CELE

WhatsApp, política móvil y desinformación: ¿cómo se dio a viralización de las noticias falsas en las elecciones brasileñas?

Mayo 2019

Facultad de Derecho
Centro de Estudios en Libertad
de Expresión y Acceso a la Información

UP
Universidad
de Palermo

WhatsApp, política móvil y desinformación: ¿cómo se dio la viralización de las noticias falsas en las elecciones brasileñas?

*João Guilherme Bastos dos Santos y Miguel Freitas**

I. Introducción

WhatsApp consiste en un diseño de red privada con encriptación punta a punta, pero los lazos interpersonales y su estructura de red en Brasil lo transformaron en una poderosa herramienta de difusión de informaciones para grandes públicos. Como alternativa a los servicios de SMS pagos, la posibilidad de utilizar servicios de mensajería móvil sin gastar datos de internet –en Brasil existen acuerdos entre operadoras de telefonía móvil para que no realicen costos al usuario– atrajo a un contingente grande de personas que no tienen acceso a la red de otro modo, lo que ayudó a la aplicación a alcanzar la cifra de 120 millones de usuarios activos en 2018. En un contexto en el que los sitios de *fact-checking* exigen contar con acceso a internet y mientras que las empresas de medios de comunicación restringen el acceso a sus productos, algunos partidos políticos han sabido utilizar esta coyuntura a su favor al difundir información con sesgo electoral en momentos clave.

Un caso emblemático ocurrió en vísperas de las elecciones presidenciales de 2014, cuando se propagó que Alberto Youssef había sido envenenado durante su arresto en la Superintendencia de la Policía Federal en Curitiba. La noticia se viralizó por WhatsApp y alcanzó smartphones del país entero en menos de 24 horas antes de la elección. Un montaje colocaba el titular en el portal de noticias G1, junto al rumor de que estaría implicado el Partido de los Trabajadores en el crimen, supuestamente para impedir una denuncia de Youssef. El rumor fue desmentido públicamente por la Policía Federal brasileña y su circulación por el G1 fue condenada por el ministro de Justicia, quien dio cuenta de la preocupación de diversos actores sobre las consecuencias de esta mentira en un escenario electoral agudo parejo y polarizado, y sobre la dificultad de dar una respuesta proporcional, veloz y eficiente al problema.

El impacto de este tipo de estrategia habría tenido un potencial mucho mayor en 2018. De acuerdo con Digital News Report 2018,[1] en 2014 el uso de smartphones para consumo de noticias en Brasil involucraba el 35% de la población, mientras que el 64% utilizaba computadoras. De acuerdo con el mismo informe en el año 2018, el 65% de los brasileños utiliza smartphones para informarse, el 62% su computadora y el 46% WhatsApp. Las autoridades y periodistas, sin embargo, subestimaron considerablemente el uso de la aplicación que permite repetir y multiplicar, de modo mucho más sistemático y eficaz, la circulación de noticias similares a las descriptas en el "caso Youssef". Esta particularidad convirtió a WhatsApp mucho más atractiva para las campañas electorales, especialmente al considerar su aspecto económico.

---

* João Guilherme Bastos dos Santos es investigador de postdoctorado en el Instituto Nacional de Ciencia y Tecnología en Democracia Digital en Brasil. Doctor en Comunicación por la Universidad del Estado de Río de Janeiro, con una pasantía de doctorado en la Universidad de Leeds en Reino Unido. Miguel Freitas es Doctor en Ingeniería Eléctrica de la Pontificia Universidad Católica de Río de Janeiro y tiene un máster en Ingeniería Eléctrica del Centro de Estudios de Telecomunicaciones - CETUC.

[1] Newman, Nic, Fletcher, Richard, Kalogeropoulos, Antonis, Levy, David A.L. y Nielsen, Rasmus Kleis, "Reuters Institute digital news report", Reuters Institute for the Study of Journalism, Oxford, University of Oxford, 2018.

Es importante resaltar que este uso de la plataforma es una distorsión específica. Las aplicaciones de comunicación seguras utilizan un cifrado de extremo a extremo como herramienta importante en la defensa del derecho a la privacidad. La innovación de la criptografía está en permitir que cualquier usuario pueda encriptar sus mensajes en su propio móvil con una clave de cifrado segura, que lo protege incluso de agencias de espionaje gubernamentales, las cuales difícilmente serían capaces de decodificar su contenido. El mensaje cifrado se descodifica solo en el dispositivo del destinatario, por lo que es inmune a las interceptaciones que pueden ocurrir durante su tráfico por la red. Esta misma dinámica, no obstante, exige medidas específicas para evitar la viralización sistemática de noticias falsas con fines de manipulación electoral, un crimen según el artículo 323 del Código Electoral brasileño.

Aunque esta tecnología esté disponible de forma pública y gratuita desde 1991,[2] diversos motivos explican por qué su adopción quedó restringida a nichos tecnológicos o de activistas. El caso Snowden, revelado por el periodista Glenn Greenwald en 2013, mostró cómo internet estaba siendo utilizada para burlar mecanismos legales de protección personal y monitorear indiscriminadamente a ciudadanos estadounidenses y extranjeros.[3] Este evento llamó la atención sobre la necesidad de proveer mecanismos para que los ciudadanos puedan defenderse del abuso de Estados que no respeten sus derechos individuales. A raíz de este contexto, el Gobierno brasileño transformó en prioridad la aprobación de la ley conocida como Marco Civil de Internet,[4] que trajo como piedra fundamental el principio de la privacidad.

Diferentes aplicaciones de mensajería instantánea han ganado popularidad con la difusión de los teléfonos inteligentes en los últimos años. Para destacar tres de los más relevantes actualmente, WhatsApp, Telegram y Signal fueron lanzados, respectivamente, en 2009, 2013 y 2014. A diferencia de los dos últimos, WhatsApp no tenía, al momento de su creación, la encriptación de extremo a extremo. Fue probablemente una decisión comercial, movida por la presión de los competidores post Snowden y de una demanda de mercado, que lleva a WhatsApp finalmente a adoptar este recurso para todos los usuarios en 2016.

Ya se han hecho revisiones acerca de los desafíos del término "falsas noticias y desinformación". Este puede referirse tanto a la propaganda política centrada en una información sesgada y a los sentimientos de hostilidad, como también a la creación de noticias y la manipulación de imágenes, incluso en otras ocasiones a un repertorio de acción política.[5] Otro punto relevante en este debate es el entendimiento de que las advertencias contra las "falsas noticias" también son movilizadas estratégicamente por diferentes actores que buscan invalidar las críticas de sus oponentes.[6] Nos interesa la concepción de las noticias falsas como un repertorio de la acción política, movilizada por los agentes políticos y los ciudadanos simpatizantes. Si estos son capaces de hacerlas virales, pueden lograr ventajas electorales en escenarios polarizados. En los últimos años, WhatsApp se ha convertido en una poderosa herramienta para este tipo de acción.

---

2 Lauzon, Elizabeth, "The Philip Zimmerman investigation: the start of the fall of export restrictions on encryption software under first amendment free speech issues", Syracuse Law Review, Nº 48, 1998, p. 1.307.

3 Greenwald, Glenn, *No place to hide: Edward Snowden, the NSA and the US surveillance state,* Nueva York, Picador-Macmillan, 2014; Freitas, Miguel, "Twister: the development of a peer-to-peer microblogging platform", *International Journal of Parallel, Emergent and Distributed Systems,* Nº 1, Vol. 31, 2016, pp. 20-33.

4 Presidencia de la República, Casa Civil, Subsecretaria para Asuntos Jurídicos, "Ley Nº 12.965, de 23 de abril de 2014", 2014, disponible en: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm, último acceso: 20 de noviembre de 2019.

5 Mendonça, Ricardo Fabrino y Freitas, Viviane Gonçalves, "Fake news e o repertório contemporâneo de ação política", VIII Congresso da Associação Brasileira de Pesquisadores em Comunicação e Política (VIII COMPOLÍTICA), Anales, Brasilia, Universidad de Brasilia, 15 al 17 de mayo, 2019.

6 Dourado, Tatiana y Gomes, Wilson, "O que são, afinal, fake news, enquanto fenômeno de comunicação política?", VIII Congresso da Associação Brasileira de Pesquisadores em Comunicação e Política (VIII COMPOLÍTICA), Anales, Brasilia, Universidad de Brasilia, 15 al 17 de mayo, 2019.

La apropiación exitosa de WhatsApp por los partidarios del Partido Social Liberal (PSL) resulta de la cooperación de diversos grupos –incluidos los votantes– y un conocimiento específico sobre la viralización sistemática de contenido. La comprensión de este fenómeno exige el reconocimiento de, al menos, dos elementos:

i) Un cambio en la forma en que se analiza esta aplicación –que supera los obstáculos relacionados a su diseño orientado al intercambio de mensajes privados– a través del cruzamiento de datos, composición de redes y análisis de flujo de contenido.

ii) La necesidad de indicar por qué el PSL obtuvo ventaja en este campo y reconocer una estrategia política que logró articular la heterogeneidad de los actores involucrados y la historicidad de sus interconexiones, al considerar los momentos en que estas técnicas fueron desarrolladas y mejoradas.

Todos estos puntos mantienen relación con trabajos conducidos desde 2011 por investigadores del grupo Tecnologías de la Comunicación y Política (TCP) de la Universidad del Estado de Río de Janeiro, vinculados al Instituto Nacional de Ciencia y Tecnología en Democracia Digital (INCTDD). En el año 2017, el grupo condujo una serie de entrevistas con profesionales de internet y campaña política, e incluyeron el tema específico de WhatsApp y noticias falsas. Estas entrevistas fueron utilizadas como base para el desarrollo de metodologías capaces de analizar el uso de WhatsApp en la diseminación de informaciones falsas y definir cuáles dinámicas hacen posible la viralización en este ambiente que no es visible públicamente. El proyecto de análisis inicial fue debatido y mejorado con la colaboración de Artur Ziviani, profesor e investigador del Laboratorio Nacional de Computación Científica (LNCC).

El uso de WhatsApp para la difusión de noticias falsas puede resultar previsible, pero el éxito de esta estrategia pareciera contar con aspectos contraintuitivos y de difícil comprensión, prácticamente un punto ciego en todas las investigaciones del área hasta entonces. Por un lado, es previsible mantener la fuente relativamente segura y hacer difícil su rastreo –escondido del escrutinio público general– pero en contacto con el segmento objetivo en particular. La segmentación es una posibilidad, incluso si la aplicación no ofrece estos datos: en versiones más sofisticadas, estos pueden ser obtenidos por algoritmos que cruzan datos de diferentes redes en línea, lo que mejora las posibilidades de su obtención en cada una de las redes tomadas individualmente. En las versiones más simples, el cruce puede darse, por ejemplo, entre números de teléfonos celulares y códigos postales –lo que implica la segmentación geográfica por calle y la información demográfica asociada– o extraído de páginas que difunden enlaces a grupos de WhatsApp en redes visibles como Facebook (los cuales pueden ser buscados directamente y cuentan con información temática asociada). Al replicar contenido de modo difícilmente rastreable y al anonimizar su fuente, WhatsApp da un paso adelante en relación a herramientas como *dark posts*[7] y *microtargeting*,[8] muchas veces utilizados por productores de noticias falsas en redes sociales en línea.

Como ha sido señalado por estudios sobre campañas políticas en línea con aumento repentino de simpatizantes,[9] la identificación de rasgos específicos de las personas o grupos de personas puede indicar mayor o menor propensión a compartir contenidos, lo que permite inferir la probabilidad de viralización. Como lo demuestra el escándalo de Cambridge Analytica, los datos sobre los rasgos de personalidad podrían extraerse del comportamiento registrado en línea en las redes sociales. Los llamados robots pueden actuar, por lo tanto, a través de la

[7] *Dark posts* son publicaciones con un amplio alcance direccionado pero que no se verán en el feed público y no serán expuestas al tráfico orgánico de la página en Facebook.

[8] *Microtarget* es el uso de los datos en línea con el objetivo de identificar segmentos muy específicos del publico y hacer campañas enfocadas en ellos.

[9] Margetts, Helen, John, Peter, Hale, Scott A. y Yasse Ri, Taha, *Political turbulence: how social media shape collective action,* Princeton y Oxford, Princeton University Press, 2016.

identificación de nichos y del envío regular de mensajes, al utilizar la propensión de estos nichos a compartir un tipo específico de información al mismo tiempo que crea una apariencia de campaña orgánica.

Por otro lado, el éxito de una acción política que se amplía rápidamente en WhatsApp es inesperado porque este limita el número de personas en cada grupo –no se agrega información social acumulada en los mensajes reenviados–. En Brasil, había un límite de reenvíos permitidos (veinte reenvíos por acción en 2018). Esto no cuenta con algoritmo de visibilidad ni herramientas de microsegmentación pagas. En resumen, todos los elementos que se consideran esenciales para la escalada rápida de los contenidos o viralización están ausentes.

Para que una información se vuelva viral sin algoritmos que aumenten su visibilidad, se necesitan personas dispuestas a compartir el mensaje cuando lo reciben, para llegar a más personas dispuestas a reenviarlo y así sucesivamente. Esto resulta en una dinámica capaz de aumentar exponencialmente la visibilidad de mensajes falsos. Sin embargo, los límites de reenvío dificultan que esta dinámica logre éxito solamente con las tasas normales de reenvío entre contactos privados individuales. En este punto, los grupos de WhatsApp dedicados a la política –incluidos los posibles receptores que pueden albergar a más de doscientos cincuenta contactos cada uno–, muchos de ellos segmentados según intereses políticos e interconectados con otros grupos, son esenciales para comprender cómo se expande la información política.

Teniendo en cuenta estos aspectos y fundamentados en estudios previos sobre el campo, probamos tres hipótesis que involucran la desinformación a gran escala vía WhatsApp, durante la campaña para las elecciones presidenciales de 2018:

H1) La lógica en red bipartita de WhatsApp, debido a su estructura de grupos interconectados, permite que la desinformación se viralice.

H2) Existen coincidencias entre la centralidad de las redes y la revelación de los grupos en ese proceso asimétrico. Esta hipótesis que utiliza métricas de algoritmo permite avanzar en nuestro análisis sobre el alcance viral de la desinformación.

H3) Se trata de un proceso en el cual la información errónea va de los nodos centrales a los periféricos, y se amplía exponencialmente a través de reenvíos entre grupos a lo largo del tiempo.

H4) Conocer los grupos en que la información falsa fue circulada primero puede dejarnos más cerca de la identificación de los autores, lo que posibilita las alianzas con esferas dedicadas al perfeccionamiento de la legislación sobre campañas políticas y la responsabilización de los posibles productores profesionales involucrados. Esto es particularmente posible en los casos en que los reenvíos involucrados en el proceso de viralización llevan indicios de su emisor inicial.

II. H1: WhatsApp como una red bipartita

Con base en la estructura interconectada de estos grupos, probamos la hipótesis orientadora de nuestros análisis, la cual fue confirmada en todo el trabajo de campo realizado durante el período electoral: más que una red de personas conectadas a través de grupos, WhatsApp está sujeto a dinámicas de redes bipartitas,[10] en las que

[10] Una red bipartita es una red en que los grupos no están formalmente asociados entre sí, pero que en la cual sus participantes en común pueden constituir conexiones y fomentar dinámicas de red.

grupos interconectados por participantes en común regulan el intercambio de datos, permiten un aumento exponencial de visibilidad y de sus lógicas de difusión viral de noticias dentro de una red sin visibilidad pública. Esta red de grupos mantiene un flujo de información encriptado, oculto al escrutinio público o responsabilidad legal, que es intenso en períodos electorales. Al transitar rápidamente en grupos diversos, estas noticias pueden fomentar olas de reenvíos que invaden nuevas olas de grupos en cada etapa. Este flujo puede mapearse a partir de la aplicación de métodos de constatación estructural y análisis de redes.

Metodológicamente, esta constatación tiene desdoblamientos relevantes. En primer lugar, además de las propiedades "topológicas", el análisis de redes reconoce propiedades "dinámicas" de viralización y "contagio". Las limitaciones de visibilidad y tamaño de grupos alejan a WhatsApp del modelo de red que caracteriza a Facebook –*preferential attachment/scale free*– en el que actores bien conectados poseen una ventaja acumulativa, al atraer más conexiones y concentrar centralidad, que lo acerca a modelos descentralizados. En este caso, el modelo de red presenta una mayor resistencia a los ataques o desactivación de nodos/grupos[11] en comparación con Facebook, lo que significa que la retirada de estos grupos de las redes sociales por decisión judicial afecta poco a poco a la dinámica de la red. En ambientes políticos polarizados, este modelo puede llevar a la composición, intencional o no, de las llamadas redes policéntricas segmentadas e integradas.[12] Este tipo de actuación en red y su asimilación por partidarios de Bolsonaro será abordada a continuación, junto con su heterogeneidad. Esto no impide que, por regla general, una persona esté intencionalmente presente en un número grande de grupos.

La viralización en este escenario ocurre porque el aumento en el establecimiento de conexiones hace más probable que personas de grupos diferentes establezcan puentes, lo que lleva a un aumento abrupto en la cantidad de personas conectadas indirectamente cada vez que un nuevo grupo o conjunto de grupos se conectan. Los límites a la cantidad de personas en cada grupo conducen a la creación de otros grupos y potencian la segmentación que caracteriza este escenario. Esta dinámica produce un momento crítico a partir del cual los elementos que componen las diversas redes o grupos pasan a estar interconectados. A esta gran red llamamos componente gigante, inicialmente identificado en tres sistemas naturales: difusión de enfermedades epidémicas en redes de contacto físico, redes neuronales y problemas de red de matriz genética.[13] La viralización de informaciones por WhatsApp sigue una lógica de contagio en que la información llega a un grupo y puede contaminar otros cuando los integrantes del grupo la reenvían. Las conexiones entre integrantes de diferentes grupos resultan en una estructura de red en la que los grupos están interconectados. Al identificar integrantes implicados en la difusión de información, logramos analizar la dinámica de circulación de información viral en esta estructura. Eso posibilita un aumento exponencial de visibilidad. Utilizamos este enfoque como alternativa a los análisis centrados en los algoritmos de visibilidad, en los rasgos de personalidad o en puntos de inflexión para acción política.[14] Así, una lógica similar de crecimiento acelerado después de un momento crítico puede estar relacionada a incrementos en la interconexión entre grupos, o sea, alteraciones en la estructura de la red.

Esta cuestión trae la posibilidad de (H2) desarrollar métodos dirigidos a las redes que rápidamente identifican qué grupos están más propensos a estar al inicio del proceso de viralización, así como (H3) los caminos preferenciales para la desinformación segmentada, puntos cruciales que impiden la propagación de esos contenidos los

---

[11] Albert, Réka y Barabási, Albert-László, "Topology of evolving networks: local events and universality", *Physical Review Letters,* Nº 85, 2000, pp. 5.234-5.237.

[12] Gerlach, Luther, "The structure of social movements: environmental activism and its opponents", en: Arquilla, John y Ronfeldt, David, *Networks and netwars: the future of terror, crime, and militancy,* Santa Monica, CA, RAND, 2001.

[13] Rapoport, Anatol y Horvath, William J., "A study of a large sociogram", *Behavioral Science,* Nº 6, 1961, pp. 279-291.

[14] Margetts *et al., op. cit.*

cuales retardan el proceso viral, que lo pueden hacer inviable. Todas las pruebas empíricas del grupo de investigación en Tecnologías de la Comunicación y Política (TCP) en la Universidad del Estado de Río de Janeiro (UERJ) confirmaron esta posibilidad. En mayo, los siete investigadores del TCP involucrados en el experimento iniciaron sus conexiones a partir de la entrada en grupos de WhatsApp en apoyo a candidatos y segmentos diferentes (incluyendo PSL, PT, PDT, REDE, PSDB y PMDB) y terminaron conectados a la misma red. Todas las pruebas empíricas de TCP-UERJ confirmaron esta posibilidad.

En el presente estudio, solo entramos en grupos sin restricciones de acceso por links de invitación, donde cualquiera puede acceder al grupo a través de un enlace. Estos fueron considerados como grupos abiertos.

La descentralización también fue confirmada, al alejarse de dinámicas de redes como Facebook (la distribución de grado indica que 8.354 perfiles están en solo un grupo, 1.107 en dos, 225 en tres, 81 en cuatro, 23 en cinco, 10 en seis, 9 en siete, 2 en ocho grupos y 1 en dieciséis). Los diversos grupos están indirectamente conectados por integrantes en común, que componen una red bipartita, lo que confirma la H1.

La posibilidad de circulación de enlaces, sin embargo, conecta los dos modelos de red, que permite a los usuarios de Facebook la distribución de enlaces a grupos de WhatsApp (con entrada condicionada al límite de miembros) y a perfiles de WhatsApp la distribución de enlaces a publicaciones en Facebook, lo cual canaliza la participación de los miembros, promueve olas de comentarios, *likes* o ataques súbitos sin que la actuación de los grupos que promocionaron esta ola sea visible. Integrantes infiltrados también son utilizados en ataques a grupos adversarios. Estos infiltrados envían links de invitación para que adversarios puedan entrar e insultar o hasta excluir miembros una vez logran tornarse administradores de grupos que atacan.

III. H2 y H3: viralización (sin algoritmo de visibilidad o información social) y el papel de las métricas de red

Reconocer el carácter viral de estas dinámicas exige la aceptación de dos premisas, cuyos despliegues metodológicos son relevantes para la comprensión del uso político de WhatsApp. En primer lugar, la viralización implica direccionalidad (relaciones asimétricas entre una fuente y sus destinatarios) y un proceso variable en el tiempo cuya progresión puede ser evaluada en etapas (y en cuáles destinatarios en un paso pueden convertirse en fuentes en el paso siguiente). Esto plantea la cuestión: ¿cómo identificar rápidamente qué grupos tienen más posibilidades de estar al inicio de este proceso? Al considerar el problema de la viralización de noticias falsas, este es el punto vulnerable de la red, ya que impedir la propagación a partir de estos puntos desacelera y puede inviabilizar la viralización.

Este enfoque es especialmente útil en casos de redes encriptadas como WhatsApp, cuyas características pueden ser válidas para comprender los modelos de red (tomando en cuenta algunas inferencias sobre los patrones que ocurren fuera de nuestro muestreo). Esto es particularmente interesante en un escenario en el que la ausencia de un número total de grupos que integran globalmente esta red prohíbe métodos que dependen de proporciones cuantitativas o representativas. La primera consecuencia importante del reconocimiento de esta asimetría es que los grupos no son equivalentes y, por lo tanto, la simple recolección y cuantificación de sus contenidos ignora las funciones estructurales de la red. Este error, combinado con la falta de medición de la representatividad en esta red privada, puede inviabilizar cualquier generalización de conclusiones cuantitativas.

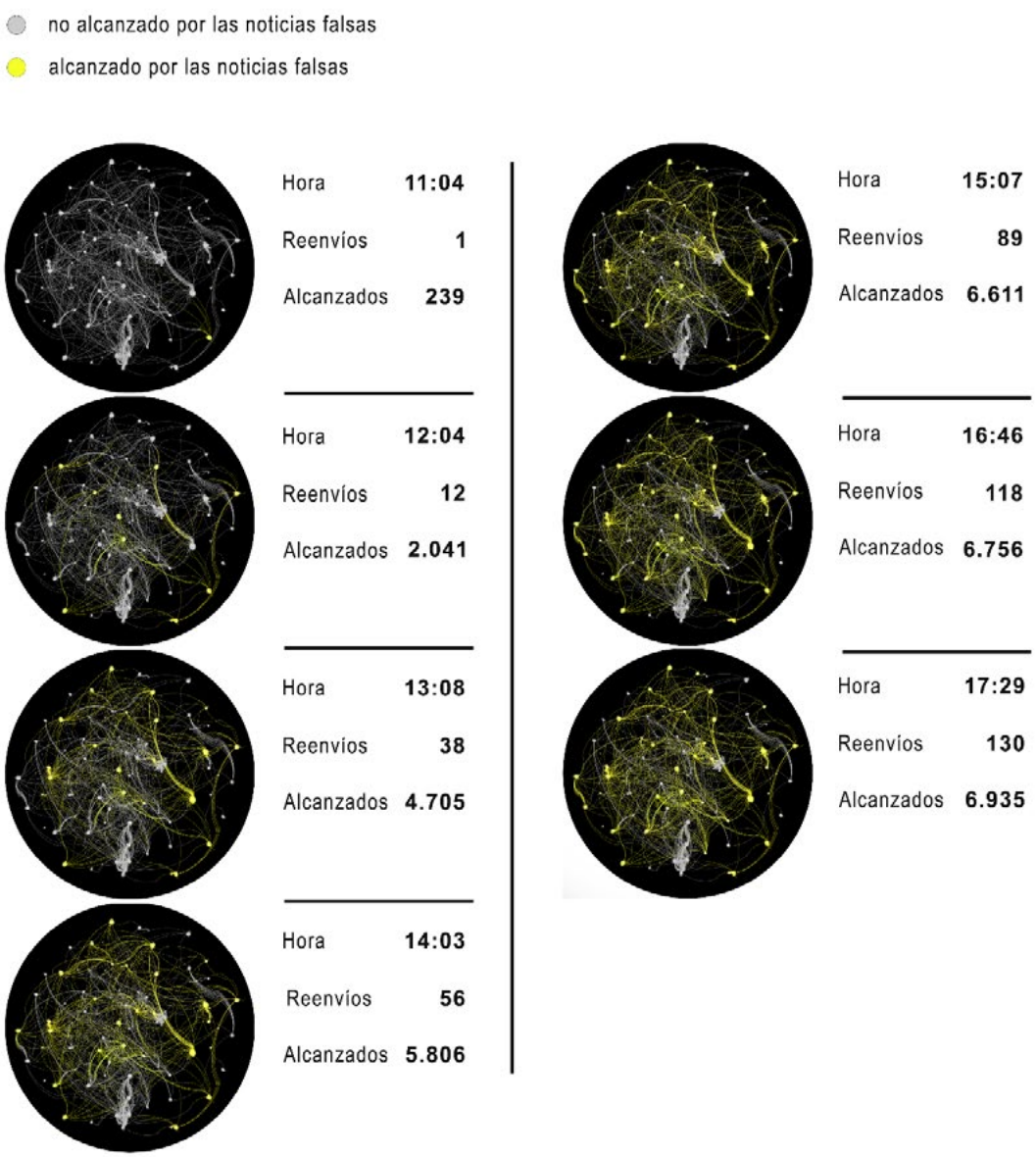
Así, nuestros criterios para entrar en grupos segmentados, de apoyo a los seis candidatos anteriormente mencionados y ya activos a principios de 2018, ocurrieron a través del acceso a enlaces en páginas favorables a esos presidenciables en Facebook. A través de estos enlaces, fuimos automáticamente agregados a otros grupos especializados durante el período preelectoral, lo que nos permitió obtener análisis que varían según el tiempo. En este aspecto, vale resaltar que considerar el tiempo también es relevante en el estudio de las redes, dado que los perfiles no pueden visualizar informaciones publicadas en el grupo antes de sus entradas.

La posición estructural de cada grupo define su relevancia en este proceso: si para circular por diversos puntos de la red una información necesariamente pasa por un grupo, este grupo tiene un nivel de centralidad en esta red, que aumenta cuando otros grupos centrales pasan a estar conectados gracias a este grupo en particular (lo que se denomina *eigenvector*, medida que varía entre 0 para grupos sin centralidad y 1 para grupos con centralidad máxima). Esto también hace que las informaciones lleguen con más probabilidad a este grupo e incrementen sus posibilidades de viralización desde el momento en que lo alcanzan. La criptografía de la fuente y los impedimentos de acceso a contenidos anteriores a la entrada del perfil en un grupo dejan las posibilidades de análisis de este fenómeno restringidas a investigadores que ya lo estuvieran acompañando cuando el fenómeno ocurrió, capaces de cruzar estos datos con informaciones sobre la estructura de la red.

La imagen siguiente muestra la viralización de una noticia falsa sobre el Tribunal Superior Electoral (TSE) en la red descripta arriba, que alcanza 6.935 de los 9.812 perfiles en pocas horas. Los aglomerados de puntos grises indican grupos sin contacto con la noticia, los amarillos muestran a aquellos que tuvieron contacto con ella y las líneas señalan conexiones entre ellos. La noticia afirma que el Tribunal Superior Electoral habría informado la anulación de 7,2 millones de votos y que 2 millones habrían sido necesarios para que el PSL venciese en la primera vuelta. La viralización por compartir trae la posibilidad de analizar las etapas de difusión y encontrar la fuente inicial y el papel de cada grupo en este proceso.

Hay una correlación clara entre la centralidad de red y la velocidad de recepción: entre los diez primeros que recibieron la noticia falsa viralizada, seis tenían una centralidad mayor que 0,90; dos, por encima de 0,85; y los restantes, una centralidad de 0,69 y 0,64. Entre los diez últimos que recibieron la noticia, tres tienen una centralidad inferior a 0,18; cuatro, entre 0,52 y 0,42; y los tres restantes tienen centralidades de 0,60, 0,73 y 0,85. Esto confirma la H2. Las acusaciones contra el TSE fueron una constante entre partidarios del PSL y son sugeridas por el propio candidato, Jair Bolsonaro, en su primera declaración a los periódicos tras el resultado de la primera vuelta. De los 438.400 mensajes textuales recogidos en esta red entre junio y octubre de 2018, 3.348 involucraban a las urnas y al TSE, la mayoría de ellas falsas.

Al identificar estándares estructurales de conexión y flujo, los modelos de red permiten utilizar una escala reducida de la red para diseños y comprender la dinámica de la red en su totalidad. Este enfoque es impositivo porque no puede ver la red completa de grupos de WhatsApp y establecer cualquier parámetro de representatividad o de confianza para métodos de cuantificación simple. Una vez que delimitamos nuestra red, el aumento en la cantidad de personas alcanzadas se desacelera en las etapas finales, a pesar de la continuidad de recursos compartidos. En el escenario real, en donde la difusión sigue a otros grupos periféricos fuera de la red analizada, esta desaceleración tardaría mucho más tiempo.



Por lo tanto, la noticia progresa preferentemente "desde" grupos con mayor centralidad "hacia" grupos periféricos. En esta lógica policéntrica, cuando otros grupos centrales son alcanzados se repite la dinámica, lo que contribuye a la viralización. En cada etapa, la multiplicación hace que la cantidad de información replicada para el siguiente conjunto de grupos sea exponencialmente mayor que la anterior, lo cual confirma la H3. Fuera de la red de grupos dedicados/especializados en política, grupos más difundidos socialmente, como familiares y otras afinidades, tienden a ser alcanzados. Esto hace que la simple cuantificación de tipos de grupos en que la noticia falsa puede ser encontrada, como los de familia, sin tener en cuenta su centralidad en la red que promovió la viralización, conduce a errores graves en la atribución de relevancia y se invierte completamente la lógica de la red.

Nuestro análisis del caso que involucra al TSE también indica que hubo sesgos de preferencia partidaria en la circulación de esta noticia. Se ha compartido 202 veces (al retornar la circulación después del último registro en los gráficos) pero, a pesar de las interconexiones entre los 90 grupos, solo 41 son alcanzados: 37 son grupos de apoyo a Bolsonaro

entre conservadores, “de derecha” o promilitares, y 4 son de discusión política sin candidato definido. La preferencia por un candidato puede indicar un filtro mayor o menor al compartir noticias específicas por los perfiles que conectan grupos favorables a este candidato a otros grupos de la red. Es importante señalar que, entre los grupos que recibieron la desinformación repetidamente, también pudimos identificar patrones que corroboran nuestra propuesta: entre los diez grupos con mayor número de repeticiones, esa información aparece en promedio 8,6 veces; la centralidad menor es de 0,84, mientras que la centralidad media es de 0,92 y seis de diez grupos poseen una centralidad superior a 0,90. El número medio de miembros del grupo es 234. Por otro lado, entre los diez grupos que recibieron solo una vez, la centralidad media es de 0,38 y la cantidad promedio de miembros por grupo es 97,25.

Al confirmar la H2, la utilización de algoritmos para la identificación de comunidades estructuradas que toman en cuenta únicamente la topología de la red (algoritmo de modularidad) fue exitosa en agrupar automáticamente partidarios de partidos diversos en categorías diferentes, identificar subgrupos entre partidarios de un mismo candidato e incluso categorizar redes de discusión política sin candidato definido –con raros errores de categorización que comprenden grupos de discusión sin candidato específico–. Esto muestra la heterogeneidad entre actores involucrados y que estas diferencias tienen desdoblamientos visibles en la estructura de la red, aun entre los que apoyan a un mismo candidato, y no solo en los filtros de noticias específicas entre los grupos.

Reconocidos los caminos preferidos, posiciones de centralidad y la dinámica favorecida por grupos segmentados más dispuestos a compartir, entendemos que la viralización en WhatsApp conlleva al menos tres etapas: primera, una etapa de producción y difusión inicial; segunda, su circulación en grupos dedicados a la política, interconectados por miembros más dispuestos a compartirla e insertarla en una dinámica de viralización; y, tercera, grupos periféricos no dedicados a la política, cuantitativamente más numerosos, aunque proporcionalmente irrelevantes en la etapa más intensa de la viralización. Al acercarse a grupos con mayor centralidad (*eigenvector*) tendemos a aproximarnos a la fuente primaria de la noticia falsa en nuestra red –no necesariamente el de toda la red de WhatsApp–. Esta centralidad puede cambiar con el tiempo, de acuerdo con las salidas de las nuevas conexiones en grupos.

Varios grupos recibieron la noticia falsa, incluso sin tener ninguna centralidad en la propagación viral. Podríamos asumir que los grupos familiares y sin afinidades políticas son más numerosos en WhatsApp que los grupos políticos. Teniendo en cuenta la relación entre la centralidad y los promotores en la dinámica viral, los grupos de familias y de afinidades reciben noticias que se volvieron virales como consecuencia y expansión, pero no causa la difusión sistemática de una noticia falsa específica. Algunos grupos familiares pueden ver la desinformación antes, que depende de su proximidad a los grupos centrales, pero tienden a ser considerablemente más pequeños que los grupos políticos en términos de miembros, y es menos probable que sus miembros estén interconectados con otros grupos políticos o con los que comparten información de manera sistemática.

Los grupos por afinidad y familiares son más numerosos y juntos pueden tener un número mayor de grupos/votantes, que hace que la desinformación llegue a las personas que no están dispuestas a buscar información política. También pueden tener efectos de “atajos cognitivos”[15] una vez que la fuente original está en general ausente, las personas dentro del grupo simplemente ven qué miembro lo publicó entre ellos. Entre las personas en grupos familiares/de afinidad que tienden a ser más familiares entre sí, el razonamiento común entre los conservadores fue: “Mi tío me lo envió sin que le pagasen por eso, entonces no hay una intervención profesional o intencional en la difusión de esta información”.

Otras contingencias ayudan a explicar las dificultades encontradas por los académicos y periodistas de investigación que intentaron dar sentido a WhatsApp en octubre, en respuesta a la participación electoral: i) no pudieron

[15] Popkin, Samuel L., *The Reasoning Voter: Communication and Persuation in Presidential Campaigns*, University of Chicago Press, 1994.

ingresar en grupos políticos que ya habían llegado al límite de participantes y ya estaban a pocas semanas de la votación; ii) cuando los grupos no estaban llenos, los periodistas dependían de los enlaces de invitación presentes fuera de WhatsApp, en un escenario de ataques mutuos que hizo que muchos grupos políticos restringieran sus invitaciones a las listas de reenvío internas; y iii) cuando finalmente ingresaron a un grupo, no tenían contenido previo a su entrada. Luego se enfrentaron a muchos filtros de cualquier contenido disponible: solamente el nuevo contenido de grupos que publican enlaces fuera de WhatsApp, por lo tanto, son grupos que no se preocupan por los "forasteros", y tampoco están llenos de miembros.

IV. H4: la arquitectura de seguridad de WhatsApp, rastreo y el caso brasileño en 2018

El potencial de las investigaciones en el auxilio de la comprensión de estos casos, junto a los desencuentros entre tecnología e iniciativas legales, motivaron en noviembre de 2018 la denuncia presentada a la Procuraduría General de la República por el investigador Miguel Freitas. La denuncia contiene indicios que podrían ser usados para identificar a los creadores de diversos contenidos falsos (*fake news*) que habían circulado en los grupos políticos de WhatsApp en el período electoral.[16] Para ello, se analizaron mensajes recogidos por otros investigadores que analizaron técnicas de viralización (cuyo primer resultado fue el análisis expuesto en los tópicos anteriores) ya divulgadas en diferentes congresos académicos,[17] seminarios y medios de comunicación, que identificaron las entradas que poseían las mejores características de trazabilidad.

En primer lugar, destacamos –entre los contenidos de los medios más populares– aquellos que fueran comprobadamente falsos y que preservaran la misma URL encriptada original por grandes períodos de tiempo, incluso en grupos diferentes. Hemos tenido éxito en el ámbito académico y, para avanzar en esta línea de investigación, los órganos competentes deberían solicitar a WhatsApp, mediante orden judicial, información sobre el usuario y/o la dirección IP responsable de la carga del archivo a los servidores de la compañía.

El material entregado consistió en dos documentos: 1) un documento con el asesoramiento técnico que demuestra el funcionamiento de la plataforma, los límites impuestos por el cifrado de extremo a extremo y el tipo de información técnicamente plausible que puede ser solicitada a la empresa; y 2) un listado de contenidos falsos/calumniosos con detalle de su propagación en los grupos monitoreados y sus identificadores únicos (*hashes* y URL) que deberían ser exigidos por la justicia. Para producir este listado de contenidos de medios rastreables, el primer paso consiste en agrupar las imágenes y los videos por similitud visual, de forma automatizada. Es común, principalmente en el caso de videos, que varias versiones del mismo contenido circulen en la red y que se diferencien solo en términos de calidad o resolución de imagen.

Entre los hallazgos, destacamos: cualquier modificación en el contenido del archivo, por menor que sea, que produce un identificador *hash* –tipo de identificador único que se genera al utilizar una función matemática, a

[16] Freitas, Miguel, "Sobre a rastreabilidade do envio de mídias na plataforma Whatsapp para o combate de crimes digitais", Río de Janeiro, 19 de noviembre de 2018.

[17] Santos, João Guilherme Bastos dos, Santos, Karina y Cardozo, Vanessa, "Cartografia do Whatsapp: a rede de apoio aos presidenciáveis nas eleições de 2018", 3º Congresso Nacional de Estudos Comunicacionais da PUC, Poços de Caldas, Minas Gerais, Conec, 2018; Santos, João Guilherme Bastos dos, Santos, Karina y Cardozo, Vanessa, "La red del 'mito' 2018: articulaciones políticas de grupos de extrema derecha en Whatsapp", Conferencia Latinoamericana de Ciencias Sociales, Buenos Aires, CLACSO, 2018.

partir del contenido del archivo– completamente diferente. En WhatsApp, los *hash* siempre se calculan a través del algoritmo SHA256 y se codifican con BASE64. Dos *hash* se encuentran dentro de la plataforma WhatsApp: el *hash* del archivo de medios original y el *hash* del archivo cifrado. Por motivos técnicos y de seguridad, cada vez que el archivo se cifra se obtiene un resultado diferente y, por lo tanto, un *hash* diferente. Esto conduce a la segunda variabilidad frecuentemente encontrada en los mensajes de WhatsApp: un mismo archivo original enviado a la red por diferentes usuarios (*upload*) produce versiones encriptadas diferentes, con diferentes *hash* correspondientes. La observación de este patrón de propagación permite discriminar casos en que el contenido original fue inicialmente distribuido por otra plataforma, por ejemplo, vía Facebook. Cuando diferentes usuarios descargan el archivo de Facebook para retransmitirlo de forma independiente dentro de WhatsApp se produce un patrón diferente.

Es posible percibir que en algunos casos, aunque el *hash* del medio (archivo jpg) sea igual en todos los mensajes mostrados, el *hash* del archivo encriptado (archivo enc) presenta versiones diferentes y, para cada una de ellas, WhatsApp produjo una URL diferente. Esta URL se puede escribir en un navegador, lo que permite que cualquier persona pueda descargar el archivo cifrado si todavía está en la red. Este archivo, sin embargo, solo se descodificará en el contenido original de la propiedad de las claves de cifrado que se reenvían en el propio mensaje y se restringen a los emisores y los destinatarios del archivo.

En las ocasiones de elevada posibilidad de que este contenido haya sido distribuido originalmente en otra plataforma, los casos marcados por varias URL no se consideran de buena trazabilidad. No sería productivo encontrar a los diferentes usuarios que hicieron la copia de una plataforma a la otra, pues –aunque la práctica es igualmente problemática– no hay ningún indicativo de que estos tendrían alguna relación con el creador original del contenido.

En este punto, la identificación de casos de viralización interna en WhatsApp es relevante. Se indican situaciones en que prácticamente todos los reenvíos se refieren al mismo *hash* de medios originales –debido a la lógica viral– y también la misma URL (y consecuentemente el mismo *hash* cifrado). Un caso de buena trazabilidad en este sentido es el supuesto mensaje entre José Sérgio Gabrielli y Fernando Haddad que combina la publicación de una "noticia-bomba" en la *Folha de São Paulo*. Se encontraron cientos de mensajes reenviados con la misma URL, lo que nos lleva a concluir que el contenido es originario de la propia plataforma WhatsApp y compartido casi exclusivamente a través de la opción de reenvío ofrecida, que activa lógicas virales de difusión. Si WhatsApp proporcionara el IP responsable del *upload* de esta URL, podríamos llegar al usuario creador de este contenido.

En respuesta oficial, sin embargo, WhatsApp se negó a almacenar los registros de carga de archivos que permitiera identificar al creador de los contenidos multimedia. La respuesta de WhatsApp es problemática desde el punto de vista legal, pues la empresa estaría posiblemente infringiendo el artículo 15 del Marco Civil que exige que los proveedores de aplicación guarden dichos registros por un plazo de seis meses, y serían obligados a suministrarlos solo por determinación judicial.

El mecanismo de rastreo aquí sugerido ofrece una ventana segura para investigaciones moderadas, es decir, sin permitir abusos ni el acceso masivo por el Estado. Esto se aplicaría también a investigaciones de diferentes naturalezas no electorales como, por ejemplo, crímenes de pedofilia. Es necesario avanzar con este debate en la sociedad, para restablecer límites y deberes de las empresas de tecnología y así estas puedan colaborar efectivamente con investigaciones judiciales legítimas sin violar los derechos individuales.

V. Consideraciones finales

La investigación logró confirmar empíricamente cuatro hipótesis: H1) WhatsApp está sujeto a lógicas de redes bipartitas gracias a su estructura de grupos segmentados interconectados; H2) la implementación de métricas pueden identificar grupos centrales en este proceso variante en el tiempo a través de diferentes etapas; H3) la (des)información va de nodos centrales hacia nodos periféricos que amplían su alcance exponencialmente y la vuelve viral; y H4) los reenvíos guardan información de remitentes iniciales que permite el rastreo sin violar la privacidad de los usuarios. El cruzamiento entre la centralidad (*eigenvector*) y la identificación de comunidad estructurada (modularidad) ofrece un mecanismo automatizado de detección de rutas preferentes para la difusión de noticias virales en cada comunidad o coalición de comunidades encontrada en la red. Replicar el uso de este algoritmo en un número mayor de redes puede avanzar y establecer modelos de red.

Creemos así contribuir para al avance de los estudios sobre herramientas como WhatsApp y repertorios de acción política en red. Si logramos identificar las rutas iniciales de las noticias y las características específicas agregadas a los reenvíos en la aplicación, podremos explorar las posibilidades que esta tecnología ofrece para que instituciones democráticas puedan ser efectivas en el cumplimiento de demandas sociales.