

El potencial *de las* redes basadas en *la* API Google Vision para *el* estudio *de* imágenes digitales *nativas*

Cómo citar este artículo: Omena, J. J., Pilipets, E., Gobbo, B., & Chao, J. (2021). El potencial de las redes basadas en la API Google Vision para el estudio de imágenes digitales *nativas*. *Diseña*, (19), Article.1.
<https://doi.org/10.7764/disena.19.Article.1>

DISEÑA | 19 |
AGOSTO 2021

ISSN 0718-8447 (impreso)
2452-4298 (electrónico)

COPYRIGHT: CC BY-SA 4.0 CL

Artículo de investigación original

Recepción

11 ENE 2021

Aceptación

08 JUL 2021

[Original English Version here](#)

Janna Joceli Omena

Center for Advanced Internet Studies (CAIS)

Elena Pilipets

Universität Klagenfurt

Beatrice Gobbo

Politecnico di Milano

Jason Chao

Universität Siegen

En este artículo presentamos los potenciales de las redes basadas en la API Google Vision para el estudio de las imágenes en línea, abordando tres modalidades importantes como parte de una metodología visual crítica: el contenido de la propia imagen, su forma específica de “audienciación” a través de referencias *web* (o metadatos de la imagen) y los sitios de circulación de la imagen. En primer lugar, definimos conceptual y técnicamente diferentes redes construidas a partir de ciertas características de visión artificial: *imagen-etiqueta*, *imagen-entidades web* e *imagen-dominio*. En segundo lugar, presentamos un diagrama de protocolo de investigación que ilustra cómo construir redes de imágenes con sus respectivas descripciones o sitios de circulación. En tercer lugar, discutimos las potencialidades de las redes de visión artificial como dispositivos de investigación, enfatizando sus (trans) formaciones relacionales de datos y sus especificidades interpretativas. Se presentarán tres diferentes estudios de caso como ejemplo. En conclusión, sostenemos que una metodología visual de este tipo requiere prácticas técnicas críticas que tengan en cuenta las múltiples capas de mediación técnica que están involucradas.

Palabras clave

Análisis de imágenes

Redes

Métodos digitales

Metodologías visuales

Visión artificial

Janna Joceli Omena—Máster en Cultura Contemporánea y Nuevas Tecnologías, Universidade NOVA de Lisboa. Becaria de investigación en el Center for Advanced Internet Studies (CAIS). Es miembro del iNOVA Media Lab y del Public Data Lab. Su investigación se centra en los métodos digitales, los estudios de redes digitales y la “tecnicidad de los medios” para apoyar la investigación social y de medios. Es editora de *Métodos Digitais: Teoria-Prática-Crítica* (iNOVA, 2019) y coordinadora del SMART Data Sprint. Algunas de sus últimas publicaciones son: “Digital Methods for Hashtag Engagement Research” (con E.T. Rabello y A.G. Mintz; *Social Media + Society*, vol. 6, n° 3) y “Call into the Platform!” (con A. Granado; *Icono14*, vol. 18, n° 1).

Elena Pilipets—Doctora en Ciencias de la Comunicación, Universität Klagenfurt. Investigadora postdoctoral en el Departamento de Medios y Comunicaciones de la Universität Klagenfurt e investigadora SMART (Social Media Research Techniques) en el iNOVA Media Lab de la Universidade NOVA de Lisboa. Sus intereses docentes y de investigación están relacionados con los estudios culturales de los medios de comunicación, la investigación en Internet y los métodos digitales. Es la actual titular del 2020 Carinthian Award for Young Social Sciences and Humanities Scholars. Entre sus publicaciones más recientes figuran “Nipples, Memes, and Algorithmic Failure: NSFW Critique of Tumblr Censorship” (con S. Paasonen; *New Media & Society*, 1ª publicación, 2020) y “Digitale Medien und Methoden. Über den methodischen Umgang mit visuellen Plattformhalten und Internet-Memes” (*Open-Media-Studies-Blog der Zeitschrift für Medienwissenschaft*).

Beatrice Gobbo—Máster en Diseño de la Comunicación, Politecnico di Milano. Estudiante de doctorado en Diseño en el Politecnico di Milano. Es miembro del DensityDesign Lab, grupo de investigación centrado en la visualización y el diseño de información. Su investigación actual se centra en el papel del diseño de comunicación y la visualización de información en el campo de la inteligencia artificial explicable. Entre sus publicaciones recientes se incluyen: “Research Protocol Diagrams as Didactic Tools to Act Critically in Dataset Design Processes” (con M. Mauri, M. A. Briones y G. Colombo; en *INTED2020 Proceedings*) y “Explaining AI through Critical Reflection Artifacts. On the Role of Communication Design within XAI” (en T. Reis, M. X. Bornschlegl, M. Angelini y M. L. Hemmje, eds.; *Advanced Visual Interfaces. Supporting Artificial Intelligence and Big Data Applications*; Springer, 2020).

Jason Chao—Máster en Big Data y Futuros Digitales, University of Warwick. Máster en Derechos Humanos, University of London. Es investigador y candidato a doctor en la Universität Siegen. Su investigación actual se centra en el desarrollo de herramientas de métodos digitales para el estudio de sensores de medios. Cuenta con experiencia en el desarrollo de *software* y en la defensa de los derechos humanos. Algunos de los *softwares* de investigación que ha desarrollado recientemente son *AppTraffic*, para el estudio del tráfico de red de las aplicaciones móviles; y *Offline Image Query and Extraction Tool* (con J. J. Omena).

El potencial de las redes basadas en la API Google Vision para el estudio de imágenes digitales nativas

Janna Joceli Omena

Center for Advanced Internet Studies (CAIS)
Bochum, Alemania
janna.joceli.omena@cais.nrw

Elena Pilipets

Universität Klagenfurt
Departamento de Medios y Comunicaciones
Klagenfurt, Austria
Elena.Pilipets@aau.at

Beatrice Gobbo

Politecnico di Milano
DensityDesign Lab
Milán, Italia
beatrice.gobbo@polimi.it

Jason Chao

Universität Siegen
Centro Colaborativo de Investigación "Medios de Cooperación"
Siegen, Alemania
teng.chao@uni-siegen.de

INTRODUCCIÓN

Al discutir las potencialidades de los métodos digitales para estudiar materiales visuales, Gillian Rose (2016) destaca la integración de las imágenes dentro de las infraestructuras digitales de "audienciación" (*audiencing*). Las máquinas forman parte de esta audiencia (Hoelzl & Marie, 2015; Paglen, 2014; Rettberg, 2020). Al organizar las imágenes en formaciones replicables, buscables y propagables, las interacciones nativamente digitales presentan tanto un proceso iterativo como un resultado distribuido (Hochman, 2014; Steyerl, 2009). La imagen ya no opera únicamente como un objeto icónico de representación, sino que se reconfigura como una superficie computable dentro de un conjunto de técnicas de aprendizaje automático, culturas de clasificación, algoritmos de recomendación y entornos de plataforma a través de los cuales se operacionaliza (Azar et al., 2021; MacKenzie & Munster, 2019; Rubinstein & Sluis, 2013). Las dinámicas que hacen que las imágenes puedan ser analizadas son específicas del sitio y están informadas por varios registros digitales (por ej., *hashtags*, marcas de tiempo y enlaces) (Gerlitz & Rieder, 2018). En consecuencia, las imágenes digitales *nativas* siempre se conectan en red a través de datos específicos de la plataforma y datos nativos de la *web*, produciendo enormes cantidades de información visual, textual y numérica "a escala" (Manovich, 2020; Parikka & Dvořák, 2021).

Este artículo se adentra en el estudio del material visual mediante métodos digitales (Rogers, 2013), una aproximación en red a la investigación visual (Colombo, 2018; Niederer & Colombo, 2019) y la visión artificial. Hasta ahora, los

métodos que combinan el análisis visual automatizado con los metadatos de las plataformas han demostrado el potencial que presentan los enfoques cuali-cuantitativos para comprender la interacción entre los patrones narrativos de las imágenes y las técnicas de distribución de contenidos dentro, y a través de, las diferentes culturas de uso de las plataformas (Niederer & Colombo, 2019). Algunos ejemplos son el mapeo semántico de ciudades (Ricci et al., 2017), los estudios sobre la circulación de imágenes (d'Andréa & Mintz, 2019), la comunicación institucional (Omena & Granada, 2020) y las *affordances* afectivas de los públicos de *hashtags* (Geboers & Van De Wiele, 2020). Las múltiples características que se pueden extraer automáticamente de una imagen mediante la visión artificial—como objetos, rostros, fondos—plantean nuevas preguntas de interpretabilidad junto con la necesidad de cambiar entre diferentes escalas de observación, referencia y comparación (Manovich, 2020).

En este artículo expandimos esta trayectoria analítica reutilizando las características de Google Vision como parte de una metodología visual crítica.¹ Discutimos (1) el contenido de la imagen en sí misma; (2) su “audienciación” específica a través de referencias obtenidas de la *web*; y (3) los sitios de circulación de imágenes (Rose, 2016). Para ello, definimos conceptual y técnicamente tres redes diferentes: la red *imagen-etiqueta*, que permite estudiar el contenido y la composición visual a través de anotaciones literales de la propia imagen; la red *imagen-entidades web*, que posiciona las imágenes en relación con el contenido de la *web*, utilizando los sistemas de clasificación de Google y el Knowledge Graph como fuente para entregar anotaciones *web*; y la red *imagen-dominio*, que agrupa las imágenes en conexión con páginas *web* o localizadores de recursos uniformes (URL) que alojan imágenes total (o parcialmente) coincidentes. Si bien reconocemos que existe incertidumbre a la hora de definir *cómo* asigna sus diferentes valores la API Google Vision, analizamos las redes de visión artificial en su potencial para situar las imágenes digitales *nativas* a través de sus (trans)formaciones relacionales de datos.

¹ Gillian Rose (2016) propone cuatro modalidades de interpretación de los materiales visuales: sitios de producción, la imagen en sí, su circulación y su “audienciación” (*audiencing*).

CARACTERIZACIÓN DE LAS REDES DE VISIÓN ARTIFICIAL

Nos referimos a las *redes de visión artificial* como una propuesta inspirada en el trabajo de Donato Ricci, Gabriele Colombo, Axel Meunier y Agata Brilli (2017), desarrollada luego por Colombo (2018). Mapeando los elementos simbólicos y materiales del debate sobre la naturaleza urbana en París, dichos autores fusionaron una colección de imágenes de Twitter—vinculadas por la palabra clave “naturaleza” (*nature*)— con una red utilizando una API de visión basada en la *web* (API de IMAGGA). Este trabajo ha inspirado a otros investigadores (Geboers & Van De Wiele, 2020; Mintz et al., 2019; Omena et al., 2020; Silva et al., 2018), quienes exploran una característica particular de la visión artificial: el etiquetado de contenido visual según tags/etiquetas predefinidas, proporcionadas por modelos de aprendizaje automatizado. Esta característica converge en lo que denominamos redes *imagen-etiqueta*,

mientras otras características permanecen aún poco estudiadas, como la detección de entidades *web* o páginas *web* con imágenes completas coincidentes que, respectivamente, permiten lo que denominamos redes de visión artificial *imagen-entidades web* e *imagen-dominio*. Estas últimas introducen nuevas articulaciones para la interpretación de imágenes en línea, incluyendo su valor potencial para los estudios de circulación de imágenes y contextualización de imágenes. A continuación, ofrecemos información conceptual y técnica acerca de qué son y de qué están hechas las redes de visión artificial, al tiempo que presentamos un *software* de investigación de código abierto creado para apoyar la API Google Vision.

¿De qué redes estamos hablando?

Nos referimos a las redes como un espacio de conexiones capturadas *en y por* los registros digitales (ver Venturini et al., 2015, 2019), pero que responden a «los procesos tecnológicos inherentes al entorno *web* y a las APIs *en los que y a través de los cuales* la comunicación en línea se estructura, se captura y se fusiona con otros registros» (Omena, en prensa; ver también Gerlitz & Rieder, 2018). Esta última situación se refiere a la *gramatización* de la plataforma, que en este sentido y más específicamente, refleja cómo las imágenes digitalmente nativas *participan en y son mediadas por* las infraestructuras tecnológicas mientras están integradas en las culturas de uso. Es crucial distinguir entre las redes directamente proporcionadas por los métodos de recopilación de datos (mediante API, *scraping* o *crawling*) y aquellas construidas sobre los datos de la *web* y el uso de una variedad de *softwares* (ver Omena & Amaral, 2019). Mientras las primeras se *diseñan a través de archivos preconfeccionados (ready-made files)*, las segundas se *construyen*. Por ejemplo, cuando es proporcionada directamente por las APIs, la red estudiada es el resultado de un buen diseño de investigación y el uso de una aplicación basada en la *web* para extraer datos (por ej., YouTube Data Tools) de una API (por ej., YouTube Data API). El archivo de salida (por ej., el formato de conjunto de datos de grafos, abreviatura de .GDF) alimenta un *software* (por ej., Gephi) para visualizar y analizar la red (por ej., las redes de conexiones de los canales de YouTube). Sin embargo, las *redes de visión artificial* deben ser construidas; por lo tanto, exigen ejecutar una lista más larga de tareas prácticas, ya que las herramientas de recopilación de datos no pueden entregar *archivos preconfeccionados* de redes de imágenes que se puedan descargar y visualizar fácilmente. Las redes de las que hablamos están hechas de imágenes digitales *nativas* que se reconstruyen a través de la visión artificial y las elecciones que el investigador tiene que hacer a lo largo del proceso de curatoría, construcción, visualización y análisis de la propia red. Entendemos entonces las *redes de visión artificial* como un conjunto de medios computacionales, datos, métodos, investigación y prácticas técnicas orquestadas por el investigador (o los investigadores). De una u otra forma, estas redes demandan cierto dominio de conocimientos y prácticas

técnicas (imponiendo desafíos desde el punto de vista metodológico), mientras introducen nuevas formas y formatos para entender la sociedad y nuestras acciones cotidianas (ver Venturini et al., 2019).

Comprender las características de la visión artificial

La visión artificial corresponde a la capacidad de la computadora para reconocer características visuales mediante técnicas algorítmicas, utilizando estos aprendizajes para identificar y clasificar objetos y escenas (Szeliski, 2021). La comunidad que desarrolla la visión artificial crea, entrena y prueba modelos de aprendizaje automático y redes neuronales para clasificar imágenes, pero el uso de esta tecnología es diverso y se adapta a diferentes propósitos. Para el mercado o las agencias gubernamentales de seguridad, ofrece servicios como la moderación de contenidos (por ej., reconociendo imágenes ofensivas o no deseadas o detectando contenidos racistas o para adultos), el análisis predictivo o la controvertida capacidad de reconocimiento facial. En la investigación social, la visión artificial ha sido útil para detectar sesgos de género y para estudiar la desinformación visual o la ideología política de las imágenes (ver Garimella & Eckles, 2020; Schwemmer et al., 2020; Xi et al., 2020). En este artículo introducimos el potencial de las redes basadas en la API Google Vision para estudiar imágenes digitales *nativas*. A continuación presentamos con más detalle las características de Google Vision, justificando el uso de la API y aclarando el papel de las redes como dispositivo de investigación.

La API, lanzada en mayo de 2017, ofrece varias funciones, de las cuales sugerimos utilizar la *detección de etiquetas* y la *detección de entidades y páginas web*. En general, las *etiquetas* y las *entidades web* proporcionan descripciones textuales de las imágenes, mientras que las *páginas web* describen dónde se encuentran las imágenes. En concreto, la *detección de etiquetas* proporciona tanto etiquetas generales como específicas, con la puntuación de confianza y la clasificación temática proporcionada por modelos de aprendizaje automático. Su nivel de especificidad ha demostrado ser el más detallado cuando se trata de la clasificación de imágenes y cuando se la compara con otras API, a saber, IBM Watson y Microsoft Azure (Mintz et al., 2019; Silva et al., 2020). Google Vision también posee características que no se encuentran en otras APIs de visión artificial: la detección de *entidades web* y *páginas web*. Esto proporciona contenido *web* a una imagen, utilizando el poder de Google Image Search para «detectar entidades temáticas como noticias, eventos o celebridades dentro de una imagen» (Google Cloud, 2017) y para encontrar imágenes idénticas o similares en una lista de URLs o páginas web.² La URL de la página, el título de la página y la URL de la imagen coincidente son fuentes que suministran imágenes coincidentes completas o parciales a las páginas *web*, mientras que Google Knowledge Graph proporciona información sobre una *entidad web* o lo que la define (una persona, un lugar o una cosa). El Knowledge Graph es «un sistema

² Nuestra experiencia encontrando poca documentación explicativa sobre Google Vision refleja un problema ya señalado por Timnit Gebru y sus colegas: la falta de documentación adecuada sobre la procedencia de los datos y sobre la creación y el uso de conjuntos de datos en la comunidad de aprendizaje automático (Gebru et al., 2020).

que entiende hechos e información sobre entidades a partir de materiales compartidos en la *web*, así como a partir de bases de datos de código abierto y con licencia», reuniendo más de 500 mil millones de datos sobre cinco mil millones de entidades (Sullivan, 2020). Esto nos indica que las entidades *web* proporcionan *referencias a una imagen obtenida del entorno web*, que va más allá del contenido de la propia imagen. Suponemos que la detección de entidades *web* y páginas *web* puede incluir los sistemas de clasificación de Google, los datos de licencias, las fuentes públicas e información fáctica recibida directamente de los propietarios de los contenidos.

¿Cómo pueden los investigadores acceder a las funciones de Google Vision y utilizarlas en la práctica? A lo largo de los años, los archivos de secuencia de comandos han sido un elemento técnico crucial para invocar a las APIs de visión artificial. Estas herramientas imponen algunos desafíos técnicos a los investigadores sociales, especialmente a los que no son desarrolladores o codificadores. En este contexto, presentamos Memespector Graphical User Interface (GUI) (Chao, 2021), una herramienta de *software* para investigar desarrollada por uno de los autores, que sirve para el estudio de grandes colecciones de imágenes a través del lente de la visión artificial. Memespector GUI se basa en el proyecto original *Memespector*, de Bernhard Rieder (posteriormente ampliado a una secuencia de comandos de Python por André Mintz), reduciendo las barreras técnicas para que los investigadores *invoken fácilmente* a la API Google Vision (como demostraremos más adelante).

Combinando las características de Google Vision y las imágenes como redes

En esta sección explicamos cómo se reordenan las características de Google Vision junto a las imágenes como redes, presentando su potencial para las metodologías visuales. La Figura 1 muestra una imagen del mosquito (*Aedes aegypti*) que transmite el virus del Zika y los *outputs* de Google Vision para detectar *etiquetas*, *entidades web* y *páginas web*. Al observar la imagen que contiene el mosquito, la función de *etiquetas* entrega una descripción literal, “insecto” e “invertebrado”, pero también produce intentos menos fiables: “araña tejedora de círculos” y “organismo”. Mientras las *etiquetas* no permiten a los usuarios ir más allá del contenido que se puede leer a través de las imágenes, las *entidades web* son capaces de describir la misma imagen reconociéndola, por ejemplo, como una “enfermedad transmitida por un mosquito, el virus del Zika”. Estas *entidades web* son responsivas a los datos extraídos de las páginas en las que se encontró la imagen del mosquito (Robinson, 2017) y a las entidades existentes de Google Knowledge Graph, descripciones arraigadas en el contenido *web* y a través de la lente de los sistemas de clasificación de la búsqueda de imágenes de Google (Google Cloud, 2017; Robinson, 2017). La descripción de la imagen adquiere diferentes significados dependiendo de la gramática tecnológica que se utilice.

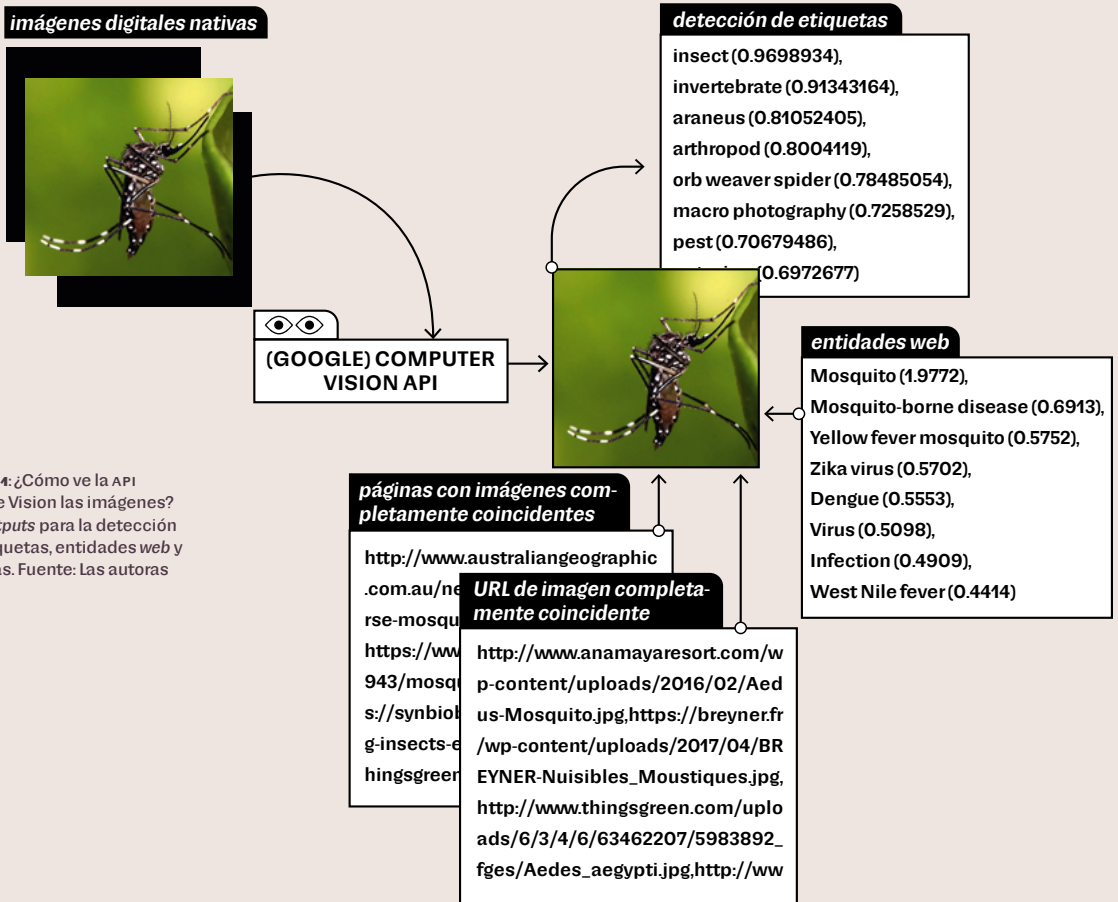


Figura 4: ¿Cómo ve la API Google Vision las imágenes? Los *outputs* para la detección de etiquetas, entidades *web* y páginas. Fuente: Las autoras

Las etiquetas van siempre acompañadas de una puntuación de confianza y son clasificadas por el índice temático, informando la probabilidad de las descripciones textuales asignadas a una imagen. Como vemos en la Figura 1, la descripción de la imagen del mosquito comienza con etiquetas genéricas ordenadas por temas que contienen la puntuación de confianza más alta, por ejemplo, “insecto” (0,9698934) y “mosquito” (1,9772), pasando a una clasificación más detallada, “artrópodo” (0,8004119) e “infección” (0,4909). Las descripciones factuales o más contextualizadas que proporcionan las *entidades web* permiten diferentes lecturas de un mismo contenido visual y pueden complementarse entre sí.

Ahora observamos la misma imagen, pero prestando atención a sus sitios de aparición mediante la *detección de páginas web*, lo que proporciona una lista de páginas *web* o URLs en las que se encuentran imágenes que coinciden total o parcialmente. Esta función ofrece una perspectiva dinámica acerca de la manera en que fluyen las imágenes a través de las plataformas (URL) y qué actores —(sub) dominios de enlace— se involucran con dicha visualidad. Por ejemplo, la imagen del

mosquito ha aparecido en revistas científicas (collections.plos.org), clínicas de salud (tratamientoendometriose.com.br), medios de comunicación (correiodoestado.com.br), redes sociales y blogs (i.pinimg.com; 1.bp.blogspot.com).

Con esto en mente, a continuación presentamos tres tipos de redes construidas a partir de los *outputs* de la API Google Vision para la detección de *etiquetas*, *entidades web* y *páginas web*, correspondientes a (1) *imagen-etiqueta*, (2) *imagen-entidades web* y (3) *imagen-dominio*. Estas redes tienen dos tipos de nodos (grafos bipartitos): mientras uno es siempre la imagen, el otro representa la característica elegida por la API de visión.

1. *Redes de imágenes y sus capas descriptivas* (clasificación de imágenes basada en etiquetas predefinidas/personalizadas). Los nodos son imágenes y etiquetas. Las conexiones implican la aparición de etiquetas en relación con las imágenes.

2. *Redes de imágenes y sus respectivas entidades web* (clasificación de imágenes basada en contenido *web*). Los nodos son imágenes y *entidades web*. Las conexiones implican la aparición de *entidades web* en relación con las imágenes.

3. *Redes de imágenes y sitios de circulación en la web* (referidas a las URLs o *páginas web* en que se encuentran las imágenes). Los nodos son imágenes y (sub)dominios. Las conexiones implican la aparición de (sub)dominios en relación con las imágenes.

La Figura 2 ilustra el aspecto que pueden tener las tres redes (sin las imágenes), mostrando la misma colección de imágenes (relacionadas con #microcefalia) que se interpretan de diferente manera según la gramática tecnológica de la API Google Vision. Una *perspectiva estática* del contenido visual sigue las capacidades descriptivas de las *etiquetas* y las *entidades web*, mientras que la *detección de páginas web* implica una *perspectiva dinámica* que sigue los sitios de circulación de las imágenes. Al leer una red de descripciones de imágenes, las etiquetas generales, seguidas de otras más específicas, determinan la posición de las imágenes. La espacialización de la red se relaciona con lo que hay *en* la imagen (detección de etiquetas) o con la manera en que se describe y contextualiza la imagen *dentro* del entorno *web* y *a través* del lente de Google (detección de *entidades web*). En la red, la posición de las imágenes pone de relieve sus descripciones literales o contextuales sin relación con la idea de movimiento. En una red de circulación de imágenes (detección de *páginas web*), la posición de los nodos está determinada por los sitios de aparición de las imágenes; en consecuencia, la interpretación de las imágenes no es estática sino dinámica. Estas perspectivas ayudan a la práctica de leer las redes de visión artificial, como demostraremos en los estudios de caso.

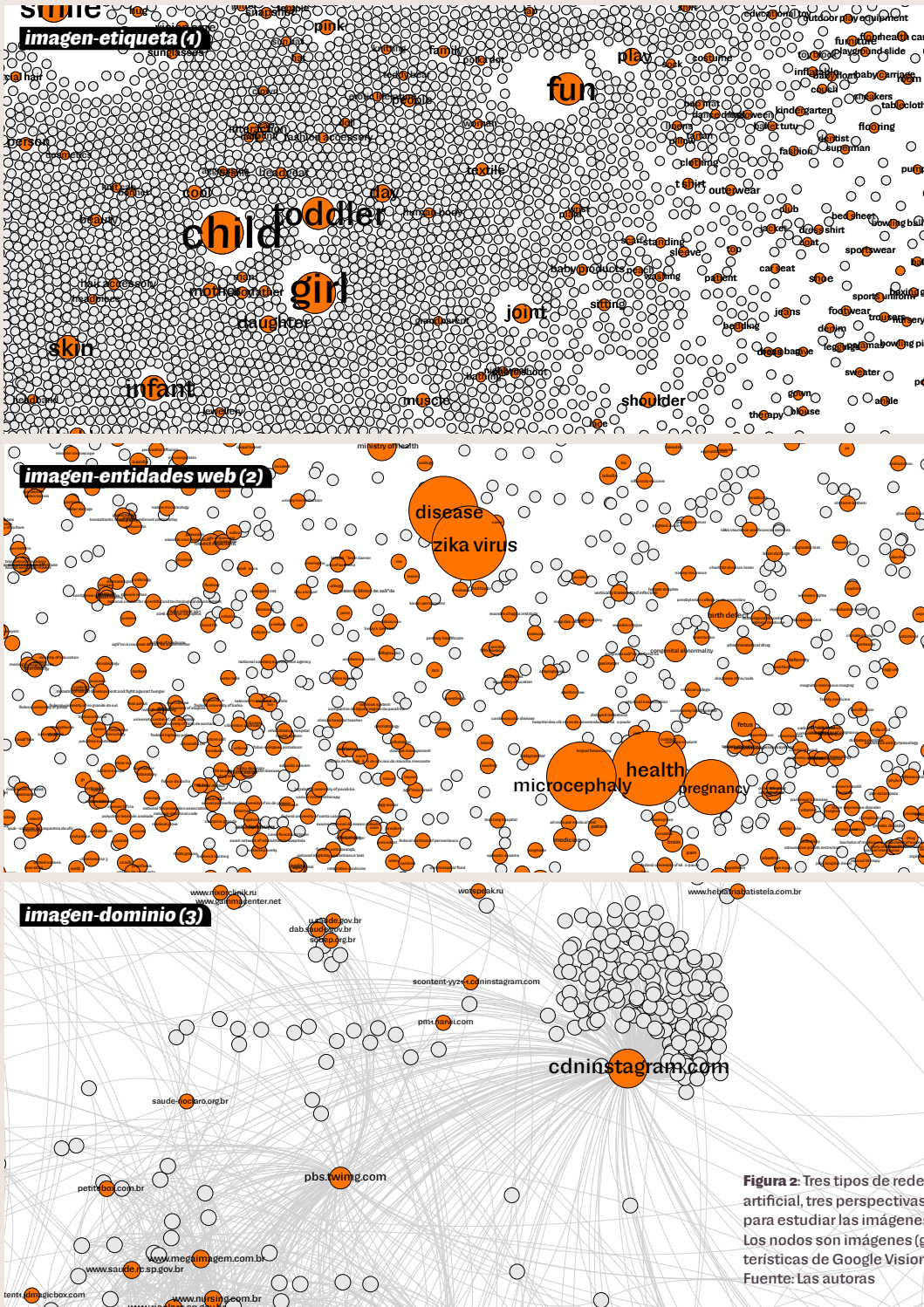


Figura 2: Tres tipos de redes de visión artificial, tres perspectivas diferentes para estudiar las imágenes en línea. Los nodos son imágenes (gris) y características de Google Vision (naranja). Fuente: Las autoras

EL DIAGRAMA DEL PROTOCOLO DE INVESTIGACIÓN

Esta sección muestra el proceso metodológico para crear redes de visión artificial, haciendo hincapié en las transformaciones de las imágenes a lo largo de este procedimiento. Este enfoque de métodos digitales requiere el compromiso de los investigadores con las prácticas técnicas (diseñar e implementar la investigación con estos métodos), invitando a los investigadores a familiarizarse con los medios computacionales en términos conceptuales, técnicos y empíricos (Omena, en prensa). Esto incluye una mirada crítica a la plataforma que aloja las imágenes analizadas y también comprender que la API de visión artificial y el *software* de investigación elegidos son medios para «expresar una voluntad y un modo de conocer» (Rieder & Röhle, 2018, p. 123).

El diagrama del protocolo (Figura 3) muestra lo que se necesita para construir e interpretar una red de visión artificial, teniendo en cuenta su contenido técnico, que refleja la actividad en línea situada, las *gruesas* capas de mediación técnica y las intervenciones de los investigadores. El protocolo pretende ser una herramienta visual autoexplicativa y autosoportada que destaca tanto el diseño del conjunto de datos como el proceso de visualización (Mauri et al., 2020; Niederer & Colombo, 2019), proporcionando también un artefacto de difusión que permite replicar el mismo método para construir otras redes sobre las características de diferentes APIs de visión artificial. Llamamos la atención (ver Figura 4) sobre el papel del investigador, quien, como argumentamos, debe dar cuenta del contenido de los medios técnicos, mientras se compromete con las prácticas técnicas. Los cinco pasos del protocolo se presentan a continuación.

El proceso comienza con el *arte de la consulta* (Figura 4), que se refiere a la capacidad del investigador para formular preguntas de investigación a través de consultas de búsqueda (Rogers, 2019; Taibi et al., 2016), dando cuenta de la formulación de consultas especificadas (por ej., #femalepresentingnipples) o subespecificadas (por ej., #microcefalia; emergencia climática) para seleccionar una lista de URLs de imágenes como punto de partida. Elegir las palabras clave es tan importante como los conocimientos técnicos requeridos cuando se utiliza un *software* de extracción o una secuencia de comandos de Python para recopilar imágenes de las plataformas digitales. Los investigadores deben reconocer que estas herramientas no pueden funcionar solas, sino que están condicionadas por sus elecciones, decisiones y conocimientos técnicos.

Una vez que se tiene a mano el archivo (o los archivos) de salida del *software* de extracción, es crucial *ver y comprobar qué metadatos tiene para ofrecer la imagen* (por ej., métricas de compromiso, marcas de tiempo, nombre de usuario, URLs) y *qué puede servir (más adelante) como información analítica valiosa*. También es aconsejable descargar lo antes posible las imágenes³ debido a la corta vida de sus URLs (Figura 5). En este caso, la máscara utilizada para nombrar los archivos

³ Usando el complemento del navegador DownThemAll! o la extensión de Firefox "Download All Images".

Figura 3: El proceso metodológico para crear redes basadas en visión artificial para estudiar imágenes en línea. Fuente: Adaptado de Omena (en prensa), diseño de Beatrice Gobbo.

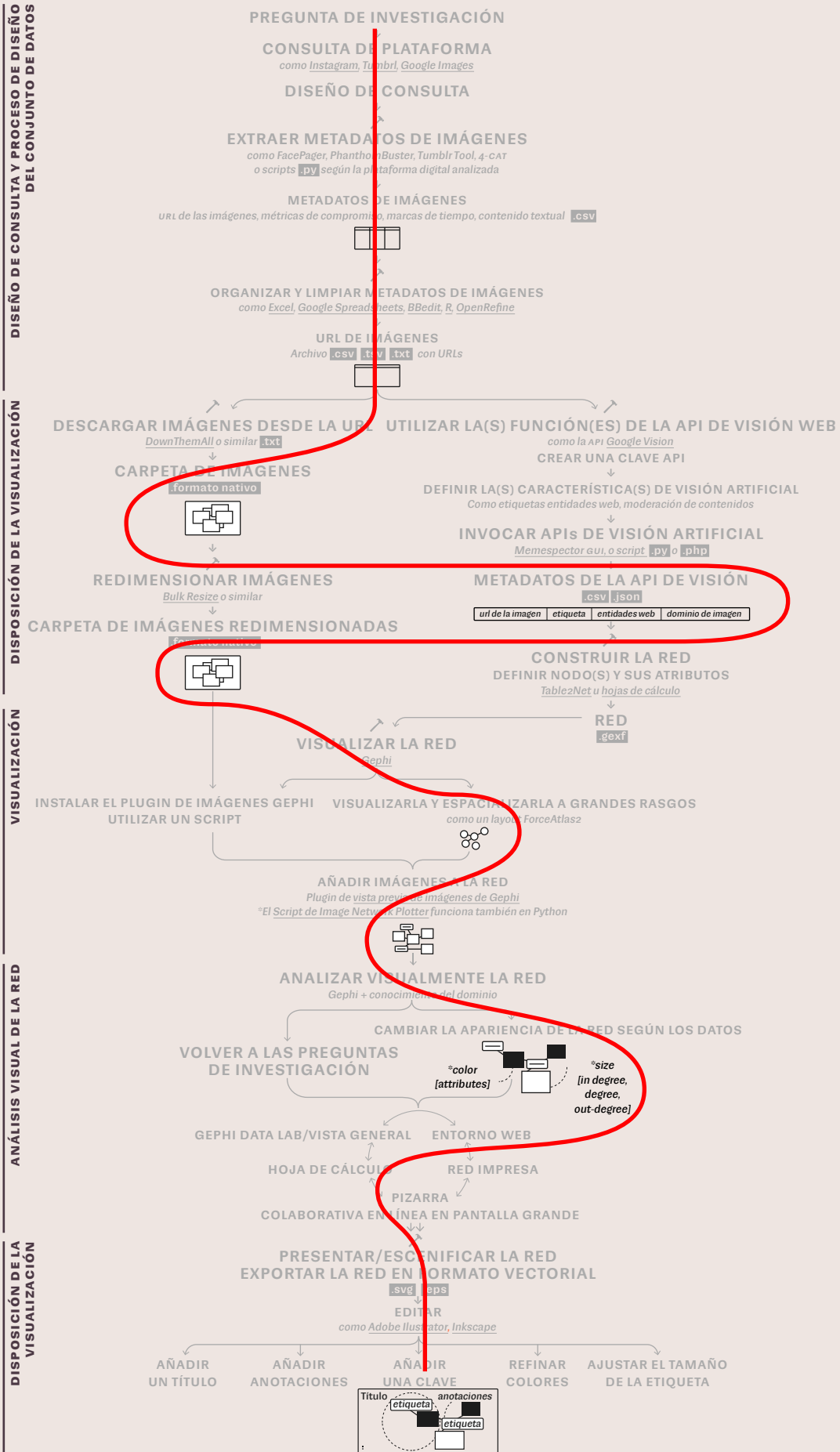


Figura 4: El arte de consultar plataformas para crear un conjunto de datos de imágenes.

DISEÑO DE CONSULTA Y PROCESO DE DISEÑO DEL CONJUNTO DE DATOS



- *intervención del investigador*
- *medio computacional*
- *output de software*
- *práctica técnica*

Figura 5: Construcción de la red con los archivos de salida del software de extracción y la API de visión.

CONSTRUCCIÓN DE LA RED



Figura 6: Visualización de la red.

VISUALIZACIÓN

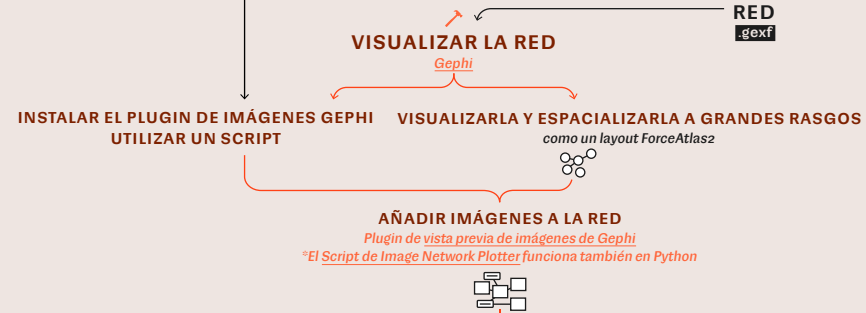


Figura 7: Análisis visual de la red y enfoque de la práctica de navegación.

ANÁLISIS VISUAL DE LA RED

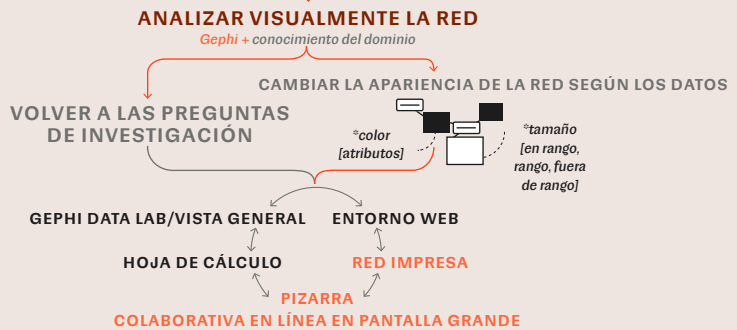


Figura 8: Presentación de la red.

DISPOSICIÓN DE LA VISUALIZACIÓN



Figuras 4-8: Adaptado de Omena (en prensa), diseño de Beatrice Gobbo.

4 La herramienta de extracción y consulta de imágenes sin conexión puede ayudar en esta tarea.

5 Por ejemplo, usando <https://bulkresizephotos.com/pt>

6 En el siguiente enlace se encuentra un tutorial para obtener un archivo de clave de autenticación de Google Vision: <https://github.com/jason-chao/memespector-gui/blob/master/doc/GetKeyFromGoogleCloud.md>

de imagen, por ejemplo, optar por `*nombre*.ext*` (id de la imagen y .JPG), facilita tanto la creación de la red como el análisis. Los investigadores pueden optar por explorar colecciones específicas de imágenes dentro de un conjunto de datos, seleccionando las imágenes según los perfiles de cuenta asociados, los dominios de los enlaces, las métricas de compromiso, la fecha de publicación (día, mes o año) o los *outputs* de la visión artificial (por ej., etiquetas o *entidades web*).⁴ Además, cuando se navega por la visualización de la red, el nombre del archivo de la imagen (por ej., `oa1781ec1998eedf38e9742acd185beb870e089e.jpg`) también es útil para filtrar y analizar las redes egocéntricas.

En el proceso de diseño del conjunto de datos, las imágenes deben organizarse en un archivo que contenga las URLs de las imágenes seleccionadas o en una carpeta de imágenes almacenada localmente. Esta última opción es la más adecuada para imágenes de redes sociales debido a la corta vida de sus URLs. El tamaño de las imágenes debe ser modificado después de la descarga.⁵ Para obtener acceso a una API de visión artificial, los investigadores deben registrarse y obtener claves de autenticación⁶ (Figura 5). A continuación, la carpeta de imágenes alojada en la computadora, o un archivo de texto que contenga la ubicación de las imágenes (aconsejable para grandes conjuntos de datos), se introduce en Memespector-GUI (Chao, 2021) para comunicarse con la API de visión y solicitarle información. Como alternativa, para invocar a las APIs de visión se pueden utilizar también los archivos de secuencia de comandos de Python o PHP. El investigador debe estar atento al archivo de configuración de la secuencia de comandos en uso (archivo de configuración) y debe conocer la manera en que su funcionamiento encaja con las características de la API de visión para poder aprovechar ambos medios técnicos.

Con los *outputs* del *software* de extracción y la API de visión, llegamos al punto en que se construye y visualiza la red (Figura 6). Aquí, las imágenes adquieren una nueva capa de significado proporcionada por la API de visión, por ejemplo, *etiquetas* y *entidades web*. Ahora podemos *ver* y *comprobar qué ofrecen los metadatos de la API de visión*. Hay que tomar decisiones cruciales al crear una red: los investigadores pueden asignar o crear atributos de nodo, pero para ello deben explorar el conjunto de datos durante un tiempo. Hecho esto, herramientas como Table2Net (Jacomy, 2013) ayudan a extraer una red de un archivo de valores separados por comas, convirtiéndolo en un formato XML de intercambio de grafos (.GEXF). Es aquí donde debemos informar cómo se establecen las conexiones entre las imágenes y las características de la API de visión, incluyendo los atributos de los nodos (numéricos o textuales).

Se requiere un archivo en formato para grafos (GEXF o GDF) para visualizar la red (Figura 6), la que adquiere forma por medio del trabajo de un algoritmo de diseño basado en fuerzas, y vida a través de su interpretación y los resultados por venir. La visualización está mediada por Gephi (Bastian et al., 2009) —o

un *software* similar— y ForceAtlas2 (Jacomy et al., 2014), que presenta una especificidad en su dibujo dirigido por fuerzas, «colocando cada nodo en función de los otros. Este proceso depende únicamente de las conexiones entre los nodos» (Jacomy et al., 2014, p. 2). Los atributos eventuales de los nodos (lo que cambia su tamaño o color) carecen de sentido en la espacialización de la red, aunque reflejan decisiones analíticas. Lo que espacializa los nodos son las conexiones ponderadas (ver Venturini et al., 2015), referidas a la naturaleza relacional de los datos de la plataforma.

En las redes de visión artificial, los nodos se espacializan en función de la manera en que la característica de visión artificial se asocia (o no) a una imagen. Aquí, los investigadores eligen los eventuales atributos de los nodos según el carácter y las características de la colección de imágenes y de acuerdo con las preguntas de investigación. Para representar las imágenes, los investigadores pueden instalar un complemento de Gephi (por ej., ImagePreview)⁷, o bien utilizar una secuencia de comandos de archivo (por ej., image-network-plotter).⁸ Para facilitar la lectura de la red, en Gephi se pueden eliminar los nodos con descripciones de imágenes ineficaces (por ej., “producto”). Al adquirir otra capa de significado mediante Gephi y ForceAtlas2, nuevos arreglos o disposiciones se imprimen en las imágenes.

El proceso analítico está guiado por el análisis visual de redes (ver Jacomy, 2021; Venturini et al., 2019). Esta técnica lleva nuestra atención a la posición, el tamaño y el color de los nodos como aspectos interpretativos clave, destacando la naturaleza relacional de los datos digitales en tándem con la tecnicidad del medio. En esta fase, la intervención del investigador (o los investigadores) es crucial. Interpretar implica navegar de ida y vuelta entre la *visión general/laboratorio de datos* de Gephi y el entorno *web*, entre la hoja de cálculo y la pantalla grande, mientras se realizan anotaciones en las versiones impresas de la red o en las pizarras en línea. Un procedimiento de navegación en el análisis no es recomendado, sino obligatorio.

Al analizar la red, los investigadores deben recordar de dónde provienen las imágenes, qué es lo que informa la posición de los nodos y sus atributos, y las especificidades de las características de la API de visión. En este caso, el investigador sabe que la metáfora visual final no representa por completo el conjunto de datos de imágenes (extraído de una plataforma digital), ni expone por completo la API de visión en todo su potencial. La red puede reflejar un *segundo orden de grammatización* co-creado con lo que representan las imágenes (y los metadatos asociados), la API de visión y los *outputs* del *software* de investigación, pero mediado por la intervención del investigador (Omena, en prensa).

En el último paso del protocolo (Figura 8), es necesario presentar y organizar adecuadamente la visualización añadiendo el título y la clave, así como convirtiendo a formato digital las anotaciones previamente escritas a mano, teniendo en cuenta cómo se experimentará la visualización (impresa o digitalmente) y las preguntas de investigación previamente formuladas.

7 Un complemento para Gephi del Yale Computer Graphics Group, disponible en <https://gephi.org/plugins/#/plugin/image-preview>

8 Un archivo de secuencia de comandos creado por André Mintz, disponible en <https://github.com/amintz/image-network-plotter>

LEYENDO REDES DE IMÁGENES CONSTRUIDAS SOBRE LA API GOOGLE VISION

Red de *imágenes-etiqueta*:

descripciones de imágenes a partir de su contenido material

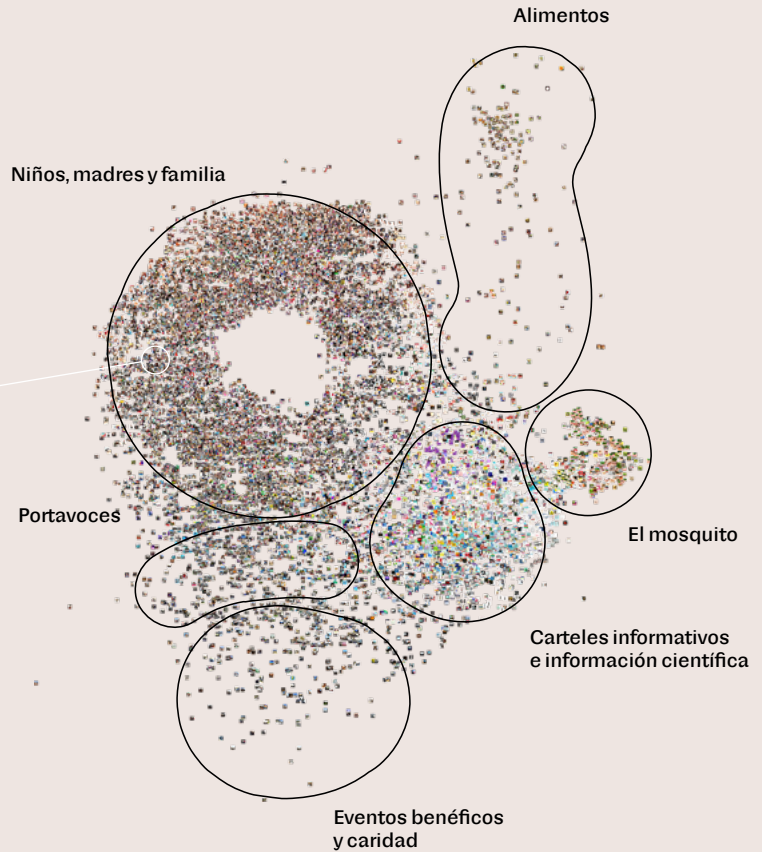
Este estudio de caso aborda la propagación del virus del Zika, que la Organización Mundial de la Salud declaró como emergencia de salud pública de interés internacional (ver Barreto et al., 2016). Este escenario epidemiológico generó advertencias a las mujeres embarazadas en Brasil y varios otros países de América Latina y el Caribe, planteando preguntas acerca de la manera en que estaban afrontando la situación las familias afectadas por la emergente infección del virus del Zika. En este sentido, nos preguntamos cómo ha reaccionado *visualmente* el público general frente al tema de la epidemia del Zika. ¿Qué pueden decir las visualidades asociadas a #microcefalia sobre las familias brasileñas afectadas por el virus?

En la figura 9, la red de *imágenes-etiqueta* reúne 10.797 imágenes publicadas por usuarios de Instagram entre junio de 2012 y octubre de 2017, quienes utilizaron el *hashtag* #microcefalia. Los datos fueron recopilados a lo largo de siete meses (de julio de 2017 a enero de 2018), utilizando Visual Tagnet Explorer para llamar a la API de la plataforma Instagram, ya obsoleta, con la consulta “microcefalia”. Los análisis fueron realizados en enero de 2018, cuando la plataforma registró un número total de 16.049 publicaciones existentes que contenían #microcefalia. La red resultante fue después utilizada para una exploración cualitativa (o reetiquetado) de los clústeres de imágenes, lo que permitió interpretar cómo se inscriben en Instagram los problemas de salud.

La fuerza del enfoque *imagen-etiqueta* reside en la capacidad de explorar, de diversas maneras, las distintas formas situadas de expresión visual que se encuentran comprometidas en la negociación pública del tema en cuestión. Como podemos ver en el caso de la red de *imágenes-etiqueta* #microcefalia (Figura 9), un grupo de imágenes se colocan juntas cuando comparten las mismas *etiquetas* u otras relacionadas, formando clústeres que abordan diferentes preocupaciones vinculadas a la epidemia del virus del Zika en Brasil (Figura 9). En Instagram, las imágenes de #microcefalia se asocian a la vida cotidiana de los niños con este trastorno neurológico. Vemos a bebés y niños pequeños con sus familias participando en actividades sociales o en sesiones de fisioterapia, además de explicaciones que ilustran las discapacidades congénitas de la microcefalia, incluyendo la imagen del mosquito que transmite el virus. El contenido de imágenes con *hashtags* específicos también muestra eventos, a portavoces y los alimentos relacionados con la microcefalia.

También es posible llevar a cabo exploraciones adicionales de los metadatos que conectan las imágenes; por ejemplo, las imágenes pueden ser dimen-

Figura 9: Red de imagen-etiqueta de microcefalia en Instagram. 10.797 imágenes con el hashtag (#microcefalia) publicadas entre junio de 2012 y octubre de 2017. Fuente: Las autoras.



sionadas según las reacciones de la audiencia (*likes*, comentarios) o el año/fecha de publicación, guiando el proceso interpretativo de manera diferente. Al leer este tipo de redes, es crucial tener en cuenta el grado de especificidad que pueda llegar a tener la API de visión en el etiquetado (cuestionando hasta qué punto esto puede ser útil) y los asuntos éticos que pueden surgir de las colecciones específicas de imágenes.

**Red de imagen-entidades web:
descripciones de imágenes a partir de contenidos web**

La figura 10 hace zoom en 500 imágenes de Tumblr con el hashtag #femalepresentingnipples (#mujermostrandopezones) clasificadas por *entidades web* de Google Vision. En esta red, cada entidad *web* visible está conectada con al menos cinco imágenes. El tamaño de los nodos de imagen corresponde al número total de reblogs, comentarios y *likes* o “notas”, destacando los memes que se hicieron virales en Tumblr en diciembre de 2018 (poco después de que la plataforma anunciara su nueva política de prohibición del contenido pornográfico). Los datos se recuperaron

tulan imágenes visualmente similares (Leetaru, 2019; Robinson, 2017). El clúster “Internet meme”, por ejemplo, especifica escenarios meméticos relacionados con el movimiento “free the nipple” (liberen el pezón), Pikachu, Lisa Simpson, Harry Potter, The Office y Star Wars, esta última predominantemente parodiando una frase de la senadora Amidala, “Así es como muere la libertad”, con más de 3.000 anotaciones (ver las Figuras 10 y 12). Al extraer información reciente sobre las imágenes del entorno *web* con la ayuda de metadatos conectados a la clasificación de cada página, este enfoque proporciona una perspectiva contextual impulsada por la *web*, que no está necesariamente informada por el contenido propio de la imagen, sino fuera de él. *Entidades web* como “Know Your Meme” o “Imgur” sugieren una cierta correlación con el significado de los sitios como archivos meméticos donde los contenidos de los memes populares y sus leyendas *web* se adaptan en tiempo real a los acontecimientos emergentes. Por lo tanto, el énfasis en el entorno *web* de la imagen nos ayuda a entender nuevos marcos de referencia, en los que obras de arte como “Retrato de una joven mujer como Santa Águeda” o “La Libertad guiando al pueblo” pueden ser *reincrustadas* en un nuevo contexto (por ejemplo, “Tumblr”) y dotadas de nuevas asociaciones (“mujer mostrando pezones”).

Por un lado, esto significa que las *entidades web* permiten a los investigadores captar las transformaciones temporalmente situadas en los temas que distribuye la *web* y las prácticas de contextualización de las que derivan las imágenes vernáculas, los GIFs de reacción y los memes de Internet. Por otro lado, a pesar de sus *affordances*, la función de entidad *web* de la API Google Vision añade cierta incertidumbre al proceso de interpretación, especialmente en lo que respecta a la lógica de adaptación del contenido. Dado que las relaciones de las imágenes, sus pies de foto textuales y los correspondientes dominios de imagen evolucionan constantemente, también lo hacen las anotaciones y las puntuaciones de confianza que proporciona la API Google Vision. En el proceso de exploración de la red, esto exige prestar atención a las descripciones *menos* fiables, ya que es aquí donde las modificaciones contextuales de las imágenes se registran primero, cada una de las cuales apunta a la relevancia temporal del tema y a la especificidad situada de las imágenes y su variada gama de asociaciones.

Red de imagen-dominio: los sitios de circulación de imágenes en la web

Este estudio de caso sirvió como herramienta metodológica para explorar las redes de circulación de imágenes. La Figura 11 ilustra las redes y cómo podemos navegar por ellas. Las dos filas representan el mismo conjunto de datos de imágenes, lo que ayuda a situar la red en diferentes entornos: su exploración en la interfaz Gephi (arriba), seguida, y en paralelo, por su exploración en archivos PDF (abajo). Vemos 992 imágenes clasificadas por Google Image Search al buscar “climate emergency” (emergencia climática), que reflejan las 100 mejores imágenes sugeridas por el bus-

cador a lo largo de los años (desde 2008 hasta 2019). Google Image Extractor retiró las URL de las imágenes de los resultados del buscador y DownThemAll! facilitó la descarga de las imágenes disponibles. En la red, los nodos son dominios de nivel superior (que señalan dónde se encontraron las imágenes) e imágenes (sugeridas por Google Image Search), conexiones que indican si una determinada imagen se encontró en las URLs.

En la descripción general de la red (Figura 11a), los dos anillos que rodean al componente principal revelan imágenes que no coinciden totalmente en la *web* (primer anillo) o imágenes que solo se encuentran en URLs específicas (segundo anillo). Cabe preguntarse qué imágenes (y dominios) son estas, o por qué no han circulado por otros lugares. Si observamos el centro de la red (Figura 11b), encontramos los dominios (delineados en rojo) en que se encontró la mayor parte de las imágenes relacionadas con la emergencia climática. Como era de esperar, vemos plataformas populares (twimg.com, fbsbx.com, pinimg.com), fuentes de información de reputación en línea (wikimedia.org) y fuentes expertas de información (esferas de blogs, por ej., blogspot.com, wp.com). Los resultados reflejan los sistemas de clasificación de Google y las directrices que siguen los calificadores de calidad de las búsquedas (Google User Content, 2020), exponiendo también la estructura jerárquica de la *web*, que entrega contenidos dispuestos en un orden determinado (ver Berners-Lee, 1995; Jacomy, 2019): contenidos conocidos por todos, contenidos conocidos por aficionados y expertos, y contenidos olvidados (Jacomy, 2019). Aquí, las redes sociales y Wikipedia pertenecen a la capa superior. Por lo tanto, al leer las redes de circulación de imágenes, los investigadores deben entender que lo que ven incluye la estructura en capas de la conectividad en línea tal como la representan los ojos de Google.

Este estudio de caso apunta a tres estrategias analíticas para explorar los sitios de circulación de imágenes. En primer lugar, identificar las imágenes que se adhieren a las plataformas o que fluyen fuera de ellas por medio de la detección visual de un grupo de imágenes que solo aparecen en un dominio. Como vemos en la Figura 11c, los diagramas y gráficos aparecen solamente en *secure.website*, mientras unas pocas imágenes se encuentran dentro de este dominio así como en Twitter y WordPress. La segunda técnica consiste en seguir a los actores principales (o de bajo perfil) dentro de la red, por ejemplo, utilizando a los vecinos de Twitter (Figura 11d) como recurso de información para analizar tanto las imágenes como los actores asociados. La tercera apunta a un análisis cualitativo de una o varias imágenes asociadas a un grupo de dominios de enlace, revelando el tipo de visualidad que comparten actores específicos, por ejemplo, observando imágenes que aparecen en un clúster de medios locales e internacionales.

La fuerza del enfoque de *imagen-dominio* radica en el estudio de la circulación de imágenes y de actores que se interesan por visualidades específicas. Al

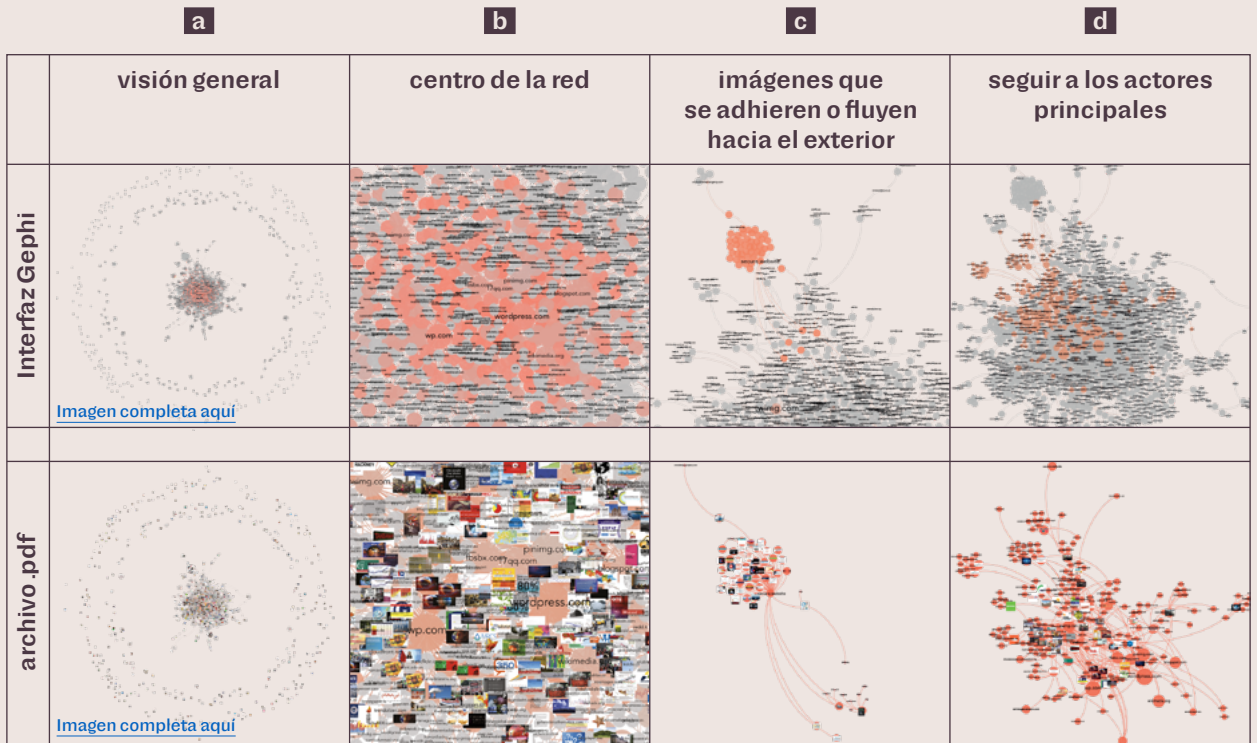







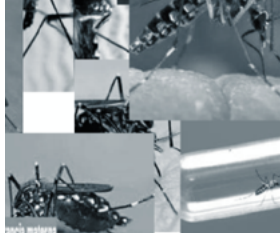






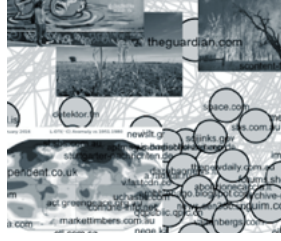


Figura 11: Red de imágenes-dominio. De izquierda a derecha, una vista ampliada que se centra en: una visión general de la red, los dominios de enlace dominantes (centro de la red), el análisis de las imágenes que se adhieren o fluyen fuera de las páginas *web*, las redes egocéntricas para seguir a los actores principales y las visualizaciones asociadas. Fuente: Las autoras.

considerar esto último, la red de *imagen-dominio* arroja luz sobre preguntas como las siguientes: ¿Cuáles son los sitios de circulación de imágenes (los sitios que albergan imágenes que coinciden total o parcialmente)? ¿Quiénes comparten los mismos contenidos visuales o similares? ¿Qué actores parecen ser más influyentes en la distribución de imágenes en la *web*? ¿Qué imágenes tienden a mantenerse dentro de las plataformas o a fluir fuera de ellas?

DISCUSIÓN Y CONCLUSIONES

En este artículo presentamos las potencialidades de las redes basadas en la API Google Vision para estudiar imágenes en línea, abarcando tres modalidades importantes que forman parte de una metodología visual crítica: la imagen en sí misma, su forma específica de “audienciación” a través de referencias obtenidas de la *web* y sus sitios de circulación (Rose, 2016). Para ello, definimos (conceptual y técnicamente) tres tipos de redes de visión artificial (Figura 12) y presentamos un protocolo de métodos digitales (Figura 3) para construir e interpretar dichas redes. A partir de los estudios de caso, identificamos los siguientes puntos de partida para comprender críticamente las redes basadas en la API Google Vision: (1) la composición de los contenidos ensamblados dentro de la imagen; (2) las referencias a una imagen obtenida a partir de contenido *web*; y (3) las páginas *web* o URLs que albergan objetos

Figura 12 (siguiente página): Lectura de redes de visión artificial según tipo, cómo/qué vemos y qué leemos. Fuente: Las autoras.

TIPO DE RED plataforma y estudio de caso	IMAGEN-ETIQUETA plataforma INSTAGRAM estudio de caso #microcefalia	IMAGEN-ENTIDADES WEB plataforma TUMBLR estudio de caso #femalepresentingnipples	IMAGEN-DOMINIO plataforma GOOGLE SEARCH IMAGES estudio de caso EMERGENCIA CLIMÁTICA
CÓMO VEMOS			
LO QUE VEMOS	<p data-bbox="262 439 502 458">Madre y su hijo con microcefalia</p>  <p data-bbox="288 738 475 757">Explicaciones científicas</p>  <p data-bbox="337 1037 427 1056">El mosquito</p>  <p data-bbox="225 1336 538 1355">Actualizaciones de noticias en los medios</p> 	<p data-bbox="642 439 834 458">Memes y GIFs de Internet</p>  <p data-bbox="637 738 840 757">Adaptaciones de Star Wars</p>  <p data-bbox="602 1037 874 1056">Fotografía de stock y liberen el pezón</p>  <p data-bbox="602 1336 874 1355">Referencias artísticas Santa Águeda</p> 	<p data-bbox="955 439 1228 477">Dominios que albergan la mayor recurrencia de aparición de imágenes</p>  <p data-bbox="955 738 1228 776">Imágenes compartidas por dominios específicos (medios informativos)</p>  <p data-bbox="995 1037 1188 1075">Dominios que comparten la misma imagen</p>  <p data-bbox="955 1336 1228 1374">Las imágenes que se adhieren y fluyen fuera de la plataforma</p> 
LO QUE LEEMOS	<p data-bbox="225 1633 534 1671">Descripciones literales de las imágenes basadas en las etiquetas.</p> <p data-bbox="225 1681 502 1700"><i>El contenido de la imagen es importante.</i></p> <p data-bbox="225 1709 521 1766"><i>Interpretación de los resultados de los modelos de aprendizaje automático para la clasificación de imágenes.</i></p>	<p data-bbox="579 1633 870 1690">Descripciones contextuales de imágenes basadas en entidades web existentes.</p> <p data-bbox="579 1700 884 1738"><i>El contenido de la imagen y el contexto fuera de ella son importantes.</i></p> <p data-bbox="579 1747 852 1785"><i>Una interpretación desde el contenido/perspectiva de la web.</i></p>	<p data-bbox="938 1633 1228 1671">Una visión dinámica de las imágenes y sus lugares de circulación.</p> <p data-bbox="938 1681 1220 1738"><i>Las imágenes totalmente emparejadas y encontradas en diferentes páginas web o URLs son importantes.</i></p> <p data-bbox="938 1747 1206 1785"><i>Una interpretación desde el contenido/perspectiva de la web.</i></p>

visuales totalmente coincidentes. Al permitir cambiar entre diferentes perspectivas del mismo conjunto de datos de imágenes, cada punto de partida ayuda a abordar la multiplicidad de imágenes digitales *nativas*. Las redes ofrecen metodologías visuales con nuevos enfoques para el análisis de imágenes.

Las redes de descripción de imágenes (organizadas a través de las funciones de *etiquetas* y *entidades web* de la API Google Vision) permiten a los investigadores, respectivamente, dar sentido a las descripciones literales y contextuales de una colección de imágenes. Las primeras informan qué hay *en* la imagen; las segundas, en tanto, proporcionan imágenes con anotaciones basadas en la *web*. En ambos tipos de redes, el proceso de interpretación es guiado por clústeres de imágenes que comparten el mismo contenido u otro similar. En las redes de circulación de imágenes, los investigadores pueden explorar clústeres de imágenes que se relacionan con una plataforma en particular y que aparecen en múltiples sitios (incluidas las plataformas principales y los sitios *web* de los medios de comunicación) o imágenes individuales (valores atípicos) que solo aparecen en sitios *web* específicos (Figura 12). Esto ayuda a detectar qué contenidos visuales se adhieren a una plataforma determinada y cuáles fluyen fuera de ella, abriendo nuevas posibilidades de análisis cualitativo de las lenguas vernáculas visuales (ver Gibbs et al., 2015; Pearce et al., 2020).

Argumentamos que comprometerse con las prácticas técnicas inherentes al enfoque de los métodos digitales por medio del diagrama del protocolo de investigación constituye un elemento esencial para construir e interpretar redes de visión artificial. La interpretación de las redes se basa en el arte de consultar las plataformas (lo que da lugar a una buena colección de imágenes), desde lo que representa una imagen hasta las interacciones que implica, dónde se aloja la imagen (por ej., redes sociales, motores de búsqueda) y cómo la hacen visible los mecanismos de la plataforma (por ej., los sistemas de clasificación). Esto implica tres desafíos a la hora de utilizar las redes de visión artificial con fines de investigación. En primer lugar, dado que el trabajo con estas redes está caracterizado por gruesas capas de mediación técnica, se requiere cierta proximidad con el *software* en su lenguaje, contenido y rendimiento (ver Omena, en prensa; Rieder, 2020). En segundo lugar, los investigadores deben comprender las prácticas técnicas y el papel de estas prácticas para aplicar con éxito la metodología (ver Figura 3). En tercer lugar, comprender cómo se reorganizan y modifican las imágenes en el proceso de análisis a través de diferentes tipos de diseño de redes. Las diferentes características de las APIs de visión artificial sugieren diferentes direcciones en el proceso de exploración y análisis de redes visuales.

Este artículo abre nuevas oportunidades para futuras investigaciones con metodologías visuales mediadas por visión artificial, que requieren sensibilidad hacia lo que la máquina percibe como relevante y la forma en que esta rele-

vancia se co-constituye tanto por medio de la práctica de los métodos digitales como a través de las prácticas de uso de la imagen en las plataformas de redes sociales y en la *web*. □

* * *

Agradecimientos

Las autoras agradecen a Giacomo Flaim y Serena Del Nero por su ayuda en la visualización de la primera versión de la red de emergencia climática; a Elaine Teixeira Rabello y Elsa Caetano por su participación en los análisis iniciales de la red de microcefalia. También queremos agradecer a Warren Pearce y a los revisores anónimos por sus valiosos comentarios.

Fuentes de financiamiento

Janna Joceli Omena recibió apoyo financiero a través de una beca del Center for Advanced Internet Studies (CAIS).

REFERENCIAS

- AZAR, M., COX, G., & IMPETT, L. (2021). Introduction: Ways of Machine Seeing. *AI & Society*. <https://doi.org/10.1007/s00146-020-01124-6>
- BARRETO, M. L., BARRAL-NETTO, M., STABELI, R., ALMEIDA-FILHO, N., VASCONCELOS, P. F. C., TEIXEIRA, M., BUSS, P., & GADELHA, P. E. (2016). Zika Virus and Microcephaly in Brazil: A Scientific Agenda. *The Lancet*, 387(10022), 919–921. [https://doi.org/10.1016/S0140-6736\(16\)00545-6](https://doi.org/10.1016/S0140-6736(16)00545-6)
- BASTIAN, M., HEYMANN, S., & JACOMY, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1), 361–362. <https://doi.org/10.1136/qshc.2004.010033>
- BERNERS-LEE, T. (1995). *Hypertext and Our Collective Destiny* [Video]. 1995 Vannevar Bush Symposium - Videos and Links - Doug Engelbart Institute. Recuperado el 20 de septiembre, 2020, de <https://www.doungelbart.org/content/view/258/000/>
- CHAO, J. (2021). Memespector GUI: Graphical User Interface Client for Computer Vision APIs [Software] (Versión 0.1) [C#] (Trabajo original publicado en 2021). <https://github.com/jason-chao/memespector-gui>
- COLOMBO, G. (2018). *The Design of Composite Images. Displaying Digital Visual Content for Social Research* [Disertación Doctoral, Politecnico di Milano]. <http://hdl.handle.net/10589/141266>
- D'ANDRÉA, C., & MINTZ, A. (2019). Studying the Live Cross-Platform Circulation of Images with Computer Vision API: An Experiment Based on a Sports Media Event. *International Journal of Communication*, 13, 1825–1845.
- GARIMELLA, K., & ECKLES, D. (2020). Images and Misinformation in Political Groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review*, 1(5). <https://doi.org/10.37016/mr-2020-030>
- GEBBERS, M. A., & VAN DE WIELE, C. T. (2020). Machine Vision and Social Media Images: Why Hashtags Matter. *Social Media + Society*, 6(2), 2056305120928485. <https://doi.org/10.1177/2056305120928485>
- GEBRU, T., MORGENSTERN, J., VECCHIONE, B., VAUGHAN, J. W., WALLACH, H., DAUMÉ III, H., & CRAWFORD, K. (2020). Datasheets for Datasets. *ArXiv*, 1803.09010 [cs].

- GERLITZ, C., & RIEDER, B. (2018). Tweets Are Not Created Equal: Investigating Twitter's Client Ecosystem. *International Journal of Communication*, 12. <https://hdl.handle.net/11245.1/4da1d406-1213-4103-8237-eef5ae786948>
- GIBBS, M., MEESE, J., ARNOLD, M., NANSEN, B., & CARTER, M. (2015). #Funeral and Instagram: Death, Social Media, and Platform Vernacular. *Information, Communication & Society*, 18(3), 255–268. <https://doi.org/10.1080/1369118X.2014.987152>
- GOOGLE CLOUD. (2017, abril 13). *Pricing | Google Cloud Vision API Documentation*. Wayback Machine. <https://web.archive.org/web/20170413081619/https://cloud.google.com/vision/docs/pricing>
- GOOGLE USER CONTENT. (2020). *General Guidelines*. <https://static.googleusercontent.com/media/guidelines.raterhub.com/pt-BR//searchqualityevaluatorguidelines.pdf>
- HOCHMAN, N. (2014). The Social Media Image. *Big Data & Society*, 1(2), 2053951714546645. <https://doi.org/10.1177/2053951714546645>
- HOELZL, I., & MARIE, R. (2015). *Softimage: Towards a New Theory of the Digital Image*. Intellect.
- JACOMY, M. (2013). Table 2 Net [Software]. <https://medialab.github.io/table2net/>
- JACOMY, M. (2019). The Web as Layers [Video]. *HYPHE - Session 4*. <https://panopto.aau.dk/Panopto/Pages/Viewer.aspx?id=48cfe5ff-5503-431b-887f-ab53007ef5c4>
- JACOMY, M. (2021). *Situating Visual Network Analysis* [Disertación Doctoral, Universidad de Aalborg]. <https://reticular.hypotheses.org/1879>
- JACOMY, M., VENTURINI, T., HEYMANN, S., & BASTIAN, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*, 9(6), e98679. <https://doi.org/10.1371/journal.pone.0098679>
- LEETARU, K. (2019, junio 1). Using Google Vision AI's Reverse Image Search To Richly Catalog Television News. *Forbes*. <https://www.forbes.com/sites/kalevleetaru/2019/06/01/using-google-vision-ais-reverse-image-search-to-richly-catalog-television-news/>
- MACKENZIE, A., & MUNSTER, A. (2019). Platform Seeing: Image Ensembles and their Invisualities. *Theory, Culture & Society*, 36(5), 3–22. <https://doi.org/10.1177/0263276419847508>
- MANOVICH, L. (2020). *Cultural Analytics*. MIT Press.
- MAURI, M., BRIONES, M. A., GOBBO, B., & COLOMBO, G. (2020). Research Protocol Diagrams as Didactic Tools to Act Critically in Dataset Design Processes. *INTED2020 Proceedings*, 9034–9043. <https://library.iated.org/view/MAURI2020RES>
- MINTZ, A., SILVA, T., GOBBO, B., PILIPETS, E., AZHAR, H., TAKAMITSU, H., OMENA, J. J., & OLIVEIRA, T. (2019). *Interrogating Vision APIs*. #SMARTDataSprint. <https://smart.inovamedialab.org/past-editions/smart-2019/project-reports/interrogating-vision-apis/>
- NIEDERER, S., & COLOMBO, G. (2019). Visual Methodologies for Networked Images: Designing Visualizations for Collaborative Research, Cross-platform Analysis, and Public Participation. *Diseña*, (14), 40–67. <https://doi.org/10.7764/disena.14.40-67>
- OMENA, J. J. (en prensa). *Digital Methods and Technicity-of-the-Mediums. From Regimes of Functioning to Digital Research* [Disertación Doctoral]. Universidade NOVA de Lisboa.
- OMENA, J. J., & AMARAL, I. (2019). Sistema de leitura de redes digitais multiplataforma. En J. J. Omena (Ed.), *Métodos Digitais: Teoria-prática-crítica* (pp. 121–140). ICNOVA.

- OMENA, J. J., & GRANADO, A. (2020). Call into the Platform! *Revista ICONO14 Revista Científica de Comunicación y Tecnologías Emergentes*, 18(1), 89–122. <https://doi.org/10.7195/ri14.v18i1.1436>
- OMENA, J. J., RABELLO, E. T., & MINTZ, A. G. (2020). Digital Methods for Hashtag Engagement Research. *Social Media + Society*, 6(3), 2056305120940697. <https://doi.org/10.1177/2056305120940697>
- PAGLEN, T. (2014, marzo 13). *Seeing Machines*. Is Photography Over? Fotomuseum Winterthur Series. <https://www.fotomuseum.ch/en/2014/03/13/seeing-machines/>
- PAIKKA, J., & DVOŘÁK, T. (2021). Introduction: On the Scale, Quantity and Measure of Images. En T. Dvořák & J. Paikka (Eds.), *Photography Off the Scale: Technologies and Theories of the Mass Image* (pp. 1–24). Edinburgh University Press.
- PEARCE, W., ÖZKULA, S. M., GREENE, A. K., TEELING, L., BANSARD, J. S., OMENA, J. J., & RABELLO, E. T. (2020). Visual Cross-platform Analysis: Digital Methods to Research Social Media Images. *Information, Communication & Society*, 23(2), 161–180. <https://doi.org/10.1080/1369118X.2018.1486871>
- PILIPETS, E., FLORES, A. M. M., FLAIM, G., SKAZEDONIG, M., SEPÚLVEDA, R., & DEL NERO, S. (2020). From “Tumblr Purge” to “Female Nipples”: Telling a Story of Platform Censorship Critique through Memes and Digital Methods. #SMARTDataSprint. <https://smart.inovamedialab.org/2020-digital-methods/project-reports/tumblr-purge-female-nipples/>
- PILIPETS, E., & PAASONEN, S. (2020). Nipples, Memes, and Algorithmic Failure: NSFW Critique of Tumblr Censorship. *New Media & Society*, 1461444820979280. <https://doi.org/10.1177/1461444820979280>
- RETTBERG, J. W. (2020). Situated Data Analysis: A New Method for Analysing Encoded Power Relationships in Social Media Platforms and Apps. *Humanities and Social Sciences Communications*, 7, Art. 5. <https://doi.org/10.1057/s41599-020-0495-3>
- RICCI, D., COLOMBO, G., MEUNIER, A., & BRILLI, A. (2017). Designing Digital Methods to Monitor and Inform Urban Policy: The Case of Paris and its Urban Nature Initiative. *International Conference on Public Policy (ICPP3)*. <https://hal.archives-ouvertes.fr/hal-01903809>
- RIEDER, B. (2015). YouTube Data Tools [Software] (Versión 1.11). <https://tools.digitalmethods.net/netvizz/youtube/>
- RIEDER, B. (2020). *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam University Press.
- RIEDER, B., & RÖHLE, T. (2018). Digital Methods: From Challenges to Bildung. En M. T. Schäfer & K. van Es (Eds.), *The Datafied Society: Studying Culture through Data* (pp. 109–124). Amsterdam University Press. <https://doi.org/10.1515/9789048531011-010>
- ROBINSON, S. (2017, junio 1). *Exploring the Cloud Vision API*. Medium. <https://medium.com/@srobtweets/exploring-the-cloud-vision-api-1af9bcf080b8>
- ROGERS, R. (2013). *Digital Methods*. MIT Press.
- ROGERS, R. (2019). *Doing Digital Methods*. SAGE.
- ROSE, G. (2016). *Visual Methodologies* (4ta ed.). Open University.
- RUBINSTEIN, D., & SLUIS, K. (2013). Concerning the Undecidability of the Digital Image. *Photographies*, 6(1), 151–158. <https://doi.org/10.1080/17540763.2013.788848>
- SCHWEMMER, C., KNIGHT, C., BELLO-PARDO, E. D., OKLOBDZIJA, S., SCHOONVELDE, M., & LOCKHART, J. W. (2020). Diagnosing Gender Bias in Image Recognition Systems. *Socius*, 6, 2378023120967171. <https://doi.org/10.1177/2378023120967171>

- SILVA, T., BARCIELA, P., & MEIRELLES, P. (2018). Mapeando Imagens de Desinformação e Fake News Político-Eleitorais com Inteligência Artificial. 3o CONEC: Congresso Nacional de Estudos Comunicacionais Da PUC Minas Poços de Caldas - Convergência e Monitoramento, 413–427. https://www.researchgate.net/publication/329525177_Mapeando_Imagens_de_Desinformacao_e_Fake_News_Politico-Eleitorais_com_Inteligencia_Artificial
- SILVA, T., MINTZ, A., OMENA, J. J., GOBBO, B., OLIVEIRA, T., TAKAMITSU, H. T., PILIPETS, E., & AZHAR, H. (2020). APIs de Visão Computacional: Investigando mediações algorítmicas a partir de estudo de bancos de imagens. *Logos*, 27(1), 25–54. <https://doi.org/10.12957/logos.2020.51523>
- STEYERL, H. (2009). In Defense of the Poor Image. *E-Flux Journal*, 10. <https://www.e-flux.com/journal/10/61362/in-defense-of-the-poor-image/>
- SULLIVAN, D. (2020, mayo 20). *A Reintroduction to our Knowledge Graph and Knowledge Panels*. Blog.Google. <https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels/>
- SZELISKI, R. (2021). *Computer Vision: Algorithms and Applications* (2da ed.). Springer.
- TAIBI, D., ROGERS, R., MARENZI, I., NEJDL, W., AHMAD, Q. A. I., & FULANTELLI, G. (2016). Search as Research Practices on the Web: The Sar-web Platform for Cross-language Engine Results Analysis. *Proceedings of the 8th ACM Conference on Web Science*, 367–369. <https://doi.org/10.1145/2908131.2908201>
- TUMBLR. (2018, diciembre). Updates to Tumblr's Community Guidelines [Tumblr]. *Tumblr Support*. <https://support.tumblr.com/post/180758979032/updates-to-tumblrs-community-guidelines>
- VENTURINI, T., JACOMY, M., & JENSEN, P. (2019). *What Do We See When We Look at Networks. An Introduction to Visual Network Analysis and Force-Directed Layouts* (SSRN Scholarly Paper ID 3378438). Social Science Research Network. <https://doi.org/10.2139/ssrn.3378438>
- VENTURINI, T., JACOMY, M., & PEREIRA, D. (2015). Visual Network Analysis: The Example of the Rio+20 Online Debate. <https://hal.archives-ouvertes.fr/hal-02305124>
- XI, N., MA, D., LIOU, M., STEINERT-THRELKELD, Z. C., ANASTASOPOULOS, J., & JOO, J. (2020). Understanding the Political Ideology of Legislators from Social Media Images. *Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM 2020*, 726–737. <http://arxiv.org/abs/1907.09594>