

Enfoques big data para la comunicación en salud: análisis de redes y análisis de sentimientos a gran escala

Big data approaches in health communication: Network analysis and large-scale supervised sentiment analysis

Carlos Arcila Calderón¹, Luz Marina Alonso², Antonio García Jiménez³

Resumen

Se exponen los dos principales enfoques metodológicos para la investigación con big data en comunicación en salud: el análisis de redes y el análisis de sentimientos. Primeramente, se explica el cambio de paradigma que está sufriendo el campo de la comunicación en salud gracias a los métodos computacionales para el análisis de datos masivos y se dan ejemplos de su uso en estudios y experiencias previas. Seguidamente, se exponen los conceptos (nodo/arista) y las principales variables de centralidad que se estudian en el análisis de red en procesos de difusión de innovaciones en salud; y, finalmente, se explica cómo ejecutar el procedimiento de análisis de sentimientos supervisado para estudiar contenidos de salud a gran escala.

Palabras clave: Comunicación en salud, big data, análisis de redes, análisis de sentimientos

Abstract

The two main methodological approaches for research with big data in health communication are presented: network analysis and the analysis of feelings. Firstly, the paradigm change that the communication field in health is experiencing, thanks to the computational methods for the analysis of massive data, and examples of its use in previous studies and experiences are explained. Next, the concepts (node / edge) and the main centrality variables that are studied in the network analysis in health innovation dissemination processes are exposed; and, finally, it is explained how to execute the supervised feelings analysis procedure to study large-scale health content

Keywords: Health communication, Big Data, Network analysis, Sentiment analysis.

Fecha de recepción: 16 de septiembre de 2017
Fecha de aceptación: 3 de diciembre de 2017

¹ Profesor del Departamento de Sociología y Comunicación. Universidad de Salamanca, España

² Profesora del Departamento de Salud Pública. Universidad del Norte. Barranquilla (Colombia).

³ Profesor de la Facultad de Ciencias de la Comunicación. Universidad Rey Juan Carlos, España.

Correspondencia: Carlos Arcila Calderón. Campus Miguel de Unamuno. Edificio FES Avda. Francisco Tomás y Valiente, s/n. 37007 Salamanca, España- carcila@usal.es

INTRODUCCIÓN

Se parte de la idea del cambio paradigmático que supone la aplicación de las diferentes técnicas asociadas al *big data* (1), teniendo también en cuenta la creciente relevancia de las diferentes interacciones entre salud, comunicación y tecnología, más allá de la simple idea de mHealth (2). Tal y como se refleja en el informe “The big-data revolution in US health care: Accelerating value and innovation” (3) son diversas las áreas de impacto potencial de este grupo de técnicas: a) la elección de los estilos de vida con vistas a promover el bienestar de los ciudadanos y una implicación activa en su propio cuidado; b) la atención basada en pruebas, lo que se ha demostrado útil para conseguir unos buenos resultados en cada paciente; c) la generación, para los trabajadores del sector sanitario, de un marco más apropiado de trabajo; d) la propuesta de enfoques sostenibles que mejoren la calidad de la atención sanitaria y reduzcan su coste; y finalmente, e) la innovación y el aumento de la productividad de la I+D en términos de descubrimiento, desarrollo y seguridad.

Una de sus expresiones más interesantes es la que se propone bajo la denominación *Social Big Data*. Bello-Orgaz, Jung y Camacho (4) vinculan el *big data* social con el análisis de grandes cantidades de información a partir de múltiples fuentes de información, con especial énfasis en los medios sociales. Este método con un enfoque interdisciplinar engloba técnicas que van desde la minería de datos y textos, la extracción terminológica, la clasificación de textos, el *machine learning*, la estadística, el *graph mining*, la lingüística y el procesamiento del lenguaje natural, la web semántica, las ontologías, etc. De hecho, una de sus aplicaciones más relevantes es la denominada *epidemic intelligence*, que puede ser definida como la identificación temprana, la valoración y la verificación de

riesgos potenciales para la salud pública, así como la difusión de las correspondientes alertas. En este sentido, se debe tener en cuenta la creciente relevancia de los medios sociales dada la cantidad y frecuencia de los mensajes que difunden. En el caso del *microblogging* Twitter, la información recogida puede derivarse de los tuits de los usuarios acerca de enfermedades y de las reacciones de la red de personas con las que están o podrían estar en contacto.

En general, esta disciplina hace acopio de procedimientos de vigilancia tecnológica tanto para el análisis reticular (nodos y aristas) de la difusión de innovaciones en salud como para el análisis automatizado y continuo de textos libres no estructurados, de la información disponible en los medios sociales, en los *blogs*, en los medios de comunicación digitales, etc. También emplea formas de identificación de mensajes relevantes, o sistemas de aprendizaje automático, por ejemplo: para clasificar tuits relacionados con la gripe; o bien modelos de regresión que permiten evaluar brotes epidémicos a partir de los mensajes existentes, incluso cuando no se pueden vincular a palabras clave o a patrones lingüísticos potencialmente relacionados.

Son destacables los sistemas de vigilancia, de rastreo y de monitorización para la detección de epidemias, como los que están en conexión con la Organización Mundial de la Salud o con el Centro Europeo para la Prevención y Control de Enfermedades. Otro ejemplo es el sistema de vigilancia de la Global Public Health Intelligence Network (GPHIN), desarrollado por la Public Health Agency de Canadá, basado en la información en la web y con una clara orientación multilingüe. Otro sistema reseñable es el

BioCaster, fundamentado en una ontología que monitoriza y clasifica los datos detectados en función de la importancia del tema, tras lo cual vincula a Google Maps mediante geo-información. Por su parte, el proyecto HealthMap se presenta como un mapa global de alerta de enfermedades que usa fuentes de diferente naturaleza, como las que proceden del Google News o las conversaciones de expertos, es un sistema automatizado que abarca diversos procedimientos de monitorización, organización, integración, visualización y diseminación de información médica. Finalmente, EpiSpider es una herramienta que extrae automáticamente información sobre brotes de enfermedades infecciosas de múltiples fuentes, incluyendo ProMedmail y sitios web médicos, y es utilizado como un sistema de vigilancia por las organizaciones de salud pública, varias universidades y organizaciones de investigación de salud.

En este artículo se explican dos de las técnicas de *big data* más empleadas en comunicación en salud, el análisis de redes y el análisis supervisado de sentimientos. Además se explica cómo puede ser utilizado el análisis de redes en los procesos de comunicación interpersonal, que son el vehículo para la difusión y adopción de innovaciones en salud. El estudio de redes en comunicación interpersonal en salud implica complejidad en la comprensión de la difusión de innovaciones y es necesario ir más allá de simples consideraciones funcionales o instrumentales, pues de esta manera se puede responder a la pregunta sobre por qué se adopta o no una innovación en el ámbito de la salud. Es así como se encuentran investigaciones que relacionan las redes de comunicación interpersonal y la influencia sobre procesos de difusión de información y persuasión, este último constructo relacionado también con el proceso de innovación (5).

En este sentido, diferentes investigadores han estudiado cómo el proceso de difusión podría estar relacionado con la fuerza de los lazos de la red. Así, en el estudio ya tradicional de Mienke y Midden (6) se tomaron dichos atributos de la red como predictores y decisión de adopción de innovación como variable a explicar. Se encontró que la información en el proceso de difusión se relacionaba con el número de lazos y no con la calidad, mientras que la decisión de adoptar estaba relacionada con la calidad de los lazos. Todo lo mencionado anteriormente muestra que distintos tipos de análisis de redes interpersonales redundan en un estudio más efectivo de su dinámica en distintos entornos (7).

Por otra parte, se explica el análisis de sentimientos, ya que permite aplicaciones de valor en el mundo de la salud tanto para los ciudadanos como para los servicios especializados y las agencias públicas (8). Su aprovechamiento está dirigido a la detección de opiniones en el seno de los diversos espacios de diálogo en el ciberespacio, especialmente en lo que se refiere al mapeamiento de diferentes dolencias. Además, pueden ser de utilidad para la industria farmacéutica, en términos de prospección y de *marketing* centrados en la salud colectiva, habida cuenta de la relevancia de las opiniones y respuestas emocionales de las personas ante determinadas situaciones o informaciones (9,10).

ANÁLISIS DE REDES DE SALUD

La tendencia a analizar las redes en relación a los nuevos cambios en salud (o innovaciones), ha sido desarrollada a lo largo del tiempo. Dentro de los modelos de comunicación que han tenido en cuenta el análisis

de redes se encuentra el de Evert Rogers (11), que hace mención a esta dinámica, considerando que las personas que interactúan en la red con un propósito lo realizan teniendo en cuenta la semejanza o la homofilia.

Con el transcurrir del tiempo y con base en modelos matemáticos se han desarrollado otros tipos de análisis de redes. También se han analizado sus atributos, así como las características intrínsecas de la red: grado nodal, centralidad, intermediación, cercanía y otras variables como densidad y tamaño. De igual manera se evalúan otras características como susceptibilidad, infecciosidad o exposición. Si se mira esta dinámica de red con un enfoque de grandes cantidades de datos, se estará planteando una potencialidad y mayor diversidad de análisis de redes que operan en los procesos de salud y enfermedad. En esa dinámica de plataformas e interacciones que existen tales como el Facebook, las redes sociales son relevantes para la planificación gubernamental y la implementación de estrategias en el área clínica y de salud pública, por lo que el modelado de los mensajes de comunicación en los medios sociales como Twitter proporciona una oportunidad para probar predicciones de comportamiento individual futuro. Analizar las redes mediante las técnicas que provee el *big data* permite potenciar, en todos los aspectos, la velocidad, la seguridad y la flexibilidad de los datos, así como la toma de decisiones.

La principal característica intrínseca en el análisis de una red de salud son los atributos de centralidad, en los cuales se reconoce que la gente en esta condición de núcleo generalmente ocupa posiciones de prestigio que influyen en el comportamiento de la red. No obstante, también se ha documentado que las personas pueden ser influidas por contactos directos con los que cuentan los nodos (12).

El concepto de centralidad realmente surge en el año 1950, trabajado por Bavelas (13) y otros académicos como Borgatti y Everett (14) que hicieron aportes importantes al análisis de redes. El término que define la posición en la red reconocida como centralidad es uno de los procedimientos más utilizados en el análisis de las redes sociales. A nivel organizacional, la pregunta que se ha reportado en estudios es: ¿quién es más central? ¿Es más central el que tiene una posición con el mayor número de conexiones? ¿Es más central el que tiene una posición más cercana a todos los demás? ¿Es más central el que tiene una posición de puente entre unos y otros? Freeman (15) se ocupó de los conceptos de grado, cercanía e intermediación nodal de las redes (16), que han sido aplicados a distintos análisis: desde el comportamiento y efecto de políticas en programas empresariales hasta su utilidad en programas innovadores en salud (17, 18).

A continuación se describen las tres principales medidas de centralidad de una red interpersonal que pueden afectar el proceso difusión y adopción de una innovación en salud.

En primer lugar, se encuentra el *grado nodal*. En este caso, los nodos de una red pueden representar eventos, ideas, objetos y personas; y los vínculos son las relaciones entre los nodos. La vinculación indica un sentido de comunidad de confianza. El grado del nodo se refiere a cada uno en particular y se define como el número de sus nodos adyacentes, o el número de relaciones que un nodo dado mantiene o al que está directamente conectado. El grado nodal es un indicativo de la actividad relacional de un nodo con sus nodos contiguos, así se podrán clasificar los nodos según su actividad y según el número de vínculos con los otros actores de la red. El mismo puede diferenciarse en térmi-

nos de *indegree* y *outdegree*, es decir, relaciones entrantes y salientes. El *indegree* es el número de lazos que se reciben y el *outdegree*, los que se envían. El grado es una característica de la medida de centralidad local y puede ser calculada en referencia a la estructura de la red. Por ejemplo, en un estudio se evalúa si una escuela genera programas para prevención de obesidad, el *oudegree* es el número de contactos de otras escuelas con los que consulta información para realizar o mejorar este tipo de programas. Igual en un estudio que evalúa compañeros sexuales, el *outdegree* es el número de compañeros sexuales los cuales son una variable importante para identificar algún riesgo (19). El *outdegree* se utiliza a veces como denominador al calcular la exposición a la red. Según Valente (20), es un indicador útil para atributos personales. Por ejemplo, en un estudio de red se puede pedir los nombres de otras personas a quienes éstas hayan consultado por algún motivo, verbigracia, por apoyo emocional; contar con ese número proporciona una medida de la red de apoyo.

En segundo lugar, se encuentra el *grado de cercanía nodal*, el cual hace referencia a lo cerca que está un actor del resto de los otros que forman el conjunto de individuos (21). Bavelas (13) desarrolló esa medida y la amplió posteriormente Sabidussi en 1966 (21), dieciséis años después y fue propuesta en función de sus distancias geodésicas (camino más cortos). Según Valente (20), se define para cada nodo como la distancia geodésica del nodo en cuestión con respecto al resto de los nodos, sea de relación directa o mediada por otros nodos; su valor es el marginal en la matriz geodésica. La centralidad de cercanía se define como una proporción donde primero se cuenta la distancia de un vértice a todos los vértices de la red. La normalización permite que la centralidad de cercanía proporcione una medida global sobre la posición del vértice de la red (19). Desde una perspectiva relacional, o

de la teoría de capital social, este recurso tiene que ver con la capacidad, primero, de alcanzar todos los nodos a partir de las relaciones, o en todo caso al máximo posible con el mínimo número de pasos (geodésica) sea directamente o por intermediarios. Se refiere a lo estrechamente vinculado que está un nodo con otro.

Finalmente, se encuentra la centralidad por intermediación nodal, que identifica la posición de un nodo en la red en términos de capacidad de conexión entre los pares de nodos de la red. También indica el potencial de un nodo para ser un intermediario de información; en este sentido, la red se interpreta como un tejido de conexiones y de interdependencias mutuas, en la que los nodos actúan como corredores. El grado de intermediación cuenta las veces que un nodo aparece en los geodésicos de todos los nodos de la red. La centralidad de intermediación de cada nodo se define como la suma de la combinación de todos los pares de nodos de la red que para comunicarse entre sí por el camino más corto (el de menos pasos) han de pasar necesariamente por el nodo en cuestión; para su reducción, el denominador es el valor de la combinación de todos los pares de nodos de la red que para comunicarse, pasen o no pasen necesariamente por dicho nodo.

ANÁLISIS DE SENTIMIENTO EN COMUNICACIÓN EN SALUD

Existe un interés creciente en el estudio de la comunicación en salud utilizando datos a gran escala derivados de distinto tipo de datos: estructurados, no estructurados y semiestructurados producidos en distintos espacios, como las clínicas, la propia comunidad o los espacios virtuales de interacción, pero todavía se hace poco con las grandes técnicas distribuidas de datos. Algunos enfoques se basan únicamente en el análisis automatizado del sentimientos

utilizando diccionarios que marcan palabras generando un valor positivo o negativo, mientras que otros métodos, como el análisis del sentimiento supervisado basado en el aprendizaje automático (22), son todavía escasos en la investigación de comunicación (23).

El análisis de sentimientos en comunicación en salud puede tener una dimensión social o de prevención, lo que es objeto de estudio en investigaciones como la de Oksanen et al. (24). Este estudio analiza las reacciones emocionales tanto pro-anorexia como anti-pro-anorexia. Recordemos que hay toda una corriente de comunidades (pro-anorexia) que fomentan la pérdida de peso y que se apoyan fundamentalmente en los lazos sociales establecidos en su interior. Y se observa que, pese a que el contenido pro-anorexia está generalizado en YouTube, los vídeos que promueven la ayuda hacia las personas que sufren este trastorno y que se oponen a este tipo de comunidades son más populares en número y consideraciones positivas. De hecho, conocer esta dinámica puede ser eficaz para las personas y profesionales que trabajan con los jóvenes.

En esta línea de trabajo también se encuentra el estudio de Mazzocut et al. (25) respecto de las conversaciones digitales sobre la relación entre la medicina complementaria y alternativa y el cáncer. Gracias al *sentiment analysis* se pudo identificar las fuentes web más empleadas, las terapias más recurrentes y las tendencias de opinión de mayor impacto. El uso en pacientes de cáncer de este tipo de medicina en Italia (país donde se centró el estudio), es generalizado al mismo tiempo que existe un alto nivel de automedicación. Este hecho puede obedecer a la falta de comunicación entre pacientes y médicos, lo que expondría a los pacientes a comportamientos de riesgos y al uso de canales informativos poco fiables.

Este enfoque también es el practicado en el artículo de Palomino et al. (26), que analiza la diseminación y las diferentes opiniones sobre un término acuñado solo a nivel social, el nature-deficit disorder, o la investigación centrada, en Twitter, sobre el sentimiento al respecto de los cigarrillos mentolados (27).

A continuación describimos el método de análisis supervisado de sentimientos, que puede ser implementado en estudios de comunicación en salud.

A diferencia del análisis de sentimiento automático, que usa diccionarios con valores preestablecidos, el análisis supervisado de sentimientos utiliza modelos generados a partir de ejemplos para predecir el sentimiento en un mensaje (28). Específicamente, esta técnica hace uso de algoritmos de *machine learning* o aprendizaje automático para generar modelos predictivos para cada sentimiento especificado (positivo, negativo, neutro, etc.). Asimismo, el análisis supervisado de sentimientos utiliza técnicas de procesamiento del lenguaje natural para poder generar los modelos solo con determinadas palabras de los contenidos de ejemplo (adjetivos o verbos, entre otros). Ambas dimensiones permiten que el análisis pueda ser implementado a gran escala para problemas de dimensión *big data* en contenidos con impacto en salud, como es el caso de la monitorización de tuits con palabras relacionadas con determinadas enfermedades o nuevos tratamientos médicos.

En todos los casos, el análisis supervisado de sentimientos comienza con la recolección y clasificación manual de contenidos de ejemplo que conformarán el corpus de entrenamiento (por ejemplo, un set de mensajes positivos y otro de negativos) del modelo y que deben ser específicos de la situación y contexto que

se quiera analizar. Aunque represente un trabajo adicional para los investigadores, esta propiedad es la que permite que este tipo de análisis pueda adaptarse con mayor calidad en un contexto de mensajes de salud; es decir, logra aumentar la precisión en la predicción de los sentimientos ya que aborda el lenguaje en su contexto lingüístico específico. En la práctica, este corpus de ejemplos se subdivide en un sub-corpus de entrenamiento propiamente, y otro sub-corpus de *test* o prueba. Con esto, se generan los modelos con los datos de entrenamiento (p. e. un 70 % del total) y se contrasta la capacidad predictiva de estos modelos con los datos de prueba (el resto), ya que estos contenidos están también etiquetados. La creación de los modelos se realiza con palabras (sus raíces o lemas), que son convertidas en vectores para que los algoritmos puedan trabajar con ellas de forma cuantitativa. Las palabras son separadas a través de la *tokenización*, con la que luego es posible escoger el tipo de palabra con el que se quiere trabajar (p. e. adjetivos) y la cantidad total de palabras/vectores que conformarán los modelos (p. e. las 5000 palabras más frecuentes en los textos de ejemplo).

CONSIDERACIONES FINALES

Todas las técnicas que engloban lo que se denomina *big data* se constituyen en pieza fundamental en la transformación de múltiples facetas comunicativas, lo que se puede observar en los crecientes vínculos entre la tecnología, la comunicación y la salud. Este artículo profundiza en dos de los encuadres metodológicos más relevantes para entender la investigación y aplicación del *big data* en comunicación en salud. Por un lado, el análisis de redes y, por otro, el análisis supervisado de sentimientos.

De igual modo se han revisado las técnicas más destacables en este cruce de dominios: por ejemplo, la vigilancia tecnológica aplicada al análisis reticular (nodos y aristas) para la innovación y las actitudes en comunicación de cuestiones de salud (como pueden ser las redes de apoyo de una persona), o el análisis automatizado de textos no estructurados, que son los habituales en los medios sociales y en la mayor parte de plataforma existentes en el ciberespacio. A su vez, se han descrito los ámbitos operativos con mayor proyección tales como la detección de brotes epidémicos, por diferentes medios, así como su valoración y verificación, junto a la transmisión de alertas e información relacionadas.

Este estudio supone un avance porque reúne y describe los desarrollos que se están desplegando en este ámbito, en un campo que es doblemente complejo. Por un lado, por el gran número y variedad de variables y técnicas que se deben tomar en su consideración. Por otro, por las dificultades inherentes a todo dominio transdisciplinar, con base tecnológica. Aunque se ha partido de planteamientos genéricos sobre salud, la clave se ha situado en la traslación de las diferentes técnicas del *big data* a la comunicación en salud. Finalmente, son muchos los aspectos que se deben seguir estudiando: por ejemplo, las limitaciones del propio *big data* en algunas de las esferas de actuación, su influencia y adaptación a los siempre cambiantes procesos comunicativos, o su capacidad para generar patrones de comportamiento comunicativo-digital de los ciudadanos en su acceso y comunicación a información sobre salud.

Conflicto de interés: ninguno.

Financiación: recursos propios.

REFERENCIAS

1. Abbasi A, Sarker S, Chiang RH. Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems* 2016;17(2):I-XXXII. Disponible en: <http://aisel.aisnet.org/jais/vol17/iss2/3/>
2. Niños-González JI, Fernández-Morales B. Comunicación, Salud y tecnología: mHealth. *Revista de Comunicación y Salud*. 2015;5:144-153. Disponible en: <http://revistadecomunicacionsalud.org/index.php/rcys/article/view/86/90>
3. Kayyali B, Knott D, Van Kuiken S. The big-data revolution in US health care: Accelerating value and innovation. *Mc Kinsey & Company*. 2013; 2(8): 1-13
4. Bello-Orgaz G, Jung JJ, Camacho D. Social big data: Recent achievements and new challenges. *Information Fusion*. 2016;28:45-59. Disponible en: <http://dx.doi.org/10.1016/j.inffus.2015.08.005>
5. Valente T, Saba W. Campaign exposure and interpersonal communication as factors in contraceptive use in Bolivia. *Journal of health communication*;2001, vol 6: 303-322
6. Mienke W, Weenig H. Communication Network influences on information diffusion and persuasion. *Journal of personality and Social Psychology*. 1991;61(5):734-742. Disponible en: <http://psycnet.apa.org/record/1992-09003-001>
7. Korteland E, Bekkers V. The diffusion of electronic service delivery innovations in dutch E-policing: The case of digital warning systems. *Public Management Review*. 2008;10(1):71-88.
8. Araujo GD, Sousa FS, Teixeira F, Mancini F, De Domenico EBL, de Paiva Guimarães M, Pisa IT. Análise de sentimentos sobre temas de saúde em mídia social. *Journal of Health Informatics*. 2012;4(3):95-99. Disponible en: <http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/195/164>
9. Plaza del Arco FM, Martín-Valdivia MT, Jiménez-Zafra SM, Molina-González MD, Martínez-Cámara E. COPOS: Corpus Of Patient Opinions in Spanish. Application of Sentiment Analysis Techniques. *Procesamiento del Lenguaje Natural*. 2016;57:83-90. Disponible en: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5340/3124>
10. Jiménez-Zafra SM, Martín-Valdivia MT, Maks I, Izquierdo R. Analysis of patient satisfaction in Dutch and Spanish online reviews. *Procesamiento del Lenguaje Natural*. 2017;58:101-108. Disponible en: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5418/3182>
11. Rogers EM. Diffusion of innovations. Fourth ed. New York: The free press; 1995.
12. Rana S, Allen N. Centrality measures to identify key stakeholders in family violence. *Psychosocial Intervention*. 2015; 24:167-176
13. Bavelas S. Communication patterns in task oriented groups. *Journal of the Acoustical Society of America*. 1950; 22:271-282.
14. Holgado D. Analyzing social networks. *Revista Hispana para el Análisis de Redes Sociales*. 2016;27(2):141-5. Disponible en: <http://dx.doi.org/10.5565/rev/redes.637>
15. Freeman LC. Centrality in social networks conceptual clarification. *Social Networks*. 1979;1(3):215-39. Disponible en: [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
16. Kessler S. Article Navigation Social Networks and Health: Models, Methods, and Applications. Thomas W. Valente. *International Journal of Epidemiology* 2011;40(6):1742-1743. Disponible en: <https://doi.org/10.1093/ije/dyr160>
17. Russo M. Cooperation Networks and Innovation: A Complex Systems Perspective to the Analysis and Evaluation of a Regional Innovation Policy Programme. *Sage Publications*. 2009;15(1):75-100
18. Kunst L, Kratzer J. Diffusion of innovation through social network of children. *Young Consumers*. 2007;8(1):36-51

19. Wasserman S, Faust K. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press; 2009.
20. Valente, T. W. *Social networks and health: Models, methods, and applications* (Vol. 1). New York: Oxford University Press: 2010.
21. Sabidussi G. The centrality index of a graph. *Psychometrika*. 1996; 31:581-603.
22. Vinodhini G, Chandrasekaran RM. Sentiment analysis and opinion mining: a survey. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2012; 2(6):282-92.
23. Van Zoonen W, Toni GLA. Social media research: The application of supervised machine learning in organizational communication research. *Computers in Human Behavior*. 2016;63:132-41.
24. Oksanen A, García D, Sirola A, Näsi M, Kakininen M, Keipi T, Räsänen P. Pro-anorexia and anti-pro-anorexia videos on YouTube: Sentiment analysis of user responses. *Journal of medical Internet research*. 2015; 17(11): e256. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4704949/>
25. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, De Paoli P, Tasso C. Web conversations about complementary and alternative medicines and cancer: content and sentiment analysis. *Journal of Medical Internet Research*. 2016;18(6):e120. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4929351/>
26. Palomino M, Taylor T, Göker A, Isaacs J, Warber S. The Online Dissemination of Nature–Health Concepts: Lessons from Sentiment Analysis of Social Media Relating to “Nature-Deficit Disorder”. *International journal of environmental research and public health*. 2016;13(1):142. Disponible en: <http://www.mdpi.com/1660-4601/13/1/142/htm>
27. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of Menthol Cigarettes Among Twitter Users: Content and Sentiment Analysis. *Journal of Medical Internet Research* 2017; 19(2):e56. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5348619/>
28. Arcila-Calderón C, Ortega-Mohedano F, Jiménez-Amores J, Trullenque, S. Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tuits basada en aprendizaje automático. *El profesional de la información (EPI)*. 2017; 26(5), 973-982.