

Analítica cultural

Lev Manovich

Massachusetts Institute of Technology (MIT),
Cambridge, Massachusetts

Traducción a cargo de GRUFFAT, Carolina
para la Cátedra de Datos (FSOC - UBA)

Resumen

Este libro presenta conceptos y métodos para el análisis computacional de datos culturales, la exploración de medios y artefactos digitalizados, y la comprensión de la asombrosa escala del crecimiento cultural en la era digital.

Palabras clave: cultura, investigación, métodos estadísticos, procesamiento de datos, medios de comunicación de masas, visualización de información

Publicación original

Nombre: Manovich, Lev

Título: Cultural analytics

Editorial: Cambridge, Massachusetts: The MIT Press

Año de publicación: 2020

ISBN 9780262037105

Clasificación LC HM623 .M365 2020

Disponible en: <https://lcn.loc.gov/2020003045>

Introducción: Cómo ver mil millones de imágenes

Sin embargo, es probable que el impacto de la computadora en las ciencias humanas sea proporcionalmente más revolucionario a largo plazo [que en las ciencias físicas y naturales]... Parte de esto tiene que ver simplemente con la voluntad de aprovechar la oportunidad, o con la predisposición, basada en el uso ya extensivo de procesos, especialmente estadísticos, facilitados por la computadora. Más aún, quizás tenga que ver con lo que una computadora, en el mismo sentido de un telescopio o un microscopio, nos permite ver. En pocas palabras, el procesamiento computacional, realizado correctamente, puede permitirnos ver relaciones y patrones en masas de datos que antes eran demasiado grandes para comprender; y las consecuencias inmediatas de una idea aplicada a los datos, si no de modo único, al menos de manera mucho más inexorable y rápida.

—Dell H. Hymes, "Introducción", en *El uso de computadoras en la Antropología (The Use of Computers in Anthropology)*, 1965¹

Este libro está situado en la intersección de la ciencia de datos, los estudios de medios y los estudios de la cultura digital. Presenta una selección de conceptos y métodos para el análisis computacional de datos culturales. Estos métodos se pueden utilizar para explorar tanto artefactos históricos digitalizados como medios digitales actuales. Si bien podemos aplicarlos a uno o varios artefactos, se vuelven especialmente importantes si queremos explorar millones de artefactos.

De hecho, la escala asombrosa de la cultura digital es lo que me motivó a comenzar a explorar estos métodos en 2005 y finalmente escribir este libro. ¿Cómo podemos entender la fotografía popular contemporánea que crece en miles de millones de imágenes por día? ¿O la música actual representada por cientos de millones de canciones compartidas por veinte millones de creadores en SoundCloud? ¿O el contenido de cuatro mil millones de tableros en Pinterest?² ¿O los patrones en los intereses intelectuales de personas en 190 países representados por 330.000 grupos de Meetup y 84.000 eventos por mes (a partir de 2019)?³ Esto también es "cultura digital" porque estos eventos físicos están habilitados por la plataforma web de Meetup. En mi opinión, la única forma posible de estudiar los patrones, tendencias y dinámicas de la cultura actual a esa escala es utilizar métodos de ciencia de datos.

No se necesita tener experiencia en ciencia de datos, programación, estadística o matemática para utilizar este libro. Mis destinatarios son investigadores académicos y estudiantes de arte, diseño, humanidades, ciencias sociales, estudios de medios, ciencia de datos e informática; profesionales que trabajan en diseño, fotografía, cine, diseño urbano, arquitectura, periodismo, museos y bibliotecas, curaduría y gestión cultural; y todos los que trabajan con las redes sociales y la web en cualquier rol (creador, bloguero, líder de proyecto, gerente, desarrollador, comercializador, etc.).

¹ Dell H. Hymes, "Introduction" en *The Use of Computers in Anthropology*, ed. Dell H. Hymes (The Hague: Mouton, 1965), 29–30. Énfasis en el original

² Comisión de Bolsa y Valores de los Estados Unidos, Declaración de registro del formulario S-1 para Pinterest, Inc., 22 de marzo de 2019,

<https://www.sec.gov/Archives/edgar/data/1506293/000119312519083544/d674330ds1.htm>

³ "About", Meetup, consultado el 12 de agosto de 2019, <https://www.meetup.com/about/>.

Incluso si uno no tiene interés en analizar conjuntos de datos culturales, se encuentra con este tipo de análisis a diario. Tal vez esté mirando su analítica de Facebook, Instagram o Weibo, o Google Analytics para su blog o sitio web, o esté utilizando un tablero de monitoreo de redes sociales en el trabajo. Y aún si no presta atención a esos datos, está interactuando constantemente con los resultados del análisis computacional cuando hace algo digital. Por ejemplo, cada vez que saca una foto, los algoritmos de la cámara del teléfono eligen automáticamente la exposición y ajustan el contraste de la foto, y también identifican el tipo de escena y los objetos en la foto⁴. El análisis computacional de los medios y las interacciones del usuario es lo que posibilita la búsqueda web, las recomendaciones, el filtrado, la personalización, las interacciones con dispositivos digitales, la publicidad conductual y otras operaciones que forman parte del "vocabulario" de la cultura digital. Por ejemplo, los motores de búsqueda como Baidu, Bing, Yandex o Google dependen del análisis computacional continuo de los contenidos de miles de millones de páginas web, imágenes en línea y otros contenidos web, para brindarle resultados relevantes.

Creo que para estar alfabetizado en una sociedad así es necesario conocer las ideas y los principios fundamentales que hacen posible tales operaciones. Este libro es una introducción amable y no técnica a algunas de estas ideas. Por lo tanto, enseña cómo se pueden explorar conjuntos de datos culturales, y también cómo piensa nuestra sociedad utilizando esos datos y algoritmos.

Mirando la cultura a través de las computadoras

En este libro se pueden encontrar numerosos ejemplos de análisis computacional de la cultura, de diversos investigadores y también de mi propio laboratorio. En lo que sigue, quiero describir dos ejemplos para mostrar las posibilidades y desafíos que plantea este tipo de análisis.

El primer ejemplo es el proyecto *Elsewhere (En algún otro lado)*, en el que trabajamos desde 2018. El proyecto investiga el crecimiento y la difusión de la cultura contemporánea teniendo en cuenta numerosas ciudades más pequeñas, en lugar de considerar solo un puñado de capitales globales. Hoy en día, los medios, estudios de investigación y diversos tipos de rankings le otorgan una atención desproporcionada a una pequeña selección de estas capitales. Por lo tanto, es fácil suponer que un puñado de ciudades "principales" continúan actuando como "centros" y el resto del mundo todavía está en la periferia y recibe con retraso las nuevas ideas. Pero ¿cuál es la imagen real? ¿Cómo cambiaron la globalización y el auge de las nuevas tecnologías de la comunicación la geografía de la cultura? ¿Podemos encontrar todas las tendencias culturales actuales en miles de ciudades más pequeñas? ¿Es posible que algunas de estas ciudades sean culturalmente más innovadoras precisamente por su distancia de la capital y su menor tamaño? ¿Quedan grandes partes del mundo hoy que no sean conscientes de estas tendencias y no innoven?

⁴ Esta función se introdujo por primera vez en el iPhone 7 de Apple en 2016. Sin embargo, incluso antes algunas aplicaciones y servicios móviles, como Google Photos, ya ofrecían detección de objetos en fotos de teléfonos móviles.

¿Cómo se ha desarrollado y difundido la cultura contemporánea en el mundo desde el comienzo de la globalización? ¿El crecimiento fue parejo o desigual, acelerándose o desacelerándose en algunos períodos? ¿Son los patrones de crecimiento los mismos para diferentes campos culturales, o cada uno de ellos se desarrolla a su modo?

Ciertamente un solo proyecto no puede responder a todas estas preguntas. El objetivo de *Elsewhere* es desarrollar y probar una nueva metodología que utiliza datos públicos sobre eventos culturales y los lugares que organizan estos eventos. La cantidad de lugares y eventos en el mundo de hoy se ha vuelto tan grande que podemos tratarlos como "big data" y utilizar métodos de la ciencia de datos para su análisis. Esta perspectiva debería permitirnos crear mapas y líneas de tiempo de la cultura contemporánea mucho más detallados que los que se proporcionan en los estudios de las industrias culturales o las listas de instituciones culturales. También elaboramos análisis de textos con todo lo que los organizadores publican sobre sus eventos: descripciones de millones de exposiciones, conferencias, talleres, festivales, reuniones de grupos de interés y otros tipos de eventos. Esto nos permitirá buscar patrones en esos temas, intereses y "palabras clave" que atraviesen las geografías y se extiendan por largos períodos de tiempo.

La Figura I.1 muestra el crecimiento en el número de eventos culturales a lo largo del tiempo utilizando plataformas y redes seleccionadas a través de las cuales se promueven u organizan dichos eventos. En conjunto, nuestro *dataset*, recopilado a partir de anuncios de eventos en seis plataformas y redes de este tipo, contiene 4.380.946 de eventos en 21.072 ciudades de 200 países en seis continentes. Estas plataformas y redes son: Behance, E-Flux, Arts and Education Network, Meetup, eventos locales de TED y TimePad (en el caso de Behance, se considera como evento al registro de una nueva cuenta de usuario).

Como vemos en los gráficos, la cantidad de eventos culturales en cada plataforma y red fue creciendo con el tiempo. Para 2006, nuestro *dataset* tiene 11.642 eventos (sumando todas las fuentes juntas); para 2009, ya cuenta con 102.211 eventos; y para el 2018 cuenta con 781.697 eventos. Por supuesto, hay una única plataforma global que reúne todos los eventos culturales, y debemos tener cuidado al generalizar los resultados obtenidos de nuestras fuentes particulares. Este proyecto ejemplifica los desafíos de utilizar fenómenos digitales para aprender sobre otros fenómenos. ¿El crecimiento que vemos en las seis fuentes de datos a lo largo del tiempo representa el crecimiento real de los eventos culturales en numerosos países? ¿O es solo una señal de la difusión de la cultura digital en sí misma, que muestra que más organizaciones en más países estaban utilizando gradualmente las plataformas y redes que elegimos mirar para promover sus eventos? ¿O simplemente vemos el efecto de que "el ganador se lo lleva todo", en el que ciertas plataformas de información se vuelven dominantes y todos comienzan a usarlas? Es probable que todos estos efectos se presenten en nuestros datos. Parte del crecimiento que vemos a lo largo del tiempo se debe a la creciente popularidad de las propias plataformas; y otra parte refleja un aumento real en el número de lugares, actores y eventos culturales.

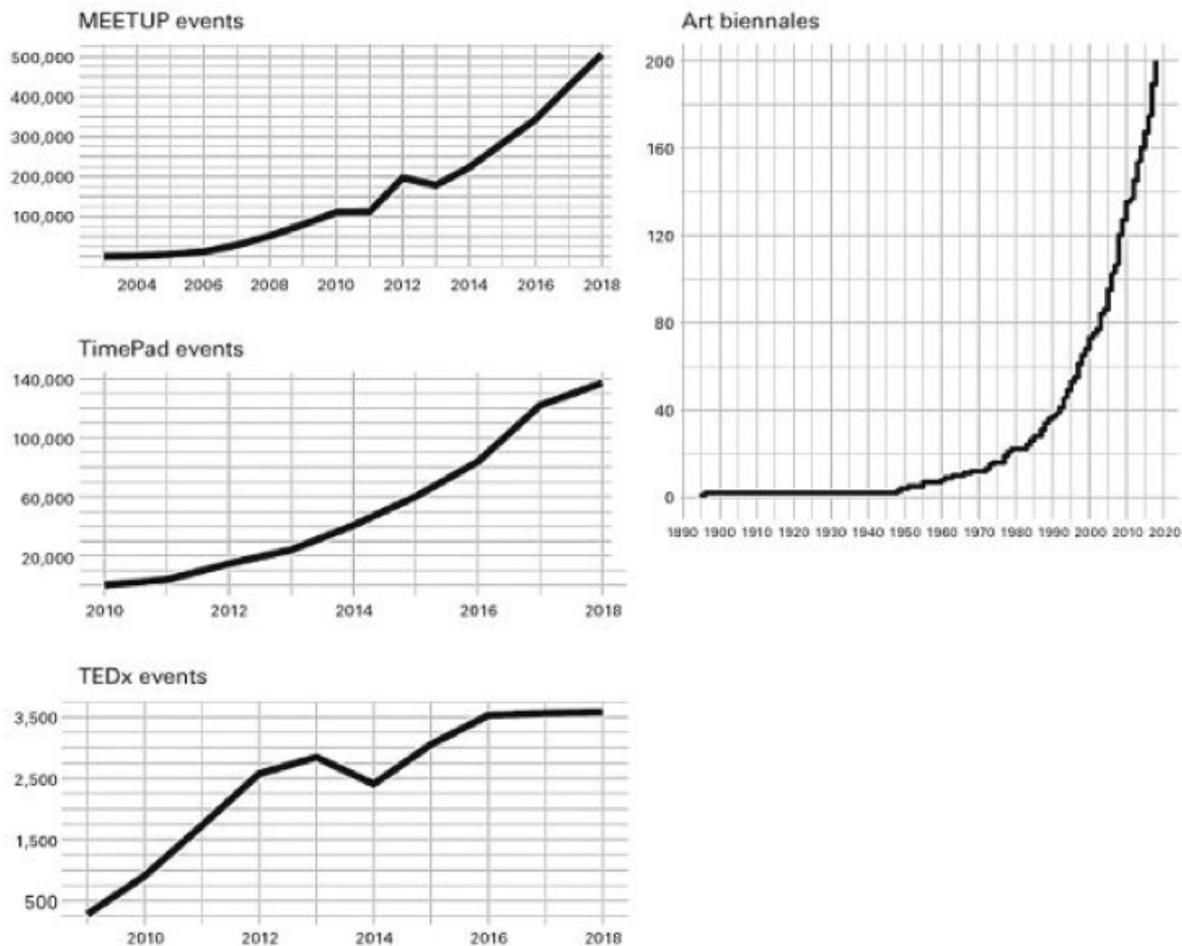


Figure I.1

Crecimiento del número de eventos culturales a lo largo del tiempo representado por fuentes de datos particulares del proyecto *Elsewhere*.

Elsewhere utiliza fechas, ubicaciones, categorías y contenido de texto de los anuncios de eventos. En el segundo ejemplo que quiero describir observamos artefactos de medios utilizando la ciencia de datos y visualizaciones. El objetivo es poder explorar varios campos culturales actuales observando muchos más productos de los que los estudios de medios analizan normalmente. El proyecto en cuestión es *Un millón de páginas de manga*, en el que Jeremy Douglass, William Huber y yo comenzamos a trabajar en 2009. La figura I.2 muestra una de las visualizaciones que creamos al explorar un *dataset* de 1.074.790 de páginas únicas provenientes de 883 series de manga. Las páginas se tomaron del sitio de manga más popular en ese momento: OneManga (<https://www.onemanga.com/>). El sitio contenía la mayoría de las páginas de estas series que los fanáticos habían escaneado y traducido a varios idiomas.

La serie más extensa disponible en OneManga se ha publicado continuamente desde 1976. Las series más populares en este sitio fueron *Naruto* (8.835 páginas para el período 1999-2009) y *One Piece* (10.562 páginas para el período 1997-2009). Además de esas series de manga tan largas, nuestro *dataset* también contiene series más breves que aparecieron en los 2000 y solo duraron de uno a tres años.

Utilizamos nuestro propio software de análisis de imágenes para medir cada página, convirtiendo ciertas características visuales seleccionadas en características numéricas. La visualización mapea las páginas en los ejes X- e Y-, según dos de estas características. El eje X- representa el desvío estándar de los valores de escala de grises de los píxeles, medidos por página. El eje Y- representa la entropía de todos los valores de escala de grises de los píxeles, también medidos por página. ¿Qué significan estas medidas en la práctica? Las páginas en la parte inferior de la visualización son las más gráficas y tienen la menor cantidad de detalles. Las páginas en la esquina superior derecha tienen muchos detalles y texturas. Las páginas con mayor contraste están a la derecha, mientras que las páginas con menos contraste están a la izquierda.

Dentro de estos cuatro extremos, encontramos todas las variaciones estilísticas posibles. Esto me sugiere que nuestro concepto básico de "estilo" puede no ser apropiado cuando consideramos grandes conjuntos de datos culturales. El concepto supone que podemos dividir un conjunto de artefactos culturales en un pequeño número de categorías discretas. En el caso de nuestro *dataset One Million Manga Pages* encontramos variaciones gráficas prácticamente infinitas. Si tratamos de dividir este espacio en categorías estilísticas discretas, cualquier intento de este tipo será arbitrario.

La visualización también muestra cuáles son las opciones gráficas más utilizadas por los artistas de manga (la parte central de la "nube" de páginas) y cuáles aparecen con mucha menos frecuencia (partes inferior e izquierda). Podemos preguntarnos por qué el manga evolucionó visualmente de maneras particulares, con algunas opciones usadas con mucha frecuencia, otras con menos frecuencia y otras casi nunca usadas. Y si queremos comprender la originalidad visual de una nueva serie de manga (representada por ciertas características visuales que podemos medir), podemos agregar sus páginas a dicha visualización y, si lo deseamos, incluso cuantificar esta originalidad. Más adelante en el libro, presento otro análisis de este *dataset*, observando las conexiones entre los estilos visuales, el género de las audiencias de manga y los géneros de manga (véanse las figuras 7.1 y 7.2 y la ilustración 9).

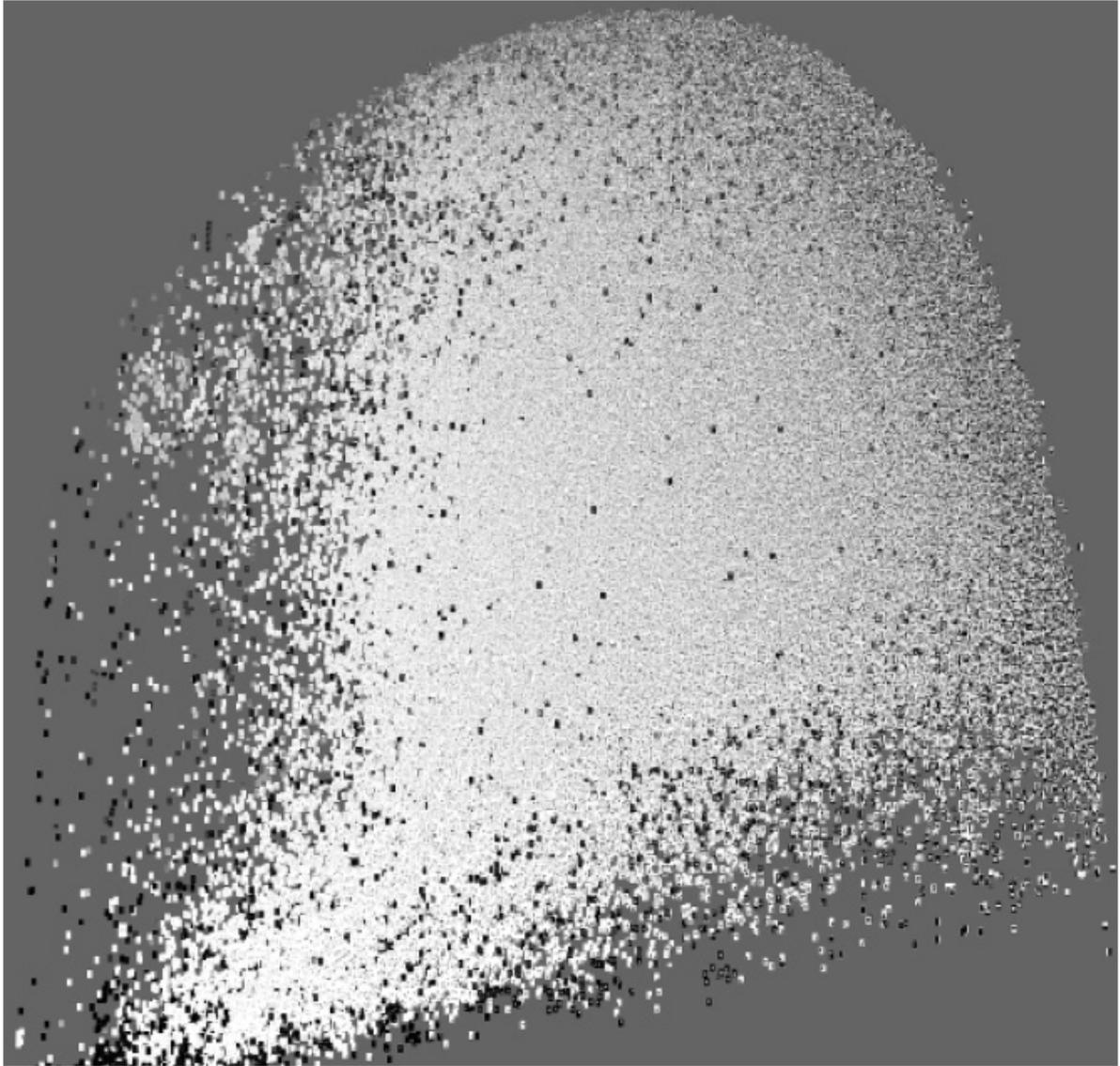


Figure I.2

Visualización de 1.074.790 páginas de manga ordenadas por dos características visuales calculadas algorítmicamente en cada página: desviación estándar de los valores de escala de grises (eje X) y entropía de los valores de escala de grises (eje Y).

Analítica cultural: cinco ideas

Empecé a pensar acerca del análisis y la visualización de patrones en la cultura digital actual a gran escala en el otoño de 2005. En ese momento, ya había ejemplos de análisis computacional de colecciones de sitios web y blogs realizados por investigadores en ciencias de la computación. Sin embargo, las ciencias de la computación son un campo enorme con docenas de subcampos. Esta investigación estaba apareciendo en revistas y conferencias de varios subcampos y no tenía su nombre propio. En las humanidades, el término *humanidades digitales*, introducido por primera vez en 2003, se estaba volviendo conocido. Pero aquí también el análisis computacional de *datasets* culturales era solo una parte de este amplio campo emergente, que también incluía la digitalización y publicación

de colecciones históricas, el uso de herramientas digitales en la enseñanza y otras actividades. El término "humanidades digitales" tenía otra limitación para mí: los académicos de humanidades digitales trabajaban casi exclusivamente con *datasets* de textos literarios históricos, en oposición a otros tipos de medios o cultura digital contemporánea, por lo que este término era a la vez demasiado amplio y demasiado limitado. Finalmente, conocía ya varios proyectos artísticos y de diseño muy impresionantes que visualizaban grandes conjuntos de datos culturales, pero este trabajo se había realizado por fuera de los departamentos académicos y tampoco tenía un solo término que lo describiera.

A mi modo de ver, el trabajo existente de análisis y la visualización de datos culturales a gran escala, que ocurría a través de varios campos, estaba creando un nuevo paradigma de investigación, aunque no tenía su nombre propio. Sentí la necesidad de un término que pudiera referirse al análisis computacional de patrones y tendencias en la *cultura digital contemporánea* (a diferencia de hacerlo solo en la cultura histórica) y que pudiera abarcar el análisis de todo tipo de medios (a diferencia de solo los textos). Cuando establecimos nuestro propio laboratorio para realizar esta investigación en la primavera de 2007⁵, se me ocurrió el término de *analítica cultural*.

Nuestro laboratorio tenía dos objetivos. El primero era práctico: utilizar métodos de las ciencias de la computación, la visualización de datos y el arte de los medios, para explorar y analizar diferentes tipos de medios contemporáneos y las interacciones de los usuarios con ellos. El segundo era teórico: nos preguntamos cómo el uso de tales métodos y grandes *datasets* de medios culturales desafía nuestras ideas modernas sobre la cultura y los métodos para estudiarla. ¿Cuáles son exactamente las nuevas posibilidades que ofrecen, y si pueden trabajar con cualquier tipo de medio? ¿Cuáles son las limitaciones de los métodos computacionales y el análisis a gran escala?

En la práctica, nuestro laboratorio se centró en analizar y visualizar patrones en *datasets* visuales, como millones de fotos compartidas en Instagram y Twitter, un millón de páginas de publicaciones de manga, decenas de largometrajes y miles de portadas de revistas. Esto llevó a que muchas personas que seguían nuestro trabajo asociaran el análisis cultural con el uso de la computación y la visualización para la exploración de grandes colecciones visuales. En los últimos años, el término *analítica cultural* (o *analítica de la cultura*) comenzó a ser utilizado de manera más amplia por muchos otros académicos, como lo ejemplifican dos simposios⁶; un programa de investigación de cuatro meses en la UCLA en 2016 que reunió a 120 investigadores líderes de universidades y laboratorios industriales⁷; una revista académica revisada por pares, *Journal of Cultural Analytics*, creada en 2016⁸; el uso del

⁵ Cultural Analytics Lab, "About", consultado el 28 de febrero de 2020, <http://lab.culturalanalytics.info/p/about.html>.

⁶ *Cultural Analytics: Computational Approaches to the Study of Culture*, simposio en la Universidad de Chicago, Chicago, 22 y 23 de mayo de 2019, http://neubauercollegium.uchicago.edu/events/uc/cultural_analytics/; *Cultural Analytics 2017*, simposio en la Universidad de Notre Dame Notre Dame, 26 y 27 de mayo de 2017, <https://sites.google.com/nd.edu/ca2017>.

⁷ "Culture Analytics program", Institute for Pure and Applied Mathematics (IPAM), UCLA, del 7 de marzo al 10 de junio de 2016, <http://www.ipam.ucla.edu/programs/long-programs/culture-analytics/>.

⁸ "About" en *Journal of Cultural Analytics*, consultado el 23 de julio de 2019, <http://culturalanalytics.org/about/>.

término para convocatorias de conferencias y listas de trabajos académicos; y una serie de programas académicos y cursos particulares de pregrado y posgrado.

En septiembre de 2019, revisé estos programas y cursos: el curso de Analítica Cultural en King's College, Londres (Reino Unido); la especialización en Analítica Cultural en la Maestría en informática y analítica de la Universidad de Carolina del Norte (EE. UU.); el certificado de posgrado en Análisis Cultural en las bibliotecas de la Universidad de Temple (EE. UU.); el Centre for Cultural Analytics del University College Dublin (Irlanda); el curso de Analítica Cultural en Dartmouth College (EE.UU.); la Maestría en datos, cultura y visualización en la Universidad ITMO (Rusia); el curso *Cultural Analytics: The Computational Study of Culture*, de la Universidad McGill (Canadá); el CulturePlex Lab de Western University (Canadá), que investiga sobre “analítica cultural e innovación digital”; el curso *Distant Reading and Cultural Analytics* en la Universidad de California, Los Ángeles (EE. UU.) y el laboratorio Cultural Data Analytics en la Universidad de Tallinn (Estonia). Las descripciones de todos estos programas y cursos utilizan el término “analítica cultural” en una variedad de formas relacionadas con el contexto institucional en el que se ofrecen (por ejemplo, un departamento de literatura, un departamento de ciencias de la información, etc.).

Mirando los artículos publicados por el *Journal of Cultural Analytics* durante sus primeros tres años, vemos, de manera similar, una variedad de temas, aunque el análisis de textos literarios predomina sobre otros tipos de medios. Estos temas incluyen novelas en inglés del siglo XIX y contemporáneas, periódicos ilustrados del siglo XIX, literatura china imperial tardía, clasificaciones de folclore, menús de restaurantes en los Estados Unidos y series de televisión estadounidenses⁹.

El término “analítica cultural” puede significar diferentes cosas para diferentes personas, y puede usarse en diferentes contextos, para realizar diferentes trabajos. Esto está bien, y no estoy interesado en controlar su uso. Este libro no intenta resumir toda la investigación de analítica cultural ni cubrir todo lo que es relevante para dicha investigación (por ejemplo, estadísticas inferenciales, diseño de experimentos, aprendizaje automático supervisado y no supervisado, análisis de texto, análisis geoespacial, análisis de datos musicales, análisis de redes, simulación basada en agentes, y otros temas que no discuto). En lugar de tratar de hacer una enciclopedia que cubra todos los temas y técnicas relevantes, decidí centrarme en un pequeño número de temas y explorarlos con mayor profundidad.

La elección de estos temas refleja las motivaciones iniciales que me llevaron a investigar cómo las computadoras nos pueden ayudar a ver la cultura contemporánea; mi experiencia desde 2008 de trabajar en más de cuarenta proyectos prácticos en nuestro laboratorio; la enseñanza práctica de analítica cultural cada año en todas mis clases universitarias desde 2006 a estudiantes de pregrado y posgrado; la realización de numerosos talleres en diferentes países; y el aprendizaje de las colaboraciones con otros investigadores académicos y diseñadores. Todos los temas que se incluyen pueden verse en el índice. La organización de los temas sigue una secuencia lógica. La parte I brinda ejemplos de analítica cultural computacional y analiza el desplazamiento de los “nuevos medios” de la

⁹ “Articles” en *Journal of Cultural Analytics*, consultado el 1 de octubre de 2019, <https://culturalanalytics.org/#>.

década de 1990 a los "más medios" de la década de 2000, lo que me motivó a comenzar a pensar en la analítica cultural. En la parte II, discuto tipos de datos culturales y técnicas para representar procesos culturales como datos computacionales. La parte III presenta conceptos para la exploración de *datasets* culturales mediante la visualización de datos y luego se enfoca en métodos desarrollados recientemente para la exploración de colecciones de imágenes y videos.

Entre todos los temas e ideas presentados en este libro, hay cinco que me apasionan más. Tomadas conjuntamente, describen una versión de la analítica cultural explorada en este libro -aunque otras versiones, visiones, usos y definiciones del término son bienvenidos-. Aquí están estas cinco ideas:

(1) Mi motivación inicial para acercarme a los métodos computacionales y la *big data* surgió de darme cuenta de que la escala de la cultura en el siglo XXI hace que sea imposible verla con los métodos existentes. Por lo tanto, en este libro *la analítica cultural se refiere al uso de métodos computacionales y de diseño -incluyendo la visualización de datos, el diseño de medios e interacciones, las estadísticas y el aprendizaje automático-, para la exploración y el análisis de la cultura contemporánea a escala*. Uno de los objetivos de estas exploraciones es *permitirnos ver lo que cientos de millones de personas en todo el mundo crean, imaginan y valoran hoy*. Esto incluye actividades culturales tanto de usuarios comunes como de profesionales y estudiantes de la industria creativa: los cientos de semanas de diseño y de la moda que ocurren cada año, los miles de festivales de cine, las decenas de miles de programas educativos, y los cientos de miles de proyectos culturales y exhibiciones que se anuncian en la web y en las redes sociales, etc. Por lo tanto, el principal objetivo práctico de la analítica cultural para mí es trabajar hacia una comprensión más inclusiva y democrática del presente cultural y también de las historias culturales. Esto significa hacer completamente visible la "larga cola" de la vida cultural, y colocar en nuestros mapas culturales a las ciudades, países, grupos, creadores individuales y artefactos que han quedado fuera tanto de las narrativas culturales tanto históricas como actuales.

El segundo objetivo es generar *nuevos conceptos teóricos* apropiados a la escala, la velocidad, la diversidad y la conectividad de la cultura digital global contemporánea. ¿En qué se diferencia esto de la teoría cultural del siglo XX? Nuestros nuevos conceptos deben ser no solo teóricos sino también *cualitativos*; es decir, deberían permitirnos medir dimensiones de la cultura digital y comparar estilos, gustos, imaginación y comportamientos culturales en muchos lugares geográficos, redes y campos creativos (esto significa que también podemos querer formalizar y cuantificar los conceptos existentes, como, por ejemplo, el estilo). Pero además debemos pensar acerca de los límites de la cuantificación, y ser sensibles a las dimensiones y aspectos de la cultura que las mediciones existentes no logran capturar.

(2) El uso de métodos de representación numérica y de análisis y visualización de datos ofrece un nuevo lenguaje para describir artefactos, experiencias y dinámicas culturales. Como argumento en el capítulo 7, los lenguajes humanos que se desarrollaron recientemente en la evolución humana no son buenos para capturar las propiedades analógicas de las experiencias culturales y sensoriales humanas. Estas limitaciones se vuelven particularmente preocupantes si queremos comparar miles, millones o miles de millones de artefactos, es decir, estudiar la cultura contemporánea en su nueva escala. En

cambio, cuando usamos números y visualizaciones podemos capturar mejor las pequeñas diferencias entre muchos artefactos y también entre grupos de artefactos (aquí la ciencia de datos, que tiene muchos métodos para caracterizar las relaciones entre cualquier cantidad de objetos -análisis de conglomerados, reducción de dimensiones, análisis de redes, etc.- se vuelve particularmente relevante).

Los números y la visualización también nos brindan un lenguaje para representar *cambios temporales graduales y continuos*. Ahora podemos describir las características de los procesos culturales que son difíciles de captar lingüísticamente; por ejemplo, cambios históricos graduales en la cultura visual durante largos períodos o cambios temporales en la forma visual en la carrera de un artista.

Dado mi compromiso con el análisis de la cultura visual, sostengo que tener un mejor lenguaje para describir sus *dimensiones analógicas* es invaluable. Las computadoras digitales, que operan con representaciones numéricas, pueden capturar mejor estas dimensiones, como el movimiento o el ritmo, que los lenguajes naturales no pueden describir adecuadamente.

(3) Si bien muchas ideas y discusiones en el libro son relevantes para trabajar con todo tipo de datos, presto especial atención a los *medios visuales*. Quiero demostrar a través de muchos ejemplos cómo podemos usar métodos computacionales y de visualización para explorar colecciones visuales, para formular preguntas culturales interesantes. La mayor parte del trabajo computacional en humanidades hasta ahora se ha centrado en textos literarios, registros de textos históricos y datos espaciales. Por el contrario, otros tipos de medios, como las imágenes fijas y en movimiento y los medios interactivos, han recibido relativamente poca atención. Esta situación está mejorando gradualmente, pero mientras escribo esto el análisis de los medios visuales sigue siendo una pequeña parte de las humanidades digitales¹⁰. Esto puede comprobarse consultando los programas de las conferencias anuales organizadas por la Alianza de Organizaciones de Humanidades Digitales, o consultando las revistas del campo. Estas limitaciones están bien resumidas por el título de un artículo publicado en 2017 en la revista *Digital Scholarship in the Humanities*: “Las humanidades digitales son fuertes en texto, livianas en visualizaciones y pobres en simulaciones.”¹¹

Esto es sorprendente porque los informáticos ya habían comenzado a desarrollar métodos para el análisis de imágenes a finales de la década de 1950. Hoy en día se implementan en numerosos servicios y dispositivos digitales, incluyendo los buscadores de imágenes web, las cámaras fotográficas y las cámaras de los teléfonos móviles, el software de edición de imágenes como Photoshop, las redes para compartir imágenes, etc. En los campos de las ciencias de la computación llamados visión por computadora y computación multimedia, los investigadores han estado publicando durante muchos años nuevos algoritmos para la detección automática del contenido de imágenes, estilos artísticos, técnicas fotográficas,

¹⁰ Conferencia Humanidades Digitales 2019, Utrecht, Países Bajos, del 9 al 12 de julio de 2019, <https://dh2019.adho.org>.

¹¹ Erik Malcolm Champion, “Digital Humanities Is Text Heavy, Visualization Light, and Simulation Poor”, *Digital Scholarship in the Humanities* 32, s1 (2017): 25-32.

géneros de televisión y video, y aplicándolos a conjuntos de datos cada vez más grandes¹². En nuestro laboratorio, hemos estado utilizando algunos de estos métodos para analizar muchos tipos de medios visuales históricos y contemporáneos -por ejemplo, veinte mil fotografías de la colección del Museo de Arte Moderno (MoMA) de Nueva York, películas de Dziga Vertov del Austrian Film Museum, dieciséis millones de imágenes compartidas en Instagram en diecisiete ciudades del mundo, 270 millones de imágenes compartidas en Twitter en todo el mundo, un millón de páginas de manga y un millón de obras de arte de la popular red de arte DeviantArt-. En este libro, me referiré a los detalles de algunos de estos proyectos y a los artículos y trabajos de otros investigadores (generalmente de informática) que analizan el contenido visual usando algoritmos.

(4) ¿Podemos explorar y estudiar colecciones de medios culturales y registros de comportamientos culturales *sin sistemas de categorías* que los lenguajes imponen a la realidad? ¿Podemos *evitar la cuantificación, las medidas y los resúmenes* que vienen con el uso de estadísticas? ¿Podemos estudiar grandes datos culturales *sin usar números*?

La respuesta aparentemente obvia a estas preguntas es no. Pero la respuesta que da la analítica cultural es sí (al menos para algunos tipos de medios, como imágenes y videos). Cualquier medida numérica, categoría lingüística, conjunto de etiquetas, redes u otras *formas de representación* -sin importar su poder descriptivo y capacidad de hacer visibles similitudes y otras formas de relaciones entre objetos- son también *formas de omisión*. Por ejemplo, las técnicas de visión por computadora actuales pueden detectar miles de tipos de objetos en fotografías¹³, pero cualquier estudiante de arte o fotógrafo profesional sabe que una foto no es simplemente una colección de objetos o figuras humanas en ella. Más que reemplazar a los observadores culturales humanos con algoritmos, la intención de la analítica cultural es aumentar nuestras habilidades humanas *proporcionándonos nuevas interfaces y técnicas para observar flujos y conjuntos de datos culturales masivos* (esto es consistente con la visión de la computación desarrollada por Douglas C. Engelbart en su famoso informe de 1962, "Aumentar el intelecto humano"¹⁴). Los humanos pueden identificar dimensiones más significativas, reconocer pequeños detalles que realmente importan, y situar a la información en más contextos que cualquier algoritmo y, a menos que el trabajo sobre la "inteligencia artificial general" progrese lo suficiente, esto seguirá siendo cierto en el futuro. Pero, ¿podemos ampliar estas capacidades humanas para aprovechar la escala de la producción y participación cultural global? En resumen: ¿Cómo podemos ver (por ejemplo) mil millones de imágenes?

(5) La analítica cultural incluye no solo la aplicación de métodos computacionales actualmente disponibles al análisis de los *datasets* y flujos de datos culturales, sino también la *evaluación crítica de estos métodos de ciencia de datos y sus supuestos*. Hoy nuestras interacciones con los medios digitales, nuestro acceso a la información y conexión entre nosotros en las redes sociales, está mediado por sistemas de software. Estos analizan

¹² Miriam Redi, Frank Z. Liu y Neil O'Hare, "Bridging the Aesthetic Gap: The Wild Beauty of Web Imagery", en *Actas de la Conferencia internacional ACM 2017 sobre recuperación multimedia*, (Nueva York: ACM, 2017), 242-250.

¹³ Clarifai informó que podía detectar once mil tipos de objetos de manera confiable en 2017. Consulte Clarifai, "Modelos", consultado el 13 de octubre de 2017, <https://www.clarifai.com/models>.

¹⁴ Douglas Engelbart, "Aumentar el intelecto humano: un marco conceptual", Proyecto SRI no. 3578, octubre de 1962, <http://dougengelbart.org/content/view/138>.

continuamente "big data cultural", es decir, el contenido de miles de millones de artefactos de medios que compartimos, nuestras interacciones online con estos artefactos y otros comportamientos físicos y en línea. En la investigación de la analítica cultural, a menudo usamos métodos similares para diferentes propósitos -por ejemplo, identificar patrones en la historia cultural, explorar el trabajo de diseñadores contemporáneos y examinar el contenido y los estilos de las fotos compartidas por miles de millones de personas *online*-. ¿Cuáles son las similitudes y diferencias entre el uso de estos métodos en la industria y en la investigación cultural? ¿Existen supuestos y objetivos integrados en los métodos que se utilizan ampliamente en la industria que debemos cuestionar si los adoptamos? ¿Qué desarrollos históricos llevaron a la popularidad de ciertos métodos en la actualidad? Abordar tales preguntas es también la parte clave de la agenda de análisis cultural.

El análisis crítico de la ciencia de datos, los algoritmos, y el uso de datos en la sociedad en general también se lleva a cabo en varios campos de investigación, entre ellos los estudios de ciencia y tecnología, las humanidades digitales, los estudios de cultura digital, los estudios críticos de datos y algoritmos¹⁵ y los estudios de software. Entre las numerosas revistas académicas de ciencias sociales y humanidades que publican artículos sobre estos temas, puedo destacar *Big Data & Society*. Le aconsejo que mire las publicaciones y conferencias en estos campos y lea los artículos que le parezcan relevantes para sus intereses.

A lo largo de este libro, discutiré lo que personalmente veo como *las líneas de investigación más interesantes y prometedoras en la analítica cultural*. Por *interesante* me refiero al análisis que nos permite pensar la cultura contemporánea de nuevas maneras y nos ayuda a cuestionar conceptos y métodos que damos por sentado para el estudio de la cultura. Algunas de esas líneas pueden ilustrarse con el trabajo existente, mientras que otras aún no se han llevado a cabo. Por lo tanto, si se va a adentrar en el análisis cultural, espero que encuentre en este libro ideas interesantes para probar cosas nuevas.

La analítica cultural es solo uno de los paradigmas que surgieron en la segunda mitad de los 2000 para aprovechar la disponibilidad de grandes cantidades de datos culturales y sociales. Estos incluyen las humanidades digitales, las ciencias sociales computacionales, la computación social, la antropología digital, la historia digital, la ciencia de las ciudades, la informática urbana y la "culturómica"¹⁶. Al mismo tiempo, grandes conjuntos de datos culturales comenzaron a analizarse y utilizarse en muchas áreas de las ciencias de la computación, como aprendizaje automático, inteligencia artificial, visión artificial, procesamiento de lenguaje natural y multimedia, así como también en ciencia de redes y estudios de comunicación. A principios de la década de 2010, comenzó el "giro cuantitativo" en la historia del arte, con el inicio de su publicación en el *International Journal for Digital Art* en 2015. En estudios cinematográficos, la primera monografía que utiliza métodos

¹⁵ David Moats y Nick Seaver, "You Social Scientists Love Mind Games": Experimenting in the 'Divide' between Data Science and Critical Algorithm Studies", *Big Data & Society* 6, no. 1, (2019), <https://doi.org/10.1177/2053951719833404>; Andrew Iliadis y Federica Russo, "Critical Data Studies: An Introduction", *Big Data & Society* 3, no. 2 (2016), <https://doi.org/10.1177/2053951716674238>.

¹⁶ Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science* 331, no. 6014 (2011): 176–182, <https://doi.org/10.1126/science.1199644>.

cuantitativos y visualización de datos para analizar obras de un solo director de cine apareció en 2018¹⁷.

En la misma década surgieron una serie de nuevas agendas de investigación para abordar el uso creciente de algoritmos, datos y sistemas de inteligencia artificial (IA) desde la perspectiva de las ciencias sociales y las humanidades. Estas áreas de investigación incluyen el comportamiento de las máquinas, y los estudios de algoritmos y de datos ya mencionados (la antología de acceso abierto de 2017 *The Datafied Society: Studying Culture through Data* recopila una serie de artículos que abordan cuestiones metodológicas y éticas relacionadas con el uso de datos y algoritmos en la investigación académica¹⁸).

La investigación que llevamos a cabo en nuestro laboratorio se desarrolló en diálogo con el trabajo de los investigadores en esos campos y todos los proyectos relevantes que vi: esa investigación no hubiera sido posible sin todo lo que fui aprendiendo de ellos. El objetivo de este libro no es trazar un límite alrededor del campo o afirmar que somos los únicos que investigamos sobre ciertos temas. Desde mi punto de vista, la analítica cultural es un instrumento para cuestionar todos los límites categóricos, por lo que no sería acertado trazar un límite alrededor de toda la investigación de analítica cultural que se lleva a cabo en muchos campos académicos y profesionales.

Analítica cultural: doce desafíos de investigación

Mi pensamiento sobre la analítica cultural y el trabajo en nuestro laboratorio se ha guiado por una lista de desafíos teóricos y prácticos que definí en la época en que se creó el laboratorio. Solo algunas de esas preguntas serán discutidas en detalle en este libro: son aquellas en las que terminé invirtiendo más energía y tiempo. Pero quiero dar la lista completa porque estas preguntas también pueden serles útiles en su trabajo. Hice una lista de los primeros ocho desafíos en 2005-2007 (preguntas 1-8 a continuación); después de años de investigación, los nuevos desafíos se vuelven evidentes (preguntas 9 a 12):

1. ¿Cómo puede ayudarnos el trabajo con “grandes cantidades de datos culturales” a cuestionar nuestros estereotipos, supuestos, conceptos y conocimientos existentes sobre las culturas?
2. ¿Cuáles son las nuevas formas de comprender y estudiar las culturas visuales y mediáticas, que permiten los métodos computacionales y los grandes conjuntos de datos?
3. ¿Cómo podemos explorar colecciones visuales masivas que pueden contener miles de millones de imágenes y videos?
4. ¿Cómo podemos combinar el análisis computacional de medios con métodos y teorías de estudios cualitativos?

¹⁷ Adelheid Heftberger, *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work* (Basel: Springer, 2018).

¹⁸ Karin van Es y Mirko Tobias Schäfer, eds., *The Datafied Society. Studying Culture through Data* (Ámsterdam: Amsterdam University Press, 2017), <https://library.oapen.org/handle/20.500.12657/31843>.

5. ¿Cómo usar enfoques computacionales para analizar experiencias y medios interactivos (p. ej., jugar un videojuego, interactuar con la aplicación de Instagram, experimentar una instalación interactiva), en lugar de artefactos de medios estáticos?

6. ¿Qué conceptos y modelos teóricos necesitamos para lidiar con la gran escala y la velocidad del contenido generado por el usuario y las interacciones del usuario *online*?

7. ¿Cómo podemos analizar y visualizar la diversidad de las culturas digitales globales contemporáneas, teniendo en cuenta las actividades de miles de millones de creadores y trillones de objetos que están creando?

8. ¿Cómo será la “ciencia de la cultura” impulsada por conjuntos masivos de datos culturales y computación, y cuáles serán sus limitaciones?

9. ¿Podemos definir medidas cuantitativas generales de la variabilidad cultural, la diversidad, el cambio temporal, la diferencia, influencia y singularidad, que serán significativas para muchos tipos de medios y diferentes períodos y culturas -en particular nuestro propio período-?

10. Dado que los métodos estadísticos y de ciencia de datos se basan en la reducción y la síntesis de datos, ¿cómo podemos además analizar computacionalmente pequeñas diferencias y detalles únicos de artefactos y experiencias individuales?

11. ¿Podemos describir las culturas -tanto objetivamente como por el modo en que las percibimos-, como distribuciones estadísticas y combinaciones de elementos, temas y estrategias? ¿O la cultura trata de “gestalts” no reducibles a sus partes? Si la respuesta a la segunda pregunta es sí, esto puede hacer imposible a la analítica cultural.

12. Supongamos que podemos detectar una pequeña cantidad de temas, estilos y técnicas culturales en miles de millones de artefactos, experiencias e interacciones culturales. Supongamos además que rastreamos estos ADN de cultura en todo el mundo, agregando otros nuevos cuando los detectamos. ¿Qué nivel de reducción es apropiado cuando extraemos una pequeña cantidad de temas de miles de millones de “objetos” culturales y qué se pierde en cada nivel? Por ejemplo, ¿cuánta información se pierde si extraemos 10.000 temas, y luego los agregamos en 1.000 temas, y luego agregamos estos datos nuevamente para tener solo 100 temas? ¿Es la reducción un enfoque erróneo? Si los autores contemporáneos quieren crear artefactos y experiencias únicos que no se puedan duplicar, ¿perderemos inevitablemente la verdadera singularidad en nuestra búsqueda para seguir tendencias más amplias?

Veo este último desafío como el más importante. ¿Deberíamos agregar grandes datos culturales y reducirlos a un número menor de estructuras -solo las ideas, temas, estilos, patrones y comportamientos que ocurren con mayor frecuencia-? Este es el paradigma que heredamos de la historia de la estadística, y muchos estudios cuantitativos de la cultura en ciencias de la computación lo siguen. En este paradigma, nos enfocamos en lo que es común entre varios objetos y no incluimos lo que ocurre con menor frecuencia. ¿O deberíamos desarrollar el paradigma opuesto: rechazar la agregación y la reducción y, en

cambio, centrarnos en la diversidad, la variabilidad y las diferencias entre numerosos artefactos, comportamientos e individuos? En este paradigma, incluimos todos los datos y prestamos especial atención a los fenómenos poco frecuentes y raros.

Tomadas conjuntamente estas doce preguntas deberían aclarar aún más de qué manera mis propias motivaciones e intereses por la analítica cultural que guían este libro difieren de los de otros investigadores que trabajan en esta área. Mis principales motivaciones al usar "big cultural data" son: cuestionar qué pensamos que sabemos sobre las culturas (1), en lugar de hacer sólo progresos técnicos en paradigmas ya bien establecidos; abordar los desafíos de pensar en nuevas formas de cultura digital (5); y comprender cómo se puede tratar su escala, velocidad y diversidad, tanto a nivel de conceptos teóricos como de características que podemos medir (6, 7, 9). Surgen otros desafíos cuando confrontamos la "cognición de datos" de la sociedad contemporánea (es decir, los supuestos y métodos de la ciencia de datos) con los tipos de sujetos para los que estos métodos no fueron diseñados: vida cultural, experiencias y artefactos (2, 3, 4) . Mi optimismo inicial sobre las posibilidades aparentemente infinitas de la analítica cultural gradualmente dio paso a una visión más realista cuando me di cuenta de las limitaciones de los enfoques estadísticos y computacionales y, en general, las limitaciones de pensar en las culturas como combinaciones de elementos que podemos rastrear (10-12).

Por lo tanto, este libro apunta a convencerlo de que estos enfoques son muy útiles y, a veces, la única forma de abordar la escala de la cultura actual, y al mismo tiempo explicar claramente lo que, al menos en la actualidad, las computadoras no pueden ver. Y esto no es porque nazcan, por así decirlo, con ceguera en el dominio estético y no se puedan curar. El problema reside en sus maestros, es decir, en nosotros. Si nosotros mismos no entendemos por qué, entre cientos de fotografías de un mismo tema, el editor de la revista elige una en particular, aunque cuando una computadora mide sus muchas características resultan ser prácticamente idénticas, entonces ¿cómo podemos esperar enseñar tales juicios estéticos a una máquina? Ciertamente, podemos alimentar con millones de ejemplos de este tipo a una red neuronal y aprenderá a predecir con cierta precisión las "mejores" fotografías, pero esa visión probabilística no es lo mismo que la comprensión.

Qué no es la analítica cultural

Habiendo definido las ideas clave y doce desafíos de investigación para la analítica cultural tal como lo veo, además necesito establecer claramente lo que no es la analítica cultural. El rápido crecimiento de muchas redes sociales y la disponibilidad de sus datos a través de Interfaces de Programación de Aplicaciones (API por sus siglas en inglés) entre aprox. 2007 y 2015 promovió una gran cantidad de investigaciones en varios campos. En nuestro laboratorio también hemos aprovechado el acceso a datos de este nuevo universo cultural. Entre 2012 y 2015, nosotros y nuestros colaboradores trabajamos con conjuntos de imágenes compartidas públicamente o información sobre ellas de Twitter, Instagram y VK, creamos una serie de visualizaciones y publicamos nuestra investigación en algunos artículos (las API de las redes proporcionaron nombres de usuario de las publicaciones, pero nunca usamos esta información en ninguna publicación o exhibición). Durante los años de las API abiertas, decenas de miles de científicos también utilizaron grandes cantidades

de datos de redes sociales. De hecho, desde hace varios años, descargar datos de Twitter y analizarlos se usa a menudo como un ejercicio en muchas clases de informática y ciencia de datos.

El crecimiento global de las redes sociales y los sitios para compartir medios después de 2007 ciertamente confirmó mis ideas sobre la necesidad de la analítica cultural formuladas anteriormente. Sin embargo, la analítica cultural no está “casada” con este tipo de medios y datos. Las redes sociales surgieron recientemente en la larga historia de los medios y es posible que no existan de la misma forma en el futuro. De hecho, en 2005 ni siquiera estaba pensando en usar los datos de las redes sociales, porque éstas aún no eran muy populares y no había ningún mecanismo para descargar sus datos. En cambio, estaba pensando en recopilar información de numerosos sitios web pertenecientes a diseñadores individuales, centros culturales, publicaciones, escuelas de arte, museos, y analizar el contenido de blogs relacionados con la cultura que ya eran muy populares. Me imaginé accediendo al contenido de tantos sitios como fuera posible en todo el mundo y visualizando los patrones cambiantes en tiempo real (ver imágenes 1 y 2).

El acceso gratuito al contenido de las principales redes sociales que estuvo disponible en unos pocos años fue importante para hacer visibles las oportunidades y los desafíos en la observación y el análisis de datos culturales a escala, pero el programa de analítica cultural no depende del acceso a las redes sociales ni a ningún otra fuente concreta. Si en algún momento en el futuro, los sitios web y las redes sociales dejan de existir en sus formas actuales, es probable que otras formas de publicación colectiva y uso compartido de medios los reemplacen.

Es casi seguro que determinadas tecnologías cambiarán, pero se mantendrán las nuevas condiciones fundamentales que ya se establecieron durante la década de 2000: la nueva escala de la cultura y la creciente presencia de la cultura en las sociedades actuales. Esto incluye tanto “más cultura” (más productores culturales, más objetos y eventos, más áreas de la sociedad donde la estética se vuelve muy importante, etc.) y “más información” acerca de la cultura (sitios web, publicaciones, conjuntos de datos). El objetivo de la analítica cultural es abordar el desafío de ver y pensar la cultura a esta nueva escala.

Analítica cultural, teoría de los medios y estudios de software

Analítica cultural es un libro de *teoría de los medios*. Sostengo que para estudiar sistemáticamente las culturas de los medios globales hoy en día, la teoría de los medios debe recurrir a la ciencia de datos. De hecho, los métodos computacionales son necesarios no solo para analizar y teorizar estas culturas globales, sino incluso para *verlas* en primer lugar.

Si confiamos solo en nuestras intuiciones basadas en el contenido que vimos en línea, las recomendaciones algorítmicas o intuiciones sobre lo que es importante, sería demasiado fácil permanecer en nuestras *burujas de filtro* cognitivo e histórico, proyectando nuestros

sesgos en el mundo¹⁹. Los campos académicos pueden tener sus propias burbujas de filtro, en las que los paradigmas de investigación establecidos pueden hacer invisibles para los académicos los tipos de actividades culturales emergentes a las que deberían prestar atención. Por ejemplo, si bien el diseño de interacciones se ha convertido en el centro de nuestras experiencias culturales diarias, en forma de aplicaciones, sitios web y dispositivos conectados, aún no se analiza lo suficiente en los estudios de medios o humanidades.

Por supuesto, los métodos computacionales y los grandes conjuntos de datos no garantizan automáticamente una mayor objetividad e inclusión. Sin embargo, pueden ayudarnos a *confrontar nuestros supuestos, prejuicios y estereotipos*. Nos permiten identificar lo que de otro modo no podríamos ver: el contenido y sus creadores que no llegan a los primeros resultados de los motores de búsqueda o recomendación, no aparecen en las listas del “top ten” o los 100 principales y, por lo tanto, permanecen invisibles.

¿Cómo se investiga la “larga cola” de la cultura digital?²⁰ Numerosos investigadores en informática han estado utilizando muestras aleatorias para seleccionar millones de publicaciones de texto, imágenes y videos compartidos en Twitter, YouTube, Instagram y otras redes, y luego analizarlos. Muestras tan grandes capturan una instantánea de la actividad global en tales redes, pero es posible que no representen bien las diferencias sustanciales en el contenido compartido en diferentes áreas geográficas o por diferentes grupos demográficos. En varios proyectos en nuestro laboratorio, seguimos una estrategia diferente: seleccionar un área geográfica pequeña y luego recopilar todo el contenido compartido en esta área. Por ejemplo, para los proyectos *On Broadway* e *Inequaligram*, analizamos los datos de todas las imágenes geolocalizadas de Instagram que se compartieron en todo Manhattan durante cinco meses²¹. No filtramos nada. No comenzamos con ningún hashtag que represente un tema en particular. No solo nos fijamos en las imágenes que tenían más likes. No separamos “arte” de “no arte”, ni las imágenes “originales” de las copias, o los “influencers” de los usuarios habituales. En cambio, todas las publicaciones con información de ubicación que indica que se compartieron en Manhattan durante este período (las 7.442.454 de imágenes, etiquetas y descripciones geolocalizadas de 1.890.585 usuarios de Instagram) se recopilaron en nuestro laboratorio y se consideraron igualmente valiosas para un análisis posterior.

Antes de que podamos “teorizar” a los medios contemporáneos, necesitamos verlos, y esto ya no es posible sin las computadoras debido a su nueva escala. Por lo tanto, en lugar de ser solo el *tema de análisis*, como en mi libro *El lenguaje de los nuevos medios*, la computación se vuelve la *herramienta práctica* para estudiar los medios en este nuevo libro.

¹⁹ Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (New York: Penguin Press, 2011).

²⁰ Chris Anderson, “The Long Tail”, *Wired*, 1 de octubre de 2004, <https://www.wired.com/2004/10/tail/>; Erik Brynjolfsson, Yu Jeffrey Hu y Michael D. Smith, “The Longer Tail: The Changing Shape of Amazon’s Sales Distribution Curve”, 22 de septiembre de 2010.

²¹ “On Broadway,” 2014; <http://on-broadway.nyc/>; “Inequaligram,” 2016, <http://inequaligram.net/>.

Desde 1984, he estado trabajando con medios digitales con diferentes roles: como animador por computadora, diseñador de gráficos en movimiento, desarrollador de software, artista de medios y profesor de arte digital. He impartido clases prácticas de arte digital desde 1992, visualización de datos desde 2006 y ciencia de datos desde 2013. También escribí herramientas de software que usamos en nuestro laboratorio para visualizar grandes conjuntos de datos visuales. Esta experiencia de diseño, programación y enseñanza de medios digitales, visualizaciones y análisis de datos se refleja en este libro. Explorar, manipular y visualizar conjuntos de datos, a mi modo de ver, es una continuación directa de la creación de arte y diseño con código que he estado haciendo desde 1984.

De igual modo, veo mi investigación de analítica cultural como una extensión directa de mi trabajo de teoría sobre los nuevos medios de los años 1990 y 2000. La diferencia es la escala de la cultura digital entonces y ahora. A mediados de los '90, la cantidad de artistas que trabajaban con algoritmos era tan pequeña que todos podíamos encontrarnos en una sola conferencia. Dos de estas conferencias anuales clave fueron el festival *Ars Electronica*, que comenzó en 1979, y el Simposio Internacional anual sobre Arte Electrónico (ISEA) que se realizó por primera vez en 1988. En el ISEA de 1994, en Helsinki, se reunieron aproximadamente 150 participantes, y esto promovió el despegue internacional del arte de los nuevos medios.

Hoy en día, cientos de miles de personas se identifican como artistas digitales, tecnólogos creativos o programadores creativos, y unos cuantos miles de millones de personas con teléfonos con cámara se han convertido en fotógrafos digitales. Una búsqueda de la frase "Cómo edito mi Instagram" en YouTube (15 de enero de 2016) arrojó 150.000 videos con consejos prácticos de usuarios de Instagram. Y la misma búsqueda del 11 de octubre de 2017 arrojó 228.000 videos. Los mejores videos de este género tienen millones de visitas cada uno²². Esta nueva escala de producción y participación en los medios exige nuevos métodos, conceptos y herramientas de investigación, y eso motiva este libro. En el mundo en el que los medios digitales son creados por unos pocos miles de millones en lugar de unos pocos miles, como era hace veinticinco años, necesitamos reinventar lo que significa estudiar la cultura.

Analítica Cultural también incorpora perspectivas del campo de los estudios de software, que se preguntan cómo el software le da forma al mundo actual. En *El software toma el mando*, escribí: "Si queremos comprender las técnicas actuales de *control, comunicación, representación, simulación, análisis, toma de decisiones, memoria, visión, escritura e interacción*, ... [debemos] considerar esta capa de software"²³. *El software toma el mando* presentó un análisis teórico e histórico de herramientas populares para la creación de medios como Photoshop y After Effects; *Analítica Cultural* analiza algunos de los conceptos y suposiciones centrales de la visión del mundo centrada en los datos. Se pregunta *cómo piensa y actúa nuestra sociedad usando datos y algoritmos*, y cómo el *análisis algorítmico del contenido y las interacciones del usuario por parte de la industria da forma a la cultura actual*.

²² "How I Edit My Instagram" (resultados de búsqueda), YouTube, consultado el 15 de enero de 2016, https://www.youtube.com/results?search_query=%22how+i+editar+mi+instagram%22.

²³ Lev Manovich, *Software Takes Command*, rev. edición (Londres: Bloomsbury Academic, 2013). Se publicó una versión anterior bajo una licencia Creative Commons en 2007, 15; cursiva en el original.

Creo que todos los miembros de la clase creativa, investigadores de medios, humanistas y científicos sociales necesitan tener una *alfabetización en ciencia de datos*: conocimiento de los principios básicos del análisis de datos, aprendizaje automático, y métodos de análisis predictivo y sus aplicaciones. ¿Por qué? Porque el software y el código basados en estos principios se emplean en toda nuestra sociedad, incluyendo la industria de la cultura digital, las empresas, las organizaciones sin fines de lucro y el gobierno. Y si la ciencia de datos aún no ha comenzado a usarse en un campo académico en particular es solo cuestión de tiempo.

En resumen, veo mi paso de los *medios* al *análisis de datos* como una progresión lógica. *El lenguaje de los nuevos medios* (escrito en 1999) describía formas de la cultura digital que surgieron en los '90. *El software toma el mando* (2007) cubrió la historia de los programas de software para la creación y edición de medios, y los nuevos lenguajes visuales que permitieron a medida que fueron adoptados a principios del siglo XXI. *Analítica Cultural* investiga la nueva etapa posterior a 2005, en la que miles de millones de personas crean medios digitales y los comparten en línea. Además, en esta etapa se le da un nuevo protagonismo al software cultural. Le delegamos más agencia. Ya no es sólo una herramienta, un medio o un asistente. En cambio, ahora se involucra en *comportamientos culturales* (por ejemplo, decidir qué nuevas publicaciones de redes sociales mostrar, mejorar la calidad de nuestras fotos, escribir artículos de noticias, etc.). Y aunque por ahora todavía escribimos publicaciones, sacamos fotos y realizamos otras acciones culturales nosotros mismos, gradualmente esas acciones pueden automatizarse por completo en el futuro (por ejemplo, en 2018, Google agregó una función de autocompletar a Gmail que completa automáticamente una respuesta de correo electrónico a medida que uno empieza a escribir; solo necesita presionar la tecla Tabulador para aceptar la sugerencia). Es por eso que todos los académicos y estudiantes culturales y de los medios deben adquirir una buena comprensión de los campos de la ciencia de datos y la IA. El capítulo 3 de este libro presenta un análisis de estos nuevos roles culturales de las computadoras; mi libro de 2018 *Estética de AI* analiza cómo el uso creciente de sistemas algorítmicos puede afectar la diversidad cultural²⁴.

Uso de este libro en clases

El contenido y la estructura de este libro reflejan mi experiencia enseñando clases prácticas de análisis cultural a diversos grupos de estudiantes. Estos incluyeron a estudiantes universitarios en arte digital, arte de los medios, ciencias de la computación e historia del arte; y estudiantes graduados en informática y humanidades, y campos de ciencias sociales como historia del arte, literatura, musicología, comunicación, economía, sociología, antropología, psicología y humanidades digitales²⁵.

Los capítulos de este libro presentan una secuencia de temas que se pueden cubrir en un semestre o en clases de un trimestre (10 a 14 semanas). Los objetivos son familiarizar a los

²⁴ Lev Manovich, *AI Aesthetics* (Moscow: Strelka Press, 2018).

²⁵ Lev Manovich, "Teaching", consultado el 20 de septiembre de 2019, <http://manovich.net/index.php/teaching>.

estudiantes con ejemplos de trabajo con conjuntos de datos culturales en humanidades, ciencias de la computación, diseño y otros campos; y explicar por qué necesitamos métodos computacionales para analizar la cultura contemporánea (parte I); aprender las operaciones conceptuales, las elecciones y las limitaciones involucradas en la creación de “datos culturales” (parte II); y comprender cómo explorar *datasets* de medios utilizando visualizaciones de datos (parte III).

¿Por qué elegí incluir estos temas y no otros? En términos más generales, ¿cuál es la diferencia entre la analítica cultural y el análisis de datos? ¿Por qué no hay capítulos en el libro que aborden el análisis de datos?

Considere el flujo de trabajo para realizar una investigación, un diseño o un proyecto artístico con datos: (1) pensar cómo se pueden analizar o representar cuantitativamente algunos temas; (2) investigar qué datos adecuados están disponibles o cómo generarlos; (3) reunir los datos; (4) utilizar métodos visuales para explorar estos datos; (5) analizar los datos utilizando métodos de estadística y ciencia de datos (es decir, estadísticas descriptivas e inferenciales, aprendizaje automático supervisado y no supervisado, análisis de series temporales, ciencia de redes, etc.); y opcionalmente (6): crear herramientas de visualización interactivas para que otros exploren estos datos, o realicen otros diseños y productos de medios.

Desde mi punto de vista, para el análisis cultural el paso 5 no es realmente diferente de lo que haría con cualquier otro dato. Y dado que hay muchos buenos libros de texto, cursos en línea y tutoriales disponibles para aprender estos métodos, además de los cursos que puede ofrecer su universidad, se pueden usar estos recursos para aprenderlos. Entonces en lugar de cubrir el material que ya existe en otros lugares, este libro trata sobre las otras partes del flujo de trabajo: 1, 2, 3, 4 y 6. En otras palabras, lo que es único en el análisis cultural no es lo que se hace con los datos, sino cómo se llega desde esta cosa elusiva que llamamos *cultura* a su representación de datos en primer lugar. Es decir, ¿cómo convertir *experiencias, eventos, acciones y medios culturales en datos*? ¿Qué se gana y qué se pierde en esta traducción? Y una vez que lo obtiene en forma de datos, ¿cómo se lo puede explorar en múltiples escalas, identificando tanto patrones únicos e infrecuentes, como otros comunes y regulares?

Sin embargo, aunque no voy a enseñarle estadística y análisis de datos en este libro, hablaré sobre los supuestos detrás de algunos de sus métodos; lo que nos permiten ver y lo que no, al menos en la actualidad. Por lo tanto, idealmente debe leer este libro en paralelo al aprendizaje de las técnicas de ciencia de datos, o después de haber aprendido algunas de ellas.

Puede leer los capítulos del libro en orden, o simplemente pasar a cualquier capítulo que le interese. He tratado de hacer que cada capítulo sea relativamente autosuficiente. A lo largo del libro, encontrará resúmenes del material ya cubierto y presentaciones de material nuevo organizado en listas numeradas. Espero que esta organización sea útil tanto si se usa el libro completo como capítulos separados en las clases.