como producto secundario, reduce la capacidad de otros para difundir su propio contenido. Dentro de estas simulaciones, se explora un conjunto de enfoques, en los cuales estos efectos se reducen y se revierten de manera efectiva.

Se proponen dos estrategias de inoculación que no dependen de que el contenido enviado sea visible para inspección o contenga otros detalles para la preservación de la privacidad de los usuarios. La característica clave en ambas estrategias es que se omite el factor de «intimidación», representado por el contenido polarizado de alto perfil. Se omite por medio de la estrategia de inoculación 1, debido a la no participación de los nodos del extremo inferior, que tienen poco que ganar en términos de *klout* y que puede inducirse mediante campañas de recomendación. También, sirve la estrategia de inoculación 2, que coloca a los usuarios en niveles de puntajes de influencia similares, independientemente de los lados polarizados. Se ha demostrado que estas dos estrategias de inoculación son aplicaciones efectivas de enfoques no invasivos y de preservación de la privacidad que ofrecen una vía para reducir la polarización. Durante las discusiones no polarizadas, o con las estrategias de inoculación vigentes, los usuarios que influyen en el nivel inferior pueden «superponerse» y, durante la «polarización», la intimidación juega un papel importante. La dinámica de polarización no se ve afectada o abordada directamente, ya que esto puede introducir complicaciones imprevistas e inducir posibles efectos contrarios a los esperados durante los intercambios agitados [22].

La dinámica para la producción de este efecto puede usarse para identificar reducciones en las capacidades de difusión de mensajes para la mayoría de los miembros de las redes sociales y para conectar esto con los líderes de discusión que utilizan contenido polarizado para aumentar su influencia relativa. El trabajo futuro implica considerar un marco similar para los conocidos «escándalos de celebridades» y cómo estos pueden elevar, en gran medida, la popularidad o la influencia de un miembro «icónico» de la sociedad. Esto puede requerir una definición y manipulación de lo que se considera un peligro para la identidad de una persona. Una extensión también podría beneficiarse al tratar con usuarios que pueden tener más o menos una serie de enlaces para crear mensajes directos a los usuarios.

METODOLOGÍA

Los intercambios de información que tienen dependencias temporales dependen de una serie de efectos de golpe. Se supone que estos eventos de una respuesta se generan gracias a un criterio de decisión local y que no son conscientes del efecto de campo medio del impacto de contenido idéntico en otra parte de la red. Las consideraciones individuales con respecto a la discreción de un nodo con respecto a la propagación de un mensaje que recibe o no constituyen un proceso complejo. El comportamiento

principal del efecto de golpe es el foco principal. La parametrización para estas decisiones se representa como un parámetro de cada nodo, que es su *valor de influencia* clasificado entre los otros nodos; l_n: $0 < l_1 \leq l_2 \leq \ldots \leq l_N$. Aunque esto puede parecer simplista, representa una lista clasificada de características de popularidad, influencia y otros efectos, como la autoridad, cuando está directamente relacionada con la respuesta de un evento de difusión de contenido.

Se supone que hay una tasa Basal para la cual los nodos generan nuevo contenido, b, como una probabilidad que es uniforme en toda la red y el tiempo. Estos eventos producen eventos *cb* distribuidos uniformemente a todos los demás nodos. Cada evento se realiza como la producción de un borde dirigido y *cb* es el número de esos bordes dirigidos. Al ser un nodo de destino para un borde, la función de respuesta de probabilidad se define para un nodo en un punto de tiempo *k*, como el siguiente:

$$r_n^{[k]} := \frac{\sum_{i=1}^{N} l_i (A^{[k]})_{(i,n)}}{1 + l_N \sum_{i=1}^{N} (A^{[k]})_{(i,n)}}. \quad \text{(Ecuación 1)}$$

Cada vez que responde un nodo, se produce una distribución exitosa de contenido relacionado supuesto ($s_n^{[k]}$), lo que crea enlaces c_r, elegidos uniformemente en toda la red. Implícito en este efecto de respuesta hay una dependencia temporal de la asociación de contenido. Está conectado con la motivación para que los miembros de la red obtengan recompensas por compartir información importante con otros miembros de manera oportuna.

Agregamos al modelo un mecanismo de recompensar la participación para estimular una respuesta exitosa en otro nodo. Esto se logra al tener un incremento en el valor de influencia de un nodo. Al aumentar el valor l_n, esto dará como resultado respuestas de mensajes más exitosas en futuros bordes dirigidos producidos. Imponemos que $e_{i,j}$, $i \notin \mathbf{j}$, $|\mathbf{j}| = c_r$ para que se produzcan bordes únicos y que no sean autodirigidos. Cualquier nodo que tenga una ventaja hacia otro que produzca una respuesta en $k + 1$, puede recibir un aumento (incremento) de su valor de influencia. El vector de influencia no se normaliza a través de los pasos de tiempo desde el denominador en la ecuación 1:

$$1 + l_N \sum_{i=1}^{N} (A^{[k]})_{in};$$

Se trata de una operación que proporciona una medida relativa. Además, no se sabe si las plataformas de redes sociales en general escalen el análisis de popularidad (números de amigos, números de *retweet*, etc.) según un agregado de red que es visible para los usuarios.

Un nodo de bajo rango, en términos de influencia percibida, incurre en el mismo incremento sobre su valor *ln* cuando induce una respuesta a través de un borde dirigido. En la práctica, esto puede diferir en cómo otros nodos ven los mensajes entrantes (flujos de contenido), ya sea digitalmente o mediante interacciones sociales tradicionales. El incremento se atribuirá de acuerdo con las respuestas exitosas inducidas, y lo hará uniformemente, denotado como ($s_n^{[k]}$), con la actualización:

$$l_{i'} = (s_n^{[k]} \times (A^{[k]})_{in} + l_i) \quad \text{(Ecuación 2)}$$

Esto proporciona la motivación para la actividad de intercambio de contenido entre usuarios y el esfuerzo para participar en la influencia sobre otros nodos en la red. Estos incrementos permitirán a un usuario tener un mayor rango de exposición de contenido y comentarios de reconocimiento.

Se puede esperar que en muchas iteraciones la diferencia máxima inicial en la influencia del nodo *max* (**l**) - *min* (**l**) sea menos significativa, ya que los incrementos son uniformes entre los usuarios. Se puede suponer, asimismo, que esta diferencia permanece constante en múltiples simulaciones

y, si *max*(l) - *min*(l) = *diff* se considera aproximadamente una constante, este valor para *k* grande se vuelve insignificante en comparación con el valor de influencia mínimo de un estado de red futuro $min(l_i) >> diff$.

Con una adición uniforme de influencia que se acumula debido al «*piggy-backing*» o «*mirroring*», desde el contenido y el *retweeting* oscuro (eliminando las etiquetas de autoría originales), se anticipa la minimización de los puntos de partida iniciales. Esto refleja la idea de un hipotético destino de estado «ergódico» uniforme para una red de usuarios con visibilidad completa durante el tiempo suficiente.

Para representar una fase de polarización, cada nodo se coloca en uno de dos grupos diferentes, las probabilidades y los *impares* y *pares*.

Debe haber una cantidad aproximadamente igual de influencia total para los nodos en cada lado. Esta ruptura en la homogeneidad afecta las probabilidades de la función de respuesta al cambiar la influencia percibida de cada nodo de un grupo diferente. Tomando el caso en el que *mod*(*n*, 2) = 0, *n* está en el grupo etiquetado «par». Las contribuciones de influencia polarizadas para su función de respuesta se convierten en:

$$\mathbf{l}_{even} = \begin{cases} l_i & i \text{ is even} \\ l_i \times (-1) & i \text{ is odd.} \end{cases} \quad \text{(Ecuación 3)}$$

Las columnas respectivas en la matriz de adyacencia utilizadas por cada cálculo de la función de respuesta se convierten en:

$$r^{[k]}_{n \in even} := \frac{\sum_{i=1}^{N} \left((A^{[k]})_{(i,n)} \cdot \mathbf{l}_{even} \right)}{1 + l_N \sum_{i=1}^{N} (A^{[k]})_{(i,n)}} \quad \text{(Ecuación 4)}$$

Aún se toma en cuenta el intercambio de las etiquetas necesarias para el caso impar. En la sección «Resultados», se explora tener esta ponderación positiva dentro del grupo y la negativa fuera del grupo, mientras se mantiene la capacidad de los nodos para enviar mensajes a través de enlaces transitorios.

ESTRATEGIA DE INOCULACIÓN 1

La extensión metodológica proporciona un método en el que los miembros de la red de intercambio de mensajes pueden aumentar su capacidad de difundir contenido a través de otros miembros. Esto está representado por sus valores de influencia, l_n, y el estado de polarización cambiará inevitablemente la distribución entre los nodos. Se describe un enfoque que busca revertir los valores de influencia relativa acumulados para producir una distribución de respuesta más uniforme. A partir de enfoques como el presentado en [20], que desarrollan un marco para reducir el efecto de las «cajas de resonancia», se puede ver que la recomendación de enlace que se centra en desviar la participación antagónica entre las partes en función del contenido es una estrategia potencialmente efectiva. Mirando los resultados experimentales de [37], que examina un estudio de caso de exposición a lados opuestos de un argumento, se propone una metodología que sigue estas instrucciones y tiene en cuenta las limitaciones prácticas.

Dado que la ecuación 4 solo tiene en cuenta los valores relativos de l_n para la probabilidad de disparo debido a la influencia externa, la estrategia de inoculación altera la asignación de los valores

de influencia en las etapas de polarización de la ecuación 3. La alteración es «neutralizar» el contenido que llega de los nodos en la mitad inferior del rango de influencia de la red:

$$\mathbf{l}_{even} = \begin{cases} l_i & i \text{ is even} \vee r_{l_i} < \lfloor {}^{N}/_{2} \rfloor \\ l_i \times (-1) & i \text{ is odd} \wedge r_{l_i} > \lfloor {}^{N}/_{2} \rfloor \end{cases} \quad \text{(Ecuación 5)}$$

Aquí, r_{li} es el rango de l_i entre otros nodos. Esta es una alteración en la ecuación 3 y modela el efecto de que solo los nodos considerados con la mitad superior de los rangos de influencia participan en el intercambio de contenido que se considera polarizante. Por lo tanto, los mensajes entre los lados polarizados que no están en la mitad superior del rango de influencia pueden tener el mismo tratamiento en el valor del contenido para la propagación.

Una aplicación práctica de esta estrategia en una plataforma real de redes sociales requiere una campaña de información de contenido promocionado para crear conciencia sobre la falta de retorno ante un desacuerdo continuo. Dado que estos usuarios pueden no participar proporcionalmente en un «intercambio acalorado» de contenido, esta situación puede ponerse al día y permitir que regrese el estado anterior. Esto depende de una financiación para una campaña, o incluso de que no haya antagonismos arraigados entre las partes. Por lo tanto, la estrategia de inoculación 2 (subsección 4.2) proporciona un enfoque que no depende de la comprensión del usuario más la buena voluntad, sino de la política de la plataforma para participar en un esfuerzo por reducir la polarización.

ESTRATEGIA DE INOCULACIÓN 2

En contraste con la estrategia propuesta para la estrategia de inoculación 1 (Subsección 4.1), esta aborda el problema de que los participantes pueden no apreciar la campaña de información dirigida a reducir los efectos de la polarización en sus éxitos de comunicación. Si bien se mantiene el requisito de que se respete la privacidad del contenido, este enfoque requiere la participación de los administradores o desarrolladores de la plataforma. El concepto detrás de la estrategia es establecer un conjunto de niveles basados en los valores de influencia. Que los usuarios estén expuestos a mensajes solamente dentro de esos niveles tampoco se separa de acuerdo con la participación de polarización. Esto está dirigido directamente a reducir el factor de «intimidación» sin eliminar el concepto de competencia de influencia que impulsa gran parte de los intercambios. Al segmentar la distribución de influencia en niveles, los usuarios tienen una barrera de entrada más baja en la conversación, pues tienen diferencias más pequeñas en los valores de influencia percibidos del contenido en función del historial de un usuario de producir diseminaciones.

Aquí modificamos la ecuación de respuesta de la ecuación 1 para representar la probabilidad de difusión al recibir contenido de otros miembros de la red dentro de su nivel predefinido. No se supone que los niveles se calculen de forma «en línea» cuando exista un mecanismo de actualización continua, pero que haya períodos en los que esto se pueda reevaluar. La localidad para las simulaciones realizadas aquí se basa en segmentar la red en 5 niveles (inspirados en la regla 80-20) donde $T = N/5$ representa el tamaño del nivel. La ecuación de respuesta (ecuación 6) se define como:

$$r^{[k]}_{n \in even, m = \lfloor {}^{n}/_{\tau} \rfloor} := \frac{\sum_{i=\tau \cdot m+1}^{N' = min(N, \tau(m+1))} (A^{[k]})_{(i,n)} \cdot \mathbf{l}_{even}}{1 + l_{min(N,\tau(m+1))} \sum_{i=\tau \cdot m+1}^{N' = min(N, \tau(m+1))} (A^{[k]})_{(i,n)}} \quad \text{(Ecuación 6)}$$

La introducción del término, $m = [n/T]$ representa el número de nivel del que es miembro el usuario. Un posible desacuerdo de los usuarios que experimentarán este cambio de plataforma puede ser que es una restricción a la libertad de asociación uniforme. Una perspectiva alternativa sobre la que los autores solo especulan es cómo se puede abordar eso es si se puede convertir en un método de *gamificación* [38], [39], en el que los niveles brindan la capacidad de esforzarse hacia otro nivel.

REFERENCIAS

[1] B. Batrinca y P. Treleaven, "Social media analytics: a survey of techniques, tools and platforms", *Ai & Society,* vol. 30, pp. 89-116, 2015.

[2] W. Fan y M. D. Gordon, "The power of social media analytics", *Commun. Acm,* vol. 57, pp. 74-81, 2014.

[3] K. Lindgren y M. G. Nordahl, "Cooperation and community structure in artificial ecosystems", *Artificial Life,* vol**.** 1, pp. 15-37, 1993.

[4] F. Shi, M. Teplitskiy, E. Duede y J. A. Evans, "The wisdom of polarized crowds", *Nature human behavior*, vol. 3, pp. 329, 2019.

[5] M. P. Fiorina y S. J. Abrams, "Political polarization in the american public", *Annu. Rev. Polit. Sci.,* vol 11, pp. 563-588, 2008.

[6] J. N. Druckman, E. Peterson y R. Slothuus, "How elite partisan polarization affects public opinion formation", *American Political Science Review*, vol. 107, pp. 57-79, 2013.

[7] M. Gentzkow, "Polarization in 2016", *Toulouse Network for Information Technology Whitepaper,* 2016.

[8] J. E. Campbell, *Polarized: Making sense of a divided America.* Princeton: Princeton University Press, 2018.

[9] M. Gentzkow, J. Shapiro, M. Taddy *et al.* "Measuring polarization in high-dimensional data: Method and application to congressional speech", *Tech. Rep*., 2016.

[10] S. R. Chakravarty *et al.* « Inequality, polarization and poverty". *Advances in distributional analysis*. *New York,* 2009.

[11] T. Mogues y M. R. Carter, "Social capital and the reproduction of economic inequality in polarized societies", *The Journal of Economic Inequality*, vol. 3, pp. 193-219, 2005.

[12] J.E. Foster y M. C. Wolfson, "Polarization and the decline of the middle class: Canada and the us", *The Journal of Economic Inequality*, vol 8, pp. 247-273, 2010.

[13] M. Büchi y F. Vogler, "Testing a digital inequality model for online political participation". *Socius*, vol 3, doi: 2378023117733903, 2017.

[14] D. M. Blei, A. Y. Ng y M. I. Jordan, « Latent dirichlet allocation », *Journal of machine Learning research,* vol 3, pp. 993-1022, 2003.

[15] M. Allahyari *et al.* "Text summarization techniques: a brief survey", *arXiv preprint arXiv:1707.02268,* 2017.

[16] Q. V. Liao, y W. T. Fu, "Can you hear me now?: mitigating the echo chamber effect by source position indicators", *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 184-196, ACM, 2014.

[17] Q. V.Liao, y W. T. Fu, "Expert voices in echo chambers: effects of source expertise indicators on exposure to diverse opinions", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2745-2754, ACM, 2014.

[18] V.V. Vydiswaran, C. Zhai, D. Roth y P. Pirolli, "Overcoming bias to learn about controversial topics", *Journal of the Association for Information Science and Technology*, vol. 66, pp. 1655-1672, 2015.
[19] S. A. Munson, S. Y. Lee y P. Resnick, "Encouraging reading of diverse political viewpoints with a browser widget", *Seventh International AAAI Conference on Weblogs and Social Media,* 2013.
[20] K. Garimella, G. De Francisci Morales, A. Gionis y M. Mathioudakis, "Reducing controversy by connecting opposing views", *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 81-90, ACM, 2017.
[21] E. Graells-Garrido, M. Lalmas y D. Quercia, "Data portraits: Connecting people of opposing views", *arXiv preprint arXiv:1311.4658,* 2013.
[22] S. Jackson, "The double-edged sword of banning extremists from social media", *SocArXiv,* 2019.
[23] D. Stasavage, "Polarization and publicity: rethinking the benefits of deliberative democracy", *The Journal of Politics*, vol. 69, pp. 59-72, 2007.
[24] M. Zimmer, "But the data is already public: on the ethics of research in Facebook". *Ethics and information technology,* vol. 12, pp. 313-325, 2010.
[25] P. Grindrod, M. C. Parsons, D. J. Higham y E. Estrada, "Communicability across evolving networks", *Physical Review E*, vol. 83, 046120, 2011.
[26] A. Gandini, "Digital work: Self-branding and social capital in the freelance knowledge economy", *Marketing theory*, vol. 16, pp. 123-141, 2016.
[27] K. T. Smith, J. Blazovich, y L. M. Smith, "Social media adoption by corporations: An examination by platform, industry, size, and financial performance", *Academy of Marketing Studies Journal*, vol. 19, pp. 127-143, 2015.
[28] A. V. Mantzaris y D. J. Higham, "A model for dynamic communicators", *European Journal of Applied Mathematics* vol. 23, pp. 659-668, 2012.
[29] A. V. Mantzaris y D. J. Higham, "Dynamic communicability predicts infectiousness", *Temporal Networks*, pp. 283-294, Springer, 2013.
[30] P. Laflin *et al*. "Discovering and validating influence in a dynamic online social network", *Social Network Analysis and Mining*, vol. 3, pp. 1311-1323, 2013.
[31] A. Rao, N. Spasojevic, Z. Li y T. Dsouza, "Klout score: Measuring influence across multiple social networks", *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2282-2289, IEEE, 2015.
[32] M. Del Vicario, A. Scala, G. Caldarelli, H. E. Stanley, y W. Quattrociocchi, "Modeling confirmation bias and polarization", *Scientific reports*, vol. 7, 40391, 2017.
[33] B. Gonçalves, N. Perra y A. Vespignani, "Modeling users' activity on twitter networks: Validation of dunbar's number", *PloS one,* vol. 6, e22656, 2011.
[34] R. I. Dunbar, "Coevolution of neocortical size, group size and language in humans", *Behavioral and brain sciences*, vol. 16, pp. 681-694, 1993.
[35] L. A. Adamic y N. Glance, "The political blogosphere and the 2004 us election: divided they blog", *Proceedings of the 3rd international workshop on Link discovery*, pp. 36-43, ACM, 2005.
[36] E. Hargittai, J. Gallo y M. Kane, "Cross-ideological discussions among conservative and liberal bloggers", *Public Choice,* vol. 134, pp. 67-86, 2008.
[37] C. A. Bail *et al*. "Exposure to opposing views on social media can increase political polarization", *Proceedings of the National Academy of Sciences,* 201804840, 2018.
[38] R. N. Landers y R. C. Callan, "Casual social games as serious games: The psychology of gamification in undergraduate education and employee training", *Serious games and edutainment applications*, pp. 399-423, Springer, 2011.
[39] S. Deterding, "Gamification: designing for motivation", *Interactions,* vol. 19, pp. 14-17, 2012.

RECONOCIMIENTOS

Esta investigación fue apoyada por el programa DARPA HR001117S0018. Los financiadores no tuvieron ningún papel en el diseño del estudio, la recopilación y el análisis de datos, la decisión de publicar o la preparación del manuscrito.

CONTRIBUCIONES DE CADA AUTOR

I.G. dirigió, además de coordinar, el esfuerzo de investigación y definió los objetivos de las investigaciones. A.V.M. exploró la definición del modelo y la redacción del borrador en papel. A.V.M., A.R. y C.E.T. trabajaron en la exploración y simulación de modelos. Todos los autores interpretaron los resultados y redactaron el documento.

CONFLICTO DE INTERESES

Los autores declaran no tener conflictos de intereses.

INFORMACIÓN ADICIONAL

La información complementaria está disponible en https://doi.org/10.1038/s41598-019-55178-8.
La correspondencia y las solicitudes de materiales deben dirigirse a I.G.
La información sobre reimpresiones y permisos está disponible en www.nature.com/reprints.

Nota del editor: Springer Nature se mantiene neutral con respecto a las reclamaciones jurisdiccionales en mapas publicados y afiliaciones institucionales.