
Capítulo 4. Modelado de temas y análisis de redes sociales en Twitter durante el Día Mundial de las Enfermedades Raras

Topic modelling and social network analysis on Twitter during World Rare Disease Day

MARTÍNEZ-MARTÍNEZ, Marta (1), MARQUÉS-SÁNCHEZ, Pilar (2), BASURTO-SAIZ, Lourdes (3), TREVISSÓN-REDONDO, Bibiana (2)

(1) Centro de Salud de Benavides de Órbigo, Gerencia Regional de Salud de Castilla y León (SACYL), Benavides de Órbigo (León), España, mmartinezm@saludcastillayleon.es; (2) SALBIS Reserch Group, Faculty of Health Sciences, Universidad de León, León, España, pilar.marques@unileon.es y btrrer@unileon.es; (3) Hospital de Leza, La Guardia, Álava, España, mlbasurto11@gmail.com

Resumen

Durante el Día Mundial de las Enfermedades Raras se recogieron las interacciones realizadas en la red social Twitter relacionadas con este acontecimiento. Se estudiaron los datos obtenidos para detectar los temas de conversación y se realizó un análisis de redes sociales para conocer los principales actores participantes y sus relaciones.

Palabras clave: Enfermedades raras, asociaciones de pacientes, Twitter, modelado de temas, análisis de redes sociales.

Abstract

During World Rare Diseases Day, the interactions carried out on the social network Twitter related to this event were collected. The data obtained was studied to detect the topics of conversation and an analysis of social networks was carried out to know the main participating actors and their relationships.

Keywords: Rare diseases, patient organizations, Twitter, topic modeling, social network analysis.

4.1. Introducción

Las enfermedades raras son aquellas que afectan a una pequeña parte de la población (menos de 1 de cada 2000 personas en el caso de Europa). A pesar de ser poco frecuentes, se estima que, a nivel europeo, más de 30 millones de personas pueden estar afectadas (EURORDIS, 2021). Las enfermedades raras constituyen un importante problema de salud pública; con frecuencia son crónicamente debilitantes, lo que obliga a los pacientes a una vida dependiente y muy limitada. Esto hace que los cuidadores principales se vean obligados a adaptarse a su nueva situación mediante cambios drásticos en su situación socioeconómica y personal (Silibello et al., 2016).

La situación de indefensión y soledad experimentada por los enfermos o familiares más cercanos, la falta de apoyo institucional y la poca información disponible (Silibello et al., 2016) constituyen un contexto propicio para la formación de las asociaciones de pacientes (o de familiares, si la enfermedad es incapacitante). Estas asociaciones cobran especial relevancia en el ámbito de enfermedades poco comunes porque, en ellas, los afectados encuentran respuestas, recursos y apoyo en personas que se encuentran en situaciones similares, además de sensibilizar al resto de la población y recaudar fondos para la investigación (Vicari & Cappai, 2016).

Las organizaciones de pacientes con enfermedades raras han ido evolucionando a lo largo del tiempo asumiendo roles no habituales en otro tipo de asociaciones de patologías más comunes. Cada vez en mayor medida, estos grupos no se limitan a ser meros espectadores y, empujados por la percepción de falta de interés por parte de los expertos, adoptan un papel protagonista intentando suplir la carencia de conocimientos acerca del tema con actividades que faciliten la participación e interacción entre familiares e investigadores (von Gizycki, 2010) y fomentando el desarrollo de innovaciones en este ámbito (Merkel et al., 2016).

Las redes sociales online (Facebook, Twitter, Instagram, etc) suponen hoy en día una herramienta muy importante para estas asociaciones, ya que les permiten obtener una visibilidad y alcance que de otra forma sería muy difícil obtener (Castillo-Esparcia & López-Villafranca, 2016). Sin embargo, el bajo número de individuos que forman este tipo de asociaciones hace que el esfuerzo que necesitan para conseguir y mantener un nivel de visibilidad adecuado sea alto.

Dentro de las actuales redes sociales en Internet, Twitter (Pershad, Hangge, Albadawi, & Oklu, 2018) es la que mayor cantidad de información puede ofrecer para el estudio de las interacciones de estas asociaciones en Internet debido al carácter abierto de sus publicaciones y a su uso cada vez más extendido en el ámbito de la salud (Sinnenberg et al., 2017). El estudio de las interacciones y dinámicas que se establecen en esta red social puede ayudar a las asociaciones de enfermedades raras a mejorar la calidad de su presencia online, obteniendo una mejor visibilidad y un mayor alcance de sus actividades en relación a todos los actores involucrados (Kloth, Deutsch, Danielson, Strack, & Law, 2019).

Para el sistema sanitario, conocer cómo relacionarse con estos enfermos y familiares puede suponer una mejora en la prestación de servicios y en el apoyo percibido. El estudio del contenido de los mensajes intercambiados en la red social Twitter encuentra cada día más aplicaciones al ámbito de la salud, suponiendo una importante fuente de información, utilizando algoritmos cada vez más precisos (Benitez-Andrades et al., 2021). Los objetivos del presente trabajo son (1) Capturar los contenidos que se generan en la red social Twitter en torno al tema del Día Mundial de las Enfermedades Raras celebrado el 28 de febrero de 2021, (2) realizar un estudio descriptivo para cuantificar el volumen de mensajes generados y los principales actores involucrados, (3) analizar las temáticas que aparecen en esos contenidos intercambiados, (4) realizar un análisis de redes sociales para estudiar las relaciones

establecidas entre los actores en la red de retuits y localizar los agentes más relevantes que participaron y (5) obtener información útil para próximos estudios.

4.2. Metodología

De las 385 asociaciones de enfermedades raras censadas en la Federación española de enfermedades raras FEDER, 258 tienen una cuenta de Twitter activa. Para realizar el presente estudio, se capturaron todos los tuits intercambiados durante el día mundial de las enfermedades raras, celebrado el 28 de febrero de 2021, que contenían uno, o más, de los siguientes términos:

DiaMundialdelasEnfermedadesRaras2021, DiaMundialdelasEnfermedadesRaras, FebreroRaro, DíaDeLasEnfermedadesRaras, DíaDeLasEnfermedadesRaras2021, DíaMundialEnfermedadesRaras, DíaMundialEnfermedadesRaras2021, DiaDeLasEnfermedadesRaras, DiaMundialEnfermedadesRaras, RareDiseaseDay, RareDiseasesDay, RareDiseasesDay2021, DiaMundialDeLasEERR, DiaMundialDeLasEERR2021, DiaMundialDeLasEERR21, DiaMundialEERR.

La captura se realizó mediante la herramienta t-hoarder (Congosto, Basanta-Val, & Sanchez-Fernandez, 2017). Una vez obtenida la información, esta se procesó y analizó con diferentes herramientas informáticas, tal como diferentes scripts realizados en lenguaje Python v.3.7 para realizar el análisis descriptivo y el modelado de temas (para el que se utilizó la técnica de *Latent Dirichlet Allocation*). La herramienta Gephi v.0.91 (Bastian, Heymann, & Jacomy, 2009) se utilizó para realizar el análisis de redes sociales y la visualización de las diferentes redes involucradas.

4.3. Resultados

Una vez capturados los tuits intercambiados, se filtraron para obtener aquellos que estuviesen escritos en idioma español, obteniéndose un total de 18046 mensajes. Estos 18046 mensajes en español fueron creados por 11741 usuarios (cuentas) únicos, de los cuales 1742 usuarios crearon tuits con contenido original (siendo el resto retuits o respuestas). De las 258 asociaciones de enfermedades raras con cuenta activa en Twitter, solamente 88 participaron con algún tipo de contenido y, de ellas, 58 crearon tuits con contenido original. La ilustración 13 muestra algunas de las medidas más relevantes obtenidas.

	Total capturado	Generados por las asociaciones
Tuits con contenido original	2287	120
Retuits	12607	225
Retuits con cita	2865	49
Respuestas	287	12
TOTALES	18046	406

Ilustración 13. Estilos permitidos y su uso

4.3.1. Cuentas más participativas

En cuanto a las cuentas que más participaron (generando mensajes de cualquier tipo), la Federación Española de Enfermedades Raras (FEDER) es el líder en cuanto al total de interacciones realizadas y

también en la creación de tuits originales, tal como se puede ver en las ilustraciones 14 y 15, respectivamente.

Cuenta	Nº mensajes
@FEDER_ONG	324
@NoInvisibles	118
@JairoFSIE	116
@DM1Steinert	98
@1asufridoramas	38
@asociacion_php (*)	37
@leo_veritas	35
@pacienteqcuanta	34
@la_discapacidad	31
@AEGHgenetica	26
@FundacionDravet (*)	26
@RetoDravet	25
@LaPorres	25
@CIBERER	24
@todosconcesped	23
@cristin48712933	23
@beatrizce78	22
@asociacion_aler (*)	22

Ilustración 14. Cuentas de twitter más participativas

Cuenta	Nº tuits originales
@FEDER_ONG	66
@AEGHgenetica	16
@asociacion_php (*)	15
@DM1Steinert	13
@GVAasaludelche	11
@leo_veritas	9
@FFAMEX	8
@cdgsindrome (*)	7
@CIBERER	7
@SaberVivir_Tv	7
@RiveKids	7
@RarasReales	6
@demamasdepapas	6
@TribunaSal	6
@SaraST733	6
@ADERAH2020	6
@RetoDravet	6
@GA1Familia (*)	6

Ilustración 15. Cuentas de twitter con más contenido original

Las cuentas que corresponden a asociaciones de enfermedades raras se encuentran señaladas con un asterisco. Como se aprecia, no son demasiadas las que aparecen en los primeros lugares, ni en cuanto a participación total (ilustración 14) ni en cuanto a generación de tuits con contenido original (ilustración 15).

4.3.2. Mensajes más retuiteados y hashtags más utilizados

La ilustración 16 muestra los mensajes más retuiteados durante el día mundial de las enfermedades raras. Además de los tuits de organizaciones, individuos o instituciones relacionadas con las estas enfermedades, se puede ver que los mensajes de cuentas de personajes e instituciones políticas (marcados con dos asteriscos) suponen una importante parte de los contenidos más retuiteados.

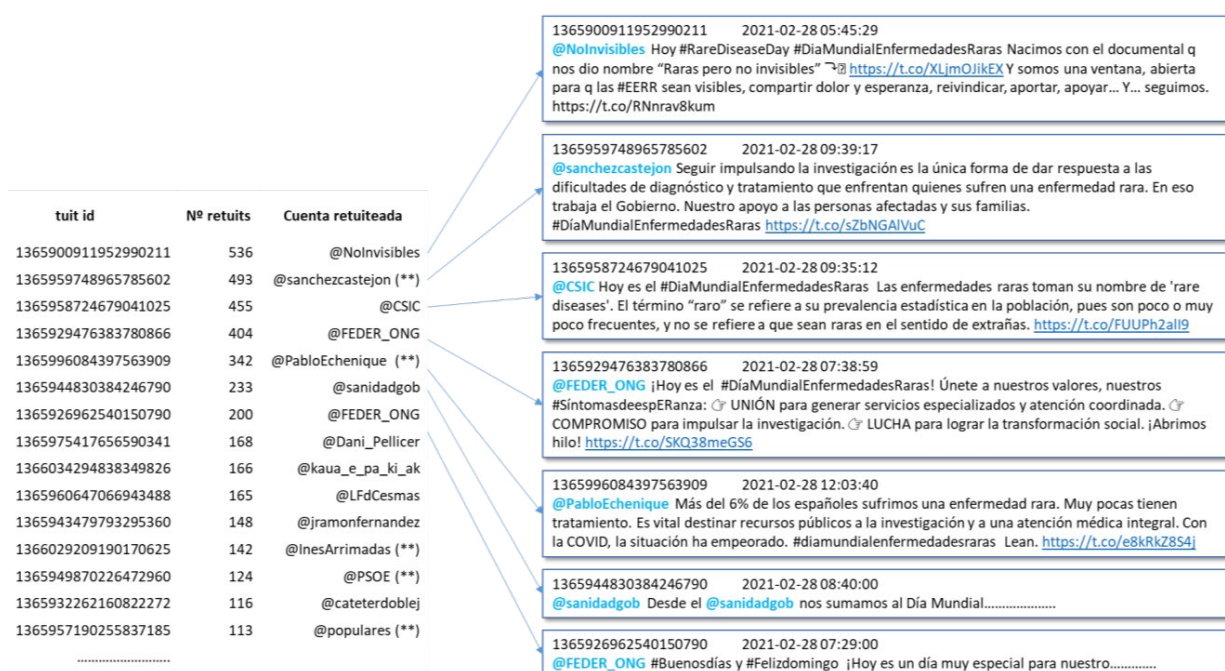


Ilustración 16. Mensajes más retuiteados

En cuanto a los hashtags más utilizados (ilustración 17) se puede comprobar que aquellos relacionados con el día mundial que se está celebrando, y con las propias enfermedades raras, son con diferencia los más utilizados. Es destacable también la presencia, entre los términos más utilizados, de un hashtag que denota contenido de crítica política.

hashtag	Nº veces utilizado
#diamundialenfermedadesraras	5151
#RareDiseaseDay	3316
#DiaMundialEnfermedadesRaras	2348
#EnfermedadesRaras	2205
#RareDiseaseDay2021	1242
#EERR	1168
#diamundialdelasenfermedadesraras	981
#DiaMundialDeEnfermedadesRaras	560
#enfermedadesraras	436
#diadelasenfermedadesraras	327
#somosFEDER	317
#sanchezmentiroso	226
#Buenosd	204
#FelizDomingo	202
#Felizdomingo	201

Ilustración 17. Hashtags más utilizados

4.3.3. Modelado de temas con Latent Dirichlet Allocation (LDA)

Con el fin de conocer los diferentes temas tratados en los mensajes intercambiados en Twitter se realizó un análisis de contenido basado en la técnica *Latent Dirichlet Allocation* (LDA) (Blei, Ng, & Jordan, 2003). La ilustración 18 muestra el resultado de este análisis para el conjunto de tuits capturados y eligiendo un número de temas igual a cuatro. Los nombres de los temas elegidos en base a las palabras asociadas se pueden ver en la última fila de la tabla.

Palabras Tema 0	Pesos Tema 0	Palabras Tema 1	Pesos Tema 1	Palabras Tema 2	Pesos Tema 2	Palabras Tema 3	Pesos Tema 3	Palabras Tema 4	Pesos Tema 4
día	595.2	apoyo	191.7	social	82.9	tratamiento	224.5	persona	256.7
mundial	247.2	visibilidad	152.2	colectivo	67.2	vida	185.1	millon	146.6
febrero	192.1	familia	132.0	ayuda	66.2	diagnóstico	154.2	cada	113.7
frecuent	94.1	patient	126.9	sanitaria	57.2	persona	102.9	año	96.8
celebra	78.6	gracia	101.0	esperanza	55.0	mejorar	87.2	españa	85.2
invis	70.2	persona	100.3	celebramo	52.2	calidad	82.2	mucha	78.2
Día mundial		Apoyo y visibilidad		Ayuda sociosanitaria		Diagnóstico y tratamiento		Personas afectadas	

Ilustración 18. Modelado de temas con LDA

4.3.4. Análisis de redes sociales

Con el fin de detectar a los principales actores participantes en el intercambio de mensajes en Twitter así como de estudiar las relaciones que se establecen entre ellos se aplicó la técnica del análisis de redes sociales (ARS) (Scott, 2000) a la red de retuits capturada. En esta red, los nodos representan cuentas de twitter y las relaciones dirigidas se interpretan como una cuenta (de la que sale la relación) que ha retuiteado un mensaje de otra (a la que llega la relación).

La ilustración 19 muestra cuales fueron las cuentas más retuiteadas. El tamaño del nodo representa el número de veces que esa cuenta fue retuiteada (es decir, el número de relaciones entrantes al nodo en la red, lo que en terminología ARS se denomina *indegree*). Como puede verse la federación FEDER, la cuenta RarasNoInvisibles y el organismo CSIC (Centro Superior de Investigaciones Científicas) conforman el núcleo central de la conversación. Hay una importante presencia de grupos relacionados con entidades y personalidades políticas que se relacionan estrechamente entre sí dentro de grupos afines.

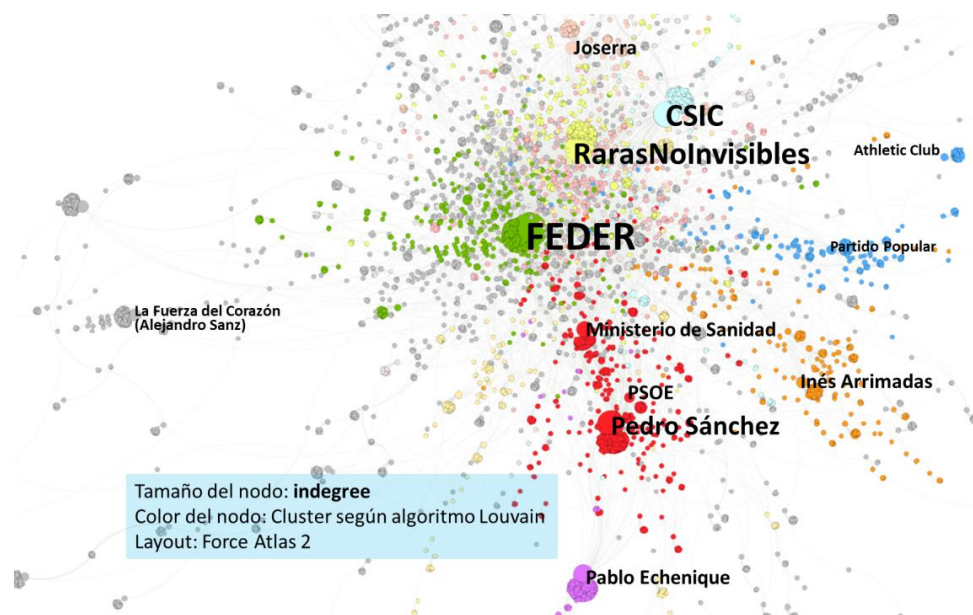


Ilustración 19. Red de retuits resaltando las cuentas más retuiteadas

La ilustración 20 muestra la misma red, pero representando mediante el tamaño de los nodos el nivel de influencia que tiene ese nodo. Este nivel de influencia se mide teniendo en cuenta no solamente los retuits recibidos por una determinada cuenta sino también la calidad de las cuentas que se relacionan con ésta. La métrica utilizada en análisis de redes sociales para medir el grado de influencia de un agente se denomina *eigenvector*. Observando la ilustración 20 vemos que las cuentas más influyentes son las mismas que más retuits recibieron (comparando con la ilustración 19).

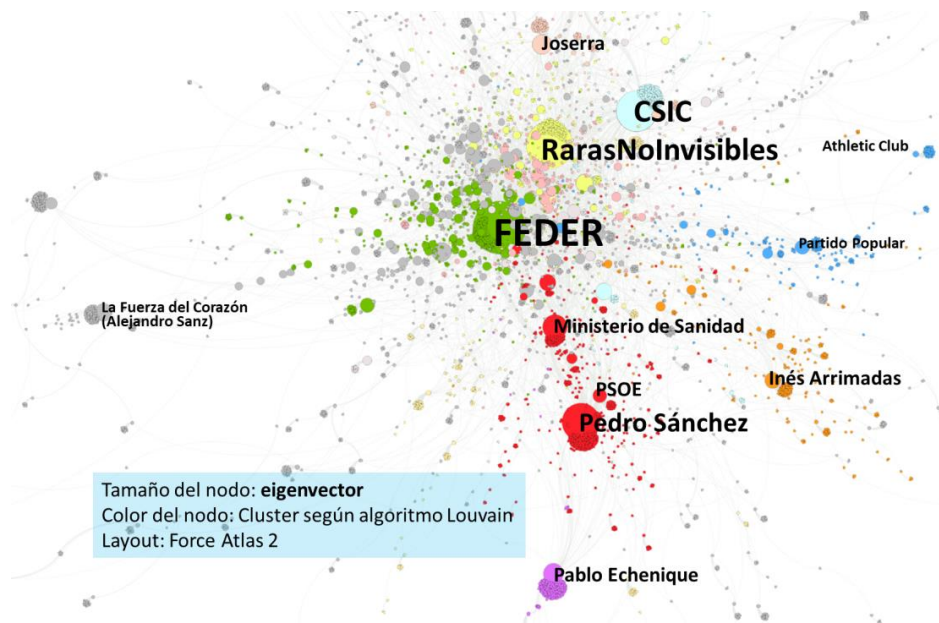


Ilustración 20. Red de retuits resaltando las cuentas más influyentes

Por último, se puede estudiar la capacidad de intermediación de las diferentes cuentas de usuario. En ARS esto se lleva a cabo calculando la métrica denominada *betweenness*. La intermediación de un nodo mide su capacidad para comunicar varios grupos de agentes relativamente separados en la red, por lo que es una métrica muy útil para conocer qué actores de la red pueden difundir una determinada información para que ésta llegue al mayor número de estos grupos. La ilustración 21 representa la misma red de retuits pero utilizando el tamaño del nodo para resaltar el nivel de intermediación del mismo.

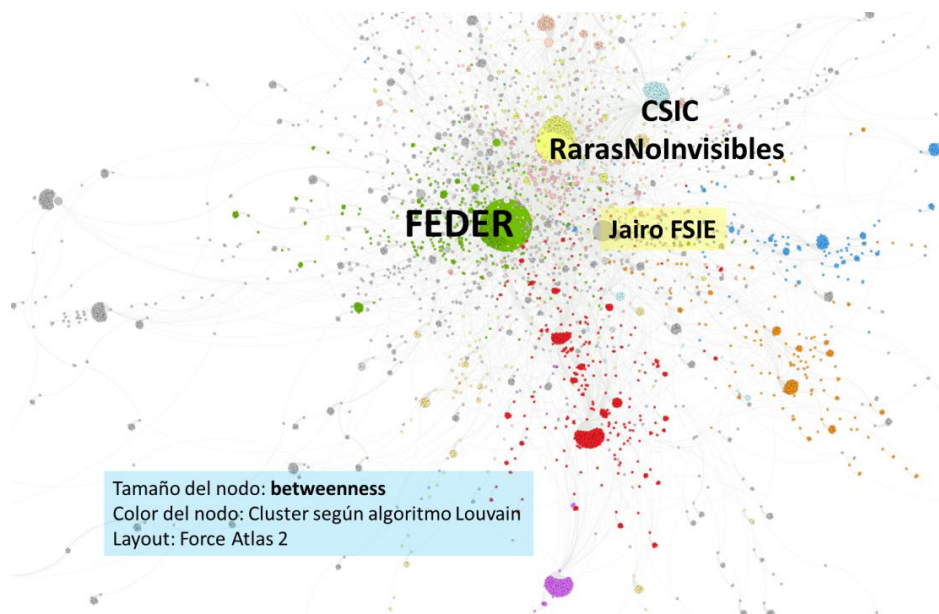


Ilustración 21. Red de retuits resaltando las cuentas con mayor capacidad de intermediación

En este caso, se puede comprobar que las cuentas relacionadas con la política o el gobierno y sus instituciones no tienen capacidad de intermediación en esta red, mientras que las cuentas de FEDER, RarasNoInvisibles y CSIC mantienen su importancia en la red respecto de esta característica. Además, aparece un actor nuevo con un importante nivel de intermediación: “Jairo FSIE”, se trata de un usuario muy activo en Twitter que habla de enseñanza y discapacidad.

4.4. Conclusiones

En este trabajo se ha realizado un análisis de los mensajes intercambiados, en idioma español, durante el día mundial de las enfermedades raras del año 2021 en la red social Twitter. Los actores principales fueron las organizaciones “paraguas” que agrupan a las asociaciones de pacientes, así como las dedicadas a promover la investigación o el apoyo social en relación con estas enfermedades. Los personajes públicos y cuentas del ámbito político también son actores relevantes este día ya que, aunque generan poco contenido, éste es difundido ampliamente por sus seguidores. Los principales temas de discusión detectados fueron los relacionados con la celebración del propio día mundial, la necesidad de concienciación de la sociedad sobre la importancia de este problema sociosanitario, la necesidad de apoyo social o el impulso de la investigación en relación con estas enfermedades. Este trabajo ha permitido comprobar que el análisis de datos procedentes de redes sociales online puede ser utilizado por los gestores y profesionales sanitarios para conseguir información relevante a la hora de mejorar los servicios prestados a estos pacientes y sus familiares.

4.5. Referencias

- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Third International Conference on Weblogs and Social Media, ICWSM 2009*. San Jose, California, USA. <https://doi.org/10.13140/2.1.1341.1520>
- Benitez-Andrades, J. A., Alija-Perez, J. M., Garcia-Rodriguez, I., Benavides, C., Alaiz-Moreton, H., Vargas, R. P., & Garcia-Ordas, M. T. (2021). BERT Model-Based Approach For Detecting Categories of Tweets in the Field of Eating Disorders (ED). In *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 586–590). IEEE. <https://doi.org/10.1109/CBMS52027.2021.00105>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Castillo-Esparcia, A., & López-Villafranca, P. (2016). Communication strategies employed by rare disease patient organizations in Spain. *Ciencia e Saude Coletiva*. <https://doi.org/10.1590/1413-81232015218.19852015>
- Congosto, M., Basanta-Val, P., & Sanchez-Fernandez, L. (2017). T-Hoarder: A framework to process Twitter data streams. *Journal of Network and Computer Applications*, 83, 28–39. <https://doi.org/10.1016/j.jnca.2017.01.029>
- EURORDIS. (2021). *What is a rare disease?* Retrieved from <https://www.eurordis.org/content/what-rare-disease>
- Kloth, Y. M., Deutsch, K. M., Danielson, K. A., Strack, J., & Law, C. (2019). What Twitter teaches us about patient-provider communication on pain. *PLoS ONE*, 14(12), 1–19. <https://doi.org/10.1371/journal.pone.0226321>
- Merkel, P. A., Manion, M., Gopal-Srivastava, R., Groft, S., Jinnah, H. A., Robertson, D., & Krischer, J. P. (2016). The partnership of patient advocacy groups and clinical investigators in the rare diseases clinical research network. *Orphanet Journal of Rare Diseases*, 11(1), 66. <https://doi.org/10.1186/s13023-016-0445-8>
- Pershad, Y., Hangge, P., Albadawi, H., & Oklu, R. (2018). Social Medicine: Twitter in Healthcare. *Journal of Clinical Medicine*, 7(6), 121. <https://doi.org/10.3390/jcm7060121>
- Scott, J. (2000). *Social Network Analysis: A Handbook* (2nd ed.). London: SAGE Publications.

-
- Silibello, G., Vizziello, P., Gallucci, M., Selicorni, A., Milani, D., Ajmone, P. F., ... Lalatta, F. (2016). Daily life changes and adaptations investigated in 154 families with a child suffering from a rare disability at a public centre for rare diseases in Northern Italy. *Italian Journal of Pediatrics*, 42(1), 76. <https://doi.org/10.1186/s13052-016-0285-0>
- Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: A systematic review. *American Journal of Public Health*, 107(1), e1–e8. <https://doi.org/10.2105/AJPH.2016.303512>
- Vicari, S., & Cappai, F. (2016). Health activism and the logic of connective action. A case study of rare disease patient organisations. *Information Communication and Society*. <https://doi.org/10.1080/1369118X.2016.1154587>
- von Gizycki, R. (2010). Contribution of rare disease patient organisations to medical education. *Orphanet Journal of Rare Diseases*, 5(Suppl 1), O25. <https://doi.org/10.1186/1750-1172-5-S1-O25>