

Cómo estudiar la construcción de la imagen de una ciudad a través de publicaciones de Instagram: una metodología aplicada a Granada

How to study the construction of a city's image through Instagram posts: a methodology applied to Granada

Fco. Javier Cantón Correa. Universidad de Granada

Doctor en Ciencias Sociales, Licenciado en Sociología y Comunicación Audiovisual y Máster en Problemas Sociales por la Universidad de Granada, ha trabajado en diversas Universidades y medios y de comunicación. Actualmente desarrolla su trabajo en Medialab UGR e investiga sobre pensamiento visual, redes sociales, comunicación y visualización de datos. ORCID: <https://orcid.org/0000-0002-8466-1679>

Artículo recibido: 30/10/2019 – Aceptado: 22/11/2019

Resumen:

Nuestro mundo digital es cada vez más visual. Aplicaciones móviles centradas en la fotografía digital como Instagram son vehículos para la creación, manipulación y difusión instantánea de imágenes. Instagram supone, por tanto, una ventana abierta a la investigación en Ciencias Sociales y Humanidades Digitales, y una oportunidad de investigar cómo los jóvenes usuarios de esta aplicación desarrollan la cultura visual en sus entornos locales a través de lenguajes visuales globales. En este trabajo se repasa la metodología usada para estudiarlo a través del análisis de la producción de Instagram en Granada a partir de una muestra de 955.564 publicaciones y 375.758 imágenes publicadas y geolocalizadas recogidas a lo largo de un año (entre abril de 2017 y de 2018), con el objetivo de mostrar cómo se construye socialmente la imagen de una ciudad.

Palabras clave:

medio urbano; fotografía; tecnología y cambio social; tratamiento digital de imágenes.

Abstract:

Our digital world is increasingly visual. Mobile applications focused on digital photography such as Instagram are vehicles for the creation, manipulation and instant dissemination of images. Therefore, Instagram is an open window to researching in Social Sciences and Digital Humanities, and an opportunity to investigate how the young users of this application are developing visual culture in their local environments through global visual languages. This work reviews the methodology used to study it through the analysis of Instagram production in Granada from a sample of 955,564 publications and 375,758 posts and geolocated images collected over a year (between April 2017 and 2018), with the aim of showing how the image of a city is socially constructed.

Keywords:

urban environment; photography; technology and social change; digital image processing.

1. Introducción

En sociedades y culturas crecientemente visuales como la actual, el texto está perdiendo relevancia ante la imagen, en la denominada Era del Post-Texto (Manjoo, 2018), a la vez que se redefine el papel de la fotografía como rito social y herramienta para la construcción de las identidades. Instagram, una aplicación móvil nacida en 2010, se ha consolidado gracias a su mezcla única de sociabilidad y edición audiovisual. Y además de ser una aplicación móvil enfocada a la fotografía y la conectividad social, Instagram supone algo más para las Ciencias Sociales y la investigación social: es una ventana abierta para observar cómo se construye socialmente nuestra cultura visual a nivel global, en entornos locales, glocalmente. Aquí se presenta una metodología para estudiar la imagen de una ciudad a través de imágenes publicadas en Instagram, tal y como se ha usado en la Tesis defendida por el autor, para presentar los diferentes análisis, resultados y conclusiones a partir de los datos extraídos. En ella se ha observado cómo se construye la visualidad de una ciudad como Granada, donde existe un foco visual de enorme importancia, la Alhambra, y a partir de esta desigualdad de base intentar comprender cómo se visibilizan e invisibilizan zonas de la ciudad, así como, a partir de determinadas afinidades y proximidades, se construyen comunidades en torno a intereses concretos a través de Instagram como red, ya no visual y creativa, sino social.

Esta investigación es una aplicación de las investigaciones y metodologías creadas por Lev Manovich y su equipo de la Iniciativa de Estudios de Software (Boy & Uitermark, 2016; Manovich, 2009, 2012, 2017; Manovich, Goddemeyer, Stefaner, & Baur, 2015; Manovich et al., 2014; Manovich & Indaco, 2017) para grandes ciudades a una ciudad mediana como es Granada, aprovechando el conocimiento del entorno local para contextualizar los datos recabados. Ha logrado ser, además, el primer estudio en España que ha usado Instagram como una herramienta científica para la observación de la construcción de la cultura visual y la imagen pública de una ciudad española de tamaño medio. Ofrece, además, una metodología sencilla y replicable que garantiza su aplicación para el estudio de cualquier otra ciudad a partir de los datos generados.

2. Metodología

Las imágenes siempre han sido un medio muy poderoso para el registro y transmisión de información y la principal herramienta en la construcción de nuestra cultura visual, compuesta de múltiples imágenes, combinadas con el texto en numerosos momentos históricos, ya sea en forma de símbolos, signos, iconos, figuras, gráficos, mapas, diseños o diagramas (Mirzoeff, 2016). La forma en la que se usan las imágenes para construir identidad y sociedad fue lo que atrajo mi interés y es lo que acabó dando pie a mi investigación, que cobra renovada importancia por el nuevo tiempo que vivimos actualmente, un tiempo postfotográfico, en el que la fotografía se ha transformado completamente. Un tiempo en el que Internet y los dispositivos móviles han supuesto

una enorme disrupción en las prácticas de creación, circulación y recepción de las imágenes (Martín Prada, 2018). Por ello se vuelve más acuciante para un correcto acercamiento a la cultura visual global partir de planteamientos metodológicos relacionados con los Estudios Visuales más actuales aunque ortodoxos (Hand, 2014a, 2014b, 2017; Rose, 2016), pero también de enfoques más cercanos a las posibilidades actuales relacionadas con el Big Data, como ocurre con los estudios realizados desde el Análisis Cultural y los Estudios de Software, enfoque desarrollado por Lev Manovich y su equipo.

A través de este trabajo pretendí analizar la visualidad de una ciudad española concreta, Granada, mostrando simultáneamente algunas de las posibilidades que existen para el análisis de datos de Instagram, con el fin de descubrir la configuración y pautas de uso de la aplicación en una ciudad de tamaño medio. El objetivo general de la investigación fue, por tanto, observar de qué manera se construía la imagen social de una ciudad y, por ende, la cultura visual a un nivel local, lo que, por un lado, visibiliza ciertos aspectos y lugares de la ciudad mientras otros permanecen invisibilizados y, por otro, provoca desigualdades territoriales entre las diferentes zonas urbanas, que muestran distintos niveles de interés visual. Para ello me dispuse a realizar, sobre una muestra de publicaciones compartidas en Instagram, un análisis sobre tres grandes áreas o variables: la ubicación o geolocalización, las métricas propias de las redes sociales (ARS o SNA en su equivalente en inglés) y el contenido, textual y visual de dichas publicaciones.

2.1. *El uso de APIs en Investigación Social*

Las interfaces de programación de aplicaciones, o Application Programming Interface suponen una oportunidad muy valiosa para las Ciencias Sociales para poder acceder a la ingente cantidad de datos producidos por aplicaciones como Twitter, Facebook o Instagram y poder así estudiar fenómenos psicológicos, sociales, económicos y políticos de una manera nunca antes vista a través de las redes sociales. Estas “puertas traseras” en los medios sociales de los que se nutren, por ejemplo, las Humanidades Digitales, pero también empresas y creadores de aplicaciones y servicios web que monetizan y rentabilizan el acceso a dichas APIs, se han incorporado muy recientemente al catálogo de métodos disponibles para las investigaciones sociales, en los que la digitalización tan solo había renovado los antiguos, adaptando los métodos clásicos al mundo digital, como, por ejemplo, los cuestionarios *online*, en el caso de las técnicas cuantitativas, o el de las “netnografías” en el de las cualitativas.

Instagram constituye un caso especial: hasta 2016, Instagram proporcionaba acceso a terceros a los datos generados por su aplicación desde su API, por lo que podía usarse para recopilar las imágenes publicadas por sus usuarios, así como sus datos y metadatos. Sin embargo, a partir de ese año, solo aquellas aplicaciones con permiso explícito de Instagram pueden usarla, pues su nueva política de empresa busca limitar aquellas aplicaciones que están generando ingresos gracias a Instagram, además de que buscan asegurarse el “toque humano” y no automatizado que resulta demasiado visible en otras redes sociales como Twitter pobladas de bots (Confessore, 2018). El solicitante de acceso a la API, por tanto, tenía que asegurar que su producto resultante iba a generar ingresos comerciales de algún tipo o que, por lo menos, había la intención de ello. Este

progresivo cierre de la API a accesos públicos sin intereses comerciales se debe a la adaptación al ecosistema Facebook, más cerrado en cuanto a accesos (aunque no parece que ello haya asegurado un mayor control por la compañía de Zuckerberg, dados los escándalos de filtración y venta de datos a terceros). La migración hacia la nueva API, denominada Instagram Graph API, integrada en los servidores de Facebook, planea la “depreciación” de la antigua en tres fases (Facebook, 2018; Instagram, 2018), al término de las cuales Facebook cerrará las capacidades de leer contenido público de la API propia de Instagram para dejar operativo el acceso solo a través de Instagram Graph API.

Para solventar este problema de acceso a la API de Instagram, se puede recurrir a páginas de terceros como Netlytic (Gruzd, 2016), un servicio web mantenido por el Social Media Lab de la Universidad de Ryerson (Toronto, Canadá) que permite a investigadores sociales monitorizar el contenido y la conexión entre usuarios en redes sociales como Twitter, Facebook o Instagram..

2.2. Procedimiento y software utilizado

Dentro de la plataforma Netlytic, puede configurarse la extracción de los datos a partir de un *hashtag* o etiqueta o a partir de unas coordenadas geográficas. Para esta investigación, se llevó a cabo una recolección de datos de Instagram de manera geolocalizada, con centro en las coordenadas 37.180976 latitud, -3.600173 longitud y un radio de 5 kilómetros durante un período de un año, entre abril de 2017 y de 2018. Las coordenadas escogidas corresponden a un punto céntrico situado en la Gran Vía granadina, cercano al mercado por la propia aplicación cuando se escoge la ubicación predeterminada denominada “Granada, España”, pero que se encuentra más próximo al centro histórico, turístico y comercial de la ciudad. Esta elección, aunque subjetiva, trataba de maximizar la recopilación de datos para el casco metropolitano de la ciudad dentro del mencionado radio de 5 kilómetros.

Posteriormente, se procedió al tratamiento y limpieza de datos a través del lenguaje estadístico R (R Development Core Team, 2008) y su IDE (entorno de desarrollo integrado) RStudio, así como aplicaciones de hojas de cálculo como Calc (de LibreOffice) o la alternativa creada por Google.

Para los análisis geográficos se utilizó el *software* QGIS 3.2 (QGIS Development Team, 2018) y para el grueso del análisis y visualización de datos se usó Tableau Desktop, excepto para el análisis del *corpus* textual, para lo que se recurrió, por una parte, al complemento de Excel para Windows denominado Meaning Cloud y, por otra, al *software* para análisis y procesamiento de *big data* Knime.

Para el análisis de redes sociales (SNA) y visualización se ha usado Gephi 0.9.2 (Gephi Consortium, 2018), un *software* también *open source* escrito en Java.

Por último, para el análisis visual, se procedió a la descarga de las imágenes de cada publicación mediante un *script* escrito en Python (Van Rossum, 2007), descargándose un total de 355.797 imágenes, lo que precisó poco más de 66 horas y un espacio digital de 75,53 GB de memoria, que fueron analizadas mediante el *software* ImageJ (Rueden

et al., 2017; Schneider, Rasband, & Eliceiri, 2012), usado en las investigaciones de Lev Manovich y del Laboratorio de Estudios del Software de la Universidad de Nueva York para generar los valores de cada imagen según el modelo de color HSB (Hue, Saturation, Brightness, es decir, Matiz, Saturación y Brillo) creado por Alvy R. Smith (1978), que permiten confeccionar visualizaciones de datos complejas con la macro ImagePlot. Una muestra de 10.000 de estas imágenes se procesó a través de Cloud Vision, un servicio del paquete de productos de la nube de Google que utiliza técnicas de *machine learning* para etiquetar automáticamente las imágenes según el contenido detectado.

La plataforma Netlytic estuvo recogiendo información en conjuntos de 100.000 observaciones, a un ritmo de muestreo aproximado de 100 publicaciones a la hora. El conjunto de datos o *dataset* resultante, que hubo que recopilar en un solo archivo CSV (datos o valores tabulados por comas, o *comma-separated values*) contenía un total de 955.564 filas (560.716 publicaciones y 394.848 comentarios a dichas publicaciones), con las siguientes columnas (variables recogidas): un código identificador único, las coordenadas de geolocalización de la publicación, las URLs de la publicación y la propia imagen (“medialink”), la fecha de publicación, el usuario o autor de la imagen subida, los usuarios que han interactuado con la publicación, el texto usado, el número de “likes” recibidos en la imagen y el nombre del filtro usado para modificar la imagen.

Con él se compilaron todos los *subsets* en un solo *dataset* completo, que fue después dividido en dos conjuntos menores para facilitar el análisis de redes, por un lado, y el de texto, por otro. En el primero se conservaron las columnas con los autores y aquellos usuarios que habían interactuado mediante comentarios con las publicaciones, para crear una lista de aristas compatible con los requerimientos de Gephi, el programa para el análisis de redes sociales que hemos usado. La red resultante total comprendía 344.433 usuarios o nodos, entretejidos en una red compuesta de 394.848 relaciones entre los nodos o aristas. En el segundo subconjunto se mantuvo la columna con el texto de cada publicación para realizar el análisis de contenido textual.

También se creó una columna para categorizar y distinguir aquellas observaciones que eran la publicación principal (“post”) de las que eran comentarios (“comment”), para segmentar cuando fuera necesario el *dataset* completo en base a esta nueva variable creada a tal efecto. Otra transformación simple que hubo que acometer, también para poder trabajar en distintas plataformas y *softwares* para análisis geográficos, fue dividir la columna de coordenadas en dos: latitud y longitud, manteniendo, además, por problemas de conversión entre unidades, el punto como separador decimal para poder garantizar la interoperabilidad y el intercambio de información entre dichas plataformas y *softwares*.

Además, una versión reducida del *dataset* principal (para aligerar su peso) fue utilizada dentro de QGIS para crear nuevas columnas que proporcionaran la información geográfica. A partir de los polígonos delimitadores y la importación de este *dataset* reducido, se procedió a clasificar cada observación, a partir de sus coordenadas geográficas, por distritos administrativos y barrios de la ciudad, con lo que, al *dataset* principal pudo añadirse otras cuatro columnas con el distrito, barrio, código de entidad y municipio (obviamente, el mismo para todos) de cada observación.

Para poder crear dentro de ImageJ, a través de la macro ImagePlot creada por Manovich y su equipo de la Iniciativa de Estudios de Software, las visualizaciones de datos denominadas “imagengramas”, compuestas por imágenes como puntos de datos en torno a ejes lineales y radiales, al estilo de gráficas de puntos tradicionales, se usó otra macro denominada ImageMeasure, incluida en la distribución de ImagePlot, que lo que hace es medir los valores de cada imagen según el modelo de color HSB (Hue, Saturation, Brightness, es decir, Matiz, Saturación y Brillo), en una escala de 0 a 255. Puede usarse el estadístico que se prefiera, pero los creadores recomiendan usar, como medida de posición central estadística para estudiar la distribución, la mediana, por ser menos sensible a los valores extremos de la distribución, que afectan más a medidas como el promedio estadístico. Se midieron, para una muestra total de 132.859 imágenes (un 37,34% de la muestra total), los valores medianos del brillo, la saturación y el matiz o valor (tono cromático), así como sus respectivas desviaciones típicas (medida de dispersión resultante de realizar la raíz cuadrada de las varianzas). Los distintos *datasets* resultantes de estas transformaciones y manipulaciones fueron integrados en una sola base de datos dentro de Tableau Desktop.

Un último *dataset* consecuencia de la manipulación del *dataset* principal fue el obtenido a raíz de tratar una pequeña muestra de 10.000 imágenes por el software de reconocimiento visual Google Cloud Vision, también mediante un script en Python para el acceso al servicio vía API, por el cual el sistema entrenado por Google permite comprender “el contenido de una imagen mediante el encapsulado de potentes modelos de aprendizaje automático”, de tal forma que “clasifica las imágenes rápidamente en miles de categorías, detecta objetos y caras determinados dentro de dichas imágenes y es capaz de leer las palabras impresas que contengan” (Google, 2018). El análisis de imágenes es tan avanzado que es posible entrenar un modelo propio para que sea capaz, por ejemplo, de etiquetar automáticamente aquellas imágenes donde aparezca un objeto concreto (por ejemplo, podríamos entrenarlo para detectar automáticamente imágenes de la Alhambra, el monumento por excelencia de Granada). Sin embargo, esta posibilidad escapaba de los conocimientos y objetivos primarios para esta investigación, pero futuras líneas de ampliación podrían usar esta capacidad. Tras un proceso de 10 horas y media, el servicio de Google fue capaz de detectar un total de 2.332 etiquetas diferentes, que fueron usadas para categorizar las publicaciones en 81.116 ocasiones, lo que nos da un promedio de 8,11 etiquetas para cada una de las 10.000 imágenes analizadas.

Por todo lo expuesto, esta investigación se encuadraría así dentro de la tendencia actual en Ciencia de Datos o *Data Science*, por cuanto se considera éste un campo que usa métodos, procesos, algoritmos y sistemas científicos para extraer conocimiento, patrones y tendencias provenientes de datos de varias formas, tanto estructurados como desestructurados (Dhar, 2013), de manera similar a la minería de datos o *data mining*. La Ciencia de Datos es, más que un conjunto de técnicas provenientes de la informática aplicadas a la Ciencia en general, un “concepto para unificar estadísticas, análisis de datos, aprendizaje automático y sus métodos relacionados”, con el objetivo de “entender y analizar fenómenos reales” a través de los datos (Hayashi, 1998). Dentro de la Ciencia de Datos, el término que se usa para hablar de la centralidad de los datos en la investigación de algún problema o fenómeno social es el de “estudio basado en datos” o

data-driven, que viene a mostrar que las características de los datos recolectados y su estructura son los que guían el desarrollo del proceso investigador.

3. Resultados

Con esta metodología y procedimiento, pudieron extraerse algunas conclusiones destacadas sobre la construcción de la cultura visual en una ciudad como Granada. Instagram, como red social, supone una muestra significativa del universo visual existente, compuesto y construido a través de las distintas miradas, visiones y representaciones sociales de la realidad compartidas públicamente por los usuarios de Instagram en Granada. Sus miradas se han visto atraídas por diversos puntos, visualmente muy potentes, y al centrar la vista en estos puntos, muchos otros han quedado fuera del campo de visión. Esta mirada es netamente social, construida colectivamente por diversas comunidades articuladas en torno a intereses variados, que van desde lo comunicativo hasta lo estéticamente placentero, pasando por el entretenimiento y el exhibicionismo. Hemos usado así a Instagram como una herramienta para observar parte del proceso de desarrollo de la socialización virtual en comunidades de usuarios, que construyen así, a partir de esta socialización secundaria, su propia identidad individual, tan social como visual, ampliando, a su vez, su creatividad audiovisual gracias a las posibilidades estéticas de la aplicación.

El monumento por excelencia de Granada, La Alhambra, que define gran parte de su identidad, no solo visual, sino también cultural, económica y social, se muestra en todo su esplendor en esta investigación, como el punto de la ciudad más fotografiado y compartido en Instagram. Y lo hace, además, de una manera aplastante. La distribución de puntos geolocalizados que ha generado nuestro conjunto de datos muestra también una considerable desigualdad territorial entre los distintos distritos y barrios que conforman la ciudad, en términos de la visualidad. Ya no es que la Alhambra concentre un punto tan alto de las publicaciones etiquetadas en Granada, es que la distancia con respecto al resto de puntos es bastante grande. Y esto se traduce en diferencias considerables entre los distritos y barrios donde se ubica la Alhambra y sus zonas aledañas y el resto de zonas urbanas. Por la distribución espacial de la ciudad de Granada, la distribución de puntos de datos geolocalizados sigue un modelo de centro-periferia, según el cual la mayor concentración de puntos se produce en los distritos céntricos y la menor, en los periféricos.

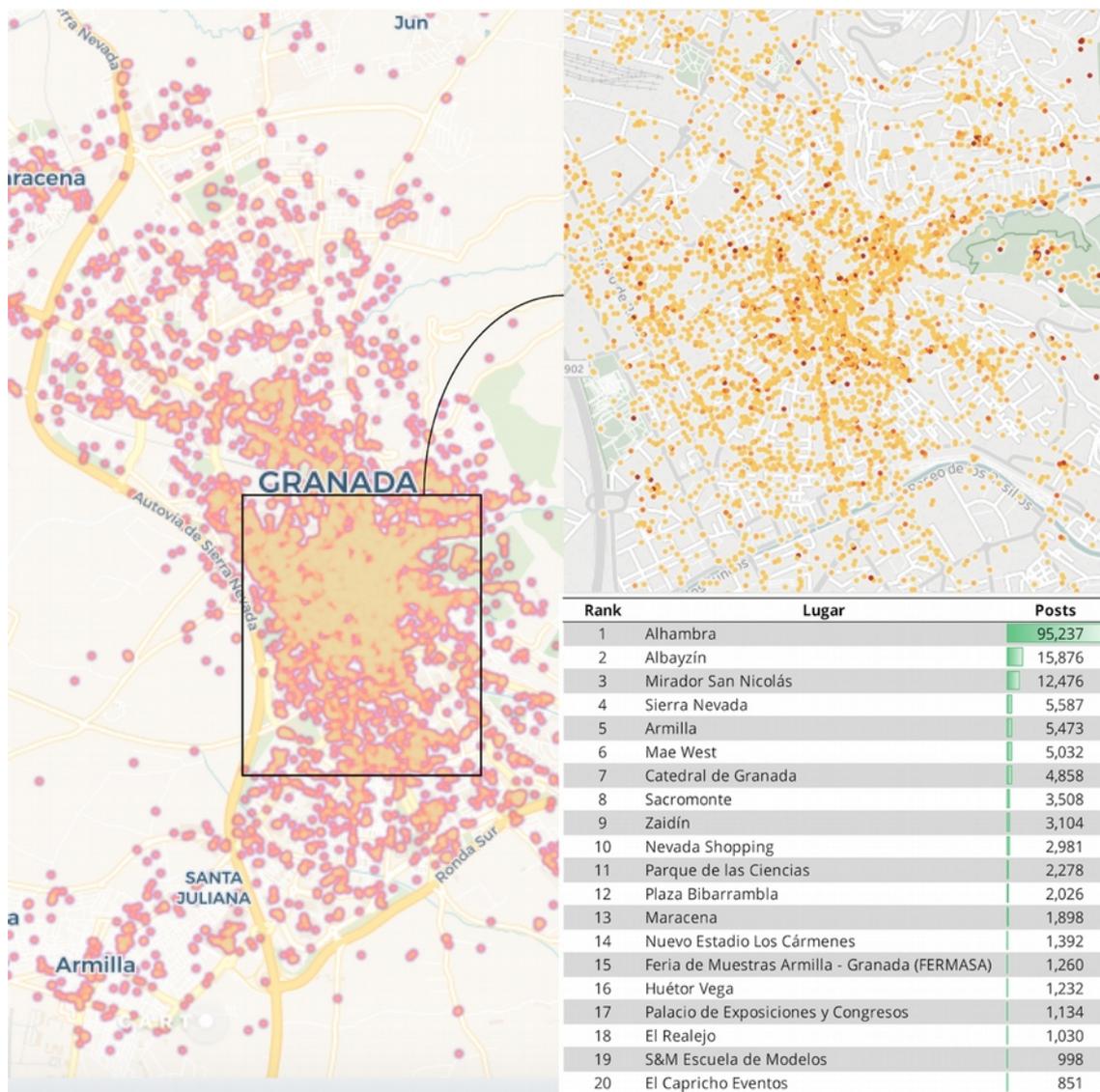


Gráfico 1.- Principales puntos visuales de Granada en número de publicaciones en Instagram. Mapa de calor general, ampliado en la zona centro, y ranking reducido con los 20 puntos principales. Fuente: Elaboración propia.

Por otro lado, las métricas procedentes del Análisis de Redes Sociales (o SNA) y extraídas de la red de usuarios de Instagram permitieron mostrar una red social muy dispersa, extensa y nada densa, con usuarios muy separados entre sí; una red, por tanto, poco homogénea y muy diversa. Es la principal conclusión que puede sacarse de los altos valores en distancia media y diámetro y los bajos coeficientes de clustering, que hablaban de una red que, aunque cumple la ley de potencia (muchos nodos con pocos grados y pocos nodos con muchos grados, es decir, cierta concentración de la desigualdad social), no puede considerarse como de “mundo pequeño”, en el que es fácil conocer a otro usuario mediante la conexión con unos pocos usuarios más. Más bien al contrario, hablamos de comunidades separadas y poco conectadas entre sí, con solo unos pocos nodos haciendo de intermediarios o “puentes” entre estas comunidades. Estas comunidades fueron detectadas mediante un análisis de modularidad, del que surgieron 46.597 comunidades, del que se seleccionaron aquellas 7 con mayor número

de nodos para analizarlas con mayor profundidad y caracterizarlas así en base a la observación cualitativa de sus miembros.

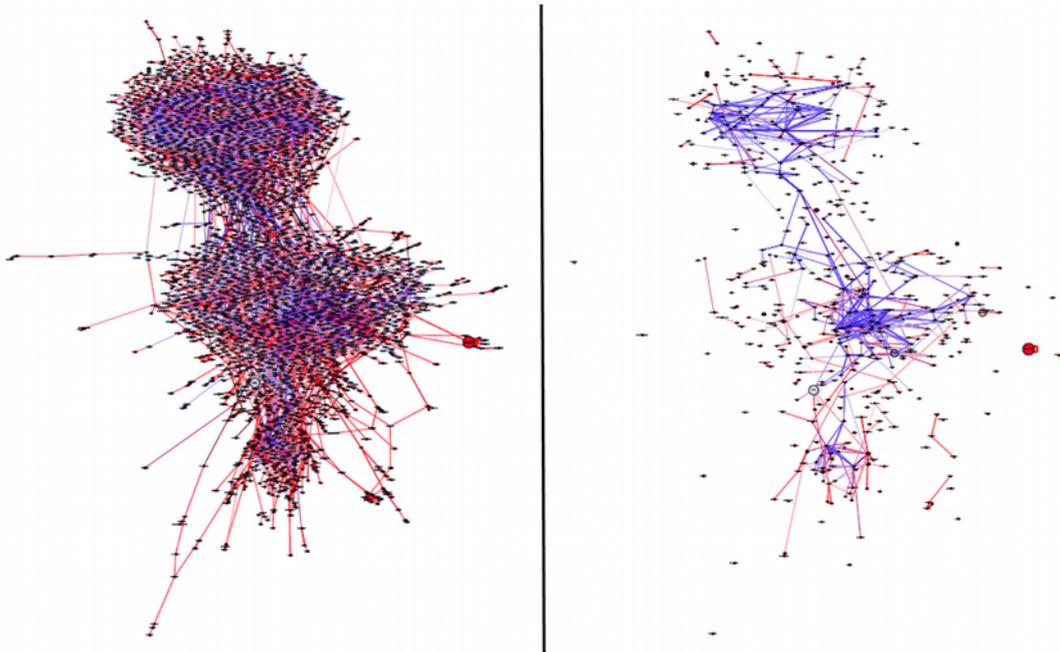


Gráfico 2.- Detalle central del grafo de la red de usuarios de Instagram de Granada. A la izquierda, filtrados por grado > 19 ; a la derecha, por grado > 70 ; ordenados mediante algoritmo ForceAtlas 2. El color representa el valor de intermediación (de azul, más alto, a rojo, más bajo). El tamaño, el valor de centralidad de vector propio. Fuente: elaboración propia.

Por último, tras el análisis geográfico y de redes sociales, pude centrarme en el análisis de contenidos, bajo la premisa de que el contenido generado en la plataforma Instagram posibilita un análisis de tipo visual, sobre la imagen compartida, y textual, sobre el texto que acompaña a la imagen y sus respectivos comentarios. Así, pudieron realizarse, tras una limpieza del corpus textual, análisis de palabras y etiquetas más utilizadas (temas de interés) o de idiomas empleados. También los tipos de filtros más usados en las imágenes o las publicaciones que habían generado mayor número de likes entre los usuarios.

Pero quizás lo más interesante pudo extraerse tras pasar una muestra de las imágenes obtenidas en esta investigación por el algoritmo de detección visual Google Cloud Vision. Con él se obtuvieron etiquetas útiles para determinar el contenido de las imágenes. Dicho análisis permitió descubrir cuáles fueron los contenidos más compartidos en Instagram en Granada: las imágenes con elementos arquitectónicos y urbanísticos mostraron un mayor peso que las imágenes con elementos humanos, ya sean grupales o individuales (selfi). Por tanto, el contenido compartido hablaba de la importancia en Granada de la fotografía paisajística y arquitectónica por encima de la personal.

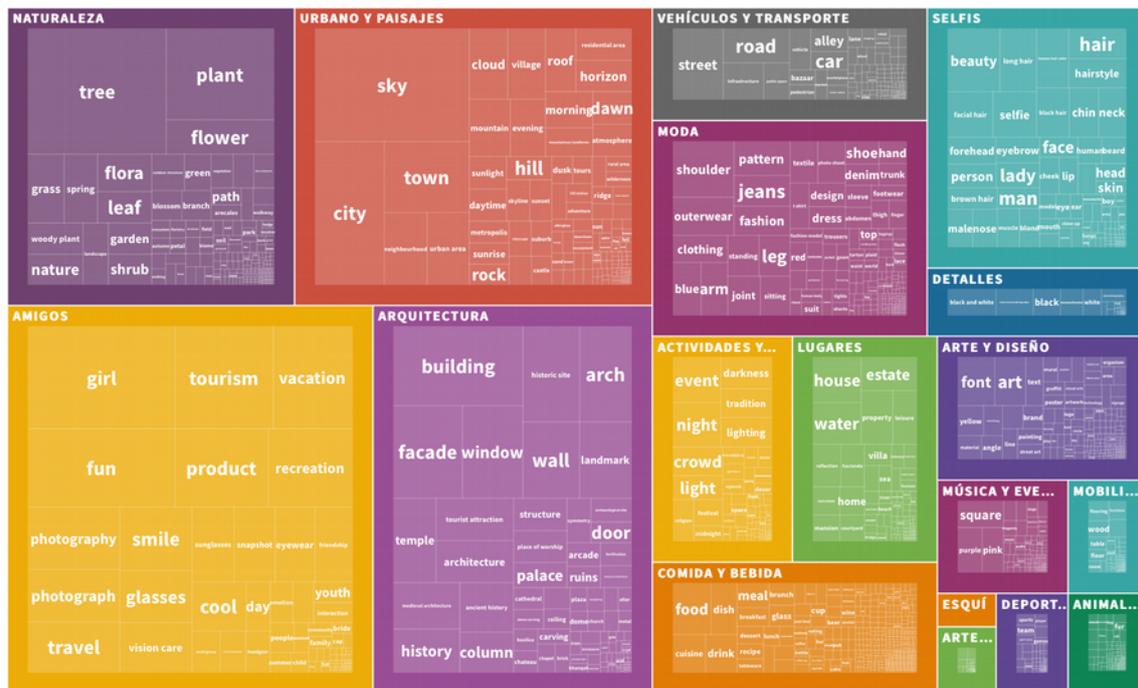


Gráfico 3.- Treemap del contenido de las imágenes compartidas en Instagram en Granada, por categorías principales de Cantón. Fuente: elaboración propia. Disponible versión interactiva en <https://public.flourish.studio/visualisation/142141/>

Por último, pudimos obtener, mediante el software creado a tal efecto por el equipo de trabajo de Lev Manovich, los datos de colorimetría según modelo HSV para las imágenes descargadas. Esta colorimetría aplicada a las imágenes se relacionaba directamente con los contenidos más compartidos, pues los colores más repetidos se asociaban con las etiquetas de contenido más frecuentes, mostrando también cómo la distribución de colores se concentraba en dos tonalidades predominantes, los tonos anaranjados y los azules. El predominio de las tonalidades azules se debía a la preponderancia en contenido de cielo, es decir, de aquellas fotos paisajísticas y arquitectónicas donde aparece el cielo azulado. Por otro lado, el tono anaranjado podía deberse a dos motivos: el color mezcla de dorado y rosáceo de los rostros que aparecen en las imágenes, el otro debido a los tonos tierra de las paredes y techos del propio monumento de la Alhambra. También destacaban las tonalidades verdosas, que suelen pertenecer a imágenes realizadas a alguna de las partes más atractivas de los jardines de la Alhambra, y las tendentes al negro, con menores valores de brillo, bien porque son la clásica estampa nocturna de la Alhambra iluminada de noche, bien porque son fotos realizadas a contraluz para remarcar la forma de los típicos arcos túmidos del monumento.

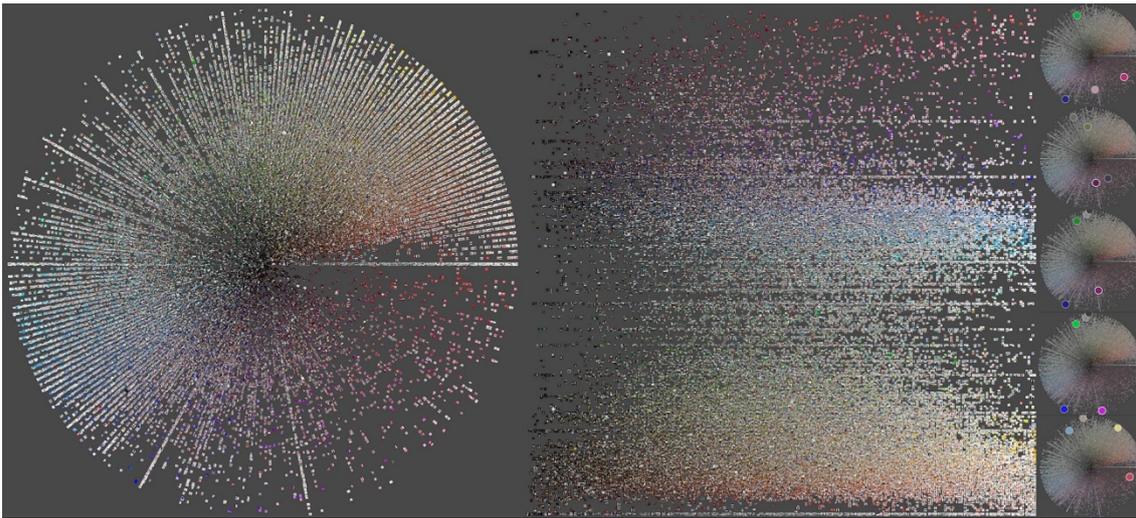


Gráfico 4.- Imagenes de Granada, radial (a) y lineal (b), realizadas a partir de 100.000 imágenes por sus valores de brillo y tono (disponible versión en alta resolución en: http://bit.ly/ig_grx_01) y extracción de su paleta cromática (c, d, e, f y g) en variaciones de colores plenos, colores claros, apagados, intensos y oscuros. Fuente: elaboración propia.

4. Discusión o Conclusiones

Resumiendo, a través de la plataforma Netlytic (Gruzd, 2016) se extrajeron casi un millón de publicaciones de manera geolocalizada durante un período de un año, entre abril de 2017 y de 2018, que se analizaron desde tres vertientes: análisis geográfico, análisis de redes y análisis de contenido, textual y visual. El tratamiento y limpieza de datos se realizó a través del lenguaje estadístico R (R Development Core Team, 2008) y diversas hojas de cálculo. Para los análisis geográficos se utilizó el software QGIS 3.2 (QGIS Development Team, 2018) y para el grueso del análisis y visualización de datos, Tableau Desktop. También se usaron las soluciones de big data Knime y Meaning Cloud para el análisis de texto. Para el análisis y visualización de redes sociales (SNA), Gephi 0.9.2 (Gephi Consortium, 2018), un software también *open source* escrito en Java. Y, por último, ImageJ (Rueden et al., 2017; Schneider et al., 2012) e ImagePlot para generar los imagenogramas resultantes. Además, una muestra de 10.000 imágenes se procesaron a través de Cloud Vision, un servicio del paquete de productos de la nube de Google que utiliza técnicas de *machine learning* para etiquetar automáticamente el contenido detectado dentro de las imágenes.

El paulatino cierre de las APIs de plataformas privativas de redes sociales, como Instagram, está provocando una dificultad creciente para la extracción de datos y su aprovechamiento con fines científicos, por lo que esta investigación ha supuesto dos aspectos que considero de relevancia: primero, que todas las técnicas y procedimientos utilizados, y mostrados aquí, suponen tan solo un pequeño muestrario de las verdaderas posibilidades de las Humanidades Digitales aplicadas y el software empleado como herramienta de investigación social a través de los datos que generamos mediante plataformas sociales; y, segundo, y no por ello menos importante, que lo que se ha mostrado aquí supone un testimonio casi arqueológico de un tiempo en el que las APIs

eran lo suficientemente accesibles para el investigador como para poder realizar análisis de datos a este nivel.

Este progresivo cierre está en total consonancia con la actual tendencia a la mercantilización de los datos generados por los usuarios para las plataformas: si los datos generados son el petróleo del siglo XXI, las empresas que los poseen, base de su generación de valor como negocios, no están dispuestas a compartírselos con otros fines que no sean los puramente comerciales. Algo que demuestra el actual proceso de petición de permisos para usar este tipo de datos: la solicitud de permisos a través de la sección de *developers* de Facebook es ya tan restringida que al investigador le merece mucho más la pena aprender a usar métodos simples de *scraping web* (técnicas utilizadas para extraer información de sitios web de manera automatizada simulando la navegación humana en la Web) para extraer los datos que necesite para su investigación social, lo que implica posicionamientos éticos que considero no entran en conflicto con el carácter público de los datos expuestos así públicamente para cualquier otro usuario de la plataforma. Las exigencias de transparencia que hacemos a las instituciones públicas gubernamentales deberían ser extensibles también a este tipo de empresas. Al fin y al cabo, la propia sociedad es la suministradora de estos datos y, al igual que ocurre con la investigación científica, buena parte de los resultados debería revertir en ella en forma de beneficios sociales, permitiendo, por ejemplo, que parte de estos datos que suponen una alta rentabilidad económica tengan, a su vez, una rentabilidad también social. Como sociedad, podemos no estar en contra de que Google haga dinero gracias a los datos que generamos, con unas prácticas éticas acordes con nuestros estándares legales, pero, además, deberíamos luchar por que el valor que Google está generando pueda servir para mejorar nuestra sociedad, en términos de eficiencia u optimización de recursos.

5. Referencias bibliográficas

- Boy, J. D., & Uitermark, J. (2016). How to Study the City on Instagram. *PLoS ONE*, 11(6). doi: 10.1371/journal.pone.0158161
- Confessore, N. (2018, enero 27). The Follower Factory. *The New York Times*. Recuperado de <https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html>
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64-73. doi: 10.1145/2500499
- Facebook. (2018). Instagram Graph API Launches and Instagram API Platform Deprecation. Recuperado de <https://developers.facebook.com/blog/post/2018/01/30/instagram-graph-api-updates>
- Gephi Consortium. (2018). *Gephi*. Recuperado de <https://gephi.org>
- Google. (2018). Análisis de contenido de imágenes con API Vision | Cloud Vision API. Recuperado de <https://cloud.google.com/vision/?hl=es>
- Gruzd, A. (2016). *Netlytic: Software for automated text and social network analysis*.

Recuperado de <https://netlytic.org>

- Hand, M. (2014a). Digitization and memory: Researching practices of adaption to visual and textual data in everyday life. En *Big Data? Qualitative Approaches to Digital Research* (pp. 205–227). Recuperado de <http://www.emeraldinsight.com/doi/pdf/10.1108/S1042-319220140000013013>
- Hand, M. (2014b). From cyberspace to the dataverse: Trajectories in digital social research. En *Big Data? Qualitative Approaches to Digital Research* (pp. 1–27). Recuperado de <http://www.emeraldinsight.com/doi/full/10.1108/S1042-319220140000013002>
- Hand, M. (2017). Visuality in Social Media: Researching images, circulations and practices. En *The SAGE Handbook of Social Media Research Methods*. SAGE.
- Hayashi, C. (1998). What is Data Science? Fundamental Concepts and a Heuristic Example. En C. Hayashi, K. Yajima, H.-H. Bock, N. Ohsumi, Y. Tanaka, & Y. Baba (Eds.), *Data Science, Classification, and Related Methods* (pp. 40-51). doi: 10.1007/978-4-431-65950-1_3
- Instagram. (2018). Instagram Developer Documentation. Recuperado de <https://www.instagram.com/developer>
- Manjoo, F. (2018, febrero 9). Welcome to the Post-Text Future. *The New York Times*. Recuperado de <https://www.nytimes.com/interactive/2018/02/09/technology/the-rise-of-a-visual-internet.html>
- Manovich, L. (2009). Cultural analytics: Visualising cultural patterns in the era of “more media”. *Domus March*.
- Manovich, L. (2012). How to compare one million images? En *Understanding digital humanities* (pp. 249–278). Recuperado de https://link.springer.com/chapter/10.1057/9780230371934_14
- Manovich, L. (2017). Cultural analytics, social computing and digital humanities. *The datafied society: studying culture through data*, 55. Recuperado de <http://www.oapen.org/download?type=document&docid=624771#page=56>
- Manovich, L., Goddemeyer, D., Stefaner, M., & Baur, D. (2015). On Broadway. Recuperado de <http://on-broadway.nyc/>
- Manovich, L., & Indaco, A. (2017). The Image of a Data City: Studying the Hyperlocal with Social Media. *Architectural Design*, 87(1), 110-117. doi: 10.1002/ad.2140
- Manovich, L., Stefaner, M., Yazdani, M., Baur, D., Goddemeyer, D., & Tifentale, A. (2014). Selfiecity. Recuperado de <http://selfiecity.net>
- Martín Prada, J. (2018). *El ver y las imágenes en el tiempo de Internet*.
- Mirzoeff, N. (2016). *Cómo ver el mundo: Una nueva introducción a la cultura visual*. Grupo Planeta Spain.
- QGIS Development Team. (2018). *QGIS Geographic Information System. Open Source Geospatial Foundation Project*. Recuperado de <https://qgis.org>
- R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Recuperado de <http://www.r-project.org>

- Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials*. Sage.
- Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1). doi: 10.1186/s12859-017-1934-z
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671-675.
- Smith, A. R. (1978). Color gamut transform pairs. *ACM Siggraph Computer Graphics*, 12(3), 12–19.
- Van Rossum, G. (2007). Python Programming Language. *USENIX Annual Technical Conference*, 41, 36.

Financiación:

Esta investigación forma parte del proyecto *Transmedialización y crowdsourcing en las narrativas de ficción y no ficción audiovisuales, periodísticas, dramáticas y literarias* (CSO2017-85965-P), financiado por el Programa Estatal de Fomento de la Investigación, Científica y Técnica de Excelencia, Subprograma Estatal de Generación de Conocimiento, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016-Convocatoria 2017.

HOW TO CITE (APA 6^a)

Cantón Correa, J. (2019). Cómo estudiar la construcción de la imagen de una ciudad a través de publicaciones de Instagram: una metodología aplicada a Granada. *Comunicación y Métodos | Communication & Methods*, 1(2), 7-20. doi: 10.35951/v1i2.22